Title: TOMATO SPOTTED WILT VIRUS

Abstract

The nucleotide sequences for tomato spotted wilt virus (TSWV) nucleocapsid is described, and transgenic plants containing the nucleocapsid nucleotide sequence from a TSMV isolate is shown to provide resistance in the transgenic plant to Toospoviruses from different serogroups. In addition, transgenic plants containing the nucleocapsid nucleotide sequence from a lettuce isolate of TSWV were produced and shown to provide (in plants producing small amounts of the nucleocapsid protein) resistance in the transgenic plant to both homologous and closely related viral isolates whereas plants producing larger amounts of the nucleocapsid protein possessed moderate levels of protection against both the homologous isolate and isolates of distantly related Impatiens necrotic spot virus (INSV).
FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

<table>
<thead>
<tr>
<th>Code</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>Austria</td>
</tr>
<tr>
<td>AU</td>
<td>Australia</td>
</tr>
<tr>
<td>BB</td>
<td>Barbados</td>
</tr>
<tr>
<td>BE</td>
<td>Belgium</td>
</tr>
<tr>
<td>BF</td>
<td>Burkina Faso</td>
</tr>
<tr>
<td>BG</td>
<td>Bulgaria</td>
</tr>
<tr>
<td>BJ</td>
<td>Benin</td>
</tr>
<tr>
<td>BR</td>
<td>Brazil</td>
</tr>
<tr>
<td>BY</td>
<td>Belarus</td>
</tr>
<tr>
<td>CA</td>
<td>Canada</td>
</tr>
<tr>
<td>CF</td>
<td>Central African Republic</td>
</tr>
<tr>
<td>CG</td>
<td>Congo</td>
</tr>
<tr>
<td>CH</td>
<td>Switzerland</td>
</tr>
<tr>
<td>CI</td>
<td>Côte d'Ivoire</td>
</tr>
<tr>
<td>CM</td>
<td>Cameroon</td>
</tr>
<tr>
<td>CN</td>
<td>China</td>
</tr>
<tr>
<td>CS</td>
<td>Czechoslovakia</td>
</tr>
<tr>
<td>CZ</td>
<td>Czech Republic</td>
</tr>
<tr>
<td>DE</td>
<td>Germany</td>
</tr>
<tr>
<td>DK</td>
<td>Denmark</td>
</tr>
<tr>
<td>ES</td>
<td>Spain</td>
</tr>
<tr>
<td>FI</td>
<td>Finland</td>
</tr>
<tr>
<td>FR</td>
<td>France</td>
</tr>
<tr>
<td>GA</td>
<td>Gabon</td>
</tr>
<tr>
<td>GB</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>GE</td>
<td>Georgia</td>
</tr>
<tr>
<td>GN</td>
<td>Guinea</td>
</tr>
<tr>
<td>GR</td>
<td>Greece</td>
</tr>
<tr>
<td>HU</td>
<td>Hungary</td>
</tr>
<tr>
<td>IE</td>
<td>Ireland</td>
</tr>
<tr>
<td>IT</td>
<td>Italy</td>
</tr>
<tr>
<td>JP</td>
<td>Japan</td>
</tr>
<tr>
<td>KE</td>
<td>Kenya</td>
</tr>
<tr>
<td>KG</td>
<td>Kyrgyzstan</td>
</tr>
<tr>
<td>KP</td>
<td>Democratic People’s Republic of Korea</td>
</tr>
<tr>
<td>KR</td>
<td>Republic of Korea</td>
</tr>
<tr>
<td>KZ</td>
<td>Kazakhstan</td>
</tr>
<tr>
<td>LI</td>
<td>Liechtenstein</td>
</tr>
<tr>
<td>LK</td>
<td>Sri Lanka</td>
</tr>
<tr>
<td>LU</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>LV</td>
<td>Latvia</td>
</tr>
<tr>
<td>MC</td>
<td>Monaco</td>
</tr>
<tr>
<td>MD</td>
<td>Republic of Moldova</td>
</tr>
<tr>
<td>MG</td>
<td>Madagascar</td>
</tr>
<tr>
<td>ML</td>
<td>Mali</td>
</tr>
<tr>
<td>MN</td>
<td>Mongolia</td>
</tr>
<tr>
<td>MR</td>
<td>Mauritania</td>
</tr>
<tr>
<td>MW</td>
<td>Malawi</td>
</tr>
<tr>
<td>NE</td>
<td>Niger</td>
</tr>
<tr>
<td>NL</td>
<td>Netherlands</td>
</tr>
<tr>
<td>NO</td>
<td>Norway</td>
</tr>
<tr>
<td>NZ</td>
<td>New Zealand</td>
</tr>
<tr>
<td>PL</td>
<td>Poland</td>
</tr>
<tr>
<td>PT</td>
<td>Portugal</td>
</tr>
<tr>
<td>RO</td>
<td>Romania</td>
</tr>
<tr>
<td>RU</td>
<td>Russian Federation</td>
</tr>
<tr>
<td>SD</td>
<td>Sudan</td>
</tr>
<tr>
<td>SE</td>
<td>Sweden</td>
</tr>
<tr>
<td>SI</td>
<td>Slovenia</td>
</tr>
<tr>
<td>SK</td>
<td>Slovakia</td>
</tr>
<tr>
<td>SN</td>
<td>Senegal</td>
</tr>
<tr>
<td>TD</td>
<td>Chad</td>
</tr>
<tr>
<td>TG</td>
<td>Togo</td>
</tr>
<tr>
<td>TJ</td>
<td>Tajikistan</td>
</tr>
<tr>
<td>TT</td>
<td>Trinidad and Tobago</td>
</tr>
<tr>
<td>UA</td>
<td>Ukraine</td>
</tr>
<tr>
<td>US</td>
<td>United States of America</td>
</tr>
<tr>
<td>UZ</td>
<td>Uzbekistan</td>
</tr>
<tr>
<td>VN</td>
<td>Viet Nam</td>
</tr>
</tbody>
</table>
TOMATO SPOTTED WILT VIRUS

Viruses in the *Tospovirus* genus infect a wide variety of plant species, particularly tobacco, peanut, vegetables and ornamental plants.

Two virus species, tomato spotted wilt virus (TSWV) and impatiens necrotic spot virus (INSV) are recognized within the Tospovirus genus.

Tomato Spotted Wilt Virus (TSWV) is unique among plant viruses in that the nucleic acid-protein complex is covered by a lipoprotein envelope and it is the only thrip transmitted virus. This virus has recently been classified as the Tospovirus genus of the *Bunyaviridae* family. TSWV virions contain a 29K nucleocapsid protein ("NP" or "N"), two membrane-associated glycoproteins (58K and 78K) and a large 200K protein presumably for the viral transcriptase [see J. Gen. Virol. 71:2207 (1991); Virol. 56:12 (1973); and J. Gen. Virol. 36:267 (1977)].

The virus genome consists of three negative-strand (-) RNAs designated L RNA (8900 nucleotides), M RNA (5400 nucleotides) and S RNA (2900 nucleotides) [see J. Gen. Virol. 36:81 (1977); J. Gen. Virol. 53:12 (1981); and J. Gen. Virol. 70:3469 (1989)], each of which is encapsulated by the NP. The partial or full-length sequences of S RNAs from three TSWV isolates reveals the presence of two open reading frames (ORF) with an ambisense gene arrangement [see J. Gen Virol. 71:1 (1990) and J. Gen. Virol. 72:461 (1991)]. The larger open reading frame is located on the viral RNA strand and has the capacity to encode a 52K nonstructural protein. The smaller ORF is located on the viral complementary RNA strand and is translated through a subgenomic RNA into the 29K NP.

The ambisense coding strategy is also characteristic of the TSWV M RNA, with the open reading frames encoding the 58K and 78K membrane-associated glycoproteins. The TSWV L RNA has been sequenced to encode a large 200K protein presumably for the viral transcriptase.

Two TSWV serogroups, "L" and "I", have been identified and characterized based on serological analysis of the structural proteins and morphology of cytopathic structures [see J. Gen Virol. 71:933 (1990) and Phytopathology 81:525 (1991)]. They have serologically conserved G1 and G2 glycoproteins, but the NP of the "I" serogroup is
serologically distinct from that of the "L" serogroup. Comparison of the NP between the "L" and "I" serogroups has shown 62% and 67% identities at nucleotide and amino acid levels, respectively [see J. Gen. Virol. 72:2597 (1991)].

TSWV has a wide host range, infecting more than 360 plant species of 50 families and causes significant economic losses to vegetables and ornamental plants worldwide. The "L" serogroup has been found extensively in field crops such as vegetables and weeds, while the "I" serogroup has been largely confined to ornamental crops.

A cucurbit isolate has recently been identified [see Plant Disease 68:1006 (1984)] as a distinct isolate because it systemically infects watermelon and other cucurbits and its NP is serologically unrelated to that of either serogroup. Although the spread of the TSWV disease can sometimes be reduced by breeding resistant plants or using non-genetic approaches, complete control of the disease by these conventional methods has generally proven to be difficult [see Plant Disease 73:375 (1989)].

Since 1986, numerous reports have shown that transgenic plants with the coat protein (CP) gene of a virus are often resistant to infection by that virus. This phenomenon is commonly referred to as coat protein-mediated protection (CPMP). The degree of protection ranges from delay in symptom expression to the absence of disease symptoms and virus accumulation. Two recent independent reports [see Biol. Technology 9:1363(1991) and Mol. Plant-Microbe Interact. 25:534 (1992)] showed that transgenic tobacco plants expressing the nucleocapsid protein (NP) gene of TSWV are resistant to infection by the homologous isolate. However, since TSWV is widespread with many biologically diverse isolates, it is very important to test the effectiveness of the transgenic plants to resist infections by different TSWV isolates. The findings of the present invention expand on those of the previous reports by demonstrating that transgenic plants according to the present invention showed resistance to two heterologous isolates of the "L" serogroup and an isolate of the "I" serogroup. We also show that resistance to the two heterologous isolates of the "L" serogroup was mainly found in plants accumulating very low, if any,
levels of NP, while transgenic plants that accumulated high levels of NP were resistant to the isolate of the "I" serogroup.

However, no resistance was observed to a Brazilian isolate, although the plants that accumulated high levels of the N protein did display a delay in symptom expression. This Brazilian isolate, designated TSWV-B has the N protein that was serologically distinct from the "L" and "I" serogroups and biologically differs from a curcurbit isolate in that the TSWV-B does not systemically infect melons or squash. Therefore, one aspect of the present invention is to characterize the TSWV-B by cloning and sequencing of its S RNA and comparisons with the published sequences of other TSWV isolates.

Various aspects of the present invention will become readily apparent from the detailed description of the present invention including the following example, figures and data.

In the Figures;

Fig. 1 depicts the strategy for cloning the NP gene from viral RNA according to the present invention;

Fig. 2 depicts the in vivo transient expression of the nucleocapsid protein (NP) gene of tomato spotted wilt virus according to the present invention in tobacco protoplasts;

Fig. 3 depicts the location of the sequenced cDNA clones in the TSWV-B S RNA according to the present invention;

Fig. 4 depicts a dendogram showing relationships among TSWV isolates according to the present invention;

Fig. 5 depicts the serological relationship of TSWV isolates described herein;

Fig. 6 depicts the correlation of the level of nucleocapsid protein (NP) accumulation in transgenic plants with the degree of resistance to TSWV isolates;

Fig. 7 depicts the TSWV-BL N coding sequences introduced into transgenic plants in accordance with one aspect of the present invention; and

Fig. 8 depicts the TSWV-BL half N gene fragments introduced into plants in accordance with one aspect of the present invention.
More specifically, figure 2 depicts transient expression of the NP gene in which the constructs were transferred into tobacco mesophyll protoplasts using polyethylene glycol (PEG). The transformed protoplasts were subsequently incubated for two days for the expression of the NP gene. Proteins were extracted from the protoplasts and tested for the NP by double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA) using antibodies against the TSWV NP. NP⁻ and NP⁺ represent the protoplasts transformed with plasmids pBI525-NP⁻ and pBI525-NP⁺, respectively. Concentration of the antibodies for coating: 5 μg/ml; dilution of the enzyme conjugate: 1:250. Data were taken 30, 60 and 90 min. after addition of substrate.

In figure 3, the five overlapping cDNA clones are shown to scale below a S RNA map of TSWV-B. These clones were synthesized with random primers from double-stranded RNA isolated from *N. benthamiana* plants infected with TSWV-B.

In figure 4, the sequences were compared using the pileup program of the GCG Sequence analysis software package. Horizontal lines are proportional to the genetic distance while vertical lines are of arbitrary length and have no significance.

More specifically, in figure 5, *N. benthamiana* Domin were infected with TSWV isolates [TSWV-BL (a lettuce isolate), Arkansas, 10W pakchoy (TSWV-10W), Begonia, and Brazil (TSWV-B)]. An infected leaf disc (0.05 gram) was ground in 12 ml of the enzyme conjugate buffer and analyzed by DAS-ELISA using antibodies raised against TSWV-BL viron (BL viron), or the NP of TSWV-BL (BL-NP), or TSWV-I (I-NP). Concentration of antibodies for coating were 1μg/ml; dilution of conjugates were 1:2000 for BL viron, 1:250 for BL-NP, and 1:1000 for I-NP. The results were taken after 10 minutes (BL), 50 minutes (BL-NP), or 30 minutes after adding substrate.

With regard to figure 6, transgenic plants were assayed in DAS-ELISA for NP accumulation with antibodies raised against the NP of TSWV-BL. Plants were read 150 min. after adding substrate and the transgenic plants were grouped into four categories: OD₄₀₅nm smaller than 0.050, OD₄₀₅nm between 0.050 to 0.200, OD₄₀₅nm between 0.200 to 0.400, and OD₄₀₅nm greater than 0.400. The OD₄₀₅nm readings of...
control NP (-) plants were from zero to 0.05. The same plants were challenged with either the Arkansas (Ark) and 10W pakchoy (10W) isolates or the Begonia isolate and the susceptibility of each plant was recorded about 12 days after inoculation. The results were pooled from fifty-one R1 NP (+) plants inoculated with the Arkansas and 10W pakchoy isolates and one hundred thirty-nine R1 NP (+) plants inoculated with the Begonia isolate. Numbers above bars represent total numbers of R1 NP (+) plants tested.

EXAMPLE I

Isolation of TSWV-BL RNAs:
The TSWV-BL isolate was purified from Datura stramonium L. as follows: the infected tissues were ground in a Waring Blender for 45 sec with three volumes of a buffer (0.033 M KH2PO4, 0.067 MK2HPO4, 0.01 M Na2SO3). The homogenate was filtered through 4 layers of cheesecloth moistened with the above buffer and centrifuged at 7,000 rpm for 15 min. The pellet was resuspended in an amount of 0.01 M Na2SO3 equal to the original weight of tissue and centrifuged again at 8,000 rpm for 15 min. After the supernatant was resuspended in an amount of 0.01 M Na2SO3 equal to 1/10 of the original tissue weight.

The virus extract was centrifuged at 9,000 rpm for 15 min. and the supernatant was carefully loaded on a 10-40% sucrose step gradient made up in 0.01 M Na2SO3. After centrifugation at 23,000 rpm for 35 min., the virus zone (about 3 cm below meniscus) was collected and diluted with two volumes of 0.01 M Na2SO3. The semi-purified virus was pelleted at 27,000 rpm for 55 min.

EXAMPLE II

Purification of TSWV and viral RNAs:
The TSWV-BL isolate [see Plant Disease 74:154 (1990)] was purified from Datura stramonium L, as described in Example I. The purified virus was resuspended in a solution of 0.04% of bentonite, 10 µg/ml of proteinase K, 0.1 M ammonium carbonate, 0.1% (w/v) of sodium diethylthiocarbamate, 1 mM EDTA, and 1% (w/v) of sodium dodecyl sulfate (SDS), incubated at 65°C for 5 min., and immediately extracted from H2O-saturated phenol, followed by another extraction.
with chloroform/isoamyl alcohol (24:1). Viral RNAs were precipitated in 2.5 volumes of ethanol and dissolved in distilled H₂O.

**EXAMPLE III**

**cDNA and PCR-based NP gene cloning:**

The first strand cDNA was synthesized from purified TSWV-BL RNAs using random primers as described by Gubler and Hoffman [see Gene 25:263 (1983)]. The second strand was produced by treatment of the sample with RNase H/DNA polymerase. The resulting double-stranded cDNA sample was size-fractionated by sucrose gradient centrifugation, methylated by EcoRI methylase, and EcoRI linkers were added. After digestion with EcoRI, the cDNA sample was ligated into the EcoRI site of pUC18, whose 5'-terminal phosphate groups were removed by treatment with calf intestinal alkaline phosphatase. *E. coli* DH5α competent cells (Bethesda Research Laboratories) were transformed and clones containing TSWV cDNA inserts were first selected by plating on agar plates containing 50 μg/ml of ampicillin, IPTG, and X-gal. Plasmid DNAs from selected clones were isolated using an alkaline lysis procedure [see BRL Focus 11:7 (1989)], and the insert sizes were determined by EcoRI restriction enzyme digestion followed by DNA transfer onto GeneScreen Plus nylon filters (DuPont). Plasmid clones that contained a TSWV-BL S RNA cDNA insert were identified as described below by hybridizing against a 32P-labelled oligomer (AGCAGGCCAAAACCTGCAAGACTTGC) complementary to the nucleotide sequence (GCAAGTTCTGCGAGTTTTGCTGCT) of the TSWV-CPNH1 S RNA [see J. Gen. Virol. 71:001 (1990)]. Several clones were identified and analyzed on agarose gels to determine the insert sizes. The clones pTSWVS-23 was found to contain the largest cDNA insert, about 1.7 kb in length.

The full-length NP gene was obtained by the use of polymerase chain reaction (PCR). First-strand cDNA synthesis was carried out at 37°C for 30 min. in a 20 μl reaction mixture using oligomer primer JLS90-46 (5'-> 3') AGCTAACCATGGTAAAGCTCAAGGAAAAGC (also used to synthesize the nucleocapsid gene of TSWV-10W) which is complementary to the S RNA in the 5' terminus of TSWV NP gene (nucleotide positions 2751 to 2773 of the TSWV-CPNH1). The reaction
mixture contained 1.5 μg of viral RNAs, 1 μg of the oligomer primer, 0.2 mM of each dNTP, 1X PCR buffer (the GeneAmp kit, Perkin-Elmer-Cetus), 20U of RNAs in Ribonuclease inhibitor (Promega), 2.5 mM of MgCl₂, and 25U of AMV reverse transcriptase (Promega Corporation). The reaction was terminated by heating at 95°C for 5 min. and cooled on ice. Then 10 μl of the cDNA/RNA hybrid was used to PCR-amplify the NP gene according to manufacturer's instructions (Perkin-Elmer-Cetus) using 1 μg each of oligomer primers JLS90-46 and JLS90-47 (5'→3'), AGCATTCATGGTTAACACACTAAGCAAGCAC (also used to synthesize the nucleotide gene of TSWV-10W), the latter oligomer being identical to the S RNA in the 3' noncoding region of the gene (nucleotide positions 1919 to 1938 of the TSWV-CPNH1). A typical PCR cycle was 1 min. at 92°C (denaturing), 1 min. at 50°C (annealing), and 2 min. at 72°C (polymerizing). The sample was directly loaded and separated on a 1.2% agarose gel. The separated NP gene fragment was extracted from the agarose gel, ethanol-precipitated and dissolved in 20 μl of distilled H₂O.

**EXAMPLE IV**

**Construction of plant expression and transformation vectors.**

The gel-isolated NP gene fragment from Example III was digested with the restriction enzyme *NcoI* in 50 μl of a reaction buffer [50 mM Tris-HCl (pH 8.0), 10 mM MgCl₂, 0.1 M NaCl] at 37°C for 3 hours, and directly cloned into *NcoI*-digested plant expression vector pB1525. The resulting plasmids were identified and designated as pB1525-NP⁺ in the sense orientation relative to the cauliflower mosaic virus (CaMV) 35S promoter, and as pB1525-NP⁻ in the reverse orientation. The ability of this expression cassette to produce the NP was determined by transient expression of the NP gene in *Nicotiana tabacum* protoplasts, as described by Pang et al [see Gene 112:229 (1992)]. The expression cassette containing the NP gene was then excised from pB1525-NP⁺ by a partial digestion with *HindIII/EcoRI* (since the NP gene contains internal *HindIII* and *EcoRI* sites), and ligated into the plant transformation vector pBIN19 (Clontech Laboratories, Inc.) that had been cut with the same enzymes. The resulting vector, pBIN19-NP⁺ and the control plasmid pBIN19 were transferred to *A. tumefaciens* strain.
LBA4404, using the procedure described by Holsters et al [see Mol. Gen. Genet. 163:181 (1978)].

Nucleotide sequence analyses of the inserts in clones pTSWV-23 and Pb1525-NP+ were determined using the dideoxyribonucleotide method, T7 polymerase (U.S. Biochemicals, Sequenase™), and the double-stranded sequencing procedure described by Siemieniak et al [see Analyt. Biochem. 192:441 (1991)]. Nucleotide sequences were determined from both DNA strands and this information was compared with the published sequences of TSWV isolates CPNH1 using computer programs available from the Genetics Computer Group (GCG, Madison, WI).

Transient expression of the NP gene in tobacco protoplasts were also prepared. Plasmid DNAs for clones pTSWVS-23 and pUC18cpphas TSWV-NP (containing the PCR-engineered NP gene insert) were isolated using the large scale alkaline method. The PCR-engineered NP gene insert was excised from clone pBIS25-NP+ by Ncol digestion to take advantage of the available flanking oligomer primers for sequencing. The expression cassette pUC18cpphas is similar to pUC18cppexp except that it utilizes the poly(A) addition signal derived from the Phaseolus vulgaris seed storage gene phaseolin. These plasmid DNAs were subjected to two CsCl-ethidium bromide gradient bandings, using a Beckman Ti 70.1 fixed angle rotor. DNA sequences were obtained using dideoxyribonucleotides and the double-stranded plasmid DNA sequencing procedure described above. Nucleotide sequence reactions were electrophoresed on one-meter long thermostated (55°C) sequencing gels and nucleotide sequence readings averaging about 750 bp were obtained. Nucleotide sequences were determined from both DNA strands of both cloned inserts to ensure accuracy. Nucleotide sequence information from the TSWV-BL S RNA isolate was compared as discussed below, with TSWV isolates CPNH1 and L3 using computer programs (GCG, Madison, WI).

The nucleotide and deduced amino acid sequences of cloned cDNA and PCR-engineered insert of TSWV-BL S RNA and their comparison with the nucleotide sequence of TSWV-CPHN1 S RNA are shown below.

The nucleotide sequence of the TSWV-BL S RNA clones pTSWVS-23
(TSWV-23) and pB1525-NP+ (TSWV-PCR) were obtained using the
double-stranded dideoxynucleotide sequencing procedure of Siemieniak,
and their sequences are compared with the relevant regions of the
nucleotide sequence of the TSWV-CPNH1 S RNA reported in GeneBank
Accession No. D00645. The nucleotide sequence of TSWV-CPNH1 S RNA
has been reported by De Haan (1990) and is represented by the following
sequence:

```
CAGCTGAAA GCAACAACAG AACTTAAAT TCTCTTGCAAG TGAATCTCT
GCTCATGCTA GCAAGAAAAC AGACTTCGACC TCTCTCAAA GCTCCACTG
ATTCTCATTT CAAAGCTGAC CTCTGCTCAA GGCTTTCAAAG GCTTTGAG
CAGCTTCCCA TCTGAGAATCT GTGCAAGATG GCAGAGATGA AAACAACAA
AACATTTTATTT TTAGCTATAG CTTGCAATCC AAACCATCAG AGTGGTGA
CAGCTTAAAT CATACTCTTG ATTTGCGAGAC ATGGATCCCA AATCGCCAAA
TGCAAAAGCTC TCTTGTGACT TACAATGAGT TTTCCTGATT TAAAGTAGCC
TTCTACATTG GCTGATGCCC CTCTATACCC CAAAGATGCG GTTCCAAATGC
TCTGAGTCA AACAGACAGA CTCTTGACAG AGTCTTTGCC AACTTACTTG
CAAGAGATG TCTATCTCTCA CACTTGTGAC AACTTGTGAC TAACTCTTG
AAAGTTGAGAT TTAGGTGAAA GAACCTTGA TTACAGTGA GAATGCTACA
AAAGGAATTA TTCTTTCATCA AAAAAGTATG AATGTCTTTC ACCTAACACA
CTACTATGT CTACTATGAA CAGCATCAAA ATCCCTCATG GGAAGATGA
CTGGCCAGA GGAAATTTAA AATTTCTTCC ACAATCTATT TCGATGGGAA
AACTTCTTGT AAAGCTGGAT TTATCGGGGA TCATAAGAGA AGAATCTAG
GGTACGGAAG GTGAGCTTC ACAGTCAAAA TAATCTTGTCT TGGTCAGCT
TTTTCTATTCT ATGTGATTTT TCTTCTTTCTT CTCTAACATT TATTATTATT
CTGTCTTCGAA ATCTCTCTG TCTAGTGAAA ACCAAAAAA
CAAAAAATTA AAAAAATTTT AAAAAAAAAT TAAAATAAAAA TAAAAATG
AAATAAAAAC AACAAAAAAT TAAAAACAGA AAAAAACAAA AGACCCGAAA
GGGCAATT TGCCCAATT TGCGTTTCTT TTTCCTTTTT TTTCCTTTTT
TTTTTATTTT TTATTTTTAT TTATTTTATT TTTATTTTTA TTTTTTTTTT
ATTTTTTTAA TGTTTTTTTT TGTTTTTTTT TGTTTTTTTT TGTTTTTTTT
AGGAAACACAC AGGAAAGCA AAATTTAATTT ATTAAACACTT ATTTAAAAAT
TGACACACTA AGGAAAGCA AGGAATAGAG ATGAGAAAGA CTATATATAT
TTTAAAGCAC ATTTATTTAT TTTGTGCAAG CTGCCGGCA AAACTTTGGA
GGATTTTGC CTGCTTTTAA ACCCAGAGA TTTCAATGAA CGTGGTAAAG
```

9
```
GGACCTAAG CTAGCAAGAT ACCTTTTACC CTTCCTCCAC TGAATCCCAT 1500
TCATTTCAAA TGCTTGTGTT TCTAGCAGAC TGCAACTTCT TCTTAGCTTG 1550
TCTTTGCGCT CATAGTCTGT TGGATGCGAT CGAGCCGCTT TGATTTTAGC 1600
ATCCCTGATAT ATAGCCAAAG CAAACACTGAT GATCAGAACG CTATCAACTG 1650
5  AAGCAATAAG AGGTTAACAGA CCTCCCAAGCA TTATGCGGAG TCTACAAGC 1700
TTGCCATCAT CGGAGTAGGA TCGTATGGCT TGAAATGAAAG GATGGGAAAC 1750
AAATTATGAT TGGATAGGAT TGAGATTCCT AGAATTCCCA GCTTCTACAA 1800
CAAACCGTGC CCTGATCAGG CTATCAGGCG TTCTGAGAGT CAGTCAGTGG 1850
CTCCTAAATCC CTCTCGAGGT TTCTCTTATG GTAATTACAC CAAAGATCAA 1900
10 ATGCTTTGCC TTTAAAAAGCT TCTTATGCTG CGACGATTCT TTTAGGAAAGT 1950
TCAGACATGA AAAAGCGTCA ATCTTACTGA TCCTGCTGAGT GCTTCCAGAA 2000
CAAAAGCTGC TCAAGTGTGA TGCCACCAGA TCCGTATCTT CCAAACCTCT 2050
AAGGGCTTGG CCTGTGTCGA ACRAAAGGAAC AATCCCTCAG TTAGAGCAGT 2100
TAAACCTTGA CATGATGATG CTAAAGTGGT TTAATAGCTT GACCGAGCTG 2150
15 AACTCTGGTG CGAAGAGTGC AACTCTGTTG CCGCAGTGCG TTCTCTAGGT 2200
TCTTTAAGTG ATGATGTTGA AGACTGAGGC TTAAGTATGG AACACAAAAAT 2250
TGACACGATTT GCTCT 2265
```

The incomplete deduced amino acid sequence of the nonstructural protein gene on TSWV-CPNH1 S RNA is provided below beginning with nucleic acid at position 1 and ending with the nucleic acid codon ending at position 783:

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

The nucleotide sequence for TSWV-23 depicted below compares closely with the TWSV sequence given above, and contains one-half of the nonstructural gene and one half of the nucleocapsid protein gene.

AAAATCCCTTC GCAGGGAAT CATCTGCTGAT GTTACGAAC AAACAAACAA 50
TGCTGAAATC TCAAGCTTTT GTCAAAGCCT CTCATGATTC TAACTGCGAG 100
CTGAGCCCTCT GCTAAGGGAT TGCAAGTGTT TTTGAGACAA TTTCTCAAC 150
GAAATTTGCAT AAGGTTCGAG GAGATGAAC AATGGAAACA TTTTAATTAT 200
CTATGCGCTG CATCCTCTAC CATACAGTGG TTGAGCAAC TGGTAAACAC 250
ACTGTTATTAT GCAAGCTGCA GCCTCCTAG CTGAAATTGTA AACAAGCTCT 300
TGGATATGCA ATAGATTTCT CGTATTCTAA GGGGCTTAAG AAGATATTGC 350
ATGATCTCTTC ATATCCGCAA AGGATTGATC ATGCTCTGTGCT GGAAGCTCAG 400
ACATCTTTGG CAGACTCTCT TTGCAAGAAT TTGAGGAAAG ATGAGCACAT 450
CTGCACTCTAG AAGACCAATG AGCTACATCC TGGAAAGTTA GATTTTAGTTG 500
AAATACCTTT GAAATCTGAT GAGAGCCTT ACAAAAGGAA ATATTTCCCTT 550
TCAAAACGCT TTGATGCTCT TCCATCTAAG ATACAAATCT TGTCCTATT 600
AGACAGCATC CAAATCCTCT CTCGAAAGAT AGACTTTGCC AGGAGAAACA 650
TTAAAAATTC TCCACATACT ATTTGAGTTG CAAATCTTTT GTTAAATCTTT 700
GATTTTAGCC GGGTTAAAAA GAAAGAATCT AAGATTAAGG AGCATTATGC 750
TGGGAGATCA AAATGATCTT GCCTGCTGCA GTTCTTCTGA ATGTTATTT 800
GTTATTTTCT TTTCTTTACT TTCATTTAT TTTCCTGTGTG TCAATCTTTTT 850
CAAATCCTTC CTGCTCTAGTA GAAACATCAA AAACAAAAT AAATTTAAAA 900
The nucleic acid sequence for TSWV-PCR according to the present invention as depicted below also compares closely with the TSWV sequence given above and covers the whole nucleocapsid protein gene.

```
TAAATTCAAA ATAAAAATAAA AATCAAAAAA TGAATATAAA GCAACAAAAA  950
AAATAAAATA CAAAAAACCA AAAAAAGATCC CGAAGGACA AGTTTGCGCA  1000
AAATTTGCGGT TTGGTTTTTG TTTTGTTTTT TTTTGTGTTTT GTTTTTTATT  1050
TTTTTTTTTTT TTTTTTTTTTT ATTTTTATTTT ATTTTTATTTT TTGTTTTTGT  1100
TTGTTTTTTT GTTTTTTTTT AAGCTACAAAA CTGAACATTTA GTATATATAT  1150
TTTTTTTTTTTTT TTTTTTTTTTTTT TTTTTTTTTTTTT TTTTTTTTTTTTT TTTTTTTTTTTTT  1200
GTAAAGGAAG CTTATATATAT ATTTTAGGCT TTTTTATAAT ATTAATCTACA  1250
GGCTGTTTGA ATCAAGGTTT GAGTGGGTG CGTTTTTTTT AAGCGGAAAC  1300
ATTTCTAAGA ACTCTTTAGG GGTCTACTG TGAATGTTCA TAGCAATACT  1350
TCCTTTAGCA TTTCAGATTG TGGGCTTAAAG TATAGCAATCA TACTTGTCTC  1400
CTCTCCATGC CTGATCTTTCA TTTAATTTAA ATGGCTTTTTCT TTTTCCAGCA  1450
GGCCCGGGCT TTGCTAAATTC TTCCTCGTGC TTTTCTCAGG TTTTTGCTGT  1500
CCGCCGAGTAC TTGGTGTTTG GATTCCTAGA TATAGCCAAAG ACAACTCTGA  1550
TCAATCTCCAA GTCCATACCT GAAGCAATTA GAGTGTGAGCTTTTACCAGC  1600
ATTATTGCAG GCTCTACAGA CTTGTCAGCA TCAAGGATGT ATTAGCATAGC  1650
TTGAGTCAAA GGCTGGGGAG CAATCTTTGA TTTGATAGTTA TTTGAGTTCCT  1700
CAGATTCAC 1709
```

```
20
TTAACACACT AAGCAGCAGC AAACAAATAAA GATAAAGAAA GTTTTTTTTATA  50
TTTTATAGGCT TTTTTTTATAT TTTAATTTAA ATGGCTTTTTCT TTTTCCAGCA  100
GGCTGTTTGA ATCAAGGTTT GAGTGGGTG CGTTTTTTTT AAGCGGAAAC  150
ATTTCTAAGA ACTCTTTAGG GGTCTACTG TGAATGTTCA TAGCAATACT  200
TCCTTTAGCA TTTCAGATTG TGGGCTTAAAG TATAGCAATCA TACTTGTCTC  250
CTCTCCATGC CTGATCTTTCA TTTAATTTAA ATGGCTTTTTCT TTTTCCAGCA  300
GGCTGTTTGA ATCAAGGTTT GAGTGGGTG CGTTTTTTTT AAGCGGAAAC  350
CTCTCCATGC CTGATCTTTCA TTTAATTTAA ATGGCTTTTTCT TTTTCCAGCA  400
GGCTGTTTGA ATCAAGGTTT GAGTGGGTG CGTTTTTTTT AAGCGGAAAC  450
CTCTCCATGC CTGATCTTTCA TTTAATTTAA ATGGCTTTTTCT TTTTCCAGCA  500
GGCTGTTTGA ATCAAGGTTT GAGTGGGTG CGTTTTTTTT AAGCGGAAAC  550
CTCTCCATGC CTGATCTTTCA TTTAATTTAA ATGGCTTTTTCT TTTTCCAGCA  600
GGCTGTTTGA ATCAAGGTTT GAGTGGGTG CGTTTTTTTT AAGCGGAAAC  650
CTCTCCATGC CTGATCTTTCA TTTAATTTAA ATGGCTTTTTCT TTTTCCAGCA  700
GGCTGTTTGA ATCAAGGTTT GAGTGGGTG CGTTTTTTTT AAGCGGAAAC  750
```
ACAAAAGTCT TTGAATTTGA ATGCTACGAG ATTTCTGATCT TCCTCAAACT 800
CAAGCTCTTT GCTTGGTGTC ACAAAGGCAA CAGTCTTTTC CTAGTGAGCC 850
TIPAACCAT 858

Together the cloned TSWV-23 insert overlaps the TSWV-PCR insert, and together they represent the 2028 nucleotides of the TSWV-BL S RNA according to the present invention. This 2028 nucleotide sequence according to the present invention contains a part of the nonstructural gene and whole nucleocapsid protein gene. The combined sequence is:

10 AAATTCTCTTT GCCAGAGAT CTCTGCTCAT GTTAGCAGAA ACAAACATCA 50
TCCTCAGTCT CAAGCTTTTT GCTAAAGCTT CTACGTGATCT TAAATTTCAG 100
CTGAGCCTTCT GCTAAGGGGT TCCTAAAGGT TTGAAGCAGA TTTCCATTCA 150
GAAATTTGCTC AAGTGTCGAG GAAGTGAAC AATAAATACA TTTTATTATAT 200
CTTATCGCTT CATACCAAC TATAACAGTG TTGAGACAGC TTTAACACCT 250
15 ACTCTTATTT GCCAAGCTCA GCTCTCCATT TGAAATGGTA AAACCTCCTTT 300
TGAATATATCA ATGAAGTTTTT CTGAATTAAA GGAACCTTCAC ACAAATTATTC 350
ATGACCTCTC ATATCCCTAA AGGAATTGTTG ATGGCTCTGCT TGCAACTCAC 400
ACATCTTTTG CACAGTCTCT TTGCAACAAC TTGCAAGAAG ATGTGACTCAT 450
CTACACCTTG AACACACAGT AGCTAACTCC TGGAAGTTAA GATTAGAGTG 500
20 AAATACCTTT GAAATCAATG GAAGACGCTT ACAAAAGGAA ATATTTCCTT 550
TCCTAAACAG TTGAATGCTT TCATCTACAT ATATAAAGTG TGICCTTTTAT 600
AGCAGACGTTC GAATCCATTT CATGAAGATT ACAGTTTGGCC AGGGAAGGAA 650
TTTTAAATCTC TCCTCAATCT ATTTGAGTTG CAAAAATCTTT GTTTAAATCTT 700
GATTAAAGCG GGAATTAATA AAAGAATGTTA GAGATTAAGG AAGCATTCCG 750
25 TCCAGGATCA AAATGATCCTT GTGTGTTGCA GCTTTTTCTA ATTTATGTAT 800
GTTTATTTTC TTTCTTTAC ATAAATATT TTTCGTTTTG TCAATTTCCTT 850
CAAAATCTCT CIGTCTAGTA GAAACCAATA AAACAAATAA AAAAAATAAAAA 900
TAAATACAAA ATAAATAAAA ATCTCAAAAA GAAATAAAAG CAACAAAAGAT 950
AATTAAAAA AAATAAAACA AAAAAAGCTC CGAAAGGACA ATTTTGGGCA 1000
30 AAATTTGGGT TTGTAAAAAT TTAAAAATTT TTAAAAATTT GTTTATTTAT 1050
TTAAAAATTTTTT TTAAAAATTTTT TTAAAAATTTTT TTAAAAATTTTTT 1100
TTGTTAATTTT GTTAAAAATT AACCAACAAC CAGAAGAAAAC AAACAAAAAT 1150
TAAACACACT TATTAAAAT TAAACAACACT AAGCAAGCAC AAAAAATAAA 1200
GATAAGAGAA GTTTATATAA TTTATAGGCT TTTTTATAAT TTTACTTACA 1250
35 GCTGCTTTTTA AGCAAGGCTT GITGTTTTTT CGTTTTTTTT AACCCCAAC 1300
ATTTCCATAAG ACTGGTAAAG GGTTCACCTG TAATGTTCCA TAGCAATACT
TCTTCTGACCT TAGAGTTGCC TGAGCTTAPG TAAGCAGCACT TACTTTTCC
CTCTTCCACCT GTGACCTTCA TCTATTCCAA ATGCTTTTCT TTTCAGCAC
GTCGCAAGCTT TTCTCTAGGC TTCCTCTGGT TCGACTTCTT TTGGGTGAT

cccgatctcc tttgaatttg cattctgata tatagcaag acacacgta
ctactccaaacct gactaactca gttctcatca tcaaggaat aatccattag

10
ttcgcaaggg tcaagtgcagt gctttccaatc ctgctgaag tttctcttagt
GGATATTITTA CCAAAATTPAA AATCCTTTTG CTTAATTTCC TCTAAATTTC
TCTGCATTT CTCTAGGAT GTGAGCAAGT AATATTGCT CATCTTGG

15
ATCTGGAGCA GGTTCACCTG AGTACAGTAC TGGAGTTGA ATGCTACACAG
ATCTCGCTTC CTTGCAAAGTT CAAGGCTTTT GCCCTGTTGC ACAAAAGCAA

This comparison showed that cDNA insert of clone pTSWVS-23 included about 760 bp of the 52 K protein viral component gene, the complete intergenic region (492 bp), and 450 bp of the NP gene (about half of the NP gene). This cloned insert had its 3'-end located exactly at an EcoRI recognition site, which suggested incomplete EcoRI methylation during the cDNA cloning procedure. Although this clone did not contain the complete TSWV-BL NP gene, its sequence was of considerable importance since it had a 450 bp overlap with the sequence of the PCR-engineered NP gene (a total of 2028 bp of the TSWV-BL S RNA is presented in the nucleotide sequence for TSWV). The sequence comparison between this TSWV-BL PCR-engineered and TSWV-CPNH1 NP genes revealed a total of 21 nucleotide differences (2.7%), eight of which encode amino acid replacements (3.1%). Since this PCR engineered NP gene was obtained using Taq polymerase, which is known to incorporate mutations, it is possible that some of these differences were introduced during PCR amplification. However, 15 of these nucleotide differences were located within the overlapping region between the TSWV-BL cDNA and PCR clones, and all but one of these nucleotide differences (position 1702 of TSWV; position 485 of TSWV-PCR)) are shared by both TSWV-BL S RNA derived clones. This
comparison clearly showed that the PCR amplification did not contribute greatly, if at all, to the difference between the nucleotide sequences of these two cloned NP gene regions. The nucleotide difference at position 1702 resulted in the amino acid replacement of Ile with Ser, and even this difference could be due to the lack of homogeneity within the TSWV-BL isolate.

EXAMPLE V

Agrobacterium-mediated transformation:

Leaf discs of Nicotiana tabacum var Havana cv 423 were inoculated with the Agrobacterium strain LBA4404 (ClonTech) containing the vector pBIN19-NP+ or the control plasmid pBIN19, by soaking overnight in a liquid culture of the Agrobacterium, and the inoculated leaf discs were incubated on non-selective MS medium for 3 days. [see Science 227:1229 (1985)]. Transformed cells were selected and regenerated in MS medium containing 300 μg/ml kanamycin and 500 μg/ml carbenicillin for shoot regeneration. Roots were induced after transfer of plantlets to hormone-free medium. Rooted transformants were transferred to soil and grown under greenhouse conditions. The MS medium contains full strength MS salt (Sigma), 30 g/l sucrose, 1 mg/l BA and 1 ml of B5 vitamins [1 mg/ml Nicotinic acid, 10 mg/ml Thiamine (HCl), 1 mg/ml Pyridoxine (HCl), 100 mg/ml Myo-Inositol]. Transgenic plants were self-pollinated and seeds were selectively germinated on kanamycin medium.

EXAMPLE VI

Serological detection of proteins:

Double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA) was used to detect the expression of NP gene in transgenic plants with polyclonal antibodies against the TSWV-BL NP. Each sample was prepared by grinding a leaf disc (about 0.05 g) from the top second leaf of the plant in 3 ml of an enzyme conjugate buffer [phosphate-buffered saline, 0.05% Tween 20, 2% polyvinylpyrrolidone 40, and 0.2% ovalbumin]. For tobacco protoplasts, the cell extracts after centrifugation were directly used for the assay. A ten- and three-fold dilutions of the samples from both transgenic plants and tobacco protoplasts were made just before DAS-ELISA.
For Western blots, a leaf disc (about 0.05 g) was ground in 0.25 ml of 2X SDS/sample buffer (0.126 M Tris buffer, 20% glycerol, 2% SDS, 2% 2-mercaptoethanol, and 0.01 mg/ml bromphenol blue). The homogenates were centrifuged and the supernatants were boiled before loading. Proteins (10-20 μl sample/lane) were separated and blotted onto a membrane. The membrane was then processed following the manufacturer's immunoselect kit instruction manual (Gibco BRL Life Technologies Inc.). Antibodies to the whole virion were preabsorbed with cell extracts from health tobacco plants [See Plant Disease 70:501 (1986)], and were used in Western blot at a concentration of 2 μg/ml.

Serological reactions of TSWV isolates (TSWV-BL, Arkansas, 10W pakchoy, Begonia or Brazil) were assayed in DAS-ELISA using antibodies raised against TSWV-BL virion, or the NP of TSWV-BL or TSWV-I.

**EXAMPLE VII**

**Inoculation of transgenic plants with TSWV isolates.**

Inocula were prepared by infecting *Nicotiana benthamiana* Domin. with different TSWV isolates and grinding infected leaves (0.5 g) of *N. benthamiana* plants (1 to 2 weeks after inoculation) in 15 ml of a buffer (0.033 M KH₂PO₄, 0.067 M K₂HPO₄ and 0.01 M Na₂SO₃). The inoculum extracts were immediately rubbed on corundum-dusted leaves of transgenic plants and the inoculated leaves were subsequently rinsed with H₂O. Because TSWV is highly unstable in vitro after grinding, each batch of inoculum was used to first inoculate NP(+) plants containing the NP gene; the last inoculated plants of each inoculum were always control NP(-) plants containing the vector sequence alone to assure that a particular virus inoculum was still infective at the end of inoculation.

Data on local lesions and systemic infections were taken 7-15 days after inoculation and expressed in the following table as the number of plants systemically infected over the number of plants inoculated, except where indicated. In this table, the data collected under "ELISA" is the data of R₀ lines from which the R₁ plants were derived; the Begonia isolate induced local lesions on the R₁ plants, and the resistance was expressed as the number of plants producing local
lesions over the number of plants inoculated; and NT indicates that there was no test.

Reactions of R₁ plants expressing the nucleocapsid protein (NP) gene of tomato spotted wilt virus (TSWV) to inoculation with TSWV isolates.

<table>
<thead>
<tr>
<th>Reactions to TSWV isolates</th>
<th>ELISA: (R₀ pl.)</th>
<th>BL</th>
<th>Arkansas</th>
<th>10W Pakchoy</th>
<th>Bologna</th>
<th>Brazil</th>
</tr>
</thead>
<tbody>
<tr>
<td>R₀ line</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP(+2)2</td>
<td>0.015</td>
<td>0/20</td>
<td>4/25</td>
<td>3/24</td>
<td>29/40</td>
<td>36/36</td>
</tr>
<tr>
<td>NP(+4)</td>
<td>0.386</td>
<td>6/30</td>
<td>21/23</td>
<td>18/21</td>
<td>9/48</td>
<td>42/42</td>
</tr>
<tr>
<td>NP(+9)</td>
<td>0.327</td>
<td>0/20</td>
<td>NT</td>
<td>20/20</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>NP(+14)</td>
<td>0.040</td>
<td>0/20</td>
<td>_</td>
<td>9/20</td>
<td>8/18</td>
<td>18/18</td>
</tr>
<tr>
<td>NP(+21)</td>
<td>0.042</td>
<td>0/15</td>
<td>5/15</td>
<td>3/15</td>
<td>2/4</td>
<td>6/6</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP(+22)</td>
<td>0.142</td>
<td>0/20</td>
<td>_</td>
<td>15/20</td>
<td>31/36</td>
<td>36/36</td>
</tr>
<tr>
<td>NP(+23)</td>
<td>0.317</td>
<td>0/20</td>
<td>_</td>
<td>16/20</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>NP(-)</td>
<td>_</td>
<td>42/42</td>
<td>24/24</td>
<td>62/62</td>
<td>66/66</td>
<td>54/54</td>
</tr>
</tbody>
</table>

As described above, the isolation of the TSWV-BL NP gene, which resides in the S RNA component of TSWV, was approached using two strategies. The cDNA cloning strategy yielded several clones containing cDNA inserts derived from TSWV-BL S RNA, as identified by hybridization against an oligomer probe complementary to the TSWV-CPNH1 S RNA. Clone pTSWVS-23 contained the longest insert, which mapped at about 1.7 kb in length. The second strategy utilized the published sequence of TSWV-CPNH1 S RNA and PCR to amplify and engineer the NP gene for expression directly from total TSWV-BL RNA. Oligomer primers JLS90-46 and -47 were synthesized, with JLS90-46 being complementary to the S RNA in the 5' coding region of the NP gene (positions 2051-2073 of the TSWV-CPNH1) while JLS90-47 being of the 3' non-coding region of the NP gene (positions 1218 to 1237 of the TSWV-CPNH1). Both of the primers contain the recognition site for the restriction enzyme Ncol for subsequent cloning, and the primer JLS90-46 has a plant consensus translation initiation codon sequence (AAXXATGG), which upon amplification was expected to fuse the translation initiation codon to the third codon (GTT) of the NP gene. Fusion of the translation initiation codon to the third codon of the
TSWV-BL NP gene was done to preserve the *N*coI recognition site while not incorporating any amino acid codons. Thus, expression of the PCR-engineered TSWV NP gene would yield a TSWV-BL NP that was two amino acids (Ser-Lys) shorter at the N-terminus than the native NP.

This specifically-amplified DNA fragment, of about 850 bp, was digested with *N*coI and cloned into the plant expression vector pB1525. The orientation of the TSWV-BL NP gene with respect to the CaMV 35S promoter was determined by restriction enzyme site mapping (EcoRI, HindIII, Aval and AiwNI). Several clones were isolated that contain the insert in the proper orientation (pB1525-NP+) and others that contain the insert in the opposite orientation (pB1525-NP−). This restriction enzyme site mapping data also showed that the inserts of clones pB1525-NP+ contained restriction enzyme sites that were identical to those found in the TSWV-CPNH1 NP gene. The expression of TSWV-BL NP gene was thus controlled by a double CaMV 35S promoter fused to the 5′-untranslated leader sequence of alfalfa mosaic virus (ALMV) of the expression vector pB1525. Expression vectors that utilize the stacked double CaMV 35S promoter elements yield higher levels of mRNA transcription than similar vectors that utilize a single 35S promoter element.

Three pB1525-NP+ clones were transiently expressed in tobacco protoplasts to confirm that the amplified DNA fragment encoded the NP. To achieve this, the clones were transferred into tobacco protoplasts by the PEG method, and after two days of incubation the expressed NP was detected by DAS-ELISA using antibodies against the whole TSWV-BL virion. High levels of NP were produced in tobacco protoplasts harboring the NP gene in plasmid pB1525-NP+; while no NP was detected in tobacco protoplasts transformed with the antisense NP sequence (pB1525-NP−).

As described previously, the PCR-engineered insert of clone pB1525-NP+ and the cDNA insert of the clone pTSWV-23 were subjected to double stranded sequencing. The sequence analysis of the cDNA and the PCR clones revealed inserts of 1.71 kb and 865 bp, respectively which, when compared with the sequence TSWV-CPNH1 S RNA, shows that cDNA insert of clone pTSWV-23 includes about 760 bp of the 52 K
protein viral component gene, the complete intergenic region (492 bp), and 450 bp of the NP gene (about one-half of the gene). This cloned insert has its 3'-end located exactly at an EcoRI recognition site suggesting incomplete EcoRI methylation during the cDNA cloning procedure. Although this clone does not contain the complete TSWV-BL NP gene, its sequence is of considerable importance since it has a 450 bp overlap with the sequence of the PCR-engineered NP gene. The sequence comparison between this TSWV-BL PCR-engineered and TSWV-CPNH1 NP genes reveals a total of 21 nucleotide differences (2.7%), eight of which encode amino acid replacements (3.1%). Since this PCR-engineered NP gene was obtained using Taq polymerase, which is known to incorporate mutations, it is possible that some of these differences were introduced during PCR amplification. However, 15 of these nucleotide differences are located within the overlapping region between the TSWV-BL cDNA and PCR clones, and all but one of these differences (position 1702) are present in both TSWV-BL S RNA derived clones. This comparison clearly shows that the PCR amplification did not contribute greatly, if at all, to the difference between the nucleotide sequences of these two NP genes. The nucleotide difference at position 1702 results in the amino acid replacement of Ile with Ser, and even this difference could be due to the lack of homogeneity within the TSWV-BL isolate.

The possibility that the nucleotide differences can be attributed to divergence among the TSWV isolates is also supported by comparisons with other sequenced regions among TSWV-CPNH1, TSWV-L3, and TSWV-Bi S RNAs. These comparisons are tabulated below:

Percent nucleotide and amino acid sequence differences for the comparison of TSWV S RNA component from isolates CPNH1, L3 and BL

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Nucleotide</th>
<th>Amino Acid</th>
<th>Intergenic Nucleotide</th>
<th>NP Gene Nucleotide</th>
<th>NP Gene Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPHN1/L3</td>
<td>68/1396b(4.9)c</td>
<td>49/464(10.6)</td>
<td>46/511(9.0)</td>
<td>24/777(3.1)</td>
<td>4/258(1.6)</td>
</tr>
<tr>
<td>CPHN1/BL</td>
<td>21/758(4.1)</td>
<td>23/251(9.2)</td>
<td>26/496(5.2)</td>
<td>19/765(2.5)</td>
<td>8/255(3.1)</td>
</tr>
<tr>
<td>L3/BL</td>
<td>38/765(5.0)</td>
<td>20/254(7.9)</td>
<td>38/498(7.6)</td>
<td>19/767(2.5)</td>
<td>4/255(1.6)</td>
</tr>
</tbody>
</table>
Comparisons are made using the sequence information available from the particular component region of TSWV-BL. The comparison for the TSWV-BL NP gene includes the combined sequence information from the cDNA clone, pTSWVS-23 and PCR-engineered insert.

Comparison numbers are total differences (nucleotides or amino acids) divided by total number of positions (nucleotides or amino acids) compared. For both nucleotide and amino acid calculation gaps, regardless of length, were counted as one mismatch.

Numbers in parentheses are percentages.

The nucleotide sequence of the NP genes from the CPNH1 and L3 isolates differ from each other by 3.1% and from the BL isolate by nearly a similar degree (2.5%). However, the NP amino acid sequences between CPNH1 and BL isolates differ by a considerably larger amount than they differ between the L3 and BL or CPNH1 and L3 isolates. The results tabulated above also reveal that the NP gene region of these TSWV isolates is subject to a higher degree of selective pressure than the 52 K protein as the differences among the amino acid sequences of the 52 K protein range between 7.9 to 10.6%, more than twice that found for the amino acid sequence of the NPs. Nucleotide sequence divergence is highest among the intergenic regions, indicating that this region is subject to less selective pressure than either genetic region.

The presence of NP gene sequences in transgenic plants was first confirmed by PCR analysis. A NP DNA fragment of about 800 bp was specifically amplified from the total DNAs of transgenic NP(+) plants using the primers homologous to sequences flanking the NP gene, whereas no corresponding fragment was detected in control NP(-) plants. Expression of the NP gene was assayed in each R0 transgenic plant by DAS-ELISA, and the results are presented in the following table:
Reactions of R0 transgenic plants expressing the nucleocapsid protein (NP) gene of tomato spotted wilt virus (TSWV) to inoculation with TSWV-BL isolate

<table>
<thead>
<tr>
<th>plant age</th>
<th>R0 clone</th>
<th>ELISA&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Lesions/leaf&lt;sup&gt;b&lt;/sup&gt;</th>
<th>NP(+) : NP(-)&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-8 leaves:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>NP(+)1</td>
<td>0.374</td>
<td>7 (199)</td>
<td>1:28</td>
</tr>
<tr>
<td></td>
<td>NP(+)2</td>
<td>0.015</td>
<td>0 (199)</td>
<td>0:199</td>
</tr>
<tr>
<td></td>
<td>NP(+)3</td>
<td>0.407</td>
<td>23 (102)</td>
<td>1:4</td>
</tr>
<tr>
<td></td>
<td>NP(+)4</td>
<td>0.386</td>
<td>2 (102)</td>
<td>1:51</td>
</tr>
<tr>
<td></td>
<td>NP(+)5</td>
<td>0.023</td>
<td>0 (124)</td>
<td>0:124</td>
</tr>
<tr>
<td>10</td>
<td>NP(+)6</td>
<td>0.197</td>
<td>35 (325)</td>
<td>1:9</td>
</tr>
<tr>
<td></td>
<td>NP(+)7</td>
<td>0.124</td>
<td>1 (325)</td>
<td>1:325</td>
</tr>
<tr>
<td>9-10 leaves:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP(+)8</td>
<td>0.344</td>
<td>36 (36)</td>
<td>1:1</td>
</tr>
<tr>
<td></td>
<td>NP(+)9</td>
<td>0.327</td>
<td>2 (20)</td>
<td>1:10</td>
</tr>
<tr>
<td>15</td>
<td>NP(+)10</td>
<td>0.406</td>
<td>34 (33)</td>
<td>1:1</td>
</tr>
<tr>
<td></td>
<td>NP(+)11</td>
<td>0.156</td>
<td>5 (20)</td>
<td>1:4</td>
</tr>
<tr>
<td></td>
<td>NP(+)12</td>
<td>0.133</td>
<td>9 (57)</td>
<td>1:6</td>
</tr>
<tr>
<td></td>
<td>NP(+)13</td>
<td>0.144</td>
<td>2 (7)</td>
<td>1:4</td>
</tr>
<tr>
<td></td>
<td>NP(+)14</td>
<td>0.040</td>
<td>0 (19)</td>
<td>0:19</td>
</tr>
<tr>
<td>20</td>
<td>NP(+)16</td>
<td>0.053</td>
<td>0 (10)</td>
<td>0:10</td>
</tr>
<tr>
<td>5-6 leaves:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP(+)20</td>
<td>0.487</td>
<td>203 (117)</td>
<td>2:1</td>
</tr>
<tr>
<td></td>
<td>NP(+)21</td>
<td>0.042</td>
<td>0 (117)</td>
<td>0:117</td>
</tr>
<tr>
<td></td>
<td>NP(+)22</td>
<td>0.142</td>
<td>0 (208)</td>
<td>0:208</td>
</tr>
<tr>
<td>25</td>
<td>NP(+)23</td>
<td>0.317</td>
<td>223 (208)</td>
<td>1:1</td>
</tr>
<tr>
<td></td>
<td>NP(+)24</td>
<td>0.051</td>
<td>0 (35)</td>
<td>0:35</td>
</tr>
<tr>
<td></td>
<td>NP(+)25</td>
<td>0.286</td>
<td>13 (35)</td>
<td>1:3</td>
</tr>
<tr>
<td></td>
<td>NP(+)26</td>
<td>0.037</td>
<td>0 (22)</td>
<td>0:22</td>
</tr>
<tr>
<td></td>
<td>NP(+)27</td>
<td>0.425</td>
<td>305 (22)</td>
<td>14:1</td>
</tr>
</tbody>
</table>

<sup>a</sup>Production of the NP in transgenic plants was assayed by double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA); concentration of antibodies against viron for coating: 1 µg/ml; dilution of conjugate to the NP of TSWV-BL: 1:250; results taken 150 min. after adding substrate; readings at 405 nm.

<sup>b</sup>Local lesions that developed on inoculated leaves were counted 7 days after inoculation. Data represent the average of three inoculated leaves. Data in parentheses are the number of lesions produced from control NP(-) plants inoculated with the same inoculum.

<sup>c</sup>The ratio of local lesions that developed on NP(+) plants transformed with pBIN19-
NP⁺ versus local lesions that developed on the control NP(-) plant when inoculated with the same inoculum.

Of the 23 NP(+) clones, 10 produced high levels of NP, 5 accumulated intermediate levels of NP, and the remaining 8 produced low levels of NP. The size of the NP expressed in transgenic plants was analyzed using Western blot. Many polypeptides from tobacco extracts were reactive to the antibodies against the whole viron even though the antibodies were pre-absorbed with extracts from healthy tobacco plants. Of those, only one band was unique to the pattern of polypeptides from tobacco plants transformed with the NP gene. This polypeptide was estimated to be around 29 kDa, which is near the expected size of the native NP. No antibody reactive-protein band of similar size was found in extracts from transgenic plants containing the vector pBIN19.

Inoculation of tobacco leaves with TSWV-BL isolate could result in either systemic infection or necrotic local lesions, depending upon weather conditions and physiological stages of plants. When R₀ plants were tested with TSWV-BL for viral resistance, TSWV-BL induced typical necrotic lesions on the inoculated leaves of control NP(-) plants 6-8 days after inoculation. However, transgenic NP(+) plants showed a spectrum of resistance to the virus when compared to control NP(-) plants. Eleven of the 23 NP(+) plants did not develop any local lesion or the number of lesions that developed was at least 20-fold less than that on the corresponding inoculated NP(-) plants. Three NP(+) plants had intermediate reactions (5- to 19-fold less lesions than controls) while the remaining 9 plants had low or no resistance. None of the inoculated NP(+) or NP(-) plants showed systemic infection. symptomless R₀ plants were monitored until the end of their life cycle, and no symptom was observed throughout their life cycles. The inoculated leaves of the symptomless NP(+) plants were checked for the presence of the virus on the leaves of C. quinoa plants. No virus was recovered from TSWV-BL-challenged leaves of highly resistant NP(+) plants, suggesting that the virus cold not replicate or spread in these NP(+) plants.
Leaf discs from selected R₀ plants were subcloned, and the regenerated plantlets were challenged by the virus. All subcloned R₀ plants displayed levels of resistance similar to their corresponding original R₀ plants.

Since TSWV is widespread and many biologically distant strains exist, the effectiveness of the transgenic plants to resist infections by different TSWV isolates were also tested. Five TSWV isolates were chosen in this study to challenge R₁ plants germinated on kanamycin-containing medium: TSWV-BL, Arkansas, 10W pakchoy, Begonia and Brazil. The first three isolates were reactive to the antibodies against the whole virion and the NP of TSWV-BL (the common TSWV "L" serogroup) (see figure 5). Begonia isolate reacted strongly to the antibodies against the NP of TSWV-I (the "I" serogroup) but not to those raised against the TSWV-BL NP, and therefore belonged to the "I" serogroup. No detectable reaction of Brazil isolate was found to the antibodies against either the NP of the TSWV-BL or the TSWV-I serogroup, and it was weakly reactive to the antibodies against the whole virion of TSWV-BL. Moreover, this isolate caused systemic mottle and crinkle on the leaves of infected tobacco and N. benthamiana, but did not infect squash or cucumbers indicating that it is a distinct isolate from the cucurbit isolate. These results indicate that this isolate may be considered to be a third serogroup.

Seedlings derived from seven R₀ lines were germinated on kanamycin medium and inoculated with the above TSWV isolates. Infectivity data were recorded daily starting seven days after inoculation. Plants inoculated with TSWV-BL, Arkansas, 10W pakchoy or Brazil isolates were rated susceptible if virus symptoms were observed on uninoculated leaves. Plants inoculated with the Begonia isolate were rated susceptible if local lesions were observed on inoculated leaves, since this isolate does not cause systemic infection in tobacco. All inoculated control NP(-) R₁ plants were susceptible to infection by these five isolates. They were systemically infected 12 days after inoculation except that transgenic R₁ plants inoculated with Begonia produced only local lesions on the inoculated leaves. However, almost all NP(+) R₁ plants were highly resistant to the homologous
isolate TSWV-BL, while much lower percentages of NP(+) R₁ plants were resistant to heterologous isolates Arkansas, 10W pakchoy and Begonia. On the other hand, all NP(+) R₁ plants from the seven transgenic lines were susceptible to the Brazil isolate, even though a slight delay (1 to 2 days) in symptom expression was observed in some of the high NP-expressing NP(+) R₁ plants from line NP(+).4.

Resistant R₁ plants remained symptomless throughout their life cycles. The inoculated leaves of seventeen symptomless NP(+) plants were checked for the presence of the virus by back inoculation on leaves of *Chenopodium quinoa* plants. No virus was recovered from the inoculated leaves of symptomless NP(+) plants, suggesting that the virus could not replicate or spread in these NP(+) plants.

The relationship between the level of NP accumulation in transgenic plants and the degree of resistance to heterologous TSWV isolates was also studied. Analysis of the data described above suggested that R₁ plants derived from R₀ lines with low levels of NP offered the best resistance to the heterologous isolates of the "L" serogroup (Arkansas and 10W pakchoy) while R₁ from a R₀ line with high level of NP were resistant to the Begonia isolate, which belongs to the "I" serogroup. For example, an average 76% of inoculated R₁ plants from low NP expressing lines NP(+) 2, 14, and 21 were resistant to infections by the Arkansas and 10W pakchoy isolates, while resistance to these isolates was observed in only 11% of similarly inoculated plants from high NP expressing lines NP(+).4, 9, and 23. On the other hand, the Begonia isolate infected 79% of R₁ plants from the low NP expressing line NP(+).2, 14, and 21 but only 19% from high NP expressing line NP(+).4.

Therefore, it was concluded that the transgenic R₁ plants expressing low levels of the NP gene were highly resistant to infection with the isolate 10W pakchoy (the "L" serogroup), but not to Begonia isolate (the "I" serogroup). In contrast, the highly NP-expressing R₁ plants were very resistant to infection by Begonia isolate but not to infection by the isolate from 10W pakchoy.

Thus, it was of interest to accurately quantitate the relation of NP expression in individual plants with resistance to the heterologous
isolates. In a number of inoculation experiments reported herein, leaf samples of transgenic plants were taken before inoculating with the Arkansas and 10W pakchoy isolates. Samples were also taken from non-inoculated leaves of plants inoculated with the Begonia isolate after observations of the apparent relation between NP expression levels and resistance were made. The latter method of sampling could be done without interference from infection by the Begonia isolate because this isolate does not cause systemic infection in tobacco nor reacts with antibodies to the TSWV-BL NP. All samples were assayed for relative NP levels by DAS-ELISA using antibodies raised to isolated NP of TSWV-BL. Figures 5 and 6 show the relation between NP levels in transgenic R1 plants (irrespective of the R0 lines they came from) and their resistance to the Arkansas and 10W pakchoy isolates or to the Begonia isolate. Nearly all transgenic R1 plants with very low or undetectable ELISA reactions (0-0.05 OD_{405nm}) were resistant to infections by the Arkansas and 10W pakchoy isolates (the "L" serogroup) but susceptible to the Begonia isolate (the "I" serogroup). In contrast, almost all R1 plants that gave high ELISA reactions (0.4-1.0 OD_{405nm}) were resistant to the Begonia isolate but susceptible to the Arkansas and 10W pakchoy isolates.

The double-stranded (ds) RNA was isolated from the N. benthamiana plants infected with TSWV-B using a combination of methods [See Acta Horticulturae 186:51 (1986), and Can. Plant Dis Surv 68:93(1988)] which have been successfully used for isolation of dsRNA from tissue infected with grapevine leafroll virus. The dsRNA was chosen for the cDNA synthesis since isolation of the virus particle from this isolate has not been possible [see Plant Disease 74:154 (1990)]. In order to make a cDNA library specific to the S RNA of TSWV-B, the double stranded S RNA was gel-purified, denatured by methyl-mercury treatment, and subjected to cDNA synthesis procedure provided by Promega using random primers. The synthesized cDNA fragments were cloned via an EcoRI adaptor into the EcoRI digested λ ZAPII (Strategene), and positive clones were identified by colony hybridization using the cDNA probes prepared by reverse transcription of gel-purified S RNA. Dozens of positive clones were analyzed on
agarose gels and only three overlapping clones containing the largest inserts (L1, L22 and L30) were selected (see figure 3), covering nearly entire TSWV-B S RNA.

The nucleotide sequences of the inserts in clones L1, L22 and L30 were determined from both DNA strands, first by the universal and reverse primers and then by the internal primers designed for sequencing the S RNA of TSWV-B. Sequencing was done using the Sanger dideoxyribonucleotide method, T7 polymerase (U.S. Biochemicals, Sequenase TM), and the double-stranded sequencing procedure described by Siemieniak [see Analyt. Biochem. 192:441 (1991)]. The sequence analyses of these clones revealed inserts of 1.994 kb, 2.368 kb and 1.576 kb, respectively, and these sequences represented 93% of the S RNA genome (see figure 3). The assembled sequence was analyzed by comparisons with sequences of TSWV isolates CONH1, L3, L, and BL using computer programs available from the Genetics Computer Group (GCG, Madison, WI).

Computer analysis showed that the assembled sequence of 2.842 kb covered the complete 52 K nonstructural protein gene, the complete intergenic region (629 bp), and 737 bp of the NP gene (only 39 N-terminal nucleotides of the N gene were not represented). In order to obtain this missing region of the N gene, a primer TTCTGGTCTTCTTCAAACTCA, identical to a sequence 62 nucleotides from the initiation codon of the N gene, was end-labeled with polynucleotide kinase to screen the cDNA library described above. Five putative clones were obtained. Sequence analysis of the five clones showed that only clones S6 and S7 contain these 39 missing nucleotides of the N gene. The latter clone also included the extreme 3' end of the S RNA.

The 5' extreme end of the S RNA was obtained using the 5' RACE System (GIBCO). Both ssRNA of TSWV-B and total RNAs isolated from tobacco plants infected with TSWV-B were used to synthesize first strand cDNA with an oligonucleotide (5'-CTGTAGCCATGAGCAAAG) complementary to the nucleotide positons 746-763 of te TSWV-B S RNA. The 3'-end of the first strand cDNA was tailed with dCTP using terminal deoxynucleotidyl transferase. Tailed cDNA was then amplified
by PCR using an anchor primer that anneals to the homopolymeric tail, and an oligonucleotide (5'-TTATATCTTCTTCT TAGGA) that anneals to the nucleotide positions 512-529 of the TSWV-B S RNA. The PCR-amplified fragment was gel-purified and directly cloned into the T-vector pT7Blue (Novagen) for sequence analysis. Eight independent clones were sequenced with an oligomer primer (5'-GTTCTGAGATTGTGCTAGT) close to the 5' region of the S RNA (nucleotide positions 40-57 of the TSWV-B S RNA). Six of the resulting clones contained the 5' extreme end of the S RNA and the 5'-terminal nucleotide sequence from these clones was identical. Thus, the complete nucleotide sequence of the TSWB-B S RNA is 3049 nucleotides in length.

Thus these two clones together with the three clones previously sequenced (L1, L22, L30, S6 and S7) covered a total of 3032 nucleotides depicted above. Comparisons with the terminal sequences of TSWV-CPNH1 and TSWV-I revealed that although the extreme 5' end of 18 nucleotides was not represented in the assembled sequence, the extreme 3'-terminus of the TSWV-B S RNA is identical to the extreme 3' end of the TSWV-I S RNA and is only one out of fifteen nucleotides different from the extreme 3' end of TSWV-CPNH1. The conservation of the terminal sequence among TSWV isolates is consistent with observations of the other members of Bunyaviridae genera, and supports the hypothesis that the terminal sequences might form stable base-paired structure, which could be involved in its replication and encapsulation.

The complete nucleotide sequence of the S RNA genome of TSWV-B (the Brazilian isolate discussed above) according to the present invention is:

\[
\begin{align*}
AGAGCAATTC & GTCTATTTTT TATTTCAAAT CGAACCTCAA CTACACAATC & 50 \\
TCACAPACTT & AATAGACA CAGACACAAG ACCACAAATG TCAATAGGTT & 100 \\
TTTTAATTC & CATATTCAG ACAAAGCCTTT CAGTTGCGG ATGACAGCA & 150 \\
TCTGTAPACT & CCACTCTGCA TATCCACGG ATTTACAGT TIOCACTCAG & 200 \\
TIOCACTCAG & CGACATAGTAC AGTGTGCTG TCAACAGGAC ACAACAGTA & 250 \\
CCTCSCSCPA & CACTCAAAA ATATACGATA TIOCATCAGT AGACMAAGGAA & 300 \\
ATTTATACG & ACACTCTCA TATCCTCAG TIGTACGATA TIGTIGTICAG & 350 \\
\end{align*}
\]
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Counter</th>
</tr>
</thead>
<tbody>
<tr>
<td>CATCAATATC AATGATTC GTCCGCACT TCTGTTTTCT TCAAGAAGAC</td>
<td>400</td>
</tr>
<tr>
<td>TTACACCAA TGAGTCAAG CATGAGGCT ATCTAAAAGT TCTTCTCCTT</td>
<td>450</td>
</tr>
<tr>
<td>GGGGATAG TGCTTTGGA AAGGGTGAG AGAGGGTGAG AGGTTTGAGA</td>
<td>500</td>
</tr>
<tr>
<td>CAGATTCCG CTCCAAGAG CAGATATAAT TCTGATGAC GATAATATAAT</td>
<td>550</td>
</tr>
<tr>
<td>CAGTTCTPA CAAGGCTAC CTCTCTCTG TCAAGAACAA TACTTACAACAA</td>
<td>600</td>
</tr>
<tr>
<td>GTCGAATGA GCCACATCG GCTTTAGGC AAGATGAAT GCTTTCTCC</td>
<td>650</td>
</tr>
<tr>
<td>TAAAGCTATG GTCTCATGAG CAGGTAGTAG TCAAGAACAA AATTCTACAC</td>
<td>700</td>
</tr>
<tr>
<td>AGATGCAAAG TATAACAGA ATCTGGAAT CTCCTCGAG CTATATTCTT</td>
<td>750</td>
</tr>
<tr>
<td>CAGATGGTA CAGAACAACAA CATTAGGCCT AACCTCACAAG CTTTTGTTAA</td>
<td>800</td>
</tr>
<tr>
<td>AGCCTTGATT GATTTCTATG TTAAGTCTG GCTTGGCCTG AGGTTTGCAA</td>
<td>850</td>
</tr>
<tr>
<td>AAGTTTGGA GCAATAGCG ATACAGAACG TCTTCAGAT TCGAGGAGAC</td>
<td>900</td>
</tr>
<tr>
<td>GAAAAGATGA AAGATTCTAC TTTCTTGAGT GAAATGAGGC CAAATACAA</td>
<td>950</td>
</tr>
<tr>
<td>CAGTGAAGAA ACAGCTTTAA ATGCCTAGCT TATAGAAGAG CACAGACCTVCG</td>
<td>1000</td>
</tr>
<tr>
<td>CATATCCPA GTCGAAGATG CCTTTGAGAT TATAATGATG TTTCTGAGAT</td>
<td>1050</td>
</tr>
<tr>
<td>CTCAGAGCG CTTACAACAC TGTCATGAT CTCCATATTC CTAAAGGAT</td>
<td>1100</td>
</tr>
<tr>
<td>TGTTCAGCT TCTCTTGACA CTCACGCTG CTTCCGAAAC TCTTCGCA</td>
<td>1150</td>
</tr>
<tr>
<td>AAGGGCTGCA AGAGATGGAG ATCATATATA CTATAAAGAC CCGAGACTA</td>
<td>1200</td>
</tr>
<tr>
<td>ACCCGACGA ACCGGGATCT AGGGAAGAA ACCGAGACT AGAGTTGAGA</td>
<td>1250</td>
</tr>
<tr>
<td>TCTACGAGA AAGAAGTATT TCTCTTCAAA AAGAAGGAA AGCTCTGAG</td>
<td>1300</td>
</tr>
<tr>
<td>TAAATGGGA GAAGAGCTCT TATTCTGATA GCCATCGAT TCTTCGAGG</td>
<td>1350</td>
</tr>
<tr>
<td>AAGTGATGGT TCTTCGAGG AGAGATGCA ATCTCTCTGT ATCTTACCC</td>
<td>1400</td>
</tr>
<tr>
<td>TATATACAGA TCCTTGCCG ACGGGAGTTT GACCAAGTG AAGAAGAAAGA</td>
<td>1450</td>
</tr>
<tr>
<td>AGTCCTTGG TAGGCTAGGC TCAAGCTGAT ATCTGAGATA AAGGCTGCTC</td>
<td>1500</td>
</tr>
<tr>
<td>ATACTACCTC AAGTGGTATT TGCAACTTGG CTTATCCGTTT ATGGITGGTTA</td>
<td>1550</td>
</tr>
<tr>
<td>TTTCTTTTAA AITCAAGGTA AGTTAGATTT CTAAGCTTTA CTTGCTTATA</td>
<td>1600</td>
</tr>
<tr>
<td>GCATTATAC AAAAAATCA AAAAAATACA AAAAAATACA AAAAAATACA</td>
<td>1650</td>
</tr>
<tr>
<td>AAAACACAAA AGAGCCCAA AAGGAGGAT TTTGTTGATT TACTCTGTTT</td>
<td>1700</td>
</tr>
<tr>
<td>TAGGCTATAG TACGCTCTT TGGTTGAGC AAAATAAGAT TGATAAATGC</td>
<td>1750</td>
</tr>
<tr>
<td>ATTACTGGA ATTAAAAGTC TCTAAAAGAG TTTCAAGAGA CAGCTTGGCC</td>
<td>1800</td>
</tr>
<tr>
<td>AAAATGGTTA TTTTGTGTTT TTTTTTGT TTGTTTTATT TTTTTATT</td>
<td>1850</td>
</tr>
<tr>
<td>TTTATTTTTA GTTATTTTTT TGTGTTTGT TTTTTATT TTTTTATT</td>
<td>1900</td>
</tr>
<tr>
<td>TCTTTTTTTT TTTTATTAT ATATATCAA CAGAACAC ACAAATTATT</td>
<td>1950</td>
</tr>
<tr>
<td>TTAATTTGGA AGATTCTACT GATTTAAGAC ACTAGCGCTG ACTTTATGAC</td>
<td>2000</td>
</tr>
<tr>
<td>ACTTAAAGAG CTTAGCTTGG CTTAACACA CIGAAGTGAA TAAAACACAA</td>
<td>2050</td>
</tr>
<tr>
<td>CTTAGTATTA TGGATCTCCTT AAMTAACACA CTTTATTTAT ATGGATCTCT</td>
<td>2100</td>
</tr>
</tbody>
</table>
GAATCAGCCT TAAAGAAGCT TTTATGGCA ACCAGGAATC TGGGCTCTT 2150
TCCTTACCTT AACCATTTCA TAGAATTGT CAGATTATC ACTGTAATAG 2200
TGCATAGCAA TCTCCTCCCT AGCAATTGGG TIGCAAGAAC TAAATATCCTT 2250
GGCATTATCT TCCCITITGT TTTATGTCG ATGATACATT GAAATATCCTT 2300
TGCTTTTAAG CACTGTCGAA ACCTTTCCCA GAGCTTCCTT AGTTGTTGAC 2350
TTTATTCCTT CAACTCCTAA CTCCTGTATC TTTATCGAT TAAATATGAC 2400
AAGAAGACAA CTGATCCTCT CGAGAGCTGC AACAGAAAGCA ATGAGAGGGA 2450
TACTACCAAC AGCAATATCA GCAAGCTCA CAGATTITGC AITGCAAGGCA 2500
GCCAGGCCTT AAGCTTTTGAC AAAGAGTGG GAGCAGATT TAHCTTTTGAT 2550
AATAGCAGAA TTTCTATGTT TGGCATGCTC TTCTAATGAC TTTACTTCTA 2600
TCATGTACAT AACGCTCTTG AAGAGCATAT CCTAAAGTCC AAACCTTCTCA 2650
GAATTTTATC TTTATGTCAG CTATCCAAA AGTAATATAC TTCTGTTTAC 2700
AACCTTATCA TGGCTCTGGC GATCTTCCAA GAAIGCAAA CATGAAAGTCA 2750
TACTCAATTT CTTATACAGG TACGAATTCT CCGGCAAGGA AGCITTTAAGG 2800
TTGAAGCCGA CCGTGTCTTG GTTCTGCTCA AACTCAACAT CIGCAGAATG 2850
AGTAAAGAGA GAGACAATGT TTCTCTTATT GAGCCTGACC TTAGACATGG 2900
TGGCACTTTA GATCTGAGCC TTCTCGAGA GATAAGATTC AAGCTGAGAA 2950
AGTCAACAC TGCTACCGCC GTGCTTTACT TATTCGATT AATGATGATG 3000
TTGTTATTGCT GAGTATTAGG TTTTTTATAA AAATGGCAGC AATGACTCT 3049

The deduced amino acid sequences of the nonstructural (single underlined above) and nucleocapsid proteins according to the present invention are:

Met Ser Ser Gly Val Tyr Glu Ser Ile Ile Gln Thr Lys Ala Ser
   5
Val Trp Gly Ser Thr Ala Ser Gly Lys Ser Ile Val Asp Ser Tyr
   20
Trp Ile Tyr Glu Phe Pro Thr Gly Ser Pro Leu Val Gln Thr Gln
   35
Leu Tyr Ser Asp Ser Arg Ser Lys Ser Ser Phe Gly Tyr Thr Ser
   50
Lys Ile Gly Asp Ile Pro Ala Val Glu Glu Glu Ile Leu Ser Gln
   65
Asn Val His Ile Pro Val Phe Asp Asp Ile Asp Phe Ser Ile Asn
   80
Ile Asp Ser Phe Leu Ala Ile Ser Val Cys Ser Asn Thr Val
   95
Asn Thr Asn Gly Val Lys His Gln Gly His Leu Lys Val Leu Ser
110
<table>
<thead>
<tr>
<th>Residue</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu Ala Gln Leu His Pro Phe Glu Pro Val Met Ser Arg Ser Glu</td>
<td>125 130 135</td>
</tr>
<tr>
<td>Ile Ala Ser Arg Phe Arg Leu Gln Glu Asp Ile Ile Pro Asp</td>
<td>140 145 150</td>
</tr>
<tr>
<td>5 Asp Lys Tyr Ile Ser Ala Ala Asn Lys Gly Ser Leu Ser Cys Val</td>
<td>155 160 165</td>
</tr>
<tr>
<td>Lys Glu His Thr Tyr Lys Val Glu Met Ser His Asn Gln Ala Leu</td>
<td>170 175 180</td>
</tr>
<tr>
<td>Gly Iys Val Asn Val Leu Ser Pro Asn Arg Asn Val His Glu Trp</td>
<td>185 190 195</td>
</tr>
<tr>
<td>Leu Tyr Ser Phe Lys Pro Asn Glu Asn Ile Glu Ser Asn Asn</td>
<td>200 205 210</td>
</tr>
<tr>
<td>Arg Thr Val Asn Ser Leu Ala Val Lys Ser Leu Leu Met Ala Thr</td>
<td>215 220 225</td>
</tr>
<tr>
<td>15 Glu Asn Asn Ile Met Pro Asn Ser Gln Ala Phe Val Lys Ala Ser</td>
<td>230 235 240</td>
</tr>
<tr>
<td>Thr Asp Ser His Phe Lys Leu Ser Leu Gln Leu Arg Ile Pro Lys</td>
<td>245 250 255</td>
</tr>
<tr>
<td>Val Leu Iys Gln Ile Ala Ile Gln Lys Leu Phe Lys Phe Ala Gly</td>
<td>260 265 270</td>
</tr>
<tr>
<td>Asp Glu Thr Gly Lys Ser Phe Tyr Leu Ser Ile Ala Cys Ile Pro</td>
<td>275 280 285</td>
</tr>
<tr>
<td>Asn His Asn Ser Val Glu Thr Ala Leu Asn Val Thr Val Ile Cys</td>
<td>290 295 300</td>
</tr>
<tr>
<td>25 Arg His Gln Leu Pro Ile Pro Lys Ser Lys Ala Pro Phe Glu Leu</td>
<td>305 310 315</td>
</tr>
<tr>
<td>Ser Met Ile Phe Ser Asp Leu Lys Glu Pro Tyr Asn Thr Val His</td>
<td>320 325 330</td>
</tr>
<tr>
<td>Asp Pro Ser Tyr Pro Gln Arg Ile Val His Ala Leu Leu Glu Thr</td>
<td>335 340 345</td>
</tr>
<tr>
<td>30 His Thr Ser Phe Ala Gln Val Leu Cys Asn Lys Leu Gln Glu Asp</td>
<td>350 355 360</td>
</tr>
<tr>
<td>Val Ile Ile Tyr Thr Ile Asn Ser Pro Glu Leu Thr Pro Ala Lys</td>
<td>365 370 375</td>
</tr>
<tr>
<td>35 Leu Asp Leu Gly Glu Arg Thr Leu Asn Tyr Ser Glu Asp Ala Ser</td>
<td>380 385 390</td>
</tr>
<tr>
<td>Lys Lys Lys Tyr Phe Leu Ser Lys Thr Leu Glu Cys Leu Pro Val</td>
<td>395 400 405</td>
</tr>
<tr>
<td>Asn Val Gln Thr Met Ser Tyr Leu Asp Ser Ile Gln Ile Pro Ser</td>
<td>410 415 420</td>
</tr>
<tr>
<td>Trp Lys Ile Asp Phe Ala Arg Gly Glu Ile Ile Arg Ile Ser Pro Gln</td>
<td>425 430 435</td>
</tr>
<tr>
<td>Ser Thr Pro Ile Ala Arg Ser Leu Leu Lys Leu Asp Leu Ser Lys</td>
<td>440 445 450</td>
</tr>
</tbody>
</table>
Ile Lys Glu Lys Lys Ser Leu Thr Trp Glu Thr Ser Ser Tyr Asp  
455 460 465  
Leu Glu;  
and  

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

As the nucleocapsid protein gene depicted above is on the viral  
complementary strand, the nucleocapsid protein gene of TSWV-B is:  

ATG TCT AAG GTC AAG CTC ACA AAA GAA AAC ATT GTC TCT CIT TTA  45  
ACT CAA TCT GCA GAT GTT GAG TTT GAA GAA GAC CAG AAC CAG GTC  90
GCA TTC AAC TTT AAG ACT TTC TGC CAG GAA AAT CTT GAC CTG ATT 135
AAG AAA ATG AGT ATC ACT TCA TGT TTG ACT TTC TTG AAG AAT CGC 180
CAA GGC ATT ATG AAA GTT GTG AAC CAA AGT GAT TTT ACT TTT GGT 225
AAG GTC AGC ATA AAG AAA AAT TCT GAA AGA GTC GTA GCT AAG GAT 270
5 ATG ACT TTC AGG AGG CTT GAT AGC ATG ATA AGA GTG AAG CTC ATA 315
GAA GAG ACT GCA AAC AAT GAG AAT CTT GCT ATT ATC ATA AAA GCA AAA 360
ATT GCC TCC CAC CCT TGG GTC CAA GCT TAC GGG CTG CCT CTG GCA 405
GAT GCA AAA TCT GTG AGA CTT GCT ATA ATG CTT GGA GGT AGT ATC 450
CCT CTC ATT GCT TCT GTT GAC AGC TTC GAG ATG ATC AGT GTT GTT 495
10 CTT GCC ATG TAT CAA GAT GCA AAG TAC AAG GAG TTA GGG ATT GAA 540
CCA ACT AAG TAC AAC ACT AAG GAA GCT CTG GGG AAG GGT TGC ACA 585
GTG CTT AAA AGC AAA GGA TTT ACA ATG GAT GAT GCA CAG ATA AAC 630
AAA GGG AAA GAA TAT GCC AAG ATG CTT AGT TCT TGC AAT CCC AAT 675
GCT AAG GGA AGC ATT GCT ATG GAC TAT TAC AGT GAT AAT CTG GAC 720
15 AAA TTC TAT GAA ATG TTT GGA GTC AAG AAA GAG GCC AAG ATT GCT 765
GTT GTC GCA TAA 777

The complete S RNA of TSWV-B should be 3049 nucleotides in length, 134 nucleotides longer than S RNA of TSWV-CPNH1. This difference was mainly attributed to the elongated intergenic region of the TSWV-B S RNA. Analysis of the sequenced region of TSWV-B S RNA revealed two open reading frames as depicted above, which is similar to other TSWV isolates. The larger one was localized on the viral RNA strand originating at nucleotide 88 and terminating at nucleotide 1491. The smaller one on the vial complementary strand was defined by an initiation codon at nucleotide 2898 and a termination codon at nucleotide 2122. The two open reading frames were separated by an intergenic region of 629 nucleotides. Comparisons of the entire sequenced TSWV-B S RNA with S RNA regions of other isolates in the following table which depicts the percent homology comparison of aligned nucleotide and amino acid sequences of the TSWV-B S RNA with those of the other isolates:
<table>
<thead>
<tr>
<th>gene</th>
<th>Overall</th>
<th>53 K protein gene</th>
<th>Intergenic</th>
<th>29 K protein</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nt</td>
<td>nt</td>
<td>aa</td>
<td>nt</td>
</tr>
<tr>
<td>B/CPNH1</td>
<td>76.4&lt;sup&gt;b&lt;/sup&gt;</td>
<td>80.0</td>
<td>86.1(78.3)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>72.4</td>
</tr>
<tr>
<td>B/L3</td>
<td>75.8</td>
<td>79.0</td>
<td>89.0(82.0)</td>
<td>76.4</td>
</tr>
<tr>
<td>B/BL</td>
<td>76.3</td>
<td>-</td>
<td>-</td>
<td>72.8</td>
</tr>
<tr>
<td>B/I</td>
<td>63.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CPNH1/L3</td>
<td>94.8</td>
<td>95.6</td>
<td>92.0(89.4)</td>
<td>89.2</td>
</tr>
<tr>
<td>CPNH1/BL</td>
<td>96.4</td>
<td>-</td>
<td>-</td>
<td>95.9</td>
</tr>
<tr>
<td>B/L3</td>
<td>75.8</td>
<td>79.0</td>
<td>89.0(82.0)</td>
<td>76.4</td>
</tr>
<tr>
<td>B/BL</td>
<td>76.3</td>
<td>-</td>
<td>-</td>
<td>72.8</td>
</tr>
<tr>
<td>B/I</td>
<td>63.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CPNH1/L3</td>
<td>94.8</td>
<td>95.6</td>
<td>92.0(89.4)</td>
<td>89.2</td>
</tr>
<tr>
<td>CPNH1/BL</td>
<td>96.4</td>
<td>-</td>
<td>-</td>
<td>95.9</td>
</tr>
</tbody>
</table>

- **a** The partial or complete S RNA sequences of isolates TSWV-CPNH1 (2.916 kb), TSWV-L3 (2.837 kb), TSWV-BL (2.037 kb) and TSWV-I (1.144 kb) were used for comparisons with the S RNA sequence of the TSWV-B (3.049 kb).
- **b** Percent similarities were calculated by Comparison of their nucleotide or predicted amino acid sequence using the program BESTFIT of the GCG Sequence analysis software package.
- **c** Percent identity is in parenthesis.

As depicted, the greatest nucleotide sequence similarity (75.8%-76.4%) was shown with the L-type isolates (CHNH1, L3 and BL). To the lesser extent, there was nucleotide sequence similarity (63%) between the TSWV-B S RNA and the S RNA of TSWV-I assigned to I serogroup.

For comparison, the sequenced S RNA regions of the L-type isolates (CHPN1, L3 and BL) shared 94.8%-96.4% nucleotide sequence similarities.

The open reading frame of 777 nucleotides encodes the N protein of 258 amino acids with a predicted molecular weight of 28700 Da. The sequence comparisons of the N open reading frame from TSWV isolates revealed that nucleotide sequences of the N genes from the isolates CPNH1, L3 and BL differs from TSWV-B by a considerably larger amount (22%-22.5%) than they differ from each other (2.7%-3.2%). Consistent to the results of the immunological analysis, the N amino acid sequences among CPNH1, L3 and BL isolates are more closely related to each other (98.8%-99.6% similarities or 96.9%-98.5% identities) than to
the TSWV-B (90.3%-91.5% similarities or 79.1%-79.9% identities). Much lower homology was observed to TSWV-I at both nucleotide (63.1%) and amino acid (69.7% similarity or 55.3% identity) levels. Except for the N open reading frame of TSWV-I that encodes 262 amino acids, the N open reading frames of the other isolates code for the 258 amino acids. Computer analysis suggested that the extra residues of TSWV-I N open reading frame resulted from the amino acid sequence insertions (residues 82 through 84 and residue 116). One potential N-glycosylation site is found at residue 68.

The second open reading frame of 1404 nucleotides encodes the nonstructural protein of 467 amino acids with a predicted molecular weight of 52566 Da. Comparisons with homologous open reading frames of TSWV-CPNH1 and TSWV-L3 showed 80% and 79% similarities at the nucleotide level, and 86.1% (or 78.3% identity) and 89% (or 82.0% identity) similarities at the amino acid level. This open reading frame contains four potential glycosylation sites, which are located in the exactly same positions as those of TSWV-CPNH1 and TSWV-L3.

The intergenic region of the TSWV-B S RNA was, due to several insertions, 126 and 41 nucleotide longer than the counterparts of TSWV-CPNH1 and TSWV-L3, respectively. The sequence analysis by the program FOLD indicated the intergenic region can form very complex and stable hairpin structure by internally base-pairing U-rich stretches with A-rich stretches of the intergenic region, which had similar stability to those produced from TSWV-CPNH1 and TSWV-L3 as indicated by minimum free energy values. This internal base-paired structure may act as a transcription termination signal.

The results tabulated above also revealed that the N protein of TSWV-B is subject to a higher degree of selective pressure than the 52 K protein; the similarities among the amino acid sequences of the 52 K protein are lower than that found for the amino acid sequence of the NPs. Nucleotide sequence divergence is highest among the intergenic regions, which indicates that this region is subject to less selective pressure than either genetic region.

The evolutionary relationships among the TSWV-B and other four TSWV isolates were analyzed and depicted in figure 4 in which the
evolutionary tree organization is consistent with the relatedness of serological data collected for these TSWV isolates. Thus, the TSWV-B, according to the present invention, is more closely related to the L-type isolates than to the I-type isolate TSWV-I, but is much less similar to the L-type isolates than the L-type isolates are to each other.

Despite a slight delay of symptom expression, transgenic plants did not show resistance to the Brazil isolate of TSWV; Serological results show that this isolate is distinct from the "L" and "I" type isolates, and biologically different from the curcubit isolate. The Brazil isolate may thus belong to still another serogroup of TSWV. In any event, infectivity results show that it is unlikely that a single NP gene will provide resistance to all isolates in the Tospovirus genus.

Transgenic plants according to the present invention that gave low or undetectable ELISA reactions (0-0.05 OD_{405nm}) were resistant to infection by the heterologous isolates (Arkansas and 10W pakchoy) of the "L" serogroup, whereas no protection against these isolates was found in plants accumulating high levels of the NP. Compared to the ELISA readings of control NP(-) plants (0.05 OD_{405nm}), these transgenic plants may produce little, if any, TSWV-BL NP. Similar results have been observed in transgenic plants, in which the CP accumulation was not detected; these were highly resistant to virus infection. The mechanism underlying this phenomenon is presently unknown. It is likely that this type of resistance might be attributed to interference of CP RNA molecules produced in transgenic plants with viral replication, presumably by hybridizing to minus-sense replicating RNA of the attacking virus, binding to essential host factors (e.g., replicase) or interfering with virion assembly.

It should be noted, however, that the resistance to the homologous TSWV-BL isolate is apparently independent of the expression levels of the NP gene. Although the relative NP levels of the individual R1 plants inoculated with TSWV-BL were not measured, it is reasonable to assume that the NP produced in these inoculated R1 plants (a total of 145 plants tested) ranged from undetectable to high.
In contrast to the case for protection against the heterologous isolates of the "L" serogroup, protection against the Begonia isolate of the TSWV-I serogroup was found in the high NP-expressing R₁ plants. Comparison of NP nucleotide sequence of the "L" serogroup with that of the "I" serogroup revealed 62% and 67% identity at the nucleotide and amino acid levels, respectively. The difference of NP genes of the two serogroups might be so great that the NP (the "L" serogroup) produced in transgenic plants acted as a dysfunctional protein on the attacking Begonia isolate of the "I" serogroup. Incorporation of this "defective" coat protein into virions may generated defective virus which inhibit virus movement or its further replication. This type of interaction is expected to require high levels of the NP for the protection. Alternatively, resistance to the Begonia isolate may also involve interference of NP transcripts produced in R₁ plants with viral replication. If this is true, more NP transcripts (due to the heterologous nature of two NP gene) may be required to inhibit replication of heterologous virus.

Although there are no obvious explanations for the results showing the relation of NP levels in individual R₁ plants to resistance to the heterologous isolates of the "L" and "I" serogroups, it is believed these are definite trends since the data were derived from a large number (190) of plants. Thus, it is believed that a measurement of CP or NP levels in individual plants may provide a more accurate way to relate NP or CP levels to resistance. By this form of data analysis, the results show that the resistance was more closely related to NP levels in each test plant than to the NP level of the R₀ line from which they were derived. For TSWV-BL Np gene in tobacco, at least, it appears that integration sites of the NP gene in plant chromosome may not be important for viral resistance.

Studies have also been conducted to determine the reaction of transgenic R₁ and R₂ tomatoes containing the nucleocapsid protein gene of TSWV-BL according to the present invention to the following isolates: Brazil (a distantly related virus), T91 (a closely related virus) and BL (a homologous isolate). In these studies, transgenic tomatoes (L. esculentum) were produced by A. tumefaciens-mediated gene transfer
of the nucleocapsid protein (N) gene of the lettuce isolate of tomato spotted wilt virus BL into germinated cotyledons using modifications of published procedures [see Plant Cell Reports 5:81 (1986)]. The tomato line "Geneva 80" was selected for transformation because it contains the Tm-22 gene which imparts resistance to TMV, thus creating the possibility of producing a multiple virus-resistant line.

Transformants were selected on kanamycin media and rooted transgenic tomatoes were potted and transferred into the greenhouse. R₁ and R₂ tomato seedlings expressed the NPT II gene, suggesting multiple insertions of this gene in the plant genome. In contrast, only 18% of the seedlings produced detectable levels of the N protein.

Nine R₁ and three R₂ lines were tested for resistance to the following three Tospovirus described, specifically TSWV-BL, TSWV-T91, and TSWV-B. Infectivity was based upon visual inspection of test plants. In those cases where plants appeared healthy except for a few rust-colored rings or insect damage, extracts from these plants were inoculated to N. benthamiana to test for the presence of the virus. As depicted in the following table, nearly all control tomato plants exhibited typical symptoms consisting of plant stunting, leaf yellow mosaic and rugosity 3 to 4 weeks after inoculations with TSWV-BL, TSWV-T91 or TSWV-B. However, only 4% of the R₁ and R₂ transgenic plants became infected with TSWV-BL, 7% with TSWV-T91, and 45% with TSWV-B.
Viral resistance in transgenic R1 and R2 tomatoes expressing the nucleoprotein gene of the lettuce strain of tomato spotted wilt virus

<table>
<thead>
<tr>
<th>Inoculating Isolatesa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant Line</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>R1 Plants:</td>
</tr>
<tr>
<td>T13-1</td>
</tr>
<tr>
<td>T13-2</td>
</tr>
<tr>
<td>T13-3</td>
</tr>
<tr>
<td>T13-4</td>
</tr>
<tr>
<td>T13-9</td>
</tr>
<tr>
<td>T13-10</td>
</tr>
<tr>
<td>T13-11</td>
</tr>
<tr>
<td>T13-12</td>
</tr>
<tr>
<td>T13-13</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
</tr>
<tr>
<td>R2 Plants:</td>
</tr>
<tr>
<td>T13-1-7</td>
</tr>
<tr>
<td>T13-1-9</td>
</tr>
<tr>
<td>T13-1-11</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
</tr>
<tr>
<td>CONTROLS</td>
</tr>
</tbody>
</table>

a plants were inoculated at the one- to two-leaf stage with 5-, 10-, or 20-fold diluted leaf extract of N. benthamiana, H423 tobacco or tomato; the same plants were re-inoculated 7 days later and symptoms were recorded after another 14 days; the reaction is expressed as number of plants with symptoms/number of plants tested

b not tested

Accordingly, the description above supports the finding that transgenic tomato plants that express the N gene of TSWV-BL show resistance to infection to TSWV-BL, to other TSWV isolates that are closely related to TSWV-BL, and to the more distantly related TSWV-B.

In further limited studies with an additional isolate, all transgenic plants were resistant to the 10W (pakchoy) isolate, whereas the controls were infected. These results show that transgenic tomatoes are better protected against closely related isolates than distantly related isolates. Unlike in transgenic tobacco and N. benthamiana expressing the TSWV-BL N gene, the level of N protein expression did not correlate with the observed protection in transgenic
tomatoes; 55% of the transgenic tomatoes were also resistant to a distantly related isolate of TSWV-B, which was not observed in transgenic tobacco and N. benthamiana plants. These discrepancies may reflect that tomato is inherently less susceptible to Tospoviruses.

In addition, studies were also conducted to determine virus distribution in a small number of plants at 5 and 7 weeks after inoculation. The distal halves from leaflets of all expanded leaves of each plant were ground and back-inoculated onto N. benthamiana. The results taken seven days after inoculation showed that virus cannot be recovered from any leaf tissue of asymptomatic transgenic plants inoculated with either TSWV-BL, -T91, or -B, confirming the visual findings reported above. In transgenic plants showing symptoms, the virus is not distributed throughout the plant. For example, a transgenic plant which could not be conclusively rated visually contained the virus in only two of the 8 leaves; the second leaves from the bottom and top of the plant. Conversely, virus present in all leaves of the infected control plant, and is absent in those of the healthy control plants.

Graft inoculations were attempted to test whether the resistant transgenic plants could become infected if virus is introduced into the vascular system. R₁ and R₂ plants that had been inoculated at 1:5, 1:10 or 1:20 dilutions of TSWV-BL, -T91, or -B were grafted onto control plants infected with the same isolates and dilutions. The 34 transgenic plants were asymptomatic after 31 days, although the non-transgenic controls were infected. After 23 days, the top 46 cm of transgenic plants had been trimmed away to induce new growth and more plant stress. Although the young, vigorously growing new shoots failed to show any symptoms on the 31st day post inoculation, 33%, 31% and 45% of TSWV-BL, -T91 and -B were showing leaf or stem symptoms, respectively at 45 days post inoculation. These results indicate that some transgenic plants are tolerant, and others are immune to infection.

Thus, according to one aspect of the present invention, transgenic plants expressing the NP gene of the TSWV-BL isolate are highly resistant to infections of both the homologous TSWV-BL isolate and heterologous isolates of the same serogroup (Arkansas and 10W
pakchoy). More significantly, the resistance is effective to Begonia isolate from other serogroups. In brief, the above clearly describes that transgenic tobacco plants expressing the nucleoprotein gene of TSWV-BL display resistance to both TSWV and INSV, and the protection appears to be mediated by the nucleoprotein against distantly related INSV and by the nucleoprotein gene ribonucleotide sequence against the homologous and closely related TSWV isolates. This is the first time broad spectrum resistance of the engineered plants to different isolates of TSWV has been shown.

While coat protein protection generally displays delay and/or reduction in infection and symptom expression, but no immunity, the present invention provided a significantly high percentage of transgenic plants which were symptom-free and free of the infective virus. Resistance of these plants under greenhouse conditions persisted throughout their life cycle, and more importantly was inherited to their progenies as shown above.

It was observed in the present invention that the transgenic plants producing little, if any, TWSV-BL NP were highly resistant to infection by the homologous isolate and other closely-related isolates within the same serogroup of TSWV, whereas no protection was found in those expressing high levels of the NP gene.

The biological diversity of TSWV is well documented and has been reported to overcome the genetic resistance in cultivated plants such as tomato. Thus, it is extremely important to develop transgenic plants that show resistant to many strains of TSWV. The present invention indicates that one method to do so would be to utilize the viral NP gene to confer this resistance, and that this resistance would be to diverse TSWV isolates. Thus, the finding of the present invention that the expression of TSWV NP gene is capable of conferring high levels of resistance to various TSWV isolates has a great deal of commercial importance.

In another series of studies, Plasmid BiN19-N+ was constructed and transferred to A. tumefaciens strain LBA4404 in accordance with Example IV, and transferred to Nicotiana benthamiana in accordance
with Example V. The nucleocapsid genes of INSV-Beg and -LI were amplified with oligomer primers INSV-A
(5'-TACAGTTGATCCATGGTTATTTCAAAATAATTTATAAGC),
hybridizing to the 5'-coding and 3'-noncoding regions of the
nucleocapsid gene of an INSV isolate, respectively. The amplified
nucleocapsid gene fragments were purified in accordance with Example
III, and digested and sequenced in accordance with Example IV.

Of a total of 24 N+ (transformed with pBIN19-N+) and 18 N-
(transformed with vector pBIN19) transgenic N. benthamiana plants
were transferred to soil and grown in the greenhouse. All N+ lines were
confirmed by PCR at leaf stages 4-5 to contain the N gene sequence.
The relative level of N protein accumulation was estimated in each
independent R0 transgenic clonal line by DAS-ELISA using antibodies of
the TSWV-BL N protein. Of the twenty-four N+ lines, two had OD405nm
readings of 0.50-1.00, seventeen between 0.02-0.10, and the remaining
five less than 0.02. Healthy N. benthamiana or transgenic N- plants
gave OD405nm readings of 0.00-0.02. All the R0 plants were self-
pollinated and the seeds from the following transgenic lines were
germinated on kanamycin (300 µg/ml) selection medium for inoculation
tests: (1) N- -2 and -6, control transgenic lines containing vector
pBIN19 alone; (2) N+ -28, a transgenic line that produced an
undetectable amount of the N protein (OD405nm = 0.005); (3) N- -21, a
transgenic line producing a low level of the N protein (OD405nm =
0.085); and (4) N+ -34 and -37, two transgenic lines accumulating high
levels of the N protein (OD405nm = 0.50-1.00). These six lines were
then analyzed by Northern hybridization; the intensity of N gene
transcripts correlated well with the levels of ELISA reactions.

Transgenic seedlings from the six R0 lines were selected by
germinating seeds on kanamycin selection medium, and these seedlings
were inoculated with the five Tospoviruses. The inoculated R1 plants
were rated susceptible if virus symptoms were observed on
uninoculated leaves. In order to exclude the possibilities of escapes,
transgenic control N- plants were always used in each inoculation of
transgenic N+ plants. In addition, each inoculum extract was always
used to first inoculate N+ plants followed by control N- plants. The results from this series of studies are depicted below:

Reactions of R1 plants expressing the nucleocapsid (N) protein gene of N. benthamiana spotted wilt virus (TSWV) to inoculation with Tospovirus

<table>
<thead>
<tr>
<th>R0 Line</th>
<th>ELISA^a</th>
<th>TSWV ISOLATE</th>
<th>INSV ISOLATE</th>
<th>TSWV-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>N- -2/-6</td>
<td>&lt;0.02</td>
<td>32/32</td>
<td>32/32</td>
<td>20/20</td>
</tr>
<tr>
<td>N+ -28</td>
<td>0.005</td>
<td>16/16</td>
<td>16/16</td>
<td>15/16</td>
</tr>
<tr>
<td>N+ -21</td>
<td>0.085</td>
<td>9/40</td>
<td>17/40</td>
<td>39/40</td>
</tr>
<tr>
<td>N+ -34</td>
<td>0.715</td>
<td>25/28c</td>
<td>28/28</td>
<td>23/28c</td>
</tr>
<tr>
<td>N+ -37</td>
<td>0.510</td>
<td>26/28c</td>
<td>22/22</td>
<td>21/28c</td>
</tr>
</tbody>
</table>

^aELISA data of R0 lines from which the R1 plants were derived;

^b30-fold diluted leaf extracts of infected N. benthamiana plants were applied to the three leaves of plants at the 3-5 leaf stages. Each extract was always used to inoculate N+ plants followed by control N- plants. Data were taken daily for at least two months after inoculation and expressed as number of plants systemically infected/number of plants inoculated;

^cindicate that nearly all susceptible R1 plants displayed a significant delay of symptom appearance.

As depicted in the above table, all R1 plants from control lines N- -2 and -6 showed systemic symptoms 5-8 days after inoculation with all the viruses tested. None of the R1 plants from line N+ -28 produced detectable levels of the N protein, and all were susceptible to these viruses except for one plant inoculated with INSV-Beg. ELISA assays of leaf discs from this N+ -28 R1 plant sampled before inoculation clearly showed that the plant identified to possess the INSV-Beg resistant phenotype did accumulate a high level of the N protein (OD405nm = 0.78 as compared to OD405nm <0.02 for all other N+ -28 R1 plants).

The low N gene expressing line N+ -21 showed the best resistance against the homologous (78%) and closely related TSWV-10W (57%) isolates and very little resistance to the two INSV isolates (3% and 10%); only three N+ -21 plants showed the resistant phenotype when inoculated with the INSV isolates. Leaf samples from these INSV-resistant N+ -21 R1 plants gave much higher ELISA reactions (OD405nm 0.5 to 1.00) and thus higher amounts of the N protein than the
susceptible N^+21 plants (OD405nm 0.02 to 0.20). The high N gene expressing lines N^+-34 and -37 showed the highest resistance to INSV isolates (18%-25%) followed by the homologous TSWV-BL isolate (7% and 11%) while none of the plants showed resistance to TSWV-10W; however, the N^+-34 and -37 R1 plants that became infected with INSV or TSWV-BL did show various lengths of delays in symptom expression. None of the R1 plants from these four transgenic N^+ lines were resistant to TSWV-B; some of the R1 plants from the N^+-34 and -37 lines showed a slight delay of symptom appearance.

In studies to determine whether the level of N protein production in N^+ R1 plants was associated with resistance to different Tospoviruses, the inoculated N^+ R1 plants in the preceding table were re-organized into four groups based on the intensity of their ELISA reactions of tissues taken before inoculation irrespective of original R0 pants. The N^+ R1 plants that expressed low levels of the N protein (0.02-0.2 OD) showed high resistance (100% and 80%) to TSWV-BL and -10W but were all susceptible to INSV-Beg and -LI, showing no detectable delay in symptom expression relative to control N^- plants. In contrast, nearly all N^+ R1 plants with high levels of the N protein (0.20-1.00 OD) showed various levels of protection against TSWV-BL, INSV-Beg and -LI, ranging from a short delay of symptom expression to complete resistance with most of these plants showing various lengths of delay in symptom development relative to control N^- plants. No protection was observed in the high expressors against TSWV-10W. In addition, none of the N^+ R1 plants were resistant to TSWV-B regardless of the level of N gene expression; however, a short delayed symptom appearance was observed in the N^+ R1 plants producing high levels of the N protein. All control N^- R1 plants and transgenic N^+ R1 plants with undetectable ELISA reactions (0 to 0.02 OD) were susceptible to all the Tospoviruses tested.

The inhibition of replication of a distantly related INSV in N. benthamiana protoplasts expressing the TSWV-BL nucleocapsid gene was also studied. In these studies, whole INSV-LI virions were used to infect protoplasts that were isolated from three transgenic lines to investigate how the products of the transgene affect replication of the
incoming virus. Viral replication was determined by measuring the level of the N protein of the infecting INSV in transgenic protoplasts using antibodies specific to the INSV N protein. DAS-ELISA analysis showed that all progenies from a given line were relatively uniform and nearly all R1 progeny gave an expression level of transgenic N gene similar to their parental transgenic line. These results allowed for the prediction of the expression level of R1 populations based on that of their parental lines. Protoplasts derived from R1 plants of the low expressor line N+\textsuperscript{-}21 supported the replication of INSV-LI whereas protoplasts from R1 plants of the higher expressor line N+\textsuperscript{-}37 did not until 42 hours after inoculation at which low levels of viral replication were observed. The same protoplasts at various time intervals (e.g. 0, 19, 30 and 42 hours) were also assayed by DAS-ELISA using antibodies specific to the TSWV-BL N protein to monitor the expression level of the transgene. As expected, protoplast from N+\textsuperscript{-}21 R1 plants produced relatively low levels (0.338-0.395 OD405nm) whereas protoplasts from N+\textsuperscript{-}37 R1 plants accumulated high levels (0.822-0.865 OD405nm). The expression level was found to be consistent at all time points.

In this aspect of the present invention it has been shown that transgenic *N. benthamiana* plants that accumulate low amounts of the TSWV-BL N protein are highly resistant to the homologous and closely related (TSVV-10W) isolates, while plants that accumulate high amounts of this protein posses moderate levels of protection against both the homologous and distantly related (INSV-Beg and INSV-LI) viruses. More importantly, these findings indicate that transgenic *N. benthamiana* plants (a systemic host of INSV) are protected against INSV-Beg and INSV-LI isolates.

As discussed above, we have shown that transgenic plants expressing the N gene of TSWV are resistant to homologous isolates, and that such plants expressing the TSWV-BL N gene are resistant to both TSWV and INSV. It has also been shown the best resistance to homologous and closely related isolates was found in transgenic plants accumulating low levels of N protein while transgenic plants with high levels of TSWV-BL N protein were more resistant to serologically distant INSV isolates. This observation led us to suspect the role of
the translated N protein product in the observed protection against homologous and closely related isolates and to speculate that either the N gene itself which was inserted into the plant genome or its transcript was involved in the protection. To test this hypothesis transgenic plants containing the promoterless N gene or expressing the sense or antisense untranslatable N coding sequence were produced. What was discovered was that both sense and antisense untranslatable N gene RNAs provided protection against homologous and closely related isolates, and that these RNA-mediated protections were most effective in plants that synthesized low levels of the respective RNA species and appears to be achieved through the inhibition of viral replication.

More specifically, the coding sequences introduced into transgenic plants is shown in figure 7. As depicted, the construct pBiN19-N contains the promoterless N gene inserted into the plant transformation vector pBIN19 (see Example IV). All other constructs contain a double 35S promoter of CaMV, a 5'-untranslated leader sequence of alfalfa mosaic virus and a 3'-untranslated/polyadenylation sequence of the nopaline synthase gene. pBi525 is a plant expression vector and is used in this study as a control; pBi525-mN contains the mutant (untranslatable) form of the N gene; pBi525-asN contains the antisense form of the untranslatable N gene. One nucleotide deletion at the 5'-terminus of the mutant N gene is indicated by the dash symbol. ATG codons are underlined and inframe termination codons in the mutant gene are shown in bold.

**EXAMPLE VIII**

Primer-directed mutagenesis and cloning of the TSWV-BL N gene was conducted as follows:

Full-length N gene was obtained by reverse transcription and polymerase chain reaction as described in Phytopathology 82:1223 (1992), the disclosure of which is incorporated *in toto* herein. The untranslatable N coding sequence was similarly generated by RT-PCR using oligomer primers A (AGCATTGGATCCATGGTTAAACACACTAAGCAAGC), which is identical to the S RNA in the 3'-noncoding region of the TSWV-BL N gene, and B (AGCTATCTAGAACCATGGATGACTCACTAAGGAAAGCATTGTG),
complementary to the S RNA in the 5’-terminus of the N gene. The latter oligomer primer contains a frameshift mutation immediately after the translation initiation codon and several termination codons to block possible translation readthroughs. The intact and mutant N gene fragments were purified on a 1.2% agarose gel as described in Example II. The gel-isolated intact and mutant N gene fragments were digested with the appropriate restriction enzyme(s) and directly cloned into BamHI/XbaI-digested plant transformation vector pBIN19 and NcoI-digested plant expression vector pBI525, respectively as described in Example IV. The resulting plasmids were identified and designated as pBIN19-N containing the intact, promoterless N gene, and pBI525-mN and pBI525-asN containing the mutant coding sequence in the sense and antisense orientations, respectively, relative to cauliflower mosaic virus 35S promoter. The translatability of the mutant N coding sequence in the expression cassette was checked by transient expression assay in Nicotiana tabacum protoplasts; and the expression cassettes containing the sense or antisense mutant N coding sequence were then excised from plasmid pBI525 by a partial digestion with HindIII/EcoRI (since the N coding sequence contains internal HindIII and ExoRI sites), and ligated into the plant transformation vector pBIN19 that had been cut with the same enzymes. The resulting vectors as well as pBIN19-N were transferred to A. tumefaciens strain LBA4404 using the procedure described in Example IV. Leaf discs of N. tabacum var Havana cv 423 were inoculated with the A. tumefaciens strain LBA4404 containing various constructs and the resulting transgenic plants were self-pollinated and seeds were selectively germinated on kanamycin medium.

PCR was performed on each R0 transgenic line as described above. The oligomer primers A and B were used to determine the presence of the N coding sequence of TSWV-BL. The oligomer primer 35S-promoter (CCCCACATTCCTCGCAAGACCC) was combined with either the oligomer primer A or B to confirm the orientation (relative to the CaMV 35S promoter) of the mutant N coding sequence inserted into the plant genome. DAS-ELISA used to detect the N protein in transgenic plants was performed using polyclonal antibodies against the TSWV-BL
N protein. For an estimation of RNA transcript level in transgenic plants by Northern blot, total plant RNAs were isolated according to Napoli [see The Plant Cell 2:279 (1990)], and were separated on a formaldehyde-containing agarose gel (10 µg/lane). The agarose gels were then stained with ethidium bromide to ensure uniformity of total plant RNAs in each lane. Hybridization conditions were as described in the GeneScreen Plus protocol by the manufacturer. Resulting signal blots were compared and normalized based on the N gene transcript band of the control lane (the mN R1 plant producing a high level of the N gene transcript) included in each blot. The transgenic plants that gave density readings (Hewlet ScanJet and Image Analysis Program) between 100 and 150 were rated as high expressors, while the plants with densities between 15 and 50 were rated as low expressors.

Inoculation of transgenic plants with Tospovirus was carried out as described above with inoculation being done at the 3-4 leaf stage except were indicated.

Tobacco protoplasts were prepared from surface-sterilized leaves derived from R1 plants [see Z. Pflanzanphysiol. 78:453 (1992) with modifications]. The isolated protoplasts (6 x 10^6 protoplasts) were transformed with 0.68 OD260nm of the purified TSWV-BL virion preparation using the PEG method [see Plant Mol. Biol. 8:363 (1987)]. The transformed protoplasts were then cultured at the final density of 1 x 10^6 protoplasts /ml in the culture medium at 26°C in the dark.

After various intervals of incubation, the cultured protoplasts were washed twice with W5 solution and lysed by osmotic shock in the enzyme conjugate buffer. Viral multiplication (replication) was estimated by measuring the N protein of the virus using DAS-ELISA.

As described, one aspect of the present invention demonstrated that transgenic tobacco producing none or barely detectable amounts of the N protein were resistant to homologous and closely related isolates. This result suggested that the observed resistance may have been due to trans interactions of the incoming viral N gene RNA with either the N gene transcript produced in the transgenic plants or the N coding sequence itself. To test whether the presence of the nuclear N gene
plays a role, transgenic P0N R0 lines and R1 plants from two P0N lines were challenged with four *Tospoviruses* (TSWV-BL, TSWV-10W, INSV-Beg and TSWV-B). Only asymptomatic plants were rated resistant while plants showing any symptoms were rated susceptible. All inoculated R0 and R1 plants were susceptible to the viruses.

To further test the possibility that the transcript of the N transgene is involved in the protection, a number of R0 transgenic plants that produced either the sense or the antisense N gene transcript but not the N protein were inoculated with the homologous isolate.

Results appear in the following table:

<table>
<thead>
<tr>
<th>Form of transgenea</th>
<th>Level of N gene RNAb</th>
<th>No. of R0 lines tested</th>
<th>No. of lines inoculatedc</th>
<th>No. of lines resistant</th>
</tr>
</thead>
<tbody>
<tr>
<td>mN</td>
<td>H</td>
<td>8</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>17</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>asN</td>
<td>nd</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>6</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>9</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>nd</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P0N</td>
<td>nd</td>
<td>12</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

a mN and asN represent plants expressing the sense and antisense untranslatable N genes, respectively; P0N represents plants containing the promoterless N gene; 
b the level of the N gene RNA was estimated in each line by Northern blots, nd indicates that the N gene transcript was not detected; 
c 30-fold diluted leaf extracts of the N. benthamiana plants infected with TSWV-BL were applied to three leaves of each plant at the 6-7 leaf stage. Each extract was first applied to all test plants followed by control healthy plants. Data were taken daily for 45 days after inoculation and only the asymptomatic plants were rated resistant.

Unlike the controls, which developed typical systemic symptoms 7 to 9 days after inoculation, 16 out of 21 mN plants and 5 out of 8 asN plants were asymptomatic throughout their life cycles. Northern blot analysis of leaf tissues sampled before inoculation showed that all the resistant R0 lines produced low levels of the sense or antisense N gene RNA, whereas the susceptible R0 lines produced either none or high levels of the RNA species. Since this data suggested that the resistance of transgenic plants to TSWV-BL was related to their relative levels of N gene transcript, transgenic progenies from four mN
and three asN R₀ lines with either high or low N gene transcript levels were selected by germination on kanamycin-containing media. These transgenic plants were tested for resistance to the four *Tospoviruses* at the 3 to 4 leaf stage, except that some R₁ plants from two asN lines were inoculated at the 6 to 7 leaf stage. The results are summarized in the following table:
<table>
<thead>
<tr>
<th>R₀ Line</th>
<th>N gene RNA&lt;sup&gt;a&lt;/sup&gt;</th>
<th>TSWV-BL</th>
<th>TSWV-10W</th>
<th>INSV-Beg</th>
<th>TSWV-B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Promoterless N gene</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P₀N-1</td>
<td>nd</td>
<td>10/10</td>
<td>10/10</td>
<td>10/10</td>
<td>10/10</td>
</tr>
<tr>
<td>P₀N-2</td>
<td>nd</td>
<td>15/15</td>
<td>10/10</td>
<td>10/10</td>
<td>10/10</td>
</tr>
<tr>
<td>N₀-3</td>
<td>nd</td>
<td>8/8</td>
<td>6/6</td>
<td>6/6</td>
<td>6/6</td>
</tr>
<tr>
<td><strong>Untranslatable N gene</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mN-2</td>
<td>H</td>
<td>20/20</td>
<td>20/20</td>
<td>20/20</td>
<td>20/20</td>
</tr>
<tr>
<td>mN-7</td>
<td>H</td>
<td>20/20</td>
<td>20/20</td>
<td>20/20</td>
<td>20/20</td>
</tr>
<tr>
<td>mN-13</td>
<td>L</td>
<td>2/20</td>
<td>4/20</td>
<td>20/20</td>
<td>20/20</td>
</tr>
<tr>
<td>mN-18</td>
<td>L</td>
<td>4/20</td>
<td>1/20</td>
<td>20/20</td>
<td>20/20</td>
</tr>
<tr>
<td>N₀-3</td>
<td>nd</td>
<td>24/24</td>
<td>32/32</td>
<td>24/24</td>
<td>24/24</td>
</tr>
<tr>
<td><strong>Antisense N gene</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>asN-1</td>
<td>L</td>
<td>20/20&lt;sup&gt;b&lt;/sup&gt;</td>
<td>20/20</td>
<td>20/20</td>
<td>20/20</td>
</tr>
<tr>
<td>asN-4</td>
<td>H</td>
<td>20/20</td>
<td>20/20</td>
<td>20/20</td>
<td>20/20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(16/16)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>(16/16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>asN-9</td>
<td>L</td>
<td>19/20</td>
<td>20/20</td>
<td>20/20</td>
<td>20/20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3/41)</td>
<td>(5/21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N₀-3</td>
<td>nd</td>
<td>16/16</td>
<td>16/16</td>
<td>16/16</td>
<td>16/16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(32/32)</td>
<td>(20/20)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Northern analysis of R₀ lines from which the R₁ plants were derived (see preceding table);

<sup>b</sup>the underlined fractions indicate that most of susceptible R₁ plants displayed a significant delay of symptom appearance;

<sup>c</sup>the fraction in parenthesis represents the inoculation data obtained from plants inoculated at the 6-7 leaf stage; the remaining data in this table were generated from plants inoculated at the 3-4 leaf stage; inoculated plants were observed daily for 45 days after inoculation.
All $R_1$ plants from high expressor lines mN-2 and mN-7 were susceptible to infections by all Tospoviruses tested, and these plants did not show a delay of symptom appearance as compared to controls. In contrast, high proportions of the $R_1$ plants from low expressor lines mN-13 and -18 were resistant to homologous (TSWV-BL) and closely related (TSWV-10W) isolates, but not resistant to infections by distantly related Tospoviruses (INSV-Beg and TSWV-B). The resistance of asN $R_1$ plants from low expressor R0 lines was markedly influenced by the TSWV isolate used for inoculation. All but one of the small $R_1$ plants (3-4 leaf stage) from low expressor lines asN-1 and -9 became infected, although there was a delay of symptom appearance, when inoculated with the homologous TSWV-BL or closely related TSWV-10W isolates. In contrast, most of the large $R_1$ plants (6-7 leaf stage) from line asN-9 were resistant to both isolates. In comparison, control $R_1$ plants and $R_1$ plants from the high expressor line such as asN-4 displayed no resistance to either of the isolates regardless of the size of test plants. Antisense RNA-mediated protection was not effective against infection by the distantly related INSV-Beg and TSWV-B isolates.

Analyses of data presented in the above two tables suggest that sense and antisense RNA-mediated protections are observed only in low expressors of the N gene. The $R_1$ asN plants that produced high levels of the antisense N gene transcript were as susceptible as control plants. In contrast, the asN low expressors displayed a delay in symptom appearance when inoculated at the 3-4 leaf stage and showed increased levels of resistant when inoculated at the 6-7 leaf stage.

Inhibition of viral replication in tobacco protoplasts expressing the sense or antisense form of untranslatable N coding sequence was also noted. In this instance, whole virion preparations of TSWV-BL were used to transfet protoplasts isolated from transgenic lines to investigate the effect of sense or antisense N gene transcript on replication of the incoming virus. Viral replication was determined by measuring the level of the N protein of the incoming virus in transfected protoplasts, and it was found that protoplasts derived from plants (mN-7 and asN-4) that produced high levels of the respective
RNA transcripts supported the replication of the virus, whereas protoplasts from mN low expressor (mN-18) did not. Protoplasts from an asN low expressor (asN-9) supported much lower levels of viral replication.

Accordingly, in this aspect of the present invention we have shown that transgenic plants expressing sense or antisense form of untranslatable N gene coding sequence are resistant to homologous (TSWV-BL) and closely related (TSWV-10W), but not to distantly related (INSV-Beg and TSWV-B) Tospoviruses. The following table provides a comparison of resistance to Tospoviruses between transgenic tobacco expressing various forms of the TSWV-BL N gene:

<table>
<thead>
<tr>
<th>Tospovirus</th>
<th>Homology to TSWV-BL N Gene</th>
<th>Form of the Transgene</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSWV-BL</td>
<td>100%</td>
<td>N R R asN P^N</td>
</tr>
<tr>
<td>TSWV-10W</td>
<td>99%</td>
<td>R R R asN S</td>
</tr>
<tr>
<td>INSV-Beg</td>
<td>60%</td>
<td>R^c S S S</td>
</tr>
<tr>
<td>TSWV-B</td>
<td>78%</td>
<td>S S S S</td>
</tr>
</tbody>
</table>

^aReactions of transgenic tobacco and N. benthamiana plants expressing the intact N gene (N) of TSWV-BL to inoculation with the four Tospoviruses are included for comparisons with inoculation results of transgenic plants containing untranslatable (mN), antisense (asN), and promoterless (P^N) N coding sequences, R = resistant, S = susceptible;

^bThe nucleotide sequences are as reported in Phytopathology 82:1223 (1992) and Phytopathology 83:728 (1993)

^cLevel of resistance may depend upon the concentration of inoculum.

These results confirm and extend the earlier aspects of the present invention for RNA-mediated protection with TSWV. Furthermore, the protection is observed in plants producing low rather than high levels of the N gene transcript, and although earlier studies reported herein indicate that tobacco plants which produced high levels of the TSWV-BL N protein displayed resistance to INSV-Beg, this additional data indicates that since resistance to INSV-Beg was not observed in transgenic plants expressing the sense or antisense form of the untranslatable of the N gene thus clearly indicating that protection against INSV-Beg is due to the presence of the N protein and not the N gene transcript. Thus, it appears that two different mechanisms are
involved in protection transgenic plants against TSWV and INSV Tospoviruses according to the present invention. One mechanism involves the N gene transcript (RNA-mediated), and another involves the N protein (protein-mediated). In addition, the results of the protoplast experiments indicate that N gene RNA-mediated protection is achieved through a process that inhibits viral replication, and the data contained in the above tables suggest that protection against the distantly related INSV-Beg isolate is conferred by the N protein of TSWV-BI, and not by the gene transcript.

Finally, further studies were conducted to provide still another aspect of the present invention - that a portion of the Tospovirus nucleoprotein gene provide protection of transgenic plants against infection by the Tospovirus. It has already been demonstrated above that the N gene RNA protects against homologous and closely related TSWV isolates while the N protein protects against the homologous isolate and distantly related INSV isolates; that N gene RNE-mediated protection is effective in plants expressing low levels of the N gene whereas N protein-mediated protection requires high levels of N protein accumulation; and that the N gene RNA-mediated protection is achieved through inhibition of viral replication. Based upon this prior data, we next set out to determine whether a portion of the N gene might work against infection by the virus. We found, as discussed below, that transgenic plants expressing about one-half of the N gene sequence is resistant to the virus.

The following describes the cloning of one-half N gene fragments of TSWV-BI in order to demonstrate this final aspect of the present invention. The first and second halves of both the translatable and untranslatable N gene were generated by reverse transcription and then PCR as described above. As depicted in figure 8, the nucleotide deletion or insertions at the 5'-terminals of the untranslatable half N gene fragments are indicated by the dash symbol; ATG codons are underlined and all possible termination codons immediately after the initiation codon of the untranslatable half N gene fragments are shown in bold.
The first half of the N gene was produced by RT-PCR using oligoprimers i (5'-TACAGTGATCCATGGTTAAGGTAATCCATAGGCTTGAC), which is complementary to the central region of the TSWV-BL N gene, and ii (5'-AGCTAACCATGGTTAAGGCTAAGGAAAGGCATTGTTCG) for the translatable or iii (5'-AGCTAATCTAGAACCATGGGACTCACTAAGGAAAGGCATTGTTCG) for the untranslatable first half N gene fragment, the latter two oligomer primers are identical to the 5' terminus of the N gene. Similarly, the second half of the N gene was produced by RT-PCR using oligomer primers iv (5'-AGCATTTGGATCCATGGTTAACACTAAACAGCAAC) which is complementary to the 3' noncoding region of the TSWV-BL N gene, and v (5'-TACAGTTCTAGAACCATGGGATGATGCAAATCTGTGAGGC) for the translatable or vi (5'-AGATTCTCTAGAACCATGGTGACTGATGCAAATCTGTGAGGC) for the untranslatable second half N gene fragment, the latter two oligomer primers are identical to the central region of the N gene. The oligomer primer iii contains a frameshift mutation immediately after the translation codon and several termination codons to block possible translation readthroughs while the oligomer primer vi contains several inframe termination codons immediately after the translation initiation codon.

The half gene fragments were purified on a 1.2% agarose gel as described above, and the gel-isolated gene fragments were digested with the restriction enzyme NcoI and directly cloned into NcoI-digested plant expression vector pBI525. The resulting plasmids were identified and designated as (1) pBI525-1N containing the first half translatable N gene, (2) pBI525-1N' containing the first half untranslatable N gene, (3) pBI525-1N containing the first half translatable N gene in the antisense orientation, (4) pBI525-2N containing the second half translatable N gene, (5) pBI525-2N' containing the second half untranslatable N gene, and (6) pBI525-2N containing the second half translatable N gene in the antisense orientation. The expression cassettes were then excised from plasmid pBI525 by digestion with HindIII/EcoRI and ligated as described above into the plant transformation vector pBIN19 that had been cut with the
same enzymes. The resulting vectors as well as plasmid pBIN19 were transferred to *A. tumefaciens* strain LBA4404, using the procedure described by Holsters supra. Leaf discs of *N. benthamiana* were inoculated with *A. tumefaciens* strain LBA4404 containing the various constructs. Transgenic plants were self-pollinated and seeds were selectively germinated on kanamycin as described above.

Analysis of transgenic plants by PCR and Northern hybridization PCR was performed on each R0 transgenic line as described previously. The oligomer primers i to vi were used to determine the presence of the N coding sequence of TSWV-BL. The oligomer primer 35S-Promoter (see Example VIII) was combined with one of the above oligomer primers to confirm the orientation (relative to the CaMV 35S promoter) of the half gene sequences inserted into the plant genome. Northern analysis was conducted as described in Example VIII.

Lettuce isolate of TSWV (TSWV-BL) was used to challenge transgenic plants. Inoculation was done using test plants at the 3-4 leaf stage as described above. To avoid the possibility of escapes, control plants were used in each experiment and each inoculum extract was used to first inoculate the transgenic plants followed by control plants.

The various constructs used in this aspect of the present invention are illustrated in figure 8. Translatable and untranslatable half N gene fragments were synthesized by RT-PCR and then cloned directly into the plant expression vector pBI525. The oligomer primers iii and vi, used for generation of untranslatable half N gene fragments by RT-PCR, contains a mutation immediately after the translation initiation codon and the resulting reading frame contains several termination codons to block possible translation readthroughs. Thus, both first and second half untranslatable N gene fragments should be incapable of producing the truncated N protein fragments when introduced into plants. Both translatable and untranslatable half N gene fragments were then placed downstream of the CaMV 35S promoter of the vector pBI525 in the sense orientation or in the antisense orientation. The expression of the half N coding sequences of TSWV-BL was thus controlled by a double CaMV 35S promoter fused to the 5'
untranslated leader sequence of alfalfa mosaic virus (ALMV) of the
expression vector pBI525. Expression vectors that utilize the stacked
double CaMV 35S promoter elements are known to yield higher levels of
mRNA transcription than similar vectors with a single 35S promoter
element. Expression cassettes were transferred from the vector
pBI525 to the pant transformation vector pBIN19. The resluting
plasmids as well as the control plasmid pBIN19 were then transferred
into A. tumefaciens strain LBA4404. Transgenic plants were obtained
with nomenclature of the transgenic lines shown in figure 8.

All the kanamycin-resistant transgenic lines were confirmed by
PCR to contain the proper N coding sequences in the expected
orientations. Each transgeinc R0 line which was grown for seeds was
then assayed using Northern blot. Six out of six 1N, four out of six 1N',
six out of six 1N'', six out of six 2N, seven out of eight 2N', and six out
of seven 2N'' transgenic R0 lines were found to produce half N gene
RNAs.

A set of transgenic R0 plants was challenged with the
homologous isolate TSWV-BL. Only asymptomatic plants were rated
resistant while the plants showing any symptom (local lesions or
systemic infections) were rated susceptible. All the inoculated R0
control plants were susceptible to the virus; in contrast, two out of
nine 1N', two out of six 1N'', four out of ten 2N', and one out of eight 2N''
R0 lines were found to be completely resistant to the virus infection.
Although none of the 1N and 2N R0 lines showed high levels of
resistance, some of those plants displayed significant delays of
symptom appearence.

Another set of transgenic R0 lines was brought to maturity for
seed production. Seedlings were germinated on kanamycin-containing
medium and inoculated with TSWV-BL. As shown in the following table,
control seedlings and seedlings from some of the transgenic lines were
susceptible to the isolate whereas seedlings from lines 1N-151, 1N'-
123, and 2N'-134 showed variojs levels of protection , ranging from
delays in symptom expression to compete resistance.
<table>
<thead>
<tr>
<th>R₀ line</th>
<th>No. plants infected/No. plants inoculated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>50/50</td>
</tr>
<tr>
<td>1N-149</td>
<td>17/17</td>
</tr>
<tr>
<td>1N-151</td>
<td>2/20</td>
</tr>
<tr>
<td>1N'-123</td>
<td>16/20</td>
</tr>
<tr>
<td>1N'-124</td>
<td>20/20</td>
</tr>
<tr>
<td>1N'-126</td>
<td>19/19</td>
</tr>
<tr>
<td>1N'-130</td>
<td>12/15</td>
</tr>
<tr>
<td>1N'-132</td>
<td>18/19</td>
</tr>
<tr>
<td>2N-155</td>
<td>20/20</td>
</tr>
<tr>
<td>2N'-134</td>
<td>0/20</td>
</tr>
<tr>
<td>2N'-135</td>
<td>19/19</td>
</tr>
<tr>
<td>2N'-142</td>
<td>20/20</td>
</tr>
<tr>
<td>2N'-143</td>
<td>20/20</td>
</tr>
</tbody>
</table>

In the above table, 30-fold diluted extracts of infected *N. benthamiana* were used to inoculate transgenic plants at the 3-4 leaf stage followed by control transgenic plants. DPI = days post inoculation.

In summary, this aspect of the present invention shows that transgenic plants expressing the first or the second half of either translatable or untranslatable N gene fragment are highly resistant to the homologous TSWV-BL isolate. This result demonstrates that a portion of the N gene is sufficient for resistance to the virus.

A listing of all nucleotide and amino acid sequences described in the foregoing description of the present invention is as follows:

SEQUENCE LISTING

(1) GENERAL INFORMATION:
   (i) APPLICANT: Dennis Gonsalves and Sheng-Zhi Pang
   (ii) TITLE OF INVENTION: Tomato Spotted Wilt Virus
   (iii) NUMBER OF SEQUENCES: 30

(2) INFORMATION FOR SEQ ID NO:1:
   (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 25 base pairs
       (B) TYPE: nucleic acid
       (C) STRANDEDNESS: single
       (D) TOPOLOGY: linear
   (ii) MOLECULE TYPE: DNA
   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
AGCAGCAGCA AGTGCCAGA CTTGC 25

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GCAAGTCTC GCAGTTTGC CTGCT 25

AGCTTAACCAT GTTACGCT ACTAGCAGAA GC 32

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ACCTTAACCAT GTTACGCT ACTAGCAGAA GC 32

AGCATTCCAT GTTACACCA CTAGACAGAC AC 32

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AGCATTCCAT GTTACACCA CTAGACAGAC AC 32

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2265 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CAAGTIGAAA GCAGAACACG AGCTGAAAT AAGIGIGGAG TGGAAATCTCT 50
GCTCATIGCA GCAGAAAACG AGCTCATCTCC TAAACTTCGA GCTTCCAGCT 100
ATTCTCATTT CAAGCTGAGC CTCTGCTTA GGGTCCCAA GGTTTIGAAG 150
TCAGACATCA AATAAGCGTC ATCTCTTCTGA TCGTGGTGCAT GTTTTTCAGA 2000
CAAAAAGTCT TGGAGTTGAA TCGTACCAAG TCTGTACCTT CCTCAAACTC 2050
AAGGCTTCTG CCTCTCTGCA ACAAGAGAAC AAGGTTTCCCT TTGACTGACTT 2100
TAACTTAGA CATGATTGCT GTAAAGTTTG TTAATGCCTT CAGGGAATGT 2150
AACCTAAGGT GTGAAAGTGG CACCTGTCGT CCGGCGACGT TTTTCTTGGT 2200
TCTGAAATG TGATGATTGA AGACTGAGTG TTAAGGTAAG AACACAAAAT 2250
TGACGGATT GTCTT 2265

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1709 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

15 (x) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AAATTCTCTT GCAGTAGGAA CTCTGTCAT GTTAGGAGAA AGAAAGTAACA 50
TGCCTPACTC TCAAGGTTTT GTGAAAGTTT CATACTGATT TAATTTCAAG 100
CTAGACCCTG GCCTAGGGGT TCGAAGGTT TTAGGCAAAC TTTTCATTCA 150
GAAATGTTCC AGGTTGTCAG GAGATGAAAC AAATAAAACA TTTTTTTAT 200

20 CTTAGGCGTG CATCCCAAACT CATACAGTGT TTAGACAGCC TTTAACCATT 250
ACTGTTATTT GCAAGCACTA GCTCGGAATT CGTAAAAGTA AAACCTCTTT 300
TGATTTACGC ATGCATTGTT CTGATTAAAG GAGCCGTAC AACATTATTC 350
ATGATCCCC ATATGCACAA AGGATTGTTT CGGCTCTGCT TTGAAATCAC 400
ACATCTTTTG CACGACTCTT TTCAACAACAG TCGAACAGAG ATGGGATCAT 450

25 CTACACCTTG AACAAAGCAG AGCTTACTCC TGGAAAGTTA GATTAAAAATG 500
AAATACACGT GATATACAGT GAAGAGCCTA ACACAGGCAA ATATTTCTTT 550
TCAAAAAAGC TTAGAATGCT TCAATCTAAC ATACAAACTA TGTCTTATT 600
AGACAGCAATCC AATAAGCTTT CATTGAAGAT AGACTTGGCC AGGGGAGAA 650
TTAAAAATTC TTCAATACCT ATTTGAGATT CAAAATCTTT TTAAATCTCT 700

30 GATTATAGCG GGAATAAAA AAGAAAGATC AAGATTAGG AAAGAATAAGC 750
TTGAGATACA AAATGCGGCT GCGTGCTCAA GCTTTCTCTA ATTAAGTTAT 800
GTTAATTTCG TCTCTTTACT TAAATTTAAT TTCTGTTTGT TCAATTTCTT 850
CATAAACCTC TCTCTTAGAG AAGACATAAA AAGAAATAAT AAAATTAAAA 900
TAAATCAAA ATAAACTAAA AATCCAAAAA TGAATAAAAAG CCAACAAAAA 950

35 AATTTAAAAA CAAAAGACA AAAAGATCC CGAAGGAGCA ATTTGCGCA 1000
AATTGCGGCT TTTTTTTGT TTTTTTGT TTTTTTTT TTTTTTTTT 1050
TTATTTTTAT TTATTTTTAT ATTATTTTTT TTTTGTGTGT 1100
TTGTTTTTAT TTATTTTTAT AAGGACAACA CACAGAGAGCA AACTTTAAT 1150
TAAGAAAACT TATTTAAAAT TTTGACACT AAGGACAACA AACAATAA 1200
GATAAAGAAA GTTTTATAATA TTTATAGGCT TTTTATAAT TTTACTTACA 1250
5
GCTGGTTTTA AGCAAGCTCT GTGAGTTTGT CTTGTGTITT AACCCCAAC 1300
ATTATAGAAG ATTTGTTAGG GTTTCACAG TAATGTCTCA TACCTACTC 1350
TCTTTAGGAT TGAGATTGCG TGGAGCTAGG TATAGGAGGA TACTCTTTCC 1400
CTCTCCTGAC CTGATCTTCA TTCAATTCAA ATCTTTTCT TTTGACACA 1450
GTGCAAACTTT TTOCTAAAGGC TTOCGCTGIGG TCAATCTTCTT TGGGTTGAT 1500
10
COGAGATTCC TTGATTTTGT CATCCGATA TATAGCAAG GACACTGTCA 1550
TCATCTCAAA GCCTATCAACT GAGCAATGA GAGGTAAGCT ACCTCCAGC 1600
AAATTGCAAA GCCTCACAGA CTTGCACAGA TCAAGAGCTT ATCCATAGGC 1650
TTAATTTAA GGGTGGGAAA CATCTTATAG TTGATTATAT TTGAGATTCT 1700
CAGATTTTCC 1709
15
(2) INFORMATION FOR SEQ ID NO:7:
(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 260 amino acids
   (B) TYPE: amino acid
   (C) STRANDEDNESS: single
20
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xii) SEQUENCE DESCRIPTION: SEQ ID NO:7:

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

(2) INFORMATION FOR SEQ ID NO:8:

20 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 858 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TTAAGCATC AAGCAAGCAG AAACAAATATT ATAAAAAGAA GCTTTATATA 50
TTATAGGTCT TTTTATATAT TTAAGCCACGG CGTGGTTTTA AGGAAACTCT 100
GGAGTTTTCG GAAACTATTG AGAGCTTCCC AGCTCCGAGC ACATTGTAAG 150
GGTATTCGGA CACATTTTCT TTTAGGACCA GTGCAACCTT TCTCTTACGC 200
TGGCTGCTGT TTTAGCAAG CAGATCTGCA ACCGATCCTG GCCCTATGCT 250
GGTCGAGTGA TTTGAGTGGG CAGACCTGTA TTTTCTACAG TTTTTTTTCA 300
GAAGCAATCA GAGGTTGAGC CCCTGAGGAG AGATTGACGC GCCCTAGACA 350
CTTGAGATCA TCAAGACGA ATCCATAGCA TTGGACTGGG GGGAAGGGAG 400
CAATCTTAGA TTTGATGAGT TGAGATTCG CAGATTTGGA ACATTTGCTG 450
AAGGCCCAG CCGCATCAG CCACTGAGGA AGTCCACGAT TCGGATAGAT 500
GGCTCAATCT CTTGCTTGGA TTTTCTATAT GCTTTTTTTA CCAAAGTTAA 550
AAATGCTTTTG CTTAATACC TTGATTGCT TCTAGAGATT CTGGAGATAT 600

62
GTCAGACATG AAATAATGCT CATCTTTTTG ATCTGCTCAA GTTTTCGAG 750
ACAAGAGTIC TTGAGGTGA ATGCTAACCAG ATTCGATCT TCCTCAAACT 800
CAAGGTCTTCT GCCTGGTTGTC AAQAAAGCAA QAATGCTTTTC CTTTAGGAGC 850
TTAACCAT 858
5  (2) INFORMATION FOR SEQ ID NO:9:
   (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH:  2028 base pairs
      (B) TYPE:   nucleic acid
      (C) STRANDEDNESS:  single
10  (D) TOPOLOGY:  linear
   (ii) MOLECULE TYPE: DNA
   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
    AAAATTCTTT CCGAGTGAAAT CTCGGCTCAT GTGAGCAGAA AACAACATA 50
    TGCCCTAACC TCAGCTTTTT GTGAAAGGTT CTACATACCGA TGGGTTTTTT 100
    15 CTGAGCCCTCT GGCTAAGGTT TCAGAAGCGA TTGGAACAGA TTTCATATCA 150
    GAAATTTGCTG AGGTGGCGAGA AGATGGAAC AAATAAAAAA TTTTTTTTTT 200
    CTTTGGGCTG CATGGCAAAAC CAAAGACCTG GTTGAACAGC TTTAAACATT 250
    ACIGATTGTTT GCAAGATGCA GCTCCCAATT CTTAAATGTA AAAAAGCTTT 300
    TGAATTATGCA ATGAATGTGTT CTTAATTTAA GAGGCTTCTAC AACAATTATTC 350
    20 ATGATCTCTCC ATATCCCAGG AGGATTGTTG ATGCTTCTGT GAAACTGAC 400
    ACACTCTTTG CACCAAGTTC TCGAACAAC TTGCAAGTGA ATGIGATCAT 450
    CTACACGCTG ACAAAACATTG AGCTGACCTC TGGAAAGTTA GAAAGTGCTG 500
    AAAAATCATT GAATATCGAT GAGAGCGCCT ACAAAAAGAA ATATTTTCTTT 550
    TCAAAAAACAT TGGAATTGCT TCAATCTAAT ATAAAAACTA TGGTTTATTT 600
    25 AGACAGCAGC CAATCCCCGT CTGGAGAGAT AGACTTTCCG AGGGAGAAA 650
    TTAATATTTTC TCCAACAATCT ATTTGAGTGA CAAAATCTTT TGAATATCTT 700
    GNTTAAACCG GATTAAAAA GAAGAAACCT AAGATTTAGG AAGCATAATGC 750
    TTTGAGGATC AAAATTGACT GCTGIGGTGA CTTTTTCTCTA ATATTGTTAT 800
    GTTTATTCTT TTTCTTCTCT GAATTTTATT TTTCTTCTTTG TCAATTTCTC 850
    30 CAATTTTCTTC CTCTGCTAGA GAAAACATAA AAACAAAAAA AAAAAATTTAA 900
    TAAACATATA ATTTAAAAAA ATCTAAAAAA TGAAATAAGGA CCAAAAAA 950
    AATTTAAAAA CAAAACAACA AAAAAAGACG CGAAGAGACA ATTTTGGCCA 1000
    AATTGIGGGT TTTTITTTTT TTTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1050
    TTATTTTTTT TTATTTTTTT AATTATGCTT TTTTATGTTT TTTGTTGTTT 1100
    35 TTTATTTTTTTT TTTTATTTTT AAGCAGACAA CACAGAAAAGC AAACCTTAT 1150
    TAAACACACT TATTTAAAAT TTAACACACT AAGCAAGCAGA AACAATTTAA 1200
63
GATAAAGAAA GCTTTATATA TTTATAGGCT TTTTTATATT TTAACCTACA 1250
GCTGCTTTTA AGCAGATTCT GGTAGTTTGT CCITTTTTTT TACCCCAARAC 1300
ATTCCATAGA ACITGTAAGA GCTTTCTATT TAATGGTTCGA TGCCAAACT 1350
TCCCTTAGCA TTTGAGATTTG TGGAGCTTAA TTAAGCAGCA TACTCTTTC 1400
5
CTTTCGCTCA TGATACCTCA TTATTTTCAA ATGCCTTTTCT TTTCAGCACA 1450
GTGCAAGCTT TTTCCTGAGGC TTCCCTGGTG TCGAATCTCT TTGGGTGAT 1500
CCCGAGATCC TTGTAATTGT GATCCGATA TATAGCCAAG ACACACTGA 1550
TCACTCTCAA GCTATCAAAT GAAGGCTAA AAGGTAAGCT ACCTCCAGC 1600
ATTATGGCAA GCTTACAGA CTGGTGAAC TCAAGAGGTA ATCCATAGCC 1650
10
TTGACCTCAA GGGTGCGAGA CAACTCTTGA TTGATAGTA TTGGAGTTCT 1700
CAGAAACCTT AGTTTCTCTA ACAAGCCCTG CAATCTCAAGC 1750
CTTTGCAGG GCTAGCAGCT GCTGCAATCT CGTCTGAAGG TTGTCTTTAT 1800
GGTAATTTTA CCAGTAATTA AATCGCTTTG CTAAATACCC TTCAATATGG 1850
TCGAGCAATT CGCTGAAATG GTAGACAGTG AATAATGCT CATCCTTTTG 1900
15
ATCTTGCTCA GGGTTTCTCG ACRAAGCTTC TGGAGTTCA ATGCATCGAG 1950
ATTCTGCTCT TGTCCTCAATGA CAGGCTCITC GCCTTTGCTG AACAAGGCAA 2000
CAATGCTTTC CGTAGTGAGC TTAAACCT 2028

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

25
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TTTCTGCCTT CTTCACACT CA 22

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CGTACGCAAT GACGAAAG 18

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 467 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: peptide

(x) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Ser Ser Gly Val Tyr Glu Ser Ile Ile Gln Thr Lys Ala Ser
5 10 15
Val Trp Gly Ser Thr Ala Ser Gly Lys Ser Ile Val Asp Ser Tyr
10 20 25 30
Trp Ile Tyr Glu Phe Pro Thr Gly Ser Pro Leu Val Gln Thr Gln
35 40 45
Leu Tyr Ser Asp Ser Arg Ser Lys Ser Ser Phe Gly Tyr Thr Ser
50 55 60
Lys Ile Gly Asp Ile Pro Ala Val Glu Glu Glu Ile Leu Ser Gln
15 65 70 75
Asn Val His Ile Pro Val Phe Asp Asp Ile Asp Phe Ser Ile Asn
80 85 90
Ile Asn Asp Ser Phe Leu Ala Ile Ser Val Cys Ser Asn Thr Val
15 95 100 105
Asn Thr Asn Gly Val Lys His Gln Gly His Leu Lys Val Leu Ser
110 115 120
Leu Ala Gln Leu His Pro Phe Glu Pro Val Met Ser Arg Ser Glu
125 130 135
Ile Ala Ser Arg Phe Arg Leu Glu Gln Glu Asp Ile Ile Pro Asp
140 145 150
Asp Lys Tyr Ile Ser Ala Ala Asn Lys Gly Ser Leu Ser Cys Val
155 160 165
Lys Glu His Thr Tyr Lys Val Glu Met Ser His Asn Gln Ala Leu
30 170 175 180
Gly Lys Val Asn Val Leu Ser Pro Asn Arg Asn Val His Glu Trp
185 190 195
Leu Tyr Ser Phe Ile Pro Asn Glu Asn Glu Ile Glu Ser Asn Asn
200 205 210
Arg Thr Val Asn Ser Leu Ala Val Lys Ser Leu Leu Met Ala Thr
35 215 220 225
Glu Asn Asn Ile Met Pro Asn Ser Gln Ala Phe Val Lys Ala Ser
230 235 240
Thr Asp Ser His Phe Lys Leu Ser Leu Glu Leu Arg Ile Pro Lys
40 245 250 255
Val Leu Lys Glu Ile Ala Ile Gln Lys Leu Phe Lys Phe Ala Gly
260 265 270
Asp Glu Thr Gly Lys Ser Phe Tyr Leu Ser Ile Ala Cys Ile Pro
275 280 285
Asn His Asn Ser Val Glu Thr Ala Leu Asn Val Thr Val Ile Cys
290  295  300
Arg His Gln Leu Pro Ile Pro Lys Ser Lys Ala Pro Phe Glu Leu
305  310  315
5 Ser Met Ile Phe Ser Asp Leu Lys Glu Pro Tyr Asn Thr Val His
320  325  330
Asp Pro Ser Tyr Pro Gln Arg Ile Val His Ala Leu Leu Glu Thr
335  340  345
His Thr Ser Phe Ala Gln Val Leu Cys Asn Lys Leu Gln Glu Asp
350  355  360
10 Val Ile Ile Tyr Thr Ile Asn Ser Pro Glu Leu Thr Pro Ala Lys
365  370  375
Leu Asp Leu Gly Glu Arg Thr Leu Asn Tyr Ser Glu Asp Ala Ser
380  385  390
15 Lys Lys Lys Tyr Phe Leu Ser Lys Thr Leu Glu Cys Leu Pro Val
395  400  405
Asn Val Gln Thr Met Ser Tyr Leu Asp Ser Ile Gln Ile Pro Ser
410  415  420
Trp Lys Ile Asp Phe Ala Arg Gly Glu Ile Arg Ile Ser Pro Gln
425  430  435
20 Ser Thr Pro Ile Ala Arg Ser Leu Leu Lys Leu Asp Leu Ser Lys
440  445  450
Ile Lys Glu Lys Lys Ser Leu Thr Trp Glu Thr Ser Ser Tyr Asp
455  460  465
25 Leu Glu;

(2) INFORMATION FOR SEQ ID NO:13:
(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 258 amino acids
   (B) TYPE: amino acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear
(ii) MOLECULE TYPE: peptide
(x) SEQUENCE DESCRIPTION: SEQ ID NO:13:
Met Ser Lys Val Lys Leu Thr Lys Glu Asn Ile Val Ser Leu Leu
5   10  15
Thr Gln Ser Ala Asp Val Glu Phe Glu Glu Asp Gln Asn Gln Val
20  25  30
35 Ala Phe Asn Phe Lys Thr Phe Cys Gln Glu Asn Leu Asp Leu Ile
440  445  450
40 Lys Lys Met Ser Ile Thr Ser Cys Leu Thr Phe Leu Lys Asn Arg
50  55  60
Gln Gly Ile Met Lys Val Val Asn Gln Ser Asp Phe Thr Phe Gly
65  70  75

66
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lys Val Thr Ile Lys Lys Asn Ser Glu Arg Val Gly Ala Lys Asp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Met Thr Phe Arg Leu Asp Ser Met Ile Arg Val Lys Leu Ile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu Glu Thr Ala Asn Asn Glu Asn Leu Ala Ile Ile Lys Ala Lys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>115</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile Ala Ser His Pro Leu Val Gln Ala Tyr Gly Leu Pro Leu Ala</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>130</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp Ala Lys Ser Val Arg Leu Ala Ile Met Leu Gly Gly Ser Ile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>140</td>
<td>145</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro Leu Ile Ala Ser Val Asp Ser Phe Glu Met Ile Ser Val Val</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>155</td>
<td>160</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu Ala Ile Tyr Gln Asp Ala Lys Tyr Lys Glu Leu Gly Ile Glu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>170</td>
<td>175</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 Pro Thr Lys Tyr Asn Thr Lys Glu Ala Leu Gly Lys Val Cys Thr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>185</td>
<td>190</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val Leu Lys Ser Lys Gly Phe Thr Met Asp Asp Ala Gln Ile Asn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>205</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys Gly Lys Glu Tyr Ala Iys Ile Leu Ser Ser Cys Asn Pro Asn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>215</td>
<td>220</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Lys Glu Ser Ile Ala Met Asp Tyr Tyr Ser Asp Asn Leu Asp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>230</td>
<td>235</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys Phe Tyr Glu Met Phe Gly Val Lys Lys Glu Ala Lys Ile Ala</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>245</td>
<td>250</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 Gly Val Ala</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3049 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGAGCAATTG GGTCAATTTT TATTTCTAAAT CGAACCCTCA CTAGGAAATTC 50
TCAGACTGT AATAAGCACA AGAGCACAAG AGCCAAATG TCATCAGGTT 100
TTTAAAGAAC GATCACTCAG ACAAGGCCTT CAGTTTGCCG ATCGAGACGC 150
TCGCTCAGGT CCATGIGGGA TTTTACTTGG ATTTATGAGT TTCCAGCTTG 200
TTTCCACTGT GITCAACAATC AGTGGTACTC TCACTCGGG AGCAAAATTA 250
GCTTGGCTA CACTCACAAT ATTGGIGGTA TTCGCTGCTG AGAGGAGAAA 300
ATTTTATCTC AGAGACACA TATCCAGAG TTTGAGAATA TTGATTGACAG 350
CATCAATAATC AATGATTCTT TCTTGSCAAC ATCTGTTTTG TOCAACACAG 400
TTAACACCAA TGAGTGGAAG CACAGGCCTC ATCTTAAAGT TCTTCTCCTT
450
GOCGAATGGC ATCCCTTGGGA ACCCTGATGAG ACCAGGTGAC AGATTGCTAG
500
CAGATCTCGG CTCCAAAGAG AAGATATAAT TCCTGATGAC AATAATATAT
550
CTGCTGTCTAA CGAGGGAGCT CTCTCCTGTG TCAAAAGAAA TACTTACAAA
600
5
GTCGAATTGA GCCACCATCA GCCTTGGAGC AAAGTGAAGT TCTTTTCCCT
650
TAACGAAGAT GTTCCACGAT TGCTGATAG TTTCAAAACCA AAITTTCAAC
700
AGATCGAAAG TAAATACAGA AGCTTAAATT CTCTTGCGAT CAAACTCTTG
750
CTCAGTGCCTA CAGAAACAA CATTATGCTT AACTCTCAAG CTITTTGTTAA
800
AGCTTCTACT GTTCTCAGTT TTAGATGAG CCTTTCGCTG AGAATTCAAA
850
10
AAAGTATGGAA GCCATATGCC ATACAGAAGC TCTTCAAGGT TCGGAGAAC
900
GAAACCGTTA AAAGTTCTTA TTTGCTTAAT GCATGCAATCC CAAATCACA
950
CAGCTGTGGA AGCATCTTAA ATGTGACGT TATATGTGAA CATCAAGTC
1000
CAATCCTTAA GTCGAAAGCT CCTTTGGAGT TATCAAGAT TTTCTTCGAT
1050
CIGAAAGGCT CTTACAACAC TGGCTGATGC CTTTCAATCG CTCAGAGGT
1100
15
TTGCTCATGCT TTTCTTGGAA CTCAACATTC CTTTGCACAA GTTCTCTGCA
1150
ACAAGCTGCA AGAGATGTTG ATCATATATA CTTAAACAG CCGTCAACTA
1200
ACCCAGCTA AGCTGGATCT AGTGGAAAGA ACCTTGAACCT ACGTGAGA
1250
TGCTTGGAAG AAGAGTTAAT TTCTTTGCAA AACACTGCAAG TGGCAGCGG
1300
TAATAGGGCA GACTATGCTT TATITGGATA GCATCAGAT TCTTCTATGG
1350
20
AAAGTACAGCT TTGCGAGAGG AGAGATCAGA ATCTCCTCTC AATCTACTCC
1400
TTTGAAGAA CTAGGCTGCA GCAGATTTT CAAGGAGGAA AAGAGGAAAG
1450
AGTCTCTGCAC TTGGAACACA TCAGCTATAG ATCTAGAATA AAAGTGGGCT
1500
ATACATACCT AGTATAGTTG TGTCACATCG CCTAAGCTTT ATGGTGTGTTA
1550
TTCTTTTAA AATCTAAAGTA AGTGGATTAC AAGTAGCTTA GTATGCCTAA
1600
25
GCATTATAC AAAAAATACA AAAAAATACA AAAAAATACA AAAAAATACA
1650
AAAACCGAAA AGAGTCCCCA AGGAGCGAT GGGGTGAGTT TACTCTGTTT
1700
TACGGATTTAG TGAATGCTTT TGGTTGAGGC AAAATAGCAT GGAACAGCTC
1750
AAATACCTGA ATTTAAGTAC CTAAAAGAGG TTTCGAAGGA CAGCTTAGGC
1800
AAAAATGCTTT TTTGTTTTTTT TTTGTTTTTT TTTGTTTTTT
1850
30
TTTATTTTTA GTTTTTTTTT TGGTTTTTT TATTTTATTT TTTTTTTTTT
1900
TTTATTTTTT TTTTTTTTAT ATATATCAAA CCAAATCCAC ACAAATATTT
1950
TPAATTACCAG GTTCTATTGT GCTTAAACAC AGCTATCTATG AATTATACAG
2000
ACCTTAACAG CGTTGTTTGA CTGTAAACACA CTTAAGCTAG TTTAACAACA
2050
CTTAGATTTA TGCATCCCTT AATTAACACA CTTTTAATTT ATGCTACCTC
2100
35
GAATCCGCTT TAAAGAGCTT TTTAATGCAAC ACCACCAATC TGGGCTCCTT
2150
68
TCCTACTCC AAACATTCCA TAGAATTTGT CAAGATTATC ACCTGTAATAG 2200
TCCTATAGCA TGTTCCCTTT AGCATGGGGA TTGCAAGAAC TTAGTATCTT 2250
GGCCATATTCT TTTCTTGTGT TTATCTGAC ATCAATGATT GAAATCTCCT 2300
TGCTTTTAAG CACTTGCCAA ACCCTCCCAA GACCTTCCCT AGGCTGTGAC 2350
TTTAGTCTGGA CAATGCTCAA CTTGCTTGGAT GTGATATGCG 2400
AAGAAGACAA CGTACGATCT CGAAGCTGCA AAGAAGACAA AGTGAAGGGA 2450
TACTACCTCC AAGCATTATA GCAAGCTGCA CAGATTTGCG ATCTGGCAGA 2500
GGGACGCGGT AAGCTTGGAC CAAAGGGGCG GACCGATTT TGCTTTTGAT 2550
AAATAGGAGA TTCTCATTGT TTGCGATGTC TTCTATGAGC TCGACTCTTA 2600
TCATGCTATC AAGCTTCCGAA AGGATGATAT CCTTAGTCCC AACTCCTTTCA 2650
GAATTTTCT TTATGTCGAC CTATTGAAA GAAAATTAC CCAGTTTGGAC 2700
AACTTTTATA ATGGCTTGGGC GATTCCTGA AAAGATCAGA CATGAAGTGA 2750
TACTCTATTT CTATTGACGG CTCAGAAATT CTCGACGAGA AGTCCTTAAAG 2800
TGGAATGCGA CTTGGTTCGT GTCTCTTTCA AACTGAAACAT CTGGAGATGG 2850
AGTGAARRGA AGACATATGT TTTCTTTTGT GACCTTGACCC TTAGACATTG 2900
TGCCAGITTTA GACCTGAGGC TTTCCTGAGA GATAAGATTC AAGGTTGAAA 2950
ACGCGCAABG TGGAGCCGGC GTCTGTTACT TATCCGTGTA ATGTTGATGT 3000
TTGTAATTTGT GAGTAAATTG TTTTGGAAA AAATGGAGA GAACTGCTCT 3049

(2) INFORMATION FOR SEQ ID NO:15:

20
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 778 base pairs
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25
(ii) MOLECULE TYPE: DNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO:15:
ATG CAA CAC CAG CAA TCT TGG CCT CTT TCT TAA CTC CAA 39
ACA TTT CAT AGA ATT TGT CAA GAT TAT CAC TGT AAT AGT 78
CCA TAG CAA TGC TTT CCT TAG CAT TGG GAT TGC AAG AAC 117
30
TAA GTA TCT TGG CAT ATT CTT TCC CTT TGT TTA TCT GTG 156
CAT CAT CCA TTG TAA AIC CTT TGC TTT TAA GCA CTG TGC 195
AAA CCT CCA CCA GAG CTT CCT TAG TGT TGT ACT TAG TGG 234
35
GTT CAA TCC CTA ACT CCT TGT ACT TTG CAT CCT GAT ATA 273
TGG CAA GAA CAA CAC TGA TCA TCT CCA AGC TGT CAA CAG 312
GTC TCA CAG ATT TTG CAT CTC CCA GAG CCA GCC CTT AAG 351
(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 18 base pairs
   (B) TYPE: Nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GTTCTGAGAT TTGCTAGT

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 18 base pairs
   (B) TYPE: Nucleic acid
   (C) STRANDEDNESS: single

(ii) MOLECULE TYPE: DNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TTATACTC TTCACGAGA

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 1401 base pairs
   (B) TYPE: Nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO:18:

ATG TCA TCA GGT GTT TAT GAA TCG ATC AIT CAG ACA AAG
GCT TCA GIT TGG GGA TCG ACA GCA TCT GST AAG TCC ATC
GTG GAT TCT TAC TGG ATT TAT GAG TTT CCA ACT GGT TCT 117
CCA CTG GGT CAA ACT CAG TGG TAC TCT GAT TCG AGG AGC 156
AAA AGT AGC TTC GGC TAC ACT TCA AAA ATT GGT GAT ATT 195
CCT GCT GTA GAG GAG GAA ATT TTA TCT CAG AAC GGT CAT 234

ATC CCA GTG TTT GAT GAT ATT GAT TTC AGC ATC AAT ATC 273
AAT GAT TCT TCT TTG GCA ATT TCT GGT TGT TCC AAC ACA 312
GTT AAC ACC AAT GGA GTG AAG CAT CAG GGT CAT CTT AAG 351
GTT CTT TCT CTT GCC CAA TTG CAT CTC TTT GAA CCT GTG 390
ATG AGC AGG TCA GAG ATT GCT AGC AGA TTC CGG CTC CAA 429

GAA GAA GAT ATA ATT CCT GAT GAC AAA TAT ATA TCT GCT 468
GCT AAC AAG GGA TCT CTC TCC TGG GTC AAG GAA CAT ACT 507
TAC AAA GTC GAA ATG AGC CAC AAT CAG GCT TTA GGC AAA 546
GIG AAT GGT CTT TCT CCT AAC AGA AAT GTT CAT GAG TGG 585
CTG TAT AGT TTT AAA CCA AAT TTC AAC CAG ATC GAA AGT 624

AAT AAC AGA ACT GTA AAT TCT CTT GCA GTC AAA TCT TTG 663
CTC ATG GCT ACA GAA AAC AAC ATT ATG CCT AAC TCT CAA 702
GCT TTT GGT AAA GCT TCT ACT GAT TCT CAT TTT AAG TTG 741
AGC CTT TGG CTC AGA ATT CCA AAA GGT TTG AAG CAA ATA 780
GCC ATA CAG AAG CTC TTT AAG TTT GCA GGA GAC GAA ACC 819

GTT AAA AGT TTC TAT TTG TCT TAT ATT GCA TGC ATC CCA AAT 858
CAC AAC AGT GTG GAA ACA GCT TTA AAT GTC ACT GAT ATA 897
TGT AGA CAT CAG CTT GCA ATC CCT AAG TCC AAA GCT CCT 936
TTT GAA TTA TCA ATG ATT TTC TCC GAT CTG AAA GAG CCT 975
TAC AAC ACT GTG CAT GAT CCT TCA TAT CCT CAA AGG ATT 1014

GTT CAT GCT TTT CCT GAG ACT CAC ACT TCC TTT GCA CAA 1053
GTT TCT TGC AAC AAG CTT CAA GAA GAT GTG ATC ATA TAT 1092
ACT ATA AAC AGC CCT GAA CTA ACC CCA GCT AAG CTG GAT 1131
CTA GTG GAA AGA ACC TTG AAC AGT GAA GAT GCT TCG 1170
AAG AAG AAG TAT TTT CTT CCT AAA ACA CTC GAA TGC TTG 1209

CCA GTA AAT GTG CAG ACT ATG TCT TAT TTG GAT AGC ATC 1248
CAG ATT CTT TCA TGA AGA GTA CTG TAT GCA AGA GAG 1287
ATC AGA ATC TCC OCT CAA TCT ACT CCT ATT GCA AGA TCT 1326
TTG CTC AAG CTG GAT TTG AGC AAG ATC AAG GAA AAG AAG 1365
TCC TTG ACT TGG GAA ACA TCC AGC TAT GAT CTA GAA 1401

(2) INFORMATION FOR SEQ ID NO:19:
(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH:  777 base pairs
   (B) TYPE: Nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

ATG TCT AAG GTC AAG CTC ACA AAA GAA AAC ATT GTC TCT CTT TTA  
ACT CAA TCT GCA GAT GTT GAG TTT GAA GAA GAC CAG AAC CAG GTC  
10  GCA TTC AAC TTT AAG ACT TTC TGT CAG GAA AAT CTT GAC CTG ATT  
   AAG AAA ATG AGT ATC ACT TCA TGT TTG ACT TTT GAG AAT CGC  
   CAA GGC ATT ATG AAA GTT GTG AAC CAA AGT GAT TTT ACT TTT GGT  
   AAG GTC ACG ATA AAG AAA AAT TCT GAA AGA GTT GGA GCT AAG GAT  
   ATG ACT TTT ACG AGG CTT GAT AGC ATG ATA AGA GTG AAG CTC ATA  
15  GAA GAG ACT GCA AAC AAT GAG AAT CTT GCT ATT ATC AAA GCA AAA  
   ATT GCC TCC CAC CCT TTT GTC CAA GCT TAC GGG CTG CCT CTG GCA  
   GAT GCA AAA TCT GTG AGA CTT GCT ATA ATG CTT GGA GGT AGT ATC  
   CCT CTC ATT GCT TCT GTC GAG AGC TCC GAG ATG ATC GAT GTC GAT  
   CTT GCC ATA TAT CAA GAT GCA AAG TAC AAG GAG TTA GGG ATT GAA  
20  CCA ACT AAG TAC AAC ACT AAG GCA GCT CTG GGG AAG GTT TGC ACA  
   GTG CTT AAA AGC AAA GGA TTT ACA ATG GAT GAT GCA CAG ATA AAC  
   AAA GGG AAA GAA TAT GCC AAG ATA CTT AGT TCT TGC AAT OCC AAT  
   GCT AAG GGA AGC ATT GCT ATG GAC TAT TAC AGT GAT AAT CTT GAC  
   AAA TTC TAT GAA ATG TTT GGA GTT AAG AAA GAG GCC AAG ATT GCT  
25  GGT GTC GCA TAA  777

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH:  40 base pairs
   (B) TYPE: Nucleic acid

30  (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TACTTATCTA GAAATCGGA CAAAGCAAAG ATTACCAAGG  40

35  (2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH:  42 base pairs

72
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:
TACAGTGGA CATGGTTAT TCCAAATAAT TTATAAAAGC AC 42

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:
AGCATGGGA CATGGTTAA CACACTAAGC AAGCAC 36

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 46 base pairs
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:
AGCTAATCTA GACCATGGGA TGACTCACA AGGAAGCAT TGGTGC 46

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:
CCCAATATCC TTGGCAAGAC CC 22

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

73
(ii) MOLECULE TYPE: DNA
(x) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TACAGTTGAT CCATGTAA A GGAATCTG AGCGTTGAC 39

(2) INFORMATION FOR SEQ ID NO:26:

5
(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 40 base pairs
    (B) TYPE: Nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

10
(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

ACCTAACCAG GTGAGGCT C AAAAGAA A GCATGCTG 40

(2) INFORMATION FOR SEQ ID NO:27:

15
(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 46 base pairs
    (B) TYPE: Nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

20
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

AGCTAATCTA GACGATG AGGACGT CACTGCTA AGGAAAGCAT TGTTGC 46

(2) INFORMATION FOR SEQ ID NO:28:

25
(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 36 base pairs
    (B) TYPE: Nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

AGCATGTTG ATGATGTA A CACACTG AGCAGAC 36

(2) INFORMATION FOR SEQ ID NO:29:

30
(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 39 base pairs
    (B) TYPE: Nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

TACAGTTCTA GAACATGGA TGATGAAAG TGCTGAGG 39

74
(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 49 base pairs
   (B) TYPE: Nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO:30:
AGATTCTCTA GACCATGGTG ACTTGAAGAG GAAAGTCTGT GAGGCTTGCG

Thus while we have illustrated and described the preferred embodiments of our invention, it is to be understood that this invention is capable of variation and modification, and we therefore do not wish to be limited to the precise terms set forth, but desire to avail ourselves of such changes and alterations which may be made for adapting the invention to various usages and conditions. Such variations and modifications, for example, would include the substitution of structurally similar nucleic acid sequences in which the difference between the sequence shown and the variation sequence is such that little if any advantages are available with the variation sequence, i.e. that the sequences produce substantially similar results as described above. Thus, changes in sequence by the substitution, deletion, insertion or addition of nucleotides (in the nucleotide sequences) or amino acids (in the peptide sequences) which do not substantially alter the function of those sequences specifically described above are deemed to be within the scope of the present invention. In addition, it is our intention that the present invention may be modified to join the N genes of various isolates that provide resistance or immunity to Tospovirus infection of plants according to the present invention into a single cassette, and to use this cassette as a transgene in order to provide broad resistance to the Tospoviruses, especially to TSWV-BL, TSWV-B, and INSV. Accordingly, such changes and alterations are properly intended to be within the full range of equivalents, and therefore within the purview of the following claims.

Having thus described our invention and the manner and a process of making and using it in such full, clear, concise and exact terms so as
to enable any person skilled in the art to which it pertains, or with
which it is most nearly connected, to make and use the same;
We Claim:

1. An isolated nucleotide sequence which is selected from the group

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACATTTCCTT</td>
<td>50</td>
</tr>
<tr>
<td>TGCCATCCTC</td>
<td>100</td>
</tr>
<tr>
<td>CTAGCTTCTC</td>
<td>150</td>
</tr>
<tr>
<td>GAAGCTGTTC</td>
<td>200</td>
</tr>
<tr>
<td>TATGGCTCGT</td>
<td>250</td>
</tr>
<tr>
<td>ACTTATGTTT</td>
<td>300</td>
</tr>
<tr>
<td>AGATTTTTCT</td>
<td>350</td>
</tr>
<tr>
<td>AGATTTTTCT</td>
<td>400</td>
</tr>
<tr>
<td>CACATTGTTT</td>
<td>450</td>
</tr>
<tr>
<td>CTACACTCTT</td>
<td>500</td>
</tr>
<tr>
<td>AAATGACTTT</td>
<td>550</td>
</tr>
<tr>
<td>CTACACACCT</td>
<td>600</td>
</tr>
<tr>
<td>AGAGACATC</td>
<td>700</td>
</tr>
<tr>
<td>GAATTTAAGG</td>
<td>750</td>
</tr>
<tr>
<td>ATGCTGATG</td>
<td>800</td>
</tr>
<tr>
<td>TTTAGTACTT</td>
<td>850</td>
</tr>
<tr>
<td>GAATACTCCT</td>
<td>900</td>
</tr>
<tr>
<td>AAATACATT</td>
<td>950</td>
</tr>
<tr>
<td>ATAAAATTAA</td>
<td>1000</td>
</tr>
<tr>
<td>AAATTCATT</td>
<td>1050</td>
</tr>
<tr>
<td>AATTTCATT</td>
<td>1100</td>
</tr>
<tr>
<td>TTTTTTATT</td>
<td>1150</td>
</tr>
<tr>
<td>TTATTTTTT</td>
<td>1200</td>
</tr>
<tr>
<td>ATAGAAGAA</td>
<td>1250</td>
</tr>
<tr>
<td>GCAAGACCA</td>
<td>1300</td>
</tr>
<tr>
<td>AGCAGGATT</td>
<td>1350</td>
</tr>
<tr>
<td>TATATTACAA</td>
<td>1400</td>
</tr>
<tr>
<td>CTATTTAAC</td>
<td>1450</td>
</tr>
<tr>
<td>GCTGAAACTT</td>
<td>1500</td>
</tr>
<tr>
<td>CCTTGCTGGT</td>
<td>1550</td>
</tr>
<tr>
<td>CCCGGCATCC</td>
<td>1600</td>
</tr>
<tr>
<td>ATTGTGGAA</td>
<td>1650</td>
</tr>
<tr>
<td>TTGGAGTTAA</td>
<td>1700</td>
</tr>
<tr>
<td>CAGATGTTCC 1709;</td>
<td></td>
</tr>
<tr>
<td>TTACGACTT</td>
<td>50</td>
</tr>
<tr>
<td>TTGACTGCTT</td>
<td>100</td>
</tr>
<tr>
<td>CTGGGTTTTT</td>
<td>150</td>
</tr>
<tr>
<td>TGACAACTT</td>
<td>200</td>
</tr>
<tr>
<td>TGGATCTAAC</td>
<td>250</td>
</tr>
<tr>
<td>TTGATCTGCA</td>
<td>300</td>
</tr>
<tr>
<td>TTCCCTGTTT</td>
<td>350</td>
</tr>
</tbody>
</table>
CATCCTGTATA TATAGCCGAG ACAACACTGA TCATCCTCAA GCTATCAGCT
400
75

GAAGCCATTA GAGTTAAGCT ACCCTGCAGC ATTAATGGCA GCTACAGCA
450

CTTTCAGCCA TCAAGGGTTA ATCCATAGGC TTGAATCAA GGGTGGGAAG
500

CATCCTTAGA TTGGATAGTA TTGGATATCT CAGATATCTT AGGTCTCTCT
550

ACAAGCTGTA CCGTCACTCA GCTATGCAGC CTCTGTGAAGG TCATGTCAGT
600

GCTGCTGACT GCTGGTGAGC TTCTCTTTAT CTAAATTTGA CCAAAGTCAA
650

AATCTTCTTG GATTGATACC TTGATATTGC TCGCAGTGTG CTTCAAGGAT
700

GAGAAAGCAG AAGGGATGTC ATCTGGATTTT GATCTGCTCA GGTGTTCGAG
750

ACAAAAAGAG TTAGAGTTGA ATGCTGCTCA ATCTTGATCTT TCTCAAGCTA
800

CAGAGCCTTT GCGTGGTCAC AACAAAGCAA CAGTCCTTTC CCTACTGACC
850

TTAACAT 858;

AAATTCTCTTT GCGATGAAAT GCTGTGCTGAT GTGAGCGAAA AACAACATCA
50

TCTCTACACT TCAAGCTTTT TCTGAAGGCTT CTAACAGTGC TAAATTCAGA
100

CCTGAGCTCT GCCTAAAGGTT TCCAAAGGGT TTAGCGAGA TTTCAATCA
150

GAAATTGTTT CAGATGGCAG GAGATGAAC AAAAAAACAA TTTATTTAT
200

CINAACGTGG TCAATCAAGA CATACAGTGC TTAGACGAC TTTAAGACCT
250

ACGGTATTGT CCGACTGACA GCTCCAAATG GTAAAGTAGT AAAAACTCTT
300

TGGATTACCA ATGAGGTCTT CIGATTTAAA GGAGCGTCTAC AGGATATATC
350

CTGATCCTTG CAACTCTGCA CAGAATGTTT ATCCTGTGCT AGGATACGT
400

ACAAAAAGAG TTAGAGTTGA ATGCTGCTCA ATCTTGATCTT TCTCAAGCTA
450

CTACACCTTG AAACCAAGATC AGTACATGCT TGGAAAGTTA GTTATTAGTG
500

AAATAAGATT GTTAAAGCAT GAGAAGCGCT ACAAAAGCAA ATATTTTGCT
550

TCAAAAACAC TTGAAGCTCT TCTAATCAAC ATCAAAACTA TGTCTATTAT
600

AGCAGACATC CAAATCCCTT CCGGAAGAT AGACTTGGGC AGGGGAGAA
650

TTAAATATTC TCCACPAATCT MTCCAGGTGT CAAATCTTTC GTTAAATCTT
700

GATTCTAGGG GAGTTAAA AAGAATATT AAGTATAAGG AAGCATATGC
750

TTGACCGATC AAGATATCTT GCTGTTCCGA CTTTTTCCTT ATTAATGTAT
800

GTTATTCTTG TTCTTTATCT TTAATGTTAT TTTCTGTTTA TCTATCTCTT
850

CAGAACTCTT CTCTTCAGTA GAAACATTA AAAAACAAAAT AAAAAAAAAT
900

TAAAAACAA AAAAAATAAA AATCAGAAAA TGAATATAAA GCAAACAAAA
950

AATTTAAAAA AAAAAACACA AAAAAAGAATC CCAAAGCCA AATTGSGCCA
1000

AATTGGGTTT TTGGTTTTGT TTTTTGTTTTT TTTTTTTTTT
1050

TTATTATTAT TTATTATTATT ATTATTATTT ATTATTATTATTT TTTATTATT
1100

TTGTATTATT AATTTATTAT AATTTATTAT AATTTATTAT AATTTATTAT
1150

TAAACAGACT TAAATTAAAT TTAACACACT AAGCAAGGC AAAAAATAAA
1200

GATAAAGAA GCTTTATATA TTAATAGGCT TTTTATTATT TTAACACTAA
1250

GGCTTGGTTA AGATACAGCT GTGAGTAGGT CCTGTTGTTT AACCCCAAC
1300

ATTATATAGA ACATGTTAGG GTGGTGAGCG TAAATGCTCA TAGCATACTC
1350

TOCTTTGAGA TTAGACGCTG TGGATAGTA TATAGGACTA TCTCTCTCTC
1400

CTTCTCTCAG CTCATCTGCA TCATTTTCAA ATGCTTTTCT TTTGCAAGCA
1450

GIGGAAACITCT TTCTCGAGGC TTCTCTGGTG TCGACTCTCT TGGGTGGGAT
1500

CCGAGAGTTC TTTGATTTTG CATCCAGTAA TATAGCCAGA AACAACATG
1550

TCAATGCAA GCGATCAGCT GAGCAATCAA GAGGAATG CTCCACAGGC
1600

ATTATGACAA GCTCCAGAGA CTTGACATCA TCAAGGGATA ATCCATAGGC
1650
AAAATTTGCT TTGTGGTTTG TTTTGTGT TTGTTGTTTT TTGGTTTTATT 1850
TTATTTTTTA GTTTATTTTT TTTTTTTTT ATTATTATTATT TTTATTATTATT 1900
TCTTTTTATT TTATTTTAT ATATTACAAA CACAACTCAG ACACAAATTT 1950
TTATTTCTCA ACTATTCTACT GTATTACAC ACTTACCTTTG ACTTATTACAC 2000
ACTTACGAG CTGTTAGCTC CTGTTACAGCA CTGGACTGGA TTAGACACCA 2050
CTTTGTTATTA TCTTCACTCCT TAAATACAGCA CATTTAAATAT ATGCTACTTCT 2100
GAATCTACCT TAAAGAAGCT TTTATAGCAAC ACCAGAATCC TTGCTCTCTC 2150
TCITACCTCC AAAATTCTCA TAGAATTCTG TAAGAATTTAT ACTGTAATTAG 2200
TGCTATTCAAG TGCTCCCTCT AAGATCTGGAA TTGGAGAGAC TAGCATCTCTT 2250
GCATATTCCT TTCCTTTATT TTAATCCTGGC ATCATCCATT GTGAATCCTC 2300
TGCTTTTAGC CACTGTCGAA ACCCTCCGCA GACCTTCCTT AGIGTGGTAC 2350
TTAGTCTGGT CAATTCCCTAA CTTCTGTGAC TTGGCTCCTT GAATATGGGC 2400
AAGAAACACA TAGTACATCT CGAAGTGGTG AACAGGACAGA ATGAGGGGGA 2450
TACTACCTCC AAAGACTTATA GCAAAGTACA CAGATTTGGC ATCTGCGACA 2500
GGCAACGCTT AAGGCTTGAC CGAAGGCTGG GAGGCAATT TGGGATTGTT 2550
AATTAGCAAGA TTCTCATTTG TTTGACCTCT TCTTATGACC TTTACCTTTTA 2600
TCAATGCTATC AAGCTCTCTG AAAGTCATAT CTTAGCTCTC AACTCTTCTTCA 2650
GAAATTTTCTT TTTACGCGAC CTATACAAAA GTAAAATCAC TTTTGTTTGC 2700
AAGCCTTTAA ATGCTCTGGA GTATCTTCAA GAAGGCTAAA CATGAGTGCA 2750
TACTATTTTT TTTTATCCTAGG TCTAGATTCT TCTGACAGAA AGTCTAAGAC 2800
TTGACTAGCA CCTGGTTCCTG GCTTCCTCA AACTCAACAT CTGCGAGATTG 2850
AGTTAAAAGA GAGACTATGT TTTCTTTTGT GAGCTTGACC TTAGCAGATTG 2900
TGGCAAGTTTA GATCTACAGCC TTTTCAGAGA GTAAGAATTC AAGTTGAGPA 2950
AGCTGAACAC TGCTAGAGGC GTGTCATACT TATCTGTTTA ATGAGATGAT 3000
TTGTAATGCTT GATGATAGGTTTTTGGATA AAATGAGCAC ATTTGCTCT 3049

2. A plant susceptible to infection by Tospoviruses which has a transgene inserted into its genome to render it resistant to infection by Tospoviruses, said transgene being selected from the group consisting of the nucleoprotein gene of TSWV-BL, TSWV-10W, INSV-LI, TSWV-B, a Tospovirus, said transgene consisting of partial or full length nucleoprotein gene sequences from TSWV-BL, TSWV-10W, TSWV-B, INSV-Beg and INSV-IL, the translatable or untranslatable sequences of said nucleoprotein gene sequences, and the sense or antisense sequences of said nucleoprotein gene sequences.

3. A method for providing a host plant with resistance to infection by Tospoviruses which comprises inserting a transgene into the host plant which gene is selected from the nucleoprotein gene of TSWV-BL, TSWV-10W, INSV-Beg, INSV-LI, TSWV-B, or mixtures of nucleotide sequences taken from the nucleoprotein gene.
Fig. 1

Viral RNAs

primer JSL90-46

reverse transcriptase

3' AGT GTA 5' cDNA (first strand)

PCR

nucleoprotein gene

cloning

EcoRI AGT GTA AMV 3' U3 5' Ss-558 HindIII

pBI525-NP+

Fig. 2

A at 405 nm

- 1/5 -
Fig. 3

Fig. 4

- TSWV-CPNH1
- TSWV-BL
- TSWV-L3
- TSWV-B
- TSWV-I
Fig. 5

A at 405nm

Health  TSWV-BL  Arkansas  10W pakchoy  Begonia  Brazil

Fig. 6

Percentage of plants infected

TSWV-BL  Begonia

0-0.05  0.05-0.2  0.2-0.4  0.4-1.0

The NP accumulated in plants (ELISA, A at 405nm)
Fig. 7

pB1525 - 35S-35S AMV 5'-UT Nos 3' -

pBIN19-N - TSWV-BL N gene -

ATGGTTAAGCTCACAAGGAAAGCATTGTGGTTGTTGACCA

ATGGATGA-CTCAGGGGAAGCATTGTTGCTTTGTTGACCA

pB1525-mN - 35S-35S AMV 5'-UT Untranslatable N gene Nos 3' -

pB1525-asN - 35S-35S AMV 5'-UT Assembled N gene Nos 3' -
FIG. 8

-5/5-
INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
IPCI: A01H 1/04; C07H 17/00; C12N 15/00
US CI: 800/205; 435/172.3; 536/23.1
According to International Patent Classification (IPC) or to both national classification and IPC.

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
U.S.: 800/205; 435/172.3; 536/23.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
APS, DIALOG

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category*</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>EP, A, 0,426,195 (GOLDBACH ET AL) 08 MAY 1991, pages 3-5 and 15-16.</td>
<td>1-3</td>
</tr>
<tr>
<td>Y</td>
<td>BIO/TECHNOLOGY, Volume 9, issued December 1991, Gielen et al, &quot;Engineered Resistance to Tomato Spotted Wilt Virus, a Negative-Strand RNA Virus&quot; pages 1363-1367, see page 1365.</td>
<td>1-3</td>
</tr>
<tr>
<td>Y</td>
<td>JOURNAL OF GENERAL VIROLOGY, Volume 70, issued 1989, de Haan et al, &quot;Molecular Cloning and Terminal Sequence Determination of the S and M RNAs of Tomato Spotted Wilt Virus&quot; pages 3469-3473, see page 3470.</td>
<td>1-3</td>
</tr>
</tbody>
</table>

[X] Further documents are listed in the continuation of Box C. See patent family annex.

Date of the actual completion of the international search
29 MARCH 1994

Date of mailing of the international search report
APR 06 1994

Name and mailing address of the ISA/US Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Fax Number: NOT APPLICABLE

Authorized officer
ELIZABETH F. MOGILWAIN

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
</table>