

(19)



(11)

**EP 3 822 353 A1**

(12)

**EUROPEAN PATENT APPLICATION**

(43) Date of publication:  
**19.05.2021 Bulletin 2021/20**

(51) Int Cl.:  
**C12N 15/67<sup>(2006.01)</sup>**

(21) Application number: **19382997.5**

(22) Date of filing: **13.11.2019**

(84) Designated Contracting States:  
**AL AT BE BG CH CY CZ DE DK EE ES FI FR GB  
 GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO  
 PL PT RO RS SE SI SK SM TR**  
 Designated Extension States:  
**BA ME**  
 Designated Validation States:  
**KH MA MD TN**

- **Universitat Politècnica de València  
46022 Valencia (ES)**

(72) Inventors:

- **DARÓS ARNAU, José Antonio  
46022 Valencia (ES)**
- **ORTOLÁ NAVARRO, Beltrán  
46022 Valencia (ES)**
- **CORDERO CUCART, María Teresa  
46022 Valencia (ES)**

(71) Applicants:

- **Consejo Superior de Investigaciones Científicas  
(CSIC)  
28006 Madrid (ES)**

(74) Representative: **Pons  
Glorieta Ruben Dario 4  
28010 Madrid (ES)**

(54) **METHOD FOR THE PRODUCTION OF DSRNA**

(57) The present invention relates to an isolated nucleotide acid sequence comprising a) the cDNAs of two strands of a target dsRNA separated by an autocatalytic intron flanked by exon fragments, and b) a plant viroid, wherein element (a) is inserted in the plant viroid sequence. The expression of this nucleotide sequence in

a host cell, such as *E. coli*, along with a tRNA ligase, allows the production of high amounts of dsRNA specific for a target gene. This dsRNA can be used in the interfering RNA technology for silencing gene expression. Additionally, the present invention also relates to a method for producing said dsRNA.

**EP 3 822 353 A1**

**Description**

5 [0001] The present invention relates to a nucleotide sequence comprising a) the cDNAs of two strands of a target double-stranded RNA (dsRNA) separated by an autocatalytic intron flanked by exons, and b) a plant viroid sequence, wherein element (a) is inserted in the plant viroid sequence. The expression of this nucleotide sequence in a host cell, such as *Escherichia coli*, along with a tRNA ligase, allows the production of large amounts of dsRNA specific for a target gene. This dsRNA can be used in the interfering RNA technology for silencing gene expression. Thus, the present invention belongs to the field of biotechnology, specifically, to the field of the production of recombinant RNA.

10 **BACKGROUND ART**

15 [0002] Gene silencing mediated by RNA or RNA interference (RNAi) is a natural defense mechanism against genetic elements, both exogenous (for example, viruses) and endogenous (for example, transposons), which in most eukaryotic organisms has evolved to become an important regulatory way of gene expression. RNAi is triggered by the presence of double-stranded RNA (dsRNA) and has the ultimate consequence of silencing or suppressing homologous messenger RNA (mRNA). Very briefly, the dsRNA is processed by the Dicer endonucleases to produce small RNA (sRNA) that is recognized by the RNase H Argonaute (Ago), which carries one of the strands of the sRNA to form the RNA-induced silencing complex (RISC). In the cell, this complex traces its target RNA based on the complementarity of the sRNA and, when it finds it, it cuts it down inactivating it or it joins it, inhibiting its translation. In some taxonomic groups such as nematodes or plants, an RNA-dependent RNA polymerase (RDR) activity uses sRNA as a primer to produce more homologous dsRNA, which feeds the cycle from its inception and amplifies the interfering signal.

20 [0003] RNAi can be efficiently induced by the administration of exogenous RNA, so it has been widely used in basic research to study gene function in many classes of organisms. At the same time, its conceptual simplicity (highly specific gene suppression based on simple base complementarity) has driven the development of a multitude of promising biotechnological applications. One of them is the use of RNA against pests: the ingestion of dsRNAs by nematodes, insects and other arthropods induces the silencing of endogenous homologous genes, causing their death or affecting their development, mobility or feeding behavior, reducing in any case the damage they cause. The silencing induced by dsRNAs has also been demonstrated in pathogens such as viruses and fungi. In this way, RNA, a natural compound that is harmless to the environment, has become a promising agent against pests and pathogens. In addition, its mechanism of action based on base complementarity makes it highly specific and easily adaptable to any new species or lineage of pest or emerging pathogen. Its use in agricultural crops has aroused great interest but its potential goes beyond agricultural biotechnology.

25 [0004] The biotechnological use of dsRNAs requires, however, systems capable of producing large quantities of this type of compound quickly and economically. The classical strategies for the production of RNAs based on chemical synthesis or in vitro transcription are not viable due to their cost and yield. The most logical alternative is the production of dsRNA in biofactory systems such as bacterial cultures, for example, of *E. coli*.

30 [0005] The patent application WO 2015/177100 discloses the production of RNA by co-expressing a tRNA ligase and a chimeric RNA molecule comprising a target RNA and a plant viroid scaffold in a host cell. This co-expression causes the production of large amounts of the chimeric RNA molecule in the host cells and may be useful, for example, in the production of RNA aptamers and other RNA molecules. Recently, a system based on elements of the biology of viroids to produce large amounts of recombinant RNA in *E. coli* has been developed (Daros et al., 2018. Sci. Rep. 8: 1904).

35 [0006] Viroids are a class of infectious agents of plants constituted exclusively by a noncoding circular RNA of relatively small size (246-401 nucleotides, in the known species to date). In spite of not coding own proteins, the viroid RNA manages to replicate itself and move around the plant, for which it has to recruit proteins and take advantage of the cellular structures of the host. The system of production of recombinant RNA in *E. coli* disclosed in Daros et al. 2018 (cited *ad supra*) is based on the co-expression of an RNA derived from the *Eggplant latent viroid* (ELVd), which contains the RNA of interest, and the chloroplastic isoform of the eggplant (*Solanum melongena* L.) tRNA ligase, an enzyme that mediates the circularization of this viroid in the infected plant. ELVd does not replicate in *E. coli*. However, when a precursor containing it is transcribed, its hammerhead ribozymes act to generate a linear monomer with the RNA of interest inserted. This monomer is recognized by the tRNA ligase and circularized. The resulting product, a hybrid molecule consisting of a highly structured viroid scaffold on which the RNA of interest is presented, is accumulated in enormous quantities in the bacterial cells as a consequence of its circularity and formation of a ribonucleoprotein complex with the tRNA ligase. With this system, quantities in the order of 100 mg of aptamers of RNA and other structured RNAs per litre of culture have been produced. However, due to the known instability of inverted DNA repeats in *E. coli*, this system was shown from the start to be incapable of producing substantial quantities of dsRNAs, the molecules necessary to induce gene silencing in pests and pathogens. Therefore, there is still the necessity of developing methods and processes that allow producing large amounts of dsRNA, overcoming the drawbacks of the systems, which already exist in the state of the art.

## SUMMARY OF THE INVENTION

[0007] Thanks to the use of autocatalytic introns in *E. coli*, the present invention allows to overcome the above-mentioned limitation of the state of the art, and to produce large quantities of dsRNAs which may be used against pests and pathogens. In addition, using these same autocatalytic introns permuted, it is possible to cleave the dsRNA of interest from the viroid scaffold on which it is transcribed, producing a pure circular dsRNA molecule without any residue of the viroid.

[0008] To solve the problem of the instability of inverted DNA repeats in *E. coli*, the repeats were separated with a cDNA corresponding to the group I intron of the 26S ribosomal RNA (rRNA) of prilled ciliated *Tetrahymena thermophila* (order Hymenostomatia). Since this intron is autocatalytic and only requires a guanosine nucleoside for its processing, once the transcript is expressed in *E. coli*, the intron is self-cleaved with extraordinary efficiency and a hairpin of dsRNA is formed closed by a loop of 20 nucleotides corresponding to two 10-nucleotides-long fragments of the exons that flank the intron. Thus, in the culture of *E. coli*, large quantities of hybrid circular molecules were produced efficiently consisting of an ELVd scaffold on which the pesticide dsRNA hairpin is presented (Figure 1).

[0009] Using this system, a 100 bp double-stranded RNA against essential genes of insect pests was produced and its pesticide activity demonstrated. However, it is understood that the commercial potential of this technology may be limited by the presence in the molecule of dsRNA of interest of a residue derived from the viroid.

[0010] In order to eliminate the viroid moiety, a new element consisting of a permuted intron-exon sequence was added to the system (Figure 2). When an intron of group I is divided in two parts, and the first is fused following the second, the self-cleavage reaction continues on the border between exon 1 and the 5' end of the intron and continues with the attack on 3' end from exon 1 to the border between the 3' end of the intron and exon 2, autocatalytically generating a circular RNA molecule formed by the two exons (Figure 3). In this way, the viroid moiety (Figure 4) is completely removed from the final dsRNA product (Figure 3).

[0011] In summary, this invention allows the production of large amounts of dsRNA in *E. coli*. To do this, the cDNAs of the two strands of the dsRNA are inserted into the cDNA of a viroid, and separated by the cDNA of a type I intron. The co-expression in *E. coli* of the resulting RNA together with the tRNA ligase of eggplant produces a large accumulation of a hybrid molecule in which the dsRNA is fused to the viroid. The autocatalytic intron disappears during the processing of the precursor. In addition, if a permuted type I intron is added to this construct, the dsRNA of interest in a circular form can be separated from the viroid molecule.

[0012] In view of the foregoing, in an aspect, the present invention relates to an isolated nucleotide sequence comprising

- a) the cDNAs of two strands of a target dsRNA separated by an autocatalytic intron flanked by exons or exon fragments, and
- b) a plant viroid sequence,

wherein element (a) is inserted in the plant viroid sequence.

[0013] Additionally, in another inventive aspect, the present invention relates to a method for producing dsRNA specific for a target gene comprising co-expressing in a host cell

- a) a nucleotide sequence according to the above-mentioned aspect, and
- b) a nucleotide sequence encoding a tRNA ligase,

wherein the nucleotide sequences of a) and b) are in the same or different nucleic acid molecule.

## DETAILED DESCRIPTION OF THE INVENTION

[0014] In an aspect, the present invention relates to an isolated nucleotide sequence, hereinafter "nucleotide sequence of the invention", comprising

- a) the cDNAs of two strands of a target dsRNA separated by an autocatalytic intron flanked by exons or exon fragments, and
- b) a plant viroid sequence,

wherein element (a) is inserted in the plant viroid sequence.

[0015] As used herein, the term "double-stranded RNA" or "dsRNA" refers to a RNA molecule comprising two nucleic acid strands containing complementary sequences each other, wherein one of the strands is called sense strand (direction 5'-3') and the other one is called antisense strand (direction 3'-5'). Thus, the nucleotide sequence of the invention comprises an inverted repeat sequence comprising the sense and antisense strands of the target RNA separated by

an intervening sequence comprising an intron type I flanked by exons or exon fragments.

[0016] As used herein, the term "target dsRNA" is the dsRNA capable of inducing the expression silencing of a specific target gene with which it shares homology. In the context of the present invention, it is considered that two nucleotide sequences share homology when the sequence identity between them is, at least, 75 % by the optimal alignment of them. Tools for determining the sequence identity between two nucleotide sequences are disclosed below. Any dsRNA molecule that needs to be generated in large amounts may be used as the target dsRNA. The target dsRNA may be non-naturally occurring. Suitable target dsRNA may include sequences up to 100 bases, up to 200 bases or up to 300 bases or more. Preferably, the target dsRNA is not derived from a plant viroid and is not naturally associated with the plant viroid sequence (i.e. the target dsRNA is heterologous). Examples of specific target genes include, without being limited to, the target genes shown in Table 1.

Table 1. Target genes useful for fighting against pests and pathogens

Target insect	Target gene	Functions
<i>Helicoverpa armigera</i>	<i>Cytochrome 450 monooxygenase CTP6AE14</i>	Degrade gossypol and increase tolerance to gossypol
	<i>GST glutathion transferase</i>	Detoxification enzymes
	<i>HaHR-3</i>	Molt-regulating transcription factor
	20-Hydroxyecdysone	Role in molting and metamorphosis
	Arginine kinase	Play role in role in cellular energy metabolism
	Chitin synthase monooxygenase	Synthesis of chitin
	Cytochrome P450	Required for tolerance against gossypol
	V-ATPase	Role in cellular energy production
	Chitinase	Role in molting and metamorphosis
<i>Diabrotica virgifera virgifera</i> LeConte	<i>Vacuolar A TPase</i>	Role in cellular energy production
	<i>DVSSJ1</i>	production of their respective smooth septate junction membrane proteins
	<i>DVSSJ2</i>	
<i>Spodoptera exigua</i>	20-Hydroxyecdysone	Role in molting and metamorphosis
<i>Manduca sexta</i>	MsCYP6B46	Nicotine degradation
<i>Nilaparvata Lugens</i>	Hexose transporter gene	Transport of glucose
	Carboxypeptidase	Hydrolysis of protein
	Trypsin like serine protease	Hydrolysis of protein
<i>Myzus persicae</i>	Rack1	Essential in foraging and feeding of the pea aphid
	MpC002	Intracellular receptor
	Hunchback (hb)	Play role in insect axial patterning
	Serine proteases (SP)	Gut proteolytic digestion
<i>Bemisia tabaci</i>	v-ATPase	Encode for v-ATPase subunit A, role in cellular energy production
	Acetylcholinesterase (AChE)	Required in signaling
	Ecdysone receptor (EcR)	Role in molting and metamorphosis

[0017] In a particular embodiment, the target dsRNA is the gene beta-actin from *Ceratitidis capitata* (diptera: *Tephritidae*). In a more particular embodiment, the cDNA sequence of the sense strand of the target dsRNA comprises, or consists of, a nucleotide sequence with a sequence identity of, at least, 70, 75, 80, 85, 90, 95, 96, 97, 98 or 99% with SEQ ID NO: 1. In a still more particular embodiment, the cDNA sequence of the sense strand of the target dsRNA comprises, or consists of, a nucleotide sequence with a sequence identity of 100% to SEQ ID NO: 1. In another particular embodiment,

the cDNA sequence of the anti-sense strand of the target dsRNA comprises, or consists of, a nucleotide sequence with a sequence identity of, at least, 70, 75, 80, 85, 90, 95, 96, 97, 98 or 99% with SEQ ID NO: 2. In a still more particular embodiment, the cDNA sequence of the sense strand of the target dsRNA comprises, or consists of, a nucleotide sequence with a sequence identity of 100% to SEQ ID NO: 2.

5 SEQ ID NO: 1

GCTGACCGAAGCACCGTTGAACCCGAAGGCTAACCGCGAGAAGATGACCCAGATCATGTTTCGAG

10 ACCTTCAACTCGCCGGCCATGTATGTGGCCATCCAG

SEQ ID NO: 2

15 CTGGATGGCCACATACATGGCCGGCGAGTTGAAGGTCGGAACATGATCTGGGTCATCTTCTCG

CGGTTAGCCTTCGGGTTCAACGGTGCTTCGGTCAGC

[0018] As used herein, the term "sequence identity" refers to the degree of similarity between two nucleotide (or amino acid) sequences obtained by the optimal alignment of the two sequences. A degree of identity expressed as a percentage will be obtained based on the number of common residues among the aligned sequences. The degree of identity between two nucleotide (or amino acid) sequences can be determined by conventional methods, for example, standard sequence alignment algorithms known in the state of the art, such as BLAST [Altschul S.F. et al. Basic local alignment search tool. J Mol Biol. 1990 Oct 5; 215(3):403-10]. The BLAST programs, for example, BLASTN, BLASTX and TBLASTX, BLASTP and TBLASTN are a part of the public domain on The National Center for Biotechnology Information (NCBI) website. Particular nucleotide sequence variants may differ from a reference sequence by insertion, addition, substitution or deletion of 1 base, 2, 3, 4, 5-10, 10-20, 20-30 or more bases.

[0019] The cDNAs of two strands of a target dsRNA in the nucleotide sequence of the invention are separated by an autocatalytic intron flanked by exons or exon fragments. As used herein, the term "autocatalytic intron" refers to self-splicing introns which do not need the action of a spliceosome to be removed from the nucleotide sequence. There are three kinds of self-splicing introns, Type I, Type II and Type III. Any of them can be used in the present invention. Thus, in a particular embodiment of the nucleotide sequence of the invention, the autocatalytic intron is an autocatalytic intron type I or an autocatalytic intron type II or an autocatalytic intron type III.

[0020] Type I introns (or Group I) are large self-splicing ribozymes that catalyze their own excision from RNA precursors. They perform two sequential transesterification reactions. The core structure consists of nine paired regions (P1-P9). An exogenous guanosine or guanosine nucleotide (exoG) first docks onto the active G-binding site located in P7, and its 3'-OH is aligned to attack the phosphodiester bond at the 5' splice site located in P1. This results in a free 3'-OH group at the upstream exon and the exoG attached to the 5' end of the intron. Then, the terminal G (omega G) of the intron replaces the exoG and occupies the G-binding site to organize the second transesterification reaction. In this reaction, the 3'-OH group of the upstream exon in P1 is aligned to attack the 3' splice site in P10, leading to the ligation of the adjacent upstream and downstream exons and release of the catalytic intron. Examples of Type I introns include, without being limited to, autocatalytic intron type I of the rRNA 26S from *Tetrahymena thermophila*, *Ankistrodesmus stipitatus* LSU rRNA intron, *Saccharomyces cerevisiae* LSU rRNA mitochondrial intron or *Podospira anserina* mitochondrial apocytochrome b intron.

[0021] Type II introns (Group II) are a large class of self-splicing ribozymes and mobile genetic elements in which, unlike Group I introns, excision occurs in the absence of a guanosine nucleotide and a lariat structure is formed. The lariat contains an A-residue branchpoint what resembles that found in lariats formed during splicing of nuclear pre-mRNA. The secondary structure of group II introns is characterized by six typical stem-loop structures, domains I to VI (D1 to DVI, or D1 to D6). The domains radiate from a central core that brings the 5' and 3' splice junctions into proximity. The proximal helix structures of the six domains are connected by a few nucleotides in the central region.

[0022] Type III introns (Group III) are a class of self-splicing ribozymes that have a conventional Group II-type DVI with a bulged adenosine, a streamlined DI, no DII-DV, and a relaxed splice site consensus. Splicing is done with two transesterification reactions with a DVI bulged adenosine as initiating nucleophile. The intron is also excised as a lariat.

[0023] In a more particular embodiment of the nucleotide sequence of the invention, the autocatalytic intron is an autocatalytic intron of type I. In a still more particular embodiment, the autocatalytic intron of type I is the autocatalytic intron type I of the rRNA 26S from *Tetrahymena thermophila* (GenBank access number V01416.1 [SEQ ID NO: 23]), or a fragment thereof with autocatalytic activity. The term "fragment" means a nucleotide sequence having one or more (e.g., several) nucleotides absent from the 5' and/or 3' ends of a reference nucleotide sequence, wherein the fragment encodes a protein having the same activity than the reference nucleotide sequence, being thus a functionally equivalent

fragment. In still more particular embodiment, the autocatalytic intron type I of the rRNA 26S from *T. thermophila* comprises, or consists of, a nucleotide sequence with a sequence identity of at least 70, 75, 80, 85, 90, 95, 96, 97, 98 or 99% with SEQ ID NO: 3. In a more particular embodiment, the autocatalytic intron comprises, or consists of, a nucleotide sequence with a sequence identity of 100% to SEQ ID NO: 3.

5 SEQ ID NO: 3

AAATAGCAATATTTACCTTTGGAGGGAAAAGTTATCAGGCATGCACCTGGTAGCTAGTCTTTAAACCAAT  
 10 AGATTGCATCGGTTTTAAAAGGCAAGACCGTCAAATTCGCGGAAAGGGGTCAACAGCCGTTTCAGTACCAAG  
 TCTCAGGGGAAACTTTGAGATGGCCTTGCAAAGGGTATGGTAATAAGCTGACGGACATGGTCTTAACCAC  
 GCAGCCAAGTCCTAAGTCAACAGATCTTCTGTTGATATGGATGCAGTTCACAGACTAAATGTGGTTCGGG  
 GAAGATGTATTCTTCTCATAAGATATAGTCGGACCTCTCCTTAATGGGAGCTAGCGGATGAAGTGATGCA  
 15 ACACTGGAGCCGCTGGGAAC TAATTTGTATGCGAAAGTATATTGATTAGTTTTGGAGTACTCG

[0024] In the present invention, the autocatalytic intron is flanked by exons or exon fragments, which may proceed from the same or different microorganism than the autocatalytic intron. The term "fragment" has been defined above. In a particular embodiment, the autocatalytic intron and the exons (or exon fragments) proceed from the same microorganism, particularly, said microorganism is *T. thermophila*. Likewise, the exons (or exon fragments) in 5' and 3' position with respect to the autocatalytic intron may proceed from the same or different microorganism. In a more particular embodiment, the exon (or exon fragment) at 5' position of the autocatalytic intron comprises, or consists of, the sequence SEQ ID NO: 4 and/or the exon (or exon fragment) at 3' position of the autocatalytic intron comprises, or consists of, the sequence SEQ ID NO: 5.

25 SEQ ID NO: 4 - ATGACTCTCT

SEQ ID NO: 5 - TAAGGTAGCC

30 [0025] Thus, in a more particular embodiment, the autocatalytic intron flanked by exons (or exon fragments) comprises, or consisting of, the sequence SEQ ID NO: 6, wherein the exons are shown below underlined.

SEQ ID NO: 6

35 ATGACTCTCTAAATAGCAATATTTACCTTTGGAGGGAAAAGTTATCAGGCATGCACCTGGTAGCTAGTCT  
 TTAAACCAATAGATTGCATCGGTTTTAAAAGGCAAGACCGTCAAATTCGCGGAAAGGGGTCAACAGCCGTT  
 CAGTACCAAGTCTCAGGGGAAACTTTGAGATGGCCTTGCAAAGGGTATGGTAATAAGCTGACGGACATGG  
 40 TCCTAACCACGCAGCCAAGTCCTAAGTCAACAGATCTTCTGTTGATATGGATGCAGTTCACAGACTAAAT  
 GTCCGGTCCGGGAAGATGTATTCTTCTCATAAGATATAGTCGGACCTCTCCTTAATGGGAGCTAGCGGATG  
 AAGTGATGCAACACTGGAGCCGCTGGGAAC TAATTTGTATGCGAAAGTATATTGATTAGTTTTGGAGTAC  
 45 TCGTAAGGTAGCC

[0026] The nucleotide sequence of the invention also comprises a plant viroid sequence. The plant viroid sequence is deemed to be the scaffold of the nucleotide sequence of the invention on which the rest of the components (the cDNAs of two strands of a dsRNA separated by an autocatalytic intron flanked by exons) are inserted. As used herein, the term "plant viroid" refers to the smallest infectious pathogens known comprising solely of a short strand of circular, single-stranded RNA that has no protein coating. The viroid replication mechanism uses the host cell enzyme RNA polymerase II or chloroplastic RNA polymerase, which catalyze "rolling circle" synthesis of new RNA using the viroid RNA as a template. Taxonomically, plant viroids are divided in two families: Family *Pospiviroidae* comprising the genera *Pospiviroid*, *Hostuviroid*, *Cocadviroid*, *Apscaviroid* and *Coleviroid*, and family *Avsunviroidae* comprising the genera *Avsunviroid*, *Pelamoviroid* and *Elaviroid*. Any viroid belonging to one of these families can be used in the context of the present invention. Examples of viroids include, without being limited to, *Potato spindle tuber viroid* (PSTVd), *Citrus exocortis viroid* (CEVd), *Chrysanthemum chlorotic mottle viroid* (CChMVd), *Hop stunt viroid* (HSVd), *Peach latent mosaic viroid* (PLMVd), *Coconut cadang-cadang viroid* (CCCVd), *Apple scar skin viroid* (ASSVd), *Coleus blumei viroid 1* (CbVd-1),

## EP 3 822 353 A1

*Avocado sunblotch viroid* (ASBVd), *Peach latent mosaic viroid* (PLMVd), and *Eggplant latent viroid* (ELVd). In case that the plant viroid sequence used in the present invention does not comprise a hammerhead ribozyme domain, i.e. it belongs to the Family *Pospiviroidae*, the enzymes responsible for strand scission of the viroids are co-expressed together with the plant viroid or the nucleotide sequence of the invention is expressed in a host cell comprising said enzymes. Alternatively, if the plant viroid sequence used in the present invention comprises a hammerhead domain, i.e. belongs to the family *Avsunviroidae*, two functional hammerhead domains, one to each side of the plant viroid sequence, have to be added to the plant viroid sequence, duplicating the hammerhead domain of the plant viroid. In a particular embodiment, the plant viroid sequence comprises a hammerhead ribozyme domain. In another particular embodiment, the plant viroid sequence is the ELVd sequence, the ASBVd sequence, the PLMVd sequence, or the CChMVd sequence.

[0027] The plant viroid sequence may comprise all or part of a plant viroid. The full-length (333 nt) genomic sequence of ELVd is publically available under the GenBank entry number AJ536613.1 GI: 29825431 (SEQ ID NO: 7). The sequence and structure of ELVd is shown in Figure 4. The hammerhead ribozyme region of ELVd corresponds to nucleotides 327 to 46 (plus) and nucleotides 153-203 (minus).

### SEQ ID NO: 7

```
1   GGGTGGTGTG TGCCACCCCT GATGAGACCG AAAGGTCGAA ATGGGGTTTC GCCATGGGTC
61  GGGACTTTAA ATTCGGAGGA TTCGTCTTTT AAACGTTCTT CCAAGAGTCC CTTCCCAAAA
121 CCCTTACTTT GTAAGTGTGG TTCGGCGAAT GTACCGTTTC GTCCTTTCGG ACTCATCAGG
181 GAAAGTACAC ACTTTCGAC GGTGGGTTTC TCGACACCTC TCCCCCTCCC AGGTACTATC
241 CCCTTCAAG GATGTGTTCC CTAGGAGGGT GGGTGTACCT CTTTTGGATT GCTCCGGCCT
301 TCCAGGAGAG ATAGAGGACG ACCTCTCCCC ATA
```

[0028] The full-length (247 nt) genomic sequence of ASBVd is publically available under the GenBank entry number NC\_001410.1 GI: 11496574 (SEQ ID NO: 8).

### SEQ ID NO: 8

```
1   TTTATTAGAA CAAGAAGTGA GGATATGATT AACTTTGTT TGACGAAACC AGGTCTGTTC
61  CGACTTTCG ACTCTGAGTT TCGACTTGIG AGAGAAGGAG GAGTCGTGGT GAACTTTTAT
121 TAAAAAATT AGTTCACTCG TCTTCAATCT CTTGATCACT TCGTCTCTTC AGGGAAAGAT
181 GGGAAGAACA CTGATGAGTC TCGCAAGGTT TACTCCTCTA TCTTCATTGT TTTTTTACAA
241 AATCTTG
```

[0029] The full-length (399 nt) genomic sequence of CChMVd is publically available under the GenBank entry number NC\_003540.1 GI: 20095240 (SEQ ID NO: 9).

### SEQ ID NO: 9

```
1   GGCACCTGAC GTCGGTGTCC TGATGAAGAT CCATGACAGG ATCGAAACCT CTTCCAGTTT
61  CGGCTTGTGT GGGAGTAAAG CTTTCGCTCT CTCCACAGCC TCATCAGGAA ACCCACTTCA
121 GGTCTCGACT GGAAGGTCGT TAAACTTCCC CTCTAAGCGG AGTAGAGGTA AATACCTCCG
181 TCCAACCCCG GGAGGAAAGG GGTGGGACC CGGAACAGAT CCAGTTCCGG TCCTTTGGAG
241 TCCATTTCTC TCGTTGGATA TTCTCCTCGG AGAAGGGAGA TGGGGCCAGT CCCAGTCCGT
301 TCGCTCTCGT AGTCACAGCC ACTGGGGAAC CTAGGCAGAT GGCTGGACGG AGTCTTAGTC
361 CACTCCAGAG GACCTTGGGT TTGAAACCCC CAAGAGGTC
```

[0030] The full-length (337 nt) genomic sequence of PLMVd is publically available under the GenBank entry number NC\_003636.1 GI: 20177433 (SEQ ID NO: 10). The hammerhead ribozyme region of PLMVd corresponds to nucleotides 282 to 335 (plus) and nucleotides 2-57 (minus).

SEQ ID NO: 10

1 GTCATAAGTT TCGTCGCATT TCAGCGACTC ATCAGTGGGC TTAGCCCAGA CTTATGAGAG  
 5 61 AGTAAAGACC TCTCAGCCCC TCCACCTTGG GGTGCCCTAT TCGGAGCACT GCAGTTCCCCG  
 121 ATAGAAAGGC TAAGCACCTC GCAATGAGGT AAGGTGGGAC TTTTCCTTCT GGAACCAAGC  
 181 GGTGGTTCC GAGGGGGGTG TGATCCAGGT ACCGCCGTAG AACTGGATT ACGACGTCTA  
 241 CCCGGGATTC AAACCCGTCC CCTCCAGAAG TGATTCTGGA TGAAGAGTCT GTGCTAAGCA  
 10 301 CACTGACGAG TCTCTGAGAT GAGACGAAAC TCTTCTT

[0031] A variant of a reference plant viroid sequence listed above (for example, the genomic sequence of ELVd, ASBVd, CChMVd or PLMVd, (SEQ ID NO: 7 to 10 respectively); or a truncated form thereof, as set out above) may comprise, or consist of, a nucleotide sequence having at least 70, 75, 80, 85, 90, 95, 96, 97, 98, 99% sequence identity to the sequence of the reference plant viroid scaffold.

[0032] In some embodiments, the plant viroid sequence may be a truncation of the full-length plant viroid sequence, for example the full-length plant viroid genome sequence of any of SEQ ID NO: 7 to 10. The plant viroid sequence may comprise, or consist of, the full-length plant viroid sequence with one or more deletions. The plant viroid sequence may be a synthetic plant viroid sequence that does not occur in nature.

[0033] In a more particular embodiment, the plant viroid sequence comprises the *Eggplant latent viroid* sequence (ELVd) which, in a still more particular embodiment, comprises, or consists of, a nucleotide sequence with a sequence identity of, at least, 70, 75, 80, 85, 90, 95, 96, 97, 98 or 99% with SEQ ID NO: 11. In another particular embodiment, the ELVd comprises, or consists of, a nucleotide sequence with a sequence identity of 100% with SEQ ID NO: 11.

SEQ ID NO: 11

CCCCAT**AG**GGTGGTGTGTGCCACCCCTGATGAGACCGAAAGGTCGAAATGGGGTTTCGCCATGGGTCTGGG  
 ACTTTAAATTCGGAGGATTCGTCTTTAAACGTTCTCCAAGAGTCCCTTCCCCAAACCCCTTACTTTGTA  
 30 AGTGTGGTTCGGCGAATGTACCGTTTCGTCTTTTCGGACTCATCAGGGAAAGTACACACTTTCCGACGGT  
 GGGTTCGTGCACACCTCTCCCCCTCCAGGTAATATCCCTTTTCAAGGATGTGTTCCTAGGAGGGTGGG  
 TGTACCTCTTTTGGATTGCTCCGGCCTTCCAGGAGAGATAGAGGACGACCTCTCCCCAT**AG**GGTGGTGTG  
 35 TGCCACCCCTGATGAGACCGAAAGGTCGAAATGGGG

Underlined: the hammerhead ribozyme sequences

Double underlined: insertion site of the inverted repeat sequence comprising the sense and antisense strands of the target RNA separated by an intervening sequence comprising an intron type I flanked by exons or exon fragments.

**Bold**: cut site of the ribozyme

[0034] In a particular embodiment of the nucleotide sequence of the invention, the plant viroid comprises a hammerhead ribozyme domain (motif), or a fragment thereof, at the 5' and 3' end of the sequence of the plant viroid. As used herein, the term "hammerhead ribozymes domain" is a small self-cleaving RNA that promotes strand scission by internal phosphoester transfer. The term "fragment" has been defined above. Suitable plant viroid sequences may comprise the hammerhead ribozyme region of the plant viroid, which may for example be located from nucleotides 327 to 46 (plus) and nucleotides 153-203 (minus) in ELVd and at the corresponding nucleotides in other plant viroids. Corresponding nucleotides in other plant viroid genomes may be identified using standard sequence alignment tools. In a particular embodiment of the present invention, the hammerhead ribozyme domain located at the 5' end comprises nucleotides 1 to 53 of SEQ ID NO: 11, and/or the hammerhead ribozyme domain located at the 3' end comprises nucleotides 334 to 386 of SEQ ID NO: 11.

[0035] The inverted repeat sequence, comprising the sense and antisense strands of the target RNA separated by an intervening sequence comprising an autocatalytic intron flanked by exons or exon fragments, may be located at any position within the plant viroid scaffold as long as it does not disrupt the viroid activity. Thus, the inverted repeat sequence may be located in any site between the viroid hammerhead ribozyme domains of the plant viroid sequence. In a more particular embodiment, the insertion site of the inverted repeat sequence is located between nucleotides 252 and 253 of SEQ ID NO: 11.

[0036] In a still more particular embodiment, the nucleotide sequence of the invention comprises, or consists of, the sequence SEQ ID NO: 12.



SEQ ID NO: 12

CCCCAT**AG**GGTGGTGTGTGCCACCCCTGATGAGACCGAAAGGTCGAAATGGGGTTTCGCCATGGGTTCGGG  
 5 ACTTTAAATTCGGAGGATTCGTCTTTAAACGTTCTCCAAGAGTCCCTTCCCCAAACCCCTTACTTTGTA  
AGTGTGGTTCGGCGAATGTACCGTTTCGTCTTTTCGGACTCATCAGGGAAAGTACACACTTTCCGACGGT  
GGGTTCGTGCACACCTCTCCCCCTCCAGGTAATATCCCTTGCTGACCGAAGCACCGTTGAACCCGAAG  
 10 GCTAACCCGCGAGAAGATGACCCAGATCATGTTTCGAGACCTTCAACTCGCCGGCCATGTATGTGGCCATCC  
AGATGACTCTCTAAATAGCAATATTTACCTTTGGAGGGAAAAGTTATCAGGCATGCACCTGGTAGCTAGT  
CTTTAAACCAATAGATTGCATCGGTTTAAAAGGCAAGACCGTCAAATTGCGGGAAAGGGGTCAACAGCCG  
TTCAGTACCAAGTCTCAGGGGAACTTTGAGATGGCCTTGCAAAGGGTATGGTAATAAGCTGACGGACAT  
 15 GGTCTAACCACGCAGCCAAGTCCTAAGTCAACAGATCTTCTGTTGATATGGATGCAGTTCACAGACTAA  
ATGTCGGTTCGGGAAGATGTATCTTCTCATAAGATATAGTCGGACCTCTCTTAATGGGAGCTAGCGGA  
TGAAGTGATGCAACACTGGAGCCGCTGGGAATAATTTGTATGCGAAAGTATATTGATTAGTTTTGGAGT  
 20 ACTCGTAAGGTAGCCCTGGATGGCCACATACATGGCCGGCGAGTTGAAGGTCTCGAACATGATCTGGGTC  
ATCTTCTCGCGTTAGCCTTCGGGTTCAACGGTGCTTCGGTCAGTCAAGGATGTTCCCTAGGAGGGT  
GGGTGTACCTCTTTTGGATTGCTCCGGCCTTCCAGGAGAGATAGAGGACGACCTCTCCCAT**AG**GGTGGT  
GTGTGCCACCCCTGATGAGACCGAAAGGTCGAAATGGGG

Underlined: the hammerhead ribozyme sequences showing in **bold the** cut site of the ribozyme.

Broken underlined.: cDNA from ELVd (C327 to G46 from GeneBank accession number AJ536613.1; SEQ ID NO: 7), showing in double underlined the insertion site of the inverted repeat sequence comprising the sense and antisense strands of the target RNA separated by an intervening sequence comprising an intron type I flanked by exons.

*Italic*: 100 bp of the  $\beta$ -actin gene of *Ceratitidis capitata*, in sense and antisense orientation.

Undulating underlined: 5' and 3' exon fragments of the autocatalytic intron of the rRNA 26S gene from *T. thermophila*.

Underlined with dots: Autocatalytic intron of the rRNA 26S gene from *T. thermophila*.

[0037] The nucleotide sequence of the invention as disclosed in previous paragraphs allows the skilled person in the art to produce large amounts of dsRNA when it is cloned in a host cell, such as, for example, *E. coli*, along with a tRNA ligase. Thus, co-expression of the nucleotide sequence of the invention with tRNA ligase is shown herein to lead to the high level production of dsRNA molecule in host cells. As the skilled person knows, the tRNA ligase is needed to catalyze the covalent joining of single-stranded 5'-hydroxyl termini of RNA to single-stranded 2'-3'-cyclic phosphodiester termini of RNA, and in this way to form the circular RNA that contains the dsRNA of interest. The tRNA ligase may be bound to the nucleotide sequence of the invention giving rise to a chimeric complex, or may be present in the vector used for cloning the nucleotide sequence in the host cell, or may be present in the host cell. In a particular embodiment, the nucleotide sequence of the invention further comprises a nucleotide sequence encoding a tRNA ligase.

[0038] The tRNA ligase may be a plant tRNA ligase, for example a plant chloroplast or plant nuclear tRNA ligase. Preferably, the tRNA ligase is a plant chloroplast tRNA ligase. The tRNA ligase may be from a eukaryote other than a plant. Suitable tRNA ligases may be homologues or orthologues of plant tRNA ligases and may be readily identified using standard techniques. In some embodiments, suitable tRNA ligases may bind to the plant viroid sequence to form a ribonucleoprotein complex. In some preferred embodiments, the plant viroid sequence and the tRNA ligase may be from the same plant species.

[0039] Suitable plant tRNA ligases are well-known in the art and include, without being limited to, *Solanum melongena* (eggplant) tRNA ligase. The amino acid sequence of *Solanum melongena* tRNA ligase is publicly available under the GenBank entry number AFK76482.1 GI: 388604525 (SEQ ID NO: 13) and the nucleotide coding sequence is publicly available under the GenBank entry number JX025157.1 GI: 388604524 (SEQ ID NO: 14).

EP 3 822 353 A1

SEQ ID NO: 13

1 MSVSHRVIYS FTHYKLYNLS SSLSSLPSRI FFPFQSPSFH TFSSLMPNNQ ERGGYEGKKW  
5 61 QVRPSSNRVP GSSSNVEPVS AATAEAITDR LKSVDITESG AQSSVPVTSL QFGSVGLAPQ  
121 SPVQHQKVIW KPKEYGTVSG APVVEAGKTP VEQKSALLSK LFKGNLLENF TVDNSTFSRA  
181 QVRATFYPKF ENEKSDQEIR TRMIEMVSKG LAIVEVTLKH SGSLFMYAGH EGGAYAKNSF  
241 GNIYTAVGVF VLGRMFREAW GTKASKKQAE FNEFLERNRM CISMELVTAV LGDHGQRPRD  
10 301 DYAVVTAVTE LGNGKPTFYS TPDVIAFCRE WRLPTNHVWL FSTRKSVTSF FAAYDALCEE  
361 GTATTVCEAL SEVADISVPG SKDHIKVQGE ILEGLVARIV KRESSEHMER VLRDFPPPPS  
421 EGEGLDLGPT LREICANRS EKQQIKALLQ SAGTAFCPNY LDWFGDENSG SHSRNADRSV  
481 VSKFLQSHPA DLYTGKIQEM VRLMREKRFP AAFKCHYNLH KINDVSSNNL PFKMVIHVYS  
15 541 DSGFRRYQKE MRHKPGLWPL YRGFFVDLDL FKVNEKKTAE MAGSNNQMVK NVEEDNSLAD  
601 EDANLMVKMK FLTYKLRTFL IRNGLSTLTK EGPSAYKSYL LRQMKIWNTS AAKQRELSKM  
661 LDEWAVYIRR KYGNKPLSSS TYLSEAEPLF EQYAKRSPQN HALIGSAGNF VKVEDFMAIV  
20 721 EGEDEEGDLE PAKDIAPSSP SISTRDMVAK NEGLIIFPPG IPGCAKSALC KEILNAPGGL

781 GDDRPVNSLM GDLIKGRYWQ KVADERRRKP YSIMLADKNA PNEEVWKQIE NMCLSTGASA  
841 IPVIPDSEGT ETNPFSIDAL AVFIFRVLHR VNHPGNLDKS SPNAGYVMLM FYHLYDGKSR  
25 901 QEFESIELIER FGSLVRIPVL KPERSPLPDS VRSIIEEGLS LYRLHTTKHG RLESTKGTIV  
961 QEWVKWEKQL RDILLGNADY LNSIQVPFEF AVKEVLEQLK VIARGEYAVP AEKRKLGSIK  
1021 FAAISLPVPE ILGLLNDLAK KDPKVGDFIK DKSMESSIQK AHLTLAHRKRS HGVTAVANYG  
1081 SFLHQKVPVD VAALLFSDKL AALEAEPGSV EGEKINSKNS WPHITLWSGA GVAAKDANTL  
30 1141 PQLLSQKAT RIDINPPVTI TGTLEFF

35

40

45

50

55

EP 3 822 353 A1

SEQ ID NO: 14

1 GTGACTGGCT CAAAGGCGGT AGTAGTGGCT CTTCTTCCCC GCTGGAATTT CTCGCCTATT  
 5 61 TTATATTGAT TGACAGCAGC CGAGTCTTCC TTTTGTGTAG AACTGAATGT CGGTTTCGCA  
 121 TAGGGTCATT TACTCTTTCA CTCATTACAA ACTCTATAAT CTCTCTTCTT CTTTATCATC  
 181 TTTGCCCTTCT AGAATCTTCT TCCCTTTTCA ATCTCCTTCC TTTCACACGT TCTCTTCACT  
 241 CATGCCCAAC AATCAGGAAA GGGGTGGTTA TGAAGGAAAA AAATGGCAAG TGAGGCCAAG  
 10 301 TTCCAATAGG GTACCAGGCT CGTCTTCAAA TGTGGAACCT GTATCTGCTG CAACTGCTGA  
 361 AGCCATTACC GACCGTCTAA AGTCCGTGGA TATTACTGAA AGTGGTGCAC AGTCTAGTGT  
 421 TCCAGTCACA TCTCTTCAGT TTGGCAGCGT TGGATTAGCA CCCCAGTCAC CTGTGCAACA  
 481 TCAAAAAGTA ATCTGGAAAC CCAAATCATA TGGAACAGTG TCTGGAGCCC CAGTGGTTGA  
 15 541 AGCTGGAAAA ACACCAGTTG AACAAAAAAG TGCTCTTTTA AGTAAATTAT TCAAGGGTAA  
 601 TTTATTGGAA AATTTTACTG TAGATAACTC AACATTCTCG AGAGCCCAAG TAAGGGCCAC  
 661 TTTCTACCCA AAATTTGAGA ATGAGAAATC AGATCAGGAG ATCAGGACAA GGATGATAGA  
 721 GATGGTCTCC AAAGGCTTGG CTATAGTCGA GGTCCACTT AAGCATTCTG GATCTCTTTT  
 20 781 TATGTATGCT GGGCATGAAG GTGGAGCATA TGCCAAGAAT AGCTTCGGGA ATATCTATAC  
 841 TGCCGTTGGC GTCTTTGTTC TTGGACGGAT GTTTCGTGAG GCATGGGGAA CTAAAGCAAG  
 901 CAAGAAGCAA GCAGAGTTCA ATGAGTTTCT TGAGCGCAAT CGTATGTGCA TATCAATGGA  
 25 961 GTTGGTCACG GCAGTGTGG GGGACCACGG ACAACGCCCA CGAGATGATT ATGCGGTTGT  
 1021 GACTGCAGTC ACGGAGTTGG GAAATGGAAA ACCAACTTTC TATTCAACTC CCGATGTAAT  
 1081 TGCTTTTTGC AGGGAATGGC GATTACCAAC AAATCATGTA TGGCTGTTCT CAACAAGGAA  
 1141 ATCAGTGA CTCTTCTTTG CTGCGTATGA TGCACTTTGC GAGGAAGGTA CAGCAACCAC  
 30 1201 CGTTTTCGAG GCTCTCAGCG AAGTTGCTGA TATTTCTGTA CCTGGATCAA AAGACCATAT  
 1261 AAAAGTGCAG GGTGAAATTT TGGAGGGTCT CGTGGCCCGC ATCGTAAAAC GTGAGAGCTC  
 1321 AGAGCATATG GAGCGGGTTC TGAGAGATTT TCCTCCTCCG CCATCAGAGG GTGAGGGTTT  
 1381 GGACCTGGGA CCTACGTTAC GTGAAATTTG TGCTGCAAAC AGATCAGAAA AGCAGCAAAT  
 35 1441 AAAGGCACTT CTTAGAGTG CTGGCACGGC TTTCTGCCCG AATTATTTGG ACTGGTTTGG  
 1501 AGATGAAAAC TCTGGTTCAC ATTCAAGAAA TGCTGATCGA TCTGTTGTCT CAAAGTTCTT  
 1561 ACAATCACAT CCTGCTGATC TTTATACAGG AAAAATACAG GAAATGGTTC GCTTGATGAG  
 1621 GGAAAAGCGC TTTCTGCTG CTTTCAAGTG TCATTATAAC TTACATAAAA TTAATGATGT  
 40 1681 ATCGAGTAAC AACCTGCCTT TCAAAATGGT GATCCATGTA TATAGTGATT CAGGCTTCCG  
 1741 CCGGTACCAG AAAGAGATGA GGCACAAACC AGGACTATGG CTTTGTATC GAGGCTTTTT  
 1801 TGTTGACCTG GATTTATTCA AGGTCAATGA GAAGAAAAC TCTGAAATGG CAGGAAGCAA  
 45 1861 CAATCAAATG GTAAAAATG TGAAGAGGA CAACAGTTTA GCTGATGAAG ATGCAAATCT  
 1921 GATGGTCAAG ATGAAATTTT TTAATTACAA GTTGAGA ACT TTTTGTATCC GTAATGGCTT

50

55

EP 3 822 353 A1

1981 GTCGACTCTT TTCAAAGAAG GACCTTCTGC GTATAAGTCT TATTACCTGA GGCAAATGAA  
2041 AATTTGGAAT ACTTCAGCAG CCAAGCAACG AGAACTCAGC AAGATGCTTG ATGAATGGGC  
5 2101 AGTATATATA CGCAGAAAAT ATGGGAACAA ACCATTGTCA TCATCCACAT ACCTAAGTGA  
2161 AGCTGAGCCT TTCCTTGAAC AATATGCAAA GCGTAGTCCA CAAAATCATG CTTTGATAGG  
2221 ATCTGCTGGA AATTTTGTCA AAGTTGAAGA TTTCATGGCT ATTGTTGAAG GAGAAGATGA  
2281 AGAGGGTGAT CTCGAGCCTG CGAAAGATAT TGCTCCTTCA AGCCCTAGTA TTTCCACCAG  
10 2341 AGACATGGTG GCAAAGAATG AGGGTCTCAT TATTTTCTTT CCAGGAATAC CAGGTTGTGC  
2401 TAAATCTGCA CTTTGTAAAG AAATACTGAA TGCTCCAGGA GGGCTTGAG ATGATCGACC  
2461 AGTTAACAGT TTAATGGGTG ATCTTATTAA AGGTAGATAT TGGCAAAAAG TTGCTGATGA  
2521 ACGTCGAAGA AAACCTTACT CGATCATGCT TGCTGACAAG AATGCACCAA ATGAGGAAGT  
15 2581 ATGGAAACAA ATTGAGAACA TGTGCCTAAG CACCGGAGCA TCTGCTATTC CAGTTATAACC  
2641 TGATTAGAA GGAAGTAAA CTAATCCATT CTCTATTGAT GCACTTGCGG TTTTATATT  
2701 CCGAGTACTT CACCGTGTCA ATCATCCGGG AAATCTTGAC AAGTCATCTC CAAATGCTGG  
2761 ATATGTGATG CTTATGTTTT ATCACCTTTA TGATGGAAAG AGCCGTCAGG AGTTCGAGAG  
20 2821 TGAGCTTATT GAACGTTTTG GATCGCTTGT CAGAATTCCT GTACTGAAAC CTGAGAGGTC  
2881 TCCTCTCCG GATTCTGTGA GGTCTATTAT CGAGGAGGGA CTCAGTCTGT ACAGACTTCA  
2941 TACAACGAAA CATGGAAGAT TGGAGTCTAC AAAAGGGACA TATGTACAAG AGTGGGTTAA  
3001 ATGGGAGAAG CAATTGAGAG ATATTCTACT TGAAAATGCA GACTATCTCA ATTCAATACA  
25 3061 GGTTCCATT GAATTTGCCG TTAAAGAAGT CCTTGAACAA CTGAAAAGTTA TTGCGAGGGG  
3121 CGAATATGCA GTGCCTGCTG AGAAGAGGAA GCTAGGATCC ATTGTATTTC CCGCTATCAG  
3181 CCTGCCAGTT CCAGAAATTC TAGGTCTTCT AAATGATCTA GCAAAGAAAAG ATCCAAAGGT  
30 3241 TGGCGATTTT ATTAAGGACA AGAGCATGGA GAGCAGCATT CAGAAGGCCC ATCTTACCCT  
3301 GGCTCACAAG AGAAGTCACG GTGTCACTGC AGTTGCCAAT TACGTTTCTT TTCTTCATCA  
3361 AAAGGTGCCA GTAGACGTGG CTGCTTTGTT GTTCTCCGAT AAATTGGCTG CACTAGAAGC  
3421 TGAGCCTGGC TCTGTTGAAG GTGAAAAGAT CAATTCTAAA AACTCATGGC CCCATATCAC  
35 3481 ATTATGGTCT GGTGCAGGAG TTGCCGCAA AGATGCCAAT AACTACCAC AGTTACTTTT  
3541 CCAAGGGAAG GCTACCCGCA TTGATATAAA TCCACCGGTC ACTATAACTG GCACTCTCGA  
3601 ATTCTTTTGA ACTCTACTTG TAATTCCTCT TGTGTGGATT TACATAATGT GAACCCTACA  
3661 TGGGTTCTGA GTGAATCATG TAATGATTCT CCTGAGTTT AGGAACAATT TTGTAGCAGT  
40 3721 CTTGCAGTTG CAGGATAGAA CTAGATCTAG AGAAAATGTA AAGATGTTA AGGCAATAAA  
3781 GGTCTTGAGC TGTGGGCATA TTGAAAATCC CCAGTAGCAT TTTCTTGAG CAAAAA  
3841 AAAAAA

45 **[0040]** Thus, in a particular embodiment of the present invention, the tRNA ligase is the eggplant tRNA ligase. In a more particular embodiment, the tRNA ligase from eggplant comprises, or consists of, a nucleotide sequence having a sequence identity of at least 70, 75, 80, 85, 90, 96, 97, 98 or 99% to SEQ ID NO: 14. In a more particular embodiment, the tRNA ligase from eggplant comprises, or consists of, a nucleotide sequence having a sequence identity of 100% to SEQ ID NO: 14.

50 **[0041]** Other suitable tRNA ligases for use in the present invention may be an orthologue from a plant species other than *Solanum melongena* (eggplant), for example *Arabidopsis thaliana* (GenBank accession number NM\_100665.3 [SEQ ID NO: 18]), *Solanum lycopersicum* (XM\_004251213.4 [SEQ ID NO: 19]), *Hevea brasiliensis* (XM\_021784835.1 [SEQ ID NO: 20]), *Prunus avium* (XM\_021975205.1 [SEQ ID NO: 21]), *Malus domestica* (XM\_029088496.1 [SEQ ID NO: 22]), *Triticum* spp. or *Oryza sativa*. Suitable orthologues may be identified using standard sequence analysis techniques.

55 **[0042]** In a particular embodiment, the plant viroid and the tRNA ligase may be from the same plant species. For example, eggplant tRNA ligase may be expressed with an ELVd sequence, or avocado tRNA ligase may be expressed with an *Avocado sunblotch viroid* (ASBVd) sequence. The dsRNA molecule and the tRNA ligase may be encoded by

the same or different nucleic acids in the host cell. The nucleic acids may be heterologous or exogenous to the host cell and may be introduced into the host cell by transformation or transfection as described below.

[0043] As a consequence of expressing the nucleotide sequence of the invention on a host cell in combination with a tRNA ligase, large quantities of hybrid circular molecules are produced efficiently comprising the plant viroid on which the dsRNA hairpin is presented. For example, more than 10 mg, more than 20 mg, more than 30 mg, more than 50 mg or more than 100 mg of dsRNA may be produced per litre of the host cell culture. The data herein, for example, shows the production of 50 mg per litre of *E. coli* culture.

[0044] However, it is desirable to avoid the presence of the plant viroid in the dsRNA molecule in order to improve the commercial potential of this technology. In order to remove the residue derived from the plant viroid, the inventors added a new element to the nucleotide sequence of the invention comprising of a permuted intron-exon sequence. When an intron of group I is divided into two parts, and the first is fused following the second, the self-cleavage reaction continues on the border between exon 1 and the 5' end of the intron and continues with the attack on 3' end from exon 1 to the border between the 3' end of the intron and exon 2, autocatalytically generating a circular RNA molecule comprised by the two exons.

[0045] Therefore, in another particular embodiment, the nucleotide sequence of the invention may further comprise permuted autocatalytic intron-exon sequences between the plant viroid sequence and the cDNAs sequences. Examples of permuted autocatalytic intron-exon sequences include, without being limited to, tRNA intron from *Anabaena* sp. strain PCC7120 or the intron from thymidylate synthase (td) gene of phage T4. In a particular embodiment, the permuted autocatalytic intron-exon sequences are the autocatalytic intron of the 26S rRNA gene of *T. thermophila*.

[0046] In a more particular embodiment, the permuted autocatalytic intron-exon sequence at 5' of the cDNA sequence of the sense strand of the target dsRNA comprises, or consists of, a nucleotide sequence with a sequence identity of at least 70, 75, 80, 85, 90, 95, 96, 97, 98, 99% with SEQ ID NO: 15. In a still more particular embodiment, the permuted autocatalytic intron-exon sequence at 5' of the cDNA sequence of the sense strand of the target dsRNA comprises, or consists of, a nucleotide sequence with a sequence identity of 100% with SEQ ID NO: 15.

SEQ ID NO: 15 (The exon is shown underlined)

CTTCTGTTGATATGGATGCAGTTCACAGACTAAATGTCGGTCGGGAAGATGTATTCTTCTCATAAGATA  
TAGTCGGACCTCTCCTTAATGGGAGCTAGCGGATGAAGTGATGCAACACTGGAGCCGCTGGGAAC TAATT  
TGTATCGCAAAGTATATTGATTAGTTTTGGAGTACTCGTAAGGTAGCC

[0047] In another particular embodiment, the permuted autocatalytic intron-exon sequence at 3' of the cDNA sequence of the antisense strand of the target dsRNA comprises, or consists of, a nucleotide sequence with a sequence identity of at least 70, 75, 80, 85, 90, 95, 96, 97, 98, 99% with SEQ ID NO: 16. In a still more particular embodiment, the permuted autocatalytic intron-exon sequence at 3' of the cDNA sequence of the sense strand of the target dsRNA comprises, or consists of, a nucleotide sequence with a sequence identity of 100% with SEQ ID NO: 16.

SEQ ID NO: 16 (The exon is shown underlined)

ATGACTCTCTAAATAGCAATATTTACCTTTGGAGGGAAAAGTTATCAGGCATGCACCTGGTAGCTAGTCT  
TTAAACCAATAGATTGCATCGGTTTTAAAAGGCAAGACCGTCAAATTGCGGGAAAGGGTCAACAGCCGTT  
CAGTACCAAGTCTCAGGGGAAACTTTGAGATGGCCTTGCAAAGGGTATGGTAATAAGCTGACGGACATGG  
TCCTAACCACGCAGCCAAGTCCTAAGTCAACAGAT

[0048] In a still more particular embodiment, the nucleotide sequence of the invention comprises the sequence SEQ ID NO: 17.

SEQ ID NO: 17

CCCCATAGGGTGGTGTGTGCCACCCCTGATGAGACCGAAAGGTCGAAATGGGGTTTCGCCATGGGTCCGG  
ACTTTAAATTCGGAGGATTCGTCTTTAAACGTTCTCTCAAGAGTCCCTTCCCCAAACCCTTACTTTGTA  
AGTGTGGTTCGGCGAATGTACCGTTTCGTCTTTTCGGACTCATCAGGGAAAGTACACACTTTCGGACGGT  
GGGTTCGTTCGACACCTCTCCCCCTCCCAGGTACTATCCCCTTCTTCTGTTGATATGGATGCAGTTCACAG  
ACTAAATGTCCGGTCCGGGAAGATGTATTCTTCTCATAAGATATAGTCCGACCTCTCCTTAATGGGAGCTA  
GCGGATGAAGTGATGCAACACTGGAGCCGCTGGGAACTAATTTGTATGCGAAAGTATATTGATTAGTTTT  
GGAGTACTCGTAAGGTAGCC *GCTGACCGAAGCACCGTTGAACCCGAAGGCTAACC* *CGGAGAAGATGACCC*  
AGATCATGTTTCGAGACCTCAACTCGCCGGCCATGTATGTGGCCATCCAGATGACTCTCTAAATAGCAAT  
ATTTACCTTTGGAGGGAAAAGTTATCAGGCATGCACCTGGTAGCTAGTCTTTAAACCAATAGATTGCATC  
GGTTTAAAAGGCAAGACCGTCAAATTCGCGGAAAAGGGGTCAACAGCCGTTCAGTACCAAGTCTCAGGGGA  
AACTTTGAGATGGCCTTGCAAAGGGTATGGTAATAAGCTGACGGACATGGTCCTAACCACGCAGCCAAGT  
CCTAAGTCAACAGATCTTCTGTTGATATGGATGCAGTTCACAGACTAAATGTCGGTCCGGGGAAGATGTAT  
TCTTCTCATAAGATATAGTCCGACCTCTCCTTAATGGGAGCTAGCCGATGAAGTGATGCAACACTGGAGC  
CGCTGGGAACTAATTTGTATGCGAAAGTATATTGATTAGTTTTGGAGTACTCGTAAGGTAGCCCTGGATG  
GCCACATACATGGCCGGCGAGTTGAAGGTCGAAACATGATCTGGGTCATCTTCTCGCGGTTAGCCTTCG  
GGTTCAACGGTGCTTCGGTCAGCATGACTCTCTAAATAGCAATATTTACCTTTGGAGGGAAAAGTTATCA  
GGCATGCACCTGGTAGCTAGTCTTTAAACCAATAGATTGCATCGGTTTTAAAGGCAAGACCGTCAAATTG  
CGGGAAAGGGGTCAACAGCCGTTCAAGTACCAAGTCTCAGGGGAAACTTTGAGATGGCCTTGCAAAGGGTA  
TGGTAATAAGCTGACGGACATGGTCTTAACCACGCAGCCAAGTCTTAAGTCAACAGAT TC AAGGATGTGT  
TCCCTAGGAGGGTGGGTGTACCTCTTTTGGATTGCTCCGGCCTTCCAGGAGAGATAGAGGACGACCTCTC  
CCCATAGGGTGGTGTGTGCCACCCCTGATGAGACCGAAAGGTCGAAATGGGG

Underlined: the hammerhead ribozyme sequences showing in **bold** the cut site of the ribozyme.

Broken underlined: cDNA from ELVd (C327 to G46 from GeneBank accession number AJ536613.1), showing in double underlined the insertion site of the inverted repeat sequence comprising the sense and antisense strands of the target RNA separated by an intervening sequence comprising an intron type I flanked by exon fragments.

*Italic*: 100 bp of the  $\beta$ -actin gene of *Ceratitis capitata*, in sense and antisense orientation.

Undulating underlined: 5' and 3' exons of the autocatalytic intron of the rRNA 26S gene from *T. thermophila*.

Underlined with dots: Autocatalytic intron of the rRNA 26S gene from *T. thermophila*. **Bold**: Permuted autocatalytic intron-exon sequence at 5' and 3' of the cDNA sequence of the sense and antisense strands, respectively, of the target dsRNA, wherein the exons are showing underlined.

**[0049]** In summary, the nucleotide sequence of the invention allows the production of large amounts of dsRNA. To do this, the cDNAs of the two strands of the target dsRNA are inserted into the cDNA of a plant viroid and separated by the cDNA of an autocatalytic intron. The co-expression of the resulting RNA together with the tRNA ligase produces a large accumulation of a hybrid molecule in which the dsRNA is fused to the viroid. The autocatalytic intron disappears during the processing of the precursor. In addition, if a permuted autocatalytic intron is added to this construction, the dsRNA of interest can be separated in a circular form from the viroid molecule. Figures 1 and 2 illustrate the process disclosed in this paragraph.

**[0050]** As the skilled person in the art understands, the nucleotide sequence of the invention may further comprise regulatory sequences, such as enhances and promoters, origins of replication, ribosomal terminators, resistance genes, etc., which allow the expression of the nucleotide sequence in a host cell.

**[0051]** In another aspect, the present invention relates to a vector, hereinafter "vector of the invention", comprising the nucleotide sequence of the invention.

**[0052]** The term "vector" in accordance with the invention means a plasmid, cosmid, virus, bacteriophage or another vector used conventionally in genetic engineering which carries the nucleotide sequence of the invention. Accordingly, the nucleotide sequence of the invention may be inserted into several commercially available vectors. Nonlimiting examples include prokaryotic plasmid vectors, such as of the pUC-series, pBluescript (Stratagene), the pET-series of

expression vectors (Novagen) or pCROTOPO (Invitrogen) and vectors compatible with an expression in mammalian cells like pREP (Invitrogen), pcDNA3 (Invitrogen), pCEP4 (Invitrogen), pMC1neo (Stratagene), pXT1 (Stratagene), pSG5 (Stratagene), EBO-pSV2neo, pBPV-1, pdBPVMMTneo, pRSVgpt, pRSVneo, pSV2-dhfr, pIZD35, pLXIN, pSIR (Clontech), pIRES-EGFP (Clontech), pEAK-10 (Edge Biosystems) pTriEx-Hygro (Novagen) and pCINeo (Promega). Samples for plasmid vectors suitable for *Pichia pastoris* include, without being limited to, the plasmids pAO815, pPIC9K and pPIC3.5K (all Invitrogen).

[0053] The nucleotide sequence of the invention may be inserted into vectors such that a (further) translational fusion with another nucleotide sequence is generated. The other nucleotide sequence may encode a protein which may, for example, increase the solubility and/or facilitate the purification of the dsRNA encoded by nucleotide sequence of the invention. To generate said fusion of nucleotides, overlap extension PCR can be applied. The products arising therefrom are termed fusion proteins and will be described further below.

[0054] The vector of the invention, further to the nucleotide sequence of the invention, may comprise a nucleotide sequence encoding the tRNA ligase. Alternatively, the present invention also encompasses a pair of vectors, the first vector comprising the nucleotide sequence of the invention, and the second vector comprising the nucleotide sequence encoding the tRNA ligase as described in previous paragraphs.

[0055] Plasmids which can be used to introduce the nucleotide sequence of the invention into a host cell are pUC18/19 (Roche Biochemicals), pBluescript II, pKK-177-3H (Roche Biochemicals), pBTac2 (Roche Biochemicals), pKK223-3 (Amersham Pharmacia Biotech), pKK-233-3 (Stratagene) or pET (Novagen).

Generally, vectors can contain one or more origins of replication (ori) and inheritance systems for cloning or expression, one or more markers for selection in the host, e.g., antibiotic resistance, and one or more expression cassettes. Suitable origins of replication include, without limiting to, the pUC, p15A, ColE1, the viral SV40 or the M13 origins of replication.

[0056] The nucleotide sequences inserted in the vector can, for example, be synthesized by standard methods. Ligation of the nucleotide sequence to transcriptional regulatory elements and/or to other nucleotide sequences can be carried out using established methods well-known by the skilled person in the art. Transcriptional regulatory elements (parts of an expression cassette) ensuring expression in prokaryotes or eukaryotic cells are well known to those skilled in the art. These elements comprise regulatory sequences ensuring the initiation of the transcription (e.g., translation initiation codon, transcriptional termination sequences, promoters, enhancers, and/or insulators) and internal ribosomal entry sites (IRES). Additional regulatory elements may include transcriptional, and/or naturally-associated or heterologous promoter regions. The regulatory elements may be heterologous regulatory elements. Preferably, the nucleotide sequence of the invention is operably linked to such expression control sequences allowing expression in prokaryotes or eukaryotic cells. The co-transfection with a selectable marker such as kanamycin, chloramphenicol or ampicillin resistance genes for culturing in *E. coli* or other bacteria allows the identification and isolation of the transfected cells. Using such markers, the cells are grown in selective medium and those with the highest resistance are the transfected cells.

[0057] In another aspect, the invention relates to a host cell, hereinafter "cell of the invention", comprising, e.g., as a result of transformation, transduction, microinjection or transfection, the nucleotide sequence or the vector of the invention.

[0058] Techniques for the introduction of nucleic acid into cells are well established in the art and any suitable technique may be employed in accordance with the particular circumstances. For eukaryotic cells, suitable techniques may include calcium phosphate transfection, DEAE-Dextran, electroporation, liposome-mediated transfection and transduction using retrovirus or other virus, e.g. adenovirus, AAV, lentivirus or vaccinia. For bacterial cells, suitable techniques may include calcium chloride transformation, electroporation and transfection using bacteriophage. Marker genes such as antibiotic resistance or sensitivity genes may be used in identifying clones containing nucleic acid of interest, as is well-known in the art. The introduced nucleic acid(s) may be on an extra-chromosomal vector within the cell or the nucleic acid may be integrated into the genome of the host cell. Integration may be promoted by inclusion of sequences within the nucleic acid or vector which promote recombination with the genome, in accordance with standard techniques.

[0059] A variety of host-expression systems may be used to express the nucleotide sequence or the vector of the invention. The host cell of the invention may be produced by introducing the nucleotide sequence or vector(s) of the invention into the host cell.

[0060] In a particular embodiment, the host cell of the invention comprises

- (a) the vector of the invention, or
- (c) a pair of vectors, the first vector comprising the nucleotide sequence of the invention, and the second vector comprising the nucleotide sequence encoding the tRNA ligase as described in previous paragraphs.

[0061] The host from which the host cell is derived may be any eukaryotic or prokaryote cell. A suitable eukaryotic host cell may be a vertebrate cell, an amphibian cell, a fish cell, an insect cell, a fungal/yeast cell, a nematode cell or a plant cell. The insect cell may be a *Spodoptera frugiperda* cell, a *Drosophila* S2 cell or a *Spodoptera* Sf9 cell, the fungal/yeast cell may be a *Saccharomyces cerevisiae* cell, *Pichia pastoris* cell or an *Aspergillus* cell.

[0062] It is preferred that the vertebrate cell is a mammalian cell such as a human cell, CHO, COS, 293 or Bowes

melanoma cell. The plant cell is preferably selected independently from a cell of Anacardium, Anona, Arachis, Artocarpus, Asparagus, Atropa, Avena, Brassica, Carica, Citrus, Citrullus, Capsicum, Carthamus, Cocos, Coffea, Cucumis, Cucurbita, Daucus, Elaeis, Fragaria, Glycine, Gossypium, Helianthus, Heterocallis, Hordeum, Hyoseyamus, Lactuca, Linum, Lolium, Lupinus, Lycopersicon, Malus, Manihot, Majorana, Medicago, Nicotiana, Olea, Oryza, Panieum, Pannasetum, Passiflora, Persea, Phaseolus, Pistachia, Pisum, Pyrus, Prunus, Psidium, Raphanus, Ricinus, Secale, Senecio, Sinapis, Solanum, Sorghum, Theobromus, Trigonella, Triticum, Vicia, Vitis, Vigna and Zea. The cell may be a part of a cell line. The cell from plant may, e.g., be derived from root, leave, bark, needle, bole or caulis.

**[0063]** Suitable prokaryotes (bacteria) useful as hosts for the invention are those generally used for cloning and/or expression include, without being limited to, *E. coli* (e.g., *E. coli* strains BL21, HB101, DH5a, XL1 Blue, Y1090, JM101, and HT115), *Salmonella typhimurium*, *Serratia marcescens*, *Burkholderia glumae*, *Pseudomonas putida*, *Pseudomonas fluorescens*, *Pseudomonas stutzeri*, *Streptomyces lividans*, *Lactococcus lactis*, *Mycobacterium smegmatis*, *Streptomyces*, *Bacillus subtilis*. Appropriate culture mediums and conditions for the above described host cells are known in the art.

**[0064]** Preferred examples for host cell to be genetically engineered with the nucleotide sequence or the vector(s) of the invention is a cell of yeast, *E. coli* and/or a species of the genus *Bacillus* (e.g., *B. subtilis*).

**[0065]** In some embodiments, the host cells may express an exogenous tRNA ligase. For example, the host cells may have been previously transformed with a nucleotide sequence encoding the tRNA ligase. A method as described herein may comprise introducing nucleic acid encoding the nucleotide sequence of the invention into host cells that express an exogenous tRNA ligase. The introduction may be followed by expression of the nucleotide sequence(s) to produce the dsRNA molecule and/or tRNA ligase. In some embodiments, host cells (which may include cells actually transformed although more likely the cells will be descendants of the transformed cells) may be cultured *in vitro* under conditions for expression of the nucleic acid, so that the encoded dsRNA molecule and/or tRNA ligase are produced. When an inducible promoter is used, expression may require the activation of the inducible promoter. The dsRNA molecules may then be recovered from the host cells or the surrounding medium.

**[0066]** In another aspect, the present invention relates to a plant, hereinafter "plant of the invention", comprising the isolated nucleotide sequence of the invention, or the vector of the invention, or the host cell of the invention.

**[0067]** Another aspect of the invention provides a recombinant host cell that expresses the dsRNA molecule and tRNA ligase, as described herein. The host cell may comprise a heterologous nucleotide sequence encoding the tRNA ligase and a heterologous nucleotide sequence encoding the dsRNA molecule (the nucleotide sequence of the invention). The nucleotide sequences may be contained in the same or separate expression vectors as described above. Another aspect of the invention provides a recombinant host cell that expresses an exogenous tRNA ligase and is suitable for transformation with a nucleotide sequence of the invention as disclosed herein. Host cells, chimeric RNA molecules and tRNA ligases are described in more detail above.

**[0068]** The expressed dsRNA, specific of a target RNA, may be isolated and/or purified after production from the host cell and/or culture medium. This may be achieved using any convenient method known in the art. Techniques for the purification and/or the isolation of RNA are well known in the art and include, for example HPLC, FPLC or affinity chromatography.

**[0069]** Following isolation and/or purification, the dsRNA molecule may subsequently be used as desired, e.g. in the formulation of a composition which may include one or more additional components, such as a pharmaceutical or a phytosanitary composition which includes one or more acceptable excipients, vehicles or carriers. dsRNAs produced as described herein may be investigated further, for example the pharmacological properties and/or activity may be determined. Methods and means of RNA analysis are well-known in the art.

**[0070]** In some embodiments, the dsRNA may be modified or adapted after isolated from the host cell. For example, the 2' -OH group of one or more nucleotides in the dsRNA may be chemically modified, for example by addition of a substituent group, such as methyl (e.g. 2'-O-methyl), halogen (e.g. 2'-Fluoro) or amine (e.g. 2'-NH<sub>3</sub>), or reduction to 2'-H (e.g. 2' deoxy).

**[0071]** As explained throughout the present description, the nucleotide sequence of the invention allows the production of high amounts of a target dsRNA when it is expressed in a host cell. Thus, in another aspect, the present invention relates to the use of the isolated nucleotide sequence of the invention, or of the vector of the invention, or of the host cell of the invention, or the plant of the invention, for producing a dsRNA specific for a target gene. The terms used for defining this aspect, as well as their particular embodiments, have been explained in previous paragraphs and can be applied to the present inventive aspect. The isolated nucleotide sequence of the invention, or of the vector of the invention, or of the host cell of the invention, or the plant of the invention, may comprise a nucleotide sequence encoding a tRNA ligase. However, as explained in previous paragraphs, the tRNA ligase (or the nucleotide sequence encoding the tRNA ligase) may be present in another vector or in the own genome of the host cell. Thus, in a particular embodiment, the invention relates to the use of the isolated nucleotide sequence of the invention, or of the vector of the invention, or of the host cell of the invention, or the plant of the invention, in combination with a tRNA ligase (or the nucleotide sequence encoding the tRNA ligase), for producing a dsRNA specific for a target gene.

**[0072]** Analogously, in another aspect, the present invention relates to a method for producing dsRNA specific for a



target gene, hereinafter "method of the invention", comprising co-expressing in a host cell

- a) the nucleotide sequence of the invention and
- b) a nucleotide sequence encoding a tRNA ligase,

5

wherein the nucleotide sequences of a) and b) are in the same or different nucleic acid molecule.

[0073] In a particular embodiment of the method of the invention, the host cell is *E. coli*.

[0074] In another particular embodiment, the tRNA ligase is the tRNA ligase from eggplant.

10

[0075] The terms "dsRNA", "host cell" and "tRNA ligase" have been defined and explained in previous aspects of the present invention. The co-expression of elements (a) and (b) cited above is carried out by culturing the host cell comprising both elements in suitable conditions. Suitable conditions for culturing a prokaryotic or eukaryotic host cell are well known to the person skilled in the art. Suitable conditions for culturing *E. coli*, are provided in the examples of the present invention. In general, suitable conditions for culturing bacteria are growing them under aeration in Luria Bertani (LB) medium. To increase the yield and the solubility of the expression product, the medium can be buffered or supplemented with suitable additives known to enhance or facilitate both. *E. coli* can be cultured from 4 to about 37 °C, the exact temperature or sequence of temperatures depends on the molecule to be overexpressed. In general, *Aspergillus* sp. may be grown on Sabouraud dextrose agar, or potato dextrose agar at about 10°C to about 40°C, and preferably at about 25°C. Suitable conditions for yeast cultures are known. The skilled person is also aware of all these conditions and may further adapt these conditions to the needs of a particular host species and the requirements of the polypeptide expressed. In case an inducible promoter controls the nucleotide sequence of the invention in the vector present in the host cell, expression of the dsRNA can be induced by addition of an appropriate inducing agent. Suitable expression protocols and strategies are known to the skilled person. Depending on the cell type and its specific requirements, mammalian cell culture can e.g. be carried out in RPMI or DMEM medium containing 10% (v/v) FCS, 2mM L-glutamine and 100 U/ml penicillin/streptomycin. The cells can be kept at 37 °C in a 5% CO<sub>2</sub>, water saturated atmosphere. Suitable expression protocols for eukaryotic cells are well known to the skilled person.

15

20

25

[0076] After culturing the host cell, the dsRNA produced may be isolated and/or purified. Methods of isolation are well-known in the art and comprise, without limitation, steps such as ion exchange chromatography, gel filtration chromatography (size exclusion chromatography), affinity chromatography, high pressure liquid chromatography (HPLC), reversed phase HPLC, or disc gel electrophoresis.

30

[0077] In another aspect, the present invention relates to a phytosanitary composition comprising a dsRNA obtained by the method of the invention.

[0078] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skilled in the art to which this invention belongs. Methods and materials similar or equivalent to those described herein can be used in the practice of the present invention. Throughout the description and claims the word "comprise" and its variations are not intended to exclude other technical features, additives, components, or steps. Additional objects, advantages and features of the invention will become apparent to those skilled in the art upon examination of the description or may be learned by practice of the invention. The following examples, drawings and sequence listing are provided by way of illustration and are not intended to be limiting of the present invention.

35

40

## BRIEF DESCRIPTION OF THE DRAWINGS

### [0079]

45

Figure 1 is a representation of the production of dsRNA in *E. coli* with the system derived from ELVd, using the autocatalytic intron of *T. termophila* to separate the inverted repeats.

Figure 2 is a scheme of the intron-exon permutation strategy (IEP) and of the self-cleavage reaction of the introns of group I, in which the ligation of the exons and the circularization of the dsRNA is produced.

50

Figure 3 is a representation of a circular double-stranded miniRNA with an extensive perfectly double-stranded region homologous to an endogenous gene of the pest or pathogen of interest, closed in both sides by the fragments of the exons.

55

Figure 4 shows the sequence and predicted structure of minimum free energy of Eggplant latent viroid (ELVd; GenBank accession number AJ536613.1 (SEQ ID NO: 7)). The hammerhead ribozyme domain and self-cleavage site are indicated on grey background and by an arrowhead, respectively. The position U245-U246 where recombinant RNAs were inserted is indicated by an arrow.

Figure 5 is a 1% agarose gel stained with ethidium bromide in which the electrophoretic migration of the empty plasmid pLELVd (lane 2) is compared to those of five independent clones of pLELVd- $\beta$ -actin (lanes 3 to 7). Lane 1 corresponds to a ladder of DNA markers whose sizes in base pairs (bp) are indicated on the left.

Figure 6 is a 5% polyacrylamide gel, containing 8 M urea, stained with ethidium bromide in which total *Escherichia coli* RNA preparations were analyzed. Lanes 1 to 4 correspond to *E. coli* clones transformed with empty pLELVd, lane 5 corresponds to a ladder of RNA markers whose size in nucleotides (nt) is indicated on the left, lanes 6 to 7 correspond to *E. coli* clones transformed with pLELVd- $\beta$ -actin. White, gray and black arrows point to ELVd RNA, ELVd-dsRNA- $\beta$ -actin, and the spliced intron RNA, respectively.

Figure 7 are two 5% polyacrylamide gels, the second containing 8 M urea, stained with ethidium bromide. A total RNA preparation from an *Escherichia coli* clone transformed with pLELVd- $\beta$ -actin was separated in the first gel (left) under non-denaturing conditions. After staining the gel, the band pointed by the arrow was cut from the gel and applied onto a second gel (right) that was run under denaturing conditions. The arrow on the right points to the spliced circular dsRNA, which is also schematized in the figure. Both gels contain RNA markers on the left.

## EXAMPLES

**[0080]** Due to the instability of DNA inverted repeats in *E. coli* cells when plasmids to produce dsRNA to induce RNAi in pests and pathogens were built, bacterial colonies bearing the expected constructs were not obtained. In order to alleviate the well-known problems with DNA inverted repeats in *E. coli*, the cDNAs corresponding to the two strands of the target dsRNAs were separated with a 433-bp cDNA (SEQ ID NO: 6) corresponding to *T. thermophila* 26S rRNA Group-I intron (GenBank accession number V01416.1 [SEQ ID NO: 23]) plus 10 bp of both native flanking exons. In this way, while the intron cDNA will separate the inverted repeat at the DNA level, the intron would self-splice from the transcript allowing formation of the dsRNA.

**[0081]** For this purpose, we amplified by PCR using the Phusion high-fidelity DNA polymerase (Thermo Scientific) two different DNAs corresponding to a 100-bp fragment of *C. capitata* beta-actin mRNA (GenBank accession number EU665679.1 [SEQ ID NO: 24]) in both orientations (SEQ ID NO: 1 and SEQ ID NO: 2). We also amplified by PCR a DNA corresponding to *T. thermophila* 26S rRNA group-I intron with fragments of flanking exons (SEQ ID NO: 6), from a plasmid obtained by gene synthesis. PCR products were separated by electrophoresis in a 1% agarose gel and the gel stained with ethidium bromide. DNAs of the expected size were eluted from the gel using silica-gel spin columns. The three DNAs were assembled between the positions T245 and T246 of the ELVd-AJ536613.1 cDNA with duplicated hammerhead ribozyme domains (SEQ ID NO: 11) in plasmid pLELVd (Daròs, J.A., Aragonés, V., and Cordero, T. (2018) *Sci. Rep.*, 8, 1904), by means of the Gibson assembly reaction using the NEBuilder HiFi DNA assembly master mix (New England Biolabs). Competent *E. coli* cells were finally electroporated with the products of the Gibson assembly reactions and transformed bacteria selected in Luria-Bertani (LB) plates containing 50  $\mu$ g/ml ampicillin. We grew five independent *E. coli* clones in liquid LB cultures and purified the corresponding plasmids. An analysis by electrophoresis in a 1% agarose gel that was stained with ethidium bromide showed all of them being the same size and exhibiting a migration delay consistent with the inserted DNAs (Figure 5, compare lane 2 with lanes 3 to 7). The expected sequence SEQ ID NO: 12) was experimentally confirmed in one of these plasmids. This plasmid was used to co-electroporate the RNase III-deficient HT115(DE3) strain of *E. coli* that lacks RNase III (Timmons, L., Court, D.L., and Fire, A. (2001) *Gene*, 263, 103-112), along with plasmid pLtRnlSm (Daròs *et al.*, 2018, cited *ad supra*) from which eggplant tRNA ligase (SEQ ID NO: 13 and SEQ ID NO: 14) is constitutively expressed. As controls, pLtRnlSm was also co-electroporated with the plasmid to express the empty ELVd (pLELVd) (Daròs *et al.*, 2018, cited *ad supra*). Nine independent colonies (four for the empty ELVd control) were picked from LB plates containing 50  $\mu$ g/ml ampicillin and 34  $\mu$ g/ml chloramphenicol and grown for 24 hours in liquid Terrific Broth (TB) medium. Total RNA was extracted from the cells using a 1:1 mix of phenol and chloroform (pH 8.0) (Daròs *et al.*, 2018), and analyzed by polyacrylamide gel electrophoresis (PAGE) in denaturing conditions (8 M urea) (Daròs *et al.*, 2018, cited *ad supra*). Two prominent bands above the 600-nt and slightly below the 400-nt RNA markers were observed in the RNA extracts from bacteria transformed with the plasmid to express the dsRNA (Figure 6, lanes 6 to 14, gray and black arrows respectively). Remarkably, these bands exhibited a fluorescence intensity comparable to those from the endogenous *E. coli* rRNAs (Figure 6, upper part of the gel), indicating a large accumulation in *E. coli*. RNA extracts from the empty ELVd controls exhibited a single prominent band above the 400-nt marker that, according to our previous analysis (Fadda, Z., Daròs, J.A., Fagoaga, C., Flores, R., and Duran-Vila, N. (2003) *J. Virol.*, 77), corresponds to the 333-nt-long circular ELVd RNA (Figure 6, lanes 1 to 4, white arrow). Circular RNAs migrate slower than the linear counterparts of the same size in denaturing conditions. The identity of the different RNA species indicated in the Figure 2 was confirmed by Northern-blot hybridization analysis using <sup>32</sup>P-labelled RNA probes complementary to the ELVd and the *T. thermophila* intron RNAs.

**[0082]** Taken together, these results indicate that, while the *T. thermophila* cDNA serves to stabilize the inverted

repeats in the *E. coli* expression plasmids, the intron very efficiently self-splices from the primary transcript in bacterial cells facilitating the accumulation of large amounts of a circular RNA product that consists of an ELVd scaffold from which the dsRNA bulges. This is schematized in Figure 1. The dsRNAs produced in *E. coli* with this method showed the same insecticide activities as the counterparts of the same sequence obtained using the standard methodologies, such as two-strand transcription from two opposite promoters.

**[0083]** Next, we designed a novel strategy to separate the active dsRNA from the undesired ELVd scaffold. For this purpose, we took advantage of the permuted intron-exon (PIE) reaction (Price, J. V, Engberg, J., and Cech, T.R. (1987) *J. Mol. Biol.*, 196, 49-60; Puttaraju, M. and Been, M. (1992) *Nucleic Acids Res.*, 20, 5357-5364). We built a new plasmid in which a cDNA corresponding to the *T. thermophila* 26S rRNA group-I intron with 10 nt of the flanking exons (SEQ ID NO: 6) was split and permuted, and the 3' fragment (SEQ ID NO: 15) inserted between the 5' moiety of ELVd and the plus strand of the dsRNA, and the 5' fragment (SEQ ID NO: 16) inserted between the minus strand of the dsRNA and the 3' moiety of the ELVd (see Figure 2). We built this plasmid by amplifying the DNAs by PCR with the Phusion high-fidelity DNA polymerase and assembling the fragment by the Gibson reaction. The sequence of the plasmid was confirmed by sequencing. The selected plasmid, along pLtrnISm to co-express the eggplant tRNA ligase, was used to electroporate competent *E. coli* cells. Transformed bacteria were selected in LB plates containing ampicillin and chloramphenicol. Next, we grew liquid cultures in TB, containing the same antibiotics, and extracted total RNA from bacteria using phenol:chloroform. The RNA preparation was analyzed by PAGE under non-denaturing and denaturing (8 M urea) conditions (Figure 7). The alteration of electrophoretic migration from non-denaturing to denaturing conditions confirmed the production in *E. coli* of a circular RNA molecule consisting of a 100-bp dsRNA locked at both ends by the exon fragments (Figure 3), as a result of the self-cleavage of the regular intron I and the permuted intron I (Figure 2).

**[0084]** In conclusion, we have worked out a new method to produce dsRNA that can be used in RNAi-based biotechnology applications. The new method is based on the activity of two versions of a self-splicing group-I intron, the second in a permuted form. The first intron stabilizes the DNA inverted repeats that are required to transcribe a dsRNA in *E. coli*, while is efficiently processed from the final transcript. The second (permuted) intron cleaves the dsRNA from the viroid scaffold producing a circular molecule consisting of the dsRNA locked at both ends by the exon fragments (Figure 3).

EP 3 822 353 A1

SEQUENCE LISTING

<110> CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS (CSIC)  
 UNIVERSITAT POLITÈCNICA DE VALÈNCIA

5

<120> METHOD FOR THE PRODUCTION OF dsRNA

<130> EP1641.1485

<160> 24

10

<170> PatentIn version 3.5

<210> 1  
 <211> 100  
 <212> DNA  
 <213> Ceratititis capitata

15

<400> 1  
 gctgaccgaa gcaccgttga acccgaaggc taaccgcgag aagatgaccc agatcatggt 60  
 cgagaccttc aactcgccgg ccatgtatgt ggccatccag 100

20

<210> 2  
 <211> 100  
 <212> DNA  
 <213> Ceratititis capitata

25

<400> 2  
 ctggatggcc acatacatgg ccggcgagtt gaaggtctcg aacatgatct gggtcatctt 60  
 ctcgcggtta gccttcgggt tcaacgggtc ttcggtcagc 100

30

<210> 3  
 <211> 413  
 <212> DNA  
 <213> Tetrahymena thermophila

35

<400> 3  
 aaatagcaat atttaccttt ggagggaaaa gttatcaggc atgcacctgg tagctagtct 60  
 ttaaaccaat agattgcatc ggtttaaag gcaagaccgt caaattgcgg gaaaggggtc 120  
 aacagccgtt cagtaccaag tctcagggga aactttgaga tggccttgca aagggtatgg 180  
 taataagctg acggacatgg tcctaaccac gcagccaagt cctaagtcaa cagatcttct 240  
 gttgatatgg atgcagttca cagactaaat gtcggtcggg gaagatgtat tcttctcata 300  
 agatatagtc ggacctctcc ttaatgggag ctagcggatg aagtgatgca aactggagc 360  
 cgctgggaac taatttgtat gcgaaagtat attgattagt tttggagtac tcg 413

40

45

50

<210> 4  
 <211> 10  
 <212> DNA  
 <213> Tetrahymena thermophila

55

<400> 4  
 atgactctct 10

EP 3 822 353 A1

<210> 5  
 <211> 10  
 <212> DNA  
 <213> Tetrahymena thermophila  
 5  
 <400> 5  
 taaggtagcc 10

<210> 6  
 <211> 433  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Autocatalytic intron flanked by exon fragments

<400> 6  
 atgactctct aatagcaat attaccttt ggagggaaaa gttatcaggc atgcacctgg 60  
 20 tagctagtct ttaaaccaat agattgcatc ggtttaaaag gcaagaccgt caaattgctg 120  
 gaaaggggtc aacagccgtt cagtaccaag tctcagggga aactttgaga tggccttgca 180  
 aagggtatgg taataagctg acggacatgg tcttaaccac gcagccaagt cctaagtcaa 240  
 25 cagatcttct gttgatatgg atgcagtca cagactaaat gtcggtcggg gaagatgtat 300  
 tcttctcata agatatagtc ggacctctcc ttaatgggag cttagcggatg aagtgatgca 360  
 aactggagc cgctgggaac taatttgtat gcgaaagtat attgattagt tttggagtac 420  
 30 tcgtaaggta gcc 433

<210> 7  
 <211> 333  
 <212> DNA  
 <213> Eggplant latent viroid

<400> 7  
 ggggtggtgtg tgccaccct gatgagaccg aaaggtcgaa atggggtttc gccatgggtc 60  
 40 gggactttaa attcggagga ttcgtccttt aaacgttctt ccaagagtcc cttccccaaa 120  
 cccttacttt gtaagtgtgg ttcggcgaat gtaccgtttc gtcctttcgg actcatcagg 180  
 45 gaaagtacac actttccgac ggtgggttcg tcgacacctc tccccctccc aggtactatc 240  
 ccctttcaag gatgtgttcc ctaggagggg ggggtgtacct cttttggatt gctccggcct 300  
 tccaggagag atagaggacg acctctcccc ata 333

<210> 8  
 <211> 247  
 <212> DNA  
 <213> Avocado sunblotch viroid

<400> 8  
 tttattagaa caagaagtga ggatagatt aaactttggt tgacgaaacc aggtctgttc 60

EP 3 822 353 A1

cgactttccg actctgagtt tcgacttggt agagaaggag gactcgtggt gaacttttat 120  
taaaaaaatt agttcactcg tcttcaatct cttgatcact tcgtctcttc agggaaagat 180  
5 ggaagaaca ctgatgagtc tcgcaaggtt tactcctcta tcttcattgt tttttataaa 240  
aatcttg 247

<210> 9  
10 <211> 399  
<212> DNA  
<213> Chrysanthemum chlorotic mottle viroid

<400> 9  
15 ggcacctgac gtcggtgtcc tgatgaagat ccatgacagg atcgaaacct cttccagttt 60  
cggtttgtgt gggagtaaag ctttcgctct ctccacagcc tcatacaggaa acccacttca 120  
ggtctcgact ggaaggtcgt taaacttccc ctctaagcgg agtagaggta aatacctccg 180  
20 tccaaccccc ggaggaaagg ggttgggacc cggaacagat ccagttccgg tcctttggag 240  
tccatttctc tcgttgata ttctcctcgg agaagggaga tggggccagt cccagtcggt 300  
tcgctctcgt agtcacagcc actggggaac ctaggcagat ggctggacgg agtcttagtc 360  
25 cactccagag gaccttgggt ttgaaacccc caagaggtc 399

<210> 10  
30 <211> 337  
<212> DNA  
<213> Peach latent mosaic viroid

<400> 10  
35 gtcataagtt tcgtcgcatt tcagcgactc atcagtgggc ttagcccaga cttatgagag 60  
agtaaagacc tctcagcccc tccacettgg ggtgccctat tcggagcact gcagttcccc 120  
atagaaaggc taagcacctc gcaatgaggt aaggtgggac ttttccttct ggaaccaagc 180  
40 ggttgggtcc gaggggggtg tgatccaggt accgccgtag aaactggatt acgacgtcta 240  
cccgggattc aaaccctcc cctccagaag tgattctgga tgaagagtct gtgctaagca 300  
cactgacgag tctctgagat gagacgaaac tcttctt 337

<210> 11  
45 <211> 386  
<212> DNA  
<213> Eggplant latent viroid

<400> 11  
50 ccccataggg tgggtgtgtc caccctgat gagaccgaaa ggtcgaaatg gggtttcgcc 60  
atgggtcggg actttaaatt cggagattc gtcctttaa cgttcctcca agagtccctt 120  
55 ccccaaacc ttactttgta agtgtggtc ggogaatgta ccgtttcgtc ctttcggact 180  
catcagggaa agtacacact ttccgacggt gggttcgtcg acacctctcc ccctcccagg 240

EP 3 822 353 A1

tactatcccc tttcaaggat gtgttcacct ggaggggtggg tgtacctctt ttggattgct 300  
 ccggccttcc aggagagata gaggacgacc tctccccata ggggtgtgtg tgccaccct 360  
 5 gatgagaccg aaaggtcgaa atgggg 386

<210> 12  
 <211> 1019  
 <212> DNA  
 10 <213> Artificial Sequence

<220>  
 <223> A plant viroid comprising the cDNAs of two strands of a target  
 dsRNA  
 15 Separated by an autocatalytic intron flanked by exon fragments  
 inserted in its nucleotide sequence.

<400> 12  
 ccccataggg tgggtgtgtgc caccctgat gagaccgaaa ggtcgaaatg gggtttcgcc 60  
 20 atgggtcggg actttaaat cggaggattc gtcctttaa cgttcctcca agagtccctt 120  
 ccccaaacc ttactttgta agtgtggttc ggcgaatgta ccgtttcgtc ctttcggact 180  
 catcagggaa agtacacact ttccgacggt gggttcgtcg acacctctcc ccctcccagg 240  
 25 tactatcccc ttgctgaccg aagcaccgtt gaaccggaag gctaaccgag agaagatgac 300  
 ccagatcatg ttcgagacct tcaactcgcc ggccatgtat gtggccatcc agatgactct 360  
 ctaaatagca atatttacct ttggagggaa aagttatcag gcatgcacct ggtagctagt 420  
 30 ctttaaacca atagattgca tcggttttaa aggcaagacc gtcaaattgc gggaaagggg 480  
 tcaacagccg ttcagtacca agtctcaggg gaaactttga gatggccttg caaagggtat 540  
 35 ggtaataagc tgacggacat ggtcctaacc acgcagcaa gtcctaagtc aacagatctt 600  
 ctggtgatat ggatgcagtt cacagactaa atgtcggtcg ggaagatgt attcttctca 660  
 taagatatag tcggacctct ccttaatggg agctagcggg tgaagtgat caacactgga 720  
 40 gccgctggga actaatttgt atgcgaaagt atattgatta gttttggagt actcgtaaag 780  
 tagccctgga tggccacata catggccggc gagttgaagg tctcgaacat gatctgggtc 840  
 atcttctcgc ggtagcctt cgggttcaac ggtgcttcgg tcagctcaag gatgtgttcc 900  
 45 ctaggagggt ggggtgtacct cttttggatt gctccggcct tccaggagag atagaggacg 960  
 acctctcccc ataggggtgt gtgtgccacc cctgatgaga ccgaaaggtc gaaatgggg 1019

<210> 13  
 <211> 1167  
 <212> PRT  
 <213> Solanum melongena

<400> 13  
 55 Met Ser Val Ser His Arg Val Ile Tyr Ser Phe Thr His Tyr Lys Leu

EP 3 822 353 A1

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



EP 3 822 353 A1

Arg Glu Ala Trp Gly Thr Lys Ala Ser Lys Lys Gln Ala Glu Phe Asn  
 260 265 270  
 5  
 Glu Phe Leu Glu Arg Asn Arg Met Cys Ile Ser Met Glu Leu Val Thr  
 275 280 285  
 Ala Val Leu Gly Asp His Gly Gln Arg Pro Arg Asp Asp Tyr Ala Val  
 10 290 295 300  
 Val Thr Ala Val Thr Glu Leu Gly Asn Gly Lys Pro Thr Phe Tyr Ser  
 305 310 315 320  
 15  
 Thr Pro Asp Val Ile Ala Phe Cys Arg Glu Trp Arg Leu Pro Thr Asn  
 325 330 335  
 His Val Trp Leu Phe Ser Thr Arg Lys Ser Val Thr Ser Phe Phe Ala  
 20 340 345 350  
 Ala Tyr Asp Ala Leu Cys Glu Glu Gly Thr Ala Thr Thr Val Cys Glu  
 25 355 360 365  
 Ala Leu Ser Glu Val Ala Asp Ile Ser Val Pro Gly Ser Lys Asp His  
 370 375 380  
 30  
 Ile Lys Val Gln Gly Glu Ile Leu Glu Gly Leu Val Ala Arg Ile Val  
 385 390 395 400  
 Lys Arg Glu Ser Ser Glu His Met Glu Arg Val Leu Arg Asp Phe Pro  
 35 405 410 415  
 Pro Pro Pro Ser Glu Gly Glu Gly Leu Asp Leu Gly Pro Thr Leu Arg  
 420 425 430  
 40  
 Glu Ile Cys Ala Ala Asn Arg Ser Glu Lys Gln Gln Ile Lys Ala Leu  
 435 440 445  
 Leu Gln Ser Ala Gly Thr Ala Phe Cys Pro Asn Tyr Leu Asp Trp Phe  
 45 450 455 460  
 Gly Asp Glu Asn Ser Gly Ser His Ser Arg Asn Ala Asp Arg Ser Val  
 50 465 470 475 480  
 Val Ser Lys Phe Leu Gln Ser His Pro Ala Asp Leu Tyr Thr Gly Lys  
 485 490 495  
 55  
 Ile Gln Glu Met Val Arg Leu Met Arg Glu Lys Arg Phe Pro Ala Ala  
 500 505 510

EP 3 822 353 A1

Phe Lys Cys His Tyr Asn Leu His Lys Ile Asn Asp Val Ser Ser Asn  
 515 520 525  
 5 Asn Leu Pro Phe Lys Met Val Ile His Val Tyr Ser Asp Ser Gly Phe  
 530 535 540  
 10 Arg Arg Tyr Gln Lys Glu Met Arg His Lys Pro Gly Leu Trp Pro Leu  
 545 550 555 560  
 Tyr Arg Gly Phe Phe Val Asp Leu Asp Leu Phe Lys Val Asn Glu Lys  
 565 570 575  
 15 Lys Thr Ala Glu Met Ala Gly Ser Asn Asn Gln Met Val Lys Asn Val  
 580 585 590  
 20 Glu Glu Asp Asn Ser Leu Ala Asp Glu Asp Ala Asn Leu Met Val Lys  
 595 600 605  
 Met Lys Phe Leu Thr Tyr Lys Leu Arg Thr Phe Leu Ile Arg Asn Gly  
 610 615 620  
 25 Leu Ser Thr Leu Phe Lys Glu Gly Pro Ser Ala Tyr Lys Ser Tyr Tyr  
 625 630 635 640  
 30 Leu Arg Gln Met Lys Ile Trp Asn Thr Ser Ala Ala Lys Gln Arg Glu  
 645 650 655  
 35 Leu Ser Lys Met Leu Asp Glu Trp Ala Val Tyr Ile Arg Arg Lys Tyr  
 660 665 670  
 Gly Asn Lys Pro Leu Ser Ser Ser Thr Tyr Leu Ser Glu Ala Glu Pro  
 675 680 685  
 40 Phe Leu Glu Gln Tyr Ala Lys Arg Ser Pro Gln Asn His Ala Leu Ile  
 690 695 700  
 45 Gly Ser Ala Gly Asn Phe Val Lys Val Glu Asp Phe Met Ala Ile Val  
 705 710 715 720  
 Glu Gly Glu Asp Glu Glu Gly Asp Leu Glu Pro Ala Lys Asp Ile Ala  
 725 730 735  
 50 Pro Ser Ser Pro Ser Ile Ser Thr Arg Asp Met Val Ala Lys Asn Glu  
 740 745 750  
 55 Gly Leu Ile Ile Phe Phe Pro Gly Ile Pro Gly Cys Ala Lys Ser Ala  
 755 760 765

EP 3 822 353 A1

Leu Cys Lys Glu Ile Leu Asn Ala Pro Gly Gly Leu Gly Asp Asp Arg  
 770 775 780

5 Pro Val Asn Ser Leu Met Gly Asp Leu Ile Lys Gly Arg Tyr Trp Gln  
 785 790 795 800

10 Lys Val Ala Asp Glu Arg Arg Arg Lys Pro Tyr Ser Ile Met Leu Ala  
 805 810 815

15 Asp Lys Asn Ala Pro Asn Glu Glu Val Trp Lys Gln Ile Glu Asn Met  
 820 825 830

20 Cys Leu Ser Thr Gly Ala Ser Ala Ile Pro Val Ile Pro Asp Ser Glu  
 835 840 845

25 Gly Thr Glu Thr Asn Pro Phe Ser Ile Asp Ala Leu Ala Val Phe Ile  
 850 855 860

30 Phe Arg Val Leu His Arg Val Asn His Pro Gly Asn Leu Asp Lys Ser  
 865 870 875 880

35 Ser Pro Asn Ala Gly Tyr Val Met Leu Met Phe Tyr His Leu Tyr Asp  
 885 890 895

40 Gly Lys Ser Arg Gln Glu Phe Glu Ser Glu Leu Ile Glu Arg Phe Gly  
 900 905 910

45 Ser Leu Val Arg Ile Pro Val Leu Lys Pro Glu Arg Ser Pro Leu Pro  
 915 920 925

50 Asp Ser Val Arg Ser Ile Ile Glu Glu Gly Leu Ser Leu Tyr Arg Leu  
 930 935 940

55 His Thr Thr Lys His Gly Arg Leu Glu Ser Thr Lys Gly Thr Tyr Val  
 945 950 955 960

60 Gln Glu Trp Val Lys Trp Glu Lys Gln Leu Arg Asp Ile Leu Leu Gly  
 965 970 975

65 Asn Ala Asp Tyr Leu Asn Ser Ile Gln Val Pro Phe Glu Phe Ala Val  
 980 985 990

70 Lys Glu Val Leu Glu Gln Leu Lys Val Ile Ala Arg Gly Glu Tyr Ala  
 995 1000 1005

75 Val Pro Ala Glu Lys Arg Lys Leu Gly Ser Ile Val Phe Ala Ala

EP 3 822 353 A1

	1010		1015		1020	
5	Ile Ser Leu Pro Val Pro Glu Ile Leu Gly Leu Leu Asn Asp Leu		1025 1030		1035	
10	Ala Lys Lys Asp Pro Lys Val Gly Asp Phe Ile Lys Asp Lys Ser		1040 1045		1050	
15	Met Glu Ser Ser Ile Gln Lys Ala His Leu Thr Leu Ala His Lys		1055 1060		1065	
20	Arg Ser His Gly Val Thr Ala Val Ala Asn Tyr Gly Ser Phe Leu		1070 1075		1080	
25	His Gln Lys Val Pro Val Asp Val Ala Ala Leu Leu Phe Ser Asp		1085 1090		1095	
30	Lys Leu Ala Ala Leu Glu Ala Glu Pro Gly Ser Val Glu Gly Glu		1100 1105		1110	
35	Lys Ile Asn Ser Lys Asn Ser Trp Pro His Ile Thr Leu Trp Ser		1115 1120		1125	
40	Gly Ala Gly Val Ala Ala Lys Asp Ala Asn Thr Leu Pro Gln Leu		1130 1135		1140	
45	Leu Ser Gln Gly Lys Ala Thr Arg Ile Asp Ile Asn Pro Pro Val		1145 1150		1155	
50	Thr Ile Thr Gly Thr Leu Glu Phe Phe		1160 1165			
55	<210> 14					
	<211> 3848					
	<212> DNA					
	<213> Solanum melongena					
60	<400> 14					
	gtgactggct caaaggcggt agtagtggct cttcttcccc gctggaattt ctcgcctatt					60
	ttatattgat tgacagcagc cgagtcttcc ttttgtgtag aactgaatgt cggtttcgca					120
	tagggtcatt tactctttca ctcattacaa actctataat ctctcttctt ctttatcatc					180
	tttgcccttct agaatcttct tcccttttca atctccttcc tttcacacgt tctcttcaact					240
	catgcccaac aatcaggaaa ggggtgggta tgaaggaaaa aatggcaag tgaggccaag					300
	ttccaatagg gtaccaggct cgtcttcaaa tgtggaacct gtatctgctg caactgctga					360
	agccattacc gaccgtctaa agtccgtgga tattactgaa agtgggtgcac agtctagtgt					420

EP 3 822 353 A1

tccagtcaca tctcttcagt ttggcagcgt tggattagca cccagtcac ctgtgcaaca 480  
tcaaaaagta atctggaaac ccaaatcata tggaacagtg tctggagccc cagtggttga 540  
5 agctggaaaa acaccagttg aacaaaaaag tgctctttta agtaaattat tcaagggtaa 600  
tttattggaa aatthttactg tagataactc aacattctcg agagcccaag taagggccac 660  
tttctacca aaatthgaga atgagaaatc agatcaggag atcaggacaa ggatgataga 720  
10 gatggctctcc aaaggcttgg ctatagtcga ggtcacactt aagcattctg gatctctttt 780  
tatgtatgct gggcatgaag gtggagcata tgccaagaat agcttcggga atatctatac 840  
tgccgttggc gtctthgttc ttggacggat gtttcgtgag gcatggggaa ctaaagcaag 900  
15 caagaagcaa gcagagttca atgagthttct tgagcgcgaat cgtatgtgca tatcaatgga 960  
gttggtcacg gcagtgthtg gggaccacgg acaacgcccc cgagatgatt atgcgthtgt 1020  
gactgcagtc acggagthtg gaaatggaaa accaactthc tathcaactc ccgatgthaat 1080  
20 tgctthtttg agggaatggc gattaccaac aaatcatgta tggctgthct caacaaggaa 1140  
atcagtgact tcctthcttg ctgcgtatga tgcactthgc gaggaaggta cagcaaccac 1200  
cgthtgcgag gctctcagcg aagthgtgta tathctgta cctggatcaa aagaccatat 1260  
25 aaaagtgcag ggtgaaatth tggagggthct cgtggccccg atcgtaaaac gtgagagctc 1320  
agagcatatg gagcggthtc tgagagatth tcctoctccg ccatcagagg gtgagggtht 1380  
ggacctggga cctacgthac gtgaaatthg tgctgcaaac agatcagaaa agcagcaaat 1440  
30 aaaggcactt cthcagagtg ctggcacggc thctgccccg aathatthgg actgththtg 1500  
agatgaaaac tctgthtcac atthcaagaaa tgctgatcga tctgthgtct caaagthctt 1560  
35 acaatcacat cctgctgatc thttatacagg aaaaatacag gaaatgthtc gcttgatgag 1620  
ggaaaagcgc thtctgctg cththcaagtg tcattataac thacataaaa thaatgatgt 1680  
atcgagthaac aacctgcctt tcaaatggg gatccatgta tatagthgatt caggctthccg 1740  
40 ccggtaccag aaagagatga ggcacaaaacc aggactatgg cctthgtatc gagctththt 1800  
tgthgactcg gaththattca aggtcaatga gaagaaaact gctgaaatgg caggaagcaa 1860  
caatcaaatg gtaaaaaatg tggaagagga caacagthta gctgatgag atgcaaatct 1920  
45 gatggthcaag atgaaatthc thactthcaa gthgagaact ththtgatcc gthaatggctt 1980  
gtcgactctt thcaaagaag gacctthctg gtataagthct ththacctga ggcaaatgaa 2040  
aaththgagat actthcagcag ccaagcaacg agaactcagc aagatgcttg atgagthggc 2100  
50 agtatatata cgcagaaaat atgggaacaa accatthgca thcatccacat acctaatgth 2160  
agctgagcct thcctthgaa aatathgcaa gcgtagthca caaatcatg cththgatagg 2220  
atctgctgga aaththgtca aagthgaaga thctcatggct atthgtthgag gagaagatga 2280  
55 agagggthgat cthcagcctg cgaaagatath tgctcctthca agccctagth ththccaccag 2340

EP 3 822 353 A1

agacatggtg gcaaagaatg agggctctcat tattttcttt ccaggaatac caggttgtgc 2400  
 taaatctgca ctttgaagg aaatactgaa tgctccagga gggcttgag atgatcgacc 2460  
 5 agttaacagt ttaatgggtg atcttattaa aggtagatat tggcaaaaag ttgctgatga 2520  
 acgtcgaaga aaaccttact cgatcatgct tgctgacaag aatgcaccaa atgaggaagt 2580  
 atggaaacaa attgagaaca tgtgcctaag caccggagca tctgctattc cagttatacc 2640  
 10 tgattcagaa ggaactgaaa ctaatccatt ctctattgat gcacttgccg tttttatatt 2700  
 ccgagtactt caccgtgtca atcatccggg aaatcttgac aagtcatctc caaatgctgg 2760  
 atatgtgatg cttatgtttt atcaccttta tgatggaaag agccgtcagg agttcgagag 2820  
 15 tgagcttatt gaacgttttg gatcgcttgt cagaattcct gtactgaaac ctgagaggtc 2880  
 tcctcttccg gattctgtga ggtctattat cgaggagga ctcagtctgt acagacttca 2940  
 tacaacgaaa catggaagat tggagtctac aaaagggaca tatgtacaag agtgggttaa 3000  
 20 atgggagaag caattgagag atattctact tggaaatgca gactatctca attcaatata 3060  
 ggttccattt gaatttgccg ttaaagaagt ccttgaacaa ctgaaagtta ttgagagggg 3120  
 25 cgaatatgca gtgcctgctg agaagaggaa gctaggatcc attgtattcg ccgctatcag 3180  
 cctgccagtt ccagaaattc taggtcttct aaatgatcta gcaaagaaag atccaaaggt 3240  
 tggcgatttc attaaggaca agagcatgga gagcagcatt cagaaggccc atcttaccct 3300  
 30 ggctcacaag agaagtcaag gtgtcactgc agttgccaat tacggttcct ttcttcatca 3360  
 aaaggtgcc a gtagacgtg ctgctttggt gttctccgat aaattggctg cactagaagc 3420  
 tgagcctggc tctgttgaag gtgaaaagat caattctaaa aactcatggc cccatatcac 3480  
 35 attatggtct ggtgcaggag ttgccgcaaa agatgccaat aactaccac agttactttc 3540  
 ccaagggag gctaccgca ttgatataaa tccaccggtc actataactg gcactctcga 3600  
 attcttttga actctacttg taattcctct tgtgtggatt tacataatgt gaaccctaca 3660  
 40 tgggttctga gtgaatcatg taatgattct cctgagtttt aggaacaatt ttgtagcagt 3720  
 cttgcagttg caggatagaa ctagatctag agaaaatgta aagatgttta aggcaataaa 3780  
 45 ggtcttgagc tgtgggcata ttgaaaatcc ccagtagcat ttttcttgag caaaaaaaaa 3840  
 aaaaaaaaa 3848  
  
 <210> 15  
 50 <211> 188  
 <212> DNA  
 <213> Tetrahymena thermophila  
  
 <400> 15  
 55 cttctgttga tatggatgca gttcacagac taaatgtcgg tcggggaaga tgtattcttc 60  
 tcataagata tagtcggacc tctccttaat gggagctagc ggatgaagtg atgcaacact 120

EP 3 822 353 A1

ggagccgctg ggaactaatt tgtatgcgaa agtatattga ttagtitttg agtactcgtg 180  
 aggtagcc 188  
 5  
 <210> 16  
 <211> 245  
 <212> DNA  
 <213> Tetrahymena thermophila  
 10  
 <400> 16  
 atgactctct aaatagcaat atttaccttt ggagggaaaa gttatcaggc atgcacctgg 60  
 tagctagtct ttaaaccaat agattgcatc ggtttaaaag gcaagaccgt caaattgcgg 120  
 15  
 gaaaggggtc aacagccggt cagtaccaag tctcagggga aactttgaga tggccttgca 180  
 aagggtatgg taataagctg acggacatgg tctaaccac gcagccaagt cctaagtcaa 240  
 cagat 245  
 20  
 <210> 17  
 <211> 1452  
 <212> DNA  
 <213> Artificial Sequence  
 25  
 <220>  
 <223> Nucleotide sequence of the invention  
 <400> 17  
 30  
 ccccataggg tgggtgtgtc caccctgat gagaccgaaa ggtcgaaatg gggtttcgcc 60  
 atgggtcggg actttaaat cggaggattc gtcctttaa cgttcctcca agagtccctt 120  
 ccccaaacc ttactttgta agtgtggttc ggcgaatgta ccgtttcgtc ctttcggact 180  
 35  
 catcagggaa agtacacact ttccgacggt gggttcgtcg acacctctcc ccctcccagg 240  
 tactatcccc ttcttctggt gatatggatg cagttcacag actaaatgtc ggtcggggaa 300  
 gatgtattct tctcataaga tatagtcgga cctctcctta atgggagcta gcggatgaag 360  
 40  
 tgatgcaaca ctggagccgc tgggaactaa tttgtatgcy aaagtatatt gattagtttt 420  
 ggagtactcg taaggtagcc gctgaccgaa gcaccgttga acccgaaggc taaccgcgag 480  
 aagatgacct agatcatggt cgagacctc aactcgcggg ccatgtatgt ggccatccag 540  
 45  
 atgactctct aaatagcaat atttaccttt ggagggaaaa gttatcaggc atgcacctgg 600  
 tagctagtct ttaaaccaat agattgcatc ggtttaaaag gcaagaccgt caaattgcgg 660  
 gaaaggggtc aacagccggt cagtaccaag tctcagggga aactttgaga tggccttgca 720  
 50  
 aagggtatgg taataagctg acggacatgg tctaaccac gcagccaagt cctaagtcaa 780  
 cagatcttct gttgatatgg atgcagttca cagactaaat gtcggtcggg gaagatgat 840  
 55  
 tcttctcata agatatagtc ggacctctcc ttaatgggag ctacgggatg aagtgatgca 900  
 aactgggagc cgctgggaac taatttctat gcgaaagtat attgattagt tttggagtac 960

EP 3 822 353 A1

	tcgtaaggta gccctggatg gccacataca tggccggcga gttgaaggtc tcgaacatga	1020
	tctgggtcat cttctcggg ttagccttcg ggttcaacgg tgcttcggtc agcatgactc	1080
5	tctaaatagc aatatttacc tttggagggg aaagttatca ggcatgcacc tggtagctag	1140
	tctttaaacc aatagattgc atcggtttaa aaggcaagac cgtcaaattg cgggaaaggg	1200
	gtcaacagcc gttcagtacc aagtctcagg ggaaactttg agatggcctt gcaaagggta	1260
10	tggttaataag ctgacggaca tggtcctaac cacgcagcca agtcctaagt caacagattc	1320
	aaggatgtgt tccctaggag ggtgggtgta cctcttttgg attgctccgg ccttccagga	1380
	gagatagagg acgacctctc cccatagggt ggtgtgtgcc acccctgatg agaccgaaag	1440
15	gtcgaaatgg gg	1452
	<210> 18	
	<211> 3820	
20	<212> DNA	
	<213> Arabidopsis thaliana	
	<400> 18	
	agaaacgacg ttgtgttcaa gttactaaat cgcgtgtgtt ttttactttt ctgcaactaa	60
25	atcgtgtttc ttcctaagaa ccctaaatct gaactgtctc atcttctccg atatcagtga	120
	gatcaataaa tcatcttctc accctagacg acagggaaac tcggggattg attcggaaaa	180
	acgagttcgt ttttctgcaa gaaaaagaga gatcacgctg agcagaagtg gcaagtaaaa	240
30	ccgaagatgg atgtccatt tgaatccggt gatagttcgg caacagttgt tgcagaagca	300
	gtaaacaacc aatttggcgg gttaagtctt aaggaaagca aactaatgc gccagtctta	360
	ccgagtcaaa ctacaagtaa ccaccgtgtt caaaatctag tatggaaacc taagtcatat	420
35	ggaacagtta gtggatcatc ttcggcaaca gaagttggta aaacatcggc tgtctcacag	480
	ataggttctt ctggagatac aaaagttggt ctcaatctga gcaaaatfff cggaggtaat	540
	ttgtttgaaa aattttcggg ggacaaatcg acttattgtc atgctcagat cagagccact	600
	ttctatccaa aatttgagaa cgaaaagact gaccaagaga taagaacaag aatgattgag	660
	atggtatcta aaggattggc aacactggag gtatctctta agcattcagg ctctctcttt	720
45	atgtacgccg gtcacaaggg gggagcatat gcaaagaaca gttttggtaa tatttatact	780
	gcggttggtg tttttgttct ttcacggatg ttcagggagg cttggggaac taaggctcca	840
	aagaaagaag cagagttaa tgatfffctt gagaaaaatc gcatgtgcat atctatggaa	900
50	ctggtaaccg ctgttcttgg agatcatggt caacgcccac tggatgatta tgtggtagtg	960
	actgctgtta ccgagttagg taatgggaag ccccgattct attcaacttc cgaataatt	1020
	tcattttgcc ggaaatggcg tctaccaca aatcatgttt ggctgttttc cacaaggaaa	1080
55	tcagtgacct ctttttctgc ggcgtttgat gcaactgtgtg aagaagggat agcaacctcg	1140



EP 3 822 353 A1

gtctgtagag ctcttgatga agtagctgat atatcagtcc cagcctcaaa ggaccatggt 1200  
 aaggtgcagg gtgagatatt agagggctctt gttgctcgta tcgtgagcag tcagagctca 1260  
 5 agagatatgg aaaatgtctt gagagatcat cctccaccac cctgtgatgg agctaactctt 1320  
 gatctgggac tcagtctaag agagatatgt gctgcccata gatctaacga gaaacagcaa 1380  
 atgagagcac ttttaaggag tgttgggcca agcttttgcc ccagtgacgt ggaatggttc 1440  
 10 ggagatgaat ctcatccgaa aagtgctgac aaatctgtta taacaaaatt cctgcaatct 1500  
 cagcctgcag attattcaac tagtaaatta caggaaatgg tacgcttgat gaaggaaaaa 1560  
 cgtcttccag ccgcattcaa gtgttaccat aattttcata gagctgagga catatcacct 1620  
 15 gataaccttt tttataaatt ggttgtccat gtgcacagcg attcaggatt taggcggttac 1680  
 cacaaagaga tgaggcacat gcctagtttg tggcctttat atcgaggttt ctttgttgat 1740  
 20 attaatttgt tcaagtcaaa caaagggagg gatctgatgg ctctgaaaag cattgataat 1800  
 gccagtgaaa atgatggctg aggagaaaag gatggtttgg ctgatgatga tgctaactta 1860  
 atgattaanaa tgaagtttct cacatacaag ttgagaacct ttttgatccg taatggccta 1920  
 25 tcaattctat ttaaagacgg ggcagcagct taaaaaactt actaccttag gcaaatgaaa 1980  
 atctggggta cgtcagatgg aaagcagaaa gaactttgca agatgcttga tgaatgggct 2040  
 gcttacataa gaagaaagtg tgggaatgat cagctatctt catcaacata cttaagtga 2100  
 30 gctgagccat tcctggagca gtacgcgaaa cggagcccaa agaaccatat tttgataggg 2160  
 tctgctggga acctagttag aactgaggac ttcttggcca ttgttgacgg tgatctagat 2220  
 gaagaaggcg atcttgtgaa gaaacaagga gtgacaccag ctactcctga gccagctgtg 2280  
 35 aaggaagctg tccaaaagga tgaggggtta attgtattct ttccaggaat tcctggatct 2340  
 gctaaatctg cactttgcaa ggagttattg aacgccccag gaggctttgg agatgatagg 2400  
 40 ccagtgcata ctctgatggg tgatcttgtc aaaggaaaat attggccaaa ggttgctgat 2460  
 gaacgtcgta aaaaaccaca atcgattatg ttggctgaca aaaatgcccc aatgaagat 2520  
 gtctggagac agattgaaga tatgtgtcgg agaactaggg cttctgcagt tccaatcgtt 2580  
 45 gctgattctg aaggaaccga taaaaatcca tattcacttg atgccttggc tgtattcatg 2640  
 ttccgtgtac ttcaacgagt taacctccg ggaaagctcg acaaggaatc ttcaaatgca 2700  
 ggttatgtgc tattgatggt ttatcacctt tatgagggaa agaaccgtaa tgaatttgag 2760  
 50 agtgaattga ttgagcgctt tggctccctc atcaagatgc cactattgaa atcagacagg 2820  
 actcctttac ctgatcctgt taaatcagtt cttgaggagg gcattgattt gttcaatctc 2880  
 55 catagtagaa gacatgggag actcgagtca acaaaaaggaa catatgcagc ggaatggacc 2940  
 aatgggaga aacaactacg agataactta gttgcaaatt ctgagtatct cagttctatt 3000

EP 3 822 353 A1

caggttccat ttgagtctat ggttcatcaa gtgcgagaag agctaaaaac aatagcgaag 3060  
 ggtgattaca agccaccaag ttcggagaaa agaaaacacg ggtctattgt tttcgctgcc 3120  
 5 atcaacttgc ctgctactca agttcacagt cttcttgaaa agttggctgc agcaaacca 3180  
 acaatgagat cttttctaga gggaaagaaa aagagcatac aggaaaaact tgaacggctct 3240  
 cacgtgacgc tcgcccacaa gagaagccat ggcgtagcaa ctgtagccag ctatagttag 3300  
 10 cacttgaaca gagaggtacc cgtagagctc accgagctca tctacaacga caagatggct 3360  
 gctctaacag cccatgttgg atctgtggac ggagagaccg tagtctccaa gaacgaatgg 3420  
 ccacatgtta cattgtggac agcgggaaggc gttactgcga aagaggccaa cacgttacct 3480  
 15 cagctttact tagaaggaaa ggcgagccgc ttgggtgatag atcctccggt gtcaatctca 3540  
 ggtcctctgg agttttctg aataacttgat taaacatgga agtttctctc ttgagggagg 3600  
 ttgctcgtgg aatgggacac atatggttgt tataataaac catttccatt gtcatgagat 3660  
 20 tttgaggtta atatatactt tacttgttca ttattttatt tgggtgttga ataaatgata 3720  
 taaatggctc ttgataatct gcattcattg agatatcaaa tatttactct agagaagagt 3780  
 25 gtcatataga ttgatggtcc acaatcaatg aaatttttgg 3820

<210> 19  
 <211> 3876  
 <212> DNA  
 30 <213> Solanum lycopersicum

<400> 19  
 tatggaaccc aattggagag tgccggttcc ccctgaaatt ttctggcttc atctgtatac 60  
 35 aacagttgag tcttcctttt ctatagcact gaatgtcggg tttcgttcac tctttcactg 120  
 ctaaactcca tctccctcac aatctctctt cttcttcttc ctcatatttc ccttctagaa 180  
 tcttctttcg ctttcaatct cgttccctca cacacagtac ctcgctattc tcttactca 240  
 40 tgccaaacaa tcagagaaag ggtggttata aagaaaaaaa atggcaagtg aggagttcca 300  
 atagggtagc aggttcatct tcaaatgtcg aacctgcac tcctgcaact actggagcca 360  
 ttgctgaccg tctaagttcc ttgaatatta ctgaaagtgg tgcacagtct agtgttccag 420  
 45 ttgcatctct tcagtttggc agcgttggat tagcacccca atcacctgtg cagcatcaaa 480  
 aagtcatctg gaaacccaag tcatatggaa cagtgtctgg agcccaaaag attgaagctg 540  
 taaaaacacc aatgaacaa aaaagtgtc ttttgagcaa attatttaag ggtagtttat 600  
 50 tggaaaattt tactgtagat aactccacat tctcaaaagc ccaaataagg gccactttct 660  
 acccaaaatt tgagaatgag aatcagatc aggagattag gacaaggatg atagagatgg 720  
 55 tctccaaagg cttggctaca gtggaggtct ctcttaagca ttctggatct ctttttatgt 780  
 atgctgggca taagggtgga gcatatgcca agaatagctt tggaatatc tatactgccg 840

EP 3 822 353 A1

ttggtgtcctt tgttcttggg cggatgtttc gtgagacatg ggtactcaa gcaagtaaga 900  
 agcaagcaga gttcaatgag tttcttgagc gcaatcgcat gtgtatatca atggagtgg 960  
 5 tcacggcagt gttgggggac cacggacaac gccacgaga tgattatgcg gttgtgactg 1020  
 cagtcacgga gttgggaagt ggaaaaccaa atttctattc aactcccgat gtaattgctt 1080  
 tttgcagaga atggagatta ccaacaaatc atatatggct gttctcaaca aggaaatcag 1140  
 10 ttacttcctt ctttgctgcg tttgatgcac tctgtgagga aggtacagca acctctgttt 1200  
 gccaggctct tgctgaagtt gctgacattt ctgtacctgg atcaaaagac catataaaag 1260  
 tgcagggtga aatthttggag ggccctcgtgg cacgcattgt aaaacgtgag agctcagagc 1320  
 15 atatggagcg ggttctaaga gatthttcctc ctctccatt agaggggtgag ggthttggact 1380  
 tgggacctac gctgctgag gthttgtgctg caaatagatc agaaaagcag caaatcaagg 1440  
 cgcttcttca gagtgctggc acggctthtct gccgaatta thttggactgg thttggagatg 1500  
 20 atgattctgg ctcacattca agaaatgctg atcgctctgt tgtgtcaaag thcttacagt 1560  
 cacatcctgc tgatthttct acaggaaaat tacaggaaat ggttcgthttg atgagggaaa 1620  
 aacgctthtc tgctgctthc aagtgttact ataactthca thaaatthaat gattthtcga 1680  
 25 gtgacaacct gcctthcaaa atggtgatcc acgtacatag tgattcaggc thccgccggt 1740  
 accagaaaaga gatgagacac aaaccaggac tgtggcctth gtatcgaggc ththttgttg 1800  
 30 acctggatth attcaaggtc aatgagaaga aaactgccga aatggttagga agtagcaatc 1860  
 aatggtaaa aatgaggaa gaggatagcc gthtagctga tgaagatgca aatthgatgg 1920  
 tcaagatgaa atctctctc tacaagttga gaactththtt gatccgtaat ggcttgtcga 1980  
 35 ctctththca agaaggccct tctgctgata aggtthatta cctgaggcaa atgaaaatat 2040  
 ggaatactthc agcagccaag caacgagaac tcagcaagat gcttgatgaa tgggcagtat 2100  
 atatacgag thaatatggg aacaaacct tgtcatcatc cacataccta agtgaagctg 2160  
 40 agccctctct tgagcaatat gcaaagtgta gtccacaaaa tcaggcgtht ataggatctg 2220  
 ctggaaatth tgtcaaagtt gaagactthca tggctattht tgaaggagaa gatgtagagg 2280  
 gtgatcttga gccgaccaa gatattgctc ctthcaagccc thaatththc agcaaagaca 2340  
 45 tggctgcaaa gaatgagggt ctcatcgtht thththccagg aataccagggt tgtgctaaat 2400  
 ctgcacttht taaggaaata ctgaatgctc caggagggt tgaagatgat cggccaatthc 2460  
 atagththaat ggggtgatct atcaaaggta gatattggca thaaagthgct gatgaagctc 2520  
 50 gaagaaaacc thactcgatc atgcttgctg acaagaatgc accaaatgag gaagtatgga 2580  
 aacaaatthga gaacatgtht ctaagcacca aagcatctgc ththccagth atacctgatt 2640  
 55 cagaaggaa tgaaatcaat ctattctcta thgatgcact ggcggtthth atathccgag 2700  
 tacttcagcg ththaatcat ccgggaaatc thgacaagtc atctthcaat gctggatthg 2760

EP 3 822 353 A1

	tgatgttaat gttttatcac ctttatgatg gaaagagccg tcaggagttc gagagtgaac	2820
	ttattgaacg ttttgatca cttgtcagaa ttctctact gaaacctgaa aggtctcctc	2880
5	ttccggattc tgtgaggtct attggtgagg agggaatcaa tctgtacagg cttcatacta	2940
	acaaacatgg aagattggag tctacaaaag ggacttttgt aaaagagtgg gttaaatggg	3000
	agaagcaatt gaggatatt ctacatggaa atgcggacta tctcaattca atacaggttc	3060
10	catttgaatt tgctgttaaa aaagtccttg aacaactgaa agctattgog aggggcgaat	3120
	atgcagctcc ttcttctgag aagaggaagc taggatccat tgtgtttgcc gctatcagcc	3180
	tgccagttcc agaaattcta ggtcttctga atgatctagc caagaaagat ctgaaggttg	3240
15	gtgatttctt gaaggacaaa agcctggaga gctgcattca gaaagcccat ttgaccctag	3300
	ctcacaagag aagtcattgg gtcactgcag ttgccaatta cggttccttt cttcaccaaa	3360
	acgtcccagt agacgtggct gctttgttgt tctccgataa attggctgcc ctagaagctg	3420
20	agcctggttc tgtcgaaggt gaaaaggtcg attctaaaaa cccatggccc cacgtcacia	3480
	tatggactgg ggcaggagct acagccaaag atgccaatc actaccacag ttactttctc	3540
	aaggggaaggc tatccgcatt gatattaatc caccogtcac tataactggc actctcgaat	3600
	tcttttgaac tctacatgta attcctcatc ttgtgtggat ttacctaag taaaccctac	3660
	atgggttctg agtgaatatc atgtaatgac actcccaggt tgtaggaaca atttttagg	3720
30	atcgtgcgct ttggctatat agcagtcctg cagttgcagg atagaaatag atttagagaa	3780
	aatgtaaaga tgttgaggca ataaaggtct taagtcgtgg gcgttttgaa atccccgatg	3840
	gcatttcttg aataaagcac accattttct accctt	3876
35		
	<210> 20	
	<211> 3860	
	<212> DNA	
	<213> Hevea brasiliensis	
40		
	<400> 20	
	gcctgctaac tccttctccc ctttttttct cctactggca gttagtagag acattgaacc	60
	ccaaaacctc tccagaaacc tccacctgca actttttgcg ctactgagtc acccactctc	120
45	tctctctccc tctctctctc tctctcctc tctctccggt ctttcggtta ctatttatag	180
	ccattgccac accaactatg cgaagcatcc tcttcacact ttctctgact ctcccgtttt	240
	ctcctccttc cctaaacctc aaatctaaaa gcttcatttt ccttcgatct tctccttctc	300
50	ttttagccct acctcgctct tcttctcttt ccatttctca tctaatcatg tctcagagaa	360
	agggcgggct tgaaggacaa cagtggaaac cgaatccaac cacggaccag ccgtctccct	420
	ccggtctggg tgcacggtt gcggcgggta ccgatagaat tgggtggttg agtatagccg	480
55	aaagcagtg cagagtcatt gttgtggagc cattgcccgt gccgtttggt aatgcgcaag	540

EP 3 822 353 A1

tggttaatca gtgtaacttg caaggtcaaa aggctatatg gaagcctatg tcctatggaa 600  
 cagtaagtgg agcttcaact gtggaagttg aaaatgtgcc ttctaatagga atgctggttg 660  
 5 atgtccagag cagtgcgtct ggtacagctt cgggtggcagc tcggaagagt agtgттаact 720  
 tgagtcaatt ctttaggggt aatctgttag aaaattttgc tgtggataat togacttatt 780  
 cacaagccca aattagagcc actttctatc ccaaatttga aatgagaag tctgatcagg 840  
 10 agattaggat cagaatgata gagatggtat ctaaagggct ggctacattg gaggtcacac 900  
 tcaagcattc tgggtctctt tttatgtatg cgggtcataa ggggtggagca tatgcaaaaa 960  
 atagcttcgg aatatatac actgctgttg gtgtctttgt tcttggtcga atgtttcatg 1020  
 15 aggcattggg aaccgcagct gcaaaaaagc aagcagagtt caatgagttt cttgagagaa 1080  
 accgaatgtg tatatcaatg gaattggtaa ctgcagttct aggagaccat gggcaacgac 1140  
 20 cccgagagga ttatgtggtg gtcacagcag tcacagagtt aggcaatgga aagccaaagt 1200  
 tctactcaac tcctgaattg attgcctttt gtcgaaaatg ggccttcca acaaatcatg 1260  
 tttggctgtt ctcaacgagg aatcagtaa cgtctttttt tgctgcctat gatgcactat 1320  
 25 gtgaagaagg aacagcaacc actgtatgta gggctcttga tgaagttgct gatatttctg 1380  
 tcccaggttc aaaagaccat ataaaggtgc aggggtgaaat cctggaggggt cttgttgctc 1440  
 ggattgtagg tcctgacagc tcaaaacaca tggaggatgt cttgagggaa catcctcctt 1500  
 30 caccagctga aagagcggac cttgatttgg gactaagtct aagagaaatt tgtgctgcta 1560  
 atagggctga tgaaaaacag caaataaaag cgctacttca gagcattggc tcttcttttt 1620  
 gccctgacaa ttcagactgg tttgggattg aagggtggtga taccattca agaaatgctg 1680  
 35 acagatctgt tgtttcaaaa tttttacaag ctcaacccac tgattattca accaagaagt 1740  
 tgcaggaat ggttcgttt ctgagagaaa gacgcttccc agctgctttt aaatgttacc 1800  
 40 ataacttcca aaaaattgat tctgtatcaa gtgacaacct tttctacaag atggttattc 1860  
 atgtccacag tgactctgga tttcggcgat atcaaaaaga gatgaggcac aagccaggggt 1920  
 tgtggccatt gtatcgaggt ttttttgttg acatcaattt attcaaggca aacaaggaga 1980  
 45 gagctgctga aattgcaaaa aataacaata acatggagga aactattaat ggcaatgatg 2040  
 ctgtatctgc aaaagatgggt atagctgacg aagatgcaaa tttaatgata aaattaaat 2100  
 ttcttacata taagttaaga acctttttga tccgcaatgg tttgtcgatt cttttcaagg 2160  
 50 atggtccatc tgcttacaag gcctactacc tgaggcagat gaaaatatgg ggcacatcag 2220  
 ctggaagca gagagaactc agcaagatgc ttgatgaatg ggctgtctac atccgaagaa 2280  
 55 aacatggaaa gaaacagttg tcatcctcaa tataccttag tgaagctgag ctttttcttg 2340  
 aacagtatgc aagtcgaagt ccagagaatc aggctctgat aggatctgct ggaagtttag 2400

EP 3 822 353 A1

ttagggctga agatttcttg gccattatag gagacagga tgaagaggt gatcttgaga 2460  
 cagagcgaga ggtggcggc ccgagcccta tgctctcagt caaggacact gttcaaaaga 2520  
 5 atgagggttt aattgtattt tttccaggaa tacctggttg tgctaagtct gctctttgca 2580  
 aggaattact aaatgccctt ggaggacttg gagatgatcg ccctgtccat agtttgatgg 2640  
 gggaccttat caaaggaaag tattggcaga aggttgctga ggagcgcagg agaaaacat 2700  
 10 attctataat gctggctgac aagaatgcac caaatgaaga agtttgaga cagattgaag 2760  
 atatgtgccg tagcaccoga gcgtcagctg ttcctgttat acctgattct gaaggaacag 2820  
 attcaaatcc attttctctt gatgcattat cagttttcat tttccgtggt cttcaacgag 2880  
 15 ttaatcatcc gggaaatctg gacaaagcat ctccaaatgc tggttacgtg cttttaatgt 2940  
 tttatcatct ttatgatggc aagagtcaac aagagtttga gagtgaatta attgaacggt 3000  
 ttggctctct agtgaagatg ccattgctga aatcggatag gagtccctct ccagatcctg 3060  
 20 tgaggtcaat tttggaagag ggaataaatc tatacgggct tcacaccaac agccatggca 3120  
 gaatggaatc taccaagggg tcatatgcaa aggaatgggc aaattgggag aagcgggtgc 3180  
 25 gtgaagttct attcagcaat gcggagtacc tcaattccat tcaggttctt tttgagtctg 3240  
 cagttaagca agtgctggaa cagctaagaa aaattgctaa ggtgaaatac acaacaccaa 3300  
 ttattgagag gagaaactt ggaacaattg tttttgctgc tatcgatttg ccagttccag 3360  
 30 aaatctgtag ttccctcaac aatttggctc agaagaacct caaggttgaa gcctttcttc 3420  
 aggacaagca cttggagcac aaccttaaga agactcactt gacccttgcc cacaagagaa 3480  
 gtcatggcgt tacagcggta gcaagttatg gactgtttct caatcagaag gtcccagtag 3540  
 35 aattgactgc acttctcttc acagataaaa tggctgccct agaagcacia ccaggttctg 3600  
 ttgatggcga aaaagtagtt tcaaaaaatg agtggccgca tattaccata tggaccgag 3660  
 atggagtagc acccaaagaa gccaacacac tgctcaatt gttctcagag gggaaagcaa 3720  
 40 ctctgtttga attcagtccg ctcatcacca tatcgggcac agtgcaattt tattgagttc 3780  
 aathtttttt ttaacaaatt ggatttgaag cttagttttt tagctgctga gaagtaattt 3840  
 45 gtataggcga aaacaaggtg 3860

<210> 21  
 <211> 4470  
 <212> DNA  
 <213> Prunus avium

50  
 <400> 21  
 tggccttgcc ctttgcagtg gaagtggcct tagagctttg gttcatcggc gtagcggag 60  
 55 gtgaaaaata tgaaattatg gatttttata gaaggaagag tcggcgtagg tgtgatcggg 120  
 gagcgggcat gaaagccgca tcgacaaccg ccattgctat gcacgaagat atcttcgagg 180

EP 3 822 353 A1

cagcccatgg atatctggat aaagatttga gctttacttt gatcggattt ctaagttcga 240  
 taattcatgg gcgtaaattc aaatgtctca aatatctatg tgagcatgaa cctcttcagc 300  
 5 ggcaaccctt ctctccagt gaggtgggtt ggctgaccag aaggacaga tgtcaagaaa 360  
 gtagataggt acgagggagg tacaggcaaa aaggaaaaca ggaaaattgg aatcccgagc 420  
 acctctctct ccctgaacc cggtcgccgc ttgttctctac ttccaagggt tataacacac 480  
 10 aaaatcccag tgcagcaacc tccataacct gccgtgtgtc tcttccactt ggctactcat 540  
 ccacattctc tctctctctc tctctctcat ataaaaatac aagcgtagaa ggttgcacgt 600  
 cgaaagcagc gattgaatgt gggcgtcaca gaaaatcctc tatgctctcg ctctccctct 660  
 15 tgccttctcg tcttcgtcta cagccctcgt ttccggaacc ttctctctcc tcaggtcctc 720  
 ctctccttct tacgcataca cagccaaacc tcgctttctc tcgctctctt tctcctcttc 780  
 caattctccc tcgcattcga ccatgcctta caaccagaga agggcgggc ttagagaaca 840  
 20 aaagtggaca gaaaaacaaa agactgacag ccaaccaacc tccacagggg attcatcagc 900  
 agctgctgtg gaagttgcca ctaacagagt tggcggattg agtctcagtg gaagcagtg 960  
 gcagaccaat gcttcagtta aaccataca gtttgggaag gtccagtcgg taaatcaggg 1020  
 25 tctagtccaa ggtcagaaag gaatctggat tcctaaatca tattcgacag tcagtggagc 1080  
 taagacaata gaagttgaag caoctgttga taaaagtaca gttggaatct caggcaatgg 1140  
 30 agctggacaa gcagcagata aaaaaccag tgttggttta agtaagttgt tcaagggatga 1200  
 cctgttggaa aatcttactg tggacaactc cacatatgct caggtgcaag tgcgcgctac 1260  
 tttctatccc aaatttgaaa atgagaagtc agatcaggag ataaggacaa ggatgataga 1320  
 35 gatggtatct aatggcttgg ctaccttggg ggtttcactt aaacactctg ggtctctttt 1380  
 tatgtacgct ggtgataagg gtggggcata tgccaaaaac agctttggga atatatacac 1440  
 tgctgtaggt gtatctgttc ttggacggat gtttcggggag gcatggggac gtgaagcagc 1500  
 40 aaaaatgcaa gcagagttca atgattttct tgagaggaac cgtgtgtgca tatcaatgga 1560  
 actcgtaact gctgttctag gagaccacgg acagcgccct aaggaagatt tcgttgtagt 1620  
 gacagcagtc acagatttgg gcaatgggaa gccaaagtcc tatgcaacct ctgagataat 1680  
 45 tgctttctgt agaaaatggc gtctaccgac aaatcatgta tggctgttct caacaaggaa 1740  
 agcagtgaca tcactttttg ctgcctttga tgctctctgc gaagaaggaa ctgcaacacc 1800  
 tgtatgcata gctcttaatg aaattgcaga tatatccata cctggttcaa aagagcatgt 1860  
 50 aaggagcagc ggtgaaatct tagagggtat tgttgctcgt attgtaagcc aggagagctc 1920  
 aaaacacatg gagaaagtcc tgaatgattt ccctcctcca ccaatggatg gagttggtct 1980  
 55 tgatttggga cctagcgtaa gggagctttg tgctgcaaat aggtcgagtg aaaagcagca 2040  
 aataaaagcg attcttgagg gtgttggctc ctctttttgc cctgaccatt cagactggct 2100

EP 3 822 353 A1

	tggaactgga	gctggtgatg	cccattccag	aaatgcagat	aataaattgg	tgctatcaaa	2160
	acttttacia	tctcatgctg	ctgatttttc	taccaccaaa	ttgcaggaaa	tgattcgttt	2220
5	gatgaaagag	aaacgttata	ctgctgcctt	taaatgttac	tataactacc	acaaaattga	2280
	ttccatatca	agtgacaacc	ttttctataa	aatggtcgtc	catgtccata	gtgattccgc	2340
	atttcgacgt	taccagaaag	agatgaggtc	aaagccagga	ttgtggcctt	tatatcgagg	2400
10	tttttttgtt	gacattaatt	tattcaaggc	aagcaaggag	agagctgctg	agatagcaaa	2460
	agacaaaagt	agcattgtgg	aagacgttag	cagtgatatg	cctggaaaat	atggattggg	2520
15	tgatgaagat	gcaaacttaa	tgataaaatt	gaaatttctt	acataataagc	tgogaacctt	2580
	tttgattcgt	aatggtttgt	caattctttt	taaagaaggt	ccagctgctt	ataaagccta	2640
	ttacttgagg	caaatgaaag	tatggggaac	ttcagcagca	aagcagaggg	aactcagcaa	2700
20	gatgcttgat	gaatgggctg	tgtatatacg	aaggaaatgt	ggaaacaaac	agctgtcgtc	2760
	atcagtgtag	ctcagtgagg	ctgagccttt	tcttgaacag	tatgcaaaac	gcagtcacaca	2820
	aaatcaggct	ctcataggat	ctgctggaaa	cttagtaagg	actgaagatt	tcttggccat	2880
25	tgttgagggg	ggaagagacg	aagaaggtga	tcttgagagg	gaccgggagg	ttgcaccatc	2940
	aagccccagt	gcctcagcca	gggacacccat	tccaaaggct	gagggctctga	ttgtcttctt	3000
	tcctggacta	cctggttctg	ctaagtcagc	actatgcaag	gaactactaa	atgccccagg	3060
30	tggaatggga	gacgatcggc	caattcaaag	cctgatgggt	gaccttatca	aaggaagata	3120
	ttggcagaag	gttgctgatg	agcgcaggag	aaaaccatat	tcaataatgc	ttgcagacaa	3180
35	gaatgcacca	aatgaagaag	tatggagaca	gatagaagac	atgtgccaca	gcaccagggc	3240
	ctccgcagtt	ccagttgttc	ctgattctga	aggaaccgat	tcaaatccgt	tttgcgttga	3300
	tgcattagcg	gttttcatgt	ttcgggtgct	tcaacgtgct	aatcatccgg	gaaaccttga	3360
40	caaggaatct	ccaaatgctg	gctatgtgct	gctaattgttc	tatcaccttt	atgagggcaa	3420
	gagtcgcccg	gagtttgacg	gcaattagt	agagcgtttt	ggctcacttg	tcaagatgcc	3480
	attgctgaaa	tctgatagga	atccattgcc	cgatcctgtg	aagtctattc	tgaggagggg	3540
45	aataaatcta	tacaagcttc	acactgccaa	acatggaaga	gtggagtcaa	caaagggaac	3600
	ctatgcaaag	gagtgggcta	aatgggagaa	acagttgctg	gatattttat	ttggcaatgc	3660
	agagtatctt	aattctgttc	aggttccctt	tgagtctgct	gttaaggacg	tatcagaaca	3720
50	gctgagaaaa	attgcacaag	gtgagtacaa	aactccagat	actgggaaga	agaagtttgg	3780
	ggccattgtt	tttgctgccg	ttagcttgcc	tgttacggaa	atcagtgatc	tgcttgataa	3840
55	tttgctgctg	aaaaatagtg	aagctggagc	tttccttaaa	gacaagcatt	tgggagaatct	3900
	taataaggct	catgttactc	ttgcacataa	gagaagtcac	ggtgttactg	ccgtggctag	3960



EP 3 822 353 A1

	ttatggcacc	ttcctccaca	aaacggcccc	tgtggatcta	actaagctgt	tcttctcgga	4020
	taaaatggct	gcacttgaag	ccagccttgg	ctctgttgaa	ggagaaagag	tagtatcgaa	4080
5	aatgagtg	ccacatgtca	ccttatggac	cgcagaaggg	gttgcagcaa	aggaggccaa	4140
	caagttaccc	cagttgcatt	cagagggaaa	ggctacctgt	attgcaattg	accacactgc	4200
	caccattgat	ggcacgctgg	aatTTTTctg	aggTTTgtat	tcattTTtaga	ggcatgatgg	4260
10	accagaccag	gccaggccag	attgacctcg	tacaatttcc	tttgatgatg	ccagtgttct	4320
	atgaagagat	agcttcaactg	ctcaaataaa	ggagaaaatt	ttgtatgtca	ggagaaggaa	4380
	atTTTgacga	gTTgattcta	ttgagaagtt	ttgtgtatat	aggcaattga	agTTTTgaaa	4440
15	tagatTTTgg	gaaaaaaaaa	atTTgaaata				4470
	<210>	22					
	<211>	4196					
20	<212>	DNA					
	<213>	Malus domestica					
	<400>	22					
	atattgtgta	tcgcggaaca	ccgaaactca	ttgcgcctct	cccctgaacc	cggtcgctgc	60
25	tacctagttc	caaggatTTT	acacccaaac	ctcagagcag	cccaggtacg	gacacgccgc	120
	aaactccaga	agttgtctct	cttccacttg	gcggtcatc	agccacatcc	gcattcgttc	180
	tctgggaaac	tctctctctc	tctctcatat	aaatatagaa	gcgtagtggg	ggcacgtcca	240
30	gagtagcaat	tatatgtggg	cgtcgcagag	aatcatctat	gctctcgctc	tcgctcatcc	300
	cctctcgtct	tcgtcagcct	tcgTTTccgg	aatTTTctct	ttcctccggg	cctcctctcc	360
	ttcctccgcc	catcctcgct	ttctccctcc	ctcgctttct	ctctcctctt	ccatctcgct	420
35	ccccattcc	agcatgcctt	acaaccagca	aaggggtggg	cgtacagAAC	aaaagtggac	480
	acaaaagcaa	aagactgaca	gcccgtcaat	gtccactggg	gattcttcag	cagctgtcac	540
40	taacagagtt	ggtgagttga	gtcttactgg	aagcaatggg	caggccaatg	cttcagttga	600
	accacacacag	tttgggaagg	tccaattggc	aaatcagggg	ccagttcaag	ggcagaaagg	660
	aatttggatg	cctaaatcat	atgtgacagt	gagcggagct	aaggcggtag	aagttgaagc	720
45	aatgTTgat	gaaagtacag	ttggaaccca	aggaaatgga	gctggacaag	cagcagctca	780
	aaaaaccaga	gTTgTTtaaa	gtaagctgTT	caatggtaat	ctgTTggaaa	atTTTactgt	840
	ggacaactcc	acttatgctc	aggtgcaaat	acgagctact	ttctatccca	aatttgaaaa	900
50	tgagaagaca	gatggggaga	ttaggacaag	gatgatagag	atggtatcta	acggcttggc	960
	taccttggag	gTTTcactta	aaactcagg	gtctctTTTT	atgtacccg	gtaataaggg	1020
	tggggcatat	gccccaaaca	gctTTgggaa	tatatatacg	gctgtagggtg	tatttgttct	1080
55	tggacgggatg	tttcgggagg	catggggacg	tgaagcacca	aaaatgcaag	aagagTTtaa	1140

EP 3 822 353 A1

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45  
 50  
 55

tgatttcctt gagaggaatc gtatgtgcat atcaatggaa cttgtaactg ctgttctagg 1200  
 agaccatgga caacgcccga aagaagattt tgctgtagtg acggcagtca cagatttggg 1260  
 caatggaaaag ccaaagttct actcaacccc tgagataatt gctttctgta gaaaatggcg 1320  
 cctaccaaca aatcatgtat ggctgttctc aacaaggaaa tcagtgacat cttttttg 1380  
 tgcctttgat gcactctgcg aagaaggaac tgcaacacct gtatgcatgg ctcttaatga 1440  
 agttgcagat atatctatac caggttcaaa agatcatgta aaggagcagg gtgaaatctt 1500  
 agagggcatt gttgctcgt a ttgtaagcca ggaaagctca gaacatgtgg aaaaagtctt 1560  
 gaaagatttc cctcctccac caatcgatgg agctgggtctt gatttgggac cgagcgcacg 1620  
 agatatttgt gctgcaaata ggtcgagtga aaagcagcaa atagaagcgc ttcttgaggg 1680  
 tgttggtccc tctttttgcc ctgaccattc ggactggttt ggagttggag ctggtgatga 1740  
 ccattccaga aacgcagata aatctgtgct atcaaaactt ttgcaatctc atcctgcaga 1800  
 tttttctacc accaaattgc aggaaatgat tccggttaatg aaagaaaagc gttttcctgc 1860  
 tgcctttaa tgttaccatg actaccacaa agttgattcc atatcaagt acaacctttt 1920  
 ctataaaatg gtcattccatg tccatagtga ctccgcattt aggcgttacc agaaagagat 1980  
 gaggtctaag ccaggactgt ggcctttata tccgagtttt tttggtgaca ttaatttatt 2040  
 caaggcaagc aaggagagag ctgctgagat agcaaaaggc aaaagcagca ttgtggaaga 2100  
 tgggaacaat ggtatgcctg gaaaatatgg attggctgat gaagatgcaa acttaatgat 2160  
 aaaactgaaa tttcttacat acaagctgcg aaccttttg attcgcgatg gtttgtcaat 2220  
 tctgtttaa gaaggtacag ctgcttacia gacctattac gagcggcaaa tgaaaatatg 2280  
 gaatacttca gaaccaaaaga agaggggaact cagaaagatg cttgatgaat gggctgtgta 2340  
 catacgaagg aatgtggaa acaaaccgct tccatcatca gtttacctca gtgaggctga 2400  
 gccttttctt gaacagtatg caaaacgcag tccacagaat cagtctctca ttggatctgc 2460  
 tggaaaacttt gtaagggctg aagatttcct gccattggt gagggaggaa gagatgaaga 2520  
 ggggtgatctt gagaggggagc aggaggttgc accatcaagc cccagtgcct caggcagggg 2580  
 cactatccca aaggctgagg gtctgattgt attctttcct ggactacctg gttctgctaa 2640  
 gtcagcatta tgcaaggaac tacttcaaga cccaggaaga atgggagatg atcgaccaat 2700  
 tcaaagcctg atgggtgacc ttatcaaagg aaaatattgg cagaagggtg ctgaagagcg 2760  
 caagaaaaaa ccatattcca taatgcttgc agacaagaat gcaccaaagc aagaagtatg 2820  
 gagacagatt gaacacatgt gcaacagcac tagggcctcc gcggttccag tctgtcctga 2880  
 ttctgaagga accgattcaa atccatttct tcttgatgca ttagcaattt tcatgtttcg 2940  
 ggtgcttcaa cgtgctaate atccaggaag ccttgacaag gaatctttaa atgctggcta 3000  
 tgtgctgcta atgttctatc acctttacga gggcaagagt cgcagagagt ttgatgggga 3060

EP 3 822 353 A1

attagttgag cgttttggct cacttgtaaa gatgccgttg ctgaagtctg ataggaatcc 3120  
 attacctgat cccgtgaaga atattctgga ggaggggaata aatctataca agcttcacac 3180  
 5 cgccaaacat ggaagattgg agtcgaccaa ggaacctat gcaaaggagt gggctaaatg 3240  
 ggagaaacag ttgcgggata ttttatttgg caatgctgag tatctgaatt ctgttcaggt 3300  
 tccttttgag tctgctgtaa aggaagtgtc agaacagctg aaaaaggtcg caaaaggtga 3360  
 10 gtacaaaact ccagatactg ggaagaggaa gttcggcgcc attgtctttg ctgccgttag 3420  
 cctgcctggt atggaagtca gtggtctgct tgatagcttg gctgtgaaac atagtgaagc 3480  
 cgaagctttc ctcaaagaca agcatttggg taaacttaac aaggctcatg ttactcttgc 3540  
 15 acacaagaga agtcatggtg ttactgccgt ggctagttac ggaaccttcc ttcacaaaac 3600  
 ggtccctgtg gatataacta agcttttctt ctccgataaa atggctgcat ttggagccag 3660  
 ccttgggtct gtcgaaggtg aaagagtagt gtctaaaaat gagtggccgc atgtcacctt 3720  
 20 atggactgca gaaggggttc cggccaagga cgccaacaac ttacccaat tgcattcggg 3780  
 ggggaaagct acctgcattg caattgacct acctgccacc attgatggca cgctggaatt 3840  
 25 tttctgaggt ttgatcggcc aggccagatt gagctcggca tacgattgga gtgaaatttc 3900  
 atttgacgat gcgggtgttc aatgaagaga gaacttaact gcttaagtga aggagaaaat 3960  
 tttgtaagcc atgaggaaaa ggaaattttg ttgagttgat ttcttttcag aggttttgta 4020  
 30 tacataaggc aattaaaatt ttaaagtctc agcgttgctt catcagcttt taccaattga 4080  
 ggggtgttta gccgcggtog catcgctcgt atctgtttac acgattcggg atgattgcgc 4140  
 agctgcaact cccccaagtt ttatgacact tgaaatggtg aactgcatt ttttta 4196  
 35  
 <210> 23  
 <211> 517  
 <212> DNA  
 <213> Tetrahymena termophila  
 40  
 <400> 23  
 tgacgcaatt caaccaagcg cgggtaaacy gcgggagtaa ctatgactct ctaaatagca 60  
 atatttacct ttggagggaa aagttatcag gcatgcacct ggtagctagt ctttaaacca 120  
 45 atagattgca tcggtttaaaggcaagacc gtcaaattgc gggaaagggg tcaacagccg 180  
 ttcagtacca agtctcaggg gaaactttga gatggccttg caaagggat ggtaataagc 240  
 tgacggacat ggtcctaacc acgcagccaa gtccctaagtc aacagatctt ctgttgatat 300  
 50 ggatgcagtt cacagactaa atgtcggtcg gggaagatgt attcttctca taagatatag 360  
 tcggacctct ccttaatggg agctagcggg tgaagtgatg caaactgga gccgctggga 420  
 55 actaatttgt atgcgaaagt atattgatta gttttggagt actcgtaagg tagccaaatg 480  
 cctcgtcatc taattagtga cgcgcatgaa tggatta 517

<210> 24  
 <211> 174  
 <212> DNA  
 <213> *Ceratitidis capitata*

5

<400> 24  
 gatgacatgg agaagatctg gcaccacacc ttctacaacg agttgcgtgt cgccccgag 60  
 gagcaccgacg ttctgctgac cgaagcaccg ttgaaccgga aggctaaccg cgagaagatg 120  
 acccagatca tgttcgagac cttcaactcg ccggccatgt atgtggccat ccag 174

10

Claims

15

1. An isolated nucleotide sequence comprising
  - a) the cDNAs of two strands of a target dsRNA separated by an autocatalytic intron flanked by exons or exon fragments, and
  - b) a plant viroid sequence,

wherein element (a) is inserted in the plant viroid sequence.
2. Isolated nucleotide sequence according to claim 1, further comprising permuted autocatalytic intron-exon sequences between the plant viroid sequence and the cDNAs sequences.
3. Isolated nucleotide sequence according to claim 2, wherein the permuted autocatalytic intron-exon sequences derive from type I introns or the intron type I of the rRNA 26S from *Tetrahymena termophila*.
4. Isolated nucleotide sequence according to anyone of claims 1 to 3, wherein the autocatalytic intron is an autocatalytic intron type I or an autocatalytic intron type II or an autocatalytic intron type III.
5. Isolated nucleotide sequence according to claim 4, wherein the autocatalytic intron type I is the autocatalytic intron type I of the rRNA 26S from *Tetrahymena termophila*.
6. Isolated nucleotide sequence according to anyone of claims 1 to 5, wherein the plant viroid sequence is the *Eggplant latent viroid* (ELVd) sequence, the *Avocado sunblotch viroid* (ASBVd) sequence, the *Peach latent mosaic viroid* (PLMVd) sequence, or the *Chrysanthemum chlorotic mottle viroid* (CChMVd) sequence.
7. Isolated nucleotide sequence according to anyone of claims 1 to 6, wherein the target RNA is a gene from a plant pathogen or pest, or the  $\beta$ -actin gene from *Ceratitidis capitata*.
8. Isolated nucleotide sequence according to anyone of claims 1 to 7, further comprising a nucleotide sequence encoding a tRNA ligase.
9. Isolated nucleotide sequence according to claim 8, wherein the tRNA ligase is the tRNA ligase from eggplant.
10. A vector comprising an isolated nucleotide sequence according to anyone of claims 1 to 9.
11. A host cell comprising an isolated nucleotide sequence according to anyone of claims 1 to 9 or a vector according to claim 10.
12. Use of an isolated nucleotide sequence according to anyone of claims 1 to 9, or a vector according to claim 10, or a host cell according to claim 11, for producing dsRNA specific for a target gene.
13. A method for producing dsRNA specific for a target gene comprising co-expressing in a host cell
  - c) a nucleotide sequence according to anyone of claims 1 to 7 and

45

50

55

d) a nucleotide sequence encoding a tRNA ligase,

wherein the nucleotide sequences of a) and b) are in the same or different nucleic acid molecule.

5 **14.** A method according to claim 13, wherein the host cell is *E. coli*.

**15.** A method according to claim 13 or 14, wherein the tRNA ligase is the tRNA ligase from eggplant.

10

15

20

25

30

35

40

45

50

55

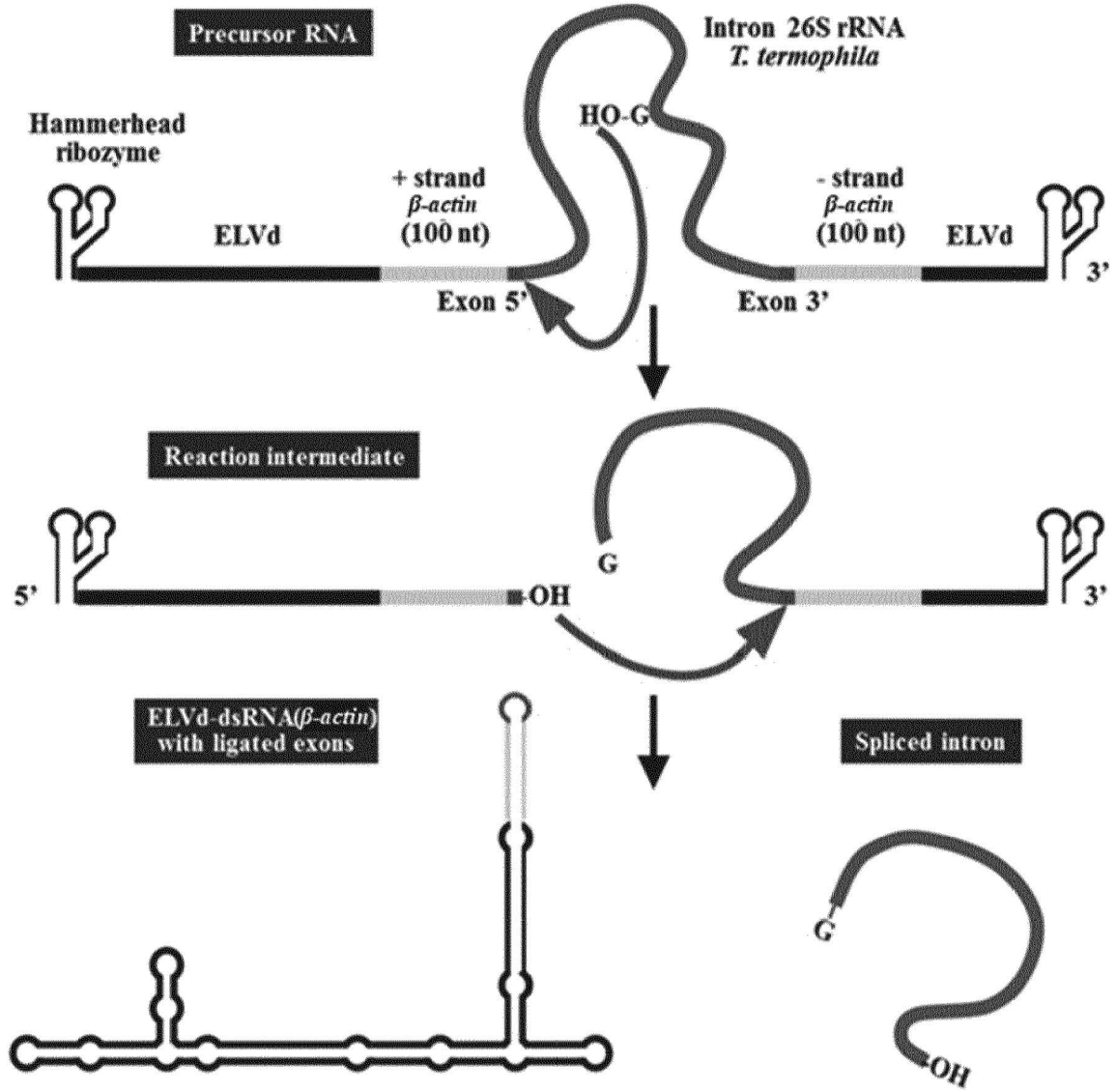


FIG. 1

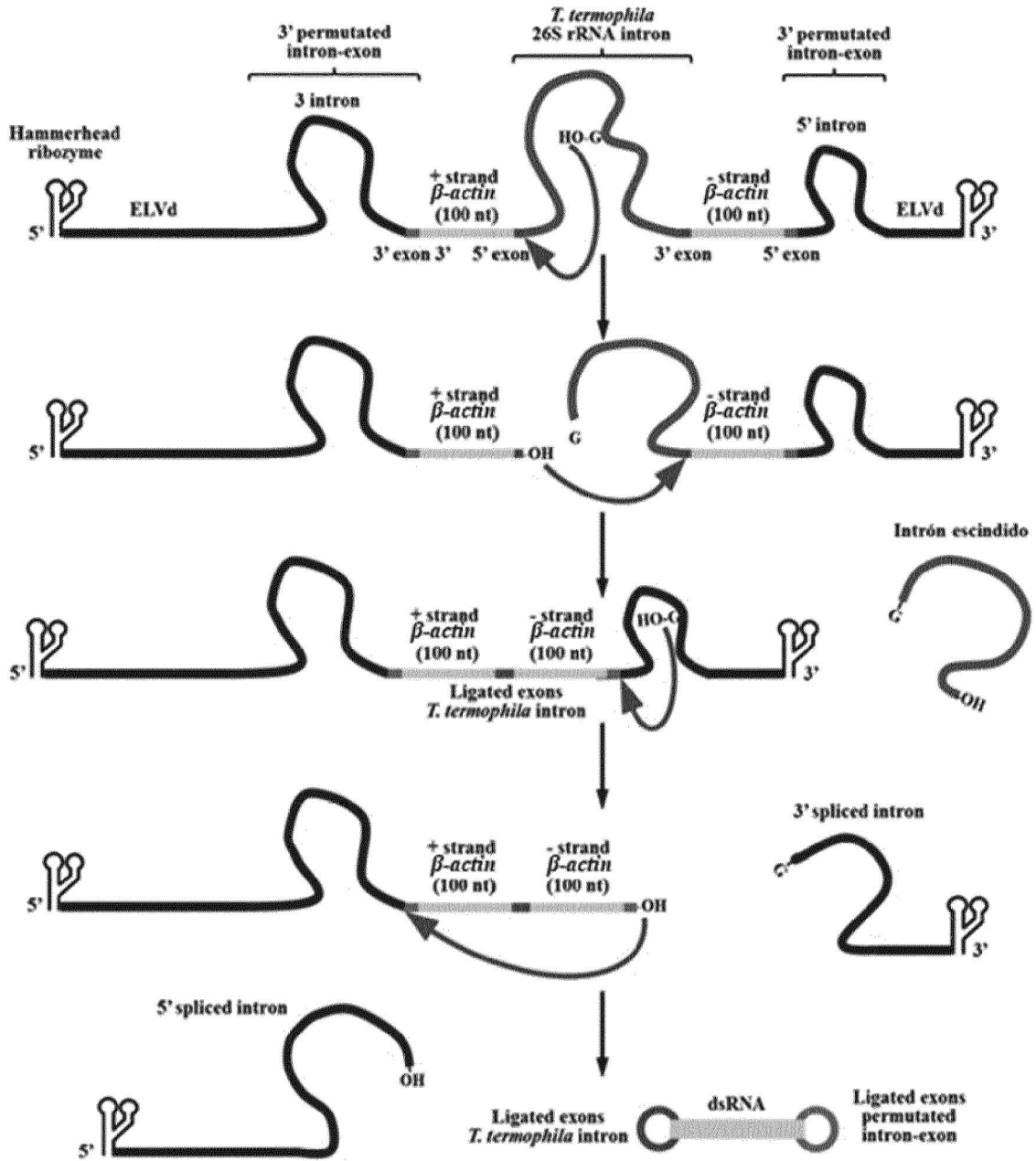
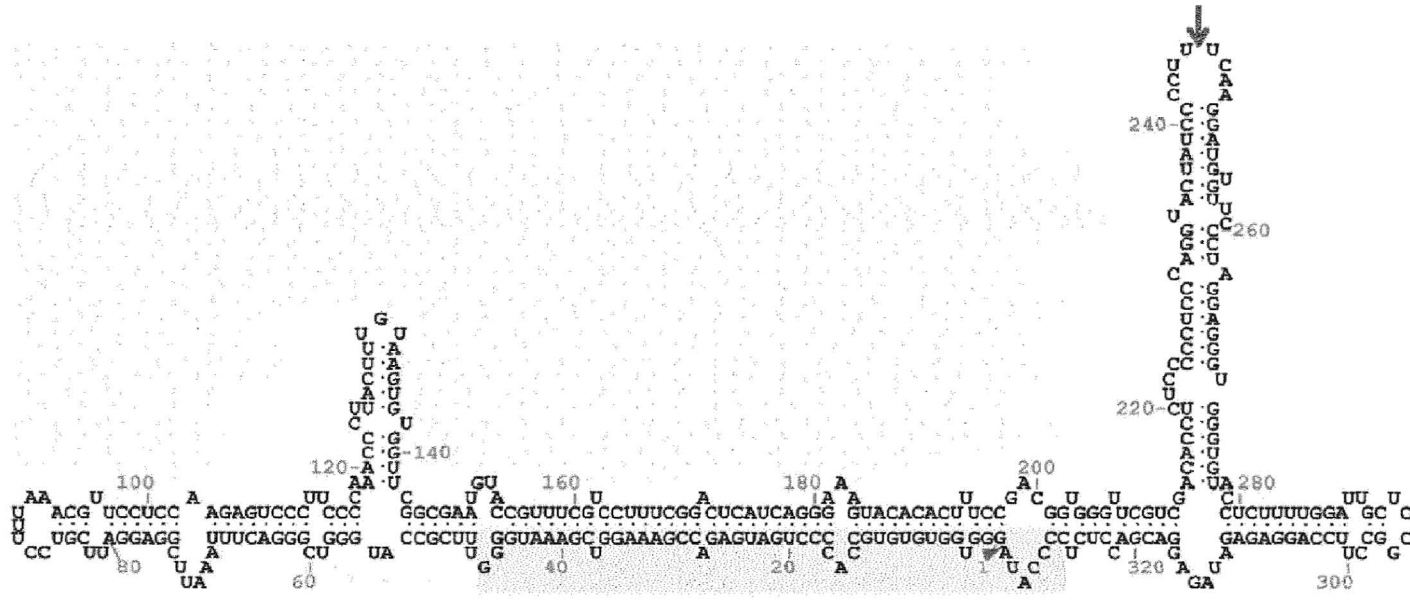


FIG. 2



FIG. 3

FIG. 4



SEQ ID NO: 7



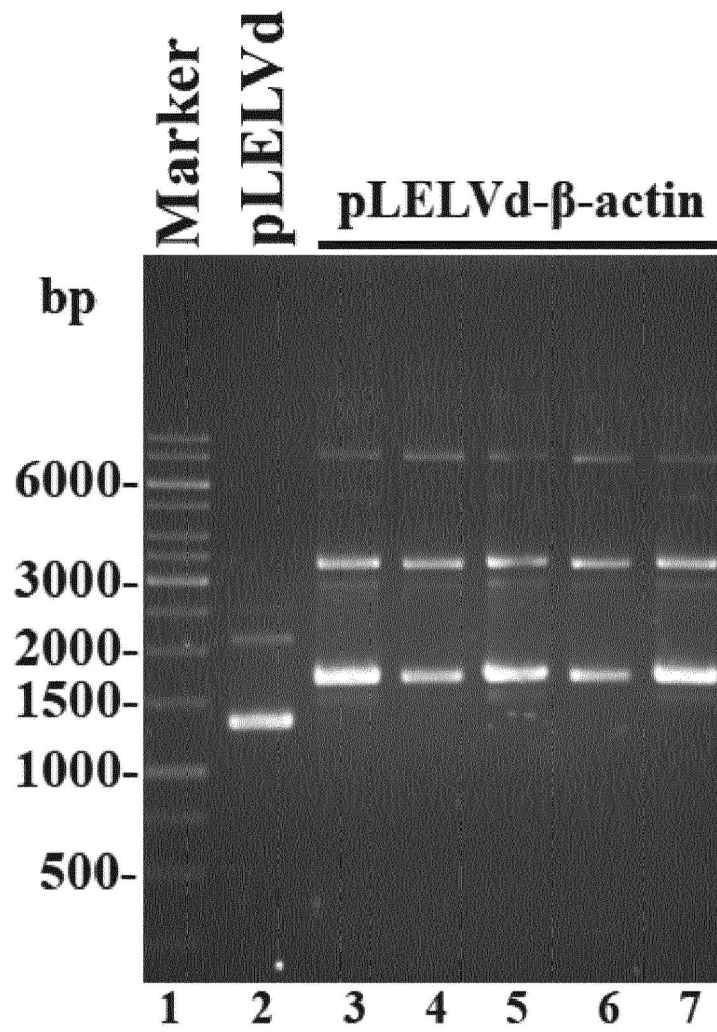


FIG. 5

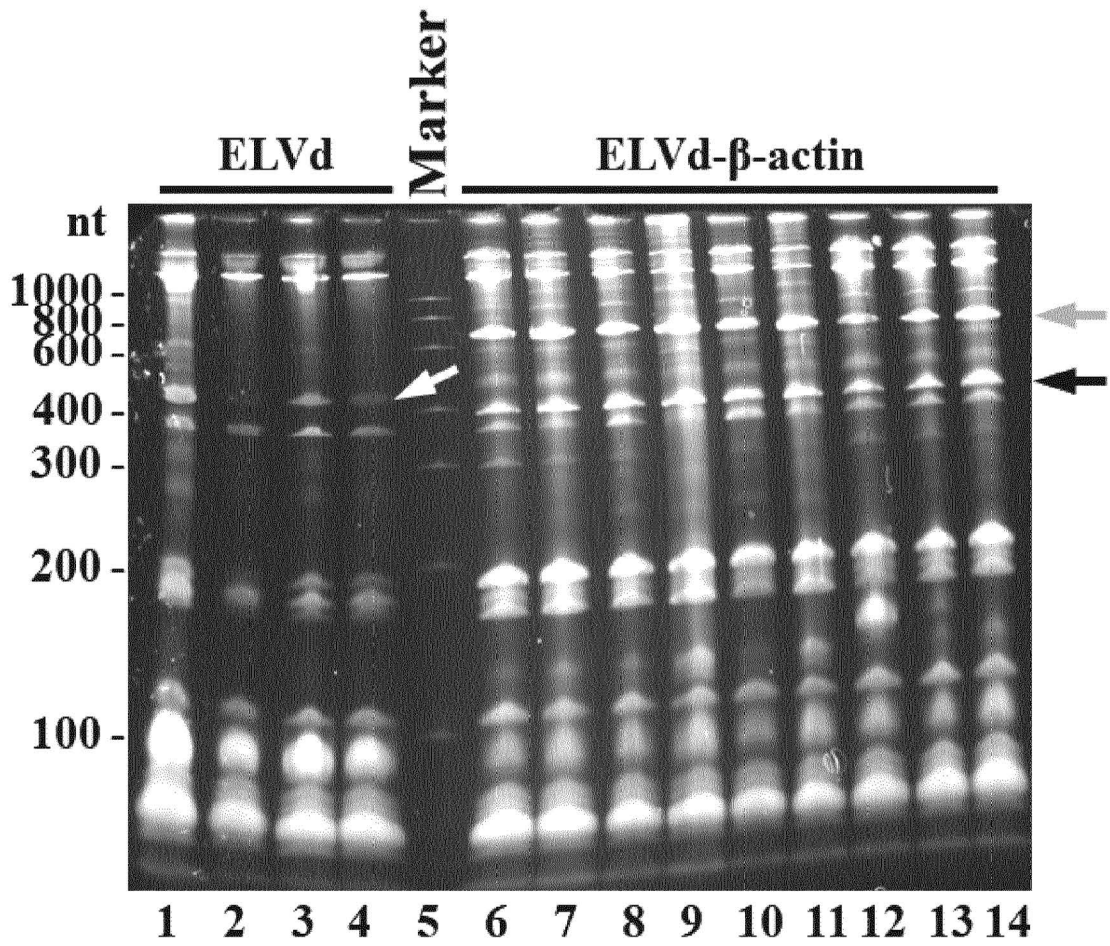


FIG. 6

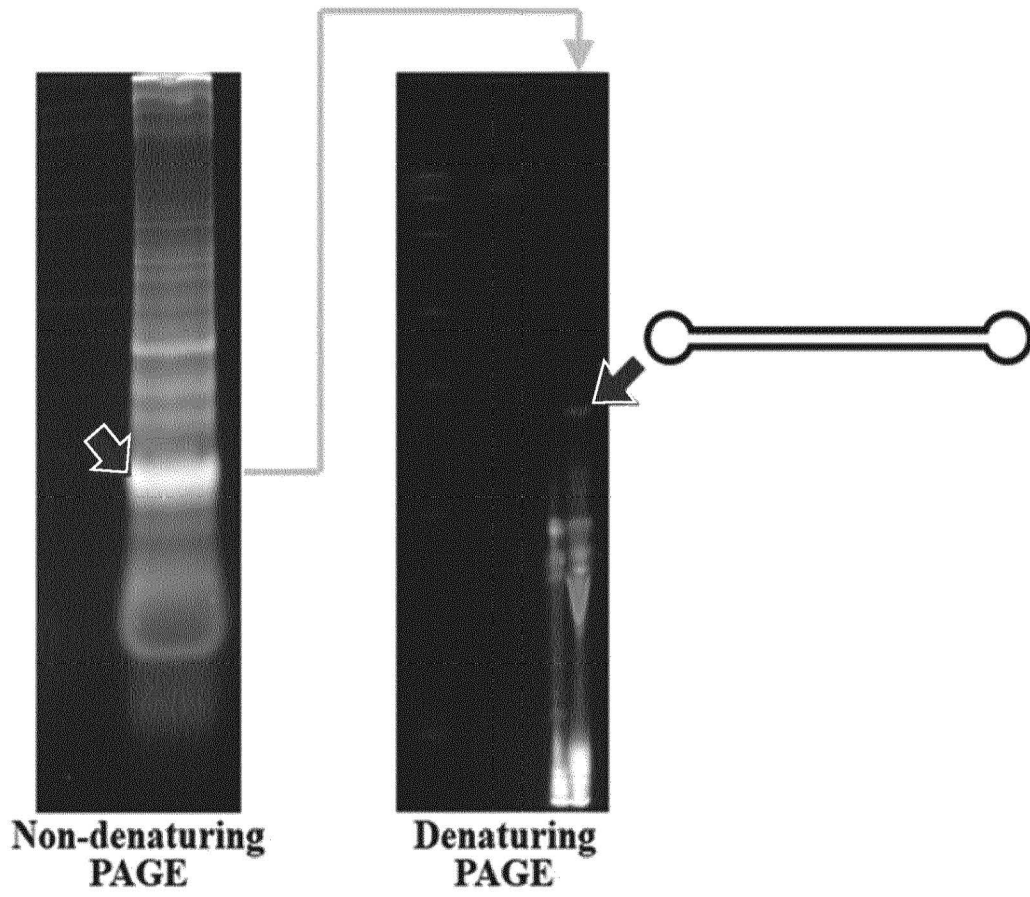


FIG. 7



EUROPEAN SEARCH REPORT

Application Number  
EP 19 38 2997

5

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
Y	JOSÉ-ANTONIO DARÒS ET AL: "A viroid-derived system to produce large amounts of recombinant RNA in Escherichia coli", SCIENTIFIC REPORTS, vol. 8, no. 1, 30 January 2018 (2018-01-30), XP055685866, DOI: 10.1038/s41598-018-20314-3 * the whole document *	1-15	INV. C12N15/67
Y	TERESA CORDERO ET AL: "Mutational Analysis of Eggplant Latent Viroid RNA Circularization by the Eggplant tRNA Ligase in Escherichia coli", FRONTIERS IN MICROBIOLOGY, vol. 9, 5 April 2018 (2018-04-05), XP55685893, ISSN: 1664-302X, DOI: 10.3389/fmicb.2018.00635 * the whole document *	1-15	
Y	FENG GUO ET AL: "In vivo selection of better self-splicing introns in Escherichia coli: The role of the P1 extension helix of the Tetrahymena intron", RNA, vol. 8, no. 5, 1 May 2002 (2002-05-01), pages 647-658, XP55685980, US ISSN: 1355-8382, DOI: 10.1017/S1355838202029011 * page 650; figure 2a *	1-15	TECHNICAL FIELDS SEARCHED (IPC) C12N
The present search report has been drawn up for all claims			
Place of search Munich		Date of completion of the search 16 April 2020	Examiner Young, Craig
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document		T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document	

EPO FORM 1503 03.82 (P04C01)

10

15

20

25

30

35

40

45

50

55



EUROPEAN SEARCH REPORT

Application Number  
EP 19 38 2997

5

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
Y	J. ROMAN ET AL: "Integration of the Tetrahymena group I intron into bacterial rRNA by reverse splicing in vivo", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, vol. 95, no. 5, 3 March 1998 (1998-03-03), pages 2134-2139, XP055685989, US ISSN: 0027-8424, DOI: 10.1073/pnas.95.5.2134 * page 2315; figure 1 *	1-15	
A	G. DINTER-GOTTLIEB: "Viroids and virusoids are related to group I introns.", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, vol. 83, no. 17, 1 September 1986 (1986-09-01), pages 6250-6254, XP055685971, US ISSN: 0027-8424, DOI: 10.1073/pnas.83.17.6250 * the whole document *	1-15	TECHNICAL FIELDS SEARCHED (IPC)
The present search report has been drawn up for all claims			
Place of search <b>Munich</b>		Date of completion of the search <b>16 April 2020</b>	Examiner <b>Young, Craig</b>
<b>CATEGORY OF CITED DOCUMENTS</b> X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document		T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document	

10

15

20

25

30

35

40

45

1

EPO FORM 1503 03.82 (P04C01)

50

55

**REFERENCES CITED IN THE DESCRIPTION**

*This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.*

**Patent documents cited in the description**

- WO 2015177100 A [0005]

**Non-patent literature cited in the description**

- DAROS et al. *Sci. Rep.*, 2018, vol. 8, 1904 [0005]
- ALTSCHUL S.F. et al. Basic local alignment search tool. *J Mol Biol.*, 05 October 1990, vol. 215 (3), 403-10 [0018]
- DARÒS, J.A. ; ARAGONÉS, V. ; CORDERO, T. *Sci. Rep.*, 2018, vol. 8, 1904 [0081]
- TIMMONS, L. ; COURT, D.L. ; FIRE, A. *Gene*, 2001, vol. 263, 103-112 [0081]
- FADDA, Z. ; DARÒS, J.A. ; FAGOAGA, C. ; FLORES, R. ; DURAN-VILA, N. *J. Virol.*, 2003, 77 [0081]
- PRICE, J. V ; ENGBERG, J. ; CECH, T.R. *J. Mol. Biol.*, 1987, vol. 196, 49-60 [0083]
- PUTTARAJU, M. ; BEEN, M. *Nucleic Acids Res.*, 1992, vol. 20, 5357-5364 [0083]