



US009388428B2

(12) **United States Patent**  
**Nasar et al.**

(10) **Patent No.:** **US 9,388,428 B2**  
(45) **Date of Patent:** **Jul. 12, 2016**

(54) **COMPOSITIONS AND METHODS RELATED TO VIRUSES OF THE GENUS *NEGEVIRUS***

(71) Applicants: **Farooq Nasar**, Albany, NY (US);  
**Rodion V. Gorchakov**, Houston, TX (US); **Andrew D. Haddow**, Springfield, MO (US); **Robert B. Tesh**, Galveston, TX (US); **Hilda Guzman**, Galveston, TX (US); **Scott C. Weaver**, Galveston, TX (US)

(72) Inventors: **Farooq Nasar**, Albany, NY (US);  
**Rodion V. Gorchakov**, Houston, TX (US); **Andrew D. Haddow**, Springfield, MO (US); **Robert B. Tesh**, Galveston, TX (US); **Hilda Guzman**, Galveston, TX (US); **Scott C. Weaver**, Galveston, TX (US)

(73) Assignee: **The Board of Regents of the University of Texas System**, Austin, TX (US)

(\* ) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **14/086,533**

(22) Filed: **Nov. 21, 2013**

(65) **Prior Publication Data**

US 2014/0227765 A1 Aug. 14, 2014

**Related U.S. Application Data**

(60) Provisional application No. 61/729,317, filed on Nov. 21, 2012.

(51) **Int. Cl.**  
**C12N 15/86** (2006.01)  
**C07K 14/005** (2006.01)

(52) **U.S. Cl.**  
CPC ..... **C12N 15/86** (2013.01); **C07K 14/005** (2013.01); **C12N 2770/00021** (2013.01); **C12N 2770/00043** (2013.01)

(58) **Field of Classification Search**  
None  
See application file for complete search history.

(56) **References Cited**

**PUBLICATIONS**

Vasilakis et al. Negevirus: a Proposed New Taxon of Insect-Specific Viruses with Wide Geographic Distribution. *Journal of Virology* 2013, 87: 2475-2488.\*

Samina et al. Isolation of viruses from mosquitoes of the Negev, Israel. *Transactions of the Royal Society of Tropical and Hygiene* (1986), 80, 471-472.\*

Glaser et al. The native Wolbachia endosymbionts of *Drosophila melanogaster* and *Culex quinquefasciatus* increase host resistance to West Nile virus infection. *PLoS ONE* 5:e11977 (2010).

Hoshino et al. Genetic characterization of a new insect flavivirus isolated from *Culex pipiens* mosquito in Japan. *Virology* 359, 405-414 (2007).

Igarashi et al. Morphological, biochemical, and serological studies on a viral agent (CFA) which replicates in and causes fusion of *Aedes albopictus* (Singh) cells. *Virology* 74, 174-187 (1976).

Locali-Fabris et al. Complete nucleotide sequence, genomic organization and phylogenetic analysis of Citrus leprosis virus cytoplasmic type. *J. General Virology* 87, 2721-2729 (2006).

Marklewitz et al. Gouleako virus isolated from West African mosquitoes constitutes a proposed novel genus in the family Bunyaviridae. *J Virol* 85, 9227-9234 (2011).

Moreira et al. A Wolbachia symbiont in *Aedes aegypti* limits infection with dengue, Chikungunya, and Plasmodium. *Cell* 139,1268-1278 (2009).

Nasar et al. Eilat virus, a newly identified host restricted alphavirus. *Proceedings of the National Academy of Sciences of the United States of America* (2012).

Rancès, E., Y. H. Ye, M. Woolfit, E. A. McGraw, and S. L. O'Neill. The relative importance of innate immune priming in Wolbachia-mediated dengue interference. *PLoS pathogens* 8:e1002548 (2012).

Sarver et al. Sindbis virus-induced cytopathic effect in clones of *Aedes albopictus* (Singh) cells. *Virology* 80:390-400. (1977).

Kuno, G. A survey of the relationships among the viruses not considered arboviruses, vertebrates, and arthropods. *Acta Virol.* 48(3), 135-43 (2004).

Stollar, V., V. Thomas. An agent in the *Aedes aegypti* cell line (Peleg) which causes fusion of *Aedes albopictus* cells. *Virology.* 64(2):367-77 (1975).

Pascon et al. The complete nucleotide sequence and genomic organization of Citrus Leprosis associated Virus, Cytoplasmic type (CiLV-C). *Virus Genes.* Jun. 2006;32(3):289-98.

Kuno, G. A survey of the relationships among the viruses not considered arboviruses, vertebrates, and arthropods. *Acta virologica* 48, 135-143 (2004).

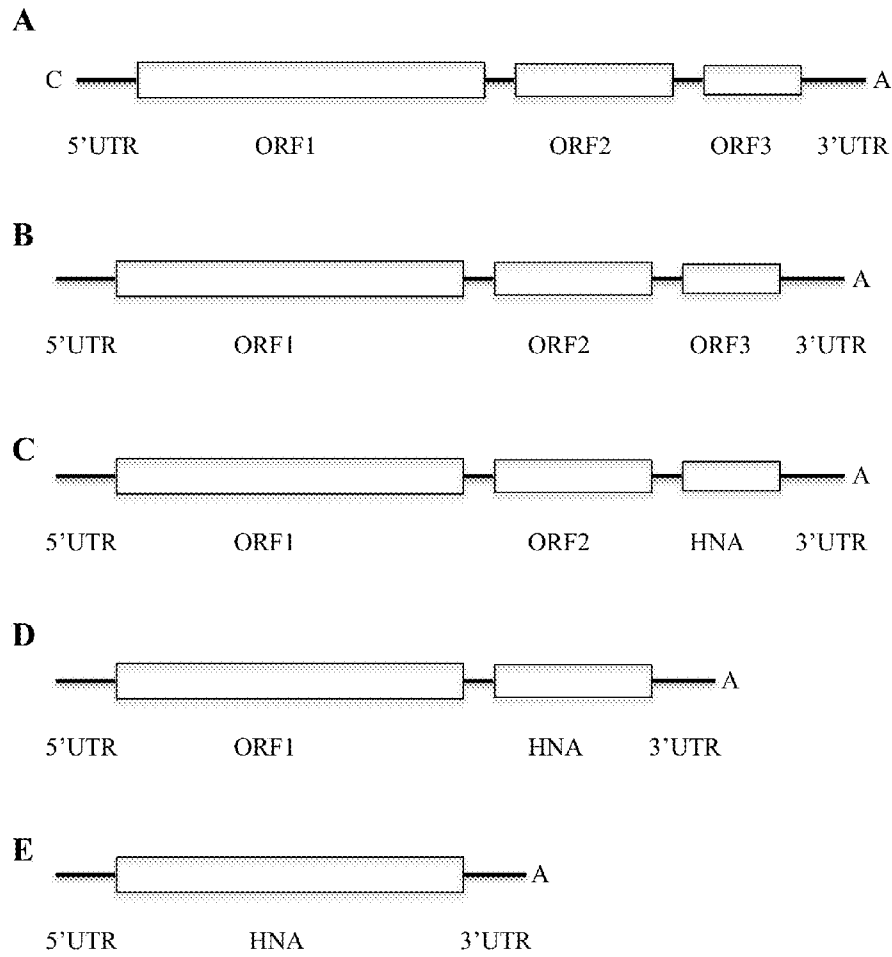
\* cited by examiner

*Primary Examiner* — Nick Zou  
(74) *Attorney, Agent, or Firm* — Norton Rose Fulbright US LLP

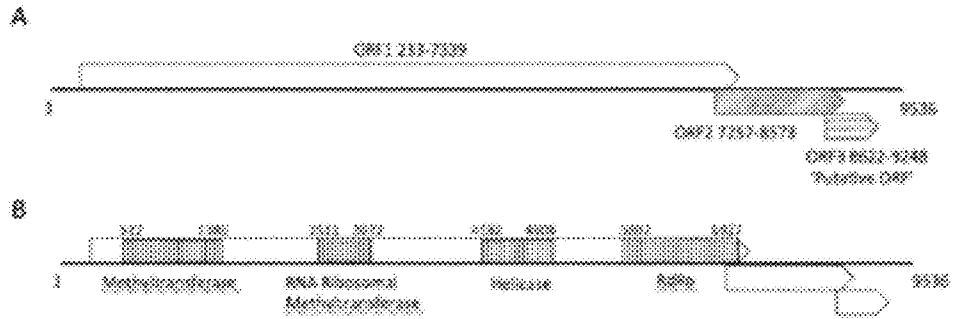
(57) **ABSTRACT**

Certain embodiments are directed to Negevirus and their polynucleotides and polypeptides.

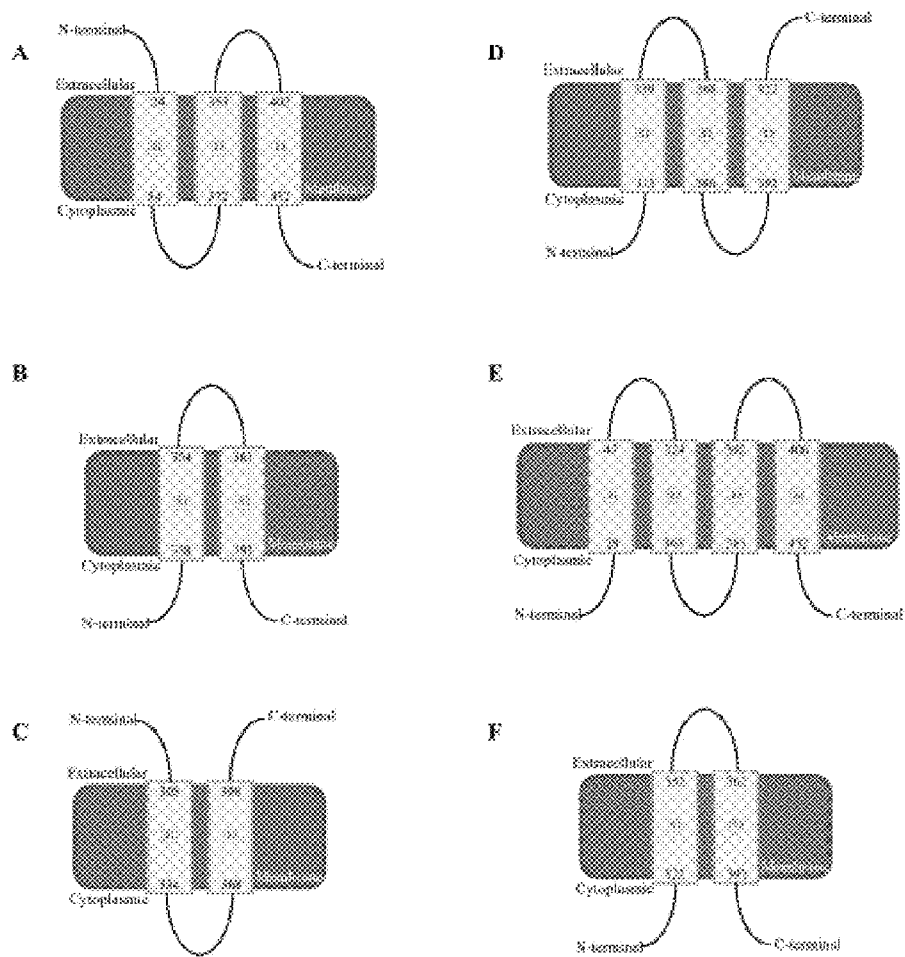
**8 Claims, 11 Drawing Sheets**



FIGs. 1A-1E



FIGs. 2A-2B



FIGs. 3A-3F

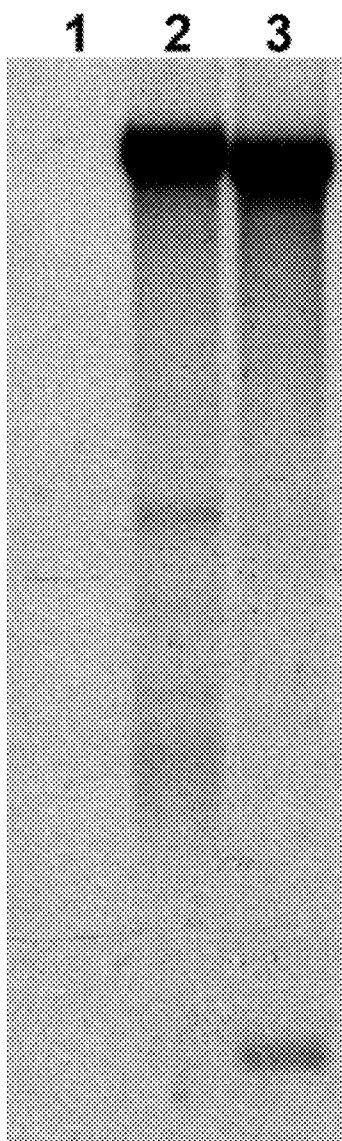


FIG. 4

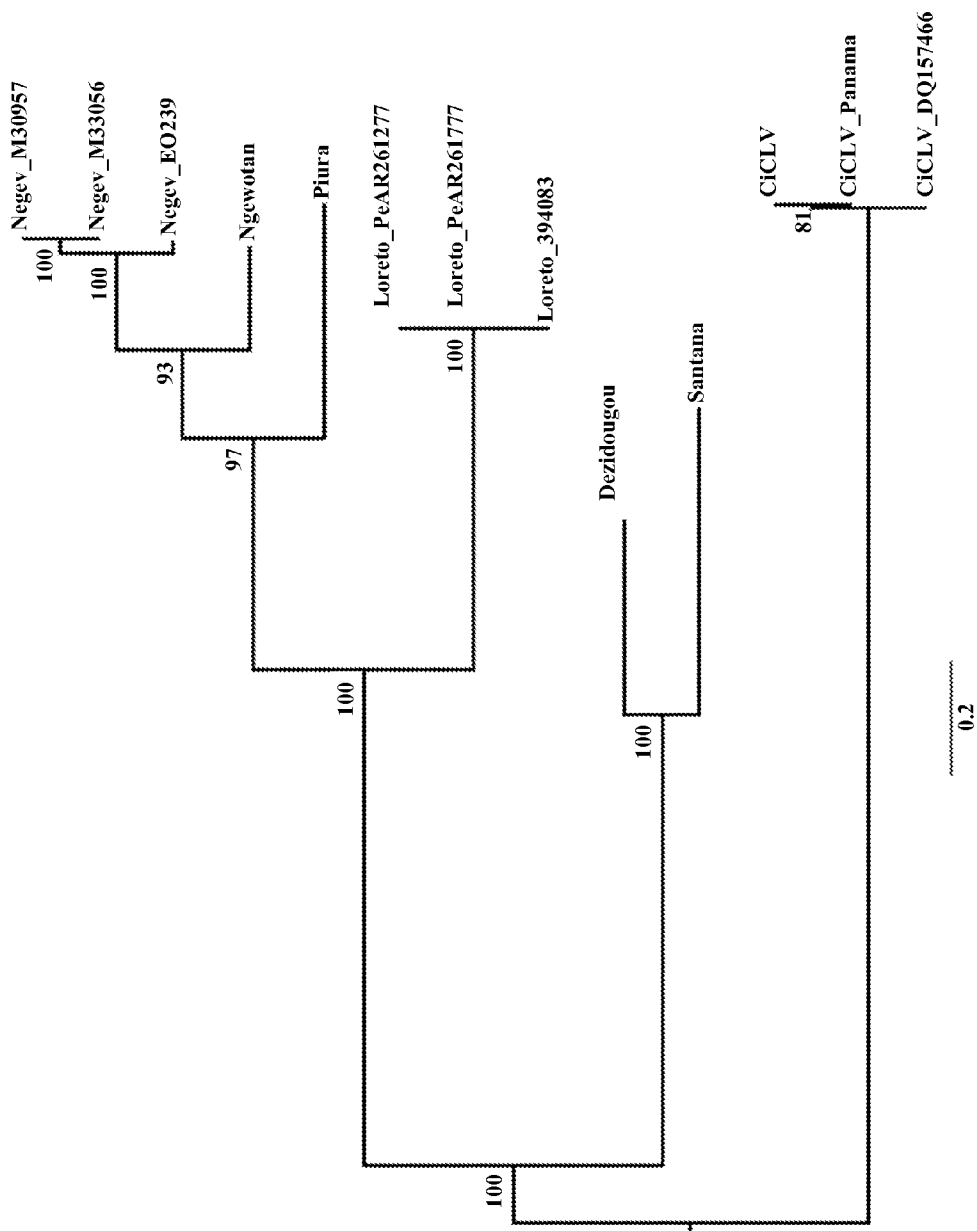


FIG. 5

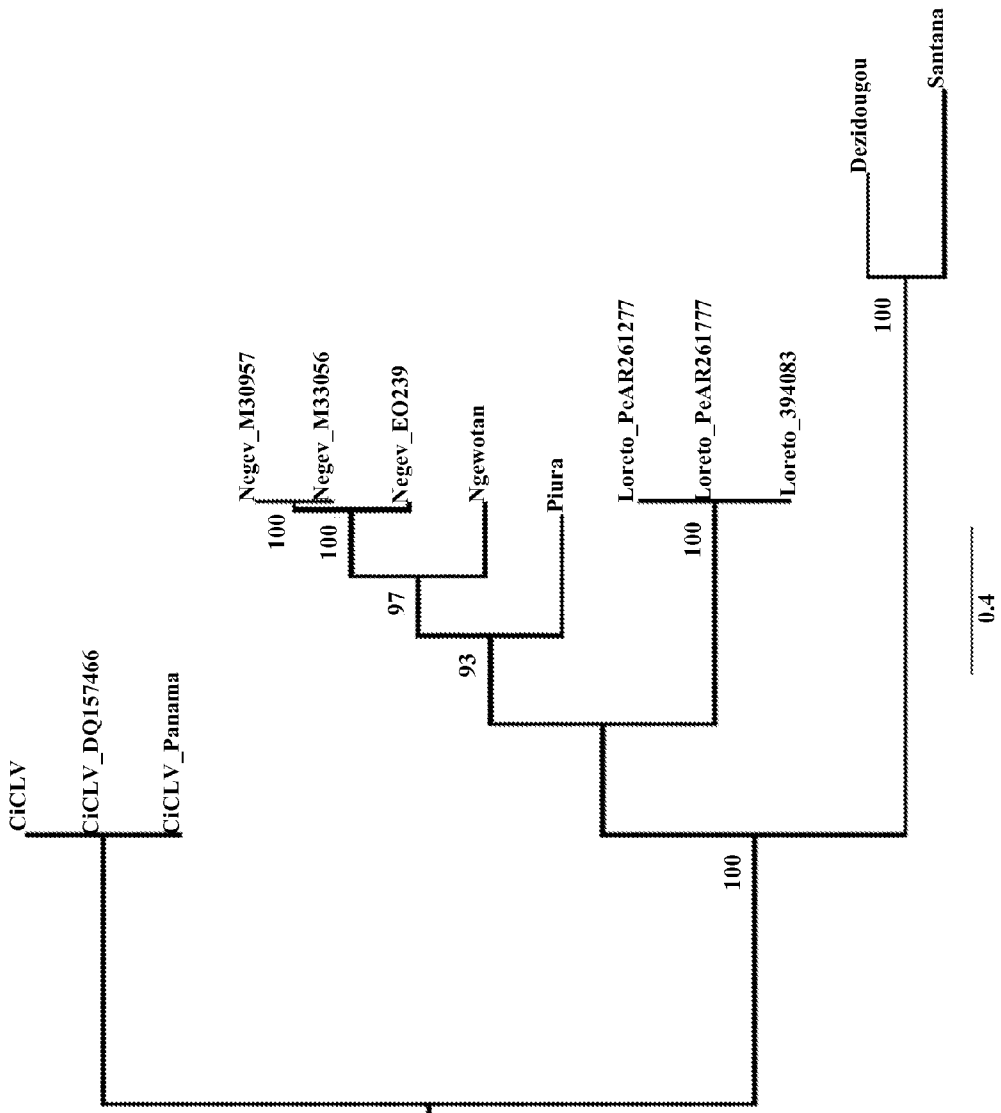


FIG. 6

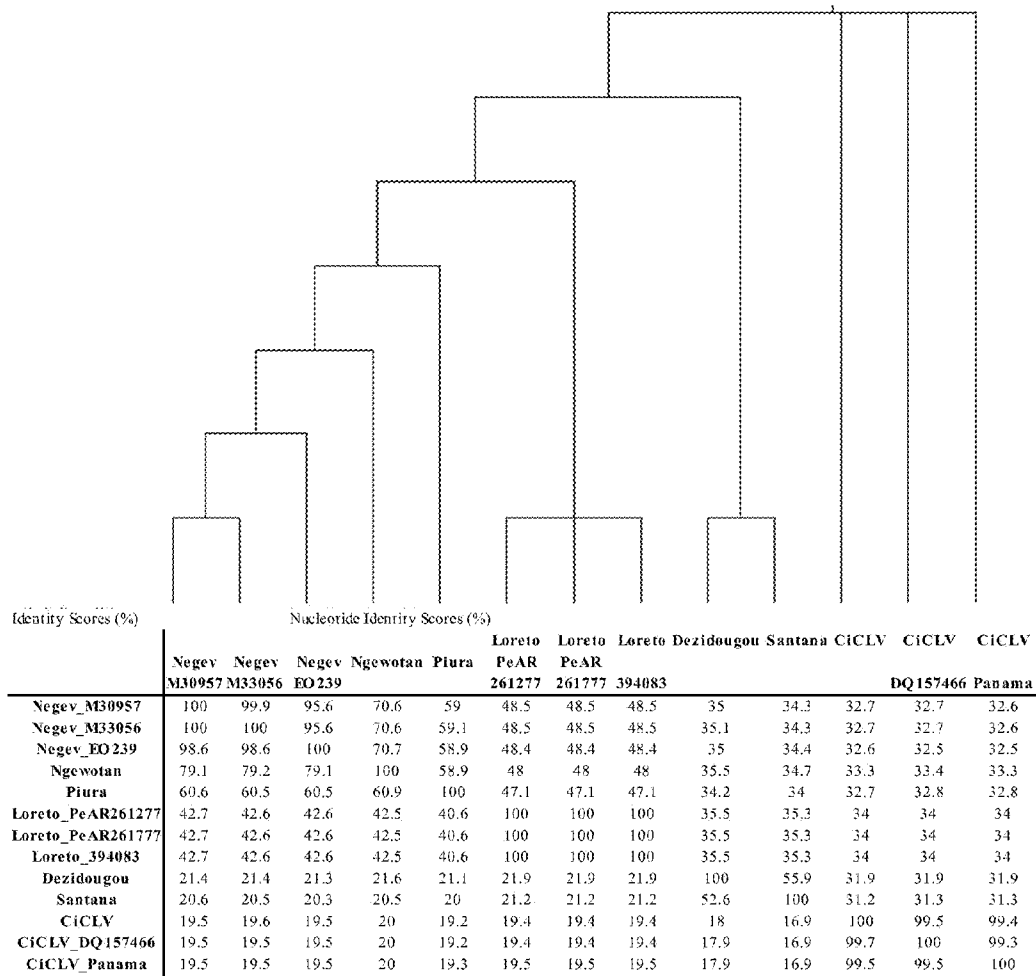


FIG. 7

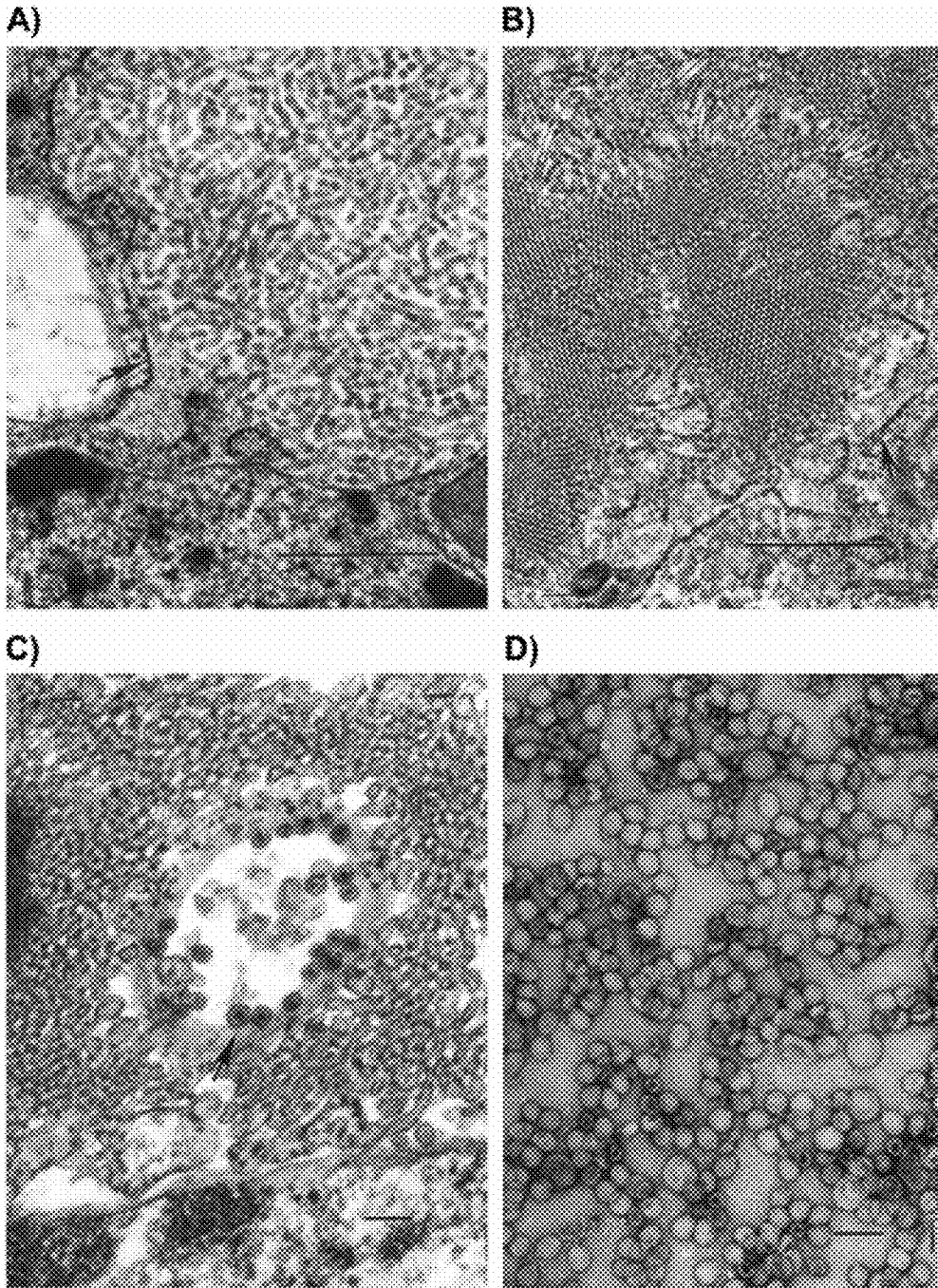


FIG. 8A-8D



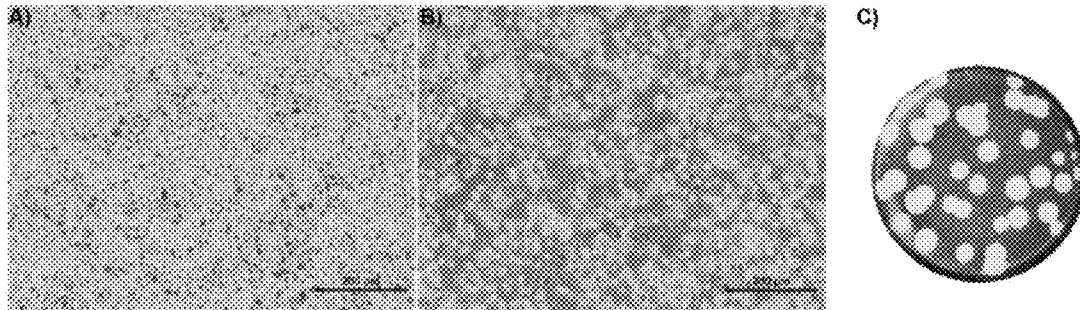


FIG. 9A-9C

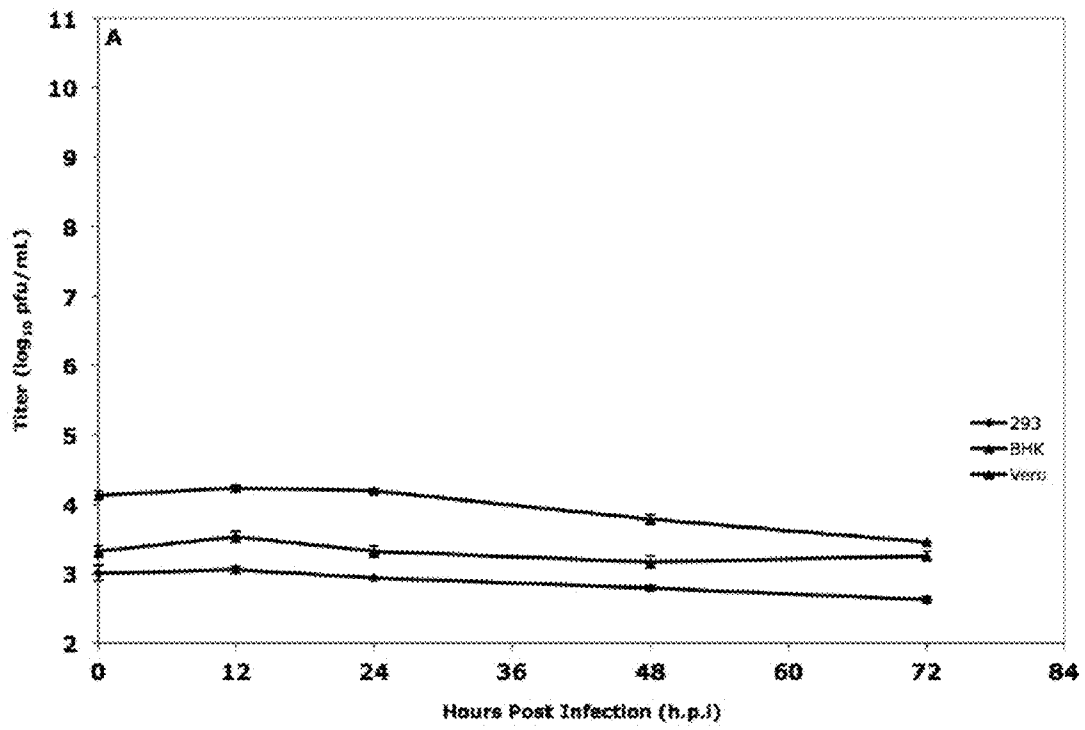


FIG. 10A

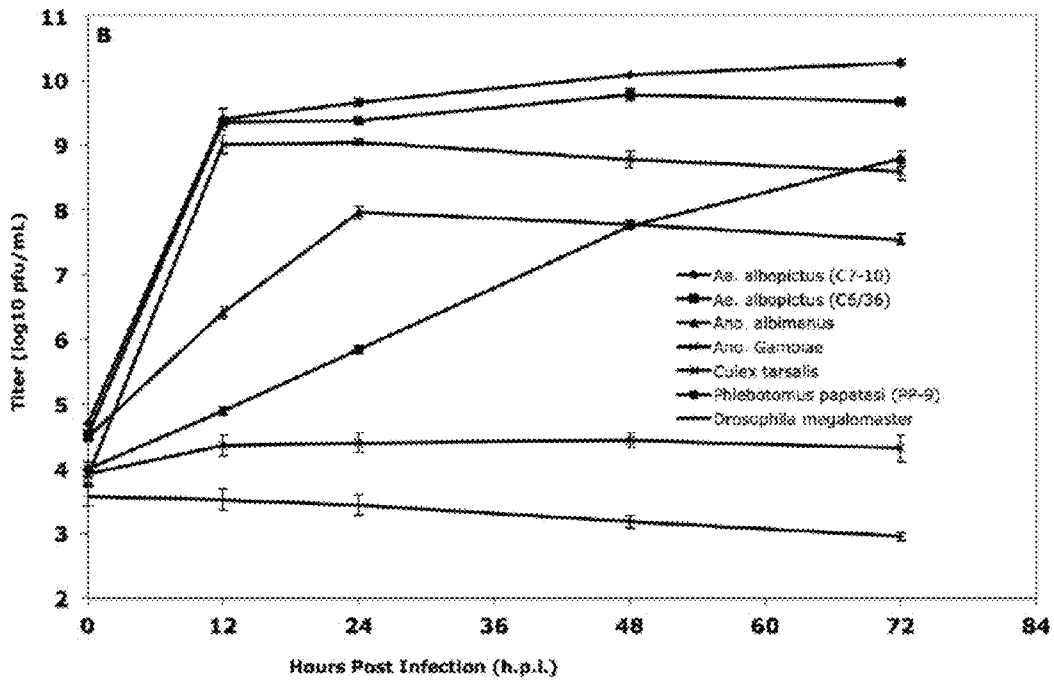


FIG. 10B

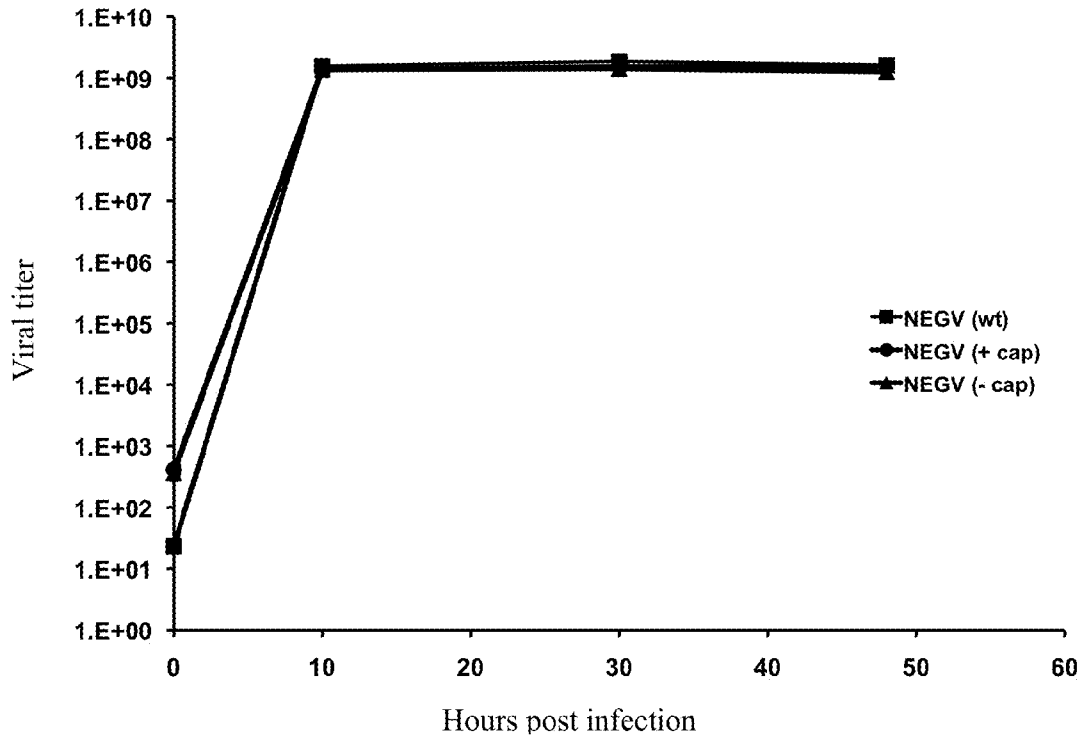


FIG. 11

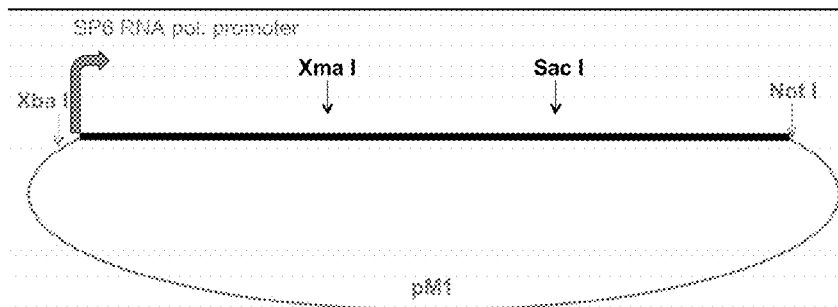


FIG. 12

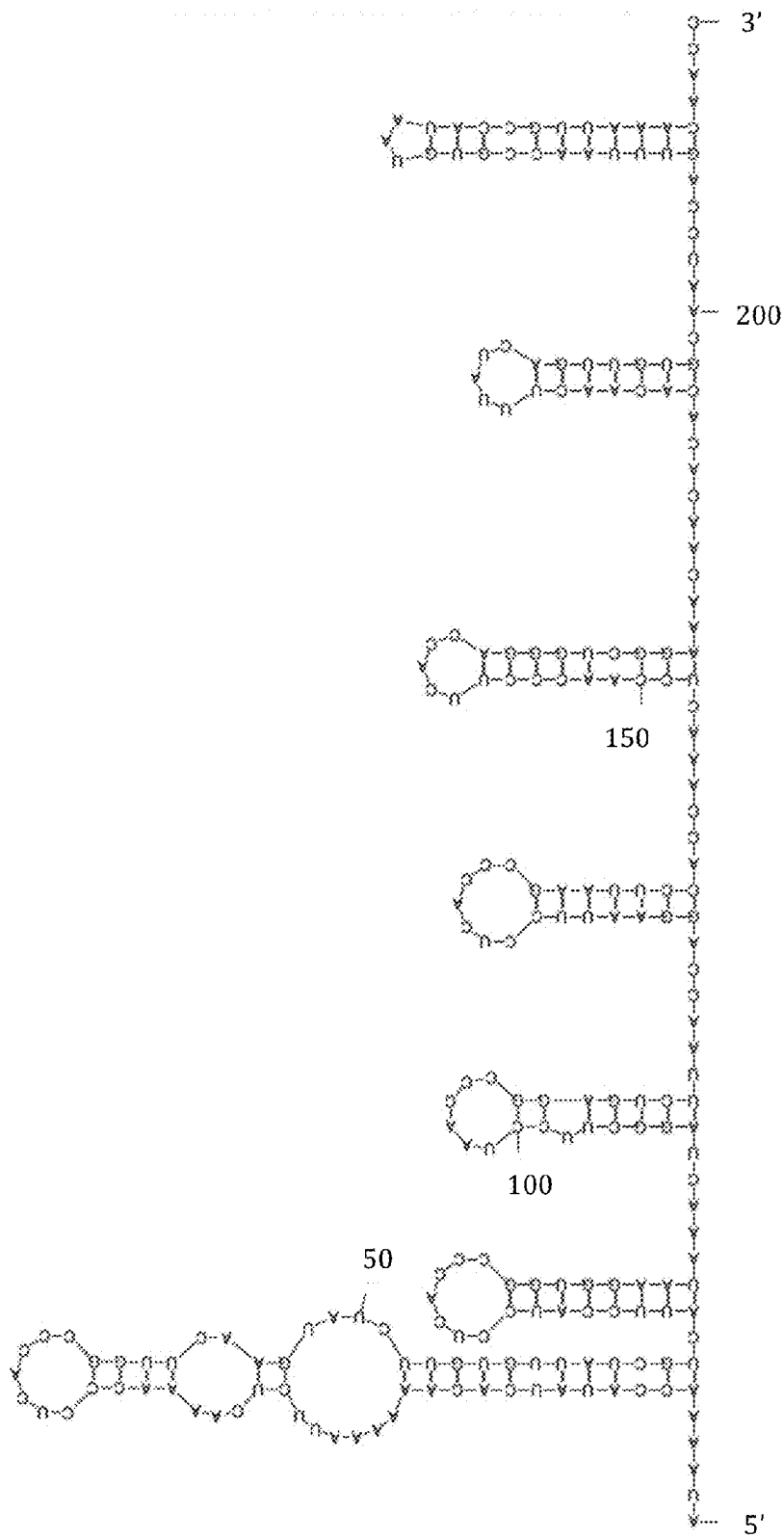


FIG. 13

1

## COMPOSITIONS AND METHODS RELATED TO VIRUSES OF THE GENUS *NEGEVIRUS*

### STATEMENT REGARDING FEDERALLY FUNDED RESEARCH

This invention was made with government support under N01-AI-30027 and HHSN272271000040I/HHSN2720004/D04 awarded by the NIH/NIAID. The government has certain rights in the invention.

### PRIORITY PARAGRAPH

This application is a non-provisional application of U.S. Provisional Patent application 61/729,317 filed Nov. 21, 2012, which is incorporated herein by reference in its entirety.

### REFERENCE TO SEQUENCE LISTING

A sequence listing required by 37 CFR 1.821-1.825 is being submitted electronically with this application. The sequence listing is incorporated herein by reference.

### BACKGROUND

During the past decade, a growing number of novel insect-specific viruses have been detected in naturally infected mosquitoes. The term “insect-specific” was initially used to describe viruses in the genus *Flavivirus* (Flaviviridae) that replicate in mosquito cells but not in vertebrate cells. Although the insect-specific flaviviruses share the same genome organization and numerous amino acid motifs with the vertebrate flaviviruses, they do not infect vertebrates nor participate in the classical arthropod-vertebrate transmission cycle of arboviruses (Kuno, 2004, *Acta virologica* 48:135-143). *Culex flavivirus* (CxFV) and cell fusing agent virus (CFAV) are probably the best-known members of the insect-specific flavivirus group (Hoshino et al., 2007, *Virology* 359: 405-414; Igarashi et al., 1976, *Virology* 74:174-187; Stollar and Thomas, 1975, *Virology* 64:367-377). Recently, an increasing number of non-flaviviral RNA viruses (negevirus, bunyaviruses, alphaviruses, nidoviruses and reoviruses) have been isolated from pools of field-collected mosquitoes, suggesting that these types of agents are quite common in mosquitoes in nature (Marklewitz et al., 2011, *J Virol* 85:9227-9234; Nasar et al., 2012, *Proc Nat Acad Sci USA* 109(36):14622-7, Epub 2012 Aug. 20; Nga et al., 2011, *PLoS pathogens* 7:e1002215; Quan et al., 2010, *Virus research* 147:17-24; Yamao et al., 2009, *Archives of virology* 154:153-158; Zirkel et al., 2011, *mBio* 2:e00077-00011).

There remains a need to identify additional viruses, as well as characterize and genetically engineer Negevirus for the benefit of mankind.

### SUMMARY

Certain embodiments described herein are directed to methods and compositions related to a novel group of insect-specific viruses, members of the genus Negevirus, that were isolated from mosquitoes and sandflies collected in Brazil, Peru, USA, Ivory Coast, Israel, and Indonesia.

In particular embodiments, the invention concerns isolated Negevirus nucleic acids, isolated Negevirus polypeptides, recombinant vectors incorporating Negevirus nucleic acid sequences and methods of using the same. The term “recombinant” may be used in conjunction with a polynucleotide or polypeptide, and generally refers to a polynucleotide or

2

polypeptide produced from a nucleic acid molecule that has been manipulated or engineered in vitro or that is the replicated product of such a molecule.

Embodiments include expression vectors engineered from Negevirus nucleic acid sequence. In certain aspects the Negevirus expression vector comprises, one or more of a 5' cap, 5'UTR, ORF1, ORF2, ORF3, 3'UTR, 3' polyadenylate segment, and heterologous nucleic acid. A Negevirus expression vector may or may not include the 5'cap. In certain embodiments, the 5'UTR of a Negevirus is included in a recombinant expression vector. A heterologous nucleic acid is a nucleic acid segment that is not encoded by the virus as isolated from nature or contains a coding region in a position, form, or context that is not naturally found in a virus.

In certain embodiments the heterologous nucleic acid can be positioned immediately 3' to ORF1, ORF2, ORF3, or the 5'UTR. In certain aspects the Negevirus expression vector comprises, 5' to 3': a 5' cap, 5'UTR, ORF1, ORF2, heterologous nucleic acid, and 3'UTR. In another aspect, the Negevirus expression vector comprises, 5' to 3': a 5'UTR, ORF1, ORF2, heterologous nucleic acid, and 3'UTR. In certain aspects, the Negevirus expression vector comprises, 5' to 3': 5'UTR, ORF1, heterologous nucleic acid, and 3'UTR. In certain aspects, the Negevirus expression vector comprises, 5' to 3': 5'UTR, heterologous nucleic acid, and 3'UTR. A Negevirus expression vector can further comprise an endogenous or heterologous polyadenylation site or polyadenylate segment 3' to the 3'UTR.

Certain aspects include a helper vector expressing one or more of ORF1, ORF2 and ORF3 polypeptides, or other polypeptides that may be needed to propagate the various expression vectors lacking one or more of ORF1, ORF2, and ORF3. The helper vector will express one or more polypeptides needed for Negevirus packaging, replication, and/or propagation.

In certain embodiments, one or more of ORF1, ORF2, and ORF3 can be replaced by a nucleic acid segment encoding a heterologous nucleic acid. In certain aspects the heterologous nucleic acid encodes a heterologous viral protein (e.g., viral structural proteins other than Negevirus structural proteins). In certain embodiments the heterologous nucleic acid can encode one or more antigens. The antigen can be a cancer antigen or a microbial antigen. In certain aspects the microbial antigen is a bacterial or a viral antigen.

In certain aspects heterologous nucleic acid expression is under the control of an endogenous Negevirus promoter. In still a further aspect, heterologous nucleic acid expression is under the control of a heterologous promoter. The heterologous nucleic acid can express a target RNA. The target RNA can function as an RNA or be translated into an encoded polypeptide(s). In certain aspects an RNA can be an inhibitory RNA or oligonucleotide such as a miRNA, siRNA, or an antisense oligonucleotide.

In certain aspects the heterologous nucleic acid encodes an insecticidal RNA, peptide, or polypeptide.

Certain aspects are directed to recombinant expression vectors comprising all or part of a nucleic acid sequence that is 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100% identical to SEQ ID NO: 1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 14, 15, 16, 17, 19, or 20. In certain aspects the expression vector comprises 10, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, 10000, 10500, 11000 consecutive nucleotides of that are at least 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100% identical to SEQ ID NO: 1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 14, 15, 16, 17, 19, or 20. In certain aspects a variant nucleotide sequence retains its func-

tion as a Negevirus. In other aspects the replication of Negevirus is attenuated in a Negev virus variant.

Embodiments also include methods of controlling an insect population by contacting insects with a Negevirus expression vector that express an insecticidal protein or nucleic acid. In certain aspects a recombinant virus as described herein can be dispersed in an environment to control insects, such as mosquitoes. In one aspect a virus can be disperse in a pond or standing water.

Certain embodiments are directed to isolated nucleic acid segments and recombinant vectors incorporating Negevirus nucleic sequences. In certain aspects the Negevirus nucleic acid sequences encode a polypeptide or peptide that includes within its amino acid sequence a contiguous amino acid sequence in accordance with, or essentially corresponding to one or more Negevirus polypeptide or segment thereof, including but not limited to SEQ ID NO:5, 10, 13, 18, or 21, or other polypeptide(s) encoded by SEQ ID NO:1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 14, 15, 16, 17, 19, or 20.

In certain aspects an isolated nucleic acid has a nucleotide sequence that is 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% identical to SEQ ID NO: 1, 3, 6, 7, 8, 11, 14, 15, 16, or 19.

The term "isolated", when used in relation to a nucleic acid molecule or sequence, refers to a nucleic acid sequence that is identified and separated from at least one contaminant nucleic acid with which it is ordinarily associated with in nature. Isolated nucleic acids are nucleic acids present in a form or setting that is different from that in which it is found in nature. In an embodiment, an isolated nucleic acid is substantially free of cellular material or culture medium when produced by recombinant DNA techniques; or chemical precursors, or other chemicals when chemically synthesized.

In certain aspects the recombinant nucleic acids are from a Negev virus (SEQ ID NO:3, 6, or 7), Piura virus (SEQ ID NO:8), Loreto virus (SEQ ID NO:11, 14, or 15), Dezidougou virus (SEQ ID NO:16), or Santana virus (SEQ ID NO:19). Further embodiments are directed to recombinant nucleic acids comprising all or part of a nucleic acid sequence that is 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100% identical to SEQ ID NO: 1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 14, 15, 16, 17, 19, or 20. In certain aspects a Negevirus has a consensus sequence of SEQ ID NO:1.

In certain aspects the Negevirus specific nucleic acids are Negevirus specific oligonucleotides of 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100 consecutive nucleotides of SEQ ID NO: 1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 14, 15, 16, 17, 19, or 20. In certain aspects the oligonucleotides are Negevirus specific oligonucleotide primers and/or oligonucleotide probes. As used herein, the phrase "Negevirus specific" means the ability of a nucleic acid molecule to recognize a nucleic acid sequence complementary to its own and to form double-helical segments through hydrogen bonding between complementary base pairs and at a level that can be distinguished from random or non-specific association.

Certain embodiments are directed to polypeptides and segments thereof encoded by a nucleic acid that is 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100% identical to SEQ ID NO:5, 10, 13, 18, or 21. In certain aspects the polypeptide comprising about or at least 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500 consecutive amino acids, including all values and ranges there between, of SEQ ID NO:5, 10, 13, 18, or 21. Further aspects include polypeptide fragments including amino acids 1, 10, 20, 30, 40, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1500, 1600,

1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400 to 10, 20, 30, 40, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, including all ranges there between. In certain aspects the polypeptide is an isolated or purified polypeptide. As used herein, the term "purified" or "to purify" refers to the removal of undesired components from a sample. In certain embodiments a polypeptide as described herein is coupled to a heterologous moiety.

Further embodiments are directed to methods for producing an antibody that specifically binds a Negevirus comprising inducing an immune response in a mammal by providing a Negevirus polypeptide to the mammal and isolating Negevirus specific antibody from the mammal. As used herein, the phrase "specifically binds" refers to the ability of an antibody to bind an amino acid sequence or structure at a level that can be distinguished from random, background, or non-specific binding.

Still further embodiments are directed to antibodies that bind a polypeptide encoded by a nucleic acid that is 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100% identical to SEQ ID NO: 5, 10, 13, 18, or 21. In certain aspects the antibody is an isolated or purified antibody. An antibody can specifically bind a Negevirus polypeptide. In certain aspects the presence of a Negevirus polypeptide can be detected by the binding of an antibody that specifically binds Negevirus.

Yet further embodiments are directed to methods for producing a Negevirus comprising introducing a Negevirus expression vector capable of producing a Negevirus into a host cell and incubating the host cell under conditions that produce Negevirus. In certain aspects the Negevirus is isolated or purified from the Negevirus producing host cell.

Certain embodiments are directed to methods for detecting a Negevirus comprising contacting a sample with a Negevirus specific reagent, measuring the binding or hybridization of the Negevirus specific reagent, and determining if the Negevirus specific reagent is specifically binding to the sample.

Further embodiments are directed to kits for producing and or detecting a Negevirus comprising a Negevirus specific antibody, peptide, nucleic acid, and/or oligonucleotide.

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence", "sequence identity", "percentage of sequence identity", and "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA sequence given in a sequence listing or may comprise a complete gene sequence. Generally, a reference sequence is at least 20 nucleotides in length, frequently at least 25 nucleotides in length, and often at least 50 nucleotides in length. Since two polynucleotides may each (1) comprise a sequence (i.e., a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window", as used herein, refers to a conceptual segment of at least 20 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 20 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions

or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith and Waterman (Smith and Waterman, *Adv. Appl. Math.* 2: 482 (1981)) by the homology alignment algorithm of Needleman and Wunsch (Needleman and Wunsch, *J. Mol. Biol.* 48:443 (1970)), by the search for similarity method of Pearson and Lipman (Pearson and Lipman, *Proc. Natl. Acad. Sci. (U.S.)* 85:2444 (1988)), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by inspection, and the best alignment (i.e., resulting in the highest percentage of homology over the comparison window) generated by the various methods is selected. The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The terms "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 85 percent sequence identity, preferably at least 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 25 50 nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of the reference sequence over the window of comparison. The reference sequence may be a subset of a larger sequence, for example, as a segment of the full-length sequences of the compositions claimed in the present invention.

Moieties of the invention, such as polypeptides, peptides, antigens, or immunogens, may be coupled or linked covalently or noncovalently to other moieties such as adjuvants, proteins, peptides, supports, fluorescence moieties, or labels. The term "conjugate" or "immunoconjugate" is broadly used to define the operative association of one moiety with another agent and is not intended to refer solely to any type of operative association, and is particularly not limited to chemical "conjugation."

The term "providing" is used according to its ordinary meaning "to supply or furnish for use." In some embodiments, the protein is provided directly by administering the protein, while in other embodiments, the protein is provided by administering a nucleic acid encoding the protein. In certain aspects the invention contemplates compositions comprising various combinations of nucleic acid, antigens, peptides, and/or epitopes.

Other embodiments of the invention are discussed throughout this application. Any embodiment discussed with respect to one aspect of the invention applies to other aspects of the invention as well and vice versa. Each embodiment described herein is understood to be embodiments that are applicable to all aspects of the invention. It is contemplated that any

embodiment discussed herein can be implemented with respect to a method or composition described, and vice versa. Furthermore, compositions and kits of the invention can be used to achieve methods described.

The use of the word "a" or "an" when used in conjunction with the term "comprising" in the claims and/or the specification may mean "one," but it is also consistent with the meaning of "one or more," "at least one," and "one or more than one."

Throughout this application, the term "about" is used to indicate that a value includes the standard deviation of error for the device or method being employed to determine the value.

The use of the term "or" in the claims is used to mean "and/or" unless explicitly indicated to refer to alternatives only or the alternatives are mutually exclusive, although the disclosure supports a definition that refers to only alternatives and "and/or."

As used in this specification and claim(s), the words "comprising" (and any form of comprising, such as "comprise" and "comprises"), "having" (and any form of having, such as "have" and "has"), "including" (and any form of including, such as "includes" and "include") or "containing" (and any form of containing, such as "contains" and "contain") are inclusive or open-ended and do not exclude additional, unrecited elements or method steps.

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating specific embodiments, are given by way of illustration only, since various changes and modifications within the spirit and scope of the specification will become apparent to those skilled in the art from this detailed description.

## DESCRIPTION OF THE DRAWINGS

The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of the specification embodiments presented herein.

FIGS. 1A-1E. Negevirus constructs. (A)-(E) are diagrams of various Negevirus constructs. C=5'cap, A=polyadenylation site, ORF1=open reading frame 1, ORF2=open reading frame 2, ORF3=open reading frame 3, HNA=heterologous nucleic acid, 5'UTR=5' untranslated region, 3' UTR=3' untranslated region, and a line between boxes represents intragenic spacers.

FIGS. 2A-2B. The genome organization and position of the open reading frames (A) for Negev E0239 (SEQ ID NO:7) and the conserved protein domains (B) again for Negev E0239. The viruses show similar genome organization and protein domains.

FIGS. 3A-3F. The predicted transmembrane domains and the orientation of the ORF2 by MAMSAT-SVM for (A) Negev virus, (B) Piura virus, (C) Loreto virus, (D) Dezidougou virus, (E) Santana virus and (F) Ngewotan virus

FIG. 4. Analysis of genomic RNA of NEG V and PIUV labeled with [<sup>3</sup>H]uridine in the presence of dactinomycin (ActD) for 12 hrs. Both viruses were purified via rate-zonal centrifugation. Viral RNA was analyzed by agarose gel electrophoresis. Lane 1=mock, 2=NEG V, 3=PIUV.

FIG. 5. Phylogenetic trees produced using Maximum likelihood methods of Negevirus genomes plus three genomes of CiCLV. The trees were rooted using the CiCLV viruses as an

outgroup. The region of the genome corresponds to the nt 626-2908 of SEQ ID NO:7 (Negev E0329), which corresponds to the helicase region of the genome. The model used was the TrN+G model with 1000 bootstraps. Bootstraps are presented on the major branches.

FIG. 6. Phylogenetic trees produced using Maximum likelihood methods of the Negevirus genomes plus three genomes of CiCLV. The trees were midpoint rooted. The region of the genome corresponds to nucleotides 4316-7309 of SEQ ID NO:7 (Negev E0329), which corresponds to the RNA-dependent RNA-polymerase of the genome. The model used was the GTR+G model with 1000 bootstraps. Bootstraps are presented on the major branches.

FIG. 7. Cladistic tree showing the relationships between the viruses along with the nucleotide identity and the protein identity for the ORF1 of the nine viruses and the RNA species one of CiCLV. Alignments were performed as proteins and then toggled back to nucleotide. The branch lengths of the tree do not reflect genetic distance but have the same topology as the trees found in FIG. 3 and FIG. 4.

FIGS. 8A-8D. Transmission electron microscopy analysis of infected cells and purified suspensions. (A) Expanded perinuclear space (arrow indicates its membrane) of an *Ae. albopictus* C6/36 cell infected with Negev (E0329) virus is filled with microtubules 20 nm in diameter and up to 160 nm long. Bar=0.5  $\mu$ m. (B) Portion of a tremendous perinuclear space-granular endoplasmic reticulum extension loaded with microtubules forming paracrystalline arrays in cross-sections in a C6/36 cell infected with ArA 20086 virus. The arrow indicates a limiting membrane with ribosomes at the outer surface. Bar=0.5  $\mu$ m. (C) Cytopathic vacuole with spherules at its periphery (arrow) surrounded by microtubules in a perinuclear space of a C6/36 cell infected with ArA 20086 virus. Bar=100 nm. (D) Negatively stained (2% uranyl acetate) suspension of purified suspension of P60 virus contains mostly particles ~50 nm in diameter. Bar=100 nm.

FIGS. 9A-9C. Cytopathic effects of Negev virus (E0329) infection in C7/10 cells. (A) Mock infected C7/10 monolayers observed with bright field microscopy at 12 hours post infection (hpi); (B) NEGV infected C7/10 cells with an MOI of 10 at 12 hpi observed under bright field microscopy; and (C) Representative plaques of Negev virus-infected C7/10 cells 36 hpi. Cells were fixed with 10% formalin and stained with crystal violet dye.

FIGS. 10A-10B. Comparative replication curves of prototype Negev virus (E0329) (A) Virus outputs from 12-72 hpi at a multiplicity of infection (MOI) of 10 by Negev E0329 in vertebrate cell lines Vero (African green monkey kidney), BHK-21 (baby hamster kidney) and HEK293 (human embryonic kidney). (B) Virus outputs from 12-72 hpi at MOI 10 by Negev virus in the following insect cell lines *Ae. albopictus* (C6/36 and C7/10), *An. albimanus*, *An. gambiae*, *Cx. tarsalis*, *P. papatasi*, and *D. melanogaster*. The limit of detection of the assay is 1.0 log 10 pfu/mL.

FIG. 11. Comparison of replication for cap+ and cap- Negevirus.

FIG. 12. Restriction map of plasmid pM1 encoding a Negevirus genome.

FIG. 13. Illustration of secondary structure of Negevirus 5'UTR.

## DESCRIPTION

The inventors describe a novel group of insect-specific viruses occurring in mosquitoes and phlebotomine sandflies distantly related phylogenetically to citrus leprosis virus C (CiLV-C), genus Cilevirus, a mite-transmitted virus causing disease in citrus plants (Locali-Fabris et al., 2012, *Genus Cilevirus*, p. 1169-1172. In King, Adams, Carstens, and Lefkowitz (ed.), *Virus Taxonomy*. Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier, San Diego; Locali-Fabris et al., 2006, *Journal of general virology* 87:2721-2729; Pascon et al., 2006, *Virus genes* 32:289-298). The inventors designate this new group of viruses as Negevirus.

### I. Negevirus

Negevirus genomes consist of single stranded, positive sense RNA with a poly-A tail. By electron microscopy, the virions appear as spherical particles with diameters of ~45-55 nm (FIGS. 8A-8D). Based on their genome organization and phylogenetic relationship, the viruses (examples of which are designated Negev virus, Ngewotan virus, Piura virus, Loreto virus, Dezidougou virus, and Santana virus) appear to form a new taxon of Negevirus. The Negevirus viruses replicate rapidly and to high titer (up to 10<sup>10</sup> PFU/ml) in mosquito cells, producing extensive cytopathic effect and plaques; but they do not appear to replicate in mammalian cells or mice. These viruses have one large open-reading-frame (ORF1) and two small ORFs, ORF2 and ORF3. The viruses appear to be nonsegmented, however they may package some mRNA. A prototype virus, Negev virus (SEQ ID NO:7), could not infect vertebrate cell lines. However, it could readily infect insect cells including mosquito cell lines encompassing three genera, as well as sandfly cells. Negev virus displayed extensive cytopathology 12 hpi and could be plagued on C7/10 cells (*Aedes albopictus*). Negev virus grew to high titers (10<sup>10</sup> pfu/mL) 48-hrs-post-infection in C7/10 cells.

A full-length infectious clone was generated using reverse genetics (FIG. 12). Interestingly, the virus could be rescued independent of the 5' cap with identical rescue efficiency to rescue with the addition of 5' cap (NEGV RNA+cap=titer of 9.0 log<sub>10</sub> pfu/mL and NEGV RNA-cap=titer of 9.0 log<sub>10</sub> pfu/mL) (FIG. 11). The 5' UTR may represent an IRES sequence that could facilitate the translation of the Negev virus genome (FIG. 13). The recombinant Negev virus displayed similar growth kinetics to wild-type virus and grew to high titers in insect cells. In vivo mosquito experiments showed that Negev virus was also able to replicate and cause disseminated infection in *Aedes aegypti* mosquitoes. The name, source, and geographic origins of viruses representative of the Negevirus group are listed in Table 1.

TABLE 1

Virus Name	Strain Designation	Mosquito Species	Collection Locality	Collection Date
Negev	M30957	<i>Culex coronator</i>	Harris Co., Texas, US	2008
Negev	M33056	<i>Culex quinquefasciatus</i>	Harris Co., Texas, US	2008
Negev	E0329	<i>Anopheles coustani</i>	Negev, Israel	1983
Piura	P60	<i>Culex</i> sp.	Paja Piura, Peru	1996
Loreto	3940-83	<i>Anopheles albimanus</i>	Lima, Peru	1983
Loreto	Pe AR 2617/77	<i>Luzomyia</i> sp.	Iquitos, Loreto, Peru	1977



TABLE 1-continued

Virus Name	Strain Designation	Mosquito Species	Collection Locality	Collection Date
Loreto	Pe AR 2612/77	<i>Culex</i> sp.	Iquitos, Loreto, Peru	1977
Dezidougou	ArA 20086	<i>Aedes aegypti</i>	Dezidougou, Ivory Coast	~1984
Santana	BeAR 517449	<i>Culex</i> sp.	Macapa, Amapa, Brazil	1992

### Genome Organization and Analysis.

The size of the positive sense, single-strand genomes of the representative viruses ranged in size from approximately 9 to 10 kilobases (kb) (SEQ ID NO:3, 6, 7, 8, 11, 14, 15, 16, and 19). Three open reading frames (ORFs) are flanked by untranslated regions (UTRs) at the 5' and 3' ends, while each ORF is separated by short intergenic regions. Using Negev virus strain E0329 as the prototype, the inventors determined that ORF1 and ORF2 encode the non-structural proteins and structural proteins, respectively. ORF prediction programs (EnzymeX, Aalsmeer, Netherlands) did not identify a function for ORF3.

#### 5'- and 3'-UTRs.

The sequences of the 5'- and 3'-UTRs vary in length among these viruses, ranging in length from 72 to 730 nt and 121 to 442 nt long, respectively. For all viruses a polyadenylation tail is present in the distal sites of each genome, varying in length from 13-52 nt in length.

#### Open Reading Frames (ORF).

A large ORF was found starting at nt 233 to 7339 of SEQ ID NO:7. Two small ORFs were identified at the 3' end of the genome (FIG. 2A). The large ORF contains putative protein domains that correspond to non-structural proteins (FIG. 2B). Using the protein domain prediction software in the BLAST suite of programs four functional domains were identified: (i) a methyltransferase domain at nt 522-1386 of SEQ ID NO:7; (ii) an RNA ribosomal methyltransferase domain at nt 2511-3072 of SEQ ID NO:7; (iii) a helicase domain at nt 4182-4908 of SEQ ID NO:7; and (iv) an RNA-dependent RNA-polymerase domain (RdRp) at nt 5802-6927 of SEQ ID NO:7 (FIG. 2B). Using the PsiPRED server available at URL <http://bioinfics.ucl.ac.uk/psipredll?program=psipred> (Buchan et al., 2010, *Nucleic acids research* 38:W563-568), ORF2 was predicted to contain three transmembrane regions according to the algorithm prediction software MEMSAT-SVM (Jones, 2007, *Bioinformatics* 23:538-544; Jones et al., 1994, *Biochemistry* 33:3038-3049; Nugent and Jones, 2009, *BMC bioinformatics* 10:159), as depicted in FIG. 3A. The presence of these transmembrane regions suggests that this protein is contained within a viral envelope and may be a glycoprotein. Analysis of all ORF2s from Negev, Piura, Loreto, Dezidougou, Santana and Ngewotan viruses showed that all contained transmembrane helices (FIG. 3A-3F). However, the number and arrangement of these helices differed for all 6 viruses, suggesting a significant difference in the arrangement of the viral surface. For ORF3, no putative functions or domain homologies were identified. Short intergenic regions ranging from 14-44 nt and 21-175 long are intersecting the junctions of ORF1/ORF2 and ORF2/ORF3, respectively.

To confirm monosegmentation, monolayers of C7/10 cells were infected with Negev E0239 and Piura P60 viruses at a MOI of 10, in the presence of [<sup>3</sup>H] uridine (20 μCi/ml). Viral RNA was isolated from purified virus and analyzed by agarose gel electrophoresis. The results showed an abundant large genomic RNA species as well as the presence of several less abundant smaller RNA species (FIG. 4).

### Phylogenetic Analysis.

Alignments of the ORF1 were created for all of the virus sequences. As initial BLAST results had indicated that the nearest viral relative was the Citrus Leptos C viruses (CiLV-C) (Locali-Fabris et al., 2006, *Journal of General Virology* 87:2721-2729), the ORF1 sequences of the newly identified sequences were aligned with the three full-length RNA sequences of CiLV-C present in GenBank. The sequences were first aligned as proteins and then toggled back to nucleotides. These alignments were run through the GUIDANCE algorithm, which shows the confidence in the alignment, to ensure that regions of the genome exhibited no more evidence of homology than random assembly of protein would show. Two regions of the ORF1, nt 626-2908 of SEQ ID NO:7 and nt 4316-7309 of SEQ ID NO:7, exhibited confidence levels sufficient to perform further phylogenetic analysis. Phylogenetic trees generated under maximum likelihood (ML) and Bayesian algorithms exhibited the same topology, and thus only the ML trees are shown (FIG. 5 and FIG. 6, respectively). Distinct viruses can be identified: Negev, Loreto, Ngewotan, Piura, Dezidougou, and Santana viruses. As the regions used to generate the phylogeny were highly conserved, the relationships presented in the trees do not necessarily reflect the diversity among the viruses. Therefore, a full-length alignment of the ORF1 polyprotein was used to determine similarities in nucleotide and protein sequence of these viruses (FIG. 7). The three Negev viruses and the three Loreto viruses exhibited nucleotide and protein identities between 95.6-100% and 98.6-100%, respectively. However, among the viruses represented in this study (excluding the (CiLV-C) the nucleotide identity ranged from 33.2-70.6% and the protein identity ranged from 20-79.2%. The 3 strains of the CiLV-C viruses were nearly identical but showed little similarity with the newly identified viruses with nucleotide identity ranging from 30.7-34.7% and protein identity ranging from 16.9-20.6%.

#### Ultrastructural Characteristics.

The most prominent ultrastructural characteristic of infected C6/36 mosquito cells was expansion of perinuclear spaces (FIG. 8A). These had a universal diameter of 20 nm. In some cells, the expansions were filled only with vesicles, in others also with microtubules so the vesicles appeared as cross-sections of the microtubules. These microtubules were up to 160 nm long and rarely even longer. In some cells, they formed paracrystalline arrays (FIG. 8B). Expanded perinuclear space filled with vesicles or tubules could occupy most of the cell volume, pushing the cytoplasm to the cell periphery as a thin rim.

The second peculiar feature of many of these viruses was the formation of cytoplasmic cytopathic vacuoles (CPVs), similar to those seen with alphaviruses, containing spherules ~50 nm in diameter at the inner surface of their limiting membrane. In alphavirus-infected cells, CPVs are modified endosomes and lysosomes in which translation, transcription, and assembly of viral nucleocapsids occurs (Froshauer et al., 1988, *Journal of Cell Biology* 107:2075-2086). Some CPVs

reached a diameter of 1.4  $\mu\text{m}$  and could be found in almost all viruses studied (Table 2) and sometimes among the microtubules of the expanded perinuclear space (FIG. 8C). In a negatively stained purified suspension of the Piura virus (P60), spherical particles with diameters of  $\sim 45$  nm and  $\sim 55$  nm were found (FIG. 8D).

TABLE 2

Strain Designation	Presence of cytoplasmic cytopathic vacuoles (CPVs)	Expansion of perinuclear space
M30957	Yes	Yes
M33056	Yes	Yes
E0329	Yes	Yes
P60	Yes	Yes
3940-83	Yes	Not Seen
Pe AR 2617/77	Yes	Not Seen
Pe AR 2612/77	Yes	Not Seen
ArA 20086	Yes	Yes
BeAR 517449	Yes	Not Seen

#### Phenotypic Characterization and Host Range.

Negev virus strain E0239 infected and produced extensive CPE on C6/36 and C7/10 cells 12 hpi (FIGS. 9A and 9B); however, no overt cytopathic effects were observed in vertebrate cell lines at either 37° C. or 28° C. up to 6-days-post-infection (data not shown). Negev virus E0239 formed 3-4 mm size plaques on C7/10 cells at 36 hpi (FIG. 9C).

Representative vertebrate [African green monkey kidney (Vero), hamster kidney (BHK-21), and human embryonic kidney (HEK-293)] as well as invertebrate [*Ae. albopictus* (C6/36 and C7/10), *An. albimanus*, *An. gambiae*, *Cx. tarsalis*, *P. papatasi*, and *D. melanogaster*] cell lines were used to determine the in vitro host range of Negev virus E0239. The cell-free supernatants of the infected cell lines were collected at 12, 24, 48, and 72 hpi and viral output was evaluated by plaque forming assay (pfu) on C7/10 cells. Negev virus E0239 failed to replicate in vertebrate cells at 37° C. (FIG. 10A) or 28° C. (data not shown), as mean replication titers remained steady or declined. However, these cell lines are permissive for replication with a wide range of arthropod-transmitted viruses (Karabatsos, 1985. International catalogue of arboviruses, including certain other viruses of vertebrates, 3rd ed. *American Society of Tropical Medicine and Hygiene*, San Antonio, Tex.). Mean replication titers of Negev virus showed significant differences in levels of replication in invertebrate cell lines (FIG. 10B). Mean replication titers peaked consistently at 24 hpi (FIG. 10B) and plateaued thereafter in all cell lines, except in *P. papatasi* where maximum titers were reached at 48 hpi and *An. gambiae* and *D. melanogaster* cells where mean replication titers remained steady or declined (FIG. 9B). Cell lysis was readily evident in *Aedes* (see FIG. 9B, depicting cytopathic effect (CPE) in C7/10 cells) and *Culex* (data not shown) cell lines at 12 hpi, whereas no overt CPE was observed in *An. albimanus* or *P. papatasi* cells at any time point (data not shown). The other 9 viruses included in this study demonstrated a similar phenotype, namely rapid growth and CPE in C6/36 cells, but no CPE in Vero or BHK-21 cells. Likewise, none of the viruses produced illness in newborn mice after intracerebral inoculation.

#### Mosquito Susceptibility Studies.

The inventors also investigated whether the prototype Negev virus E0329, could infect and disseminate after ingestion in two important mosquito vectors, *Ae. aegypti* and *Ae. albopictus*. As shown in Table 3, when *Ae. aegypti* ingested various concentrations of Negev virus, the level of infection varied in a dose dependent manner. At the highest dose of  $10^9$  pfu/mL, 91% of midguts were infected, decreasing to 57%

and 8%, for the  $10^7$  and  $10^5$  pfu/mL doses, respectively. Table 3 also shows that the percentage of mosquitoes with virus dissemination (total number of disseminated infections divided by total engorged mosquitoes) also decreased from 73% to 50% for the higher doses, to 0% for the lowest dose. Furthermore, dissemination rates from the infected midguts (total number of disseminated infections divided by total number of infected mosquitoes), ranged from 80% to 87.5% for the highest two doses, to 0% for the lowest dose. In contrast to *Ae. aegypti*, *Ae. albopictus* mosquitoes were relatively refractory to oral infection with Negev virus midgut and disseminated infection rates in the 5-6% range for all doses (Table 3). No mortality other than regularly observed attrition was observed in any of the mosquitoes during the 14-day incubation. Thus, infection with Negev virus did not appear to have a deleterious effect on the insects.

TABLE 3

Species	Dose (pfu)	% infected	% disseminated
<i>Aedes aegypti</i>	$10^9$	91 (20/22)	73 (16/22)
	$10^7$	57 (8/14)	50 (7/14)
	$10^5$	8 (2/25)	0 (0/25)
<i>Aedes albopictus</i>	$10^9$	5 (1/20)	5 (1/20)
	$10^7$	6 (2/35)	6 (2/35)

In addition to their broad geographic distribution, the Negev viruses appear to infect a wide range of hematophagous insects (mosquitoes of the genera *Culex*, *Aedes*, and *Anopheles* as well as sand flies of the genus *Lutzomyia*). The three isolates of Negev and of Loreto viruses were each made from pools of 3 different insect genera and/or species from two different localities. This suggests that these viruses are not species-specific and may have a broad host range among *Diptera*. All of the viruses described herein were obtained from hematophagous insects collected during arbovirus surveillance studies.

Recombinant Negevirus can be produced (1) entirely using cDNAs or RNAs or (2) a combination of cDNAs transfected into a host cell, or (3) cDNAs or RNAs transfected into a cell, which is further infected with an expression vector providing in trans supplemental components or activities needed to produce either an infectious or non-infectious (e.g., virus-like particle) recombinant Negevirus. In certain embodiments, RNA can also be used to produce a Negevirus. Using any of these methods, the minimum components required are an RNA molecule containing the cis-acting signals for (1) encapsidation of the genomic RNA by the Negevirus or other structural protein(s), and (2) replication of the genomic RNA.

For any gene or heterologous nucleic acid contained within the engineered Negevirus genome, the gene would be flanked by the appropriate transcription initiation and termination signals that will allow expression of those genes and production of protein products. Particularly a heterologous nucleic acid, which is a nucleic acid that is typically not encoded by a Negevirus as isolated from nature or contains a Negevirus coding region in a position, form, or context that it typically is not found, e.g., a chimeric protein. In certain aspects a heterologous nucleic acid can encode an insecticidal protein or nucleic acid.

To produce "non-infectious" engineered Negevirus, the engineered Negevirus must have the minimal replicon elements. This produces virus particles that are budded from the cell, but are non-infectious particles. To produce "infectious" particles, the virus particles must additionally comprise proteins that can mediate virus particle binding and fusion, such

as through the use of an attachment protein or receptor ligand, and proteins for encapsidation of the genome.

In certain aspects, the first step in generating a recombinant Negevirus is expression of an RNA that is a genomic equivalent from an expression vector, e.g., a cDNA. The RNA is packaged in a virus, which can be recovered. If any Negevirus protein is absent from the recombinant RNA genome, then it is typically supplied in trans.

For preparing "non-infectious Negevirus" particles, the procedure may be the same as above, except that the expression vector transfected into the cells would contain some but not all Negevirus genes. The expression vector of non-infectious Negevirus particles may additionally contain a gene encoding a reporter protein or a therapeutic/heterologous nucleic acid, e.g., an insecticidal nucleic acid.

Transfected cells are usually incubated for at least 24 hr at a desired temperature, usually about 37° C. For non-infectious virus particles, the supernatant is collected and the virus particles isolated. For infectious virus particles, the supernatant containing virus is harvested and transferred to fresh cells. The fresh cells are incubated for approximately 48 hours, and the supernatant is collected.

The virus or virus-like particle can then be isolated and/or purified. The terms "isolation" or "isolating" a Negevirus means the process of culturing and purifying the virus particles such that very little cellular debris remains. One example would be to take the virion containing supernatant and pass them through a 0.1-0.2 micron pore size filter (e.g., Millex-GS, Millipore) to remove the virus and cellular debris. Alternatively, virions can be purified using a gradient, such as a sucrose gradient. Recombinant Negevirus particles can then be pelleted and suspended in whatever excipient or carrier is desired. Titers can be determined by indirect immunofluorescence using antibodies specific for particular proteins.

A "suitable cell" or "host cell" means any cell that would permit assembly of the recombinant Negevirus. Methods for identifying such a cell are described herein. The terms "cell," "cell line," and "cell culture" may be used interchangeably. All of these terms also include their progeny, which is any and all subsequent generations. It is understood that all progeny may not be identical due to deliberate or inadvertent mutations. In the context of expressing a heterologous nucleic acid sequence, "host cell" refers to a prokaryotic or eukaryotic cell, and it includes any transformable organism that is capable of replicating a vector and/or expressing a heterologous gene encoded by a vector. A host cell can, and has been, used as a recipient for vectors or viruses. A host cell may be "transfected" or "transformed," which refers to a process by which exogenous nucleic acid, such as a modified protein-encoding sequence, is transferred or introduced into the host cell. A transformed cell includes the primary subject cell and its progeny.

Host cells may be derived from prokaryotes or eukaryotes, including yeast cells, insect cells, and mammalian cells, depending upon whether the desired result is replication of the vector or expression of part or all of the vector-encoded nucleic acid sequences. Numerous cell lines and cultures are available for use as a host cell, and they can be obtained through the American Type Culture Collection (ATCC), which is an organization that serves as an archive for living cultures and genetic materials ([www.atcc.org](http://www.atcc.org)). An appropriate host can be determined by one of skill in the art based on the vector backbone and the desired result. A plasmid or cosmid, for example, can be introduced into a prokaryote host cell for replication of many vectors. Many host cells from various cell types and organisms are available and would be known to one of skill in the art. Similarly, a viral vector may

be used in conjunction with either a eukaryotic or prokaryotic host cell, particularly one that is permissive for replication or expression of the vector.

Host and host cells include but are not limited to mosquitoes and sandflies. The term "mosquito" as used herein concerns any type of mosquito (e.g., *Anopheles*, *Aedes*, and *Culex*), including but not limited to members of the group Anophelinae (*Anopheles*, *Bironella*, and *Chagasia*) and Culicinae (*Aedeomyia*, *Aedes*, *Armigeres*, *Ayurakitia*, *Borachinda*, *Coquillettia*, *Culex*, *Culiseta*, *Deinocerites*, *Eretmapodites*, *Ficalbia*, *Galindomyia*, *Haemagogus*, *Heizmannia*, *Hodgesia*, *Isostomyia*, *Johnbelkinia*, *Kimia*, *Limatus*, *Lutzia*, *Malaya*, *Mansonia*, *Maorigoeldia*, *Mimomyia*, *Onirion*, *Opifex*, *Orthopodomyia*, *Psorophora*, *Runchomyia*, *Sabethes*, *Shannoniana*, *Topomyia*, *Toxorhynchites*, *Trichoprosopon*, *Tripteroides*, *Udaya*, *Uranotaenia*, *Verrallina*), as well as cells thereof. In certain aspects the host or host cell is Tiger mosquitoes, *Aedes albopictus* (e.g., C6/36 and C7/10 cells), AP-61 line of *Aedes pseudoscutellaris* (AP-61 cells), *Aedes aegypti*, *Aedes aboriginis*, *Aedes cantator*, *Aedes sierrensis*, *Aedes sollicitans*, *Aedes squamiger*, *Aedes sticticus*, *Aedes vexans*, *Culex coronator*, *Culex pipiens*, *Culex quinquefasciatus*, *Culex vishnui*, *Culex tarsalis*, *Anopheles coustani*, *Anopheles albimanus*, *Anopheles gambiae*, and *Anopheles quadrimaculatus*.

Sandfly is a colloquial name for any species or genus of flying, biting, blood-sucking Dipteran encountered in sandy areas. Sandflies belong to the fly family Psychodidae. Sandflies include, but are not limited to members of the *Brumptomyia*, *Lutzomyia*, *Warileya*, *Phlebotomus*, *Sergentomyia*, or *Sauroleishmania* genus or cell derived therefrom. In certain aspects *Phlebotomus papatasi* cells are used as a host cell.

## II. Negevirus Nucleic Acids

The present invention includes polynucleotides isolatable from cells or viruses that are capable of expressing all or part of a Negevirus, Negevirus nucleic acid, and/or a Negevirus protein or polypeptide. Some embodiments concern all or parts of a viral genome that has been specifically mutated or altered to generate a virus or viral polypeptide. The polynucleotides may encode a peptide or polypeptide containing all or part of a viral and/or heterologous amino acid sequence, e.g., a therapeutic or insecticidal polypeptide. Recombinant proteins can be purified from cells expressing nucleic acids described herein.

As used herein, the term "RNA, DNA, or nucleic acid segment" refers to a RNA, DNA, or nucleic acid molecule that has been isolated from host cell RNA or DNA or other contaminants. Included within the term "nucleic acid" are polynucleotides and recombinant vectors, including, for example, plasmids, viruses, and the like.

As used in this application, the term "Negevirus polynucleotide" refers to any member of the Negevirus group as defined herein. The term "Negev virus" refers to a particular subset (e.g., species or variant) of Negevirus. Other members Negevirus include Piura virus, Loreto virus, Dezidougou virus, Santana virus, and Ngewotan virus.

The term "cDNA" is intended to refer to DNA prepared using RNA as a template. There may be times when the full or partial viral genome is prepared as a cDNA. As will be understood by those in the art, the term nucleic acid includes genomic sequences, cDNA sequences, and smaller engineered nucleic acid segments that express, or may be adapted to express, proteins, polypeptides, domains, peptides, fusion proteins, and variants. A nucleic acid encoding all or part of a native or modified polypeptide may contain a contiguous

nucleic acid of: 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 441, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, 1010, 1020, 1030, 1040, 1050, 1060, 1070, 1080, 1090, 1095, 1100, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 9000, 10000, or more nucleotides, nucleosides, or base pairs corresponding to SEQ ID NO:3, 6, 7, 8, 11, 14, 15, 16, or 19.

The nucleic acid segments used in the present invention, regardless of the length of the coding sequence, may be combined with other nucleic acid sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant nucleic acid protocol.

A nucleic acid sequence may encode a full-length polypeptide sequence with additional heterologous coding sequences, for example, to allow for purification of the polypeptide, transport, secretion, and/or post-translational modification. A tag or other heterologous polypeptide may be added to a polypeptide-encoding sequence, wherein "heterologous" refers to a polynucleotide or encoded polypeptide or segment thereof that is not typically found associated with or encoded by the naturally occurring virus.

In a non-limiting example, one or more nucleic acid construct may be prepared that include a contiguous stretch of nucleotides identical to or complementary to a particular Negevirus viral segment, such as a Negevirus ORF1, ORF2, and/or ORF3. A nucleic acid construct may be at least 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 15000, or 20000 nucleotides in length (including all intermediate lengths and intermediate ranges). It will be readily understood that "intermediate lengths" and "intermediate ranges," as used herein, means any length or range including or between the quoted values (i.e., all integers including and between such values).

The nucleic acid segments used in the present invention encompass nucleic acids that are 70, 75, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% identical (including all values and ranges there between) to SEQ ID NO:2, 3, 6, 7, 8, 11, 14, 15, 16, or 19. Such sequences may arise as a consequence of codon redundancy and functional equivalency that are known to occur naturally within nucleic acid sequences and the proteins thus encoded. Alternatively, functionally equivalent proteins or peptides may be created via the application of recombinant DNA technology, in which changes in the protein structure may be engineered, based on considerations of the properties of the amino acids being exchanged. Changes designed by human may be introduced through the application of site-directed mutagenesis techniques, e.g., to introduce improvements to the antigenicity of a protein, to modulate toxic effects, or to increase the efficacy of any treatment involving the protein or a virus comprising such protein.

It also will be understood that this invention is not limited to the particular nucleic acid and amino acid sequences of these identified sequences. Recombinant vectors and isolated nucleic acid segments may therefore variously include

Negevirus-coding regions themselves, coding regions bearing selected alterations or modifications in the basic coding region, or they may encode larger polypeptides that nevertheless include Negevirus-coding regions, or may encode biologically functional equivalent proteins or peptides that have variant amino acids sequences.

#### A. Mutagenesis of Negevirus Polynucleotides

In various embodiments, the Negevirus polynucleotide may be altered or mutagenized. Alterations or mutations may include insertions, deletions, point mutations, inversions, and the like and may result in the modulation, activation and/or inactivation of certain proteins or molecular mechanisms, as well as altering the function, location, or expression of a gene product, in particular rendering a gene product non-functional. Where employed, mutagenesis of a polynucleotide encoding all or part of a Negevirus may be accomplished by a variety of standard, mutagenic procedures (Sambrook et al., 2001). Mutation is the process whereby changes occur in the quantity or structure of an organism. Mutation can involve modification of the nucleotide sequence of a single gene, blocks of genes or whole genomes. Changes in single genes may be the consequence of point mutations which involve the removal, addition or substitution of a single nucleotide base within a DNA sequence, or they may be the consequence of changes involving the insertion or deletion of large numbers of nucleotides.

Insertional mutagenesis is based on the inactivation of a gene via insertion of a known nucleic acid fragment. Because it involves the insertion of some type of nucleic acid fragment, the mutations generated are generally loss-of-function, rather than gain-of-function mutations. However, there are several examples of insertions generating gain-of-function mutations. Insertional mutagenesis may be accomplished using standard molecular biology techniques.

Chemical mutagenesis offers certain advantages, such as the ability to find a full range of mutations with degrees of phenotypic severity, and is facile and inexpensive to perform.

Radiation mutagenesis is the result of biological molecules being degraded by ionizing radiation. Ionizing radiation causes DNA damage, generally proportional to the dose rate.

Random mutagenesis may be introduced using error prone PCR. The rate of mutagenesis may be increased by performing PCR in multiple tubes with dilutions of templates. One particularly useful mutagenesis technique is alanine scanning mutagenesis in which a number of residues are substituted individually with the amino acid alanine so that the effects of losing side-chain interactions can be determined, while minimizing the risk of large-scale perturbations in protein conformation.

In vitro scanning saturation mutagenesis provides a rapid method for obtaining a large amount of structure-function information including: (i) identification of residues that modulate ligand binding specificity, (ii) a better understanding of ligand binding based on the identification of those amino acids that retain activity and those that abolish activity at a given location, (iii) an evaluation of the overall plasticity of an active site or protein subdomain, (iv) identification of amino acid substitutions that result in increased binding.

Polynucleotide may also be modified using site-directed mutagenesis. Structure-guided site-specific mutagenesis represents a powerful tool for the dissection and engineering of protein-ligand interactions (Wells, 1996; Braisted et al., 1996). The technique provides for the preparation and testing of sequence variants by introducing one or more nucleotide sequence changes into a selected DNA.

## B. Vectors

A Negevirus genome, or segment thereof, can be encoded by a nucleic acid vector. The term "vector" is used to refer to a carrier nucleic acid molecule into which an exogenous nucleic acid sequence can be inserted for introduction into a cell where it can be replicated. Vectors include plasmids, cosmids, viruses (bacteriophage, animal viruses, and plant viruses), and artificial chromosomes (e.g., YACs).

A vector can encode modified or non-modified polypeptide sequences comprising heterologous sequences such as a tag or targeting molecule. Useful vectors encoding such fusion proteins include pIN vectors, vectors encoding a stretch of histidines; and pGEX vectors, for use in generating glutathione S-transferase (GST) soluble fusion proteins for later purification and separation or cleavage. A targeting molecule is one that directs the modified polypeptide to a particular organ, tissue, cell, or other location in a subject's body. Alternatively, the targeting molecule alters the tropism of an organism, such as Negevirus for certain cell types.

The term "expression vector" refers to a vector containing a nucleic acid sequence coding for at least part of a gene product capable of being transcribed. In some cases, RNA molecules are translated into a protein, polypeptide, or peptide. In other cases, these sequences are not translated, for example, in the production of antisense molecules or ribozymes. Expression vectors can contain a variety of "control sequences," which refer to nucleic acid sequences necessary for the transcription and possibly translation of an operably linked coding sequence in a particular host organism. In addition to control sequences that govern transcription and translation, vectors and expression vectors may contain nucleic acid sequences that serve other functions as well.

An expression vector typically includes a promoter and/or enhancer. A "promoter" is a control sequence that is a region of a nucleic acid sequence at which initiation and rate of transcription are controlled. It may contain genetic elements that bind regulatory proteins and molecules, such as RNA polymerase and other transcription factors. The phrases "operatively positioned," "operatively coupled," "operatively linked," "under control," and "under transcriptional control" mean that a promoter is in a correct functional location and/or orientation in relation to a nucleic acid sequence to control transcriptional initiation and/or expression of that sequence. A promoter may or may not be used in conjunction with an "enhancer," which refers to a cis-acting regulatory sequence involved in the transcriptional activation of a nucleic acid sequence.

Other elements that can be included in an expression vector are initiation signals and internal ribosome binding sites. A specific initiation signal also may be required for efficient translation of coding sequences. These signals include the ATG initiation codon or adjacent sequences. Exogenous translational control signals, including the ATG initiation codon, may need to be provided. An exogenous translational control signals and initiation codons can be either natural or synthetic. In certain embodiments of the invention, the use of internal ribosome entry sites (IRES) elements are used to create multigene, or polycistronic, messages. IRES elements are able to bypass the ribosome-scanning model of 5' methylated Cap dependent translation and begin translation at internal sites. In certain aspects, the 5' UTR can be used as an IRES element.

The vectors or constructs of the present invention will generally comprise at least one termination signal. A "termination signal" or "terminator" is comprised of the RNA sequences involved in specific termination of an RNA transcript by an RNA polymerase. Thus, in certain embodiments

a termination signal that ends the production of an RNA transcript is contemplated. A terminator may be necessary in vivo to achieve desirable message levels.

In expression, particularly eukaryotic expression, one will typically include a polyadenylation signal to effect proper polyadenylation of the transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and/or any such sequence may be employed. Preferred embodiments include the SV40 polyadenylation signal and/or the bovine growth hormone polyadenylation signal, convenient and/or known to function well in various target cells. Polyadenylation may increase the stability of the transcript or may facilitate cytoplasmic transport.

In order to propagate a vector in a host cell, it may contain one or more origins of replication sites (often termed "ori"), which is a specific nucleic acid sequence at which replication is initiated. Alternatively an autonomously replicating sequence (ARS) can be employed if the host cell is yeast.

In certain embodiments of the invention, cells containing a nucleic acid construct of the present invention may be identified in vitro or in vivo by including a marker in the expression vector. Such markers would confer an identifiable change to the cell permitting easy identification of cells containing the expression vector. Generally, a selectable marker is one that confers a property that allows for selection. A positive selectable marker is one in which the presence of the marker allows for its selection, while a negative selectable marker is one in which its presence prevents its selection. An example of a positive selectable marker is a drug resistance marker.

Usually the inclusion of a drug selection marker aids in the cloning and identification of transformants, for example, genes that confer resistance to neomycin, puromycin, hygromycin, DHFR, GPT, zeocin and histidinol are useful selectable markers. In addition to markers conferring a phenotype that allows for the discrimination of transformants based on the implementation of conditions, other types of markers including screenable markers such as GFP, whose basis is colorimetric analysis, are also contemplated. Alternatively, screenable enzymes such as herpes simplex virus thymidine kinase (tk) or chloramphenicol acetyltransferase (CAT) may be utilized. One of skill in the art would also know how to employ immunologic markers, possibly in conjunction with FACS analysis. The marker used is not believed to be important, so long as it is capable of being expressed simultaneously with the nucleic acid encoding a gene product. Further examples of selectable and screenable markers are well known to one of skill in the art.

## C. Insecticidal Peptides and Polypeptides

In certain aspects a heterologous nucleic acid is toxic or encodes an insecticidal polypeptide/peptide. Insecticidal peptides include toxins such as the omega-ACTX-1 family of insecticidal toxins (e.g., omega-ACTX-Hv1a), *Bacillus thuringiensis* toxins, neurotoxins, juvenile hormone esterases and variants thereof. In certain aspects the heterologous nucleic acid can encode a polypeptide that disrupts insect propagation, e.g., oostatic peptides such as trypsin modulating oostatic factor (TMOF) and neuropeptide F (NPF)) and variants thereof. Various examples of suitable Insecticidal peptides are discussed in greater detail below.

## Juvenile Hormone Esterases.

The term "juvenile hormone esters" as used herein includes mutants or analogs of the naturally occurring proteins. Several useful JHE mutants are described in PCT Application WO 94/03588 published Feb. 17, 1994. Juvenile hormone esters (JHE) are insect proteins, which appears at critical times in the insect's life. It appears to present no risk

to other groups of organisms. It is nonlethal to an individual cell, which allows and perhaps encourages viral replication; yet the enzyme will fatally disrupt the normal development of the organism. U.S. Pat. No. 5,098,706 provides an example of the administration of an affinity purified JHE enzyme to insects, which results in anti juvenile hormone activity. Such anti-juvenile hormone activity is effectively lethal.

#### *Bacillus thuringiensis* Insecticidal Peptides.

Insecticidal peptides or toxins obtained from any of a variety of *Bacillus thuringiensis* strains can be used in certain aspects. These toxins are well known. For example, U.S. Pat. Nos. 4,448,885 and 4,467,036 describe the expression of *Bacillus thuringiensis* crystal protein in *Escherichia coli*.

#### Insect Toxins.

Numerous insect toxins may be used to carry out the present invention, some of which are described in U.S. Pat. No. 6,162,430. For example, the insect toxin may be a neurotoxin derived from or similar to an arthropod or other invertebrate toxin, such as a scorpion toxin, a wasp toxin, a snail toxin, a mite toxin, or a spider toxin. A useful scorpion toxin is, for example, AaIT from *Androctonus australis*. A useful snail venom is that from the snail *Conus querciones* (Olivera et al., (1990) *Science*, 249:257-263).

Yet another suitable toxin affects insect sodium channels in a manner very similar to the effect of  $\alpha$ -toxins on mammalian sodium channels. This neurotoxin was derived from a yellow scorpion *Leirus quinquestriatus hebraeus*, LqhoAIT and LqhlT2 (Eitan et al., (1990) *Biochemistry*, 29:5941-5947).

#### TMOF and TMOF Analogs.

Serine esterases such as trypsin and trypsin-like enzymes (collectively referred to herein as "TTLE") are important components of the digestion of proteins by insects. In the mosquito, *Aedes aegypti*, an early trypsin that is found in the midgut of newly emerged females is replaced, following the blood meal, by a late trypsin. A female mosquito typically weighs about 2 mg and produces 4 to 6  $\mu$ g of trypsin within several hours after ingesting a blood meal. Continuous biosynthesis at this rate would exhaust the available metabolic energy of a female mosquito; as a result, the mosquito would be unable to produce mature eggs, or even to find an oviposition site. To conserve metabolic energy, the mosquito regulates TTLE biosynthesis with a peptide hormone named Trypsin Modulating Oostatic Factor (TMOF). Mosquitoes produce TMOF in the follicular epithelium of the ovary 12-35 hours after a blood meal; TMOF is then released into the hemolymph where it binds to a specific receptor on the midgut epithelial cells, signaling the termination of TTLE biosynthesis. This regulatory mechanism is not unique for mosquitoes; flesh flies, fleas, sand flies, house flies, dog flies and other insect pests which need protein as part of their diet have similar regulatory mechanisms. PCT Patent Application WO 00/63233, published Oct. 26, 2000, discloses examples of TMOF analogs.

#### D. Immunogenic Peptides and Polypeptides

Certain embodiments are directed to recombinant Negevirus compositions that can induce an immune response when administered to a subject. An immunogenic composition can comprise a replication-competent, attenuated, recombinant Negevirus or a vector encoding the same. In certain embodiments, the immunogenic composition contains a recombinant Negevirus described herein. Certain aspects are directed to a recombinant Negevirus as described herein. In a further aspect a recombinant Negevirus comprises a heterologous nucleic acid encoding one or more antigens that stimulate an immune response when administered to a subject. In certain embodiments a heterologous nucleic acid encodes a heterologous antigen. As used herein, the term "antigen" or "targeted

antigen" refers to a substance that is capable of being the target of an immune response. An antigen may be the target of, for example, a cell-mediated and/or humoral immune response of a subject administered or provided an immunogenic composition described herein. The term "antigen" or "targeted antigen" encompasses includes, but is not limited to all or part of viral antigens, tumor-specific or -related antigens, bacterial antigens, parasitic antigens, allergens, and the like.

#### E. Nucleic Acid Detection

In addition to their use in directing the expression of Negevirus proteins, polypeptides and/or peptides, the nucleic acid sequences disclosed herein have a variety of other uses. For example, they have utility as probes or primers for embodiments involving nucleic acid hybridization or amplification. They may be used in diagnostic or screening methods. Detection of nucleic acids encoding Negevirus or Negevirus polypeptides are encompassed by the invention.

The use of a probe or primer of between 13 and 100 nucleotides, preferably between 17 and 100 nucleotides in length, or in some aspects of the invention up to 1-2 kilobases or more in length, allows the formation of a duplex molecule that is both stable and selective. Molecules having complementary sequences over contiguous stretches greater than 20 bases in length are generally preferred, to increase stability and/or selectivity of the hybrid molecules obtained. One will generally prefer to design nucleic acid molecules for hybridization having one or more complementary sequences of 20 to 30 nucleotides, or even longer where desired. Such fragments may be readily prepared, for example, by directly synthesizing the fragment by chemical means or by introducing selected sequences into recombinant vectors for recombinant production.

Accordingly, the nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of DNAs and/or RNAs or to provide primers for amplification of DNA or RNA from samples. Depending on the application envisioned, one would desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of the probe or primers for the target sequence.

In certain embodiments, it will be advantageous to employ nucleic acids of defined sequences of the present invention in combination with an appropriate means, such as a label, for determining hybridization. A wide variety of appropriate indicator means are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of being detected. In certain embodiments, one may desire to employ a fluorescent label or an enzyme tag such as urease, alkaline phosphatase or peroxidase, instead of radioactive or other environmentally undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known that can be employed to provide a detection means that is visibly or spectrophotometrically detectable, to identify specific hybridization with complementary nucleic acid containing samples.

The term "primer," as used herein, is meant to encompass any nucleic acid that is capable of priming the synthesis of a nascent nucleic acid in a template-dependent process. Typically, primers are oligonucleotides from ten to twenty and/or thirty base pairs in length, but longer sequences can be employed. Primers may be provided in double-stranded and/or single-stranded form, although the single-stranded form is preferred.

A number of template dependent processes are available to amplify the oligonucleotide sequences present in a given template sample. One of the best-known amplification meth-

ods is the polymerase chain reaction (referred to as PCR™), which is described in detail in U.S. Pat. Nos. 4,683,195, 4,683,202 and 4,800,159, each of which is incorporated herein by reference in their entirety.

Nucleic acids can be visualized using a variety of methods, including staining of a gel with ethidium bromide and visualization of bands under UV light. Alternatively, if the amplification products are integrally labeled with radio- or fluorometrically-labeled nucleotides, the separated amplification products can be exposed to x-ray film or visualized under the appropriate excitatory spectra.

In particular embodiments, detection is by Southern blotting and hybridization with a labeled probe. The techniques involved in Southern blotting are well known to those of skill in the art. One example of the foregoing is described in U.S. Pat. No. 5,279,721, incorporated by reference herein, which discloses an apparatus and method for the automated electrophoresis and transfer of nucleic acids. Other methods of nucleic acid detection that may be used in the practice of the instant invention are disclosed in U.S. Pat. Nos. 5,840,873, 5,843,640, 5,843,651, 5,846,708, 5,846,717, 5,846,726, 5,846,729, 5,849,487, 5,853,990, 5,853,992, 5,853,993, 5,856,092, 5,861,244, 5,863,732, 5,863,753, 5,866,331, 5,905,024, 5,910,407, 5,912,124, 5,912,145, 5,919,630, 5,925,517, 5,928,862, 5,928,869, 5,929,227, 5,932,413 and 5,935,791, each of which is incorporated herein by reference.

### III. Negevirus Polypeptides

Certain embodiments are directed to Negevirus polypeptides. Negevirus polypeptides include the encoded polyproteins and polypeptides as well as segments thereof. Negevirus polypeptides include those having the amino acid sequences set forth in SEQ ID NO:5, 10, 13, 18, or 21. In certain aspects, a Negevirus polypeptide is a variant of one or more of the polypeptides corresponding to SEQ ID NO:5, 10, 13, 18, or 21. In certain aspects of the invention these viral polypeptides and/or variants thereof can be comprised in a proteinaceous composition. Proteinaceous compositions include viral particles, virus-like particles, immunostimulatory compositions, and other compositions having one or more viral polypeptide components. Polypeptide variant(s) can be engineered or selected for a modification in one or more physiological or biological characteristics, such as host cell range, host cell specificity, toxicity to non-target cells or organs, replication, cytotoxicity to a target cell, infectivity, manufacturing parameters, size of virus particle, stability of viral particles, in vivo clearance, immunoreactivity, and the like. These polypeptide variants can be engineered by using a variety of methodologies known in the art, including various mutagenesis techniques described see below.

When the present application refers to the function or activity of viral proteins or polypeptides, it is meant to refer to the activity or function of that viral protein or polypeptide under physiological conditions, unless otherwise specified. Determination of which molecules possess certain activity may be achieved using assays familiar to those of skill in the art, such as infectivity assays, protein binding assays, plaque assays and the like.

As used herein, a “protein” or “polypeptide” refers to a molecule comprising polymer of amino acid residues. In some embodiments, a wild-type version of a protein or polypeptide are employed, however, in many embodiments of the invention, all or part of a viral protein or polypeptide is absent or altered so as to render the virus more suitable for a particular use. The terms described above may be used interchangeably herein.

A “modified protein” or “modified polypeptide” or “variant protein” or “variant polypeptide” refers to a protein or polypeptide whose chemical structure or amino acid sequence is altered with respect to the wild-type or a reference protein or polypeptide. In some embodiments, a modified protein or polypeptide has at least one modified activity or function (recognizing that proteins or polypeptides may have multiple activities or functions). The modified activity or function may be reduced, diminished, eliminated, enhanced, improved, or altered in some other way (such as infection specificity) with respect to that activity or function in a wild-type protein or polypeptide, or the characteristics of virus containing such a polypeptide. It is contemplated that a modified protein or polypeptide may be altered with respect to one activity or function yet retain unaltered activity or function in other respects. Alternatively, a modified protein may be completely nonfunctional or its cognate nucleic acid sequence may have been altered so that the polypeptide is no longer expressed at all, is truncated, or expresses a different amino acid sequence as a result of a frameshift, fusion, or other modification.

In certain embodiments the size of a recombinant protein or polypeptide may comprise, but is not limited to, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, 700, 725, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975, 1000, 1100, 1200, 1300, 1400, 1500, 1750, 2000, 2250, 2500 or greater amino molecule residues, and any range derivable therein, of a polypeptide corresponding to SEQ ID NO:5, 10, 13, 18, or 21. It is contemplated that polypeptides may be modified by truncation, rendering them shorter than their corresponding unaltered form or by fusion or domain shuffling which may render the altered protein longer.

The term “proteinaceous composition” encompasses amino molecule sequences comprising at least one of the 20 common amino acids in naturally synthesized proteins, or at least one modified or unusual amino acid. Proteinaceous compositions may be made by any technique known to those of skill in the art, including the expression of proteins, polypeptides, or peptides through standard molecular biological techniques, or the chemical synthesis of proteinaceous materials.

Amino acid sequence variants of the polypeptides of the present invention can be substitutional, insertional or deletion variants. A mutation in a gene encoding a viral polypeptide may affect 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500 or more non-contiguous or contiguous amino acids (i.e., segment) of a polypeptide, as compared to a wild-type or unaltered polypeptide or other reference polypeptides. Various polypeptides encoded by Negevirus may be identified by reference to and alignment with GenBank Accession Numbers and the related public database entries.

Deletion variants lack one or more residues of the native, unaltered, or wild-type protein SEQ ID NO:5, 10, 13, 18, or

21. Individual residues can be deleted, or all or part of a domain (such as a catalytic or binding domain) can be removed. A stop codon can be introduced (by substitution or insertion) into a nucleic acid sequence to generate a truncated protein.

An insertion mutation typically involves the addition of material at a non-terminal position in a polypeptide. A chimeric polypeptide is a particular type of insertion mutant that include homologous or similar portions of a related protein in place of the related portion of a target protein. This may include the insertion of an immunoreactive epitope or simply one or more residues. Terminal additions, typically called fusion proteins, may also be generated.

Substitution variants typically contain the exchange of one amino acid for another at one or more sites within the protein, and may be designed to modulate one or more properties of the polypeptide, with or without the loss of other functions or properties. Substitutions may be conservative, that is, one amino acid is replaced with one of similar shape and charge. Conservative substitutions are well known in the art and include, for example, the changes of: alanine to serine; arginine to lysine; asparagine to glutamine or histidine; aspartate to glutamate; cysteine to serine; glutamine to asparagine; glutamate to aspartate; glycine to proline; histidine to asparagine or glutamine; isoleucine to leucine or valine; leucine to valine or isoleucine; lysine to arginine; methionine to leucine or isoleucine; phenylalanine to tyrosine, leucine or methionine; serine to threonine; threonine to serine; tryptophan to tyrosine; tyrosine to tryptophan or phenylalanine; and valine to isoleucine or leucine. Alternatively, substitutions may be non-conservative such that a function or activity of the polypeptide is affected. Non-conservative changes typically involve substituting a residue with one that is chemically dissimilar, such as a polar or charged amino acid for a non-polar or uncharged amino acid, and vice versa.

The term "functionally equivalent codon" is used herein to refer to codons that encode the same amino acid, such as the six codons for arginine or serine, and also refers to codons that encode biologically equivalent amino acids (Amino Acids/Codons): Alanine (Ala)/GCA, GCC, GCG, or GCU; Cysteine (Cys)/UGC or UGU; Aspartic acid (Asp)/GAC or GAU; Glutamic acid (Glu)/GAA or GAG; Phenylalanine (Phe)/UUC or UUU; Glycine (Gly)/GGA, GGC, GGG, or GGU; Histidine (His)/CAC or CAU; Isoleucine (Ile)/AUA, AUC, or AUU; Lysine (Lys)/AAA or AAG; Leucine (Leu)/UUA, UUG, CUA, CUC, CUG, or CUU; Methionine (Met)/AUG; Asparagine (Asn)/AAC or AAU; Proline (Pro)/CCA, CCC, CCG, or CCU; Glutamine (Gln)/CAA or CAG; Arginine (Arg)/AGA, AGG, CGA, CGC, CGG, or CGU; Serine (Ser)/AGC, AGU, UCA, UCC, UCG, or UCU; Threonine (Thr)/ACA, ACC, ACG, or ACU; Valine (Val)/GUA, GUC, GUG, or GUU; Tryptophan (Trp)/UGG; and Tyrosine (Tyr)/UAC or UAU.

It also will be understood that amino acid and nucleic acid sequences may include additional residues, such as additional N- or C-terminal amino acids or 5' or 3' sequences, and yet still be essentially as set as forth herein, including having a certain biological activity.

The following is a discussion based upon changing of the amino acids of a Negevirus protein to create an equivalent, or even an improved, molecule. For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activ-

ity, certain amino acid substitutions can be made in a protein sequence, and in its underlying DNA coding sequence, and nevertheless produce a protein with like properties. It is thus contemplated by the inventors that various changes may be made in the nucleic acid sequences of Negeviruses without appreciable loss of biological utility or activity of interest, as discussed below.

In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring a biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982). It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like.

#### IV. Antibodies

The isolated viruses can be used to produce or detect antibodies specific for one or more Negeviruses. Certain embodiments are directed to an isolated antibody that specifically binds a Negevirus polypeptide. In certain aspects the Negevirus specific antibody binds a segment of a polypeptide encoded by a Negevirus ORF1, ORF2, or ORF3. Again, any given antibody can recognize and bind one of a number of possible epitopes present in the polypeptide; thus only a unique portion of a polypeptide (having the epitope) needs to be present in an assay to determine if the antibody specifically binds the polypeptide.

By "an antibody that specifically binds" a Negevirus polypeptide or protein is meant an antibody that selectively binds to an epitope on any portion of the Negevirus peptide such that the antibody binds specifically to the corresponding Negevirus polypeptide without significant background. Specific binding by an antibody further means that the antibody can be used to selectively remove the target polypeptide from a sample comprising the polypeptide or and can readily be determined by radioimmunoassay (RIA), bioassay, or enzyme-linked immunosorbent (ELISA) technology. An ELISA method effective for the detection of the specific antibody-antigen binding can, for example, be as follows: (1) bind the antibody to a substrate; (2) contact the bound antibody with a sample containing the antigen; (3) contact the above with a secondary antibody bound to a detectable moiety (e.g., horseradish peroxidase enzyme or alkaline phosphatase enzyme); (4) contact the above with the substrate for the enzyme; (5) contact the above with a color reagent; (6) observe the color change.

An antibody can include antibody fragments such as Fab fragments, which retain the binding activity. Antibodies can be made as described in, e.g., Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1988). Briefly, purified antigen can be injected into an animal in an amount and in intervals sufficient to elicit an immune response. Antibodies can either be purified directly, or spleen cells can be obtained from the animal. The cells are then fused with an immortal cell line and screened for antibody secretion. Individual hybridomas are then propagated as individual clones serving as a source for a particular monoclonal antibody.

Optionally, the antibody is a monoclonal antibody. The term monoclonal antibody as used herein refers to an antibody from a substantially homogeneous population of antibodies, i.e., the individual antibodies comprising the population are identical except for possible naturally occurring mutations that may be present in minor amounts. Monoclonal antibodies may be prepared using hybridoma methods, such as those described by Kohler and Milstein, *Nature*, 256:495 (1975) or Harlow and Lane, *Antibodies, A Laboratory*



Manual, Cold Spring Harbor Publications, New York (1988). In a hybridoma method, a mouse or other appropriate host animal is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes may be immunized in vitro.

Additionally provided is a method of screening a cell for infectivity by Negevirus, comprising contacting the cell with Negevirus and detecting the presence of Negevirus in the cells. Negevirus particles can be detected using any standard physical or biochemical methods. For example, physical methods that can be used for this detection include DNA based methods such as (1) polymerase chain reaction (PCR) for viral RNA or (2) direct hybridization with labeled probes, and immunological methods such as by (3) antibody directed against the viral structural or non-structural proteins. Reporter genes can also be utilized to detect cells that transduce Negevirus. For example,  $\beta$ -gal, green fluorescent protein or luciferase can be inserted into a recombinant Negevirus. The cell can then be contacted with the recombinant Negevirus, either in vitro or in vivo and a colorimetric assay could detect a color change in the cells that would indicate transduction of Negevirus in the cell. Additional detection methods are outlined in Fields, Virology, Raven Press, New York, N.Y. 1996.

#### V. Kits

Also provided are kits for performing nucleic acid assays and/or immunoassays as described herein. The kits comprise a Negevirus specific reagent. The Negevirus specific reagent can be a polynucleotide, such as a nucleic acid primer or probe; or a polypeptide, peptide or antibody. The kit can further comprise an assay substrate (e.g., a plate, a membrane, and a well) and/or other reagents. Optionally, the kit can further comprise a control sample.

Optionally, the kit can comprise isolated antibodies described herein. The kit can further comprise a Negevirus specific antibody. Optionally, the kit can comprise an assay substrate. Optionally, the kit can further comprise a control sample. The control sample can comprise one or more Negevirus.

#### VI. Examples

The following examples as well as the figures are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples or figures represent techniques discovered by the inventors to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

##### Example 1

##### Negevirus

#### A Proposed New Taxon of Insect-Specific Viruses with Wide Geographic Distribution

##### A. Materials and Methods Viruses.

All viruses used in this study were obtained from the World Reference Center for Emerging Viruses and Arboviruses

(WRCEVA) at the University of Texas Medical Branch. The proposed names, original sources and geographic origins of the viruses included in the study are listed in Table 1.

Negev (NEGV) strain E0329 (SEQ ID NO:7), the prototype strain of Negev virus, was initially isolated from a pool of *Anopheles coustani* mosquitoes collected in the Negev Desert, Israel in 1983.

Negev strains M30957 (SEQ ID NO:3) and M33056 (SEQ ID NO:6) were isolated at the WRCEVA from pools of *Culex coronator* and *Cx. quinquefasciatus* mosquitoes, respectively, collected in Houston, Harris County, Tex. in 2008.

Piura (PIUV) strain P60 (SEQ ID NO:8) was isolated at the Naval Medical Research Unit #6, Lima, Peru from a pool of *Culex* species mosquitoes collected in Piura, Peru in 1996.

Loreto (LORV) strain 3940-83 (SEQ ID NO:11) was isolated from a pool of *Anopheles albimanus* mosquitoes collected in Lima, Peru in 1983.

Loreto strain PeAR 2612/77 (SEQ ID NO: 15) was isolated from a pool of *Culex* spp. mosquitoes collected in Iquitos, Loreto, Peru in 1977.

Loreto strain 2617/77 (SEQ ID NO: 14) was isolated from a pool of phlebotomine sandflies (*Lutzomyia* spp.) collected in Iquitos, Peru in 1977.

Dezidougou (DEZV) strain ArA 20086 (SEQ ID NO: 16) was isolated at the Institute Pasteur, Dakar, Senegal from a pool of *Aedes aegypti* collected in Dezidougou, Ivory Coast in 1987.

Santana (SANV) strain BeAR 517449 (SEQ ID NO: 19) was isolated at the WRCEVA from a pool of *Culex* species mosquitoes originally collected in Santana, Amapa, Brazil in 1992.

Ngewotan (NWTV) strain JKT 9982 was isolated at the Naval Medical Research #2, Jakarta, Indonesia from a pool of *Cx. vishnui* collected at Wotan, Central Java, Indonesia in 1981.

All viruses were initially isolated from triturated pools of field-collected mosquitoes collected during arbovirus surveillance studies. The mosquito homogenates were inoculated into cultures of the C6/36 line of *Ae. albopictus* cells (Igarashi, 1978, *Journal of General Virology* 40:531-544) or the AP-61 line of *Ae. Pseudoscutellaris* (Varma et al., 1974, *Transactions of the Royal Society of Tropical Medicine and Hygiene* 68:374-382). After inoculation, cultures were maintained in incubators at a constant temperature of 28 C and observed at regular intervals for evidence of viral cytopathic effect (CPE)(Kim et al., 2009, *Virology* 386:154-159).

Before sequencing, all virus stocks were grown in cultures of the C6/36 clone of *Ae. albopictus* cells (Igarashi, 1978, *Journal of General Virology* 40:531-544), obtained from the American Type Culture Collection (ATCC), Manassas, Va. Infection was characterized by detachment of cells and cell lysis. Plaque assays were performed using the C7/10 (LTC-7) clone of *Ae. albopictus* cells (Sarver and Stollar, 1977, *Virology* 80:390-400).

Nucleotide Sequence Accession Numbers.

The genomic sequences of CiLV-C are available in GenBank Accession numbers NC 008169, DQ388512, and DQ157466, and were included in the phylogenetic analyses.

Cell Lines Utilized for Virus Replication Kinetics.

African green monkey kidney (Vero), baby hamster kidney (BHK-21), human embryonic kidney (HEK-293), *Drosophila melanogaster* and *Ae. albopictus* (C6/36 and C7/10) cell lines were obtained from the ATCC. *Anopheles albimanus*, *An. gambiae*, *Cx. tarsalis* and *Phlebotomus papatasi* cells were obtained from the WRCEVA (Bello Garcia et al., 1995, *Memorias do Instituto Oswaldo Cruz* 90:547-551; Chao and Ball, 1976, p. 263-266. In Kurstak and Maramo-

rosch (ed.), *Invertebrate Tissue Culture Applications in Medicine, Biology and Agriculture*. Academic press, New York; Marhoul and Pudney, 1972, *Transactions of the Royal Society of Tropical Medicine and Hygiene* 66:183-184). Monolayers of Vero, BHK-21 and HEK-293 were grown at 37° C. in Dulbecco's minimal essential medium (DMEM) (4.5 g/L D-Glucose) with 10% heat-inactivated fetal bovine serum (FBS) and 1% penicillin/streptomycin. C6/36, C7/10, *An. albimanus*, *An. gambiae*, and *Cx. tarsalis* were grown at 28° C. in Dulbecco's minimal essential medium (DMEM) (4.5 g/L D-Glucose) with 10% heat-inactivated FBS, 1% penicillin/streptomycin and 1% tryptose phosphate broth (TPB). *Phlebotomus papatasi* cells were maintained in Schneider's medium (Sigma, St. Louis, Mo.) supplemented with 10% fetal bovine serum and Penicillin (100 units/mL)-Streptomycin (100 µg/ml). All 3 vertebrate and invertebrate cell lines were propagated in 6-well plates, and infected with a multiplicity of infection (MOI) of 10 in duplicate. Plates with vertebrate cells were incubated for one hour with periodic gentle rocking at 37° C., whereas plates containing invertebrate cells were incubated at 28° C. After three washes with PBS to remove unabsorbed virus, 2 ml of complete cell media were then added to each well, and plates were incubated at 28° C. or 37° C. for the invertebrate or vertebrate cell lines respectively. Virus from individual dishes was harvested at designated time-points for 3 days post infection (p.i.), clarified by low speed centrifugation, and the virus titer was determined by plaque assay in C7/10 cells. Virus yield at each time-point was recorded as PFU/cell, represented as the ratio of the total amount of virus present in the sample by the number of cells originally infected.

#### Virus Purification.

Virus purification was performed as previously described (Sherman and Weaver, 2010, *J Virol* 84:9775-9782). Virus was amplified on C7/10 cells at MOI of 0.5, harvested 48 hpi and clarified by centrifugation at 2,000 g for 10 min. Virus was precipitated overnight at 4° C. by adding polyethylene glycol and NaCl to 7% and 2.3% (wt/vol) concentrations, respectively. Virus was pelleted by centrifugation at 4,000×g for 30 min at 4° C., and the precipitate was then resuspended in TEN buffer (0.05 M Tris-HCl, pH 7.4, 0.1 M NaCl, 0.001 M EDTA) and loaded onto a 20-to-70% continuous sucrose (wt/vol) gradient in TEN buffer and centrifuged at 270,000 g for 1 hr. Following centrifugation, the visible virus band was harvested using a Pasteur pipette and centrifuged 4 times through an Amicon Ultra-4 100-kDa-cutoff filter (Millipore) and resuspended in 1 mL of TEN buffer.

#### Transmission Electron Microscopy.

For ultrastructural analysis, infected cells C6/36 were fixed for at least 1 hr in a mixture of 2.5% formaldehyde prepared from paraformaldehyde powder, and 0.1% glutaraldehyde in 0.05M cacodylate buffer pH 7.3 to which 0.03% picric acid and 0.03% CaCl<sub>2</sub> were added. The monolayers were washed in 0.1 M cacodylate buffer. Cells were scraped and processed further as a pellet. The pellets were post-fixed in 1% OsO<sub>4</sub> in 0.1M cacodylate buffer pH 7.3 for 1 hr, washed with distilled water, and en bloc stained with 2% aqueous uranyl acetate for 20 min at 60° C. The pellets were dehydrated in ethanol, processed through propylene oxide and embedded in Poly/Bed 812 (Polysciences, Warrington, Pa.). Ultrathin sections were cut on Leica EM UC7 ultramicrotome (Leica Microsystems, Buffalo Grove, Ill.), stained with lead citrate and examined in a Philips 201 transmission electron microscope at 60 kV.

Purified virus particles were also allowed to adhere to a Formvar carbon-coated copper grid for 10 min, negatively stained with either 2% aqueous uranyl acetate for 30 sec or

2% phosphotungstic acid with pH adjusted to 6.8 with 1N KOH (30 sec), and then examined in the electron microscope.

#### Plaque Assay.

Virus titrations were performed on confluent C7/10 cell monolayers in 6-well plates. Duplicate wells were inoculated with 0.1-ml aliquots of serial 10-fold dilutions of virus in growth medium. An additional 0.4 mL of growth media was added to each well to prevent cell desiccation, and virus was adsorbed for 2 hrs. Following incubation, the virus inoculum was removed by aspiration, and cell monolayers were overlaid with 3 mL of a medium 1:1 mixture of 2% tragacanth and 2×MEM with 5% FBS, 2% tryptose phosphate broth solution, 2% of a 100× solution of penicillin and streptomycin. Cells were incubated at 28° C. in 5% CO<sub>2</sub> for 2 days to allow plaque development and then the overlay was removed and monolayers were fixed with 3 mL of 10% formaldehyde in PBS for 30 mins. Cells were subsequently stained with 2% crystal violet in 30% methanol for 5 min at RT; excess stain was removed under running water and plaques were counted and recorded as the number of plaques per ml of inoculum.

#### Experimental Infection of Mosquitoes with Negev Virus.

Laboratory colonies of *Ae. aegypti* and *Ae. albopictus* were used for experimental infections. The progenitors of both colonies were originally collected in Thailand and had been maintained in an insectary for about 10 generations. Six-to-10 days after emergence, cohorts of 100 females of each species were allowed to feed on artificial blood meals containing three different concentrations (5, 7, and 9 log<sub>10</sub> PFU/ml) of Negev virus strain EO-239 made by serially diluting a virus stock of known titer in defibrinated sheep blood (Colorado Serum Company, Denver, Colo.) in MEM. Artificial blood meals were placed in vials covered with mouse skin and were warmed to 37° C. using a Hemotek feeder (Discovery Workshops, Accrington, UK). Mosquitoes were allowed to feed for one hour, and were then cold-anesthetized on ice for sorting. Engorged females were removed and placed in cages and maintained with 10% sucrose at 28° C. with a relative humidity of ~70%. Fourteen days after feeding, mosquitoes were cold-anesthetized, and the legs and wings were removed. Mosquito bodies and legs/wings were put in individual tubes containing 500 µL MEM with 10% FBS and a stainless steel bead for trituration. Each body and leg/wing sample was homogenized for 4 minutes using a Mixer Mill 300 (Retsch, Haan, Germany). Samples were centrifuged for 10 min at 5,000 rpm, and 100 ml of each sample supernatant were inoculated into individual 24-well plates containing C7/10 cells. Cultures were maintained with 2 ml of medium at 28° C. and 5% CO<sub>2</sub>. CPE observed in C7/10 cultures was used as a surrogate indicator for the presence of virus.

#### Genome Sequencing.

All virus sequences were obtained using 454 pyrosequencing (Roche Life Sciences, Branford, Conn.), except for the Dezidougou genomic sequences which were obtained by Illumina sequencing (Illumina, San Diego, Calif.).

#### Pyrosequencing.

RNA was extracted from virus stocks using TRIzol LS (Invitrogen, Carlsbad, Calif., USA) and treated with DNase I (DNA-Free, Ambion, Austin, Tex., USA). cDNA was generated using the Superscript II system (Invitrogen) employing random hexamers linked to an arbitrary 17-mer primer sequence (Palacios et al., 2007, *Emerg Infect Dis* 13:73-81). Resulting cDNA was treated with RNase H and then randomly amplified by PCR with a 9:1 mixture of primer corresponding to the 17-mer sequence and the random hexamer linked 17-mer primer (Palacios et al., 2007, *Emerg Infect Dis* 13:73-81). Products greater than 70 base pairs (bp) were selected by column chromatography (MinElute, Qiagen,

Hilden, Germany) and ligated to specific adapters for sequencing on the 454 Genome Sequencer FLX (454 Life Sciences, Branford, Conn., USA) without fragmentation (Cox-Foster et al., 2007, *Science* 318:283-287; Margulies et al., 2005, *Nature* 437:376-380; Palacios et al., 2008, *N Engl J Med* 358:991-998). Software programs accessible through the analysis applications at the GreenePortal website at URL [tako.cpmc.columbia.edu/Tools/](http://tako.cpmc.columbia.edu/Tools/) were used for removal of primer sequences, redundancy filtering, and sequence assembly. Sequence gaps were completed by RT-PCR amplification using primers based on pyrosequencing data. Amplification products were size-fractionated on 1% agarose gels, purified (MiniElute, Qiagen, Hilden, Germany), and directly sequenced in both directions with ABI PRISM Big Dye Terminator 1.1 Cycle Sequencing kits on ABI PRISM 3700 DNA Analyzers (Perkin-Elmer Applied Biosystems, Foster City, Calif.). The terminal sequences for each virus were amplified using the Clontech SMARTer RACE kit (Clontech, Mountain View, Calif., USA). Genome sequences were verified by Sanger dideoxy sequencing using primers designed from the draft sequence to create products of 1,000 bp with 500 bp overlaps.

#### Illumina Sequencing.

Viral RNA (0.1-0.2 µg) was fragmented by incubation at 94° C. for eight minutes in 19.5 µl of fragmentation buffer (Illumina 15016648). First and second strand synthesis, adapter ligation and amplification of the library were performed using the Illumina TruSeq RNA Sample Preparation kit under conditions prescribed by the manufacturer (Illumina, San Diego, Calif.). Samples were tracked using the “index tags” incorporated into the adapters as defined by the manufacturer. Cluster formation of the library DNA templates was performed using the TruSeq PE Cluster Kit v3 (Illumina, San Diego, Calif.) and the Illumina cBot workstation using conditions recommended by the manufacturer. Paired end 50 base sequencing by synthesis was performed using TruSeq SBS kit v3 (Illumina, San Diego, Calif.) on an Illumina HiSeq 1000 using protocols defined by the manufacturer. Cluster density per lane was 645-980 k/mm<sup>2</sup> and post filter reads ranged from 148-178 million per lane. Base call conversion to sequence reads was performed using CASAVA-1.8.2. Virus sequences were edited and assembled using the SeqMan and NextGen modules of the DNASTar Lasergene 7 program (Bioinformatics Pioneer DNASTar, Inc., Madison, Wis.). In certain cases, pre-filtering of reads to remove host sequence enhanced the assembly process.

#### RNA Analysis.

C7/10 cell monolayers were infected with Negev and Piura viruses at a MOI of 10. [<sup>3</sup>H]uridine (20 µCi/ml) was added 1-hour post infection (hpi) or 24 hpi respectively, and incubated for additional 24 hours. Supernatants were harvested and virus was purified via rate-zonal centrifugation (see below). Viral RNA was isolated by TRIzol LS (Invitrogen, Grand Island, N.Y.), denatured with glyoxal in dimethyl sulfoxide and analyzed by agarose gel electrophoresis, using previously described conditions (Gorchakov et al., 2004, *J Virol* 78:61-75).

#### Genomic Analysis.

The genome of Negev strain E0329 was used to determine protein domains. The genome was translated into proteins and then submitted to NCBI conserved domain prediction tool at URL [ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi](http://ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). The nucleotide and protein identities for the ORFs of all of the ten viruses generated were determined in EnzymeX (EnzymeX, Aalsmeer, Netherlands).

#### Phylogenetic Analysis.

Completed genomes of the 10 sequences were first aligned using translated protein sequences before being toggled back to nucleotides while maintaining the alignment. To determine areas of alignment that had sufficient confidence to determine phylogenetic relationships the alignment was run using the GUIDENCE software (Penn et al., 2010, *Nucleic Acids Research* 38:W23-28; Penn et al., 2010, *Molecular Biology and Evolution* 27:1759-1767). Areas with sufficient confidence were selected for further phylogenetic analysis. The Phylogenetic analyses were undertaken using PAUP\* version 4.0.10b (Swofford, D. 2000. PAUP\*. Phylogenetic Analysis Using Parsimony (\* and other methods). Version 4. Sinauer Associates, Sunderland, Mass.). The optimal evolutionary model for each data set was estimated from 56 models implemented using Modeltest version 3.06 (Posada and Crandall, 1998, *Bioinformatics* 14:817-818). An optimal maximum likelihood (ML) tree was then estimated using the appropriate model and a heuristic search with tree-bisection-reconstruction branch swapping and 10 replicates, estimating variable parameters from the data, where necessary. Bootstrap replicates were calculated for each dataset under the same models mentioned above. Bayesian analysis was undertaken using MrBayes v3.1 (Huelsenbeck and Ronquist, 2001, *Bioinformatics* 17:754-755; Ronquist and Huelsenbeck, 2003, *Bioinformatics* 19:1572-1574), and datasets were run for five hundred thousand generations until they reached congruence. The models used were HKY+G and HKY+I+G.

## SEQUENCE LISTING

The patent contains a lengthy “Sequence Listing” section. A copy of the “Sequence Listing” is available in electronic form from the USPTO web site (<http://seqdata.uspto.gov/?pageRequest=docDetail&DocID=US09388428B2>). An electronic copy of the “Sequence Listing” will also be available from the USPTO upon request and payment of the fee set forth in 37 CFR 1.19(b)(3).

The invention claimed is:

1. A nucleic acid expression vector comprising 5' to 3' (a) Negevirus 5'UTR, (b) Negevirus ORF1, (c) heterologous nucleic acid, and (d) Negevirus 3'UTR.

2. The expression vector of claim 1, further comprising a Negevirus ORF2 3' to Negevirus ORF 1 and 5' to the heterologous nucleic acid.

3. The expression vector of claim 1, further comprising a 5'cap.

4. The expression vector of claim 1, wherein the heterologous nucleic acid encodes an insecticidal polypeptide.

5. The expression vector of claim 1, wherein the heterologous nucleic acid is under the control of a heterologous promoter.

6. The nucleic acid expression vector of claim 1 encoding a Negevirus nucleic acid having a consensus sequence of SEQ ID NO: 1.

7. A method for producing a recombinant Negevirus comprising (a) inserting an expression vector of claim 1 into an isolated eukaryotic cell, and (b) incubating the eukaryotic cell expressing a recombinant Negevirus genome under conditions such that a recombinant Negevirus particle is formed.

8. A kit for producing a recombinant Negevirus comprising an expression vector of claim 1 and at least one cell culture reagent.

\* \* \* \* \*