

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:

PCT/EP2015/001399
9 July 2015 (09.07.2015)

(22) International Filing Date:

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

14176624.6 10 July 2014 (10.07.2014) EP

(71) Applicant: **DEUTSCHES KREBSFORSCHUNGSZENTRUM** [DE/DE]; Im Neuenheimer Feld 280, 69120 Heidelberg (DE).(72) Inventors: **DE VILLIERS-ZUR HAUSEN**; Eichenstr. 1, 69483 Wald-Michelbach (DE). **ZUR HAUSEN, Harald**; Eichenstr. 1, 69483 Wald-Michelhausen (DE). **GUNST, Karin**; Obergasse 2, 69493 Hirschberg (DE). **WHIT-****LEY, Corinna**; Abteistr. 6, 68519 Vierheim (DE). **PÉREZ, Iranzu Lamberto**; Gundolfstr. 7, 69120 Heidelberg (DE).(74) Agent: **HUBER & SCHÜSSLER**; Trudering Str. 246, 81825 München (DE).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,

[Continued on next page]

(54) Title: HCBI, MSBI, MSSI AND CMI SEQUENCES AS AN EARLY MARKER FOR THE FUTURE DEVELOPMENT OF CANCER AND DISEASES OF THE CNS AND AS A TARGET FOR THE TREATMENT AND PREVENTION OF THESE DISEASES

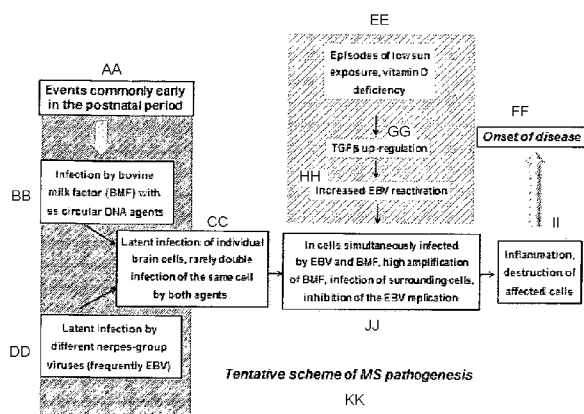


Fig. 13

- AA Événements survenus généralement tôt dans la période postnatale
 BB Infection par le BMF (bovine milk factor) avec des ADN circulaires simple brin
 CC Infection latente de cellules cérébrales particulières, rarement double infection de la même cellule par les deux agents
 DD Infection latente par différents virus du groupe de l'herpès (fréquemment l'EBV)
 EE Épisodes de faible exposition au soleil, carence en vitamine D
 FF Installation de la maladie
 GG Régulation à la hausse du TGF-β
 HH Réactivation accrue de l'EBV
 II Inflammation, destruction des cellules affectées
 JJ Dans les cellules simultanément infectées par l'EBV et le BMF, forte amplification du BMF, infection des cellules environnantes, inhibition de la réplication de l'EBV
 KK Schéma, à confirmer, de la pathogenèse de la sclérose en plaques

(57) Abstract: Described are HCBI (Healthy Cattle Blood Isolate), MSBI (Multiple Sclerosis Brain Isolate), MSSI (Multiple Sclerosis Serum Isolate) and CMI (Cow Milk Isolate) nucleotide sequences as well as probes and primers comprising part of said nucleotide sequences and antibodies against polypeptides encoded by said nucleotide sequences. Said compounds are useful as early markers for the future development of cancer and diseases of the CNS (Multiple sclerosis MS, Prion-linked diseases, amyotrophic lateral sclerosis, transmissible spongiforme encephalitis, Parkinson's disease, Alzheimer disease) and should represent targets for treatment and prevention.



TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

HCBI, MSBI, MSSSI and CMI sequences as an early marker for the future development of cancer and diseases of the CNS and as a target for the treatment and prevention of these diseases

5

The present invention relates to HCBI (Healthy Cattle Blood Isolate), MSBI (Multiple Sclerosis Brain Isolate), MSSSI (Multiple Sclerosis Serum Isolate) and CMI (Cow Milk Isolate) nucleotide sequences as well as probes and primers comprising part of said nucleotide sequences and antibodies against polypeptides encoded by said nucleotide sequences. Finally, the present invention relates to the use of said compounds as an early marker for the future development of diseases such as cancer and diseases of the CNS and as a target for treatment and prevention of these diseases.

15

Background

Several epidemiological analyses conducted in recent decades indicate that the long-term consumption of "red" meat processed by different ways (including smoked or air-dried meat and meat as component of sausages consumed rare, undercooked or grilled) can be regarded as a risk factor for colon cancer (World Cancer Report 2007, zur Hausen 2012). "Red" meat is regarded as comprising beef, pork, mutton, lamb and goat meat, in contrast to "white" meat (poultry meat/fish).

25

Thus far, chemical carcinogenic substances being produced during roasting, grilling, barbecuing, smoking and air-drying were blamed as risk factors for cancer. However, often the fact was disregarded that the same substances are also produced in comparable concentrations during analogous ways of preparation of poultry meat/fish. Accordingly, this does not support the assumption that these chemical substances play an exclusive role as regards the development of colon cancer. Since, in

30

35

addition, the current epidemiological analyses suggest that beef is the main risk factor it has been postulated that an additional species-specific - presumably infectious - factor contributes to the triggering of this type of cancer (zur
5 Hausen, 2012). The results of the correlation of analyses of the global spreading of domesticated bovine species with the global incidence of colon cancer seem to suggest that the consumption of meat of bovine species stemming from European/Asian cattle (*Bos taurus*) but not from breedings of
10 zebu, water buffalo or yak might be of importance as a main risk factor (zur Hausen, 2015)

Thus, the technical problem underlying the present invention is to identify specific nucleotide sequences that might be
15 associated with diseases such as cancer or diseases of the CNS and, thus, to provide means for diagnosis and therapy.

Brief Description of the Present Invention

20

The solution to said technical problem is achieved by providing the embodiments characterized in the claims.

During the experiments resulting in the present invention sera
25 of cattle were screened for infectious agents - starting from the assumption that the presence in sera is also indicative for the presence of these agents in "red" meat. Sera from healthy cows were screened and new viral nucleic acid components could be isolated. The DNA sequences and open reading frames of
30 several of these components showed a recognizable relationship to two sequences which were already described for transmissible spongiform encephalopathies (TSE) for TSE-diseases of sheep, cattle and humans.

The TSE isolates have also been suspected to play a role in cancer induction (Manuelidis, 2011), thus, it is reasonable to assume that the viral sequences described might be associated with the development of diseases like cancer, specifically colon and breast cancers but also Hodgkin's disease and others, and diseases of the CNS (Multiple sclerosis MS, amyotrophic lateral sclerosis, transmissible spongiforme encephalopathies/ Prion-linked diseases, Parkinson's disease, Alzheimer disease).

Brief description of the drawings

Abbreviations for Figures 1-4: Rep = replication protein; CP = capsid protein

CMI: cattle milk isolate

HCBI: healthy cattle blood isolate

MSCI: MS brain isolate

MSSI: MS serum isolate

Sphinx: slow progressive hidden infection of variable (X) latency

Figure 1:

Group 1 - Isolates from bovine serum, bovine milk and multiple sclerosis (MS) brain samples (post mortem) - all related to Sphinx1.76

(a) HCBI6.252 (2522 bp) (Healthy Cattle Blood Isolate; identical to CMI1.252 (Cow Milk Isolate)).

(b) HCBI6.159 (1591 bp): A deletion in HCBI6.252 apparently resulted in HCBI6.159 - both were rescued independently from the bovine serum using back-to-back primer, i.e. circular DNA

molecules and not an artefact. This could indicate that these infectious organisms have 2 forms - a large and small molecule related to or complementing each other

(c) MSBI1.176 (Multiple Sclerosis Brain Isolate) (1766 bp) -
5 98% similar to Sphinx1.76

(d) MSBI2.176 (1766 bp) - isolated from the same MS-brain sample as MSBI1.176

(e) CMI1.252 (2523 bp) (Cow Milk Isolate; identical to HCBI6.252) 9

10 (f) CMI2.214 (2148 bp)

(g) CMI3.168 (1687 bp)

(h) CMI4.158 (1583 bp)

The isolates were all generated by using back-to-back primers
15 designed on the replication gene of Sphinx1.76.

Primers (several isolates were isolated twice by applying both primer pairs independently).

Nn (forward GGATTAATGCCAATGATCC), Xn (reverse CTTTGCCTGTTTCTCT
CG), and/or No (forward GAGGACGAATTAATATTACAAGTC), Xo (reverse
20 GTTCTCGTTTTCTTGGTAA)

Figure 2:

Group 2: Isolate HCBI7.228 (2280bp) (Healthy Cattle Blood Isolate) related to Sphinx2.36

25 This sample was generated by PCR from bovine serum using back-to-back primers

Primers: Nd (forward CAGATTGCAAAGCCTGTAATTCAT), Xd (reverse CTAAGGCAGATCAACACAGGGATA)

Figure 3:

Group 3: Isolates from bovine serum and multiple sclerosis serum and MS brain samples (post mortem) - distantly related to known ssDNA viruses

Nucleotide sequences of (a) HCBI8.215 (2152 bp) and (b) HCBI9.212 (2121 bp) (Healthy Cattle Blood Isolate) as well as (c) MSSI2.225 (2259 bp) (Multiple Sclerosis Serum Isolate). An identical sequence (MSBI3.224) to MSSI2.225 was also isolated from an MS-brain sample.

These HCBI isolates were generated as follows:

Serum samples were subjected to ultracentrifugation on Optiprep gradient centrifugation. The gradient was fractionated, DNA extracted from fractions and subjected to rolling circle amplification. Products were digested with BamHI or EcoRI enzymes. Products were cloned and sequenced. Complete genomes were obtained by amplification of original DNA samples with back-to-back primers designed on the respective gradient clones. Complete genomes were sequenced in full.

The MSSI isolate (as well as the clone MSBI3.224 from the MS-brain) was obtained after rolling circle amplification of the extracted DNA from MS serum and MS brain, restriction enzyme digestion, cloning of the product and subsequent sequencing. The complete genome was obtained by amplification with back-to-back primers designed on the original clone. The complete genome was sequenced in full.

Figure 4:

Group 4 - distantly related to *Psychrobacter* spp plasmid: Nucleotide sequence of MSSI1.162 and putative open reading frames

MS-serum isolate MSSI1.162 (group 4) was generated by rolling circle amplification of a multiple MS-serum sample. The product

WO 2016/005054 PCT/EP2015/001399
was digested with restriction enzyme HindIII, cloned and
sequenced.

Figure 5:

5 Genome organisation of 4 groups of isolates

The isolates are grouped according to their sequence similarity to the Sphinx1.76 genome (group 1), Sphinx2.36 (group 2), myco-like Gemycircularviruses (group 3) and Psychrobacter spp. Plasmid (group 4).

10 (A) group 1

(B) groups 2,3 and 4

Figure 6:

15 Alignment of a replication gene/ iteron-like repeat region
between 8 isolates and Sphinx1.76

Figure 7:

20 Schematic outline of latent infection of different types of
brain cells with Herpes type genomes and BMF factor

Figure 8:

Spontaneous reactivation of Herpes DNA in a cell concomitantly
infected by Herpes and BMF DNA

Amplification of BMF and inhibition of Herpes DNA replication.

25

Figure 9:

Schematic outline of the recognition of foreign antigens by the
immune system, T cell response

30 Figure 10:

Mononuclear inflammatory cells surrounding a small vein in an
early lesion

Figure 11:

Advancing age of the lesion (plaque) on the left with normal
5 white matter on the right

Macrophages are present in the lesion (arrows) and at the interphase.

Figure 12:

10 Schematic outline of the pathogenesis concept for multiple
sclerosis

EBV is used as an example for the role of Herpes-type viruses.

Figure 13:

15 Tentative Scheme of MS pathogenesis

Detailed description of the present invention

20 In the present application a new concept for the pathogenesis of multiple sclerosis and cancer is presented: Interaction of an Amplifying Virus and the amplified DNA of a Bovine Milk (or serum) Factor (BMF)

25 Introduction

The incidence of multiple sclerosis (MS) increased in several parts of the world (reviewed in Kurtzke, 2000, Alcalde-Cabero et al., 2013). This increase has been mainly attributed to environmental factors. Migrants from high to lower risk areas
30 retain the MS risk of their birth place only if they are at least age 15 at migration, frequently interpreted to be due to an infection acquired during early childhood (reviewed in

WO 2016/005054 PCT/EP2015/001399
Kurtzke, 2000). Clustering of cases and the geographic epidemiology has also been widely discussed: A rising incidence of MS was noted in females linked to urbanization (Kotzamani et al., 2012).

5

Demyelination is a characteristic feature of MS lesions. Four fundamentally different patterns of demyelination were found, defined on the basis of myelin protein loss, the location and extension of plaques, the patterns of oligodendrocyte
10 destruction, and the immunopathological evidence of complement activation (Lucchinetti et al., 2000, Metz et al., 2014). At a given time point of the disease the patterns of demyelination were heterogeneous between patients; but they were homogenous within multiple active lesions from the same patient,
15 potentially pointing to different contributing factors.

Two of the risk factors seem to deserve special attention: the relatively consistent results pointing to a possible role of different, predominantly herpes-group viruses, and the
20 consumption of fresh cow milk, potentially including other dairy products (see below). In addition, Vitamin D deficiency plays a role as a risk factor.

- Viral Risk Factors

25 Apparently all Herpes virus types share two properties which seem to be relevant for the subsequent discussion:

During their persistence in a latent stage, spontaneous reactivation may occur, in part regulated by specific gene functions, partly also by epigenetic mechanisms (reviewed in
30 Nicoll et al., 2012, Grinde, 2013). Reactivation may also be triggered by interaction with extracellular cytokines, such as transforming growth factor β .

Spontaneous induction of a lytic cycle has been observed for virtually every human pathogenic Herpes virus type. The high antibody titers against Epstein-Barr virus structural proteins in EBV-positive Burkitt's lymphomas and nasopharyngeal cancers (reviewed in Henle, W. and Henle, G., 1977) may serve here as one example. Reactivations of human Herpes virus type 6, Varicella-Zoster virus, Herpes simplex virus and others are not rare events and may affect a number of different cell types (Hu Knox et al., 2000).

10

A second remarkable property of herpes virus infection represents the amplification of various double- or single-stranded small DNA virus genomes upon infection of cells containing such DNAs in a latent state. This has been noted for human and monkey Polyoma viruses, JC and SV40, for human and bovine Papilloma viruses, as well as for single-stranded Adeno-associated (AAV) and Anello-/TT-viruses (Schlehofer and zur Hausen, 1990, Heilbronn et al., 1993, Borkosky et al., 2012). The Herpes-group viruses used in these studies were Herpes simplex virus, human cytomegalovirus and Epstein-Barr virus. The potential to induce amplification of latent small viral DNA genomes is also shared by Adeno- and Vaccinia viruses (Schlehofer and zur Hausen, 1990). The helper effect of Herpes- and adenovirus-induced amplification of parvoviruses has been intensively studied for adeno-associated viruses. The replication of the latter seems to depend on this helper effect but in turn leads to a reduction of Herpes- or adenovirus replication due to the preferential amplification of small viral DNA (Schlehofer et al., 1983, Matz et al., 1984, Bantel-Schaal and zur Hausen, 1988, Schmitt et al., 1989, Schlehofer and zur Hausen, 1990, Heilbronn et al., 1990a, Heilbronn et al., 1990b).

WO 2016/005054 PCT/EP2015/001399
Spontaneous induction of Herpes-group viruses and the
amplification of latent small viral DNA form the basis for the
subsequent postulation of the mechanism underlying MS
development.

5

- Bovine milk factor

Several reports noted a correlation between consumption of non-
pasteurized cow milk and MS development (*Murray TJ. 1976,*
Sepčić et al., 1993, Malosse and Perron, 1993), whereas others
10 stressed long-time consumption of cow milk as a risk factor, in
particular when consumed in the early phase of life (*Agranoff*
and Goldberg, 1974, Christensen, 1975, Warren, 1984, Butcher,
1976, 1986, Winer et al., 2001, Munger et al., 2011a).

15 If a specific factor in cow milk exists which increases the
risk for MS development, one can anticipate a protective role
of long-term breast-feeding. Long-term breast-feeding (for six
months and more) has indeed repeatedly been reported as having
a protective effect for MS development (*Christensen, 1975,*
20 *Warren, 1984, Tarrats et al., 2002, Conradi et al., 2013*). The
existence of a cow milk factor would also not exclude a
specific genetic predisposition for the development of MS. A
monogenic predisposition for MS has been reported in a
chromosomal localization close to BRCA1 (*Holzmann et al.,*
25 *2013*).

- Vitamin-D deficiency

A role of vitamin-D deficiency has repeatedly been implicated
for the initiation of MS (*reviewed in Ascherio et al. 2012,*
30 *2013*).

A convincing relationship between vitamin D deficiency and
Epstein Barr Virus reactivation originates from early studies

on EBV reactivation by transforming growth factor beta (TGF- β).

A serum factor, purified and labeled as Epstein-Barr virus-inducing factor (EIF) (Bauer et al., 1982) proved to be identical to the subsequently described TGF- β molecule (Frolik et al., 1983, Bauer et al., 1991). TGF- β in turn is negatively regulated by activated vitamin D receptors (Isik et al., 2012, Ito et al., 2013, Zerr et al., 2014). This could very well explain the season-related preferential onset of MS and of exacerbations.

The relationship between low vitamin D and EBV reactivation is further supported by studies describing a correlation between low vitamin D and elevated immunoreactivity against Epstein-Barr virus prior to the clinical manifestation of multiple sclerosis (Munger et al., 2011b, Décard et al., 2012) and a higher rate of EBV excretion of EBV-positive MS patients in comparison to EBV-positive healthy controls (Yea et al., 2013).

Thus, at least two factors have been implicated as potential etiological contributors for both diseases of the CNS (e.g. MS) and cancer (e.g. colon and breast cancer): vitamin D deficiency and the reactivation of various herpes group viruses, mainly Epstein-Barr virus (EBV), human herpes virus type 6, and varicella-zoster virus. According to the present invention the identification of several novel types of small circular single-stranded DNAs, presumably of viral origin, from cattle sera of the present invention and commercially available dairy products, show a unifying concept. The inventors have demonstrated in the present invention that co-infection of cells with herpes-group viruses and small single-stranded or double-stranded DNA viruses results in an substantial amplification of small viral DNA with partial inhibition of the herpes virus. Some of the molecules identified in dairy cattle sera and milk are distantly related to DNA reported in prion-

linked brain lesions and have been found in two autopsy lesions of patients with multiple sclerosis. The amplification of these single-stranded DNA molecules by reactivation of a co-latently persisting herpes virus genome could result in their
5 amplification and evoking a local immune response resulting in destruction of the affected brain cells. This model could in part explain the North-South incidence gradient of multiple sclerosis, which is thought to be linked to vitamin-D deficiency and herpes virus reactivation (c.f. Fig. 12 and 13).

10

In addition, the full-length genomes of the isolates from MS brains and sera were isolated and re-circularized before transfection into the human cell line 293TT. Transfected cells were harvested on day 3 and total RNA extracted using the miRNA
15 Easy Kit (Qiagen). Samples were further purified (DNase digestion, ribosome removal) and subjected to high throughput RNA sequencing. RNA transcripts have been obtained for MSBI1,176, MSBI2,176 (Group 1), MSSI2.225 (Group 3) and MSSI1.162 (Group 4). The RNA transcripts clearly show that the
20 isolates replicate in human cells.

The inventors consider Vitamin D deficiency and herpes virus reactivations as risk factors also for breast and colon cancers. Reactivation of dual latent infections within the same
25 cell, outlined above for multiple sclerosis pathogenesis, could therefore also play a particular important role in the aetiology of these cancers.

Thus, it is considered by the inventors that multiple sclerosis
30 (MS) and also the other below mentioned diseases result from

- Latent infections of the same cell with two different infectious agents, one of them most likely a herpes-type virus (in particular EBV, HHV-6, VZV, but also HSV, HHV-7),

the other one acquired by bovine milk consumption (bovine milk factor - BMF) as the first event. Each of them latently infects individual cells, but occasionally genomes of both agents occur within the same cell.

- 5 • Reactivation of the herpes-type virus most frequently, but not only, Epstein Barr virus (EBV) to a lytic cycle as a second precondition. For EBV this is probably linked to increased levels of transforming growth factor β (TGF β) which is negatively regulated by activated vitamin D receptors;
- 10 • As third event, amplification of BMF, resulting in partial suppression of Herpes-type DNA synthesis and formation of BMF particles or spreading of its nucleic acid to neighbouring cells via neuronal interconnections;
- 15 • This is followed by an infection of neighbouring cells with expression of BMF protein;
- 20 • Finally, T-cell response against BMF leads to the destruction of affected cells and in case of MS to plaque formation. This supports the clinical observation of the focal appearance of lesions, commonly starting from a central vein and the intensive localized immune response in early lesions.

25 Transmissions of agents present in milk or dairy products may lead to latent infections in human brain cells followed by amplification of these agents in case of co-latency and spontaneous induction of a Herpes virus DNA or Herpes virus - like DNA. Potential BMF candidate agents are described in Examples 2-5 and the accompanying figures.

30

The presence of presumably circular single stranded DNA related to Sphinx-sequences, Anello-, Circo-, and Gemycircularvirus

families, as well as *Psychrobacter* species in cattle sera (see Examples 2-5), in commercially available milk samples, as well as in florid MS lesions and MS serum permits the development of a concept for MS pathogenesis. It integrates observations of involvement of Herpes-type viruses, most prominently of EBV, of their property to amplify small double- and single-stranded DNA viruses, of viral cow milk factors, of vitamin-D deficiency, the EBV inducing property of TGF β , and the partial season dependence of MS onset and of exacerbation in the course of disease. This concept is schematically outlined in Figures 7-13.

An initial dual latent infection of the same or closely flanking cells by a herpes virus genome and the postulated BMF, followed by a trigger for Herpes virus reactivation and the subsequent preferential amplification of single-stranded DNA are defined as the primary event. In the case of latent EBV infection, vitamin D deficiency with the subsequent up-regulation of TNF- β as an EBV-inducing factor could be the important trigger for up-regulation. Probably abortive infection of neighbouring cells with viral antigen expression results in an active T-cell response and the destruction of the affected cells. The frequently described seasonality of MS onset and of new rounds of MS exacerbations, the repeatedly reported North-South gradient of MS incidence should reflect the degree of sun-light exposure, negatively correlating vitamin D levels with TGF β concentration and EBV reactivation.

Thus, the inventors anticipate the presence of different BMF sequences also in susceptible normal human brain cells in a latent form. The remarkable heterogeneity of the BMF isolates may also find its reflection in variations for pathologic characteristics of MS in humans (*Lucchinetti et al., 2000, Metz*

et al., 2014). It would not be too surprising if eventually "high" and "low" risk types will be identified, in a certain analogy to human papillomavirus pathogenicity (zur Hausen, 1985). The majority of those carriers will not develop MS, since the latter should require latent co-infection of a BMF-positive cell with a Herpes-type virus and spontaneous induction of the latter. This should be a rare event, increasing, however, under conditions resulting in frequent Herpes virus reactivations.

10

As a final point, it should be of interest to also apply this concept to other presumably autoimmune diseases and certain cancers occurring at increased frequency under conditions of vitamin D deficiency. As far as cancers are concerned, this specifically accounts for colon- and breast cancer, and possibly for ovarian, prostate, pancreatic cancer and lung cancers.

15

The risk for insulin-dependent diabetes mellitus has been repeatedly linked to cow milk consumption and vitamin D deficiency (reviewed in Scott, 1990, in Grant, 2006, in Hyppönen et al., 2010). The latter system seems to come particularly close to the MS situation.

20

Accordingly, the present invention relates to an HCBI, MSBI, MSSI or CMI polynucleic acid comprising:

25

(a) a nucleotide sequence depicted in any one of Figures 1 to 4;

(b) a nucleotide sequence having at least 90% identity to a nucleotide sequence of (a);

30

(c) a fragment of a nucleotide sequence of (a) or (b);

(d) a nucleotide sequence being complementary to a nucleotide sequence of (a), (b) or (c); or

(e) a nucleotide sequence which is redundant as a result of the degeneracy of the genetic code compared to any of the above-given nucleotide sequences.

5 The term "polynucleic acid" refers to a single-stranded or double-stranded nucleic acid sequence. A polynucleic acid may consist of deoxyribonucleotides or ribonucleotides, nucleotide analogues or modified nucleotides or may have been adapted for diagnostic or therapeutic purposes. A polynucleic acid may also
10 comprise a double stranded cDNA clone which can be used, for example, for cloning purposes.

The HCBI, MSBI, MSSSI or CMI polynucleic acids of the invention can be prepared according to well-known routine methods, for
15 example, by (a) isolating the entire DNA or RNA from a sample, (b) detecting the HCBI, MSBI, MSSSI or CMI sequence by hybridization or PCR and (c) cloning of the HCBI, MSBI, MSSSI or CMI sequence into a vector.

20 Also included within the present invention are sequence variants of the polynucleic acid of the invention containing either deletions and/or insertions of one or more nucleotides, especially insertions or deletions of one or more codons, mainly at the extremities of oligonucleotides (either 3' or 5') and
25 which show at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to said polynucleic acid sequences of the invention. Polynucleic acid sequences according to the present invention which are similar to the sequences as shown in Figures 1 to 4 can be characterized and isolated according to any of the
30 techniques known in the art, such as amplification by means of sequence-specific primers, hybridization with sequence-specific probes under more or less stringent conditions, sequence

determination of the genetic information of HCBI, MSBI, MSSI or CMI etc.

5 The present invention also provides fragments of the nucleotide sequences of the present invention described above that harbour a replication gene which codes for a replication protein. An autonomous replicating nucleotide sequence comprises a nucleotide sequence of the replication gene or a fragment thereof which is capable of inducing autonomous replication.

10 Replication protein represents an endonuclease which binds single-stranded DNA inducing a single-stranded cut at or near the origin of replication (Wolds, 1997). The skilled person can derive at such fragments capable of inducing autonomous replication without undue experimentation. Such fragments may

15 have a length of at least 45, at least 55, or at least 65 nt.

The person skilled in the art can easily determine which nucleic acid sequences are related to a nucleotide sequence of Figures 1 to 4 or which fragments are still capable of replicating

20 autonomously by using standard assays.

The present invention also provides polynucleic acid sequences which are redundant as a result of the degeneracy of the genetic code compared to any of the above-given nucleotide sequences.

25 These variant polynucleic acid sequences will thus encode the same amino acid sequence as the polynucleic acids they are derived from.

The HCBI, MSBI, MSSI or CMI polynucleic acids of the invention

30 might be present as an extrachromosomal episome, might be integrated into the host's genome and/or might be linked to a host cell DNA.

The present invention also relates to an oligonucleotide primer

35 comprising or consisting of part of a polynucleic acid as

defined above, with said primer being able to act as primer for specifically sequencing or specifically amplifying HCBI, MSBI, MSSI or CMI polynucleic acid of the invention.

5 The term "primer" refers to a single stranded DNA oligonucleotide sequence capable of acting as a point of initiation for synthesis of a primer extension product which is complementary to the nucleic acid strand to be copied. The length and the sequence of the primer must be such that they
10 allow priming the synthesis of the extension products. Preferably the primer is about 5-50 nucleotides. Specific length and sequence will depend on the complexity of the required DNA or RNA targets, as well as on the conditions of primer use such as temperature and ionic strength.

15

The fact that amplification primers do not have to match exactly with a corresponding template sequence to warrant proper amplification is amply documented in the Literature. The amplification method used can be, for example, polymerase chain
20 reaction (PCR), ligase chain reaction (LCR), nucleic acid sequence-based amplification (NASBA), transcription-based amplification system (TAS), strand displacement amplification (SDA) or amplification by means of Qb replicase or any other suitable method to amplify nucleic acid molecules using primer
25 extension. During amplification, the amplified products can be labelled either using labelled primers or by incorporating labelled nucleotides.

Labels may be isotopic (^{32}P , ^{35}S , etc.) or non-isotopic (biotin, digoxigenin, etc.). The amplification reaction is repeated
30 between 20 and 70 times, advantageously between 25 and 45 times.

Any of a variety of sequencing reactions known in the art can be used to directly sequence the viral genetic information and
35 determine the ORF by translating the sequence of the sample into

the corresponding amino acid sequence. Exemplary sequencing reactions include those based on techniques developed by Sanger or Maxam and Gilbert. It is also contemplated that a variety of automated sequencing procedures may be utilized when performing the subject assays including sequencing by mass spectrometry (see, for example: PCT publication WO 94/16101). It will be evident to one skilled in the art that, for example the occurrence of only two or three nucleic bases needs to be determined in the sequencing reaction.

10

Preferably, these primers are about 5 to 50 nucleotides long, more preferably from about 10 to 25 nucleotides. Most preferred are primers having a length of at least 13 bases.

15 The present invention also relates to an oligonucleotide probe comprising or consisting of part of an HCBI, MSBI, MSSSI or CMI polynucleic acid as defined above, with said probe being able to act as a hybridization probe for specific detection of a HCBI, MSBI or CMI polynucleic acid according to the invention.

20

The probe can be labelled or attached to a solid support.

The term "probe" refers to single stranded sequence-specific oligonucleotides which have a sequence which is complementary to the target sequence of an HCBI, MSBI, MSSSI or CMI polynucleic acid to be detected.

25

Preferably, these probes are about 5 to 50 nucleotides long, more preferably from about 10 to 25 nucleotides. Most preferred are probes having a length of at least 13 bases.

30

The term "solid support" can refer to any substrate to which an oligonucleotide probe can be coupled, provided that it retains its hybridization characteristics and provided that the background level of hybridization remains low. Usually the solid

35

substrate will be a microtiter plate, a membrane (e.g. nylon or nitrocellulose) or a microsphere (bead). Prior to application to the membrane or fixation it may be convenient to modify the nucleic acid probe in order to facilitate fixation or improve the hybridization efficiency. Such modifications may encompass homopolymer tailing, coupling with different reactive groups such as aliphatic groups, NH₂ groups, SH groups, carboxylic groups, or coupling with biotin or haptens.

10 The oligonucleotides according to the present invention, used as primers or probes may also contain or consist of nucleotide analogues such as phosphorothioates, alkylphosphorates or peptide nucleic acids or may contain intercalating agents. These modifications will necessitate adaptations with respect to the conditions under which the oligonucleotide should be used to obtain the required specificity and sensitivity. However, the eventual results will be essentially the same as those obtained with the unmodified oligonucleotides.

20 The introduction of these modifications may be advantageous in order to positively influence characteristics such as hybridization kinetics, reversibility of the hybrid-formation, biological stability of the oligonucleotide molecules, etc.

25 The polynucleic acids of the invention may be comprised in a composition of any kind. Said composition may be for diagnostic, therapeutic or prophylactic use.

The present invention also relates to a recombinant expression vector comprising an HCBI, MSBI, MSSI or CMI polynucleic acid of the invention as defined above operably linked to prokaryotic, eukaryotic or viral transcription and translation control elements as well as host cells containing such vector.

35 The term "vector" may comprise a plasmid, a cosmid, an

artificial chromosome, a phage, or a virus or a transgenic non-human animal. Particularly useful for vaccine development may be HCBI, MSBI, MSSI or CMI recombinant molecules, BCG or adenoviral vectors, as well as avipox recombinant viruses.

5

The term "recombinant expression" used within the context of the present invention refers to the fact that the polypeptides of the present invention are produced by recombinant expression methods be it in prokaryotes, or lower or higher eukaryotes as discussed in detail below.

10

The term "host cell" refers to cells which can be or have been, used as recipients for a recombinant vector or other transfer polynucleotide, and include the progeny of the original cell which has been transfected.

15

It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation or recombination.

20

The term "lower eukaryote" refers to host cells such as yeast, fungi and the like. Lower eukaryotes are generally (but not necessarily) unicellular. Preferred lower eukaryotes are yeasts, particularly species within *Saccharomyces*, *Schizosaccharomyces*, *Kluiveromyces*, *Pichia* (e. g. *Pichia pastoris*), *Hansenula* (e. g. *Hansenula polymorph*), *Schwaniomyces*, *Schizosaccharomyces*, *Yarrowia*, *Zygosaccharomyces* and the like. *Saccharomyces cerevisiae*, *S. carlsbergensis* and *K. lactis* are the most commonly used yeast hosts, and are convenient fungal hosts.

25
30

The term "higher eukaryote" refers to host cells derived from higher animals, such as mammals, reptiles, insects, and the like. Presently preferred higher eukaryote host cells are derived from Chinese hamster (e. g. CHO), monkey (e. g. COS and

35

Vero cells), baby hamster kidney (BHK), pig kidney (PK15), rabbit kidney 13 cells (RK13), the human osteosarcoma cell line 143 B, the human cell line HeLa and human hepatoma cell lines like Hep G2, the 293TT cell line (Buck et al., 2004) and
5 insectcell lines (e.g. Spodoptera frugiperda). The host cells may be provided in suspension or flask cultures, tissue cultures, organ cultures and the like. Alternatively the host cells may also be transgenic non-human animals.

10 The term "prokaryotes" refers to hosts such as E. coli, Lactobacillus, Lactococcus, Salmonella, Streptococcus, Bacillus subtilis or Streptomyces. Also these hosts are contemplated within the present invention.

15 The segment of the HCBI, MSBI, MSSI or CMI DNA encoding the desired sequence inserted into the vector sequence may be attached to a signal sequence. Said signal sequence may be that from a non-HCBI, MSBI, MSSI or CMI source, but particularly preferred constructs according to the present invention contain
20 signal sequences appearing in the HCBI, MSBI, MSSI or CMI genome before the respective start points of the proteins.

Higher eukaryotes may be transformed with vectors, or may be infected with a recombinant virus, for example a recombinant
25 vaccinia virus. Techniques and vectors for the insertion of foreign DNA into vaccinia virus are well known in the art, and utilize, for example homologous recombination. A wide variety of viral promoter sequences, possibly terminator sequences and poly(A)-addition sequences, possibly enhancer sequences and
30 possibly amplification sequences, all required for the mammalian expression, are available in the art. Vaccinia is particularly preferred since vaccinia halts the expression of host cell proteins. For vaccination of humans the avipox and Ankara Modified Virus (AMV) are particularly useful vectors.

Also known are insect expression transfer vectors derived from baculovirus Autographa californica nuclear polyhedrosis virus (AcNPV), which is a helper-independent viral expression vector.

5 Expression vectors derived from this system usually use the strong viral polyhedrin gene promoter to drive the expression of heterologous genes. Different vectors as well as methods for the introduction of heterologous DNA into the desired site of baculovirus are available to the person skilled in the art for
10 baculovirus expression. Also different signals for posttranslational modification recognized by insect cells are known in the art.

The present invention also relates to a polypeptide having an
15 amino acid sequence encoded by an HCBI, MSBI, MSSI or CMI polynucleic acid as defined above, or a part or an analogue thereof being substantially similar and biologically equivalent.

The term "polypeptide" refers to a polymer of amino acids and
20 does not refer to a specific length of the product. Thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not refer to or exclude post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the
25 like. Included within the definition are, for example, polypeptides containing one or more analogues of an amino acid (including, for example, unnatural amino acids, peptide nucleic acid (PNA), etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally
30 occurring and non-naturally occurring.

The polypeptides according to the present invention contain preferably at least 3, preferably 4 or 5 contiguous HCBI, MSBI, MSSI or CMI amino acids, 6 or 7 preferably however at least 8

WO 2016/005054 PCT/EP2015/001399
contiguous HCBI, MSBI, MSSSI or CMI amino acids, at least 10 or
at least 15.

The polypeptides of the invention, and particularly the
5 fragments, can be prepared by classical chemical synthesis. The
synthesis can be carried out in homogeneous solution or in solid
phase. The polypeptides according to this invention can also be
prepared by means of recombinant DNA techniques.
The present invention also relates to a method for production of
10 a recombinant polypeptide as defined above, comprising: (a)
transformation of an appropriate cellular host with a
recombinant vector, in which a polynucleic acid or a part
thereof as defined above has been inserted under the control of
the appropriate regulatory elements, (b) culturing said
15 transformed cellular host under conditions enabling the
expression of said insert, and (c) harvesting said polypeptide.

The present invention also relates to an antibody raised upon
immunization with at least one polypeptide as defined above,
20 with said antibody being specifically reactive with any of said
polypeptides, and with said antibody being preferably a
monoclonal antibody. The term „antibody“, preferably, relates to
antibodies which consist essentially of pooled monoclonal
antibodies with different epitopic specificities, as well as
25 distinct monoclonal antibody preparations. Monoclonal antibodies
are made from an antigen containing, e.g., a polypeptide encoded
by an HCBI, MSBI, MSSSI or CMI polynucleic acid of the invention
or a fragment thereof by methods well known to those skilled in
the art. As used herein, the term „antibody“(Ab) or „monoclonal
30 antibody“ (Mab) is meant to include intact molecules as well as
antibody fragments (such as, for example, Fab and F(ab')₂
fragments) which are capable of specifically binding to protein.
Fab and F(ab')₂ fragments lack the Fc fragment of intact
antibody, clear more rapidly from the circulation, and may have
35 less non-specific tissue binding than an intact antibody. Thus,

these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies useful for the purposes of the present invention include chimerical, single chain, and humanized antibodies.

5

The present invention also relates to diagnostic and therapeutic approaches using cell-mediated immune responses.

Preferably, the antibody or antigen binding fragment thereof carries a detectable label. The antibody/fragment can be directly or indirectly detectably labeled, for example, with a radioisotope, a fluorescent compound, a bioluminescent compound, a chemiluminescent compound, a metal chelator or an enzyme. Those of ordinary skill in the art will know of other suitable labels for binding to the antibody, or will be able to ascertain such, using routine experimentation.

The present invention also relates to a diagnostic kit for use in determining the presence of an HCBI, MSBI or CMI polynucleic acid or polypeptide of the invention, said kit comprising a primer, a probe, and/or an antibody of the invention. Said kit may have any format well known to the person skilled in the art, e.g. can be an ELISA-based kit.

The present invention also relates to a method for the detection of an HCBI, MSBI, MSSSI or CMI polynucleic acid according to the invention present in a biological sample, comprising: (a) optionally extracting sample polynucleic acid, (b) amplifying the polynucleic acid as described above with at least one primer as defined above, optionally a labelled primer, and (c) detecting the amplified polynucleic acids.

The term "polynucleic acid" can also be referred to as analyte strand and corresponds to a single- or double-stranded polynucleic acid molecule.

The term "labelled" refers to the use of labelled nucleic acids. This may include the use of labelled nucleotides incorporated during the polymerase step of the amplification or labelled
5 primers, or by any other method known to the person skilled in the art.

The present invention also relates to a method for the detection of an HCB1, MSBI, MSS1 or CMI polynucleic acid according to the
10 invention present in a biological sample, comprising: (a) optionally extracting sample polynucleic acid, (b) hybridizing the polynucleic acid as described above with at least one probe as defined above, and (c) detecting the hybridized polynucleic acids.

15 The hybridization and washing conditions are to be understood as stringent and are generally known in the art. However, according to the hybridization solution (SSC, SSPE, etc.), these probes should be hybridized at their appropriate temperature in order
20 to attain sufficient specificity.

According to the hybridization solution (SSC, SSPE, etc.), these probes should be stringently hybridized at their appropriate temperature in order to attain sufficient specificity. However,
25 by slightly modifying the DNA probes, either by adding or deleting one or a few nucleotides at their extremities (either 3' or 5'), or substituting some non-essential nucleotides (i. e. nucleotides not essential to discriminate between types) by others (including modified nucleotides or inosine) these probes
30 or variants thereof can be caused to hybridize specifically at the same hybridization conditions (i.e. the same temperature and the same hybridization solution). Also changing the amount (concentration) of probe used may be beneficial to obtain more specific hybridization results. It should be noted in this
35 context, that probes of the same length, regardless of their GC

content, will hybridize specifically at approximately the same temperature in TMACI solutions.

Suitable assay methods for purposes of the present invention to
5 detect hybrids formed between the oligonucleotide probes and the
HCBI, MSBI, MSSI or CMI polynucleic acid sequences in a sample
may comprise any of the assay formats known in the art, such as
the conventional dot-blot format, sandwich hybridization or
reverse hybridization. For example, the detection can be
10 accomplished using a dot blot format, the unlabelled amplified
sample being bound to a membrane, the membrane being
incorporated with at least one labelled probe under suitable
hybridization and wash conditions, and the presence of bound
probe being monitored.

15 An alternative and preferred method is a "reverse" dot-blot
format, in which the amplified sequence contains a label. In
this format, the unlabelled oligonucleotide probes are bound to
a solid support and exposed to the labelled sample under
20 appropriate stringent hybridization and subsequent washing
conditions. It is to be understood that also any other assay
method which relies on the formation of a hybrid between the
polynucleic acids of the sample and the oligonucleotide probes
according to the present invention may be used.

25 The present invention also relates to a method for detecting a
polypeptide encoded by an HCBI, MSBI, MSSI or CMI polynucleic
acid of the present invention or an antibody against said
polypeptide present in a biological sample, comprising: (a)
30 contacting the biological sample for the presence of such
polypeptide or antibody as defined above, and (b) detecting the
immunological complex formed between said antibody and said
polypeptide.

35 The immunoassay methods according to the present invention may

utilize antigens from different domains of the new and unique polypeptide sequences of the present invention. It is within the scope of the invention to use for instance single or specific oligomeric antigens, dimeric antigens, as well as combinations of single or specific oligomeric antigens. The HCBI, MSBI, MSSSI or CMI antigens of the present invention may be employed in virtually any assay format that employs a known antigen to detect antibodies or cell-mediated immune responses. Thus, the present invention also encompasses the detection of cell-mediated immune responses against HCBI, MSBI, MSSSI or CMI antigens and the application of therapeutic interferences based on cell-mediated immune responses against HCBI, MSBI, MSSSI or CMI antigens.

Of course, an assay format that denatures the HCBI, MSBI, MSSSI or CMI conformational epitope should be avoided or adapted. A common feature of all of these assays is that the antigen is contacted with the body component suspected of containing HCBI, MSBI, MSSSI or CMI antibodies under conditions that permit the antigen to bind to any such antibody present in the component. Such conditions will typically be physiologic temperature, pH and ionic strength using an excess of antigen. The incubation of the antigen with the specimen is followed by detection of immune complexes comprised of the antigen.

Design of the immunoassays is subject to a great deal of variation, and many formats are known in the art. Protocols may, for example, use solid supports, or immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, enzymatic, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the immune complex are also known; examples of which are assays which utilize biotin and avidin or streptavidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

The immunoassay may be in a heterogeneous or in a homogeneous format, and of a standard or competitive type. In a heterogeneous format, the polypeptide is typically bound to a solid matrix or support to facilitate separation of the sample from the polypeptide after incubation. Examples of solid supports that can be used are nitrocellulose (e. g., in membrane or microtiter well form), polyvinyl chloride (e. g., in sheets or microtiter wells), polystyrene latex (e. g., in beads or microtiter plates, polyvinylidene fluoride (known as Immulon), diazotized paper, nylon membranes, activated beads, and Protein A beads. The solid support containing the antigenic polypeptides is typically washed after separating it from the test sample, and prior to detection of bound antibodies. Both standard and competitive formats are known in the art.

In a homogeneous format, the test sample is incubated with the combination of antigens in solution. For example, it may be under conditions that will precipitate any antigen-antibody complexes which are formed. Both standard and competitive formats for these assays are known in the art.

In a standard format, the amount of HCBI, MSBI, MSSSI or CMI antibodies in the antibody-antigen complexes is directly monitored. This may be accomplished by determining whether (labelled) anti-xenogeneic (e. g. anti-human) antibodies which recognize an epitope on anti-HCBI, -MSBI, -MSSSI or -CMI antibodies will bind due to complex formation. In a competitive format, the amount of HCBI, MSBI, MSSSI or CMI antibodies in the sample is deduced by monitoring the competitive effect on the binding of a known amount of labelled antibody (or other competing ligand) in the complex.

Complexes formed comprising anti-HCBI, -MSBI, -MSSSI or -CMI antibody (or in the case of competitive assays, the amount of

competing antibody) are detected by any of a number of known techniques, depending on the format. For example, unlabeled HCBI, MSBI, MSSSI or CMI antibodies in the complex may be detected using a conjugate of anti-xenogeneic Ig complexed with
5 a label (e. g. an enzyme label).

In an immunoprecipitation or agglutination assay format the reaction between the HCBI, MSBI, MSSSI or CMI antigens and the antibody forms a network that precipitates from the solution or
10 suspension and forms a visible layer or film of precipitate. If no anti-HCBI, -MSBI, -MSSSI or -CMI antibody is present in the test specimen, no visible precipitate is formed.

There currently exist three specific types of particle
15 agglutination (PA) assays. These assays are used for the detection of antibodies to various antigens when coated to a support. One type of this assay is the hemagglutination assay using red blood cells (RBCs) that are sensitized by passively adsorbing antigen (or antibody) to the RBC. The addition of
20 specific antigen/antibodies present in the body component, if any, causes the RBCs coated with the purified antigen to agglutinate.

To eliminate potential non-specific reactions in the
25 hemagglutination assay, two artificial carriers may be used instead of RBC in the PA. The most common of these are latex particles.

The solid phase selected can include polymeric or glass beads,
30 nitrocellulose, microparticles, microwells of a reaction tray, test tubes and magnetic beads. The signal generating compound can include an enzyme, a luminescent compound, a chromogen, a radioactive element and a chemiluminescent compound. Examples of enzymes include alkaline phosphatase, horseradish peroxidase and
35 beta-galactosidase. Examples of enhancer compounds include

WO 2016/005054 PCT/EP2015/001399
biotin, anti-biotin and avidin. Examples of enhancer compounds
binding members include biotin, anti-biotin and avidin.

The above methods are useful for evaluating the risk of
5 developing diseases like cancer or an autoimmune disease due to
the deleterious effects of the presence of a subgenomic HCBI,
MSBI, MSSSI or CMI polynucleotide sequence by itself or linked to
a particular host gene or gene fragment within the patient's
cells and allow taking appropriate counter measures.

10 Thus, the present invention also relates to an antisense
oligonucleotide or iRNA specific for the HCBI, MSBI, MSSSI or CMI
virus polynucleic acid of the invention.

15 The generation of suitable antisense oligonucleotides or iRNAs
includes determination of a site or sites within the HCBI, MSBI,
MSSSI or CMI polynucleic acid for the antisense interaction to
occur such that the desired effect, e.g., inhibition of
expression of the polypeptide, will result. A preferred
20 intragenic site is (a) the region encompassing the translation
initiation or termination codon of the open reading frame (ORF)
of the gene or (b) a region of the mRNA which is a "loop" or
"bulge", i.e., not part of a secondary structure. Once one or
more target sites have been identified, oligonucleotides are
25 chosen which are sufficiently complementary to the target, i.e.,
hybridize sufficiently well and with sufficient specificity, to
give the desired effect. In the context of this invention,
"hybridization" means hydrogen bonding, which may be Watson-
Crick, Hoogsteen or reversed Hoogsteen hydrogen bonding, between
30 complementary nucleoside or nucleotide bases. "Complementary" as
used herein, refers to the capacity for precise pairing between
two nucleotides. For example, if a nucleotide at a certain
position of an oligonucleotide is capable of hydrogen bonding
with a nucleotide at the same position of a DNA or RNA molecule,
35 then the oligonucleotide and the DNA or RNA are considered to be

complementary to each other at that position. The oligonucleotide and the DNA or RNA are complementary to each other when a sufficient number of corresponding positions in each molecule are occupied by nucleotides which can hydrogen bond with each other. Thus, "specifically hybridizable" and "complementary" are terms which are used to indicate a sufficient degree of complementarity or precise pairing such that stable and specific binding occurs between the oligonucleotide and the DNA or RNA target. It is understood in the art that the sequence of an antisense compound does not need to be 100% complementary to that of its target nucleic acid to be specifically hybridizable. An antisense compound is specifically hybridizable when binding of the compound to the target DNA or RNA molecule interferes with the normal function of the target DNA or RNA to cause a loss of utility, and there is a sufficient degree of complementarity to avoid non-specific binding of the antisense compound to non-target sequences under conditions in which specific binding is desired, i.e., in the case of therapeutic treatment.

"Oligonucleotide" (in particular in the context of antisense compounds) refers to an oligomer or polymer of ribonucleic acid (RNA) or deoxyribonucleic acid (DNA) or mimetics thereof. This term includes oligonucleotides composed of naturally-occurring nucleobases, sugars and covalent internucleoside (backbone) linkages as well as oligonucleotides having non-naturally-occurring portions which function similarly. Such modified or substituted oligonucleotides are often preferred over native forms because of desirable properties such as, for example, enhanced cellular uptake, enhanced affinity for nucleic acid target and increased stability in the presence of nucleases. While antisense oligonucleotides are a preferred form of the antisense compound, the present invention comprehends other oligomeric antisense compounds, including but not limited to oligonucleotide mimetics such as are described below. The

antisense compounds in accordance with this invention comprise from about 8 to about 50 nucleobases (i.e. from about 8 to about 50 linked nucleosides). Particularly preferred antisense compounds are antisense oligonucleotides, even more preferably those comprising from about 15 to about 25 nucleobases. Antisense compounds include ribozymes, external guide sequences (EGS), oligonucleotides (oligozymes), and other short catalytic RNAs or catalytic oligonucleotides which hybridize to the target nucleic acid and inhibit its expression. The antisense compounds also include an iRNA comprising a sense sequence and an antisense sequence, wherein the sense and antisense sequences form an RNA duplex and wherein the antisense sequence comprises a nucleotide sequence sufficiently complementary to the nucleotide sequence of an HCBI, MSBI, MSSI or CMI polynucleic acid of the present invention.

Alternatively, the invention provides a vector allowing to transcribe an antisense oligonucleotide of the invention, e.g., in a mammalian host. Preferably, such a vector is a vector useful for gene therapy. Preferred vectors useful for gene therapy are viral vectors, e.g. adenovirus, adeno-associated virus, herpes simplex virus, vaccinia, or, an RNA virus such as a retrovirus. Preferably, the retroviral vector is a derivative of a murine or avian retrovirus. Examples of such retroviral vectors which can be used in the present invention are: Moloney murine leukemia virus (MoMuLV), Harvey murine sarcoma virus (HaMuSV), murine mammary tumor virus (MuMTV) and Rous sarcoma virus (RSV). Most preferably, a non-human primate retroviral vector is employed, such as the gibbon ape leukemia virus (GaLV), providing a broader host range compared to murine vectors. Since recombinant retroviruses are defective, assistance is required in order to produce infectious particles. Such assistance can be provided, e.g., by using helper cell lines that contain plasmids encoding all of the structural genes of the retrovirus under the control of regulatory sequences

within the LTR. Suitable helper cell lines are well known to those skilled in the art. Said vectors can additionally contain a gene encoding a selectable marker so that the transduced cells can be identified. Moreover, the retroviral vectors can be modified in such a way that they become target specific. This can be achieved, e.g., by inserting a polynucleotide encoding a sugar, a glycolipid, or a protein, preferably an antibody. Those skilled in the art know additional methods for generating target specific vectors. Further suitable vectors and methods for in vitro- or in vivo-gene therapy are described in the literature and are known to the persons skilled in the art; see, e.g., WO 94/29469 or WO 97/00957. The HCBI, MSBI, MSSSI or CMI polynucleotide sequences of the invention may also serve as a suitable vector itself, either composed solely of rearranged HCBI, MSBI, MSSSI or CMI sequences or of chimeric HCBI, MSBI, MSSSI or CMI host cell DNA sequences. In addition, the nucleotide sequences of the invention may be used for the construction of artificial chromosomes.

In order to achieve expression only in the target organ, the DNA sequences for transcription of the antisense oligonucleotides can be linked to a tissue specific promoter and used for gene therapy. Such promoters are well known to those skilled in the art.

Within an oligonucleotide structure, the phosphate groups are commonly referred to as forming the internucleoside backbone of the oligonucleotide. The normal linkage or backbone of RNA and DNA is a 3' to 5' phosphodiester linkage. Specific examples of preferred antisense compounds useful in the present invention include oligonucleotides containing modified backbones or non-natural internucleoside linkages. Oligonucleotides having modified backbones include those that retain a phosphorus atom in the backbone and those that do not have a phosphorus atom in the backbone. Modified oligonucleotide backbones which can

result in increased stability are known to the person skilled in the art, preferably such modification is a phosphorothioate linkage.

5 A preferred oligonucleotide mimetic is an oligonucleotide mimetic that has been shown to have excellent hybridization properties, and is referred to as a peptide nucleic acid (PNA). In PNA compounds, the sugar-backbone of an oligonucleotide is replaced with an amide containing backbone, in particular an
10 aminoethylglycine backbone. The nucleobases are retained and are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone.

Modified oligonucleotides may also contain one or more
15 substituted or modified sugar moieties. Preferred oligonucleotides comprise one of the following at the 2' position: OH; F; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; O-, S- or N-alkynyl; or O-alkyl-O-alkyl, wherein the alkyl, alkenyl and alkynyl may be substituted or unsubstituted C₁ to C₁₀ alkyl or
20 C₂ to C₁₀ alkenyl and alkynyl. A particularly preferred modified sugar moiety is a 2'-O-methoxyethyl sugar moiety.

Antisense-oligonucleotides of the invention may also include nucleobase modifications or substitutions. Modified nucleobases
25 include other synthetic and natural nucleobases such as 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine
30 and 2-thiocytosine etc., with 5-methylcytosine substitutions being preferred since these modifications have been shown to increase nucleic acid duplex stability.

Another modification of the oligonucleotides of the invention
35 involves chemically linking to the oligonucleotide one or more

moieties or conjugates which enhance the activity, cellular distribution or cellular uptake of the oligonucleotide. Such moieties include lipid moieties such as a cholesterol moiety, cholic acid, a thioether, a thiocholesterol, an aliphatic chain, e.g., dodecandiol or undecyl residues, a phospholipid, a polyamine or a polyethylene glycol chain, or adamantane acetic acid, a palmityl moiety, or an octadecylamine or hexylamino-carbonyl-oxycholesterol moiety.

10 The present invention also includes antisense compounds which are chimeric compounds. "Chimeric" antisense compounds or "chimeras," in the context of this invention, are antisense compounds, particularly oligonucleotides, which contain two or more chemically distinct regions, each made up of at least one monomer unit, i.e., a nucleotide in the case of an oligonucleotide compound. These oligonucleotides typically contain at least one region wherein the oligonucleotide is modified so as to confer upon the oligonucleotide increased resistance to nuclease degradation, increased cellular uptake, and/or increased binding affinity for the target nucleic acid. An additional region of the oligonucleotide may serve as a substrate for enzymes capable of cleaving RNA:DNA or RNA:RNA hybrids. By way of example, RNase H is a cellular endonuclease which cleaves the RNA strand of an RNA:DNA duplex. Activation of RNase H, therefore, results in cleavage of the RNA target, thereby greatly enhancing the efficiency of oligonucleotide inhibition of gene expression. Consequently, comparable results can often be obtained with shorter oligonucleotides when chimeric oligonucleotides are used, compared to phosphorothioate deoxyoligonucleotides hybridizing to the same target region. Chimeric antisense compounds of the invention may be formed as composite structures of two or more oligonucleotides, modified oligonucleotides, oligonucleosides and/or oligonucleotide mimetics as described above. Such compounds have also been referred to in the art as hybrids or gapmers.

The present invention also relates to a pharmaceutical composition comprising an antibody or antisense oligonucleotide of the invention and a suitable excipient, diluent or carrier.

5 Preferably, in a pharmaceutical composition, such compound as described above is combined with a pharmaceutically acceptable carrier. "Pharmaceutically acceptable" is meant to encompass any carrier, which does not interfere with the effectiveness of the biological activity of the active ingredient and that is not

10 toxic to the host to which it is administered. Examples of suitable pharmaceutical carriers are well known in the art and include phosphate buffered saline solutions, water, emulsions, such as oil/water emulsions, various types of wetting agents, sterile solutions etc. Such carriers can be formulated by

15 conventional methods and the active compound can be administered to the subject at an effective dose.

An "effective dose" refers to an amount of the active ingredient that is sufficient to prevent the disease or to affect the

20 course and the severity of the disease, leading to the reduction or remission of such pathology. An "effective dose" useful for treating and/or preventing these diseases or disorders may be determined using methods known to one skilled in the art.

25 Administration of the suitable compositions may be effected by different ways, e.g. by intravenous, intraperitoneal, oral, subcutaneous, intramuscular, topical or intradermal administration. The route of administration, of course, depends on the kind of therapy and the kind of compound contained in the

30 pharmaceutical composition. The dosage regimen will be determined by the attending physician and other clinical factors. As is well known in the medical arts, dosages for any one patient depends on many factors, including the patient's size, body surface area, age, sex, the particular compound to be

35 administered, time and route of administration, the kind of

WO 2016/005054 PCT/EP2015/001399
therapy, general health and other drugs being administered
concurrently.

In a preferred embodiment of the present invention, the disease
5 that can be prevented/treated is cancer, preferably breast
cancer, ovarian cancer, lung cancer, prostate cancer, pancreatic
cancer, Hodgkin's disease, colorectal cancer or colon cancer or
a disease of the CNS, preferably Alzheimer's disease or multiple
sclerosis (MS), amyotrophic lateral sclerosis, Parkinson's
10 disease, or transmissible spongiforme encephalopathies/Prion-
linked diseases. In addition, due to a similarity of risk
factors between MS and diabetes mellitus, the latter condition
is also included. The terms "cancer" and "disease of the CNS"
also comprise early stages of said diseases.

15

The present invention also relates to a vaccine for immunizing a
mammal against an HCBI, MSBI, MSSI or CMI infection, comprising
at least one polypeptide or HCBI, MSBI, MSSI or CMI polynucleic
acid as defined above or corresponding VLP (virus-like particle)
20 or peptide/protein/DNA complexes, in a pharmaceutically
acceptable carrier. It also involves molecular and immunological
tests in animals (in particular cattle) and within their
products (e.g. milk and dairy products).

25 It may also include small chemicals for targeted therapy derived
from the analysis of structural components of these agents.

A "vaccine" is an immunogenic composition capable of eliciting
protection against HCBI, MSBI, MSSI or CMI, whether partial or
30 complete. A vaccine may also be useful for treatment of an
already infected individual, in which case it is called a
therapeutic vaccine.

The term "therapeutic" refers to a composition capable of
35 treating HCBI, MSBI, MSSI or CMI infection or diseases linked to

this infection. The term "effective amount" refers to an amount of epitope-bearing polypeptide sufficient to induce an immunogenic response in the individual to which it is administered, or to otherwise detectably immunoreact in its intended system (e. g., immunoassay). Preferably, the effective amount is sufficient to effect treatment, as defined above. The exact amount necessary will vary according to the application. For vaccine applications or for the generation of polyclonal antiserum/antibodies, for example, the effective amount may vary depending on the species, age, and general condition of the individual, the severity of the condition being treated, the particular polypeptide selected and its mode of administration, etc. Effective amounts will be found within a relatively large, non-critical range. An appropriate effective amount can be readily determined using routine experimentation. Preferred ranges of proteins for prophylaxis of HCBI, MSBI, MSSSI or CMI caused diseases are 0.01 to 100 µg/dose, preferably 0.1 to 50 µg/dose. Several doses may be needed per individual in order to achieve a sufficient immune response and subsequent protection against an HCBI, MSBI, MSSSI or CMI infection and an HCBI, MSBI or CMI linked disease, respectively.

Pharmaceutically acceptable carriers include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the vaccine. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, and amino acid copolymers. Such carriers are well known to those of ordinary skill in the art.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: aluminium hydroxide (alum), N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP) as found in U.S. Patent No. 4,606,918, N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-

hydroxy-phosphoryloxy)-ethylamine (MTP-PE) and RIBI, which
contains three components extracted from bacteria,
monophosphoryl lipid A, trehalose dimycolate, and cell wall
5 Skeleton (MPL +TDM + CWS) in a 2% squalene/Tween 80 emulsion.
Any of the 3 components . MPL, TDM or
CWS may also be used alone or combined 2 by 2. Additionally,
adjuvants such as Stimulon (Cambridge Bioscience, Worcester, MA)
or SAF-1 (Syntex) may be used. Further, Complete Freund's
10 Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA) may be
used for non-human applications and research purposes.

The immunogenic compositions typically will contain
pharmaceutically acceptable vehicles, such as water, saline,
15 glycerol, ethanol, etc. Additionally, auxiliary substances, such
as wetting or emulsifying agents, pH buffering substances,
preservatives, and the like, may be included in such vehicles.

Typically, the immunogenic compositions are prepared as
20 injectables, either as liquid solutions or suspensions. Solid
forms suitable for solution in, or suspension in, liquid
vehicles prior to injection may also be prepared. The
preparation also may be emulsified or encapsulated in liposomes
for enhanced adjuvant effect. The proteins may also be
25 incorporated into Immune Stimulating Complexes together with
saponins, for example Quil A (ISCOMS).

Immunogenic compositions used as vaccines comprise a "sufficient
amount" or "an immunologically effective amount" of the proteins
30 of the present invention, as well as any other of the above
mentioned components, as needed. "Immunologically effective
amount" means that the administration of that amount to an
individual, either in a single dose or as part of a series, is
effective for treatment, as defined above. This amount varies
35 depending upon the health and physical condition of the

individual to be treated, the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors.

5 It is expected that the amount will fall in a relatively broad range that can be determined through routine trials. Usually the amount will vary from 0.01 to 1000 µg/dose, more particularly from 0.1 -100 µg/dose.

10

The following examples are intended to illustrate, but not to limit the invention. While such examples are typical of those that might be used, other methods known to those skilled in the art may alternatively be utilized.

15

Example 1

Material and Methods

20 (A) Fractionation of bovine sera on density-sedimentation gradients with subsequent cloning

Initially, pools of 5 sera from a total of 120 bovine sera were subjected to Optiprep-(iodixanol)-density gradient ultracentrifugation after prior benzonase treatment to remove
25 all free DNA and RNA (Buck et al., 2005). Protein-associated DNA was extracted from fractions (Qiagen PCR Purification Kit) and 1µl DNA/fraction subjected to RCA (rolling circle amplification) in a solution of 50µM Exo-resistant random primers (Thermo Scientific), 3.2µmol each dNTPs (Takara) and
30 10U phi29 polymerase (Biolabs). Restriction digested products (EcoR1 or BamH1) were separated by agarose gel electrophoresis, eluted and cloned into vector pUC19 prior to sequencing.

(B) Rolling circle amplification of DNA extracted from sera, cow milk or brain tissue: DNA was extracted by phenol-chloroform from milk and post mortem brain tissue and sera from MS patients. DNA from all serum samples was extracted using the High Pure Viral Nucleic Acid Kit (Roche). RCA (rolling circle amplification) with random primers on DNA from protein-associated fractions, restriction digestion, cloning and sequencing of resulting fragments (refer above). Abutting primers used were designed either on the individual isolated DNA sequences, as well as on the replication genes of Sphinx1.76 or Sphinx2.36 and used in inverted PCR on RCA amplified DNA from single bovine sera and cow milk, as well as sera from multiple sclerosis patients and post mortem multiple sclerosis brain samples.

Example 2

Concept for the pathogenesis of multiple sclerosis:

Isolation of circular DNA molecules (bovine agents) from bovine serum, cow milk and multiple sclerosis brain (Group 1)

The epidemiology of colon cancer suggested the involvement of an infectious factor present in red meat derived from cattle of European/Asian descent (zur Hausen, 2012; zur Hausen, 2015) and cow milk consumption has been suspected to play a role in multiple sclerosis. In attempts to isolate these putative factors, sera from 120 healthy 5-year old cows were obtained from the Veterinary Faculty of the University of Leipzig and analyzed for the presence of circular episomal DNA. Since the first isolates HCBI6.252 (Healthy Cattle Blood Isolate)(2522bp) and HCBI6.159 (1591bp) revealed a distant relationship to DNA related to sequences found in brain

lesions of animals linked to prion-associated conditions (Manuelidis, 2011). The inventors concomitantly analysed 8 sera (from patients in relapse), 2 CSF and 1 PBMC from MS patients, as well as 12 biopsies from post mortem brain tissue for Sphinx-related sequences. Two circular DNA molecules related to Sphinx1.76 (1758bp acc.no. HQ444404) were isolated from one MS brain sample - MSBI1.176 (Multiple Sclerosis Brain Isolate) (1766bp) and MS2.176 (1766bp). Since there is an elevated MS risk after cow milk consumption, the inventors investigated commercially available pasteurized milk for the presence of related DNA. Indeed, they isolated episomal single-stranded DNA molecules from all 4 milk samples (CMI1.252 (Cow Milk Isolate), CMI2.214, CMI3.168 and CMI4.158) (HCBI6.252 and CMI1.252 are near identical). This was taken as an indication that milk excretion of these agents is indeed occurring.

The inventors used 2 primer pairs designed on Sphinx1.76 for inverted PCR on all human and bovine samples. These primer pairs were: forward 5'-GGATTAATGCCAATGATCC-3' (nt 721-739), reverse 5'-CGAGAGAAACAGGCAAAG-3' (nt703-720) and forward 5'-GAGGACGAATTAATATTACAAGTC-3' (nt868-891), reverse 5'-TTACCAAGAAAAGCGAGAAC-3' (nt848-867). The resulting sequences are all distantly similar (ranging from 79% - 98%) to the Sphinx1.76 isolate. HCBI6.159 apparently evolved from a deletion in HCBI6.252 (deletion of nt 1129-2060 in HCBI6.252) as their overlapping sequences are identical. HCBI6.159 was isolated independently from HCBI6.252 and it is therefore highly unlikely that it is an artifact. MSBI1.176 is 98% identical to Sphinx1.76, but the nature (patterns) of the single sequence differences are such that these can be regarded as two separate agents. As the Sphinx1.76 construct was not available in the inventor's laboratory, it could not have resulted from laboratory contamination. The inventors

WO 2016/005054 **PCT/EP2015/001399**
isolated a second very distantly Sphinx1.76-related (but identical in size) circular DNA molecule MSBI2.176 from the same brain biopsy.

5 The large ORFs of the isolate of group 1 encode for replication protein (ProtSweep, del Val et al., 2007) sharing high similarity between them. Another common feature is the presence of iteron-like tandem repeats (3x22nt plus 17/18nt of the repeat in each isolate). Alignment of this repeat region
10 indicates only single nucleotide variation in the core (Figure 6). These iteron-like repeats may constitute binding sites for Rep proteins (Chattoraj, 2000, Dziewit et al., 2013).

Nucleotide sequence accession number: The complete sequences
15 of 8 isolates have been deposited in the EMBL Databank under the acc. no.:

CMI1.2522	Acc no. LK931487
CMI2.214	Acc no. LK931488
CMI3.168	Acc no. LK931489
20 CMI4.158	Acc no. LK931490
MSBI1.17	Acc no. LK931491
MSBI2.176	Acc no. LK931492
HCBI6.252	Acc no. LK931493
HCBI6.159	Acc no. LK931494

25 In this context, diseases of the CNS (e.g. Multiple sclerosis MS, amyotrophic lateral sclerosis, transmissible spongiforme encephalopathies/ Prion-linked diseases, Parkinson's disease, Alzheimer disease) are also highly interesting since the
30 similar sequences described by Manuelidis are primarily found in the CNS.

Example 3**Isolation of a bovine factor from bovine serum (group 2)**

5 DNA from bovine serum was extracted followed by rolling circle amplification of the DNA. PCR amplification using abutting primers (forward (nt2313-2336) 5'CTAATGCAGATCAACACAGGGATA-3' and reverse (nt2312-2291) 5'-GAATTACAGGCTTTGCAATCTG-3') designed on the replication gene of Sphinx2.36 (2364bp, acc no. HQ444405)
10 (Manuelidis 2011) resulted in one circular DNA isolate HCBI7.228 (2280bp, acc no. LK931498) (group 2).

Example 4

15 **Isolation of sequences from bovine sera and multiple sclerosis serum distantly related to a novel group of single-stranded DNA viruses, the myco-like Gemycircularviruses (Group 3)**

A large number of circular ssDNA viruses have recently been
20 identified by metagenomic analyses (Rosario et al., 2012a). The group of Gemycircularviruses has been isolated from fungi, faeces of various animals and plants leaves (Rosario et al., 2012b; Sikorski et al. 2013). The inventors report the isolation of the genomes of 3 novel viruses distantly related to this
25 virus group.

Two sequences HCBI8.215 (2152bp) and HCBI9.2112 (2121bp) were isolated after density gradient fractionation of bovine serum pools. Fragments obtained after restriction digest (BamH1 and
30 EcoR1) were cloned into pUC19. Their full-length genomes were verified by inverted PCR using abutting primers. One multiple sclerosis serum isolate MSSI2.225 rescued using abutting primers and inverted PCR was related to the myco-like Gemycircularviruses isolated from cattle sera (HCBI8.215 and
35 HCBI9.212). The inventors subsequently isolated a sequence

MSBI3.224 from a post-mortem-MS affected brain tissue which proved to be identical to MSSI2.225.

5 The genome organization of all 3 isolates revealed a putative
spliced replication protein coded for on the negative strand and
the coat protein (CP) on the positive strand. The CP was
arginine rich and a similarity to the TTV ORF1 protein was
indicated in a DomainSweep analysis (del Val et al., 2007).
Putative rolling circle motifs I, II and III and a Walker B
10 motif of each were identified as follows: HCBI8.215 (LLTYA,
HLHAFVD, YAIKD; VFDDI), for HCBI9.212 (LLKMP, HYHIYLG, YVGKD;
VFDDI) and for MSSI2.225 (LLTYP, HLHAFVD, YAIKD; IFDDF). The GRS
motifs were AVFDVGGFHPNISITK, TAFDYFGAHGNIKSIR and
RAFDEGCHPNVSPSR respectively. The nona-nucleotide motif for
15 both HCBI8.215 and MSSI2.225 is (TAATGTTAT) and for HCBI9.212
(TAATATTAT).

These 3 isolates probably constitute a novel group of viruses as
their capsid proteins share similarities with the ORF1 of TT
20 viruses from the family *Anelloviridae*, and not to that of the
plant virus family *Geminiviridae* as in the Gemycircularviruses.
The isolation of these 3 Gemycircularvirus-related genomes
directly from animal and diseased human tissue links them to
have some etiological relevance for diseases such as multiple
25 sclerosis.

HCBI8.215	Acc no. LK931483
HCBI9.212	Acc no. LK931484
MSSI2.225	Acc no. LK931485

30

35

Bacterial plasmid-related circular DNA from human serum (group 4)

5 *Psychrobacter* species are frequently present as food contaminants and have been isolated from human tissues including brain, cerebrospinal fluid and blood. It is considered as an opportunistic human pathogen (Caspar et al., 2013; Lloyd-Puryear et al., 1991). The inventors report the isolation of a circular
10 DNA molecule, MSSII.162, from serum taken from a multiple sclerosis patient during relapse.

Rolling circle amplification and restriction digest were performed on DNA extracted from sera from patients with multiple
15 sclerosis. The resulting fragment was cloned into vector pUC19. The full-length genome was verified by inverted PCR amplification using primers designed on the sequence identified initially: forward primer 5'-GACTTCTGATTGATTGATGCCTG-3' and reverse 5'-CCTGTTGAATACCGCTTAAATACT-3'. All products were
20 sequenced by primer walking. MSSII.162 (multiple sclerosis serum isolate) (1627bp) (Acc no. LK931486) shares 66% nucleotide similarity by BlastN analysis to the pRWF102 plasmid of the *Psychrobacter* species PRwf-1. The putative protein (321 amino acids) encoded by the large ORF shows weak similarity to
25 replication protein family of *E. coli* as analysed by ProtSweep (del Val et al., 2007).

This new isolate is only distantly related to known *Psychrobacter* species and their plasmids. It may therefore
30 represent a yet unknown human pathogen.

Summary:

The presence of presumably infectious agents and their nucleic acids in the serum of healthy cows should imply that the same particles are also present in red meat.

The inventors isolated 13 novel single-stranded DNA molecules from cattle serum and milk and MS brain tissue and sera. These isolates are grouped in 4 groups according to their sequence similarity to the Sphinx1.76 genome (group 1), Sphinx2.36 genome (group 2), myco-like Gemycircularviruses (group 3) and *Psychrobacter* spp. plasmid (group 4). Their genome organizations are presented in Figure 5 A,B. The main feature of all the sequences is the presence of a replication-associated protein encoding ORF.

All the isolates are presumably single-stranded DNA because of the bias of RCA towards amplification of single-stranded DNA (del Solar et al., 1998). A taxonomic classification of the isolates is, at this stage, not possible. Only group 3 can be defined as clearly related to the myco-like Gemycircularviruses. This relatively novel group of viruses have been identified by metagenomics and isolated mainly from faeces of a very large variety of animals and insects, as well as plants and fungi. They share some characteristics with the Geminiviruses of plants (*Sikorski et al., 2013*). Isolation from tissues of animals or humans has not yet been described. The putative replication-associated protein of all 3 isolates described in this application has an arginine-rich region with a similarity to human single-stranded TT viruses.

Infection of human cells by such agents should evoke a strong immune reaction, quite distinct from human TT viruses, where reasonable evidence for vertical transmission has been obtained (reviewed in zur Hausen and de Villiers, 2014). This could

explain the high susceptibility to environmental factors for MS development during the first 15 years of life: primary infection may initially lead to rounds of replication and spreading of BMF probably via blood cells, eventually resulting in latent brain cell infection. This initial infection should induce an immune response, probably neutralizing the agent in subsequent rounds of infection prior to entry of the brain. The isolates reported here seem to represent excellent candidates for the postulated bovine milk factor (BMF).

10

A high variability in size was noted in group 1. The circular isolate HCBI6.159 seem to have evolved from HCBI6.252 through deletion of 931 nucleotides from the latter. The isolates all possess a replication gene and have an iteron-like repeat region in common (*Dziewit et al., 2013*). Alignment of this region between 8 isolates and Sphinx1.76 reveals a central identical core (Figure 6). Group 2 and 4 isolates do not have repeat regions.

15

20 The "Sphinx" sequences (*Manuelidis, 2011*) show high homologies to plasmid sequences of the bacterium *Acinetobacter* (*Vallenet et al., 2008; Longkumer et al., 2013*). The sequences obtained in the present invention also exhibit striking homologies to the corresponding plasmid sequences. Although a large number of plasmids have been isolated and sequenced from *Acinetobacter*, thus far none of them corresponded exactly to the bovine and human sequences reported in this invention. Interestingly, a group of scientists in the UK published serological data over a period of years pointing to an increased selective formation of antibodies against *Acinetobacter* proteins but not against other bacterial antigens obtained from patients suffering from multiple sclerosis (see review article: *Ebringer et al., 2012*). These results could not be confirmed by the group of Chapman (*Chapman et al., 2005*). However, it has to be stressed that the group of Chapman used a different strain of *Acinetobacter*

30

35

(*Acinetobacter calcoaceticus*). Unequivocal results were obtained by the group of Ebringer for three strains of *Acinetobacter* (*Acinetobacter lwoffii*, *A. radioesistens* and a specific isolate, *A. 11171*). However, the results obtained for *A. junii* 17908 were less impressive and significant reactivity was hardly detectable (Hughes et al., 2001). These results suggest that we are dealing with strain-specific reactivities wherein this sero-reactivity is due to strain-specific plasmids exhibiting homologies to the DNA sequences obtained in the present invention.

The isolate MSS11.162 (group 4) has similarity to a plasmid of the *Psychrobacter* spp. *Psychrobacter* species have been considered as an opportunistic human pathogen (Caspar et al., 2013) and has been isolated from a case of meningitis (Lloyd-Puryear et al., 1991). These bacteria have repeatedly been reported as contaminants during and after cold-storage of meat (de Filippis et al., 2013) and were frequently isolated from milk and a variety of cheeses (Coton et al., 2012).

It is of interest to note that Manuelidis reported two "Sphinx-structures", labeled as "large" (2.36) and "small" (1.76) Sphinx. Although most of the present sequences substantially differed from her isolates, the inventors also obtained large and small Sphinx-like sequences from the same probes. Circular HCBI6.159 seems to have evolved from HCBI6.252 through a deletion of 931 nucleotides from the latter. It cannot be excluded that the other larger isolates may have smaller counterparts which were not isolated. It remains, however, to be determined whether the two structures found here persist within the same protein coat or complement each other.

The isolation of DNA of similar, in part even identical single-stranded circular nucleic acids from cattle sera, commercially available cow milk and florid MS tissues argues in favour of the concept outlined above.

References

- 5 Buck CB, Pastrana DV, Lowy DR, Schiller JT. Efficient intracellular assembly of papillomaviral vectors. J. Virol. 2004; 78:751-757.
- 10 Chapman MD, Hughes LE, Wilson CD, Namnyak S, Thompson EJ, Giovannoni G. No evidence for production of intrathecal immunoglobulin G against *Acinetobacter* or *Pseudomonas* in multiple sclerosis. Eur Neurol. 2005; 53(1):27-31.
- 15 de Villiers EM, Borkosky SS, Kimmel R, Gunst K, and Fei JW. (2011) The diversity of Torque teno viruses: In vitro replication leads to the formation of additional replication-competent subviral Molecules. J Virol 2011; 85(14):7284-7295
- 20 Ebringer A, Hughes L, Rashid T, Wilson C. *Acinetobacter* Immune Responses in Multiple Sclerosis: Etiopathogenetic Role and Its Possible Use as a Diagnostic Marker. Arch Neurol. 2005; 62:33-36.
- 25 Ebringer A, Rashid T, Wilson C. The role of *Acinetobacter* in the pathogenesis of multiple sclerosis examined by using Popper sequences. Med Hypotheses. 2012; 78(6):763-769.
- 30 Hughes, L.E., Bonell, S., Natt, R.S., Wilson, C., Tiwana, H., Ebringer, A., Cunningham, P., Chamoun, V., Thompson, E.J., Croker, J., and Vowles, J. Antibody responses to *Acinetobacter* spp. and *Pseudomonas aeruginosa* in multiple sclerosis: prospects for diagnosis using the myelin-*Acinetobacter*-neurofilament antibody index. Clin. Diagn. Laboratory Immunol. 2001; 8: 1181-1188.

Longkumer T, Kamireddy S, Muthyala VR, Akbarpasha S, Pitchika GK, Kodetham G, Ayaluru M, Siddavattam D. Scientific Reports 2013; 3:2240.

Manuelidis L. Nuclease resistant circular DNAs copurify with infectivity in scrapie and CJD. J. Neurovirol. 2011; 17:131-145.

Wold, MS. Replication protein A: heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. Ann. Review Biochem. 1997; 66:61-92

Vallenet D, Nordmann P, Barbe V, Poirol L, Mangenot S, Bataille E, Dossat C, Gas S, Kreimeyer A, Lenoble P, Oztas S, Poulain J, Segurens B, Robert C, Abergel C, Claverie J-M, Raoult D, Medigue C, Weissenbach J, Cruveiller S. Comparative analysis of Acinetobacters: three genomes for three lifestyles. PLoS One 2008; 3(3):e1805-e1805.

Xu B, Zhi N, Hu G, Wan Z, Zheng X, Liu X, Wong S, Kajigaya S, Zhao K, Mao Q, Young NS. Hybrid DNA virus in Chinese patients with seronegative hepatitis discovered by deep sequencing. Proc Natl Acad Sci U S A. 2013; 110: 10264-9.

zur Hausen H. Red meat consumption and cancer: Reasons to suspect involvement of bovine infectious factors in colorectal cancer. Int J Cancer 2012; 130:2475-2483.

zur Hausen H., de Villiers, E.-M., Prenatal Infections with Subsequent Immune Tolerance could explain the Epidemiology of Common Childhood Cancers, World Cancer Report, IARC, Lyon, 2014, pp. 261-265.

Abdel-Hag NM, Asmar BI. Human herpes virus 6 (HHV6) infection. Indian J Pediatr. 2004; 71: 89-96.

Agranoff BW, Goldberg D. Diet and the geographical distribution of multiple sclerosis. Lancet II 1974; 1061.

Alcalde-Cabero E, Almazán-Isla J, García-Merino A, de Sá J, de Pedro-Cuesta J. Incidence of multiple sclerosis among European Economic Area populations, 1985-2009: the framework for monitoring. BMC Neurol. 2013; 13: 58-80.

Alenda R, Alvarez-Lafuente R, Costa-Frossard L, Arroyo R, Mirete S, Alvarez-Cermeño JC, Villar LM. Identification of the major HHV-6 antigen recognized by cerebrospinal fluid IgG in multiple sclerosis. Eur J Neurol. 2014 Apr 12. doi: 10.1111/ene.12435.

Almohmeed YH, Avenell A, Aucott L, Vickers MA. Systematic review and meta-analysis of the sero-epidemiological association between Epstein Barr virus and multiple sclerosis. PLoS One. 2013; 8: e61110.

Angelini DF, Serafini B, Piras E, Severa M, Coccia EM, Rosicarelli B, Ruggieri S, Gasperini C, Buttari F, Centonze D, Mechelli R, Salvetti M, Borsellino G, Aloisi F, Battistini L. Increased CD8+ T cell response to Epstein-Barr virus lytic antigens in the active phase of multiple sclerosis. PLoS Pathog. 2013; 9: e1003220.

Ascherio A, Munger KL, Lünemann JD. The initiation and prevention of multiple sclerosis. Nat Rev Neurol. 2012; 8: 602-12.

Ascherio A. Environmental factors in multiple sclerosis. Expert Rev. Neurother. 2013; 13 (12s) 3-9.

Ashtari F, Jamshidi F, Shoormasti RS, Pourpak Z, Akbari M. Cow's milk allergy in multiple sclerosis patients. J Res Med Sci. 2013; 18 (Suppl 1): S62-5.

Bager P, Nielsen NM, Bihrmann K, Frisch M, Wohlfart J, Koch-Henriksen N, Melbye M, Westergaard T Sibship characteristics and risk of multiple sclerosis: a nationwide cohort study in Denmark. Am J Epidemiol. 2006; 163: 1112-7.

Bantel-Schaal U, zur Hausen H. Adeno-associated viruses inhibit SV40 DNA amplification and replication of herpes simplex virus in SV40-transformed hamster cells. Virology 1988; 164: 64-74.

Banwell B, Bar-Or A, Cheung R, Kennedy J, Krupp LB, Becker DJ, Dosch HM; Wadsworth Pediatric Multiple Sclerosis Study Group. Abnormal T-cell reactivities in childhood inflammatory demyelinating disease and type 1 diabetes. Ann Neurol. 2008; 63: 98-111.

Baranzini SE, Mudge J, van Velkinburgh JC, Khankhanian P, Khrebtukova I, Miller NA, Zhang L, Farmer AD, Bell CJ, Kim RW, May GD, Woodward JE, Caillier SJ, McElroy JP, Gomez R, Pando MJ, Clendenen LE, Ganusova EE, Schilkey FD, Ramaraj T, Khan OA, Huntley JJ, Luo S, Kwok PY, Wu TD, Schroth GP, Oksenberg JR, Hauser SL, Kingsmore SF. Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. Nature. 2010; 464: 1351-6.

Bauer G, Höfler P, zur Hausen H. Epstein-Barr virus induction by a serum factor. I. Induction and cooperation with additional inducers. Virology. 1982; 121: 184-94.

Bauer G, Götschl M, Höfler P. Tumor-promoting activity of Epstein-Barr-virus-inducing factor transforming growth factor type beta (EIF/TGF-beta) is due to the induction of irreversible transformation. Int J Cancer. 1991; 47: 881-8.

Ben Fredj N, Rotola A, Nefzi F, Chebel S, Rizzo R, Caselli E, Frih-Ayed M, Di Luca D, Aouni M. Identification of human herpes viruses 1 to 8 in Tunisian multiple sclerosis patients and healthy blood donors. J Neurovirol. 2012; 18 :12-19.

5

Beretich BD, Beretich TM. Explaining multiple sclerosis by ultraviolet exposure: a geospatial analysis. Mult Scler 2009; 15: 891-8.

10 Birkeland SA, Storm HH, Lamm LU, Barlow L, Blohmé I, Forsberg B, Eklund B, Fjeldborg O, Friedberg M, Frödin L, et al. Cancer risk after renal transplantation in the Nordic countries, 1964-1986. Int J Cancer. 1995 Jan 17;60(2):183-9.

15 Borkosky SS, Whitley C, Kopp-Schneider A, zur Hausen H, de Villiers EM. Epstein-Barr virus stimulates torque teno virus replication: a possible relationship to multiple sclerosis. PLoS One. 2012; 7: e32160.

20 Brecht I, Weissbrich B, Braun J, Toyka KV, Weishaupt A, Buttmann M. Intrathecal, polyspecific antiviral immune response in oligoclonal band negative multiple sclerosis. PLoS One. 2012; 7: e40431.

25 Buchanan R, Bonthius DJ. Measles virus and associated central nervous system sequelae. Semin Pediatr Neurol. 2012; 19: 107-14.

Butcher PJ, The distribution of multiple sclerosis in relation to the dairy industry and milk consumption. N Z Med J. 1976; 83: 427-30.

30

Butcher PJ. Milk consumption and multiple sclerosis—an etiological hypothesis. Med Hypotheses. 1986; 19: 169-78.

Caserta MT, Mock DJ, Dewhurst S. Human herpes virus 6. Clin Infect Dis. 2001; 33: 829-33. Review.

Christensen JC. Multiple sclerosis: some epidemiological clues to etiology.

Acta Neurol Latinoam. 1975; 21: 66-85. Review.

Christensen T. Human herpes viruses in MS. Int MS J. 2007; 14: 41-7. Review.

Conradi S, Malzahn U, Paul F, Quill S, Harms L, Then Bergh F, Ditzenbach A, Georgi T, Heuschmann P, Rosche B. Breastfeeding is associated with lower risk for multiple sclerosis. Mult Scler. 2013; 19: 553-8.

Cusick MF, Libbey JE, Fujinami RS. Multiple sclerosis: autoimmunity and viruses: Curr. Opin. Rheumatol. 2013; 25: 496-501.

Décard BF, von Ahsen N, Grunwald T, Streit F, Stroet A, Niggemeier P, Schottstedt V, Riggert J, Gold R, Chan A. Low vitamin D and elevated immunoreactivity against Epstein-Barr virus before first clinical manifestation of multiple sclerosis. J Neurool Neurosurg Psychiatry. 2012; 83: 1170-3.

Dewhurst S. Human herpes virus type 6 and human herpes virus type 7 infections of the central nervous system. Herpes. 2004; 11 Suppl 2: 105A-111A. Review.

Disanto G, Morahan JM, Barnett MH, Giovannoni G, Ramagopalan SV. The evidence for a role of B cells in multiple sclerosis. Neurology 2012; 78: 823-32.

Engels EA, Biggar RJ, Hall HI, Cross H, Crutchfield A, Finch JL, Grigg R, Hylton T, Pawlish KS, McNeel TS, Goedert JJ. Cancer risk in people infected with human immunodeficiency virus in the United States. Int J Cancer. 2008; 123: 187-94.

5

Ferrò MT, Franciotta D, Prella A, Bestetti A, Cinque P. Active intrathecal herpes simplex virus type 1 (HSV-1) and human herpes virus-6 (HHV-6) infection at onset of multiple sclerosis. J Neurovirol. 2012; 18: 437-4

10

Fotheringham J, Jacobson S. Human herpes virus 6 and multiple sclerosis: potential mechanisms for virus-induced disease. Herpes 2005; 12: 4-9.

15

Fraussen J, Vrolix K, Martinez-Martinez P, Losen M, De Baets MH, Stinissen P, Somers V. B cell characterization and reactivity analysis in multiple sclerosis. Autoimmun Rev. 2009; 8: 654-8.

20

Frolik CA, Dart LL, Meyers CA, Smith DM, Sporn MB. Purification and initial characterization of a type beta transforming growth factor from human placenta. Proc Natl Acad Sci U S A. 1983; 80: 3676-80.

25

Gaitán MI, Sati P, Inati SJ, Reich DS. Initial investigation of the blood-barrier in MS lesions at 7 tesla. Mult Scler. 2013; 19: 1068-73.

30

Geeraedts F, Wilczak N, van Binnendijk R, De Keyser J. Search for morbillivirus proteins in multiple sclerosis brain tissue. Neuroreport. 2004; 19: 15: 27-32.

Grant WB. Epidemiology of disease risks in relation to vitamin D insufficiency. Prog Biophys Mol Biol. 2006; 92: 65-79.

- Grinde B. Herpes viruses: latency and reactivation - viral strategies and host response. J Oral Microbiol. 2013; doi: 10.3402/jom.v5i10.22766.
- 5 Guggenmos J, Schubart AS, Ogg S, Andersson M, Olsson T, Mather IH, Linington C. Antibody cross-reactivity between myelin oligodendrocyte glycoprotein and the milk protein butyrophilin in multiple sclerosis. J. Immunol. 2004; 172: 661-8.
- 10 Hawkes CH, Giovannoni G, Keir G, Cunningham M, Thompson EJ. Seroprevalence of herpes simplex virus type 2 in multiple sclerosis. Acta Neurol Scand. 2006; 114: 363-7.
- 15 Heilbronn R, Bürkle A, Stephan S, zur Hausen H. The adeno-associated virus rep gene suppresses herpes simplex virus-induced DNA amplification. J Virol. 1990a; 64: 3012-8.
- 20 Heilbronn R, Weller SK, zur Hausen H. Herpes simplex virus type 1 mutants for the origin-binding protein induce DNA amplification in the absence of viral replication. Virology. 1990b; 179: 478-81.
- 25 Heilbronn R, Albrecht I, Stephan S, Bürkle A, zur Hausen H. Human cytomegalovirus induces JC virus DNA replication in human fibroblasts. Proc Natl Acad Sci U S A. 1993; 90: 11406-10.
- 30 Henle W, Henle G. Evidence for an etiologic relation of the Epstein-Barr virus to human malignancies. Laryngoscope. 1977; 87: 467-73.
- Hollis BW, Roos BA, Draper HH, Lambert PW. Vitamin D and its metabolites in human and bovine milk. J Nutr. 1981; 111: 1240-8.

Holzmann C, Bauer I, Meyer P. Co-occurrence of multiple sclerosis and cancer in a BRCA1 positive family. Eur J Med Genet. 2013; 56: 577-9.

5 Hsiao FC, Tai AK, Deglon A, Sutkowski N, Longnecker R, Huber BT. EBV LMP-2A employs a novel mechanism to transactivate the HERV-K18 superantigen through its ITAM. Virology. 2009; 385: 261-6.

Hu Knox KK, Brewer JH, Henry JM, Harrington DJ, Carrigan DR.
10 Human herpes virus 6 and multiple sclerosis: systemic active infections in patients with early disease. Clin Infect Dis. 2000; 31: 894-903.

Hyppönen E. Vitamin D and increasing incidence of type 1 diabetes-evidence for an association? Diabetes Obes Metab. 2010; 12: 737-43.
15

Isik S, Ozuguz U, Tutuncu YA, Erden G, Berker D, Acar K, Aydin Y, Akbaba G, Helvaci N, Guler S. Serum transforming growth
20 factor-beta levels in patients with vitamin D deficiency. Eur J Intern Med. 2012; 23: 93-7.

Ito I, Waku T, Aoki M, Abe R, Nagai Y, Watanabe T, Nakajima Y, Ohkido I, Yokoyama K, Miyachi H, Shimizu T, Murayama A,
25 Kishimoto H, Nagasawa K, Yanagisawa J. A nonclassical vitamin D receptor pathway suppresses renal fibrosis. Clin Invest. 2013; 123: 4579-94.

James E, Dobson R, Kuhle J, Baker D, Giovannoni G, Ramagopalan SV. The effect of vitamin D-related interventions on multiple sclerosis relapses: a meta-analysis. Mult Scler. 2013; 19: 1571-9.
30

Kakalacheva K, Münz C, Lünemann JD. Viral triggers of multiple sclerosis. Biochim Biophys Acta. 2011; 1812: 132-40.
35

Karner W, Bauer G. Activation of a varicella-zoster virus-specific IgA response during acute Epstein-Barr virus infection. J Med Virol. 1994; 44: 258-62.

5

Koch-Henriksen N, Sorensen PS. Why does the north-south gradient of incidence of multiple sclerosis seem to have disappeared on the northern hemisphere? J Neurol Sci. 2011; 311: 58-63.

- 10 Koch-Henriksen N, Stenager E, Laursen B. The use of epidemiological multiple sclerosis registers in research: the Danish MS Registry. Acta Neurol Scand Suppl. 2012; 195: 7-12.

- 15 Kotzamani D, Panou T, Mastorodemos V, Tzagournissakis M, Nikolakaki H, Spanaki C, Plaitakis A. Rising incidence of multiple sclerosis in females associated with urbanization. Neurology. 2012; 78: 1728-35.

- 20 Kurtzke JF. Multiple sclerosis in time and space-geographic clues to cause. J Neurovirol. 2000; 6 Suppl 2: S134-40.

Kurtzke JF. Epidemiology in multiple sclerosis: a pilgrim's progress. Brain. 2013; 136: 2904-17.

- 25 Kuusisto H, Hyöty H, Kares S, Kinnunen E, Elovaara I. Human herpes virus 6 and multiple sclerosis: a Finnish twin study. Mult Scler. 2008; 14: 54-8.

- 30 Latif N, Rana F, Guthrie T. Breast cancer and HIV in the era of highly active antiretroviral therapy: two case reports and review of the literature. Breast J. 2011; 17: 87-92.

Lebrun C, Debouverie M, Vermersch P, Clavelou P, Rumbach L, de Seze J, Wiertlevski S, Defer G, Gout O, Berthier F, Danzon A.

WO 2016/005054 PCT/EP2015/001399
Cancer risk and impact of disease-modifying treatments in
patients with multiple sclerosis. Mult Scler. 2008; 14: 399-405.

Lemire JM, Adams JS, Sakai R, Jordan SC. 1 alpha,25-
5 dihydroxyvitamin D3 suppresses proliferation and immunoglobulin
production by normal human peripheral blood mononuclear cells. J
Clin Invest 1984; 74: 657-61.

Libbey JE, Cusick MF, Fujinami RS. Role of pathogens in multiple
10 sclerosis. Intern Rev Immunol. Early Online 1-18, 2013; DOI:
10.3/08830185.2013.823422

Liu H, Fu Y, Li B, Yu X, Xie J, Cheng J, Ghabrial SA, Li G, Yi
X, Jiang D. Widespread horizontal gene transfer from circular
15 single-stranded DNA viruses to eukaryotic genomes. BMC Evol
Biol. 2011 Sep 26;11:276. doi: 10.1186/1471-2148-11-276.

Longkumer T, Kamireddy S, MuthyalaVR, Akbarpasha S, Pitchika GK,
Kodetham G, Ayaluru M, Siddavattam D. Acinetobacter phage genome
20 is similar to Sphinx 2.36, the circular DNA copurified with TSE
infected particles. Sci Rep 2013; 3: 2240.doi:
10.1038/srep02240.

Lucchinetti C, Brück W, Parisi J, Scheithauer B, Rodriguez M,
25 Lassmann H.Heterogeneity of multiple sclerosis lesions:
implications for the pathogenesis of demyelination. Ann Neurol.
2000; 47: 707-17.

Magliozzi R, Serafini B, Rosicarelli B, Chiapetta G, Veroni C.,
30 Reynolds R, Aloisi F. B-cell enrichment and Epstein-Barr virus
infection in inflammatory cortical lesions in secondary
progressive multiple sclerosis. J Neuropathol Exp Neurol 2013;
72: 29-41.

- Malosse D, Perron H. Correlation analysis between bovine populations, other farm animals, house pets, and multiple sclerosis prevalence. Neuroepidemiology. 1993; 12: 15-27.
- 5 Marrie RA, Wolfson C. Multiple sclerosis and varicella zoster virus infection: a review. Epidemiol Infect. 2001; 127: 315-25. Review.
- 10 Matz B, Schlehofer JR, zur Hausen H. Identification of a gene function of herpes simplex virus type 1 essential for amplification of simian virus 40 DNA sequences in transformed hamster cells. Virology. 1984; 134: 328-37.
- 15 Mesliniene S, Ramrattan L, Goldings S, Sheikh-Ali M. Role of vitamin D in the onset, progression and severity of multiple sclerosis. Endocr Pract. 2013; 19: 129-36.
- 20 Metz I, Weigand SD, Popescu BF, Frischer JM, Parisi JE, Guo Y, Lassmann H, Brück W, Lucchinetti CF. Pathologic heterogeneity persists in early active multiple sclerosis lesions. Ann Neurol. 2014 Apr 26. doi: 10.1002/ana.24163.
- 25 Midgard R, Glattre E, Grønning M, Riise T, Edland A, Nyland H. Multiple sclerosis and cancer in Norway. A retrospective cohort study. Acta Neurol Scand. 1996; 93: 411-5.
- 30 Mirandola P, Stefan A, Brambilla E, Campadelli-Fiume G, Grimaldi LM. Absence of human herpes virus 6 and 7 from spinal fluid and serum of multiple sclerosis patients. Neurology. 1999; 53: 1367-8.
- 35 Müller K, Heilmann C, Poulsen LK, Barington T, Bendtzen K. The role of monocytes and T cells in 1,25-dihydroxyvitamin D3 mediated inhibition of B cell function in vitro. Immunopharmacology 1991; 21: 121-8.

Munger KL, Chitnis, T, Frazier AL, Giovannucci E, Spiegelman D, Ascherio A. Dietary intake of vitamin D during adolescence and risk of multiple sclerosis. J Neurol. 2011a; 258: 479-85.

5

Munger KL Levin LI, O'Reilly EJ, Falk KI, Aschewrio A. Anti-Epstein-Barr virus antibodies as serological markers of multiple sclerosis: a prospective study among United States military personnel. Mult Scler 2011b; 17: 1185-93.

10

Murray TJ. An unusual occurrence of multiple sclerosis in a small rural community. Can J Neurol Sci. 1976; 3: 163-6.

15

Nicoll MP, ProençaJT, Efstathiou S. The molecular basis of herpes simplex virus latency. FEMS Microbiol Rev. 2012; 36: 684-705.

20

Nielsen NM, Rostgaard K, Rasmussen S, Koch-Henriksen N, Storm HH, Melbye M, Hjalgrim H. Cancer risk among patients with multiple sclerosis: a population-based register study. Int J Cancer. 2006; 118: 979-84.

25

Nora-Krukle Z, Chapenko S, Logina I, Millers A, Platkajis A, Murovska M. Human herpes virus 6 and 7 reactivation and disease activity in multiple sclerosis. Medicina (Kaunas). 2011; 47: 527-31.

30

Nordal HJ, Vandvik B, Norrby E. Multiple sclerosis: local synthesis of electrophoretically restricted measles, rubella, mumps and herpes simplex virus antibodies in the central nervous system. Scand J Immunol. 1978; 7: 473-9.

35

O'Gorman C, Lin R, Stankovich J, Broadley SA, Modelling genetic susceptibility to multiple sclerosis with family data. Neuroepidemiology. 2013; 40: 1-12.

Ohara Y. Multiple sclerosis and measles virus. Jpn J Infect Dis. 1999; 52: 198-200. Review.

- 5 Olival GS, Lima BM, Sumita LM, Serafim V, Fink MC, Nali LH, Romano CM, Thomaz RB, Cavenaghi VB, Tilbery CP, Penalva-de-Oliveira AC. Multiple sclerosis and herpes virus interaction. Arq Neuropsiquiatr. 2013; 71: 727-30.
- 10 Ongrádi J, Rajda C, Maródi CL, Csiszár A, Vecsei L. A pilot study on the antibodies to HHV-6 variants and HHV-7 in CSF of MS patients. J Neurovirol. 1999; 5: 529-32.
- 15 Opsahl ML, Kennedy PG Investigating the presence of human herpes virus 7 and 8 in multiple sclerosis and normal control brain tissue. J Neurol Sci. 2006; 240: 37-44.
- 20 Ordoñez G, Pineda B, Garcia-Navarrete R, Sotelo J. Brief presence of varicella-zoster viral DNA in mononuclear cells during relapses of multiple sclerosis. Arch Neurol. 2004; 61: 529-32.
- 25 Owens GP, Gilden D, Burgoon MP, Yu X, Bennett JL. Viruses and multiple sclerosis. Neuroscientist. 2011; 17: 659-76. Review.
- Pakpoor J, Pakpoor J, Disanto G, Giovannoni G, Ramagopalan SV. Cytomegalovirus and multiple sclerosis risk. J Neurol. 2013; 260: 1658-60.
- 30 Petersen T, Møller-Larsen A, Ellermann-Eriksen S, Thiel S, Christensen T. Effects of interferon-beta therapy on elements in the antiviral immune response towards the human herpes viruses EBV, HSV, and VZV, and to the human endogenous retroviruses HERV-H and HERV-W in multiple sclerosis. J Neuroimmunol. 2012; 35 249: 105-8.

Pierrot-Deseilligny C, Souberbielle JC. Contribution of vitamin D insufficiency to the pathogenesis of multiple sclerosis. *Ther Adv Neurol Disord*. 2013; 6: 81-116.

5

Pisacane A, Impagliazzo N, Russo M, Valiani R, Mandarini A, Florio C, Vivo P. Breast feeding and multiple sclerosis. *BMJ*. 1994; 308: 1411-2.

- 10 Pohl D, Rostasy K, Jacobi C, Lange P, Nau R, Krone B, Hanefeld F. Intrathecal antibody production against Epstein-Barr and other neurotropic viruses in pediatric and adult onset multiple sclerosis. *J Neurol*. 2010; 257: 212-6.

- 15 Provvedini DM, Tsoukas CD, Deftos LJ, Manolagas SC. 1 alpha,25-dihydroxyvitamin D3-binding macromolecules in human B lymphocytes: effects on immunoglobulin production. *J Immunol* 1986; 136: 2734-40.

- 20 Rall GF. Measles virus 1998-2002: progress and controversy. *Annu Rev Microbiol*. 2003; 57: 343-67. Review.

Rima BK, Duprex WP. Molecular mechanisms of measles virus persistence. *Virus Res*. 2005; 111: 132-47. Review.

25

Rosecrans R, Dohnal JC. Seasonal vitamin D changes and the impact on health risk assessment. *Clin Biochem*, 2014; pii: S0009-9120. doi: 10.1016/j.clinbiochem.2014.02.004.

- 30 Ross RT. The varicella-zoster virus and multiple sclerosis. *J Clin Epidemiol*. 1998; 51: 533-5. Review.

- Ruprecht K, Obojes K, Wengel V, Gronen F, Kim KS, Perron H, Schneider-Schaulies J, Rieckmann P. Regulation of human endogenous retrovirus W protein expression by herpes simplex
- 35

WO 2016/005054 PCT/EP2015/001399
virus type 1: implications for multiple sclerosis. J Neurovirol.
2006; 12: 65-71.

Salzer J, Nyström M, Hallmans G, Steenlund H, Wadell G,
5 Sundström P. Epstein-Barr virus antibodies biobank samples. Mult
Scler. 2013; 19: 1587-91.

Sanders V, Felisan S, Waddell A, Tourtellotte W. Detection of
herpesviridae in postmortem multiple sclerosis brain tissue and
10 controls by polymerase chain reaction. J Neurovirol. 1996; 2:
249-58.

Schlehofer JR, Gissmann L, Matz B, zur Hausen H. Herpes simplex
virus-induced amplification of SV40 sequences in transformed
15 Chinese hamster embryo cells. Int J Cancer. 1983; 32: 99-103.

Schlehofer JR, zur Hausen H. Adenovirus infection induces
amplification of persistent viral DNA sequences (simian virus
40, hepatitis B virus, bovine papillomavirus) in human and
20 rodent cells. Virus Res. 1990;17: 53-60.

Schmitt J, Schlehofer JR, Mergener K, Gissmann L, zur Hausen H.
Amplification of bovine papillomavirus DNA by N-methyl-N'-nitro-
N-nitrosoguanidine, ultraviolet irradiation, or infection with
25 herpes simplex virus. Virology 1989; 172:73-81.

Scott FW. Cow milk and insulin-dependent diabetes mellitus: is
there a relationship? Am J Clin Nutr. 1990; 51:489-491.

30 Sepcic J, Mesaros E, Materljan E, Sepic-Grahovac D. Nutritional
factors and multiple sclerosis in Gorski Kotar, Croatia.
Neuroepidemiology. 1993; 12: 234-40.

Simpson S Jr, Taylor B, Dwyer DE, Taylor J, Blizzard L, Ponsonby
35 AL, Pittas F, Dwyer T, van der Mei I. Anti-HHV-6 IgG titer

WO 2016/005054 PCT/EP2015/001399
significantly predicts subsequent relapse risk in multiple
sclerosis. Mult Scler. 2012; 18: 799-806.

Sotelo J, Corona T. Varicella zoster virus and relapsing
5 remitting multiple sclerosis. Mult Scler Int. 2011; 2011:
214763.

Sun LM, Lin CL, Chung CJ, Liang JA, Sung FC, Kao CH.Increased
breast cancer risk for patients with multiple sclerosis: a
10 nationwide population-based cohort study. Eur J Neurol. 2013;
Sep 19. doi: 10.1111/ene.12267.

Sundqvist E, Bergström T, Daialhosein H, Nyström M, Sundström P,
Hillert J, Alfredsson L, Kockum I, Olsson T. Cytomegalovirus
15 seropositivity is negatively associated with multiple sclerosis.
Mult Scler. 2013 Sep 2. [Epub ahead of print].

Sundström P, Juto P, Wadell G, Hallmans G, Svenningsson A,
Nyström L, Dillner J, Forsgren L. An altered immune response to
20 Epstein-Barr virus in multiple sclerosis: a prospective study.
Neurology. 2004; 62: 2277-82.

Sutkowski N, Chen G, Calderon G, Huber BT. Epstein-Barr virus
latent membrane protein LMP-2A is sufficient for transactivation
25 of the human endogenous retrovirus HERV-K18 superantigen. J
Virol. 2004; 78: 7852-60.

Svejgaard A. The immunogenetics of multiple sclerosis.
Immunogenetics. 2008; 60: 275-86.

30 Tai AK, Luka J, Ablashi D, Huber BT. HHV-6A infection induces
expression of HERV-K18-encoded superantigen. J Clin Virol. 2009;
46: 47-8.

Tan IL, van Schijndel RA, Pouwell PJ, van Walderveen MA, Reichenbach JR, Manoliu RA, Barkhof F. MR venography of multiple sclerosis. *AJNR Am Neuroradiol*. 2000; 21: 1029-42.

- 5 Tarrats R, Ordoñez G, Rios C, Sotelo J. Varicella, ephemeral breastfeeding and eczema as risk factors for multiple sclerosis in Mexicans. Acta Neurol Scand. 2002; 105: 88-94..

- 10 Taus C, Pucci E, Cartechini E, Fié A, Giuliani G, Clementi M, Menzo S. Absence of HHV-6 and HHV-7 in cerebrospinal fluid in relapsing-remitting multiple sclerosis. *Acta Neurol Scand*. 2000; 101: 224-8.

- 15 Turcanova VL, Bundgaard B, Höllsberg P. Human herpes virus-6B induces expression of the human endogenous retrovirus K18-encoded superantigen. *J Clin Virol*. 2009; 46: 15-9.

- Virtanen JO, Jacobson S. Viruses and multiple sclerosis. *CNS Neurol Disord Drug Targets*. 2012; 11: 528-44. Review.
- 20 Warren TR. The increased prevalence of multiple sclerosis among people who were born and bred in areas where goitre is endemic. Med Hypotheses. 1984; 14: 111-4.

- 25 Waubant E, Mowry EM, Krupp L, Chitnis T, Yeh EA, Kuntz N, Ness J, Belman A, Milazzo M, Gorman M, Weinstock-Guttman B, Rodriguez M, James JA. Antibody response to common viruses and human leukocyte antigen-DRB1 in pediatric multiple sclerosis. *Mult Scler*. 2013; 19: 891-5.

- 30 Wikström J. Studies on the clustering of multiple sclerosis in Finland. *Riv Patol Nerv Ment*. 1976; 97: 199-204.

- 35 Winer S, Astsaturon I, Cheung RK, Schrade K, Gunaratnam L, Wood DD, Moscarello MA, O'Connor P, McKerlie C, Becker DJ, Dosch HM.

T cells of multiple sclerosis patients target a common environmental peptide that causes encephalitis in mice. J Immunol. 2001; 166: 4751-6.

5 Wu H, Li T, Zeng M, Peng T. Herpes simplex virus type 1 infection activates the Epstein-Barr virus replicative cycle via a CREB-dependent mechanism. Cell Microbiol. 2012 ; 14: 546-59.

Yea C, Tellier R, Chong P, Westmacott G, Marrie RA, Bar-Or A, Banwell B; Canadian Pediatric Demyelinating Disease Network. Epstein-Barr virus in oral shedding of children with multiple sclerosis. Neurology. 2013; 81: 1392-9.

Zerr P, Vollath S, Palumbo-Zerr K, Tomcik M, Huang J, Distler A, Beyer C, Dees C, Gela K, Distler O, Schett G, Distler JH. Vitamin D receptor regulates TGF- β signalling in systemic sclerosis. Ann Rheum Dis. 2014 Jan 21. doi: 10.1136/annrheumdis-2013-204378.

zur Hausen H. Genital papillomavirus infections. Prog Med Virol. 1985; 32: 15-21.

zur Hausen, H. and de Villiers, E.M., Dairy cattle serum and milk factors contributing to the risk of colon and breast cancers, Int. J. Cancer, 2015, Feb. 3 doi: 10.1002/ijc.29466

25 Zwart SR, Mehta SK, Ploutz-Snyder R, Bourbeau Y, Locke JP, Pierson DL, Smith SM. Response to vitamin D supplementation during Antarctic winter is related to BMI, and supplementation can mitigate Epstein-Barr virus reactivation. J Nutr. 2011; 141: 30 692-7.

References (Examples 2-5)

Buck CB, Pastrana DV, Lowy DR, Schillier JT. 2005. Generation of HPV pseudovirions using transfection and their use in neutralization assays. *Methods Mol Med* 119: 445-462.

- 5 Caspar Y, Recule C, Pouzol P, Lafeuillade B, Mallaret M, Mairin M, Croize J. *Psychrobacter arenosus* bacteremia after blood transfusion, France. *Emerging Infectious Diseases* 2013; 19: 1118-1120.
- 10 Coton M, Delbes-Paus C, Irlinger F, Desmasures N, Le Fleche A, Stahl V, Montel MC, Coton E. Diversity and assessment of potential risk factors of Gram-negative isolates associated with French cheeses. *Food Microbiol.* 2012; 29:88-98.
- 15 De Filippis F, La Storia A, Villiani F, Ercolini D. Exploring the sources of bacterial spoilers in beefsteaks by culture-independent high-throughput sequencing. *PLoS One* 2013; 8:e70222.
- del Solar G, Giraldo R, Ruiz-Echevarria MJ, Espinosa M, Diaz-
20 Orejas R. Replication and control of circular bacterial plasmids. *Microbiol Mol. Biol. Rev.* 1998; 62:434-464.
- de Villiers EM, Borkosky SS, Kimmel R, Gunst K, Fei JW. The diversity of torque teno viruses: in vitro replication leads to
25 the formation of additional replication-competent subviral molecules. *J Virol.* 2011; 85: 7284-95.
- de Villiers, EM, zur Hausen, H. Concept for the pathogenesis of multiple sclerosis (I): interaction of an amplifying virus and a
30 helper-dependent bovine milk factor. (submitted)
- Funk M, Gunst K, Lucansky V, Müller H, zur Hausen H, de Villiers E-M., Isolation of protein-associated circular DNA from healthy cattle serum, *Genome Announcements*, 2(4): e00846-14, 2014

Dziewit L, Cegielski A, Romaniuk K, Uhrynowski W, Szych A, Niesiobedzki P, Zmuda-Baranowska MJ, Zdanowski MK, Bartosik D. Plasmid diversity in arctic strains of Psychrobacter spp. Extremophiles. 2013; 17: 433-44

- 5 Ebringer A, Rashid T, Wilson C. Bovine spongiform encephalopathy, multiple sclerosis, and creutzfeldt-jakob disease are probably autoimmune diseases evoked by Acinetobacter bacteria. Ann N Y Acad Sci. 2005; 1050: 417-28.
- 10 Lloyd-Puryear M, Wallace D, Baldwin T, Hollis DG. Meningitis caused by Psychrobacter immobilis in an infant. J Clin. Microbiol. 1991; 29: 2041-2042.
- Liu H, Fu Y, Li B, Yu X, Xie J, Cheng J, Ghabrial SA, Li G, Yi
15 X, Jiang D. Widespread horizontal gene transfer from circular single-stranded DNA viruses to eukaryotic genomes. BMC Evol Biol. 2011; 11: 276. doi: 10.1186/1471-2148-11-276.
- Longkumer T, Kamireddy S, MuthyalaVR, Akbarpasha S, Pitchika GK,
20 Kodetham G, Ayaluru M, Siddavattam D. Acinetobacter phage genome is similar to Sphinx 2.36, the circular DNA copurified with TSE infected particles. Sci Rep 2013; 3: 2240.doi: 10.1038/srep02240.
- 25 Martin DP, Biagini P, Lefeuvre P, Golden M, Roumagnac P, Varsani A. Recombination in eukaryotic single stranded DNA viruses. Viruses. 2011; 3: 1699-738.
- Rosario K, Duffy S, Breitbart M. 2012a. A field guide to
30 eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. Arch. Virol. 157: 1851-1871.
- Rosario K, Dayaram A, Marinov M, Ware J, Kraberger S, Stainton D, Breitbart M, Varsani A. 2012b. Diverse circular ssDNA viruses

WO 2016/005054 PCT/EP2015/001399
discovered in dragonflies (Odonata: Epiprocta). J Gen. Virol.
93:2668-2681.

Sikorski A, Massaro M, Kraberger S, Young LM, Smalley D, Martin
5 DP, Varsani A. 2013. Novel myco-like DNA viruses discovered in
the faecal matter of various animals. Virus Res. 177:209-216.

zur Hausen H, de Villiers EM. Prenatal Infections with
Subsequent Immune Tolerance Could Explain the Epidemiology of
10 Common Childhood Cancers. World Cancer Report 2014, IARC Lyon,
pp 261-265.

de Villiers E-M, zur Hausen H. Concept for the pathogenesis of
15 multiple sclerosis (I): Interaction of an amplifying virus and a
helper-dependent bovine milk factor. (submitted).

del Val C, Ernst P, Falkenhahn M, Fladerer C, Glatting KH, Suhai
S, Hotz-Wagenblatt A. 2007. ProtSweep, 2DSweep and DomainSweep:
20 protein analysis suite at DKFZ. Nucleic Acids Res. 35 (Web
Server issue):W444-450.

Chattoraj DK. 2000. Control of plasmid DNA replication by
iterons: no longer paradoxical. Mol. Microbiol. 37: 467-476.

25 zur Hausen H. 2001. Proliferation-inducing viruses in non-
permissive systems as possible causes of human cancers. Lancet
357: 381-384.

30 Whitley C, Gunst K, Müller H, Funk M, zur Hausen H, de Villiers
E-M., Novel replication-competent circular DNA molecules from
healthy cattle serum, milk and multiple sclerosis-affected
human brain tissue, Genome Announcements 2(4): e00849-14, 2014.

WO 2016/005054 **PCT/EP2015/001399**
Lamberto I, Gunst K, Müller H, zur Hausen H, de Villiers E-M.
Mycovirus-like DNA virus sequences from cattle serum, human
brain and serum from multiple sclerosis patients, Genome
Announcements 2(4): e00848-14, 2014

5

Gunst K, zur Hausen H, de Villiers E-M., Isolation of bacterial
plasmid-related replication-associated circular DNA from serum
sample of a multiple sclerosis patient, Genome announcement
2(4), e00847-14, 2014

10

Claims

1. An HCBI, MSBI, MSSI or CMI polynucleic acid comprising:

5 (a) a nucleotide sequence depicted in any one of Figures 1 to 4;

(b) a nucleotide sequence having at least 90% identity to a nucleotide sequence of (a);

(c) a fragment of a nucleotide sequence of (a) or (b);

10 (d) a nucleotide sequence being complementary to a nucleotide sequence of (a), (b) or (c); or

(e) a nucleotide sequence which is redundant as a result of the degeneracy of the genetic code compared to any of the above-given nucleotide sequences.

15 2. An oligonucleotide primer comprising part of an HCBI, MSBI, MSSI or CMI polynucleic acid of claim 1, said primer being capable of acting as primer for specifically sequencing or specifically amplifying the nucleic acid of a certain HCBI, MSBI, MSSI or CMI isolate containing a nucleotide sequence in claim 1.

25 3. An oligonucleotide probe comprising part of an HCBI, MSBI, MSSI or CMI polynucleic acid of claim 1, said probe being capable of acting as a hybridization probe for specific detection of the nucleic acid of a certain HCBI, MSBI, MSSI or CMI isolate containing a nucleotide sequence of claim 1.

30 4. An expression vector comprising an HCBI, MSBI, MSSI or CMI polynucleic acid of any one of claims 1 to 3 operably linked to prokaryotic, eukaryotic or viral transcription and translation control elements.

35 5. A host cell transformed or modified with an expression vector according to claim 4.

6. A polypeptide being encoded by an HCBI, MSBI, MSSSI or CMI polynucleic acid of claim 1.

5 7. An antibody or antigen binding fragment thereof specifically binding to a polypeptide of claim 6.

8. Use of a primer according to claim 2, a probe according to claim 3, a polypeptide of claim 6, or an antibody or fragment
10 thereof according to claim 7 for the preparation of a diagnostic composition for the diagnosis of a predisposition or an early stage of cancer, a disease of the CNS or diabetes.

9. Use according to claim 8, wherein the cancer is breast
15 cancer, ovarian cancer, lung cancer, prostate cancer, colorectal cancer or colon cancer, and the disease of the CNS is Multiple sclerosis MS, amyotrophic lateral sclerosis, transmissible spongiforme encephalopathies/Prion-linked diseases, Parkinson's disease or Alzheimer disease.

20

10. A method for the detection of an HCBI, MSBI, MSSSI or CMI polynucleic acid according to claim 1 in a biological sample, comprising: (a) optionally extracting sample polynucleic acid, (b) amplifying the polynucleic acid as described above with at
25 least one primer according to claim 2, optionally a labelled primer, and (c) detecting the amplified polynucleic acid.

11. A method for the detection of an HCBI, MSBI, MSSSI or CMI polynucleic acid according to claim 1 in a biological sample, comprising: (a) optionally extracting sample polynucleic acid, (b) hybridizing the polynucleic acid as described above with at
30 least one probe according to claim 3, optionally a labelled probe, and (c) detecting the hybridized polynucleic acid.

12. A method for detecting a polypeptide of claim 6 or an antibody of claim 7 present in a biological sample, comprising: (a) contacting the biological sample for the presence and/or concentration of such polypeptide or antibody as defined above, and (b) detecting the immunological complex formed between said antibody and/or said polypeptide.

13. An antisense oligonucleotide reducing or inhibiting the expression of an HCBI, MSBI, MSSI or CMI polynucleic acid of claim 1 or a vector containing said antisense oligonucleotide.

14. A pharmaceutical composition comprising the antibody or antigen binding fragment thereof of claim 7 or the antisense oligonucleotide of claim 16 and a suitable pharmaceutical carrier.

15. A vaccine comprising an HCBI, MSBI, MSSI or CMI polynucleic acid of claim 1 or a polypeptide according to claim 6.

16. The vaccine of claim 15, which comprises a VLP or protein/DNA or polypeptide/DNA complex or specific proteins or attenuated infectious agents.

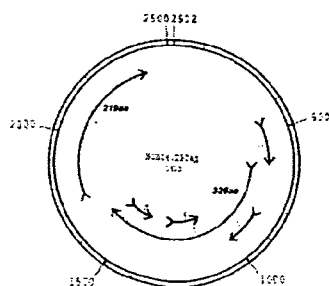
17. Use of an HCBI, MSBI, MSSI or CMI polynucleic acid of claim 1 as a lead component for the development of a medicament for prevention or treatment of cancer, a disease of the CNS or diabetes.

18. Use according to claim 17, wherein the cancer is breast cancer, ovarian cancer, lung cancer, prostate cancer, colorectal cancer or colon cancer and the disease of the CNS is Multiple sclerosis MS, amyotrophic lateral sclerosis, transmissible spongiforme encephalopathies/Prion-linked diseases, Parkinson's disease or Alzheimer disease.

*HCBI6.252 (2522bp)**(identical to CMI1.252)**(ri55.NnXn.2300.7)*

TTTTGTTTTTCAAAGGAAGTAATTAGCGAGTGTCTACGAGCGACACAATGAAAGTTCGATTATTCCTCCTCTAA
 AAAACCGCTTTTAAAAATATTGGTTGCTAGATGGTTTTAACTATCGTGAGCTTTTGCTTTTAAAAATACACGAGC
 AAAGCGAGTTCATAGTCGCTTTTGATCTGCTCTTACTAACAGCTCATTTTTAATGGGCGCTTAGTAAGAGATGCT
 TTTGGGGTTTCGGGGCTTGCCCTGAACAAGGTCACAAAGTGGAATTTATCACGGTAGTGAAAAAGTACCCTTTG
 TGTATCCTTGCTTTATGCTTTTTAAACCTTTGAACCTTTCCCTTAATTTGCAGAAATTGCCCTCGACTAAGCTT
 GCTTAGTCAAAAAAGTTTGAGCAAAGCGAAAACATAGGGCAATTTTCATGATGAAAATGGGCTTTTAAAGTTTTT
 AAATGCTTTTAAATGCTTTTAGACATGTTGAAATGCAAGCCTAGCAAGGCATACAGAGGACATACCCCTACGTTT
ACCGATCAATACCCCTACGTTTACCGATCAATACCCCTACGTTTACCGATCAATACCCCTACGTTTACCTTGCAT
 ATCACTACAAACAATATTAATGTAGTAATAAATTCAAATAATAAATGTAGGTTATGAGCGATTTAATAGTAAAA
 GATAACGCCCTAATGAACGCTAGTTACAACCTAGCTTTGGTTGAACAGCGTTTAATCCTTTTAGCTATATTGGAA
 GCGAGAGAAACAGGCAAAGGGATTAATGCCAATGATCCCTTAACAGTTTCATGCAAGTAGCTATATCAATCAATTT
 AACGTGGAAAGACATACGGCATATCAAGCATTAAAAGATGCTTGCAAAGACTTGTTTGCACGTCAATTTAGTTAC
 CAAGAAAAGCGAGAACGTGGACGAATTAATATTACAAGTCGATGGGTTTCACAAATTGGTTATATGGACGATACA
 GCAACCGTTGAAATTATTTTTGCCCCTGCGGTTGTTCTTTGATTACACGTTTAGAGGAACAATCACTCAATAC
 GATATTGAACAAATTAGCGAGTTATCGAGTGCTTATGCTGTCCGCCTATATGAGTTACTAATTTGTTGGCGTAGC
 ACAGGAAAAACCCCAATTATTGATCTAACAGAGTTTAGAAAGCGTCTAGGTGTTTTAGATACCGAATACACTAGA
 ACAGATAACTTGAAGATGCGAGTGATTGAGCTAGGACTTAAACAAATCAACGAGCATACGGACATTACAGCAAGC
 TACGAACAGCACAAAAAAGGGCGGACAATCACAGGATTCTCATTCAAGTTTAAGCAGAAGAAGAAAACAGGGGCG
 GAAATGcCAAAAAATAGCGATTCTAGCCACATATCGAAAAACCTAGTCAAATCCCTGCCAACATCGCAAAACAG
 CCTGAAAAATGcCAAAAAAGATGATTTAGGGCACAGAGCAAGCAAGATTACAGGGTTAATTATGTCTGAATGGTTT
 AGCAGATCGGTTTAAACGTGGTGATGAGTCAGTAATAGACATGATGAAACGCATCAAAGAGGAAATTACAACTGA
 TACAACAGCAGACCAATGGGAAAACAAGCTAGAAGAATTTGGCGTGATTTTTCAGTCATGATCAATTTCCCGAAG
 GACATATAAAAAATAATGTAGTTAAATATAGTAGTTATAATATTGACTACAATATTTAACTACTATATTGTAGTC
 TTAAATCAACTTACATAGAATTTAAATTATGTCTCTTCGCCAAATTTAGAACAACTCTCTCTTTGATCAAACTGA
 TGAAAGCTCTACTGTGCCACTATCAATTAGACTTCCAAACAACCTGAATAATGAGCTTGAAGAACTGACTATAAC
 ACTAGATAAATCCAAGAGTTTTCTACTCCTAGAGTTTCAATTAAGGCTGGAATTAGTGAGACTAGCAATATGTTAGA
 AGACAAAGCGAAGAACCCTATACTTGAAAGTCCAAGAGATTCTAGTGACTTTACAACCTATAAGAAAGTTTATGTT
 GAACACTAATTTCAATAACAATAAAGCATCACACATGACTATGTTGGAGAATCAAGAAGCTGCTGCATTTCTGTAA
 AGGTTGGAAAGAATATATCTGTAATTTAGAAGAAGGCGATAAGGTTTACCTTTATCAAAGTGGCGTTGGTTTTGT
 AGCATCTGGTATTGTCACTGGTGACTTAATCAAAAGTGAGTATCGTGGACAACCAGAAGAAAAATATTCTAAAAA
 ATAAAGGATTTTAAAAATAGGATTTAAAGCAATTTCTGCAAAGAGTTTAAAGACCTTAAGTGGCGGTGGTGCTAGC
 TTTCGTACAACAATGAGCGAGCTAAAGCCTGAACAATGTAGAAAATTAGATATGGAATAGAAAAACGGATGAAG
 TCAGTAAGTTAAAAATATGAGTTGTTGTTTTAGTGTGTACGTAGCAATCACTTAGAAGTTCCCTTTGTATTACT
 GTCTAGCAATGGTTTTTAAAAAGCGTCACTACTGATAGTGACGCTTT

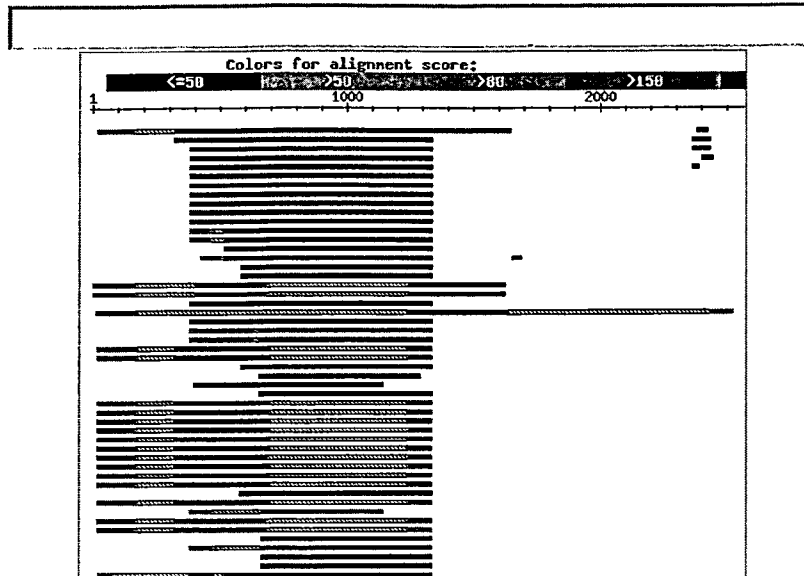
ATACCCCTACGTTTACCGATCA - repeat 3x + 17nt



86% nucleotide similarity to TSE-associated circular DNA isolate *Sphinx 1.76*

ORFs: 326aa – 85% similarity to Replicase 3-like protein of *Sphinx1.76*

Figure 1(a)

*219aa – regulation of transcription**Blastn – Geall:*

```
>>>nrnuc:GI_313507477 Gi|313507477|gb|HQ444404.1| TSE-associ... 1552 0.0
>>>nrnuc:GI_462476960 Gi|462476960|gb|APGK01000458.1| Dendro... 731 0.0
>>>nrnuc:GI_586961986 Gi|586961986|gb|CP006964.1| Acinetobac... 475 1e-129
```

```
>>>>nrnuc:GI_313507477 Gi|313507477|gb|HQ444404.1| TSE-associated circular DNA isolate
Sphinx 1.76, complete sequence. 0/0
Length=1758
Score = 1552 bits (1720), Expect = 0.0
Identities = 1152/1334 (86%), Gaps = 13/1334 (0%)
Strand=Plus/Plus
```

```
Query 320 TTTTAAACCTTTGAACTTTTCCT-TAATTTGCAGAAATTGCCCTCGACTAAGCTTGCT 378
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 277 TTCTAAACCTTTGACTTCTTCCCATTGTTTGCAGAAATTGCCCTCGA---AGCTTGCT 333

Query 379 TAGTCAAAAAGTTTGAGCAAAGCGAAAACATAGGGCAATTTTCATGATGAAAATGGGCT 438
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 334 TAGTCAAAAAGTTTGAGCAAAGCGAAAACATAGGGCAATTTTCATGAAGAAATGGGCT 393

Query 439 TTTAAAGTTTTTAAATGCTTTTAAATGCTTTTAGACATGTTGAAATGCAAGCCTAGCAAG 498
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 394 TTTAAAGTTTTTAAATG-TTTTAAATGCTTTTAGACATGCTAAGAAGCCACACAGCAAG 452

Query 499 GCATACAGAGGACATACCCCTACGTTTACCGATCAATACCCCTACGTTTACCGATCAATA 558
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 453 GCATACAGAGGACATTCTCCTACGTTTACCGATCAATACCC--CGTTTACCGATCAATA 510

Query 559 CCCCTACGTTTACCGATCAATACCCCTACGTTTACCTTGCATATCACTACAAACAATATT 618
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 511 CCCCTACGTTTACCGATCAATACCCCTACGTTTACCTTGCATATACTACAAAGAATACT 570

Query 619 AATGTAGTAATAACTTCAAAATAATAAATGTAGGTTATGAGCGATTTAATAGTAAAAGAT 678
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 571 AGTGTAGTA-TAACTTCAAAAGAATAATTGTAGGTTATGAGCGATTTAATAGTAAAAGAT 629

Query 679 AACGCCCTAATGAACGCTAGTTACAACCTAGCTTTGGTTGAACAGCGTTTAATCCTTTTA 738
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 630 AACGCCCTAATGAATGCTAGTTATAACTTAGCTTTGGTTGAACAGAGGTTAATTCTATTA 689

Query 739 GCTATATTGGAAGCGAGAGAGAAACAGGCAAAGGATTAAATGCCAATGATCCCTTAACAGTT 798
      || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 690 GCAATCATAGAAGCGAGAGAGAAACAGGCAAAGGATTAAATGCCAATGATCCTCTTACGTT 749

Query 799 CATGCAAGTAGCTATATCAATCAATTTAACGTGGAAAGACATACGGCATATCAAGCATTA 858
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Sbjct 750 CATGCAGGTAGCTATATCAATCAATTTAACGTACAAAGGCATACGGCATATCAAGCCCTC 809

Query 859 AAAGATGCTTGGCAAAGACTTGTTCACGTCATTTAGTTACCAAGAAAAGCGAGAACGT 918
|||||

Sbjct 810 AAAGATGCTTGTAAAGACTTGTTCGCCGTCATTCAGTTACCAAGAAAAGCGAGAACGA 869

Query 919 GGACGAATTAATATTACAAGTCGATGGGTTTCACAAATTGGTTATATGGACGATACAGCA 978
|||||

Sbjct 870 GGACGAATTAATATTACAAGTCGATGGGTTTCGCAAATTGGCTATATGGACGATACAGCA 929

Query 979 ACCGTTGAAATTATTTTGCCCTGCGGTTGTTCTTTGATTACACGTTTAGAGGAACAA 1038
|||||

Sbjct 930 ACCGTTGAGATTATTTTGCCCTGCGGTTGTTCTCTGATTACACGGCTAGAGGAACAG 989

Query 1039 TTCCTCAATACGATATTGAACAAATTAGCGAGTTATCGAGTGCTTATGCTGTCCGCTA 1098
|||||

Sbjct 990 TTCACCCAGTACGATATTGAGCAAATTAGCGGTTTATCGAGTGCATATGCTGTTTCGTATG 1049

Query 1099 TATGAGTTACTAATTTGTTGGCGTAGCACAGGAAAACCCCAATTATTGATCTAACAGAG 1158
|| || | |||

Sbjct 1050 TACGAACTGCTGATTGTTGGCGTAGCACAGGCAAACACCAATTATTGAGCTAGACGAG 1109

Query 1159 TTTAGAAAGCGTCTAGGTGTTTTAGATACCGAATACACTAGAACAGATAACTTGAAGATG 1218
|||||

Sbjct 1110 TTTAGAAAGCGAATAGGTGTTTTAGATACTGAATACACTAGAACAGATAATTTAAAGATG 1169

Query 1219 CGAGTGATTGAGCTAGGACTTAAACAAATCAACGAGCATACGGACATTACAGCAAGCTAC 1278
| ||| |||||

Sbjct 1170 CAAGTTATTGAATTAGCCCTAAACAAATCAACGAACATACTGACATCACAGCAAGCTAT 1229

Query 1279 GAACAGCACAAAAAGGGCGGACAATCACAGGATTCTCATTCAAGTTTAAGCAGAAG-AA 1337
|||||

Sbjct 1230 GAACAACACAAAAAGGGCGAGTGATTACCGGATTCTCATTTCATGTTTAAGCACAAGAAA 1289

Query 1338 GAAAACAGGGGCGGAAATGCCAAAAATAGCGATTCTAGCCACATATCGAAAAACCTAG 1397
|||||

Sbjct 1290 CAAAACA-GCGATAAAACGCTGATACTAACGCTTCTAGCCACGTATCGTAAACATAG 1348

Query 1398 TCAATCCCTGCCAACATCGCAAAACAGCCTGAAAATGCCAAAAAGATGATTAGGGCA 1457
|||||

Sbjct 1349 TCAATCCCTACCAACATTGTAACAGCCTGAAAACGCCAAATGAGCGATTAGAACAA 1408

Query 1458 CAGAGCAAGCAAGATTACAGGGTTAATTATGTCGAATGGTTTAGCAGATCGGTTTAAACG 1517
|||||

Sbjct 1409 TAGAGCGAGCCGTGTTACAGGGGAAATAATGCGAAATCGTCTGTCAGATCGGTTTAAACA 1468

Query 1518 TGGTGATGAGTCAGTAATAGACATGATGAAACGCA-TCAAAGAGGAAATTACAACGATA 1576
|| |||||

Sbjct 1469 AGGCGATGAATCAGCAATCGACATGATGAAACGTATTCAAAG-TGAAATAATAACCGATG 1527

Query 1577 CAACAGCAGACCAATGGGAAAACAAGCTAGAAGAATTTGGCGTGATTTTCAGTCATGAT 1636
|||

Sbjct 1528 CAATAGCAGACAGTGGGAAAGCAAACCTGGAGGAGTTTGGCGTGGTTTTTTAGTCATGA- 1586

Query 1637 CAATTTCCCGAAGG 1650
| |||||

Sbjct 1587 CGATTTCCCGAAGG 1600

Score = 228 bits (252), Expect = 2e-55
Identities = 143/153 (93%), Gaps = 1/153 (0%)
Strand=Plus/Plus

Query 22 ATTAGCGAGTGTCTACGAGCGACACAATGAAAGTTCGATTATTCCTCCTCTAAAAAAC 81
|||||

Sbjct 22 ATTAGCGAGTGTCTACGAGCGACTCAATGAAAGTTCGATTATTCCTCCTCTGAAAAAC 80

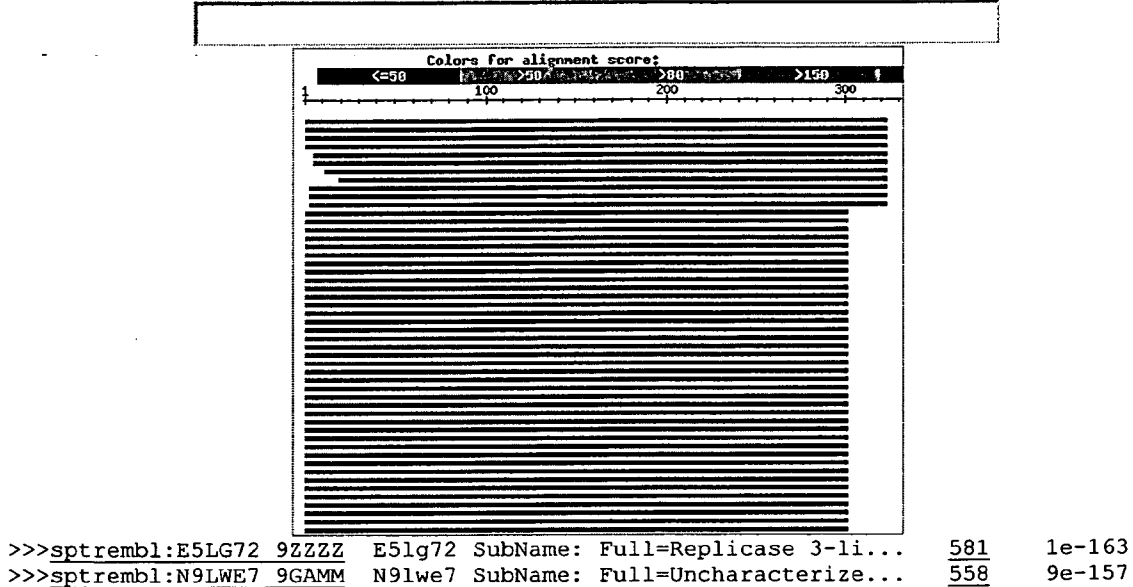
Query 82 GCTTTTAAAAATATTGGTTGCTAGATGGTTTTAACTATCGTGAGCTTTTGCTTTTAAAA 141
|||||

Sbjct 81 GCTTTTAAAAATATTGGCTGCTAGATGGTTTTTACTATAGTGAGGTTTGTCTTTAAAA 140

Query 142 TACACGAGCAAAGCGAGTTCATAGTCGCTTTTG 174
|||||

Sbjct 141 AACACGAGCAAAGCGAGTTCATAGTTGCTTTTG 173

BlastP – Uniprot:
ORF 326aa:



>>>sptrembl:E5LG72_9ZZZZ E5lg72 SubName: Full=Replicase 3-like protein *Sphinx1.76*; 5/2013
Length=324
Score = 581 bits (1498), Expect = 1e-163, Method: Compositional matrix adjust.
Identities = 278/324 (85%), Positives = 294/324 (90%), Gaps = 0/324 (0%)

Query 1 MSDLIVKDNALMNASYNLALVEQRLILLAILEARETGKGINANDPLTVHASSYINQFNVE 60
Sbjct 1 MSDLIVKDNALMNASYNLALVEQRLILLAILEARETGKGINANDPLTVHAGSYINQFNVQ 60

Query 61 RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSQIGYMDTATVEIIFAPAVVP 120
Sbjct 61 RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSQIGYMDTATVEIIFAPAVVP 120

Query 121 LITRLEEQFTQYDIEQISELSSAYAVRLYELLCWRSTGKTPIIDLTEFRKRLGVLDTHEY 180
Sbjct 121 LITRLEEQFTQYDIEQISGLSSAYAVRMYELLCWRSTGKTPIIELDEFKRIGVLDTHEY 180

Query 181 TRTDNLKMRVIELGLKQINEHTDITASYEQHKKGRTITGFSFKFKQKKKTGAEMPKNSDS 240
Sbjct 181 TRTDNLKM+VIEL LKQINEHTDITASYEQHKKGRTITGFSF FK KK+ + P + S 240

Query 241 SPHIEKPSQIPANIAKQPENAKKDDLHRSKITGLIMSNGLADRFKRGDESVIDMMKRI 300
Sbjct 241 SPRIVKHSQIPTNIVKQPENAKMSDLEHRASRVTGEIMRNRLSDRFKQGDESAIDMMKRI 300

Query 301 KEEITTDTTADQWENKLEEFVGF 324
+ EI TD ADQWE+KLEEFVGF
Sbjct 301 QSEIITDAIADQWESKLEEFVGVF 324

>>>sptrembl:N9LWE7_9GAMM N9lwe7 SubName: Full=Uncharacterized protein;. 12/2013
Length=324
Score = 558 bits (1439), Expect = 9e-157, Method: Compositional matrix adjust.
Identities = 267/324 (82%), Positives = 290/324 (89%), Gaps = 0/324 (0%)

Query 1 MSDLIVKDNALMNASYNLALVEQRLILLAILEARETGKGINANDPLTVHASSYINQFNVE 60
Sbjct 1 MSELIVKDNALMNASYNLALVEQRLILLAILEARESGKGINANDPLTVHAESYINQFVGH 60

Query 61 RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSQIGYMDTATVEIIFAPAVVP 120
Sbjct 61 RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSQIY+D+TATVEIIFAPAVVP 120

Query 121 LITRLEEQFTQYDIEQISELSSAYAVRLYELLCWRSTGKTPIIDLTEFRKRLGVLDTHEY 180
Sbjct 121 LITRLEEQFTQYDIEQISGLSSAYAVRMYELLCWRSTGKTPVIELGEFRKRIEVLDTHEY 180

Query 181 TRTDNLKMRVIELGLKQINEHTDITASIEQHKKGRTITGFSFKFKQKKKTGAEMPKNSDS 240
 RTDNLKMRVIEL LKQIN+HTDITA+YEQHKKGRTITGFSFKFKQK+KT E PKNSDS
 Sbjct 181 IRTDNLKMRVIELALKQINDHTDITATYEQHKKGRTITGFSFKFKQKRKTELETPKNSDS 240

Query 241 SPHIEKPSQIPANIAKQPENAKDDLGHASKITGLIMSNGLADRFKRGDESVIDMMKRI 300
 +K +IP N+ KQENA DL HRASKITGLIMSN L+DRFK+ DES++ MM RI
 Sbjct 241 DVPKQKSVEIPTNLVKQPENANMSDLQHRASKITGLIMSNRLSDRFKQSDSIMQMARI 300

Query 301 KEEITTTDTADQWENKLEEFVIF 324
 + EIT + A+QWENKLEEFVIF
 Sbjct 301 QSEITDEAIANQWENKLEEFVIF 324

General Description References Copyright Links Sequence

General Information about the Entry

Entry name	SPTREMBL:E5LG72_9ZZZZ
Prim. accession #	E5LG72
Created	8-FEB-2011
Last sequence update	version 1 8-FEB-2011
Last annotation update	version 8 1-MAY-2013

Description and Origin of the Protein

Keywords	
Description	subname: full=replicase 3-like protein;
Gene name(s)	
Organism source	TSE-associated circular DNA.
Organism classification	unclassified sequences.
NCBI Taxonomy ID	932710;

References

1.	Manuelidis, L.; Nuclease resistant circular DNAs copurify with infectivity in scrapie and CJD. J. Neurovirol. 0:0 (2010)
	Position nucleotide sequence.
	Comment strain=sphinx 1.76;

Copyright

Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms> Distributed under the Creative Commons Attribution-NoDerivs License

Database Cross-references

EMBL	HQ444404; ADR65123.1; -; Genomic_DNA.
GO	GO:0005727; C:extrachromosomal circular DNA; IEA:InterPro. GO:0003887; F:DNA-directed DNA polymerase activity; IEA:InterPro. GO:0006270; P:DNA replication initiation; IEA:InterPro.
Gene3D	1.10.10.10; -; 2.
InterPro	IPR000525; Initiator_Rep_prot. IPR011991; WHTH_DNA-bd_dom.
Pfam	PF01051; Rep_3; 1.

Sequence information

Characteristics	Length: 324 aa, molecular weight: 37186 Da, CRC64 check sum: 3F0A0F966DD3DCAF
	MSDLIVKDNA LMNASYNLAL VEQRLILLAI IEARETGKGI NANDPLTVHA GSYINQFNQ 60 RHTAYQALKD ACKDLFARQF SYQEKREGR INITSRWVSQ IGYMDDTATV EIIFAPAVVP 120 LITRLEEQFT QYDIEQISGL SSAYAVRMYE LLICWRSTGK TPIELDEFR KRIGVLDTEY 180 TRTDNLKMQV IELALKQINE HTDITASIEQ HKKGRVITGF SFMFKHKKQN SDKTPDTNAS 240 SPRIVKHSQI PTNIVKQPEN AKMSDLEHRA SRVTGEIMRN RLSDRFKQGD ESAIDMMKRI 300

QSEIITDAIA DQWESKLEEF GVVF

324
//**ProtSweep**

A Protein Identification Tool

Results for /home/vir088/rindneu/tsekompl/CMI1.252p.316.pep - identical to HCBI6.252
326.pepin the first 300aa

Number of residues: 316 Molecular weight: 35.91 kD
Charge: 8 Isoelectric point: 10.04
more...

Prediction of protein localization:

cyto	13 %
extr	10 %
cyto nucl	10 %

Identified Protein & Transcript:

Protein:
id: sptrembl:E5LG72_9ZZZZ
valuation: homologous
scientific organism: Dasypus novemcinctus
common organism: Nine-banded armadillo
abstract: E5lg72 SubName: Full-Replicase 3-like protein; 5/2013Length=324
isoform:

GO information:

accession	group	name	evidence
GO:0005727	cellular component	extrachromosomal circular DNA	IEA:InterPro
GO:0003887	molecular function	DNA-directed DNA polymerase activity	IEA:InterPro
GO:0006270	biological process	DNA replication initiation	IEA:InterPro

Interpro information:

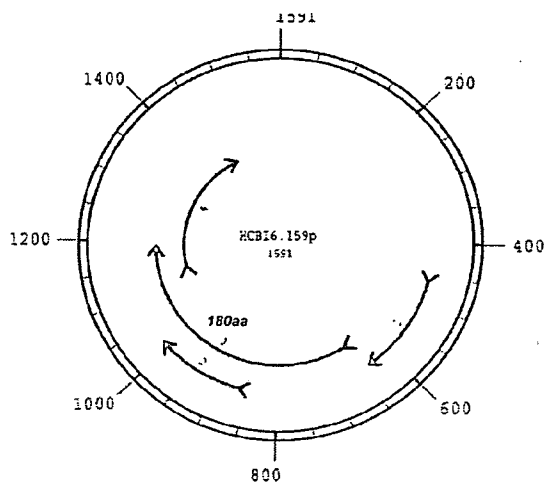
id	short name
IPR000525	Initiator_Rep_prot
IPR011991	WTHH_DNA-bd_dom

HCBI6.159 (1591bp)**(ri55.NnXn.1600.5)**

TTTTGTTTTTCAAAGGAAGTAATTGGCGAGTGTCTACGAGCGACACAATGAAAGTTCGATTATTCCTT
 CCTCTAAAAAACCGCTTTTAAAAATATTGGTTGCTAGATGGTTTTAACTATCGTGAGCTTTTGCTTTT
 AAAAATACACGAGCAAAGCGAGTTCATAGTCGCTTTTGATCTGCTCTTACTAACAGCTCATTTTTAAT
 GGGCCGTTAGTAAGAGATGCTTTTGGGGTTTCGGGGCTTGCCCTGAACAAGGTCACAAAGTGGGAATT
 TATCACGGTAGTGAAAAAGTACCCCTTGTGTATCCTTGCTTTATGCTTTTTAAACCTTTGAACTTTTC
 CCTTAATTTGCAGAAATTGCCCTCGACTAAGCTTGCTTAGTCAAAAAAGTTTGAGCAAAGCGAAAAC
 ATAGGGCAATTTTCATGATGAAAATGGGCTTTTAAAGTTTTTAAATGCTTTTAAATGCTTTTAGACAT
 GTTGAAATGCAAGCCTAGCAAGGCACACAGAGGACATACCCCTGCGTTTACCGATCAATACCCCTACG
TTTACCGATCAATACCCCTACGTTTACCGATCAATACCCCTACGTTTACCTTGCATATCACTACAAAC
 AATATTAATGTAGTAATAACTTCAAAATAATAAATGTAGGTTATGAGCGATTTAATAGTAAAAGATAA
 CGCCCTAATGAACGCTAGTTACAACCTTAGCTTTGGTTGAACAGCGTTTAAATCCTTTTAGCTATATTGG
 AAGCGAGAGAAACAGGCAAAGGGATTAATGCCAATGATCCCTTAACAGTTCATGCAAGTAGCTATATC
 AATCAATTTAACGTGGAAAGACATACGGCATATCAAGCATTAAAGATGCTTGCAAAGACTTGTTTGC
 ACGTCAATTTAGTTACCAAGAAAAGCGAGAACGTGGACGAATTAATATTACAAGTCGATGGGTTTCAC
 AAATTGGTTATATGGACGATACAGCAACCGTTGAAATTAATTTTTGCCCTGCGGTTGTTCCTTTGATT
 ACACGTTTAGAGGA "ACAATTCACCTCAATACGATATTGAACAAATTAGCGAGTTATCGAGTGCTTATG
 CTGTCCGCCTATATGAGTTACTAATTTGTTGGCGTAGCACAGGACTATGTTGGAGAATCAAGAAGCTG
 CTGCATTTTCGTAAAGGTTGGAAAGAATATATCTGTAATTTAGAAGAAGGCGATAAGGTTTACCTTTAT
 CAAAGTGGCGTTGGTTTTGTAGCATCTGGTATTGTCACCTGGTGACTTAATCAAAGTGAGTATCGTGG
 ACAACCAGAAGAAAAATATTCTAAAAAATTAAAGGATTTTAAATAGGATTTAAAGCAATTTCTGCAA
 AAGAGTTTAAGACCTTAACCTGGCGGTGGTGCTAGCTTTCGTACAACAATGAGCGAGCTAAAGCCTGAA
 CAATGTAGAAAATTAGATATGGAAATAGAAAACGGATGAAGTCAGTAAGTTAAAAATATGAGTTGTT
 GTTTTAGTGTGTACGTAGCAATCACTTAGAAGTTCCCTTTGTATTACTTGTCTAGCAATGGTTTTAA
 AAAGCGTCACTACTGATAGTGACGCTTT

ATACCCCTACGTTTACCGATCA - repeat 3x + 17nt

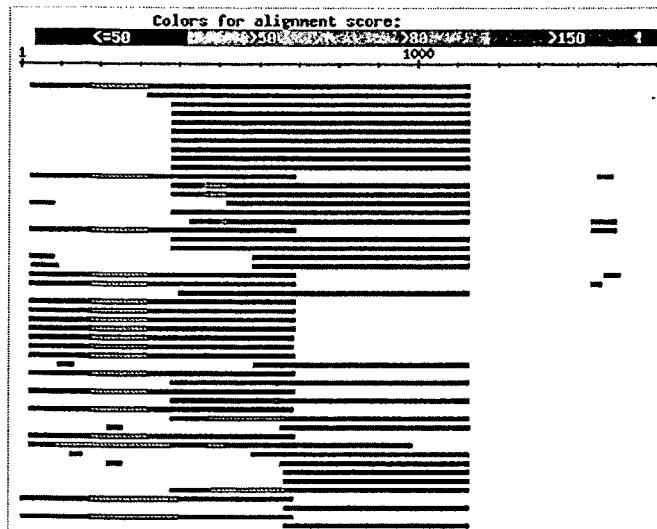
" - sequence of HCBI6.159 is same as HCBI6.252 except that nt 1132-2163 has been deleted. Both sequences were generated by the same back-to-back primers.



89% nucleotide similarity to TSE-associated circular DNA isolate *Sphinx* 1.76

ORF 180aa - 96% similarity to similarity to Replicase 3-like protein of *Sphinx* 1.76

Figure 1(b)

Blastn-Geall:

```
>>>nrnuc:GI_313507477  Gi|313507477|gb|HQ444404.1| TSE-associ... 1070 0.0
>>>nrnuc:GI_462476960  Gi|462476960|gb|APGK01000458.1| Dendro... 560 2e-155
```

```
>>>>nrnuc:GI_313507477 Gi|313507477|gb|HQ444404.1| TSE-associated circular DNA isolate
Sphinx 1.76, complete sequence. 0/0
Length=1758
```

```
Score = 1070 bits (1186), Expect = 0.0
Identities = 728/812 (89%), Gaps = 8/812 (0%)
Strand=Plus/Plus
```

```
Query 320 TTTTAAACCTTTGAACTTTTCCTT-TAATTTGCAGAAATTGCCCTCGACTAAGCTTGCT 378
          || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 277 TTCTAAACCTTTGACTTCTTCCCATTGTTTGCAGAAATTGCCCTCGA---AGCTTGCT 333

Query 379 TAGTCAAAAAAGTTTGAGCAAAGCGAAACATAGGGCAATTTTCATGATGAAAATGGGCT 438
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 334 TAGTCAAAAAAGTTTGAGCAAAGCGAAACATAGGGCAATTTTCATGAAGAAATGGGCT 393

Query 439 TTTAAAGTTTTTAAATGCTTTTAAATGCTTTTAGACATGTTGAAATGCAAGCCTAGCAAG 498
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 394 TTTAAAGTTTTTAAATG-TTTTAAATGCTTTTAGACATGCTAAGAAGCCACACAGCAAG 452

Query 499 GCACACAGAGGACATACCCCTGCGCTTTACCGATCAATACCCCTACGTTTACCGATCAATA 558
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 453 GCATACAGAGGACATTCTCTACGTTTACCGATCAATACCC--CGTTTACCGATCAATA 510

Query 559 CCCCTACGTTTACCGATCAATACCCCTACGTTTACCTTGCATATCACTACAAACAATATT 618
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 511 CCCCTACGTTTACCGATCAATACCCCTACGTTTACCTTGCATATACTACAAAGAATACT 570

Query 619 AATGTAGTAATAACTTCAAAATAATAAATGTAGGTTATGAGCGATTAAATAGTAAAAGAT 678
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 571 AGTGTAGTA-TAAGTTCAAAAGAATAATTGTAGGTTATGAGCGATTAAATAGTAAAAGAT 629

Query 679 AACGCCCTAATGAACGCTAGTTACAACCTAGCTTTGGTTGAACAGCGTTTAAATCCTTTTA 738
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 630 AACGCCCTAATGAATGCTAGTTATAACTTAGCTTTGGTTGAACAGAGGTTAATCTATTA 689

Query 739 GCTATATTGGAAGCGAGAGAAACAGGCAAAGGGATTAATGCCAATGATCCCTTAACAGTT 798
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 690 GCAATCATAGAAGCGAGAGAAACAGGCAAAGGGATTAATGCCAATGATCCTCTTACGGTT 749

Query 799 CATGCAAGTAGCTATATCAATCAATTTAACGTGGAAGACATACGGCATATCAAGCATT 858
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 750 CATGCAGGTAGCTATATCAATCAATTTAACGTACAAGGCATACGGCATATCAAGCCCTC 809

Query 859 AAAGATGCTTGCAAAGACTTGTTTGCACGTCAATTTAGTTACCAAGAAAAGCGAGAACGT 918
```

```

Query   318   CTTTTTAAACCTTTGAACCTTTTCCCTTAATTTGCAGAAATTGCCCTCGACTAAGCTT-G   376
          |||||
Sbjct  1380   CTTTTTAAACCTTTGACTTTCCCTATCCTTTGCATAAATTGCCCTT-GACGTAGCAAAG   1322

```

```
>>>sptrembl:E5LG72 92ZZZ E5lg72 SubName: Full=Replicase 3-li... 324 7e-87
```

>>>sptrembl:N9LWE7_9GAMM N9lwe7 SubName: Full=Uncharacterize... 302 5e-80
>>>sptrembl:N9JLM2_ACIBA N9jlm2 SubName: Full=Uncharacterize... 286 2e-75

>>>>sptrembl:E5LG72_9ZZZZ E5lg72 SubName: Full=Replicase 3-like protein;. 5/2013
Length=324
Score = 324 bits (831), Expect = 7e-87, Method: Compositional matrix adjust.
Identities = 154/159 (96%), Positives = 157/159 (98%), Gaps = 0/159 (0%)

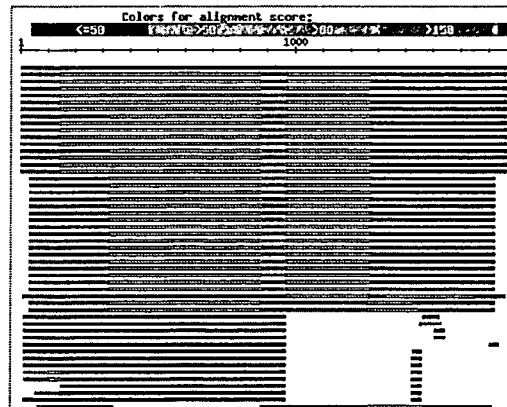
Query 1 MSDLIVKDNALMNASYNLALVEQRLILLAILEARETGKGINANDPLTVHASSYINQFNVE 60
MSDLIVKDNALMNASYNLALVEQRLILLAI+EARETGKGINANDPLTVHA SYINQFNV+
Sbjct 1 MSDLIVKDNALMNASYNLALVEQRLILLAILEARETGKGINANDPLTVHAGSYINQFNVQ 60
Query 61 RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSQIGYMDDTATVEIIFAPAVVP 120
RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSQIGYMDDTATVEIIFAPAVVP
Sbjct 61 RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSQIGYMDDTATVEIIFAPAVVP 120
Query 121 LITRLEEQFTQYDIEQISELSSAYAVRLYELLICWRSTG 159
LITRLEEQFTQYDIEQIS LSSAYAVR+YELLICWRSTG
Sbjct 121 LITRLEEQFTQYDIEQISGLSSAYAVRMYELLICWRSTG 159

>>>>sptrembl:N9LWE7_9GAMM N9lwe7 SubName: Full=Uncharacterized protein;. 12/2013
Length=324
Score = 302 bits (773), Expect = 5e-80, Method: Compositional matrix adjust.
Identities = 144/159 (90%), Positives = 151/159 (94%), Gaps = 0/159 (0%)

Query 1 MSDLIVKDNALMNASYNLALVEQRLILLAILEARETGKGINANDPLTVHASSYINQFNVE 60
MS+LIVKDNALMNASYNL LVEQRLILLAI+EARE+GKGINANDPLTVHA SYINQF V
Sbjct 1 MSELIVKDNALMNASYNLDLVEQRLILLAIVEARESGKGINANDPLTVHAESYINQFGVH 60
Query 61 RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSQIGYMDDTATVEIIFAPAVVP 120
R TAYQALKDACKDLFARQFSYQEKREGR NITSRWVSI Y+D+TATVEIIFAPAVVP
Sbjct 61 RTTAYQALKDACKDLFARQFSYQEKREGRANITSRWVSIAYIDETATVEIIFAPAVVP 120
Query 121 LITRLEEQFTQYDIEQISELSSAYAVRLYELLICWRSTG 159
LITRLE+QFTQYDIEQIS LSSAYAVR+YELLICWRSTG
Sbjct 121 LITRLEKQFTQYDIEQISGLSSAYAVRMYELLICWRSTG 159

>>>>sptrembl:N9JLM2_ACIBA N9jlm2 SubName: Full=Uncharacterized protein;. 12/2013
Length=325
Score = 286 bits (733), Expect = 2e-75, Method: Compositional matrix adjust.
Identities = 134/160 (83%), Positives = 150/160 (93%), Gaps = 1/160 (0%)

Query 1 MSDLIVKDNALMNASYNLALVEQRLILLAILEARETGKGINANDPLTVHASSYINQFNVE 60
MS+LIVKDNALMNASYNLALVEQRLILLAI+EARE+G+GINANDPLTVHA SYINQF VE
Sbjct 1 MSELIVKDNALMNASYNLALVEQRLILLAILEARESGRGINANDPLTVHAESYINQFKVE 60
Query 61 RHTAYQALKDACKDLFARQFSYQEKREGR-INITSRWVSQIGYMDDTATVEIIFAPAVV 119
+HTAYQALKDACKDLFARQFSYQ++ +G N+TSRW+SQIGY ++ ATV+++FAPAVV
Sbjct 61 KHTAYQALKDACKDLFARQFSYQQRSAKGN+TNMTSRWISQIGYTENEATVQLVFAPAVV 120
Query 120 PLITRLEEQFTQYDIEQISELSSAYAVRLYELLICWRSTG 159
PLITRLEEQFT+YDIEQIS LSSAYAVRLYELLICWRSTG
Sbjct 121 PLITRLEEQFTKYDIEQISLSSAYAVRLYELLICWRSTG 160

Blastn - Geall:

>>>nrnuc:GI_313507477 Gi|313507477|gb|HQ444404.1| TSE-associ... 2491 0.0
>>>nrnuc:GI_347602273 Gi|347602273|gb|JN541306.1| Uncultured... 886 0.0

>>>>nrnuc:GI_313507477 Gi|313507477|gb|HQ444404.1| TSE-associated circular DNA isolate
Sphinx 1.76, complete sequence. 0/0
Length=1758

Score = 2491 bits (2762), Expect = 0.0
Identities = 1419/1439 (98%), Gaps = 5/1439 (0%)
Strand=Plus/Plus

Query	1	AAGCTTGCTTAGTCAAAAAAGTTTGAGCAAAGCGAAAACATAGGGCAATTTTCATGAAGA	60
Sbjct	325	AAGCTTGCTTAGTCAAAAAAGTTTGAGCAAAGCGAAAACATAGGGCAATTTTCATGAAGA	384
Query	61	AATTGGGCTTTTAAAGTTTAAATGCTTTTAAATGCTTTTAGACATGCTAAGAAGCCCCA	120
Sbjct	385	AATTGGGCTTTTAAAGTTTAAATG-TTTTAAATGCTTTTAGACATGCTAAGAAGCCCCA	443
Query	121	CACAGCAAGGCATACAGAGGACATTCTCTACGTTTACCGATCAATACCCCTACGTTTAC	180
Sbjct	444	CACAGCAAGGCATACAGAGGACATTCTCTACGTTTACCGATCAATACCCC--CGTTTAC	501
Query	181	CGATCAATACCCCTACGTTTACCGATCAATACCCCTACGTTTACCTTGCGTATAACTACA	240
Sbjct	502	CGATCAATACCCCTACGTTTACCGATCAATACCCCTACGTTTACCTTGCGTATAACTACA	561
Query	241	AAGAATACTAGTGTAGTAATAACTTCAAAAGAATAATGTAGGTTATGAGCGATTTAATA	300
Sbjct	562	AAGAATACTAGTGTAGTA-TAACTTCAAAAGAATAATGTAGGTTATGAGCGATTTAATA	620
Query	301	GTAAAAGATAACGCCCTAATGAATGCTAGTTATAACTTAGCTTTGGTTGAACAGAGGTTA	360
Sbjct	621	GTAAAAGATAACGCCCTAATGAATGCTAGTTATAACTTAGCTTTGGTTGAACAGAGGTTA	680
Query	361	ATTCTATTAGCAATCATAGAAGCGAGAGAAACAGGCAAAGGATTAATGCCAATGATCCT	420
Sbjct	681	ATTCTATTAGCAATCATAGAAGCGAGAGAAACAGGCAAAGGATTAATGCCAATGATCCT	740
Query	421	TTACAGTTCATGCAAGTAGCTATATCAATCAATTTAACGTAGAAAGGCATACGGCATAT	480
Sbjct	741	CTTACGGTTCATGCAGGTAGCTATATCAATCAATTTAACGTACAAAGGCATACGGCATAT	800
Query	481	CAAGCCCTCAAAGATGCTTGTAAGACTTGTTTGCCCGTCAATTCAGTTACCAAGAAAAG	540
Sbjct	801	CAAGCCCTCAAAGATGCTTGTAAGACTTGTTTGCCCGTCAATTCAGTTACCAAGAAAAG	860
Query	541	CGAGAACGAGGACGAATTAATATTACAAGTCGATGGGTTTCGCAAATTGGCTATATGGAC	600
Sbjct	861	CGAGAACGAGGACGAATTAATATTACAAGTCGATGGGTTTCGCAAATTGGCTATATGGAC	920
Query	601	GATACAGCAACCGTTGAGATTATTTTGGCCCTGCGGTTGTTCTCTGATTACACGGCTA	660
Sbjct	921	GATACAGCAACCGTTGAGATTATTTTGGCCCTGCGGTTGTTCTCTGATTACACGGCTA	980

Query 661 GAGGAACAGTTCACCCAGTACGATATTGAGCAAATTAGCGGTTTATCGAGTGCATATGCT 720
|||||
Sbjct 981 GAGGAACAGTTCACCCAGTACGATATTGAGCAAATTAGCGGTTTATCGAGTGCATATGCT 1040
Query 721 GTTCGTATGTACGAACTGCTGATTGTTGGCGTAGCACAGGCAAAACACCAATTATTGAG 780
|||||
Sbjct 1041 GTTCGTATGTACGAACTGCTGATTGTTGGCGTAGCACAGGCAAAACACCAATTATTGAG 1100
Query 781 CTAGACGAGTTTAGAAAGCGAATAGGTGTTTGTAGATACTGAATACACTAGAACAGATAAT 840
|||||
Sbjct 1101 CTAGACGAGTTTAGAAAGCGAATAGGTGTTTGTAGATACTGAATACACTAGAACAGATAAT 1160
Query 841 TTAAAGATGCGAGTTATTGAATTAGCCCTAAAACAAATCAACGAACATACAGACATCACA 900
|||||
Sbjct 1161 TTAAAGATGCAAGTTATTGAATTAGCCCTAAAACAAATCAACGAACATACTGACATCACA 1220
Query 901 GCAAGCTATGAACAACACAAAAAAGGCGAGTGATTACAGGATTCTCATTCAAGTTTAAG 960
|||||
Sbjct 1221 GCAAGCTATGAACAACACAAAAAAGGCGAGTGATTACCGGATTCTCATTTCATGTTTAAG 1280
Query 961 CACAAGAAACAAAAACAGCGATAAAACGCCAAAAAATAGCGATTCTAGCCACGTATCGTA 1020
|||||
Sbjct 1281 CACAAGAAACAAAAACAGCGATAAAACGCCCTGATACTAACGCTTCTAGCCACGTATCGTA 1340
Query 1021 AAACATAGTCAAATCCCTACCAACATTGTAAAACAGCCTGAAAACGCCAAAAATGAGCGAT 1080
|||||
Sbjct 1341 AAACATAGTCAAATCCCTACCAACATTGTAAAACAGCCTGAAAACGCCAAAAATGAGCGAT 1400
Query 1081 TTAGAACATAGAGCGAGCCGTGTTACAGGGGAAATAATGCGAAATCGTCTGTCAGATCGG 1140
|||||
Sbjct 1401 TTAGAACATAGAGCGAGCCGTGTTACAGGGGAAATAATGCGAAATCGTCTGTCAGATCGG 1460
Query 1141 TTTAAACAAGGCGATGAATCAGCAATCGACATGATGAAACGTATTCAAAGTGAAATAATA 1200
|||||
Sbjct 1461 TTTAAACAAGGCGATGAATCAGCAATCGACATGATGAAACGTATTCAAAGTGAAATAATA 1520
Query 1201 ACCGATGCAATAGCAGACCAGTGGGAAAGCAAACCTGGAGGAGTTTGGCGTGGTTTTTAG 1260
|||||
Sbjct 1521 ACCGATGCAATAGCAGACCAGTGGGAAAGCAAACCTGGAGGAGTTTGGCGTGGTTTTTAG 1580
Query 1261 TCATGACGATTTCCCGAAGGGCGCACTTAGCCATTGAGAAAAATCTTCGATTTTTTCAAT 1320
|||||
Sbjct 1581 TCATGACGATTTCCCGAAGGGCGCACTTAGCCATTGAGAAAAATCTTCGATTTTTTCAAT 1640
Query 1321 GGAAGTCCGTGGGGGTAAACCCCTCAACCCCAAAAGCAAAAACACTGTAATCAGGGAAAA 1380
|||||
Sbjct 1641 GGAAGTCCGTGGGGGTAAACCCCTCAACCCCAAAAGCAAAAACACTGTAATCAGGGAAAA 1700
Query 1381 AACATTTTTGATTTTGATCCTTGTTTGTCTACTCGTAGACACTCGTTTGTGTTTGCTCTT 1439
|||||
Sbjct 1701 AACATTTTTGATTTTGATCCTTGTTTGTCTACTCGTAGACACTCGTTTGTGTTTGCTCTT 1758

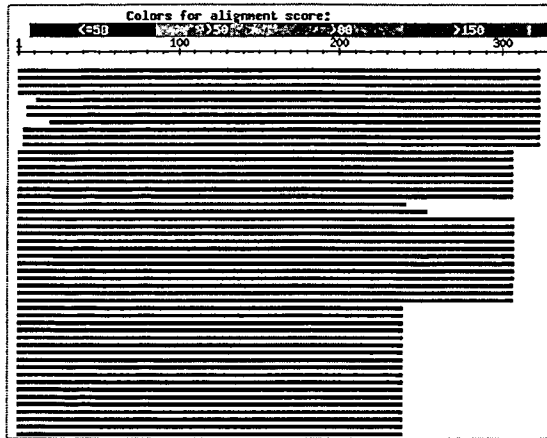
Score = 587 bits (650), Expect = 2e-163
Identities = 325/325 (100%), Gaps = 0/325 (0%)
Strand=Plus/Plus

Query 1440 TCTAGAATTACAAAAAAGATATTAGCGAGTGCTACGAGCGACTCAATGAAAGTTCGAT 1499
|||||
Sbjct 1 TCTAGAATTACAAAAAAGATATTAGCGAGTGCTACGAGCGACTCAATGAAAGTTCGAT 60
Query 1500 TATTCGCCCTCTGGAAAACCGCTTTTAAAAATATTGGCTGCTAGATGGTTTTTACTATAG 1559
|||||
Sbjct 61 TATTCGCCCTCTGGAAAACCGCTTTTAAAAATATTGGCTGCTAGATGGTTTTTACTATAG 120
Query 1560 TGAGGTTTTGCTTTTaaaaaaCACGAGCAAAGCGAGTTCATAGTTGCTTTTGCTTGTTT 1619
|||||
Sbjct 121 TGAGGTTTTGCTTTTAAAAAACACGAGCAAAGCGAGTTCATAGTTGCTTTTGCTTGTTT 180
Query 1620 TCGGGTCTTAGGGGAAATCCCTAACAAAGTCCTCGAATATCAAAATGTGGCTACATTTTG 1679
|||||
Sbjct 181 TCGGGTCTTAGGGGAAATCCCTAACAAAGTCCTCGAATATCAAAATGTGGCTACATTTTG 240
Query 1680 TATATACGGGTAGGCTTGCTTATTTGATtttttttCTTCTAAACCTTTGACTTCTTCCCC 1739
|||||
Sbjct 241 TATATACGGGTAGGCTTGCTTATTTGATTTTTTTTCTTCTAAACCTTTGACTTCTTCCCC 300
Query 1740 ATTGTTTGCAGAAATTGCCCTCGA 1764

Sbjct 301 |||||
ATTGTTTGCAGAAATTGCCCTCGA 325

BlastP –Uniprot:

ORF 324aa:



```
>>>sptrembl:E5LG72 9ZZZZ E5lg72 SubName: Full=Replicase 3-li... 658 0.0
>>>sptrembl:N9LWE7 9GAMM N9lwe7 SubName: Full=Uncharacterize... 569 5e-160
```

```
>>>sptrembl:E5LG72_9ZZZZ E5lg72 SubName: Full=Replicase 3-like protein;. 5/2013
Length=324
Score = 658 bits (1698), Expect = 0.0, Method: Compositional matrix adjust.
Identities = 316/324 (97%), Positives = 319/324 (98%), Gaps = 0/324 (0%)
```

Query	1	MSDLIVKDNALMNASYNLALVEQRLILLAIIEARETGKGINANDPLTVHASSYINQFNVE	60
		MSDLIVKDNALMNASYNLALVEQRLILLAIIEARETGKGINANDPLTVHASYINQFNV+	
Sbjct	1	MSDLIVKDNALMNASYNLALVEQRLILLAIIEARETGKGINANDPLTVHAGSYINQFNVQ	60
Query	61	RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSQIGYMDDTATVEIIFAPAVVP	120
		RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSQIGYMDDTATVEIIFAPAVVP	
Sbjct	61	RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSQIGYMDDTATVEIIFAPAVVP	120
Query	121	LITRLEEFTQYDIEQISGLSSAYAVRMYELLCWRSTGKTPIIELDEFKRKIGVLDTEY	180
		LITRLEEFTQYDIEQISGLSSAYAVRMYELLCWRSTGKTPIIELDEFKRKIGVLDTEY	
Sbjct	121	LITRLEEFTQYDIEQISGLSSAYAVRMYELLCWRSTGKTPIIELDEFKRKIGVLDTEY	180
Query	181	TRTDNLKMRVIELALKQINEHTDITASYEQHKKGRVITGFSFKHKKQNSDKTPKNSDS	240
		TRTDNLKM+VIELALKQINEHTDITASYEQHKKGRVITGFSFKHKKQNSDKTP + S	
Sbjct	181	TRTDNLKMQVIELALKQINEHTDITASYEQHKKGRVITGFSFMFKHKKQNSDKTPDTNAS	240
Query	241	SPRIVKHSQIPTNIVKQPENAKMSDLEHRASRVTEIMRNRLSDRFKQGDESAIDMMKRI	300
		SPRIVKHSQIPTNIVKQPENAKMSDLEHRASRVTEIMRNRLSDRFKQGDESAIDMMKRI	
Sbjct	241	SPRIVKHSQIPTNIVKQPENAKMSDLEHRASRVTEIMRNRLSDRFKQGDESAIDMMKRI	300
Query	301	QSEIITDAIADQWESKLEEFQVVF	324
		QSEIITDAIADQWESKLEEFQVVF	
Sbjct	301	QSEIITDAIADQWESKLEEFQVVF	324

>>>sptrembl:N9LWE7_9GAMM N9lwe7 SubName: Full=Uncharacterized protein;. 12/2013
Length=324

Score = 569 bits (1467), Expect = 5e-160, Method: Compositional matrix adjust.
Identities = 272/324 (83%), Positives = 294/324 (90%), Gaps = 0/324 (0%)

Query	1	MSDLIVKDNALMNASYNLALVEQRLILLAIIEARETGKGINANDPLTVHASSYINQFNVE	60
		MS+LIVKDNALMNASYNL LVEQRLILLAI+EARE+GKGINANDPLTVHA SYINQF V	
Sbjct	1	MSELIVKDNALMNASYNLDLVEQRLILLAIIVEARESGKGINANDPLTVHAESYINQFGVH	60
Query	61	RHTAYQALKDACKDLFARQFSYQEKRRERGRINITSRWWSQIGYMDDTATVEIIFAPAVVP	120
		R TAYQALKDACKDLFARQFSYQEKRRERGR NITSRWWSQI Y+D+TATVEIIFAPAVVP	
Sbjct	61	RTTAYQALKDACKDLFARQFSYQEKRRERGRANITSRWWSQIAYIDETATVEIIFAPAVVP	120
Query	121	LITRLEEQFTQYDIEQISGLSSAYAVRMYELLICWRSTGKTPIIELDEFKRKIGVLDTEY	180
		LITRLE+OFTQYDIEQISGLSSAYAVRMYELLICWRSTGKTP+IEL EFRKRKIGVLDTEY	

Sbjct 121 LITRLEKQFTQYDIEQISGLSSAYAVRMYELICWRSTGKTPVIELGEFRKRIGVLDTEY 180

Query 181 TRTDNLKMRVIELALKQINEHTDITASIEQHKKGRVITGFSFKFKHKQNSDKTPKNSDS 240
RTDNLKMRVIELALKQIN+HTDITA+YEQHKKGR ITGFSFKFK K++ +TPKNSDS

Sbjct 181 IRTDNLKMRVIELALKQINDHTDITATYEQHKKGRITGFSFKFKQKRKTELETPKNSDS 240

Query 241 SPRIVKHSQIPTNIVKQPENAKMSDLEHRASRVGTGEIMRNRLSDRFKQGDSEAIMMKRI 300
K +IPTN+VKQPENA MSDL+HRAS++TG IM NRLSDRFKQ DES + MM RI

Sbjct 241 DVPKQKSVEIPTNLVKQPENANMSDLQHRASKITGLIMSNRLSDRFKQSDSIMQMMARI 300

Query 301 QSEIITDAIADQWESKLEEFVGVF 324
QSEI +AIA+QWE+KLEEFVGF

Sbjct 301 QSEITDEAIAQWENKLEEFVGF 324

ProtSweep

A Protein Identification Tool

Results for /home/vir088/rindneu/tsekompl/MSB11.176.324.pep

Number of residues: 324	Molecular weight: 37.29 kD
Charge: 1	Isoelectric point: 8.14
more...	
Prediction of protein localization:	
cyto	13 %
cyto nucl	10.5 %
extr	10 %

Identified Protein & Transcript:

Protein:

id: sptrembl:E5LG72 9ZZZZ

valuation: homologous

scientific organism: Tupaia belangeri

common organism: Common tree shrew

abstract: E5lg72 SubName: Full=Replicase 3-like protein;. 5/2013Length=324

isoform:

GO information:

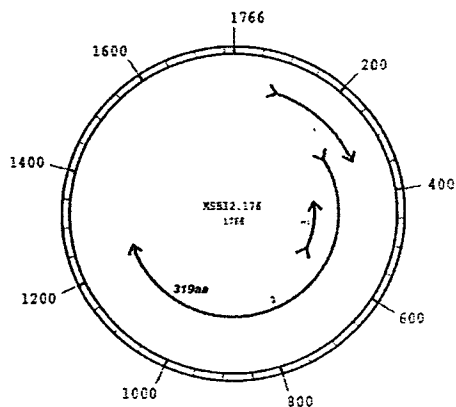
accession	group	name	evidence
GO:0005727	cellular component	extrachromosomal circular DNA	IEA:InterPro
GO:0003887	molecular function	DNA-directed DNA polymerase activity	IEA:InterPro
GO:0006270	biological process	DNA replication initiation	IEA:InterPro

Interpro information:

id	short name
IPR000525	Initiator_Rep_prot
IPR011991	WTH_DNA-bd_dom

MSB12.176 (1766bp)**(ms36.NnXn.1800.6)**

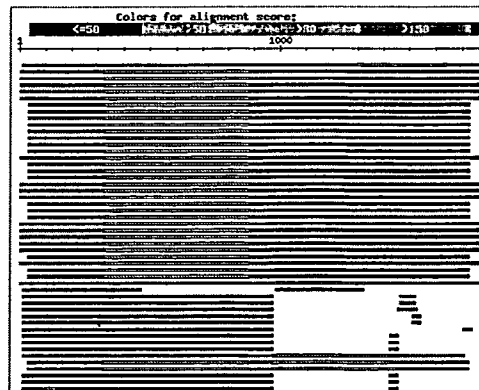
AAGCTTGCTTAGTCAAAAAAGTTTGAGCAAAGCGAAAACATAGGGCAATTTTCGTGATGAAAATGGGCTTTTAATGTTTTTAA
 ATGCTTTTAAATGCTTTTAGACATGCTGAAACGCTTATATAGCAATGCATACAGAGGACATATGTCCACGTTTACCTATCAAT
 ATGTCCACGTTTACCTATCAATATGTCCACGTTTACCTATCAATATGTCCACGTTTACCTTGCCTATTTACACACATTTAAAT
 AGTATGGATATATCCAATGAAATTATAATGTGTACTTATGAGCAAATTAGTAGTGAAAGACAATGCCCTAATGAACGCTAGTT
 ACAACTTGGATCTCGTTGAACAGCGTTTAATTTCTATTGGCAATCATCGAAGCAAGAGAATCAGGCAAAGGAATTAATGCAAAT
 GACCCGCTTACGGTTCATGCAGAGAGTTATATCAATCAATTTGGTGTTTCATCGAGTAACGTCATATCAAGCTCTCAAAGATGC
 TTGTGATAACTTGTGTTGCACGTCAATTCAGCTACCAATCCAAAAGTGAAAAGGGAACATACAAAATCATCGTTCACGTTGGG
 TTAGTGAAATTATTTACATTGATACAGAAGCAACAGTAAAAATAATTTGCACCTGCTATTGTCCCACTGATTACAAGGCTA
 GAAGAACAGTTTACCAAGTATGATATTGAGCAAATTAGTGATTTATCGAGTGCTTATGCAATTCGCTTATACGAGTTATTGAT
 TTGCTGGCGTAGCACAGGGAAAACACCAATTATTGGGCTAGGCGAATTTAGAAATCGGGTGGTGTGTTAGATAGTGAATATC
 ATCGAATTGCACACTTGAAAGAACGAGTTATTGAACATTCAATTAAACAGATTAACGAGCATACCGACATCACAGCCACCTAC
 GAACAGCACAAAAAGGGCGGACAATCACAGGATTTTCATTCAAGTTTAAGCAGAAGAAGCCCAACAAGCCGAAATTGCTAC
 AGAAACGCCAAAAACAGCCACGAATGACCCAGACACGACAAAACCCCTTACAGAGCCTCAGATCGCAAAATACAGCATGATTC
 TGTGCAAACTAGGCAGTATTTAGACTTGAGTAACCTCCAGACTATCCAGCTTTTGCAAAATTGGATTGGGAACATTTTGAGG
 AACCTGAAAAAGCAGATGAACAAATAGCAAAACGGATTTTCACAGCATTGAAAACAGAAACCGACTACAGCAAGAAAACTA
 ATTTTGTAGTTGTGATGGGTTTCCCGAAATAACATGAAGGGCGCACTTACGCAAAATTTTGTACGCCAAATTTTGCAAGTA
 CGGTCAGGGAAACCCGACACCCCAAGCAAAACACTGTAATCAGGGAAAAAACATTTTGTATTTGATCCTTGTTTGTCA
 CTCGTAGACACTCGTTTTGTGCTCTTTCTAGAATTACAAAAAGATATTAGCGAGTGCTACGAGCGACTCAATGAAAG
 TTCGATTATTCCCCCTCTGGAAAACCGCTTTTAAAAATATTGGCTGCTAGATGGTTTTTACTATAGTGAGGTTTTGCTTTTAA
 AAAACACGAGCAAAGCGAGTTCATAGTTGCTTTTGTGTTTTCGGGTCTTAGGGGAAATCCCCTAACAAAGTCCTCGAATAT
 CAAAATGTGGCTACATTTGTATATACGGGTAGGCTTGCTTATTGATTTTTTTCTTCTAAACCTTTGACTTCTTCCCATT
 GTTTCAGAAATGCCCCCGACC

Repeat – 3x 22nt + 18nt

78% nucleotide similarity in 984nt and 95-99% to 431nt of TSE-associated circular DNA isolate Sphinx 1.76

ORF: 319aa – initiation replication-B protein 65% amino acid similarity to Acinetobacter radioresistens

Figure 1(d)

Blastn – Geall:

```
>>>nrmuc:GI_462476960 Gi|462476960|gb|APGK01000458.1| Dendro... 1943 0.0
>>>nrmuc:GI_347602255 Gi|347602255|gb|JN541288.1| Uncultured... 1449 0.0
>>>nrmuc:GI_462476960 Gi|462476960|gb|APGK01000458.1| Dendroctonus ponderosae Seq01000458,
whole genome shotgun sequence. 0/0
Length=1455
Score = 1943 bits (2154), Expect = 0.0
Identities = 1202/1284 (93%), Gaps = 1/1284 (0%)
Strand=Plus/Minus
```

```
Query 11 AGTCAAAAAAGTTTGAGCAAAGCGAAAACATAGGGCAATTTTCGTGATGAAAATGGGCTT 70
Sbjct 1318 AGTCAAAAAAGTTTGAGAGAAGCGAAAACATAGGGCAATTT-CATGGTGAGAATGGGCTT 1260

Query 71 TTAATGTTTTTAAATGCTTTTAAATGCTTTTAGACATGCTGAAACGCTTATATAGCAATG 130
Sbjct 1259 TTAAGGTTTTTAAAGCTTTTAAATGCTTTTAGACATGCTGAAACGACCACACAGCAAGG 1200

Query 131 CATACAGAGGACATATGTCCACGTTTACCTATCAATATGTCCACGTTTACCTATCAATAT 190
Sbjct 1199 CATACAGAGGGCATATGTCCACGTTTACCTATCAATATGTCCACGTTTACCTATCAATAT 1140

Query 191 GTCCACGTTTACCTATCAATATGTCCACGTTTACCTTGCCTATTTACACACATTTAAATA 250
Sbjct 1139 GTCCACGTTTACCTATCAATATGTCCACGTTTACCTTGCATAGTCACACATATTTAAATA 1080

Query 251 GTATGGATATATCCAATGAAATTATAATGTGTACTTATGAGCAAATTAGTAGTGAAAGAC 310
Sbjct 1079 ATATGGATATATCCAATAAAATTATAATGTGTACTTATGAGTAGGTTAGTAGTGAAAGAC 1020

Query 311 AATGCCCTAATGAACGCTAGTTACAACCTGGATCTCGTTGAACAGCGTTTAAATCTATTG 370
Sbjct 1019 AATGCCCTAATGAACGCTAGTTATAACTGGATCTCGTTGAACAGCGTTTAAATCTATTG 960

Query 371 GCAATCATCGAAGCAAGAGAATCAGGCAAAGGAATTAATGCAAATGACCCGCTTACGGTT 430
Sbjct 959 GCAATCATCGAAGCGAGGGAGTCAGGCAAAGGAATTAACGCCAATGACCCCTCTACAATC 900

Query 431 CATGCAGAGAGTTATATCAATCAATTTGGTGTTCATCGAGTAACTGCATATCAAGCTCTC 490
Sbjct 899 CATGCAGAGAGCTACATCAATCAATTTGGTGTTCATCGAGTAACTGCATATCAAGCTCTC 840

Query 491 AAAGATGCTTGTGATAACTTGTTCACGTCGAATTCAGCTACCAATCCAAAAGTGAAAAA 550
Sbjct 839 AAAGATGCTTGTGATAACCTGTTGCACGTCGAATTCAGCTATCAGTCCAAAAGTGAAAAA 780

Query 551 GGAACATACAAAATCATCGTTTACGTTGGGTTAGTGAAATTATTACATTGATACAGAA 610
Sbjct 779 GGAACATACAAAATCATCGTTTACGTTGGGTTAGTGAAATTATTACATCGACACAGAA 720

Query 611 GCAACAGTAAAAATAATATTTGCACCTGCTATTGTCCACTGATTACAAGGCTACAAGAA 670
Sbjct 719 GCAACAGTAAAAATAATCTTTGCACCTGCTATCGTCCCACTGATTACAAGATTAGAAGAA 660
```

Query	1641	CCTAACAAAGTCCTCGAATATCAAAATGTGGCTACATTTTGTATATACGGGTAGGCTTGCT	1700
Sbjct	1455	CCTAACAAAGTCCCCGAATATCAAAATGTGGCTACATTTTGTATATACGGGTAGGCTTGCT	1396
Query	1701	TATTTGAttttttttC-TTCTAAACCTTT-GACTTCTTCCCCATTGTTTGCAGAAATGCC	1758
Sbjct	1395	TATTGAATCTAATTCCTTTTAAACCTTTCGACT--TTCCCTATCCTTTGCATAAATTGC	1338
Query	1759	CCCCGAC	1765
Sbjct	1337	CCCTGAC	1331

Query	881	ATTAACGAGCATACCGACATCAGGCCACCTACGAACAGCACAAAAAGGGCGGACAATC	940
Sbjct	1	ATCAACGAGCACACCGACATCAGGCCACCTACGAACAGCACAAAAAGGGCGGACAATC	60
Query	941	ACAGGATTTTCATTCAAGTTTAAGCAGAAGAAGCCCAAACAAGCCGAAATTGCTACAGAA	1000
Sbjct	61	ACAGGATTTTCATTCAAGTTTAAGCAGAAGAAGCCCAAACAAGCCGAAATTGCTACAGAA	120
Query	1001	ACGCCAAAAACAGCCACGAATGACCCAGACACGACAAAACCCCTTACAGAGCCTCAGATC	1060

Sbjct 121 |||||
ACGCCAAAAACAGCCACGAATGACCCAGACACGACAAAAACCCCTTACAGAGCCTCAGATC 180

Query 1061 GCAAAATACAGCATGATTCTGTGCAAACTAGGCAGTATTTTCAGACTTGAGTAACTTCCCA 1120
|||||

Sbjct 181 GCAAAATACAGCATGATTCTGTGCAAACTAGGCAGTATTTTCAGACTTGAGTAACTTCCCA 240
|||||

Query 1121 GACTATCCAGCTTTTGCAAATTGGATTGGGAACATTTTGAGGAACCCCTGAAAAAGCAGAT 1180
|||||

Sbjct 241 GACTATCCAGCTTTTGCAAATTGGATTGGGAACATTTTGAGGAACCCCTGAAAAAGEAGAT 300
|||||

Query 1181 GAACAAATAGCAAAACGGATTTTCACAGCATTGAAAACAGAAACCGACTACAGCAAGAAA 1240
|||||

Sbjct 301 GAACAAATAGCAAAACGGATTTTCACAGCATTGAAAACAGAAACCGACTACAGCAAGAAA 360
|||||

Query 1241 AACTAATTTTGTAGTTGTGATGGGTTTCCCGAAATAACATGAAGGGCGCACTTACGCA-- 1298
|||||

Sbjct 361 AACTAATTTTGTAGTTGTGATGGGTTTCCCGAAATAACATGAAGGGCGCACTTAGCCATT 420
|||||

Query 1299 -----AAATTTTGTCTACGCCAAATTTTGCAGTACGGTCAGGGAAACCCCGACACCCCA 1353
||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Sbjct 421 GAGAAAAATCTTCGATTTTTTCAAT---GGAAGTCCGTGGGGGTAAACCCCTCAACCCCA 477
|||||

Query 1354 AAAGCAAAACACTGTAATCAGGGAAAAAACATTTTGTATTTGATCCTTGTGTCACCT 1413
|||||

Sbjct 478 AAAGCAAAACACTGTAATCAGGGAAAAAACATTTTGTATTTGATCCTTGTGTCACCT 537
|||||

Query 1414 CGTAGACACTCGTTTGTGTTTGTCTCTTTCTAGAATTACAAAAAAGATATTAGCGAGTGT 1473
|||||

Sbjct 538 CGTAGACACTCGTTTGTGTTTGTCTCTTTCTAGAATTACAAAAAAGATATTAGCGAGTGT 597
|||||

Query 1474 CTACGAGCGACTCAATGAAAGTTCGATTATTCGCCCTCTGGAACCGCTTTTAAAAATA 1533
|||||

Sbjct 598 CTACGAGCGACTCAATGAAAGTTCGATTATTCGCCCTCTGGAACCGCTTTTAAAAATA 657
|||||

Query 1534 TTGGCTGCTAGATGGTTTTTACTATAGTGAGGTTTGTCTTTTaaaaaaCACGAGCAAAG 1593
|||||

Sbjct 658 TTGGCTGCTAGATGGTTTTTACTATAGTGAGGTTTGTCTTTTAAAAAACACGAGCAAAG 717
|||||

Query 1594 CGAGTTCATAGTTGCTTTTGTCTTTTTCGGGTCTTAGGGGAAATCCCTAACAGTCCT 1653
|||||

Sbjct 718 CGAGTTCATAGTTGCTTTTGTCTTTTTCGGGTCTTAGGGGAAATCCCTAACAGTCCT 777
|||||

Query 1654 CGAATATCAAAATGTGGCTACATTTTGTATATACGGGTAGGCTTGCTTATTTGATttttt 1713
|||||

Sbjct 778 CGAATATCAAAATGTGGCTACATTTTGTATATACGGGTAGGCTTGCTTATTTGATTTTTT 837
|||||

Query 1714 ttCTTCTAAACCTTTGACTTCTTCCCCATTGTTTGCAGAAA-TGCCCCCGAC 1765
|||||

Sbjct 838 TTCTTCTAAACCTTTGACTTCTTCCCCATTGTTTGCAGAAATGCCCCCTCGAC 890
|||||

Score = 576 bits (638), Expect = 3e-160
Identities = 322/324 (99%), Gaps = 0/324 (0%)
Strand=Plus/Plus

Query 1 AAGCTTGCTTAGTCAAAAAAGTTTGAGCAAAGCGAAACATAGGGCAATTTTCGTGATGA 60
|||||

Sbjct 892 AAGCTTGCTTAGTCAAAAAAGTTTGAGCAAAGCGAAACATAGGGCAATTTTCGTGATGA 951
|||||

Query 61 AAATGGGCTTTTAAATGTTTTTAAATGCTTTTAAATGCTTTTACACATGCTGAAACGCTTA 120
|||||

Sbjct 952 AAATGGGCTTTTAAATGTTTTTAAATGCTTTTAAATGCTTTTACACATGCTGAAACGCTTA 1011
|||||

Query 121 TATAGCAATGCATACAGAGGACATATGTCCACGTTTACCTATCAATATGTCCACGTTTAC 180
|||||

Sbjct 1012 TATAGCAATGCATACAGAGGACATATGTCCACGTTTACCTATCAATATGTCCACGTTTAC 1071
|||||

Query 181 CTATCAATATGTCCACGTTTACCTATCAATATGTCCACGTTTACCTTGCGTATTTACACA 240
|||||

Sbjct 1072 CTATCAATATGTCCACGTTTACCTATCAATATGTCCACGTTTACCTTGCGTATTTACACA 1131
|||||

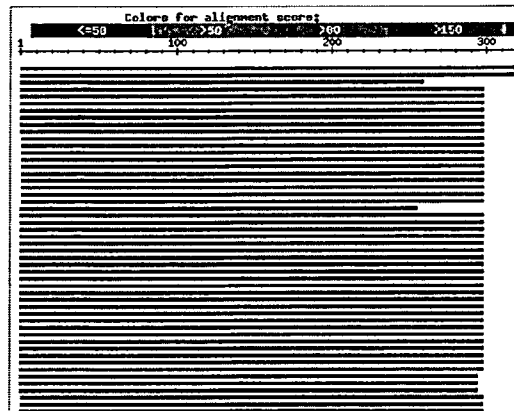
Query 241 CATTTAAATAGTATGGATATATCCAATGAAATTATAATGTGTACTTATGAGCAAATTAGT 300
|||||

Sbjct 1132 CATTTAAATAGTATGGATATATCCAATGAGATTATAATGTGTACTTATGAGCAAATTAGT 1191
|||||

Query 301 AGTGAAAGACAATGCCCTAATGAA 324
|||||

Sbjct 1192 AGTGAAAGACAATGCCCTAATCAA 1215
|||||

BlastP – Uniprot:
ORF 319aa:



```
>>>sptrembl:D0T8X0_ACIRA D0t8x0 SubName: Full=Initiator RepB... 435 1e-119
>>>sptrembl:C6RK77_ACIRA C6rk77 SubName: Full=Initiator RepB... 435 1e-119

>>>>sptrembl:D0T8X0_ACIRA D0t8x0 SubName: Full=Initiator RepB protein;. 9/2013
Length=313
Score = 435 bits (1118), Expect = 1e-119, Method: Compositional matrix adjust.
Identities = 208/319 (65%), Positives = 252/319 (78%), Gaps = 7/319 (2%)
```

```
Query 1 MSKLVVKDNALMNASYNLDLVEQRLILLAIIEARESGKGINANDPLTVHAESYINQFGVH 60
M +LVVKDNAL+NASYNLDLVEQRLILLAI+EAARESGKGINANDPL VHAESYINQF VH
Sbjct 1 MRELVVKDNALINASYNLDLVEQRLILLAIIVEARESGKGINANDPLVVHAESYINQFNVH 60

Query 61 RVTAYQALKDACDNLFARQFSYQSKSEKGNIQNHRSRWVSEIIYIDTEATVKIIFAPAIV 120
R TAYQALKDAC +LF RQFSYQ +++GN+++RSRWVSEI Y+D EA VK+IFAPAIV
Sbjct 61 RNTAYQALKDACHDLFVRQFSYQKVNDRGNVEHYRSRWVSEIGYVDNEAIVKLIFAPAIV 120

Query 121 PLITRLEEQFTKYDIEQISDLSSAYAIRLYELLICWRSTGKTPPIIGLGEFNRNVGVLDSE 180
PLITRLEEQFTKY+++Q+S L+SAYA+RLYELLI WRSTGKTP+I L +FR R+G+L++E
Sbjct 121 PLITRLEEQFTKYELQVSQLTSAYAVRLYELLIARSTGKTPVIELADFRKRIGILETE 180

Query 181 YHRIAHLKERVIEHSIKQINEHTDITATYEQHKKGRITITGFSFKFKQKKPKQAEIATETP 240
Y R+ K V+E +I QINEHTDI YEQHKKGR+I GFSF F QK+ K+ E
Sbjct 181 YKRMERFKTSVLELAINQINEHTDINVAYEQHKKGRSIVGFSFNFSQKEKKK---ILEKA 237

Query 241 KTATNDPDTTKPLTEPQIAKYSMILCKLGSISDLSNFPDYPAFANWIGNILRNPEKADEQ 300
+ + + K LTE QI KYS +L KL +SDLS F DY +F+ WIGNILR PE +
Sbjct 238 QVS----EGFKKLEAQITKYSTVLSKLHELSDLSTFDYQSFISIWIGNILREPESVREF 293

Query 301 IAKRIFTALKTETDYSKKN 319
A+RIF++L TD++ N
Sbjct 294 TAERIFSSLFKRTDFASPN 312
```

```
>>>>sptrembl:C6RK77_ACIRA C6rk77 SubName: Full=Initiator RepB protein;. 9/2013
Length=313
Score = 435 bits (1118), Expect = 1e-119, Method: Compositional matrix adjust.
Identities = 208/319 (65%), Positives = 252/319 (78%), Gaps = 7/319 (2%)
Query 1 MSKLVVKDNALMNASYNLDLVEQRLILLAIIEARESGKGINANDPLTVHAESYINQFGVH 60
M +LVVKDNAL+NASYNLDLVEQRLILLAI+EAARESGKGINANDPL VHAESYINQF VH
Sbjct 1 MRELVVKDNALINASYNLDLVEQRLILLAIIVEARESGKGINANDPLVVHAESYINQFNVH 60

Query 61 RVTAYQALKDACDNLFARQFSYQSKSEKGNIQNHRSRWVSEIIYIDTEATVKIIFAPAIV 120
R TAYQALKDAC +LF RQFSYQ +++GN+++RSRWVSEI Y+D EA VK+IFAPAIV
Sbjct 61 RNTAYQALKDACHDLFVRQFSYQKVNDRGNVEHYRSRWVSEIGYVDNEAIVKLIFAPAIV 120

Query 121 PLITRLEEQFTKYDIEQISDLSSAYAIRLYELLICWRSTGKTPPIIGLGEFNRNVGVLDSE 180
PLITRLEEQFTKY+++Q+S L+SAYA+RLYELLI WRSTGKTP+I L +FR R+G+L++E
Sbjct 121 PLITRLEEQFTKYELQVSQLTSAYAVRLYELLIARSTGKTPVIELADFRKRIGILETE 180

Query 181 YHRIAHLKERVIEHSIKQINEHTDITATYEQHKKGRITITGFSFKFKQKKPKQAEIATETP 240
```

```

Sbjct 181 Y R+ K V+E +I QINEHTDI YEQHKKGRI GFSF F QK+ K+ E
YKRMERFKTSVLELAINQINEHTDINVAYEQHKKGSI VGFNFNQKEKKK---ILEKA 237

Query 241 KTATNDPDTTKPLTEPQIAKYSMILCKLGSISDLNFPDYPAFANWIGNILRNPEKADEQ 300
+ + + K LTE QI KYS +L KL +SDLS F DY +F+ WIGNILR PE +
Sbjct 238 QVS----EGFKKLTEAQITKYSTVLSKLHELSDLSTFQDYQSFSIWIGNILREPESVRFE 293

Query 301 IAKRIFTALKTETDYSKKN 319
A+RIF++L TD++ N
Sbjct 294 TAERIFSSLFKRTDFASPN 312

```

ProtSweep

A Protein Identification Tool

Results for /home/vir088/rindneu/tsekompl/MSBI2.176.319.pep

Number of residues: 319	Molecular weight: 36.37 kD
Charge: 4	Isoelectric point: 9.07

more...

Prediction of protein localization:

cyto	18 %
extr	5 %
nucl	5 %

Identified Protein & Transcript:

Protein:

id: sptrembl:C6RK77 ACIRA

valuation: similar

scientific organism: Acinetobacter radioresistens

abstract: C6rk77 SubName: Full=Initiator RepB protein;. 9/2013Length=313

isoform:

GO information:

accession	group	name	evidence
GO:0005727	cellular component	extrachromosomal circular DNA	IEA:InterPro
GO:0003887	molecular function	DNA-directed DNA polymerase activity	IEA:InterPro
GO:0006270	biological process	DNA replication initiation	IEA:InterPro

Interpro information:

id	short name
IPR000525	Initiator_Rep_prot
IPR011991	WTH_DNA-bd_dom

*CM11.252 (2523bp)**(CMI = Cow milk isolate)**(mil.NoXo.2500.7)***IDENTICAL TO HCBI6.252**

TTAAAAATATTGGTTGCTAGATGGTTTTAACTATCGTGGAGCTTTTGCTTTTAAAAATACACGAAGCAAAGCGAGTTCATAGT
 CGCTTTTGATCTGCTCTTACTAACAGCTCATTTTTAATGGGCCGTTAGTAAGAGATGCTTTTGGGGTTTCGGGGCTTGCCCTG
 AACAAAGTCAAAAAGTGGGAATTTATCACGGTAGTGAAAAAGTACCCTTTGTGTATCCTTGCTTTATGCTTTTAAACCTTTG
 GACTTTTCCCTTAATTTGCAGAAATTGCCCTCGACTAAGCAAGCTTAGTCAAAAAGTTTGAGCAAAGCGAAAACATAGGGC
 AATTTTCATGATGAAAATGGGCTTTTAAAGTTTTTAAATGCTTTTAAATGCTTTTAGACATGTTGAAATGCAAGCCTAGCAAG
 GCATACAGAGGACATACCCCTACGTTTACCGATCAATACCCCTACGTTTACCGATCAATACCCCTACGTTTACCGATCAATAC
 CCCTACGTTTACCTTGCATATCACTACAAACAATATTAATGTAGTAATAAATTCAAAATAATAAATGTAGGTTATGAGCGATT
 TAATAGTAAAGGATAACGCCCTAATGAACGCTAGTTACAACCTTAGCTTTGGTTGAACAGCGTTAATCCTTTTAGCTATATTG
 GAAGCGAGAGAAAACAGGAAAAGGGATTAATGCCAATGACCCCTTAACAGTTCATGCAAGTAGCTATATCAATCAATTTAACGT
 GGAAAGACATACGGCATATCAAGCATTAAAAGATGCTTGCAAGACTTGTTTGCACGTCAATTTAGTTACCAAGAAAAGCGAG
 AACGAGGACGAATTAATATTACAAGTCGATGGGTTTCACAAATTGGTTATATGGACGATACAGCAACCGTTGAAATTATTTTT
 GCCCTGCGGTTGTTCCCTTGATTACAGCTTTAGAGGAACAATTCATCAATACGATATTGAACAAATTAGCGAGTTATCGAG
 TGCTTATGCTGTCCGCTATATGAGTTACTAATTTGTTGGCGTAGCACAGGAAAACCCCAATTATTGATCTAACAGAGTTTA
 GAAAGCGTCTAGGCGTTTATAGATACCGAATACACTAGGACAGATAAATTGAAGATGCGAGTGATTGAGCTAGGACTTAAACAA
 ATCAACGAGCATACGGACATTACAGCAAGCTACGAACAGCACAAAAAGGGCGGACAATCACAGGATTCTCATTCAAGTTTAA
 GCAGAAGAAGAAAACAGGGGCGGAAATGCCAAAAATAGCGATTCTAGCCACATATCGAAAAACCTAGTCAAATCCCTGCCA
 ACATCGCAAAACAGCCTGAAAATGCCAAAAAGATGATTTAGGGCACAGAGCAAGCAAGATTACAGGGTTAATTATGTCGAAT
 GGTTTAGCAGATCGGTTTAAACGTGGTGATGAGTCAGTAATAGACATGATGAAACGCATCNGAGGAAATTACAACGATACAC
 ACAGCAGACCAATGGGAAAACAAGCTAGAAGAATTTGGCGTGATTTTTCAATCATGATCAATTTCCCGAAGGACATATAAAAA
 ATAATGTAGTTAAATATAGTAGTTATAATATTGACTACAATATTTAACTACTATATTGTAGTCTTAAATCAACTTACATAGAA
 TTTAAATTATGTCTCTTCGCCAAATTTTAGAACATCTCTCTTGATCAAACCTGATGAAAGCTCTACTGTGCCACTATCAATT
 AGACTTCCAAACAACCTGAATAATGAGCTTGAAGAAGTACTATAACACTAGATAAATCCAAGAGTTTTCTACTCCTAGAGTT
 CATTAAAGGCTGGAATTAGTGAGACTAGCAATATGTTAGAAGACAAAGCGAAGAACCCTATACTTGAAAGTCCAAGAGATTCTA
 GTGACTTTACAACCTATAAGAAGTTTATGTTGAACACTAATTTCAATAACAATAAAGCATCACACATGACTATGTTGGAGAAT
 CAAGAAGCTGCTGCATTTTCGTAAGGTTGGAAGAATATATCTGTAATTTAGAAGAAGGCGATAAGGTTTACCTTTTATCAAAG
 TGGCGTTGTTTTGTAGCATCTGGTATTGTCACGTGACTTAATCAAAAGTGAGTATCGTGGACAACCAGAAGAAAAATATT
 CTAAAAAATTAAAGGATTTTAAAAATAGGATTTAAAGCAATTTCTGCAAAAGAGTTTAAAGACCTTAACCTGGCGGTGGTGCTAGC
 TTTCTGTAACAATGAGCGAGCTAAAGCCTGAACAATGTAGAAAATTAGATATGGAATAGAAAACGGATGAAGTCAGTAAG
 TTAAAAATATGAGTTGTTTGTAGTGTGTACAGTAGCAATCACTTAGAAGTTCCCTTTGTATTACTTGTCTAGCAATGGTTT
 TAAAAAGCGTCACTACTGATAGTGACGCTTTTTTTGTTTTTCAAAGGAAGTAATTAGCGAGTGTCTACGAGCGACACAATGAA
 AGTTTCGATTATTCCTCTCTAAAAAACCGCTT

Repeat – 3x22nt + 17nt

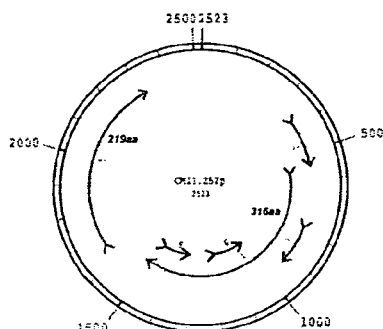
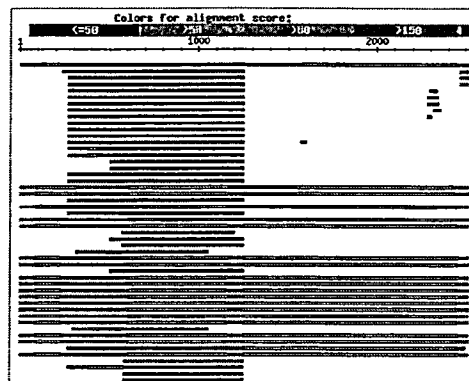
85-90% nucleotide similarity to TSE-associated circular DNA isolate *Sphinx 1.76*ORFs: 316aa – replication protein 85% similarity to *Sphinx1.76*

Figure 1(e)

*219aa – regulation of transcription**Blastn – Geall:*

```
>>>nrmuc:GI_313507477 Gi|313507477|gb|HQ444404.1| TSE-associ... 1519 0.0
>>>nrmuc:GI_462476960 Gi|462476960|gb|APGK01000458.1| Dendro... 733 0.0
>>>nrmuc:GI_586961986 Gi|586961986|gb|CP006964.1| Acinetobac... 489 4e-134

>>>>nrmuc:GI_313507477 Gi|313507477|gb|HQ444404.1| TSE-associated circular DNA isolate
Sphinx 1.76, complete sequence. 0/0
Length=1758
Score = 1519 bits (1684), Expect = 0.0
Identities = 1144/1334 (85%), Gaps = 14/1334 (1%)
Strand=Plus/Plus
Query 237 TTTTAAACCTTTGGACTTTT--CCCTTAATTTGCAGAAATTGCCCTCGACTAAGCAAGC 294
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 277 TTCTAAACCTTTG-ACCTCTTCCCATTTGTTGCAGAAATTGCCCTCGA---AGCTTGC 332

Query 295 TTAGTCAAAAAAGTTTGAGCAAAGCGAAAACATAGGGCAATTTTCATGATGAAAATGGGC 354
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 333 TTAGTCAAAAAAGTTTGAGCAAAGCGAAAACATAGGGCAATTTTCATGAAGAAATGGGC 392

Query 355 TTTTAAAGTTTAAATGCTTTTAAATGCTTTTAGACATGTTGAAATGCAAGCCTAGCAA 414
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 393 TTTTAAAGTTTAAATG-TTTTAAATGCTTTTAGACATGCTAAGAAGCCACACAGCAA 451

Query 415 GGCATACAGAGGACATACCCCTACGTTTACCGATCAATACCCCTACGTTTACCGATCAAT 474
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 452 GGCATACAGAGGACATTCTCTACGTTTACCGATCAATACCCC--CGTTTACCGATCAAT 509

Query 475 ACCCCTACGTTTACCGATCAATACCCCTACGTTTACCTTGCATATCACTACAAACAATAT 534
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 510 ACCCCTACGTTTACCGATCAATACCCCTACGTTTACCTTGCATATAACTACAAAGAATAC 569

Query 535 TAATGTAGTAATAACTTCAAATAATAAATGTAGGTTATGAGCGATTAAATAGTAAAGGA 594
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 570 TAGTGTAGTA-TAACTTCAAAGAATAAATGTAGGTTATGAGCGATTAAATAGTAAAGGA 628

Query 595 TAACGCCCTAATGAACGCTAGTTACAACCTAGCTTTGGTTGAACAGCGTTAATCCTTTT 654
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 629 TAACGCCCTAATGAATGCTAGTTATAACTTAGCTTTGGTTGAACAGAGGTTAATTCTATT 688

Query 655 AGCTATATTGGAAGCGAGAGAAACAGGAAAAGGGATTAATGCCAATGACCCCTTAACAGT 714
      ||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 689 AGCAATCATAGAAGCGAGAGAAACAGGCAAAGGGATTAATGCCAATGATCCTCTTACGGT 748

Query 715 TCATGCAAGTAGCTATATCAATCAATTTAACGTGGAAGACATACGGCATATCAAGCATT 774
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 749 TCATGCAGGTAGCTATATCAATCAATTTAACGTACAAAGGCATACGGCATATCAAGCCCT 808

Query 775 AAAAGATGCTTGCAAAGACTTGTTTGACGTCATTTAGTTACCAAGAAAAGCGAGAACG 834
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 809 CAAAGATGCTTGTAAGACTTGTTTGCCCGTCAATTCAGTTACCAAGAAAAGCGAGAACG 868
```

Query 835 AGGACGAATTAATATTACAAGTCGATGGGTTTCACAAATTGGTTATATGGACGATACAGC 894
|||||
Sbjct 869 AGGACGAATTAATATTACAAGTCGATGGGTTTCGCAAATTGGCTATATGGACGATACAGC 928
Query 895 AACCGTTGAAATTATTTTGGCCCTGCGGTTGTTCTTTGATTACACGTTTAGAGGAACA 954
|||||
Sbjct 929 AACCGTTGAGATTATTTTGGCCCTGCGGTTGTTCTCTGATTACACGGCTAGAGGAACA 988
Query 955 ATTCACCTCAATACGATATTGAACAAATTAGCGAGTTATCGAGTGCTTATGCTGTCCGCCT 1014
|||||
Sbjct 989 GTTACCCAGTACGATATTGAGCAAATTAGCGGTTTATCGAGTGCATATGCTGTTCTGAT 1048
Query 1015 ATATGAGTTACTAATTGTTGGCGTAGCACAGGAAAAACCCCAATTATTGATCTAACAGA 1074
|||
Sbjct 1049 GTACGAACTGCTGATTGTTGGCGTAGCACAGGAAAACACCAATTATTGAGCTAGACGA 1108
Query 1075 GTTTAGAAAGCGCTCTAGGCGTTTGTAGATACCGAATACACTAGGACAGATAACTTGAAGAT 1134
|||||
Sbjct 1109 GTTTAGAAAGCGAATAGGTGTTTGTAGATACTGAATACACTAGAACAGATAATTTAAAGAT 1168
Query 1135 GCGAGTGATTGAGCTAGGACTTAAACAAATCAACGAGCATACGACATTACAGCAAGCTA 1194
|||
Sbjct 1169 GCAAGTTATTGAATTAGCCCTAAAACAAATCAACGAACATACTGACATCAGCAAGCTA 1228
Query 1195 CGAACAGCACAAAAAGGGCGGACAATCAGAGATTCTCATTCAAGTTAAGCAGAAG-A 1253
|||||
Sbjct 1229 TGAACAAACAAAAAGGGCGAGTGATTACCGGATTCTCATTGTTAAGCACAAGAA 1288
Query 1254 AGAAAAACAGGGCGGAAATGCCAAAAATAGCGATTCTAGCCACATATCGAAAAACCTA 1313
|||
Sbjct 1289 ACAAACA-GCGATAAACGCGTACTAAGCTTCTAGCCACGTATCGTAAACATA 1347
Query 1314 GTCAATCCCTGCCAACATCGCAAAACAGCCTGAAAATGCCAAAAAGATGATTTAGGGC 1373
|||||
Sbjct 1348 GTCAATCCCTACCAACATTGTAAACAGCCTGAAAACGCCAAATGAGCGATTTAGAAC 1407
Query 1374 ACAGAGCAAGCAAGATTACAGGGTTAATTATGTCGAATGGTTTAGCAGATCGGTTTAAAC 1433
|||
Sbjct 1408 ATAGAGCGAGCCGTGTACAGGGGAAATAATGCGAAATCGTCTGTGATCGGTTTAAAC 1467
Query 1434 GTGGTGATGAGTCAGTAATAGACATGATGAAACGCATCNNGAG-GAAATTACAACTGATA 1492
|||
Sbjct 1468 AAGGCGATGAATCAGCAATCGACATGATGAAACGTATTCAAAGTGAAATAATAACCGATG 1527
Query 1493 CAACAGCAGACCAATGGGAAAACAAGCTAGAAGAATTTGGCGTGATTTTTCAGTCATGAT 1552
|||
Sbjct 1528 CAATAGCAGACCAAGTGGGAAAGCAAACCTGGAGGAGTTTGGCGTGGTTTTTTAGTCATGA- 1586
Query 1553 CAATTTCCCGAAGG 1566
|||
Sbjct 1587 CGATTTCCCGAAGG 1600

Score = 122 bits (134), Expect = 2e-23
Identities = 83/91 (91%), Gaps = 2/91 (2%)
Strand=Plus/Plus

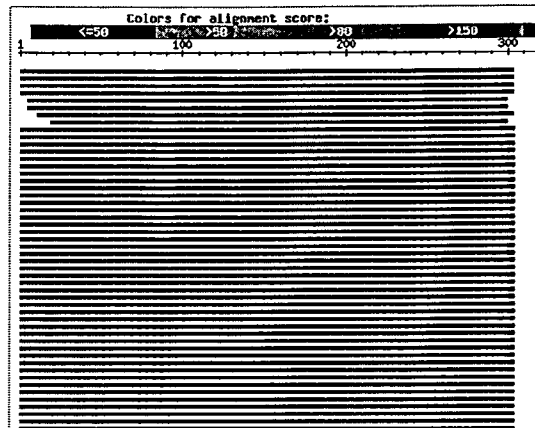
Query 1 TTAAAAATATTGGTTGCTAGATGGTTTTAACTATCGTGGAGCTTTTGCTTTTAAAAATAC 60
|||||
Sbjct 85 TTAAAAATATTGGCTGCTAGATGGTTTTTACTATAGT-GAGGTTTGTCTTTAAAAAAAC 143
Query 61 ACGAAGCAAAGCGAGTTCATAGTCGCTTTTG 91
|||
Sbjct 144 ACG-AGCAAAGCGAGTTCATAGTTGCTTTTG 173

Score = 95.1 bits (104), Expect = 3e-15
Identities = 60/64 (93%), Gaps = 1/64 (1%)
Strand=Plus/Plus

Query 2460 ATTAGCGAGTGTCTACGAGCGACACAATGAAAGTTTCGATTATTCCTCTCTAAAAAAC 2519
|||||
Sbjct 22 ATTAGCGAGTGTCTACGAGCGACTCAATGAAAGTTTCGATTATTCCTCTCTGAAAAAC 80
Query 2520 GCTT 2523
|||
Sbjct 81 GCTT 84

>>>>nrruc:GI_462476960 Gi|462476960|gb|APGK01000458.1| Dendroctonus ponderosae Seq01000458,

whole genome shotgun sequence. 0/0
Length=1455
Score = 733 bits (812), Expect = 0.0
Identities = 790/1033 (76%), Gaps = 21/1033 (2%)
Strand=Plus/Minus

ORF 316aa:**BlastP – Uniprot:**

```
>>>sptrembl:E5LG72_9ZZZZ E5lg72 SubName: Full=Replicase 3-li... 545 1e-152
>>>sptrembl:N9LWE7_9GAMM N9lwe7 SubName: Full=Uncharacterize... 523 4e-146

>>>>sptrembl:E5LG72_9ZZZZ E5lg72 SubName: Full=Replicase 3-like protein;. 5/2013
Length=324
Score = 545 bits (1403), Expect = 1e-152, Method: Compositional matrix adjust.
Identities = 261/304 (85%), Positives = 276/304 (90%), Gaps = 0/304 (0%)

Query 1 MSDLIVKDNALMNASYNLALVEQRLILLAILAEARETGKGINANDPLTVHASSYINQFNVE 60
      MSDLIVKDNALMNASYNLALVEQRLILLAI+EARETGKGINANDPLTVHA SYINQFNV+
Sbjct 1 MSDLIVKDNALMNASYNLALVEQRLILLAILAEARETGKGINANDPLTVHAGSYINQFNVQ 60

Query 61 RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSIQGYMDDTATVEIIFAPAVVP 120
      RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSIQGYMDDTATVEIIFAPAVVP
Sbjct 61 RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSIQGYMDDTATVEIIFAPAVVP 120

Query 121 LITRLEEQFTQYDIEQISELSSAYAVRLYELLICWRSTGKTPIIDLTEFRKRLGVLDTEY 180
      LITRLEEQFTQYDIEQIS LSSAYAVR+YELLICWRSTGKTPII+L EFRKR+GVLDTEY
Sbjct 121 LITRLEEQFTQYDIEQISGLSSAYAVRMYELLICWRSTGKTPIIELDEFKRKRGVLDTEY 180

Query 181 TRTDNLKMRVIELGLKQINEHTDITASIEQHKKGRITGFSFKFKQKKKTGAEMPKNSDS 240
      TRTDNLKM+VIEL LKQINEHTDITASIEQHKKGR ITGFSF FK KK+ + P + S
Sbjct 181 TRTDNLKMVIELALKQINEHTDITASIEQHKKGRVITGFSFMFKHKKQNSDKTPDTNAS 240

Query 241 SPHIEKPSQIPANIAKQPENAKKDDLGHRAKITGLIMSNGLADRFKRGDESVIDMMKRI 300
      SP I K SQIP NI KQPENAK DL HRAS++TG IM N L+DRFK+GDES IDMMKRI
Sbjct 241 SPRIVKHSQIPTNIVKQPENAKMSDLEHRASRVTEIMRNRLSDRFKQGDESAIDMMKRI 300

Query 301 XRKL 304
      ++
Sbjct 301 QSEI 304
```

```
>>>>sptrembl:N9LWE7_9GAMM N9lwe7 SubName: Full=Uncharacterized protein;. 12/2013
Length=324
Score = 523 bits (1347), Expect = 4e-146, Method: Compositional matrix adjust.
Identities = 250/304 (82%), Positives = 272/304 (89%), Gaps = 0/304 (0%)

Query 1 MSDLIVKDNALMNASYNLALVEQRLILLAILAEARETGKGINANDPLTVHASSYINQFNVE 60
      MS+LIVKDNALMNASYNL LVEQRLILLAI+EARE+GKGINANDPLTVHA SYINQF V
Sbjct 1 MSELIVKDNALMNASYNLALVEQRLILLAILAEARESGKGINANDPLTVHAESYINQFGVH 60

Query 61 RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSIQGYMDDTATVEIIFAPAVVP 120
      R TAYQALKDACKDLFARQFSYQEKREGR NITSRWVSIQI Y+D+TATVEIIFAPAVVP
Sbjct 61 RTTAYQALKDACKDLFARQFSYQEKREGRANITSRWVSIQIAYIDETATVEIIFAPAVVP 120
```

```
Query 121 LITRLEEQFTQYDIEQISELSSAYAVRLYELLCWRSTGKTPIIDLTEFRKRLGVLDTEY 180
          LITRLE+QFTQYDIEQIS LSSAYAVR+YELLCWRSTGKTP+I+L EFRKR+GVLDTEY
Sbjct 121 LITRLEKQFTQYDIEQISGLSSAYAVRMYELLCWRSTGKTPVIELGEFRKRIGVLDTEY 180

Query 181 TRTDNLKMRVIELGLKQINEHTDITASYEQHKKGRTITGFSFKFKQKKKTGAEMPKNSDS 240
          RTDNLKMRVIEL LKQIN+HTDITA+YEQHKKGRTITGFSFKFKQK+KT E PKNSDS
Sbjct 181 IRTDNLKMRVIELALKQINDHTDITATYEQHKKGRTITGFSFKFKQKRKTELETFKNSDS 240

Query 241 SPHIEKPSQIPANIAKQPENAKKDDLGHRAKITGLIMSNGLADRFKRGDESVIDMMKRI 300
          +K +IP-N+ KQPENA DL HRASKITGLIMSN L+DRFK+ DES++ MM RI
Sbjct 241 DVPKQKSVEIPTNLVKQPENANMSDLQHRASKITGLIMSNRLSDRFKQSDSIMQMARI 300

Query 301 XRKL 304
          ++
Sbjct 301 QSEI 304
```

>>>>sptrembl:N9JLM2_ACIBA N9jlm2 SubName: Full=Uncharacterized protein;. 12/2013
Length=325

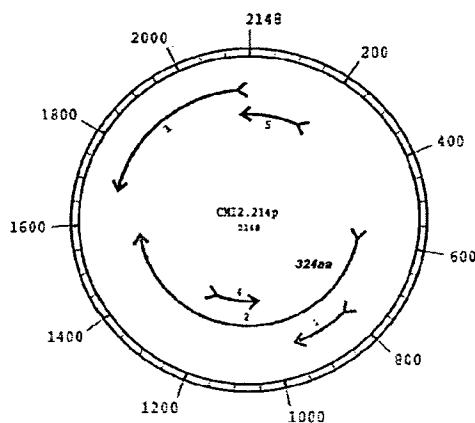
Score = 481 bits (1239), Expect = 1e-133, Method: Compositional matrix adjust.
Identities = 239/305 (78%), Positives = 270/305 (88%), Gaps = 1/305 (0%)

CM12.214 (2148bp)

(*mi2.NnXn.2000.8*)

TTGTACTGAATTAATGAACGAGTGTCTACGAGTGGCATAAAGAAAGTTCGCTTATTGCCCTCTGAAAAACCGCTTTTCTAT
ATCTAATTTTCATGAAAATTTAACTACTGTGAGCTTTTGCTTTTGGAAAGACACGAGCGAAGCGAGTGCCATTTTGTCTCTT
GCTTTTGGGGTTTGGGGCTTGCCCCAACAGGGCACAGAGTAGGAATTTTACACGGTAGTGAAAAAGTCCCCTCTGTGTCTCT
TGCCCTATGCTTTTAAACCTTTGAATTTTCCCTGTAATTTGAAGAAATTGCCCTCGACTAAGCTTGCTTAGTCAAAAAAG
TTTGAGCAAAGCGAAAACATAGGGCAATTTTCATAATAAAAAATGGGCTTTTAAAGTTTTTAAATGCTTTTAAATGCTTTTGA
CATGCTGAAATGCCTACACAGCAAGGCATACAGAGGACTACTCCTAGGTTTACCCACCAATACTCCTAGGTTTACCTACCA
TACTCCTAGGTT TACCTACCAATACTCCTAGGTTTACCTTTGCGTAATCCTAAAAGATATAATACAAAGGAACACTAGGTTT
AATATTAATAAGTAAATGAGCGATTTAATAGTAAAGATAATGCCTTAATGAACGCTAGTTATAACCTAGACTTAGTTGAGC
AGCGTTTAATCTCTTGCGTATCCTTGAGGCGAGAGAAACAGGCAAAGGGATTAATGCCAATGATCCTTAACCGTTCATGCT
GAAAGCTATATCAATCAATTTGGTGTAGCTAGACAGACTGCATATCAAGCATTAAAGATGCTTGTAAGGACTTATTTGCACG
CCAATTTAGTTTACCAAGAAAAACGTGAGCGTGGGCGAGCTAATATTACAAGTCGTTGGGTATCGCAAATGCTTACATTGATG
AAACGGCAACAGTTGAGGTTATTTTTGCCCTGCCGTGTTCCACTGATTACACGATTAGAGGAACAATTCACCAATACGAT
ATTGAACAAATTAGCGGTTTATCGAGTGCCTATGCCGTCCGTCTATATGAATTATTGATTTGTTGGCGTAGTACAGGGAAAAAC
ACCTGTTATTGAAC TAGCAGAGTTTAGAAAGCGTTTAGGTGTTTAAATGATGAATATACTAGATCGGATAATTTCAAAAAAT
GGATTATTGAAAACCTATTAAACAAATAAACGAGCATACCGACATCACAGCAAGCTACGAACAGCACAAAAAAGGGCGGACA
ATCACAGGTTTTTCATTCAAGTTTAAAGCAGAAGAAAAAACAGAACCTGAAACGCCAAAAAACAGCGATTCTAGCCAACGTAT
CGAAAAGCCTAGTCAATCCCTGCCAACATCGTAAAAACAACCTGAAAACGCAAATTTGAGCGATTTACAGCATCGAGCAAGCA
AGATTACAGGGTTAATCATGTCTAACCGTCTATCAGATCGGTTTAAACAAGGTGATGAATCAATCATGCAGATGATGGCACGT
ATTCAAAGTGAGATTACAACCGATTCAATAGCCGACCAATGGCAATCAAATTAGAAGAATTTGGCGTAGTTTTTTAGTCTGT
CACGTAGCAATCACGTAGAAAATTATGCTTGAAAGGCTTGCTGTTATTACATTTCAAAAAGCGTCACACTGATAGTGACGCT
TTTCTTAGTTTGCCTGTGAATTAACAGTGAGTGATAGACCTCTTTGGTTAGGACTTCATTAACCTTTTCATGCCATTCTT
TAATTATTTTTGTCGTTTCTGCATTTTCACTAGAACAAATACTAATGTTTTTCTTGGCTGAGGTAGTGTTCCTCCCAATG
TCCCAATCCAATACAGAATAGGGCGTCTGAAAACAATCCAATCGGTTTCACTTTCATCACCATTGACTTCATAAGTCCACC
AATTACCTGAACATAATGCAATTAATTTGCCACCATTTTAAATAGCAACAATATCATTCACTTTTATTCTGTTAAAAAGTCCCT
CTTTTTCAACCTGTCCTCCAGCCCATTTTCTAAACCTATAATTTTTTGTCTCTAAAATTGAATGTAAACTTCCGAAAAG
CTTTGATCATCAGGGTGCATCTGCATATGCCAATATTGAGTCATGTTTTTTTTCCATAGGTAATTATGTTTGT

Repeat – 3x22nt + 17nt



79% nucleotide similarity to TSE-associated circular DNA isolate *Sphinx* 1.76

ORF: 324aa – replication protein 86% similarity to *Sphinx*1.76

Figure 1(f)

Query	1188	ATAAACGAGCATACCGACATCACAGCAAGCTACGAACAGCACAAAAAGGGCGGACAATC	1247
Sbjct	1	ATAAACGAGTACACCGACATCATAGCAAGCTACGAACAGCACAAAAAGGGCGGACAATC	60
Query	1248	ACAGGGTTTTTCATTCAAGTTTAAAGCAGAAGaaaaaaCAGAACCTGAAACGCCAAAAAAC	1307
Sbjct	61	ACAGGGTTTTTCATTCAAGTTTAAAGCAGAAGAAAAAACAGAACCTGAAACGCCAAAAAAT	120
Query	1308	AGCGATTCTAGCCAACGTATCGAAAAGCCTAGTCAAATCCCTGCCAACATCGTAAACAA	1367
Sbjct	121	AGCGATTCTAGCCAACGTATCGAAAAGCCTAGTCAAATCCCTGCCAACATCGTAAACAA	180
Query	1368	CCTGAAAACGCAAATTTGAGCGATTTACAGCATCGAGCAAGCAAGATTACAGGGTTAATC	1427
Sbjct	181	CCTGAAAACGCAAATTTGAGCGATTTACAGCATCGAGCAAGCAAGATTACAGGGTTAATC	240
Query	1428	ATGTCTAACCGTCTATCAGATCGGTTTAAACAAGGTGATGAATCAATCATGCAGATGATG	1487
Sbjct	241	ATGTCTAACCGTCTATCAGATCGGTTTAAACAAGGTGATGAATCAATCATGCAGATGATG	300
Query	1488	GCACGTATTCAAAGTGAGATTACAACCGATTCAATAGCCGACCAATGGCAATCAAATTA	1547
Sbjct	301	GCACGTATTCAAAGTGAGATTACAACCGATTCAATAGCCGACCAATGGCAATCAAATTA	360
Query	1548	GAAGAATTTGGCGTAGTTTTTTAGTCTGTACGTAGCAATCAGTAGAAAATTATGCTTG	1607
Sbjct	361	GAAGAATTTGGCGTAGTTTTTTAGTCTGTACGTGGCAATCAGTAGAAAATTATGTTTG	420
Query	1608	AAAGGCTTGCTGTTATTACATTTCAAAAAGCGTCACTACTGATAGTGACGCTTTTCTTAG	1667
Sbjct	421	AAAGGCTTGCTGTTATTACATTTCAAAAAGCGTCACTACTGATAGTGACGCTTTTCTTAG	480
Query	1668	TTTGCCTGTTGAATTAACAGTGAGTGATAGACCTCTTTTGTTAGGACTTCATTAACCT	1727
Sbjct	481	TTTGCCTGTTGAATTAACAGTGAGTGATAGACCTCTTTTGTTAGGACTTCATTAACCT	540
Query	1728	TTTCATGCCATTCTTTAATTATTTTTGTCGTTTCTGCATTTTCACTAGAACAAATAACTA	1787
Sbjct	541	TTTCATGCCATTCTTTAATTATTTTTGTCGTTTCTGCATTTTCACTAGAACAAATAACTA	600
Query	1788	ATGTTTTTCTTTGGCTGAGGTAGTGTTTTCCCCCAATGTCCCAATCCAATACACGAATAG	1847
Sbjct	601	ATGTTTTTCTTTGGCTGAGGTAGTGTTTTCCCCCAATGTCCCAATCCAATATACGAATAG	660

```
Query 1848 GCGCTCTGAAAACAATCCAATCGGTTTCACTTTTCATCACCATTGACTTCATAAGCTCCAC 1907
          |||
Sbjct 661 GCGCTCTGAAAACAATCCAATCGGTTTCACTTTTCATCACCATTGACTTCATAAGCTCCAC 720

Query 1908 CAATTACCTGAACTAATGCAATTAATTTGCCACCATTTTAAATAGCAACAATATCATTCA 1967
          |||
Sbjct 721 CAATTACCTGAACTAATGCAATTAATTTGCCACCATTTTAAATAGCAACAATATCATTCA 780

Query 1968 CTTTCATTCTGTAAAAAGTCCTCTTTTCAACCTGCTCCAGCCCCATTTTCCTAAAC 2027
          |||
Sbjct 781 CTTTCATTCTGTAAAAAGTCCTCTTTTCAACCTGCTCCAGCCCCATTTTCCTAAAC 840

Query 2028 CTATAAttttttGTTCTCTAAAATTGAATGTAAAACCTCCGAAAAGCTTTGATCATCAG 2087
          |||
Sbjct 841 CTATAATTTTTTGTCTCTAAAATTGAATGTAAAACCTCTGAAAAGTTTGTATCATCAG 900

Query 2088 GGTGCATCTGCATATGCCAATATTGAGTCATGtttttttCCATAGGTAATTATGTTTGT 2147
          |||
Sbjct 901 GGTGCATCTGCATATGCCAATATTGAGTCATG-TTTTTTCCATAGGTAATTATGTTTGT 959

Query 2148 T 2148
          |
Sbjct 960 T 960
Score = 165 bits (182), Expect = 2e-36
Identities = 128/151 (84%), Gaps = 1/151 (0%)
Strand=Plus/Plus

Query 1 TTGACTGAATTAATGAACGAGTGTCTACGAGTGGCATAAAGAAAGTTCGCTTATTGCC 60
          |||
Sbjct 961 TTGACTGAATTAATGAACGAGTGTCTACGAGTGGCATAAAGAAAGTTCGCTTATTGCC 1020

Query 61 -CTCTGAAAAACCGCTTTTCTATATCTAATTTTCATGAAAATTTTAACTACTGTGAGCTT 119
          |||
Sbjct 1021 TCTTTGAAAAACCGCTTTTAAAAATACTGGCTGCTCGATAATTTTACTATCGAGTGTCT 1080

Query 120 TTGCTTTTGAAAAGACACGAGCGAAGCGAGT 150
          |||
Sbjct 1081 TTGCTTTTGAAAATAACGAGCGAAGCGAGT 1111
Score = 69.8 bits (76), Expect = 1e-07
Identities = 44/48 (91%), Gaps = 0/48 (0%)
Strand=Plus/Minus
Query 1188 ATAAACGAGCATACCGACATCACAGCAAGCTACGAACAGCACAAAAAA 1235
          |||
Sbjct 1257 ATAAACGAGTACACCGACATCACAGCCACCTACGAACAGCACAAAAAA 1210

>>>>nrnuc:GI_313507477 Gi|313507477|gb|HQ444404.1| TSE-associated circular DNA isolate
Sphinx 1.76, complete sequence. 0/0
Length=1758
Score = 1301 bits (1442), Expect = 0.0
Identities = 1248/1577 (79%), Gaps = 43/1577 (2%)
Strand=Plus/Plus

Query 19 CGAGTGTCTACGAGTGGCATAAAGAAAGTTCGCTTATTGCCCTCTGAAAAACCGCTTTT 78
          |||
Sbjct 27 CGAGTGTCTACGAGCGACTCAATGAAAGTTCGATTATTCCCCCTCTGAAAAACCGCTTTT 86

Query 79 TCTA-TATCTAATTTTCATGAAAATTTTAACTACTGTGAGCTTTTGCTTTTGAAAAGACAC 137
          |||
Sbjct 87 AAAAATAT-TGGCTGCTAGATGGTTTTACTATAGTGAGGTTTGTCTTTAAAAAAACAC 145

Query 138 GAGCGAAGCGAGTGCCATTTTGTCTTTGCTT---TTGGGGTTT---GGGGCTTGCCCC 191
          |||
Sbjct 146 GAGCAAAGCGAGTTCA--TAGTTGCTTTTGTCTTTTGGGGTCTTAGGGGAAATCCCT 203

Query 192 AACAGGGCACAGAGTAGGAATTT-----TACACGGTAGTGAAAAAGTCCCTCTGTGTGT 246
          |||
Sbjct 204 AACAAGTCCCTCGAATATCAAAATGTGGCTACATTTT-GTATATACGGGTAGGCT-TGCTT 261

Query 247 CCTTGCCCTATGCTTTT-TAAACCTTTGAATTTTCCCTGTAATTTGAAGAAATTGCCCC 305
          |||
Sbjct 262 ATTTGATTTTTTTTCTTAAACCTTTGACTTCTTCCCATTGTTTGCAGAAATTGCCCC 321

Query 306 TCGACTAAGCTTGCTTAGTCAAAAAAGTTTGAGCAAAGCGAAACATAGGGCAAttttca 365
          |||
Sbjct 322 TCGA---AGCTTGCTTAGTCAAAAAAGTTTGAGCAAAGCGAAACATAGGGCAATTTTCA 378

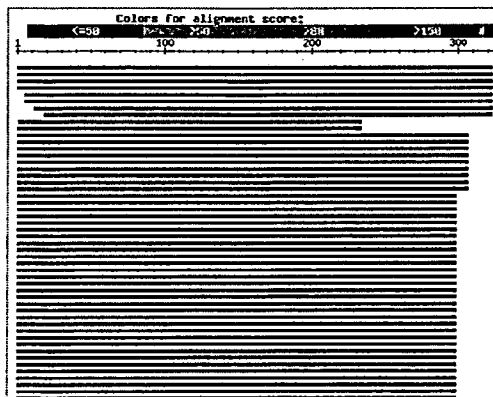
Query 366 taataaaaaatgggcttttaaatgttttaaatgcttttaaatgcttttaGACATGCTGAAA 425
```


Query 1557 GGCGTAGTTTTTTAGTC 1573
||||| |||||
Sbjct 1566 GGCGTGGTTTTTTAGTC 1582

>>>nrruc:GI_462476960 Gi|462476960|gb|APGK01000458.1| Dendroctonus ponderosae Seq01000458,
whole genome shotgun sequence. 0/0
Length=1455
Score = 711 bits (788), Expect = 0.0
Identities = 782/1032 (75%), Gaps = 20/1032 (1%)

ORF 324aa:

BlastP – Uniprot:



>>>sptrembl:N9LWE7_9GAMM N9lwe7 SubName: Full=Uncharacterize... 589 6e-166
>>>sptrembl:E5LG72_9ZZZZ E5lg72 SubName: Full=Replicase 3-li... 562 7e-158
>>>sptrembl:N9LWE7_9GAMM N9lwe7 SubName: Full=Uncharacterized protein;. 12/2013
Length=324
Score = 589 bits (1518), Expect = 6e-166, Method: Compositional matrix adjust.
Identities = 280/324 (86%), Positives = 304/324 (93%), Gaps = 0/324 (0%)

Query 1 MSDLVKDNALMNASYNLDLVEQRLILLAILEARETGKGINANDPLTVHAESYINQFGVA 60
Sbjct 1 MS+LIVKDNALMNASYNLDLVEQRLILLAI+EARE+GKGINANDPLTVHAESYINQFGV 60
Query 61 RQTAYQALKDACKDLFARQFSYQEKREGRANITSRWVSQIAYIDETATVEVIFAPAVVP 120
Sbjct 61 RTTAYQALKDACKDLFARQFSYQEKREGRANITSRWVSQIAYIDETATVEIIFAPAVVP 120
Query 121 LITRLEEQFTQYDIEQISGLSSAYAVRLYELLICWRSTGKTPVIELAEFRKRLGVLNDEY 180
Sbjct 121 LITRLEKQFTQYDIEQISGLSSAYAVRMYELLICWRSTGKTPVIELGEFRKRIGVLDTEY 180
Query 181 TRSDNFKKWIENPIKQINEHTDITASYEQHKKGRTITGFSFKFKQKKKTEPETPKNSDS 240
Sbjct 181 R+DN K +IE +KQIN+HTDITA+YEQHKKGRTITGFSFKFKQK+KTE ETPKNSDS 240
Query 241 SORIEKPSQIPANIVKQPENANLSDLQHRASKITGLIMSNRLSDRFKQGDDESIMQMMARI 300
Sbjct 241 +K +IP N+VKQPENAN+SDLQHRASKITGLIMSNRLSDRFKQ DESIMQMMARI 300
Query 301 QSEITTDSDIADQWQSKLEEFQGVVF 324
Sbjct 301 QSEIT ++IA+QW++KLEEFQGV+F 324
Query 301 QSEITDEAIAQWENKLEEFQGVIF 324

>>>sptrembl:E5LG72_9ZZZZ E5lg72 SubName: Full=Replicase 3-like protein;. 5/2013
Length=324
Score = 562 bits (1449), Expect = 7e-158, Method: Compositional matrix adjust.
Identities = 269/324 (83%), Positives = 290/324 (89%), Gaps = 0/324 (0%)

Query 1 MSDLVKDNALMNASYNLDLVEQRLILLAILEARETGKGINANDPLTVHAESYINQFGVA 60
Sbjct 1 MSDLVKDNALMNASYNL LVEQRLILLAI+EARETGKGINANDPLTVHA SYINQF V 60
Query 61 RQTAYQALKDACKDLFARQFSYQEKREGRANITSRWVSQIAYIDETATVEVIFAPAVVP 120

Sbjct 61 R TAYQALKDACKDLFARQFSYQEKREGR NITSRWVSI Y+D+TATVE+IFAPAVVP 120
RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSIQGYMDDTATVEIIFAPAVVP 120

Query 121 LITRLEEQFTQYDIEQISGLSSAYAVRLYELICWRSTGKTPVIELAEFRKRLGVLNDEY 180
LITRLEEQFTQYDIEQISGLSSAYAVR+YELICWRSTGKTP+IEL EFRKR+GVL+ EY 180

Sbjct 121 LITRLEEQFTQYDIEQISGLSSAYAVRMYELICWRSTGKTPIELDEFKRIGVLDTEY 180

Query 181 TRSDNFKKWIENPIKQINEHTDITASIEQHKKGRITGFSFKQKKTEPETPKNSDS 240
TR+DN K +IE +KQINEHTDITASIEQHKKGR ITGFSF FK KK+ +TP + S 240

Sbjct 181 TRTDNLKMQVIELALKQINEHTDITASIEQHKKGRVITGFSFMFKKKQNSDKTPDTNAS 240

Query 241 SQRIEKPSQIPANIVKQPENANLSDLQHRASKITGLIMNRLSDRFKQGDSEIMQMARI 300
S RI K SQIP NIVKQPEN A +SDL+HRAS++TG IM NRLSDRFKQGDSE + MM RI 300

Sbjct 241 SPRIVKHSQIPTNIVKQPENAKMSDLEHRASRVTEIMRNRLSDRFKQGDSEAIMMKRI 300

Query 301 QSEITDTSIADQWQSKLEEFVVF 324
QSEI TD+IADQW+SKLEEFVVF 324

Sbjct 301 QSEITDAIADQWESKLEEFVVF 324

>>>sptrembl:N9JLM2_ACIBA N9jlm2 SubName: Full=Uncharacterized protein;. 12/2013
Length=325

Score = 517 bits (1332), Expect = 3e-144, Method: Compositional matrix adjust.
Identities = 257/325 (79%), Positives = 291/325 (89%), Gaps = 1/325 (0%)

ProtSweep

A Protein Identification Tool

Results for /home/vir088/rindneu/tsekompl/CMI2.214p.324.pep

Number of residues: 324 Molecular weight: 37.11 kD
Isoelectric point: 7.51
more...
Prediction of protein localization:
cyto 17 %
cyto_nucl 12.5 %
extr 6 %
nucl 6 %

Identified Protein & Transcript:

Protein:
id: sptrembl:N9LWE7 9GAMM
valuation: homologous
scientific organism: Monodelphis domestica
common organism: Short-tailed gray opossum
abstract: N9lwe7 SubName: Full=Uncharacterized protein;. 12/2013Length=324
isoform:

GO information:

accession	group	name	evidence
GO:0005727	cellular component	extrachromosomal circular DNA	IEA:InterPro
GO:0003887	molecular function	DNA-directed DNA polymerase activity	IEA:InterPro
GO:0006270	biological process	DNA replication initiation	IEA:InterPro

Interpro information:

id	short name
----	------------

CM13.168 (1687bp)**(mi3.NnXn.1600.5)**

GGATTAATGCCAATGATCCCTTAACAGTTCATGCAAGTAGCTATATCAATCAATTTAACGTTGAAAGACATACGGCATATCAA
GCATTAAAAGATGCTTGTAAAGGACTTATTTGCACGCCAATTTAGTTACCAAGAAAAACGTGAGCGTGGGCGAGCTAATATTAC
AAGTCGTTGGGTATCGCAAATTGCTTACATTGATGAAACGGCAACAGTTGAGGTTATTTTTGCCCTGCCGTTGTTCCACTGA
TTACACGATTAGAGGAACAATTCACCCATACGATATTGAACAAATTAGCGGTTTATCGAGTGCCTATGCAGTTTCGCTTATAT
GAGTTATTGATCTGCTGGAGAACAACAGGAAAGACACCAGTTCCTTGACTTAACAGAGTTTAGAAAGCGTCTAGGTGTTTTAGA
TACCGAATACACTAGAACAGATAACTTGAAGATGCGAGTGATTGAGCAGTCACTCAAGCAAATAACAAGCATACCGACATCA
CAGCGAGCTACGAACAGCACAAAAAAGGCGGACAATCACAGGATTCTCATTCAAGTTTAAGCAGAAGAAGAAAAACAGAGCCA
GAAACGCCAAAAACAACGATTCTGGCGTTTCTAAGCCGAAAACAGTAGAAATCCCTGCCGAAGTCGTAAACAGCCTAAAAA
CACCAATTTGAGCGATTTAGAGAAACGAGTAAGGATGATCACAGGGCAATTGCTAAAAATAACCTGGCAAGTCGTTTTCAAC
ATGGGAATGAATCACCCTGGATATGATGAAACGTATTCAAAGTGAAATCACATCAGACGAACTGCCGATCTATGGCAAAAC
AACTTGAATCAATGGGTGTGGTTTTTTAGTCATGACGAATTTCCCGAAGGGCGCACTTTCCCAAATATCGTTTGAGTGCCT
TTTTCTATCGAAAAAGTAAAAAGACAAAGAAAAATTTCTTTGCTTTTTTTCTGTTTTTTCAAAAAAAGGGATTAGCGAGTGTC
TACGAGCGACACAATGAAAGTTCGATTATTTCCCCCTCTGAAAAACCGCTTTTTTGTATCTGATTTCAAGAAAATTTAACTA
TCGTGCGCTTTTGCTTTTGTAGAAGACACGAGCGAAGCGAGTGCCATTTTTGCTTTTGCTTTTGGGGTTTGGGGCTT
GCCCAACAAGGCACAGAGTAGGAATTTACACGGTAGTGGAAAAGTCCCCTCTGTGTGCTTGTCTTAATCTTTTTAAAC
CTTTGATTCTTCACCGTAATTTGCTAAAAATTGCCCGACTGAGTGAAGCGAAGTCAAAAAAGTTTGAGCAAAGCGAAAAACAT
AGGGCAATTTTCGTGATGAAAATGGGCTTTTAAGGTTTTTAATGCTTTTAAATGCTTTTAAATGCTTTTAGACACGCTGAAACGCAAGCCTA
GCAAGGTATACAGAGGGCATTACCTACGTTTACCGATCAATACCCCTACGTTTACCGATCAATACCCCTACGTTTACCGATC
AATACCCCTACGTTTACCTTGCATATACTACAAAATATAATAATGTAGTAATAACTTCAAATAATAACGTAGGTATGAG
CGATTTAATAGTAAAAGATAACGCCCTAATGAACGCTAGTTACAACCTAGCTTTGGTTGAACAGCGTTTAAATCTATTGGCTA
TCCTAGAGGCGAGAGAAACAGGCAAAG

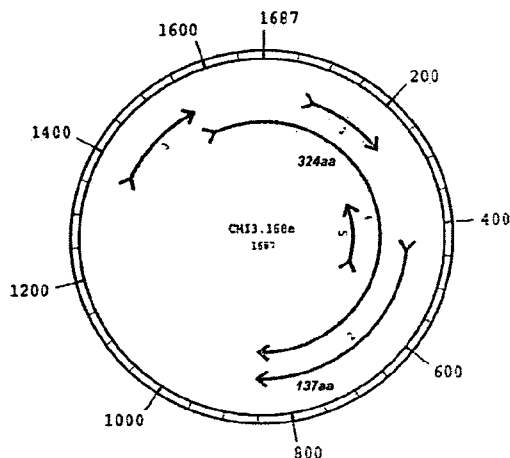
Repeat - ATACCCCTACGTTTACCGATCA - 3x22nt + 12nt**79% nucleotide similarity to TSE-associated circular DNA isolate Sphinx 1.76****ORF: 324aa – replication protein 80% similarity to Spinx1.76****137aa – replication protein 59% similarity to Spinx1.76**

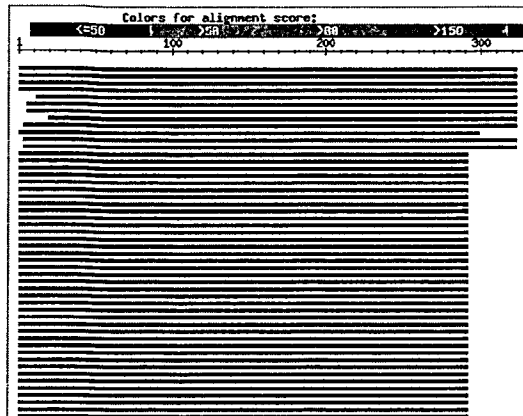
Figure 1(g)

Query	1	GGATTAATGCCAATGATCCCTTACAGTTTCATGCAAGTAGCTATATCAATCAATTAAACG 	60
Sbjct	721	GGATTAATGCCAATGATCCTCTTACGGTTCATGCAGGTAGCTATATCAATCAATTAAACG 	780
Query	61	TTGAAAGACATACGGCATATCAAGCATTAAAAGATGCTTGTAAGGACTTATTTGCACGCC 	120
Sbjct	781	TACAAAGGCATACGGCATATCAAGCCCTCAAAGATGCTTGTAAGGACTTGTTTGCCCGTC 	840
Query	121	AATTTAGTTACCAAGAAAAACGTGAGCGTGGGCGAGCTAATATTACAAGTCGTTGGGTAT 	180
Sbjct	841	AATTCAGTTACCAAGAAAAGCGAGAACGAGGACGAATTAATATTACAAGTCGATGGGTTT 	900
Query	181	CGCAAATTGCTTACATTGATGAAACGGCAACAGTTGAGGTTATTTTTGCCCTGCCGTTG 	240
Sbjct	901	CGCAAATTGCTTATATGGACGATACAGCAACCGTTGAGATTATTTTTGCCCTGCCGTTG 	960
Query	241	TTCCACTGATTACAGATTAGAGGAACAATTCACCCAATACGATATTGAACAAATTAGCG 	300
Sbjct	961	TTCTCTGATTACACGGCTAGAGGAACAGTTCACCCAGTACGATATTGAGCAAATTAGCG 	1020
Query	301	GTTTATCGAGTGCCTATGCAGTTCGCTTATATGAGTTATTGATCTGCTGGAGAACAACAG 	360
Sbjct	1021	GTTTATCGAGTGCATATGCTGTTCTGATGTACGAACTGCTGATTGTTGGCGTAGCACAG 	1080
Query	361	GAAAGACACCAGTTCCTTGACTTAAACAGAGTTTAGAAAGCGTCTAGGTGTTTTAGATACCG 	420
Sbjct	1081	GCAAAACACCAATTATTGAGCTAGACGAGTTTAGAAAGCGAATAGGTGTTTTAGATACTG 	1140
Query	421	AATACACTAGAACAGATAACTTGAAGATGCGAGTGATTGAGCAGTCACTCAAGCAAATAA 	480
Sbjct	1141	AATACACTAGAACAGATAATTTAAAGATGCAAGTTATTGAATTAGCCCTAAAACAAATCA 	1200
Query	481	ACAAGCATACCGACATCACAGCGAGCTACGAACAGCACAAAAAGGGCGGACAATCACAG 	540
Sbjct	1201	ACGAACATACTGACATCACAGCAAGCTATGAACAACACAAAAAGGGCGAGTGATTACCG 	1260
Query	541	GATTCTCATTCAAGTTTAAAGCAGAAG~AAGAAAACAGAGCCAGAAACGCCAAAAACAAC 	599

Query	971	ttca-aaaaaagGGATTAGCGAGTGTCTACGAGCGACACAATGAAAGTTCGATTATTTC	1029
Sbjct	8	TTCACAAAAAGATATTAGCGAGTGTCTACGAGCGACTCAATGAAAGTTCGATTATT-CC	66
Query	1030	CCCTCTGAAAAACCGCTTTTTTGTATCT-GATTTCAAGAAAATTTTAACTATCGTGC	1088
Sbjct	67	CCCTCTGAAAAACCGCTTTTAAAAATATTGGCTGCTAGATGGTTTTTACTATAGTGAGGT	126
Query	1089	TTTGCTTTTGAGAAGACACGAGCGAAGCGAGTGCCAttttgcctttgcttttgcctttt	1148
Sbjct	127	TTTGCTTTTAAAAAAACACGAGCAAAGCGAGTTC-ATAGTTGCTTTTGCTTGT--TTTCG	183
Query	1149	ggggtttggggcttgcCCC-AACAAGGCACAGAGTAGGAATTT-----TACACGGTAGTG	1202
Sbjct	184	GGTCTTAGGGGAAATCCCTTAACAAGTCTCGAATATCAAAATGTGGCTACATTTT-GTA	242
Query	1203	GAAAAGTCCCTCTGTGTGTCCTTGCTCTAATCTTTTTAAACCTTTGA-TTCTTCACCG	1261
Sbjct	243	TATACGGGTAGGCT-TGCTTATTGATTTTTTTTCTTCTAAACCTTTGACTTCTTCCCCA	301
Query	1262	TAATTTGCTAAAAATTGCCCGACTGAGTGAAGCGAAGTCAAAAAAGTTTGAGCAAAGCG	1321
Sbjct	302	TTGTTTGC-AGAAATTGCCCTC--GAAGCTTGCTTAGTCAAAAAAGTTTGAGCAAAGCG	358
Query	1322	AAAACATAGGGCAATTTTCGTGATGAAAATGGGCTTTTAAGGTTTTTAAATGCTTTTAA	1381
Sbjct	359	AAAACATAGGGCAATTTTCATGAAGAAATTGGGCTTTTAAAGTTTTTAAATG-TTTTAA	417
Query	1382	TGCTTTTAGACACGCTGAAACGCAAGCCTAGCAAGGTATACAGAGGCATTTACCTACGT	1441
Sbjct	418	TGCTTTTAGACATGCTAAGAAGCCACACAGCAAGGCATACAGAGGACATTCTCCTACGT	477
Query	1442	TTACCGATCAATACCCCTACGTTTACCGATCAATACCCCTACGTTTACCGATCAATACCC	1501
Sbjct	478	TTACCGATCAATACCCC--CGTTTACCGATCAATACCCCTACGTTTACCGATCAATACCC	535
Query	1502	CTACGTTTACCTTGCAATATACTACAAAATATAAATGTAGTAATAACTTCAAAATAAT	1561
Sbjct	536	CTACGTTTACCTTGCGTATACTACAAAGAATACTAGTGTAGTA-TAACTTCAAAAGAAT	594
Query	1562	AAACGTAGGTTATGAGCGATTAAATAGTAAAAGATAACGCCCTAATGAACGCTAGTTACA	1621
Sbjct	595	AATTGTAGGTTATGAGCGATTAAATAGTAAAAGATAACGCCCTAATGAATGCTAGTTATA	654
Query	1622	ACTTAGCTTTGGTTGAACAGCGTTTAAATCTATTGGCTATCCTAGAGGCGAGAGAAACAG	1681
Sbjct	655	ACTTAGCTTTGGTTGAACAGAGGTTAATCTATTAGCAATCATAGAAGCGAGAGAAACAG	714
Query	1682	GCAAAG 1687	
Sbjct	715	GCAAAG 720	

ORF 324aa:

BlastP – Uniprot:



```
>>>sptrembl:N9LWE7_9GAMM N9lwe7 SubName: Full=Uncharacterize... 550 5e-154
>>>sptrembl:E5LG72_9ZZZZ E5lg72 SubName: Full=Replicase 3-li... 540 3e-151
>>>sptrembl:N9JLM2_ACIBA N9jlm2 SubName: Full=Uncharacterize... 523 5e-146
```

```
>>>>sptrembl:N9LWE7_9GAMM N9lwe7 SubName: Full=Uncharacterized protein;. 12/2013
Length=324
Score = 550 bits (1416), Expect = 5e-154, Method: Compositional matrix adjust.
Identities = 261/324 (80%), Positives = 288/324 (88%), Gaps = 0/324 (0%)
```

```
Query 1 MSDLIVKDNALMNASYNLALVEQRLILLAILEARETGKGINANDPLTVHASSYINQFNVE 60
MS+LIVKDNALMNASYNL LVEQRLILLAI+EARE+GKGINANDPLTVHA SYINQF V
Sbjct 1 MSELIVKDNALMNASYNLDLVEQRLILLAIVEARESGKGINANDPLTVHAESYINQFGVH 60

Query 61 RHTAYQALKDACKDLFARQFSYQEKRRERGRANITSRWVSQIAYIDETATVEVIFAPAVVP 120
R TAYQALKDACKDLFARQFSYQEKRRERGRANITSRWVSQIAYIDETATVE+IFAPAVVP
Sbjct 61 RTTAYQALKDACKDLFARQFSYQEKRRERGRANITSRWVSQIAYIDETATVEIIFAPAVVP 120

Query 121 LITRLEEQFTQYDIEQISGLSSAYAVRLYELLCWRRTTGKTPVLDLTFEVRKRLGVLDTEY 180
LITRLE+QFTQYDIEQISGLSSAYAVR+YELLCWR+TGKTPV++L EFRKR+GVLDTEY
Sbjct 121 LITRLEKQFTQYDIEQISGLSSAYAVRMYELLCWRSTGKTPVIELGEFRKRIGVLDTEY 180

Query 181 TRTDNLKMRVIEQSLKQINKHTDITASIEQHKKGRITITGFSFKFKQKKKTEPETPKNND 240
RTDNLKMRVIE +LKQIN HTDITA+YEQHKKGRITITGFSFKFKQK+KTE ETPKN+DS
Sbjct 181 IRTDNLKMRVIELALKQINDHTDITATYEQHKKGRITITGFSFKFKQKRKTELETPKNSDS 240

Query 241 GVSKPKTVEIPAEVVKQPKNTNLSLEKRVRMITGAIKNNLASRFQHGNEPLDMMKRI 300
V K K+VEIP +VKQP+N N+SDL+ R ITG I N L+ RF+ +ES + MM RI
Sbjct 241 DVPKQKSVEIPTNLVKQPENANMSDLQHRASKITGLIMSNRLSDRFKQSDSIMQMMARI 300

Query 301 QSEITSDETADLWQNKLESMGVVF 324
QSEIT + A+ W+NKLE GV+F
Sbjct 301 QSEITDEAIANQWENKLEEFVIF 324
```

```
>>>>sptrembl:E5LG72_9ZZZZ E5lg72 SubName: Full=Replicase 3-like protein (Sphinx);. 5/2013
Length=324
Score = 540 bits (1391), Expect = 3e-151, Method: Compositional matrix adjust.
Identities = 254/324 (78%), Positives = 283/324 (87%), Gaps = 0/324 (0%)
```

```
Query 1 MSDLIVKDNALMNASYNLALVEQRLILLAILEARETGKGINANDPLTVHASSYINQFNVE 60
MSDLIVKDNALMNASYNLALVEQRLILLAI+EARETGKGINANDPLTVHA SYINQFN+
Sbjct 1 MSDLIVKDNALMNASYNLALVEQRLILLAIEARETGKGINANDPLTVHAGSYINQFNQV 60

Query 61 RHTAYQALKDACKDLFARQFSYQEKRRERGRANITSRWVSQIAYIDETATVEVIFAPAVVP 120
RHTAYQALKDACKDLFARQFSYQEKRRERGR NITSRWVSQI Y+D+TATVE+IFAPAVVP
Sbjct 61 RHTAYQALKDACKDLFARQFSYQEKRRERGRINITSRWVSQIGYMDTATVEIIFAPAVVP 120
```

```

Query 121 LITRLEEQFTQYDIEQISGLSSAYAVRLYELLCWRTTGKTPVLDLTFEFRKRLGVLDTHEY 180
          LITRLEEQFTQYDIEQISGLSSAYAVR+YELLCWR+TGKTP+++L EFRKR+GVLDTHEY
Sbjct 121 LITRLEEQFTQYDIEQISGLSSAYAVRMYELLCWRSTGKTPIIELDEFKRKRIGVLDTHEY 180

Query 181 TRTDNLKMRVIEQSLKQINKHTDITASIEQHKKGRTITGFSFKFKQKKKTEPETPKNNDS 240
          TRTDNLKM+VIE +LKQIN+HTDITASIEQHKKGR ITGFSF FK KK+ +TP N S
Sbjct 181 TRTDNLKMQVIELALKQINEHTDITASIEQHKKGRVITGFSFMFKKKQNSDKTPDTNAS 240

Query 241 GVSKPKTVEIPAEEVVKQPKNTNLSDLKRVRMITGAIKNNLASRFQHGNEPLDMMKRI 300
          K +IP +VKQP+N +SDLE R +TG I +N L+ RF+ G+ES +DMMKRI
Sbjct 241 SPRIVKHSQIPTNIVKQFENAKMSDLEHRASRVGTGEIMRNRLSDRFKQGDSEAIMMKRI 300

Query 301 QSEITSDETADLWQNKLESMGVVF 324
          QSEI +D AD W++KLE GVVVF
Sbjct 301 QSEIITDAIADQWESKLEEFVVVF 324

```

ProtSweep**A Protein Identification Tool**

Results for /home/vir088/rindneu/tsekompl/CMI3.168.324.pep

Number of residues: 324	Molecular weight: 36.97 kD
Charge: 5	Isoelectric point: 9.63

more...

Prediction of protein localization:

cyto	12 %
cyto_nucl	10 %
extr	6 %
nucl	6 %
mito	6 %

Identified Protein & Transcript:**Protein:**id: sptrembl:N9LWE7 9GAMM

valuation: similar

scientific organism: Acinetobacter sp

abstract: N9lwe7 SubName: Full=Uncharacterized protein; 12/2013Length=324

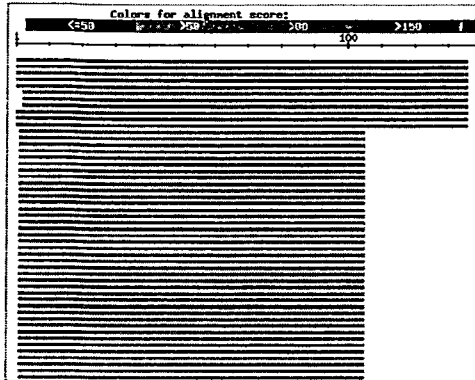
isoform:

GO information:

accession	group	name	evidence
GO:0005727	cellular component	extrachromosomal circular DNA	IEA:InterPro
GO:0003887	molecular function	DNA-directed DNA polymerase activity	IEA:InterPro
GO:0006270	biological process	DNA replication initiation	IEA:InterPro

Interpro information:

id	short name
IPR000525	Initiator_Rep_prot
IPR011991	WTH_DNA-bd_dom

ORF 137aa:**BlastP – Uniprot:**

```
>>>sptrembl:N9LWE7_9GAMM N9lwe7 SubName: Full=Uncharacterize... 191 1e-46
>>>sptrembl:N9JLM2_ACIBA N9jlm2 SubName: Full=Uncharacterize... 190 2e-46
>>>sptrembl:N8ZNY0_9GAMM N8zny0 SubName: Full=Uncharacterize... 190 2e-46
>>>sptrembl:S7WS39_9GAMM S7ws39 SubName: Full=DNA replicatio... 190 2e-46
>>>sptrembl:E5LG72_9ZZZZ E5lg72 SubName: Full=Replicase 3-li... 171 9e-41

>>>>sptrembl:N9LWE7_9GAMM N9lwe7 SubName: Full=Uncharacterized protein;. 12/2013
Length=324
Score = 191 bits (484), Expect = 1e-46, Method: Compositional matrix adjust.
Identities = 91/137 (66%), Positives = 108/137 (78%), Gaps = 0/137 (0%)

Query 1 MRVIEQSLKQINKHTDITASIEQHKKGRITITGFSFKFKQKKKTEPETPKNNDSGVSKPKT 60
MRVIE +LKQIN HTDITA+YEQHKKGRITITGFSFKFKQK+KTE ETPKN+DS V K K+
Sbjct 188 MRVIELALKQINDHTDITATYEQHKKGRITITGFSFKFKQKRKTELETPKNSDSVVPKQKS 247

Query 61 VEIPAEVVKQPKNTNLSLEKRVRMITGAIKNNLASRFQHGNEsplDMMKRIQSEITSD 120
VEIP +VKQP+N N+SDL+ R ITG I N L+ RF+ +ES + MM RIQSEIT +
Sbjct 248 VEIPNLVQKQENANMSDLQHRASKITGLIMSNRLSDRFKQSDSIMQMMARIQSEITDE 307

Query 121 ETADLWQNKLESMGVVF 137
A+ W+NKLE GV+F
Sbjct 308 AIANQWENKLEEFGVIF 324

>>>>sptrembl:N9JLM2_ACIBA N9jlm2 SubName: Full=Uncharacterized protein;. 12/2013
Length=325
Score = 190 bits (483), Expect = 2e-46, Method: Compositional matrix adjust.
Identities = 103/137 (75%), Positives = 113/137 (82%), Gaps = 0/137 (0%)

Query 1 MRVIEQSLKQINKHTDITASIEQHKKGRITITGFSFKFKQKKKTEPETPKNNDSGVSKPKT 60
MRVIE +LKQIN +TDITASIEQHKKGR ITGFSFKFKQKKKTE ETPKN+DSGVSK K
Sbjct 189 MRVIELALKQINDYTDITASIEQHKKGRVITGFSFKFKQKKKTEAETPKNSDSGVSKQKP 248

Query 61 VEIPAEVVKQPKNTNLSLEKRVRMITGAIKNNLASRFQHGNEsplDMMKRIQSEITSD 120
+EIP EVVKQPKN NLSLE R ITG I +N L RF+ +ESPL+MMKRIQSEIT+D
Sbjct 249 IEIPTEVVKQPKNANLSLEHRASRITGEIMRNRLTDRFKQADESPLEMMKRIQSEITTD 308

Query 121 ETADLWQNKLESMGVVF 137
AD W+NKLE GVVVF
Sbjct 309 TIADQWENKLEEFGVVF 325

>>>>sptrembl:N8ZNY0_9GAMM N8zny0 SubName: Full=Uncharacterized protein;. 12/2013
Length=325
Score = 190 bits (483), Expect = 2e-46, Method: Compositional matrix adjust.
Identities = 103/137 (75%), Positives = 113/137 (82%), Gaps = 0/137 (0%)
```

```

Query 1 MRVIEQSLKQINKHTDITASIEQHKKGRITITGFSFKFKQKKKTEPETPKNNDSGVSKPKT 60
MRVIE +LKQIN +TDITASIEQHKKGR ITGFSFKFKQKKKTE ETPKN+DSGVSK K
Sbjct 189 MRVIELALKQINDYTDITASIEQHKKGRVITGFSFKFKQKKKTEAETPKNSDSGVSKQKP 248

Query 61 VEIPAEVVKQPKNTNLSLEKRVRMITGAIKNNLASRFQHGNEsplDMMKRIQSEITSD 120
+EIP EVVKQPKN NLSLE R ITG I +N L RF+ +ESPL+MMKRIQSEIT+D
Sbjct 249 IEIPTEVVKQPKNANLSLEHRASRITGEIMRNRLTDRFKQADESPLEMMKRIQSEITD 308

Query 121 ETADLWQNKLESMGVVF 137
AD W+NKLE GVVF
Sbjct 309 TIADQWENKLEEFVVVF 325

```

```

>>>>sptrembl:S7WS39_9GAMM S7ws39 SubName: Full=DNA replication protein;. 12/2013
Length=314
Score = 190 bits (483), Expect = 2e-46, Method: Compositional matrix adjust.
Identities = 103/137 (75%), Positives = 113/137 (82%), Gaps = 0/137 (0%)

```

```

Query 1 MRVIEQSLKQINKHTDITASIEQHKKGRITITGFSFKFKQKKKTEPETPKNNDSGVSKPKT 60
MRVIE +LKQIN +TDITASIEQHKKGR ITGFSFKFKQKKKTE ETPKN+DSGVSK K
Sbjct 178 MRVIELALKQINDYTDITASIEQHKKGRVITGFSFKFKQKKKTEAETPKNSDSGVSKQKP 237

Query 61 VEIPAEVVKQPKNTNLSLEKRVRMITGAIKNNLASRFQHGNEsplDMMKRIQSEITSD 120
+EIP EVVKQPKN NLSLE R ITG I +N L RF+ +ESPL+MMKRIQSEIT+D
Sbjct 238 IEIPTEVVKQPKNANLSLEHRASRITGEIMRNRLTDRFKQADESPLEMMKRIQSEITD 297

Query 121 ETADLWQNKLESMGVVF 137
AD W+NKLE GVVF
Sbjct 298 TIADQWENKLEEFVVVF 314

```

```

>>>>sptrembl:E5LG72_9ZZZZ E5lg72 SubName: Full=Replicase 3-like protein;. 5/2013
Length=324
Score = 171 bits (434), Expect = 9e-41, Method: Compositional matrix adjust.
Identities = 82/137 (59%), Positives = 100/137 (72%), Gaps = 0/137 (0%)

```

```

Query 1 MRVIEQSLKQINKHTDITASIEQHKKGRITITGFSFKFKQKKKTEPETPKNNDSGVSKPKT 60
M+VIE +LKQIN+HTDITASIEQHKKGR ITGFSF FK KK+ +TP N S K
Sbjct 188 MQVIELALKQINEHTDITASIEQHKKGRVITGFSFMFKKKQNSDKTPDTNASSPRIVKH 247

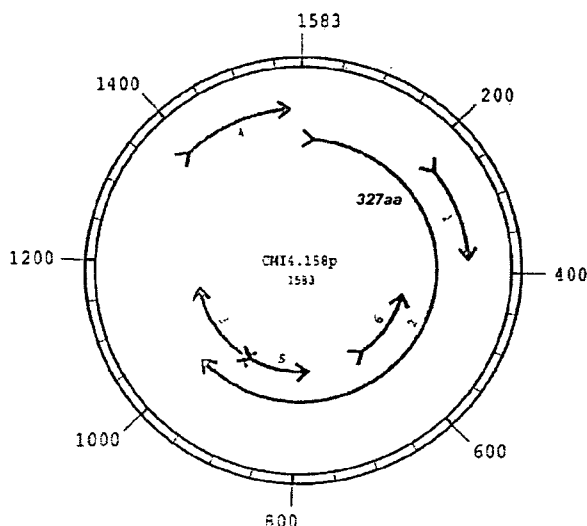
Query 61 VEIPAEVVKQPKNTNLSLEKRVRMITGAIKNNLASRFQHGNEsplDMMKRIQSEITSD 120
+IP +VKQP+N +SDLE R +TG I +N L+ RF+ G+ES +DMMKRIQSEI +D
Sbjct 248 SQIPTNIVKQPENAKMSDLEHRASRVTEIMRNRLSDRFKQGDSEIDMMKRIQSEIITD 307

Query 121 ETADLWQNKLESMGVVF 137
AD W++KLE GVVF
Sbjct 308 AIADQWESKLEEFVVVF 324

```

CM14.158 (1583bp)***(mil.NoXo.1500.5)***

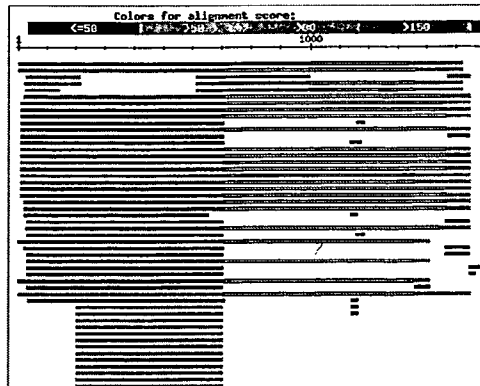
GAGGACGAATTAATATTACAAGTCGTTGGGTTTCACAAATTGGTTATATGGACGATACAGCAACCGTTGAGGTTATTTTTGCC
 CCTGCGGTTGTTCATTGATTACAAGACTGGAGGAACAATTCACCCAGTACGATATTGAGCAAATTAGCGGTTTATCTAGTGC
 TTATGCCGTCCTGTATGAATTATTGATCTGTTGGCGTAGCACAGGAAAAACGCCCGTCATAGAAGTAGCAGAGTTTAGAA
 AACGAATAGGTGTTTTAGATACTGAATACACGAGAACAGATAAATTAAAGATGCGAGTTATTGAGCTAGCACTCAAGCAGATC
 AACGAACATACAGACATCACAGCAAGCTATGAACAGCACAAAAAAGGGCGGACAATCACTGGGTTCTCATTCAAGTTTAAGCA
 GAAGAAGAAAAACAGAGCTAGAAACGCCAAAAAACAGCGATTCTAGCCACATATCGAAAAACCTAGTCAAATCCCTACCAACA
 TCGTAAACAGCCTGAAAACGCCAAAAAGACGATTAGGGCATAGAGCAAGTAAGATTACAGGTTTAATCATGTCCAATGGT
 TTAGCAGATCGGTTTAAACGTGGTGATGAGTCCGTCATGAATATGATGAAACGCATCAAAGAAGAAATTACAACCGATGCAAC
 AGCCGATCTATGGCAAAACAACTTGAATCAATGGTGTGGTTTTTTAGTCATGACTAAATTGGCTATGCAATGTCTCCGCTCT
 TTTTCTAAGAACCGTTTTTGTCTTTTCAAATCGCGCTCGCGGGTGTCTTTACTGGTATATGCGCTTTTTCGTGCTGCTTTT
 CAAAAATGTGGCGATAGCCAACATTCTCGATTTTGATCTTGATTTTGGGGTTTATAGGCTCTCCCTAACAGACATAGAGACA
 CAACCTCGCTAGAGGGGTCTCTATGTGCTTGGCCCTTAATTAATAAACTAAATAAAATTTTACCCGTTTTTGTCTTAAAAAAG
 GCAGGTTTTTCTTTGTCTTACTTTTGACTTTGATTTTCTCTGTCTTTTGTAGTTACCCACAAGAATTTTCATGTGGTTTTG
 GGCTTTTAAGGTTTTTAAATGCTTTTAAATGCTTTTAAACATGCTGAAACGTAAGCATAGCAAGGTATACAGAGTTTATTCTC
 CTACGTTTACCGATCAATACCCCTACGTTTACCGATCAATACCCCTACGTTTACCGATCAATACCCCTACGTTTACCGATCAA
TACCCCTACGTTTACGTTTGCATATACTACAAAAACAATAATGTAGTAATAACTTCAAAATAATAAGAGTAGGTTATGAGCG
ATTTAATAGTAAAGATAACGCCCTAATGAACGCTAGTTACAACCTAGCTTTGGTTGAACAGCGTTTAATTCTTTTAGCTATA
 TTGGAAGCGAGAGAAACAGGAAAAGGGATTAATGCCAATGATCCCTAACAGTTCATGCAAATAGCTATATCAATCAATTTAA
 CGTGGAAGACATACGGCATATCAAGCATTAAAGATGCCTGCAAAGACTTGTGTCACGTCAATTCAGTTACCAAGAAAAGC
 GAGAAC

Repeat – 3x22nt + 17nt

85% nucleotide similarity to TSE-associated circular DNA isolate Sphinx 1.76

ORF 327aa replication protein 80% similarity to Sphinx 1.76

Figure 1(h)

Blastn - Geall:

>>>nrnuc:GI_313507477 Gi|313507477|gb|HQ444404.1| TSE-associ... 1128 0.0
>>>nrnuc:GI_462476960 Gi|462476960|gb|APGK01000458.1| Dendro... 477 2e-130

>>>>nrnuc:GI_313507477 Gi|313507477|gb|HQ444404.1| TSE-associated circular DNA isolate
Sphinx 1.76, complete sequence. 0/0
Length=1758
Score = 1128 bits (1250), Expect = 0.0
Identities = 857/1005 (85%), Gaps = 5/1005 (0%)
Strand=Plus/Plus

Query	1	TCAAAATAATAAGAGTAGGTTATGAGCGATTTAATAGTAAAAGATAACGCCCTAATGAAC	60
Sbjct	585	TCAAAAGAATAATTGTAGGTTATGAGCGATTTAATAGTAAAAGATAACGCCCTAATGAAT	644
Query	61	GCTAGTTACAACCTTAGCTTTGGTTGAACAGCGTTAATTCTTTTAGCTATATTGGAAGCG	120
Sbjct	645	GCTAGTTATAACTTAGCTTTGGTTGAACAGAGGTTAATTCTATTAGCAATCATAGAAGCG	704
Query	121	AGAGAAACAGGAAAAGGGATTAATGCCAATGATCCCTAACAGTTCATGCAATAGCTAT	180
Sbjct	705	AGAGAAACAGGCAAAGGGATTAATGCCAATGATCCTCTTACGGTTCATGCAGGTAGCTAT	764
Query	181	ATCAATCAATTTAACGTGGAAAGACATACGGCATATCAAGCATTAAAAGATGCCTGCAAA	240
Sbjct	765	ATCAATCAATTTAACGTACAAAGGCATACGGCATATCAAGCCCTCAAAGATGCTTGTA	824
Query	241	GACTTGTGTTGCACGTCAATTCAGTTACCAAGAAAAGCGAGAACGAGGACGAATTAATATT	300
Sbjct	825	GACTTGTGTTGCCCCGTCAATTCAGTTACCAAGAAAAGCGAGAACGAGGACGAATTAATATT	884
Query	301	ACAAGTCGTTGGGTTTCACAAATTGGTTATATGGACGATACAGCAACCGTTGAGGTTATT	360
Sbjct	885	ACAAGTCGATGGGTTTCGCAAATTGGCTATATGGACGATACAGCAACCGTTGAGATTATT	944
Query	361	TTTGCCCTGCGGTTGTTCCATTGATTACAAGACTGGAGGAACAATTCACCCAGTACGAT	420
Sbjct	945	TTTGCCCTGCGGTTGTTCCCTCTGATTACACGGCTAGAGGAACAGTTCACCCAGTACGAT	1004
Query	421	ATTGAGCAAATTAGCGGTTTATCTAGTGCTTATGCCGTCCTGCTGTATGAATTATTGATC	480
Sbjct	1005	ATTGAGCAAATTAGCGGTTTATCGAGTGCATATGCTGTTCTGTATGTACGAACGCTGATT	1064
Query	481	TGTTGGCGTAGCACAGGAAAAACGCCCGTCATAGAACTAGCAGAGTTAGAAAACGAATA	540
Sbjct	1065	TGTTGGCGTAGCACAGGCAAAACACCAATTATTGAGCTAGACGAGTTAGAAAACGAATA	1124
Query	541	GGTGTGTTTAGATACTGAATACACGAGAACAGATAACTTAAAGATGCGAGTTATTGAGCTA	600

Query	1345	AAttttcatgtggtt-ttgggcttttaagggtttttaaatgcttttaaatgcttttaaACA	1403
Sbjct	371	AATTTTCATGAAGAAATTGGGCTTTTAAAGTTTTTAAATG-TTTTAAATGCTTTTAGACA	429
Query	1404	TGCTGAAACGTAAGCATAGCAAGGTATACAGAGTTCATTCTCCTACGTTACCGATCAAT	1463
Sbjct	430	TGCTAAGAAGCCACACAGCAAGGCATACAGAGGACATTCTCCTACGTTACCGATCAAT	489
Query	1464	ACCCCTACGTTTACCGATCAATACCCCTACGTTACCGATCAATACCCCTACGTTTACC	1522
Sbjct	490	ACCC--CGTTTACCGATCAATACCCCTACGTTACCGATCAATACCCCTACGTTTACC	546

Score = 477 bits (528), Expect = 2e-130
Identities = 538/713 (75%), Gaps = 17/713 (2%)

Colors for alignment scores					
(SSD)	0	50	100	150	200
	140	200	260	320	380

>>>sptrembl:E5LG72 9ZZZZ E5lq72 SubName: Full=Replicase 3-li... 574 2e-161

>>>sptrembl:N9LWE7_9GAMM N9lwe7 SubName: Full=Uncharacterize... 563 3e-158

>>>>sptrembl:E5LG72_9ZZZZ E5lg72 SubName: Full=Replicase 3-like protein;. 5/2013
Length=324

Score = 574 bits (1479), Expect = 2e-161, Method: Compositional matrix adjust.
Identities = 276/324 (85%), Positives = 294/324 (90%), Gaps = 0/324 (0%)

```
Query 1  MSDLVKDNALMNASYNLALVEQRLILLAILEARETGKGINANDPLTVHANSYINQFNVE 60
          MSDLVKDNALMNASYNLALVEQRLILLAI+EARETGKGINANDPLTVHA SYINQFNV+
Sbjct 1  MSDLVKDNALMNASYNLALVEQRLILLAILEARETGKGINANDPLTVHAGSYINQFNVQ 60

Query 61  RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSQIGYMDDTATVEVIFAPAVVP 120
          RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSQIGYMDDTATVE+IFAPAVVP
Sbjct 61  RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSQIGYMDDTATVEIIFAPAVVP 120

Query 121 LITRLEEQFTQYDIEQISGLSSAYAVRLYELLICWRSTGKTPVIELAEFRKRIGVLDTEY 180
          LITRLEEQFTQYDIEQISGLSSAYAVR+YELLICWRSTGKTP+IEL EFRKRIGVLDTEY
Sbjct 121 LITRLEEQFTQYDIEQISGLSSAYAVRMYELLICWRSTGKTPIELDEFKRKRIGVLDTEY 180

Query 181 TRTDNLKMRVIELALKQINEHTDITASIEQHKKGRITITGFSFKFKQKKKTELETPKNSDS 240
          TRTDNLKM+VIELALKQINEHTDITASIEQHKKGR ITGFSF FK KK+ +TP + S
Sbjct 181 TRTDNLKMQVIELALKQINEHTDITASIEQHKKGRVITGFSFMFKHKKQNSDKTPDTNAS 240

Query 241 SPHIEKPSQIPTNIVKQPENAKKDDLGHASKITGLIMSNGLADRFKRGDESVMNMMKRI 300
          SP I K SQIPTNIVKQPENAK DL HRAS++TG IM N L+DRFK+GDES ++MMKRI
Sbjct 241 SPRIVKHSQIPTNIVKQPENAKMSDLEHRASRVTEIMRNRLSDRFKQGDSEAIMMMKRI 300

Query 301 KEEITTDATDLWQNKLESMVWFF 324
          + EI TDA AD W++KLE F
Sbjct 301 QSEIITDAIADQWESKLEEFVGVF 324
```

>>>>sptrembl:N9LWE7_9GAMM N9lwe7 SubName: Full=Uncharacterized protein;. 12/2013
Length=324

Score = 563 bits (1452), Expect = 3e-158, Method: Compositional matrix adjust.
Identities = 271/324 (83%), Positives = 292/324 (90%), Gaps = 0/324 (0%)

```
Query 1  MSDLVKDNALMNASYNLALVEQRLILLAILEARETGKGINANDPLTVHANSYINQFNVE 60
          MS+LIVKDNALMNASYNL LVEQRLILLAI+EARE+GKGINANDPLTVHA SYINQF V
Sbjct 1  MSELIVKDNALMNASYNLDLVEQRLILLAIVEARESGKGINANDPLTVHAESYINQFGVH 60

Query 61  RHTAYQALKDACKDLFARQFSYQEKREGRINITSRWVSQIGYMDDTATVEVIFAPAVVP 120
          R TAYQALKDACKDLFARQFSYQEKREGR NITSRWVSQI Y+D+TATVE+IFAPAVVP
Sbjct 61  RTTAYQALKDACKDLFARQFSYQEKREGRANITSRWVSQIAYIDETATVEIIFAPAVVP 120

Query 121 LITRLEEQFTQYDIEQISGLSSAYAVRLYELLICWRSTGKTPVIELAEFRKRIGVLDTEY 180
          LITRLE+QFTQYDIEQISGLSSAYAVR+YELLICWRSTGKTPVIEL EFRKRIGVLDTEY
Sbjct 121 LITRLEKQFTQYDIEQISGLSSAYAVRMYELLICWRSTGKTPVIELGEFRKRIGVLDTEY 180

Query 181 TRTDNLKMRVIELALKQINEHTDITASIEQHKKGRITITGFSFKFKQKKKTELETPKNSDS 240
          RTDNLKMRVIELALKQIN+HTDITA+YEQHKKGRITITGFSFKFKQK+KTELETPKNSDS
Sbjct 181 TRTDNLKMRVIELALKQINDHTDITATYEQHKKGRITITGFSFKFKQKRKTELETPKNSDS 240

Query 241 SPHIEKPSQIPTNIVKQPENAKKDDLGHASKITGLIMSNGLADRFKRGDESVMNMMKRI 300
          +K +IPTN+VKQPEN A DL HRASKITGLIMSN L+DRFK+ DES+M MM RI
Sbjct 241 DVPKQKSVEIPTNLVKQPENANMSDLQHRASKITGLIMSNRLSDRFKQSDSEIMQMMARI 300

Query 301 KEEITTDATDLWQNKLESMVWFF 324
          + EIT +A A+ W+NKLE F
Sbjct 301 QSEITDEAIANQWENKLEEFVGVIF 324
```

>>>>sptrembl:N9JLM2_ACIBA N9jlm2 SubName: Full=Uncharacterized protein;. 12/2013
Length=325

Score = 517 bits (1331), Expect = 3e-144, Method: Compositional matrix adjust.
Identities = 260/325 (80%), Positives = 287/325 (88%), Gaps = 1/325 (0%)

ProtSweep

A Protein Identification Tool

Results for /home/vir088/rindneu/tsekompl/CMI4.158.327.pep

Number of residues: 327	Molecular weight: 37.51 kD
Charge: 2	Isoelectric point: 8.59

more...

Prediction of protein localization:

cyto_nucl	9.5 %
cyto	9 %
nucl	8 %

Identified Protein & Transcript:

Protein:

id: sptrembl:E5LG72-92222

valuation: homologous

scientific organism: Vicugna pacos

common organism: Alpaca

abstract: E5lg72 SubName: Full=Replicase 3-like protein; 5/2013Length=324

isoform:

GO information:

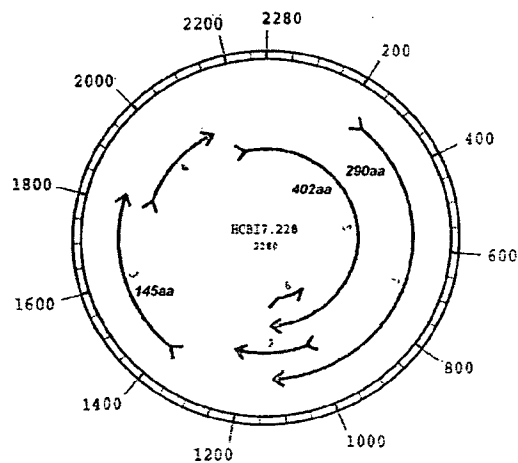
accession	group	name	evidence
GO:0005727	cellular component	extrachromosomal circular DNA	IEA:InterPro
GO:0003887	molecular function	DNA-directed DNA polymerase activity	IEA:InterPro
GO:0006270	biological process	DNA replication initiation	IEA:InterPro

Interpro information:

id	short name
IPR000525	Initiator_Rep_prot
IPR011991	WHTH_DNA-bd_dom

HCBI7.228 (2280bp)**(ri71.NdXd.2500.4G)**

CTAAGGCAGATCAACACAGGGATAGAATAACACGTTTTGGGATTTTGAAACATAGAGCGAAGCTACAAGAACAATATTTGTGG
 ACGCAGGTTGATTTCAAATCTGAAGGCGAAGACGATCTATCCAATAAGGCTCTTAAGGCTGCAACAAAATTTAAAGGTTGTGG
 CCAGTTTCTTATATTAGAAATTACTACACTATCGACCAAGTTAAGCTCGAAAAATTCACGTATGTGGACAGCATTTGCTAT
 GCCCAATGTGTGCAGGTATTCGCGCTGCTCGCTCAATGAAACGCTATTTAGATCGTATCCATGAATTAATGCGTCAGAATCCG
 CGTTTAAAGCCTGTTTTGATCACTTAACGGTCAAGAACGGGGAAGACTTAGAAGAACGCTATAATCACCTTACAAGCTCATT
 TAGAAAGCTTCTATCGCGTTATAGGGATTACAAGAAAAAGGGTTGGGGGTTTAACCAATTCTGTAAATTTGATGGTGCTTTCT
 ATACCACTGAATACCTTACAACGAAAAGACAAAAAATGGCATCCGCATATTCATATTTTCGCTTTACTCAATGAATGGATA
 GACCAGGAAGAATTATCTGAGACCTGGCATGACATTACCCTGGATTCTTATATCATCGATATTCGTAGAGTTAAAAAACCAA
 AGAACACGGCTATAGCAAAGCTGTGCTGAGGTATGTAAATACGCTCTTAAGTTTGGCGATCTATCAGTTGAAAACACTTGGG
 ATGCATTTTAAACGCTTAAGGGTAAACGCTTACAGGCTCTTTCGGTCTATGCATGGAGTTAAATTCAGATAAGGCCACG
 GATGATATGCCACTGGAGGAGCTGCCGTACTTAGAAATGTTCTATCGTTTTCGTGTGATAAAAATCCTACTACAACCTAGA
 AATAACAAAAGACGTAACGCCACAAACGAATCTCGACAGAGTGAGGAGGCGGCACCAACGACCGACAGACTGGAGAGAATG
 AGTTGAAGGCGCAGGATGAAGCGAGGACGTTGCCGGTGCGAGGCGACGCAGTCACGGAACGACCGCAGCAAGTCCATGAATAT
 TTTGACCCATTTGATATTGAGAGAGAAGAATACGACTTCTACCCTTAAGGGGTTGACCCCTTAAACCCAAAGGGCGCACTT
 ATGCAAACTTTTCAAGTTTCGCGAGTGCTCCCACCAGTAACAGAGGGCGCGGAGTGCGCCCGAACTGACATCATCTAAATTTCA
 AAACCTTCTCTCCCTGATTACAAGCTCCCTTTTGCCTTTCTCCGGAAAGTTTAGCCATCCCAAACGGGTGGCGACGCATTGA
 CAATTATAATTTAATCTGCAATATTCATCGGACATTCATAAATAATCTGATTGGAATATTAAATATGCTTTTTAAATCTCTAG
 GCGTTGCAATCGCTACAAATAATGCTTCTAACTCTGACGAAGACTTGAAACCAATTACTATCCGTTTAGCTGATAGTGAAATT
 CAGGCGTATGACACTGTTGCAAAATCAATGGGTCTAACTCGTCAAGATTTTCTATCACACGTTATTCTGTTCCAACTTCCGACA
 GGCTTTAAAGAGTTTGTAGTGGGTATACAAGTTCAAATCCTTCTACACCCTTATCTGTTTTGTTTAATTCAAATACGGACA
 ATGACGAAGTTAAGCTCAAACCTTCCAACCTTCTCTCATCTATTTCTCAAGAATTAATGGACGAAGAAGAAAAAGAAATTCAT
 GAAGCAATCGAAAACCATATGAATACTTCTATGCATATGCAACTCGTCCAAAAGGCATCTTTGTTGAGAAGGATGATTCAAA
 ATGATTAAGCTTGAAGGGATCGTTTTAAACGTTATTTACCCAGCAAGGTGGTAAAGATAAAAAACAGGTGATCAATTTGCTGA
 CCGTGACAAGGTGCAAATTTTAGGTGCTATGGATCTGCCTAACGGTGACGTTAAAAATGAGTTATTTACGTTATCTGTAGAAA
 ACTATCGGGAATTTAAAGACTTCTTAAATCAGAAAATCTGTATTGCTGTTGGTGCAATGGCAAGTGGTGCAGTGTATTTTTT
 TACGTTGCTAAAGGCGCAAAGCCTGTTTTAGCAGAAATGATGTGAAACACGAAAAAACAGTGGGTAGCCTTGGCTTCTCG
 CTGTTTTTTCTTTCTTTGATACTATTAATAAGTGGGGATGAAATTGACTAAAAAGACAATAAAAAAGCCCATATTTCTG
 ACACTTTGGCGGGTGATGAATTACAGGCTTTGCAATCTG



78% nucleotide similarity to TSE-associated circular DNA isolate *Sphinx* 2.36

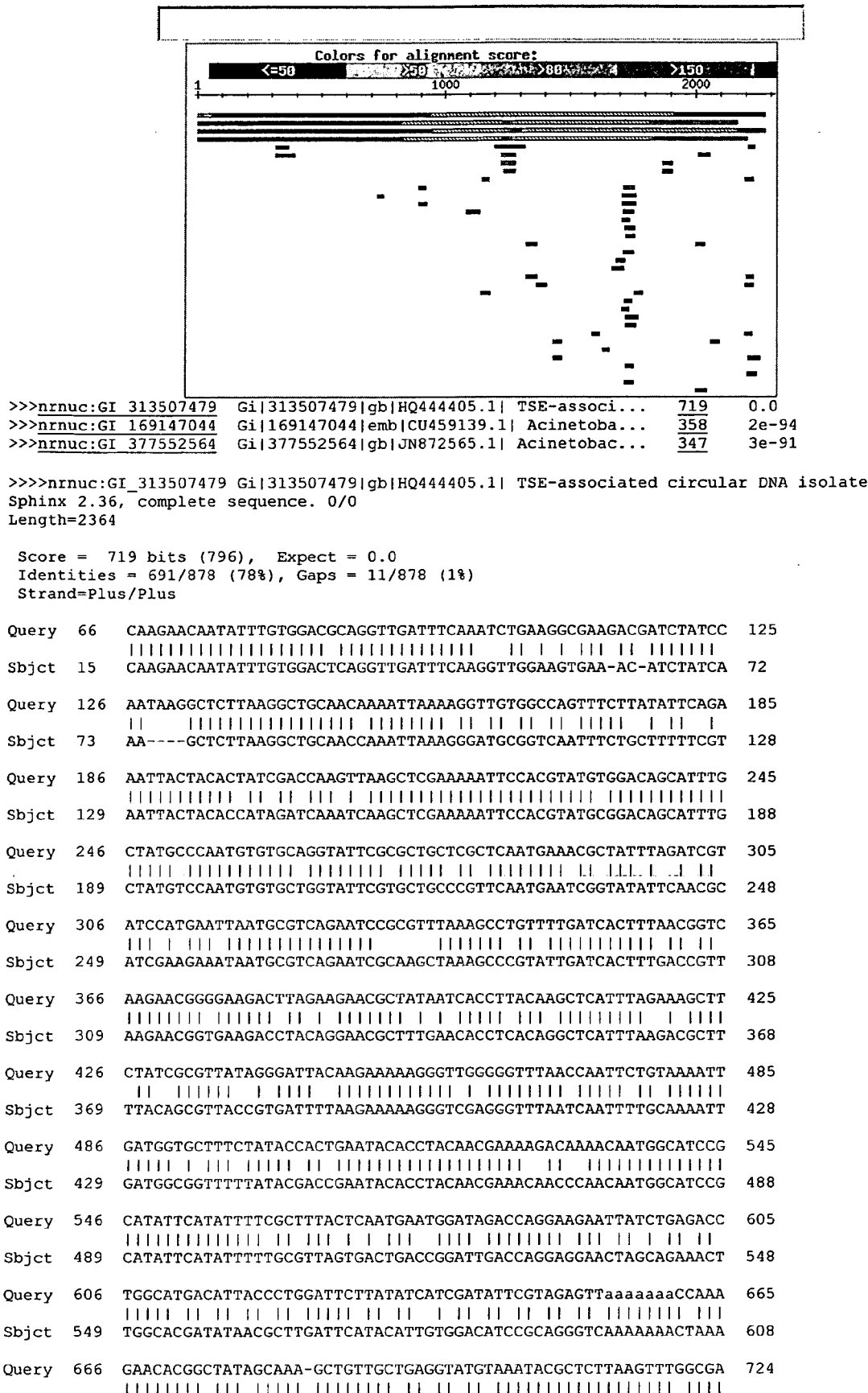
ORFs:

402aa – 81% in 232aa of the replication protein of *Sphinx* 2.36

290aa – 89% in 168aa of the replication protein of *Sphinx* 2.36

145aa – vague similarity to capsid protein of narcissus mosaic virus (nmv)

Figure 2

Blastn - Geall:

Sbjct	609	GAACACGGATAT-GCAAAGGCTGTTGCCGAAGTCTGCAAAACGCTCTTAAGTTTAGCGA	667
Query	725	TCTATCAGTTGAAAACACTTGGGATGCA-TTTTTAACGCTTAAGGGTAACGTCTTACAG 	783
Sbjct	668	TCTATCCACTGAGAAAACCTTTCAAGCATTTTTTGACCCTTAAGGGCAAAGGCTTACAG	727
Query	784	GCTCTTTCGGTTCTATGCATGGAGTTAAAAATTCCAGATAAGGCCACGGATGATATGCCAC 	843
Sbjct	728	GTTCATTCCGGCTCTATGCATGGTGTA AAAATTCTCGAAAGCGGACCCGATGAAATGCCTA	787
Query	844	TGGAGGAGCTGCCGTACTTAGAAATGTTCTATCGTTTCGTTGTGATAAAAAATCCTACT 	903
Sbjct	788	AAGAGGA ACTTCCATATCTTGAGCTGCTCTATCGTTTCGTTTTTGGTGAAAGGCTTATT	847
Query	904	ACAACCTAGAAAATAACAAAAGACGTAACGCC--ACAAA	939
Sbjct	848	ACAACCTAGAGTTAACTAAGGACGTAAGCCGAACAAA	885

Score = 315 bits (348), Expect = 2e-81
Identities = 294/371 (79%), Gaps = 2/371 (0%)
Strand=Plus/Plus

Query	1911	CGTGACAAGGTGCAAATTTTAGGTGCTATGGATCTGCCTAACGGTGACGTTAAAAATGAG	1970
Sbjct	1943	CGTGACAAGGTGCAAATTCCTTGCTCGCTGGAGTTGCCAACGGTGAGATCAAACATGAG	2002
Query	1971	-TTATTTACGTTATCTGTAGAAAACATATCGGGACTTTAAAGACTTCTTAAATCAGAAAAT	2029
Sbjct	2003	CTTGTTGAC-TTAACGGTTGAAGATGCTCGCATCTTTGAGCCATTCAAACACAAGGTAAT	2061
Query	2030	CTGTATTGCTGTTGGTGCAATGGCAAGTGGTCGCAATGTTATTTTTACGTTGCTAAAGG	2089
Sbjct	2062	TAGCATTTCATGCGGTGCTATGGCTATCGGTGCAAAATGTTGTTTTTATGTTCGAAAAGG	2121
Query	2090	CGCAAAGCCTGTTTTAGCAGAAATGATGTGAAACACGaaaaaaaaCAGTGGGTTAGCCCTT	2149
Sbjct	2122	TGCAAAGCCTGTTTTAGCAGATGTAATGTAATTCATGAAAAAAAAACAGTGAGTTAGGCTT	2181
Query	2150	GGCTTCTCGCTGTTTTTCTTTCTCTTGATACTATTAACATAAAGTGGGGATGAAATTGAC	2209
Sbjct	2182	GCCGACTCGCTGTTTTTCTTTACTTGATACTATTAACATAAAGTGGGGACAAAATTGAC	2241
Query	2210	TAAAAAAGACAATAAAAAAGCCATAATTCTGACACTTTGGCGGGTGATGAATTACAGGC	2269
Sbjct	2242	TAAAAAAGACAATAAAAAAGCCATAATTCTGATGCTTTGGCGGGCGACGAATTACAGGC	2301
Query	2270	TTTGCAATCTG	2280
Sbjct	2302	TTTGCAATCTG	2312

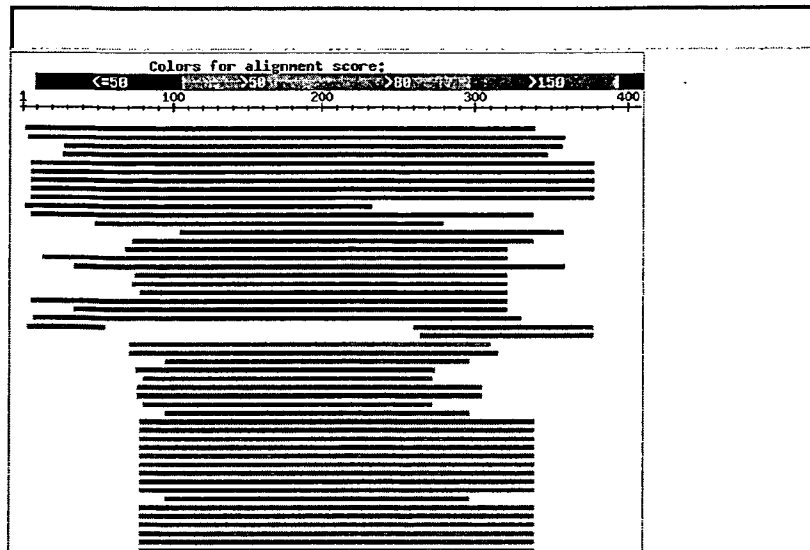
Score = 77.0 bits (84), Expect = 8e-10
Identities = 48/52 (92%), Gaps = 0/52 (0%)
Strand=Plus/Plus

Query 1 CTAAGGCAGATCAACACAGGGATAGAATAACACGTTTTGGGATTTTGAACA 52
|||||
Sbjct 2313 CTAATGCAGATCAACACAGGGATAGAATATCAGTTTTGGACTTTTGAACA 2364

```
>>>nrruc:GI_169147044 Gi|169147044|emb|CU459139.1| Acinetobacter baumannii str. AYE
plasmid p4ABAYE, complete genome. 0/0
Length=2726
```

Score = 358 bits (396), Expect = 2e-94
Identities = 580/832 (69%), Gaps = 17/832 (2%)
Strand=Plus/Plus

Query	3	AAGGCAGATCAACACAGGGATAGAATAACACGTTTTGGGATTTTGAACATAGAGCGAAG	62
Sbjct	115	AAGG-AGATCAACATAGGGATAGAATAACACGTTTTGGCATTTTGAACATAGATCGAAG	173
Query	63	CTACAAGAACAAATATTGTGGACGCAGGTGATTTCAA-----TCTGAAGCGGAAG--	114
Sbjct	174	CAACAAGAAAACTATTATTTTCGTTAGCTAAGATTAAAGAAAATTATCATGCCGATGTA	233

BlastP – Uniprot:**ORF 402aa:**

>>>sptrembl:H6VTN7_ACIBA H6vtn7 SubName: Full=Replication pr... 476 8e-132
>>>sptrembl:K5PFB3_ACIBA K5pfb3 SubName: Full=Replication pr... 459 1e-126

>>>>sptrembl:H6VTN7_ACIBA H6vtn7 SubName: Full=Replication protein;. 11/2013
Length=378

Score = 476 bits (1225), Expect = 8e-132, Method: Compositional matrix adjust.
Identities = 222/342 (64%), Positives = 270/342 (78%), Gaps = 4/342 (1%)

```
Query 3 LTKKDNKKAHNSDTLAGDELQALQSAKADQHRDRITRFGILKHAQLQEYLWTQVDFKS 62
+ KKDNNKKAH S+ LAG E+QALQSA ADQHRDRITRF LKHAQL QE YL+T FK
Sbjct 1 MLKKDNKKAHISEALAGGEIQALQSANADQHRDRITRFATLKHAQNQENYLFLLAKFKE 60

Query 63 EGEDDLN-----KALKAATKLKGGCQFLIFRNYTIDQVKLEKHFVCGQHLLCPMCAGIR 118
E D+ N KALK+A KL CG +L+F+N+YTI +VKL K CGQHLLCP CA IR
Sbjct 61 NYEKDVKNESIKALKSAQKLNECGNYLLFKNFYITIGEVKLSKLRTCGQHLLCPFCAAIR 120

Query 119 AARSMKRYLDRIHELMRQNPRLKPVLTITLVKNGEDLEERYNHLTSSFRKLLSRYRDYKK 178
A+R++++Y++RI ++++N +LKPVLITITLVKNG DL ER HL SFR LL R RDY+K
Sbjct 121 ASRAIQKYVERIDQVLQENRKLKPVLTITLVKNGSDLAERSEHLMKSFRLLERRRDYK 180

Query 179 KGWGFNQFCKIDGAFYTTEYTYNEKTKQWHPHIHIFALLNEWIDQEELSETWHDITLDSY 238
KG GFN+FC+ GA Y+ E T+NEKT +WHPHIH+FAL+++WIDQ+E SE WH +T DS
Sbjct 181 KGRGFNEFCVKQAMYSYENTFNEKTGEWHPHIHFALVDQWIDQEFSEYWHSLTGDSM 240

Query 239 IIDIRRVKKTKEHGYSKAVAEEVCKYALKFGDLSVENTWDAFLTLKGRLTGSFGSMHGVK 298
++D+RR +K K +GYSKA AEVCKYALKFGDLSVE TW+AF LKGRLTGSFG + GVK
Sbjct 241 VVDVRRARKEKGYGYSKAAAEVCKYALKFGDLSVEKTWEAFKVLKGRLTGSFGLLWGVK 300

Query 299 IPDKATDDMPLEELPYLEMFYRFVFDKKSYYNLEITKDVTPQ 340
IPD TDDMP E+LPYLEM Y+F + KKSYY+L IT+ V PQ
Sbjct 301 IPSMTDDMPSEDLPYLEMFLYKFAYSKKSYYDLLITRHVEPQ 342
```

>>>>sptrembl:K5PFB3_ACIBA K5pfb3 SubName: Full=Replication protein;. 11/2013
Length=421

Score = 459 bits (1181), Expect = 1e-126, Method: Compositional matrix adjust.
Identities = 221/360 (61%), Positives = 274/360 (76%), Gaps = 5/360 (1%)

```
Query 5 KKDNNKKAHNSDTLAGDELQALQSAKADQHRDRITRFGILKHAQLQEYLWTQVDFKSEG 64
+KD KK S TL G + + S +ADQHRDRITRFGILKHAQL QE YLWT +K
Sbjct 2 QKDIKKPILLSSTLVGGDNKGSNSVEADQHRDRITRFGILKHAQLQEYLWTQVDFKSEG 61

Query 65 EDDLN-----KALKAATKLKGGCQFLIFRNYTIDQVKLEKHFVCGQHLLCPMCAGIRAA 120
++D N +A K+A KL GCG FL+F+N+YTIQVKL KFHVC QHLLCP CAGIRA+
Sbjct 62 QNDQPNESVRATKSAQKLLGCGNLLFKNFYITIDQVKLSKHFVCNQHLLCPFCAGIRAS 121
```

Query 121 RSMKRYLDRIHELMRQNPRLKPVLTITLVKNGEDLEERYNHLTSSFRKLLSRYRDYKKKG 180
++++Y +R+ E++ + LKPVLIT TVKNG DL ER +HL SFR L+ R RDY KKG
Sbjct 122 KAIQKYTERVDEVLSSKRTLPVLITFTVKNGTDLGERSHLMKSFRITLMERRRDYLLKG 181

Query 181 WGFNQFCKIDGAFYTTTEYTYNEKTKQWHPHIHIFALLNEWIDQEELSETWHDITLDSYII 240
GFN+FKKI+GA Y+ E TYNE+TK+WHPH+H+FALL++WIDQEELS+ WH IT DS I+
Sbjct 182 RGFNEFCKINGAMYSYENTYNEQTKWHPHLMFALLDDWIDQEELSQYWSITGDSMIV 241

Query 241 DIRRVKKTKEHGYSKAAEVCKYALKFGDLSVENTWDAFLTLKGKRLTGSFSGSMHGVKIP 300
DIR+VKK K GYSKA AEVCKYALKFGDLSVENTW+AF LKGKRL+G+FGS++GVKIP
Sbjct 242 DIRKVKKEKGLGYSKAAAEVCKYALKFGDLSVENTWEAFKILKGKRLSGAFSGSLYGVKIP 301

Query 301 DKATDDMPLE-ELPYLEMFYRFVFDKKSYYNLEITKDVTPTQNSRQSEEAAPTDTTGEN 359
+ D+MP E +LPYLEM Y+FVF KKSYY+L +T+ V P+ + E T GE+
Sbjct 302 ENLADEMPDENDLPYLEMLYKFVFGKKSYYDLTMTTRHVEPRAQDDRPMEKELRRPTEGES 361

>>>>sptrembl:E5LG73_9ZZZZ E5lg73 SubName: Full=Replicase 1-like protein (*Sphinx2.36*);. 4/2011
Length=232

Score = 382 bits (980), Expect = 2e-103, Method: Compositional matrix adjust.
Identities = 189/232 (81%), Positives = 207/232 (89%), Gaps = 2/232 (0%)

Query 49 LQEQYLWTQVDFKSEGEDDLNKAATKLKGCQFLIFRNYTIDQVKLEKHFHVCQGH 108
+QEQLWTQVDFK E + KALKAATKLKGCQFL+FRNYTIDQ+KLEKHFHVCQGH
Sbjct 1 MQEQYLWTQVDFKVGSETSI--KALKAATKLKGCQFLFRNYTIDQIKLEKHFHVCQGH 58

Query 109 LLCPMCAGIRAARSMKRYLDRIHELMRQNPRLKPVLTITLVKNGEDLEERYNHLTSSFRK 168
LLCPMCAGIRAARSM RY+ RI E+MRQN +LKPVLITLVKNGEDL+ER+ HLT SF+
Sbjct 59 LLCPMCAGIRAARSMNRYIQRIEIMRQNRKLPVLITLVKNGEDLQERFEHLTGSKFT 118

Query 169 LLSRYRDYKKKGWGFNQFCKIDGAFYTTTEYTYNEKTKQWHPHIHIFALLNEWIDQEELSE 228
LL RYRD+KKKG GFNQFCKIDG FYTTEYTYNE T+QWHPHIHIFAL+ + IDQEEL+E
Sbjct 119 LLQRYRDFKKKGRGFNQFCKIDGGFFTYTNETTQWHPHIHIFALVTDRIIDQEELAE 178

Query 229 TWHDTLDSYIIDIRRVKKTKEHGYSKAAEVCKYALKFGDLSVENTWDAFL 280
TWHDTLDSYI+DIRRVKKTKEHGY+KAAEVCKYALKF DLS E T+ AF
Sbjct 179 TWHDTLDSYIVDIRRVKKTKEHGYAKAAEVCKYALKFSDLSTEKTFQAFF 230

ProtSweep
A Protein Identification Tool

Results for /home/vir088/rindneu/tsekompl/HCB17.228.402.pep

General Information:

Number of residues: 402

Molecular weight: 47.20 kD

Charge: 2

Isoelectric point: 8.03

[more...](#)

Prediction of protein localization:

cyto	20.5 %
cyto_nucl	12 %
pero	4 %

Identified Protein & Transcript:

Protein:

id: [sptrembl:H6VTN7](#) ACIBA

valuation: similar

scientific organism: *Acinetobacter baumannii*

abstract: H6vtn7 SubName: Full=Replication protein; 11/2013Length=378

isoform:

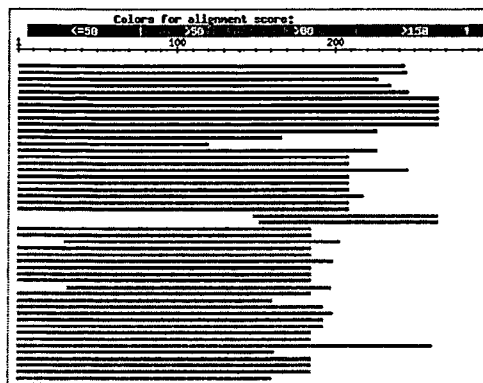
Features & Functions:

GO INFORMATION:

ACCESSION	GROUP	NAME	EVIDENCE
GO:0005727	cellular component	extrachromosomal circular DNA	IEA:InterPro
GO:0003677	molecular function	DNA binding	IEA:InterPro
GO:0006260	biological process	DNA replication	IEA:InterPro

Del Val C., Ernst P., Falkenhahn M., Fladerer., Glatting K.H., Suhai S., Hotz-Wagenblatt A.: ProtSweep, 2Dsweep and DomainSweep: protein analysis suite at DKFZ. Nucleic Acids Research, 2007, 1-7

ORF 290aa:



```
>>>sptrembl:N9TLPO_9GAMM N9tlp0 SubName: Full=Uncharacterize... 343 5e-92
>>>sptrembl:C6RQ49_ACIRA C6rq49 SubName: Full=Putative Repli... 336 7e-90
>>>sptrembl:H6VTN7_ACIBA H6vtn7 SubName: Full=Replication pr... 333 7e-89
```

```
>>>>sptrembl:N9TLPO_9GAMM N9tlp0 SubName: Full=Uncharacterized protein; Flags: Fragment;.
12/2013
Length=399
```

Score = 343 bits (880), Expect = 5e-92, Method: Compositional matrix adjust.
Identities = 159/247 (64%), Positives = 198/247 (80%), Gaps = 2/247 (0%)

```
Query 1 MCAGIRAARSMKRYLDRIHELMRQNPRLKPVLTITLVKNGEDLEERYNHLTSSFRKLLSR 60
CAGIRA++S+++Y DR+ E++ +N +LKPVLIT TVKNG DL ER +HL SFR LL R
Sbjct 89 FCAGIRASKSIQKYTDVRVDEVLKSKNRKLKPVLTITFTVKNGTDLGERSHLLIKSFRTLLER 148

Query 61 YRDYKKKGWGFNQFCKIDGAFYTTTEYTYNEKTKQWHPHIHIFALLNEWIDQEELSETWHD 120
RDY KKG GFN+FKIDGA Y+ E T+NE TK+WHPH+H+FALL+EWIDQE+LSE WHD
Sbjct 149 RRDYLKKGRGFNEFCKIDGAMYSYENTFNESTKEWHPHLHLFALLDEWIDQEQLSELWHD 208

Query 121 ITLDSYIIDIRRVKKTKEHGYSKAVAEVCKYALKFGDLSVENTWDAFLT LKGKRLTGSFG 180
IT DSYI+D+R+VKK+K+ GY+KA AEVCKYALKFGDL+VENTW+AF LKGKRL+G+FG
Sbjct 209 ITGDSYIVDVRKVKKSKDTGYAKAAAEVCKYALKFGDLTVENTWEAFKILKGKRLSGAFG 268

Query 181 SMHGVPKIPDKATDDMPLEELPYLEMFYRFVFDKKSYYNLEITKDVTPQT--NSRQSEEA 238
S++GVKIPD DD+ +LPYLEM Y+FVF KKSYY+L +T+ V PQ N E+
Sbjct 269 SLYGVKIPDNPADDIDDSLPYLEMVYKFVFGKKSYYDLTMTRHVEPQGDNDNDNEEKEG 328

Query 239 PTTDTTG 245
T+D G
Sbjct 329 ATSDRGG 335
```

```
>>>>sptrembl:C6RQ49_ACIRA C6rq49 SubName: Full=Putative Replication protein; Flags: Fragment;.
9/2013
Length=303
```

Score = 336 bits (861), Expect = 7e-90, Method: Compositional matrix adjust.
Identities = 155/245 (63%), Positives = 192/245 (78%), Gaps = 0/245 (0%)

Query 2 CAGIRAARSMKRYLDRIHELMRQNPRLKPVLTITLVKNGEDLEERYNHLTSSFRKLLSR 61
CA IRA+R++++Y++RI +++++N +LKPVLITLTVKNG DL ER HL SFRKLL R
Sbjct 9 CAAIRASRAIQKYVERIDQVLQENRKLKPVLTITLVKNGSDLAERSEHLMKSFRKLLDRR 68

Query 62 RDYKKKGWGFNQFCKIDGAFYTTTEYTYNEKTKQWHPHIHIFALLNEWIDQEELSETWHD 121
RDY+KKG GFN+FCK+ GA Y+ E T+NEKT +WHPHIH+FAL+++WIDQ+E SE WH +
Sbjct 69 RDYEKKGRGFNEFCKVQGAMYSYENTFNEKTGEWHPHIHMFALVDQWIDQDEFSEYWHSL 128

Query 122 TLDSYIIDIRRVKKTKEHGYSKAVA EVCKYALKFGDLSVENTWDAFLT LK GKRLTGSFGS 181
T DS ++D+RR KK-K HGYSKA AEVCKYALKFGDLSVE TW+AF LK GKRLTGSFG
Sbjct 129 TGDSMVVDVRRAKKEKGHGYSKAAAEVCKYALKFGDLSVEKTWEAFKVLK GKRLTGSFGL 188

Query 182 MHGVKIPDKATDDMPLEELPYLEM FYRFVFDKKSYYNLEITKDVT PQTNSRQSEEAAPT 241
+ GVKIPD TDDMP E+LPYLEM Y+F + KKSYY+L IT+ V P+ E+ P
Sbjct 189 LWGVKIPD TDDMP EEDLPYLEM LYKFAYGKKSYYDLLITRHVEPKAKDDNDEKEEPRQ 248

Query 242 DTTGE 246
GE
Sbjct 249 RRGGE 253

>>>>sptrembl:H6VTN7_ACIBA H6vtn7 SubName: Full=Replication protein;. 11/2013
Length=378

Score = 333 bits (853), Expect = 7e-89, Method: Compositional matrix adjust.
Identities = 149/228 (65%), Positives = 186/228 (81%), Gaps = 0/228 (0%)

Query 1 MCAGIRAARSMKRYLDRIHELMRQNPRLKPVLTITLVKNGEDLEERYNHLTSSFRKLLSR 60
CA IRA+R++++Y++RI +++++N +LKPVLITLTVKNG DL ER HL SFR LL R
Sbjct 115 FCAAIRASRAIQKYVERIDQVLQENRKLKPVLTITLVKNGSDLAERSEHLMKSFRLLER 174

Query 61 YRDYKKKGWGFNQFCKIDGAFYTTTEYTYNEKTKQWHPHIHIFALLNEWIDQEELSETWHD 120
RDY+KKG GFN+FCK+ GA Y+ E T+NEKT +WHPHIH+FAL+++WIDQ+E SE WH
Sbjct 175 RRDYEKKGRGFNEFCKVQGAMYSYENTFNEKTGEWHPHIHMFALVDQWIDQDEFSEYWH 234

Query 121 ITLDSYIIDIRRVKKTKEHGYSKAVA EVCKYALKFGDLSVENTWDAFLT LK GKRLTGSFG 180
+T DS ++D+RR +K K +GYSKA AEVCKYALKFGDLSVE TW+AF LK GKRLTGSFG
Sbjct 235 LTGDSMVVDVRRARKEKGYGYSKAAAEVCKYALKFGDLSVEKTWEAFKVLK GKRLTGSFG 294

Query 181 SMHGVKIPDKATDDMPLEELPYLEM FYRFVFDKKSYYNLEITKDVT PQ 228
+ GVKIPD TDDMP E+LPYLEM Y+F + KKSYY+L IT+ V PQ
Sbjct 295 LLWGVKIPDSMTDDMP SEDLPYLEM LYKFAYS KKSYYDLLITRHVEPQ 342

>>>>sptrembl:E5LG73_9ZZZZ E5lg73 SubName: Full=Replicase 1-like protein (*Sphinx2.36*);.
4/2011
Length=232

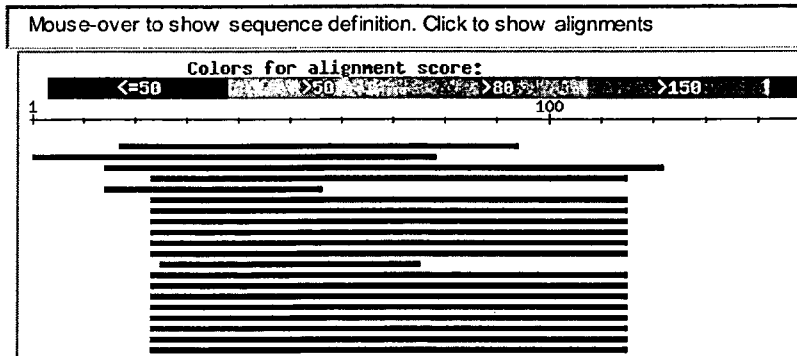
Score = 271 bits (694), Expect = 2e-70, Method: Compositional matrix adjust.
Identities = 136/168 (80%), Positives = 150/168 (89%), Gaps = 0/168 (0%)

Query 1 MCAGIRAARSMKRYLDRIHELMRQNPRLKPVLTITLVKNGEDLEERYNHLTSSFRKLLSR 60
MCAGIRAARSM RY+ RI E+MRQN +LKPVLITLTVKNGEDL+ER+ HLT SF+ LL R
Sbjct 63 MCAGIRAARSMNRYIQRIEIMRQNRKLKPVLTITLVKNGEDLQERFEHLTGSFKTLLQR 122

Query 61 YRDYKKKGWGFNQFCKIDGAFYTTTEYTYNEKTKQWHPHIHIFALLNEWIDQEELSETWHD 120
YRD+KKG GFNQFCKIDG FYTTEYTYNE T+QWHPHIHIFAL+ + IDQEEL+ETWHD
Sbjct 123 YRDFKKKGRGFNQFCKIDGGFYTTEYTYNETTQWHPHIHIFALVTDRI DQEELAEWHD 182

Query 121 ITLDSYIIDIRRVKKTKEHGYSKAVA EVCKYALKFGDLSVENTWDAFL 168
ITLDSYI+DIRRVKKTKEHGY+KAVA EVCKYALKF DLS E T+ AF
Sbjct 183 ITLDSYIVDIRRVKKTKEHGYAKAVA EVCKYALKFSDLSTEKTFQAFF

ORF 145aa:



```
>>>swissprot:CAPSD NMV P15100 narcissus mosaic virus (nmv). ... 38.9 0.85
>>>sptrembl:G0L9E1 ZOEGA G019e1 SubName: Full=Uncharacterize... 38.9 0.94
>>>>swissprot:CAPSD NMV P15100 narcissus mosaic virus (nmv). recname: full=coat protein;
altname: full=capsid protein; short=cp;. 10/2013
Length=240
```

Score = 38.9 bits (89), Expect = 0.85, Method: Compositional matrix adjust.
Identities = 28/78 (35%), Positives = 40/78 (51%), Gaps = 8/78 (10%)

```
Query 18  SDEDLKPITIRLADSEIQAYDTVAKSMGLTRQDFLSHVIRSNFRQALKEFVVGYTSSNPS 77
          +D ++P + R   EIQA+ T AK  L RQ +   + ++F  LK  VG  S P
Sbjct 167  ADGLIRPPSQR----EIQAHS- AKYGALARQRYR---METSFPPLKSLTVGSAVSTPC 218

Query 78  TPLSVLFNSNTDNDEVKL 95
          TPL  L N N + ++KL
Sbjct 219  TPLKHLQNCNRNTSKLKL 236
```

```
>>>>sptrembl:G0L9E1_ZOEGA G019e1 SubName: Full=Uncharacterized protein;. 5/2013
Length=291
```

Score = 38.9 bits (89), Expect = 0.94, Method: Compositional matrix adjust.
Identities = 28/84 (33%), Positives = 42/84 (50%), Gaps = 5/84 (5%)

```
Query 1  MLEKSLGVAIATNNASNSDEDLKPITIRLADSEIQAYDTVA-KSMGLTRQD----FLSHV 55
          LF+S+ V I TNN  ++ L  +++ L  S I  Y TV  K M L R+D  F +
Sbjct 83  FLFRSIAVLITNNWDMTEIFLFALSVLLLISCIVYYTVPDKEMILNRRDGLVSFPAPT 142

Query 56  IRSNFRQALKEFVVGYTSSNPSTP 79
```

(ril.10E.1.5)

[illegible]

104aa – replication-associated protein 63% similarity to Gemycircularvirus 8

Figure 3(a)

Query	1820	AGTCCTCCAGCGACAACATCACCATCCTTGATCGCGTAATCATAATGGACTTGAGGGTCG	1879
Sbjct	1893	AGCCCTCCAGCGACCACGCTCTCCATCTTTGATCGCATAGTCATAACCGCCTTCCGGCCTT	1952
Query	1880	CGCTTAG-TAATTGAGATGTTAGGGTGAAAACCTCCAACATCAAAGACAGCAGCACGTCT	1938
Sbjct	1953	CCTCGAGATGCTTCA-ATGTTAGGGTGGAAGCCTCCAACATCAAACACATCGGGTCTTCG	2011
Query	1939	TGATTGAAATCGTCGTCCAAAATCCACAAAAGCGTGCAGATGAGTTCGCCATCAGAGTG	1998
Sbjct	2012	GGATCGAAACTTTCCGGGAGAAATCGCAGAAAGCATGGAGATGAGTACCCCCATCAGCGTG	2071
Query	1999	GGCTTCTCTTCCAATGATGCATTCTCCTCCAAGTCCACCAATGTGTGAAACAATAGTGAA	2058
Sbjct	2072	ATGCTCTCGTCCAATGATACACTACCGCCAAGGAATGCCAGATGGTCATTAAACGGCCCA	2131
Query	2059	TGGGTCGAGGCCGCCGATTGAGCGTAGGTGAGAAGCGCATATCGGGCATAGAAGCGAAA	2118
Sbjct	2132	TGCATCAAGGTGCGCCGCACTGTGCGTAGGTAAGGAGGACATAGCGTGCTTGAATTGAAA	2191
Query	2119	T	2119
Sbjct	2192	T	2192

>>>>nrrnuc:GI_537368005 Gi|537368005|gb|KF371642.1| Gemycircularvirus 1b isolate P22,
complete genome. 0/0
Length=2160

Score = 114 bits (126), Expect = 3e-21
Identities = 197/284 (69%), Gaps = 4/284 (1%)
Strand=Plus/Plus

```
Query 1823 CCTCCAGCGACAACATCACCATCCTTGATCGCGTAATCATAATGGACTTGAGGGTCGCGC 1882
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 1825 CCTCCAGCGACCACGTCTCCATCTTTGATCGCATAGTCATAACCGCCTTCCGGGCTTCCG 1884

Query 1883 TTAGTAATTGAG--ATGTTAGGGTGGAAACCTCCAACATCAAAGACAGCAGCAGCTCTTG 1940
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 1885 TAAG--ATGGAACAAGGTTCCGGATGGCGGCTCCAACATCAAAGACATCGGATCTTCTGG 1942

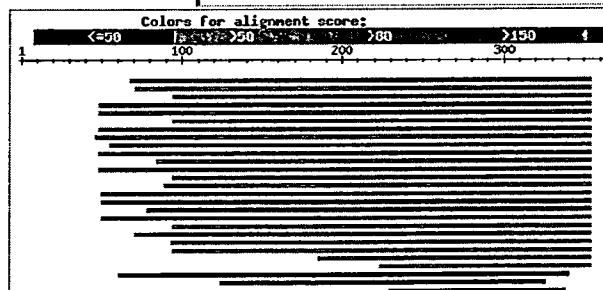
Query 1941 ATTGAAATCGTCGTCCAAAATCCACAAAAGCGTGCAGATGAGTTCCGCCATCAGAGTGGG 2000
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 1943 ATCGAAACTTCCGGGAGAAATCGACGAAAGCATGGAGATGAGTACCTCCATCAGCATGCA 2002

Query 2001 CTTCTCTTCCAATGATGCATTCTCCTCCAAGGTCAACCAATGTGTGAAACAATAGTGAATG 2060
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 2003 ATTCTCGTCCAATGATACACTCACCGCCAAGAAATGCCATATGGTCATTAACGGCCCATG 2062

Query 2061 GGTGAGGCCGCCGCGATTGAGCGTAGGTGAGAAGCGCATATCGG 2104
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 2063 CATCAAGGTCGCCGCACTGTGCGTAGGTAAGGAGGGCATAACGG 2106
```

BlastP - Uniprot:

ORF 359aa:



>>>sptrembl:K0A2I9_9VIRU K0a2i9 SubName: Full=Putative coat ... 129 3e-27
>>>sptrembl:S5RC76_9VIRU S5rc76 SubName: Full=Putative coat ... 124 9e-26
>>>sptrembl:K0A169_9VIRU K0a169 SubName: Full=Putative coat ... 123 1e-25

>>>sptrembl:K0A2I9_9VIRU K0a2i9 SubName: Full=Putative coat protein (Dragonfly-associated virus 1);. 5/2013

Length=290

Score = 129 bits (323), Expect = 3e-27, Method: Compositional matrix adjust.
Identities = 94/289 (32%), Positives = 141/289 (48%), Gaps = 18/289 (6%)

```
Query 69 STFRRRSF--RNGRSRTTRYSRPRTNVRVRNLASRKRCDNMICTPIDDTGTPQTPGPV 126
          S+ +RRS+ R SR T + RT+ RR+ N++S+K R+N + G PG
Sbjct 17 SSKRRRSYVGRRLSSRRTTKTVMKRTSRRRILNISSQKKRNNAPVNFNVLGESPIPGSK 76

Query 127 SMDASSVYGFIPTARTAGTQLESGERVNPSSAQRKSKCYVRGYGETVEIQTDASP 186
          M + ++ PTA T T + S SA R + Y RG E E+ T +
Sbjct 77 VMGSDRTTYLLWRPTALTDDTDIYS-----SASRGAQQIYWRGIKERAEGVTSSGAA 128

Query 187 WVWRRIVFSTIGLASQFDPGLLYNNDETRGRSLFNIRSGTQQASILEGVVFRYLFQGT 246
          W WRRIVF L GL+ N + + G+ R++ + SG A+ L + +LFQG+
Sbjct 129 WRWRRIVFEAKSLNIN---GLISNVETSAGFNAMVEL-SGDPPAN-LRNSLESFLFQGS 183

Query 247 EGVGDWNVMTATTAPRHVRVYSDRRRVIQSPNSEGGTIRLSKHYDGINRSIIYDDVEIGA 306
          GVDW V A R+ SD+ R +QS N G + + +N++++Y+D E+G
Sbjct 184 VGVGDWNTVFNAKVDNNRCRLSDKVRHLQSGNDRGRFYTF-RQWIPLNKTMYNDDEVG- 241
```

Query 307 GAKATNAIASGTPYGNMGDLFVLDFFSSASDDETSGASFQPHGKFYWHE 355
 K +N AS G MGD++V D F + D + +F P YWHE
 Sbjct 242 DDKQSNTFASPGRKG-MGDVYVYDMFQCLTSDAINTLNFNQPATLYWHE 289

>>>>sptrembl:S5RC76_9VIRU S5rc76 SubName: Full=Putative coat protein;. 12/2013
 Length=304

Score = 124 bits (310), Expect = 9e-26, Method: Compositional matrix adjust.
 Identities = 93/295 (31%), Positives = 136/295 (46%), Gaps = 29/295 (9%)

Query 72 RRRSFRNGRSRTTRYSRPRTNVRVRNLASRKCRDNM-----ICTPIDDTGTPQTPG 124
 RRRS+ R RRY R +T +R+ N++S K RDNM + PI+ G G
 Sbjct 27 RRRSYPTTR---RRYPR-KTLAKRILNISSTKKRDNMMSYTNVTVANPIN--GQTYAAG 79

Query 125 PVSMDASSVYGFIPTARTAGTQLESGERVNAPS-SAQRKSKCYVRGYGETVEIQTD 183
 P + Y F + PTAR + + N P A R + CY+RG E + ++T+
 Sbjct 80 PAVLRGDQTYIFAWIPTARDL---IRADGEPNTPVWLAGRTRQTCYMRGLKEAMNVRTNS 136

Query 184 ASPWVWRRIVFSTIG--LASQFDPGLLYNNDETRGRSLFNIRSGTQQASILEGVVFRY 241
 S W WRR I F+ G L G + + G R++ N+ + L +
 Sbjct 137 GSSWQWRRICFTMKGPNLIGYATTGSNLWQETSNGMVRTMNNVNT-----KPLGNEILAQ 191

Query 242 LFQGTGEGVDWANVMTATTAPRHVRVYSDRRRVIQSPNSEGGTIRLSKHYDGINRSIIYDD 301
 LF+G + DW NV A T + DR I S N E G +R KH+ +N++++Y+D
 Sbjct 192 LFRGAQNQDNLVFNATQDTTKFTIKYDRTISIASGN-EQGLVRTYKHWYPMKNLVYND 250

Query 302 VEIGAGAKATNAIASGTPYGNMGDLFVLDFFSSASD-DETSASFQPHGKFYWHE 355
 E G + +G P MGD +V+DF A D + + P YWHE
 Sbjct 251 DENADGVLTSYSTNGKP--GMGDYVVDFFQPADGADASDTLALLPEATLYWHE 303

>>>>sptrembl:K0A169_9VIRU K0a169 SubName: Full=Putative coat protein;. 5/2013
 Length=289

Score = 123 bits (309), Expect = 1e-25, Method: Compositional matrix adjust.
 Identities = 89/267 (33%), Positives = 131/267 (49%), Gaps = 27/267 (10%)

Query 95 RVRNLASRKCRDNMICTPIDDTG-TPQTPGPVSMDSVYGFIPTARTAGTQLESGE 153
 RR+ N++S+K RDNMI TG P G ++ A+ ++ TAR +
 Sbjct 43 RRILNVSSKKKRDNMINSVFSPTGINPAVRGFIATGAAPTM-LVWCATAR-----D 92

Query 154 RV----NAPSSAQRKSKCYVRGYGETVEIQTDASPWVWRRIVFSTIGLASQFDPGLLY 209
 RV N + + R CY+RG E + I T++ + W+WRR I F+ GL S G
 Sbjct 93 RVSNASNPTADSVRESQTCYMRGLKEHIRINTNNPTSWSLWRRICFTAKGLFSGS--GTSA 150

Query 210 NNDET-RGWGRSLFNIRSGTQQASILEGVVFRYLFQGTGEGVDWANVMTATTAPRHVRVYS 268
 ++ ET GW R+L + ++GT + L + LF GT GVDW + TA T +V++
 Sbjct 151 DSVETSNGWVRALMD-QTGSFSGTSLTNL---LFAGTNGVDWNTFTANTDSTNVKIMY 205

Query 269 DRRRVIQSPNSEGGTIRLSKHYDGINRSIIYDDVEIGAGAKATNAIASGTPYGNMGDLFV 328
 D+ I S S G IR K + G+N++I+Y + E G G G P MGD +V
 Sbjct 206 DKTFTINSVGS--GVIRDYKMWGMNKNIVYGEDENGEINNRYSTIGRP--GMGDYV 261

Query 329 LDFSSASDDETSGASFQPHGKFYWHE 355
 +D +S S +F YWHE
 Sbjct 262 VDLLASNSSAAADTLTFSTEATLYWHE 288

ProtSweep
A Protein Identification Tool

Results for /home/vir088/rindneu/gemi/HCBI8.215.359.pep

Number of residues: 359	Molecular weight: 40.45 kD
Charge: 25	Isoelectric point: 11.19

[more...](#)

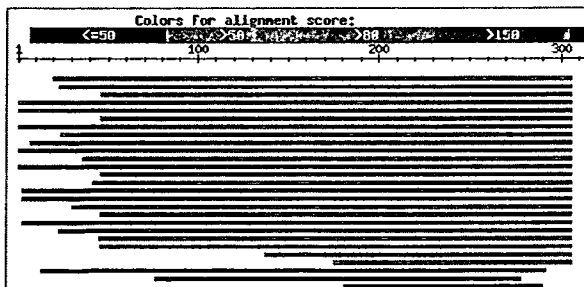
Prediction of protein localization:

extr	19%
E.R.	6%
plas	3%
E.R. mito	3%

Identified Protein & Transcript:

Protein:
Id: <u>sptrembl:T1YSG7_9VIRU</u>
valuation: putative
scientific organism: Gemycircularvirus 1c
abstract: T1ysg7 SubName: Full=Capsid protein; 2/2014Length=307
isoform:

ORF 311aa:



```
>>>sptrembl:K0A2I9_9VIRU K0a2i9 SubName: Full=Putative coat ... 128 3e-27
>>>sptrembl:S5RC76_9VIRU S5rc76 SubName: Full=Putative coat ... 124 7e-26
>>>sptrembl:K0A169_9VIRU K0a169 SubName: Full=Putative coat ... 123 1e-25
```

```
>>>>sptrembl:K0A2I9_9VIRU K0a2i9 SubName: Full=Putative coat protein (Dragonfly-associated
circular virus 1);. 5/2013
Length=290
```

```
Score = 128 bits (322), Expect = 3e-27, Method: Compositional matrix adjust.
Identities = 94/289 (32%), Positives = 141/289 (48%), Gaps = 18/289 (6%)
```

```
Query 21 STFRRRSF--RNGRSRTTRRYSRPRTNVRRVRNLSRKRCDNMICTPIDDTGTPQTPGPV 78
S+ +RRS+ R SR T+ RT+ RR+ N++S+K R+N + G PG
Sbjct 17 SSKRRRSYVGRRLSSRRTTKTVMKRTSRRRILNISSQKKRNNAPVNFNYLGESPIPGSK 76

Query 79 SMDASSVYGFIFSPARTAGTQLESGERVNPSSAQRKSKCYVRGYGETVEIQTDDASP 138
M + ++ PTA T T + S SA R + Y RG E E+ T +
Sbjct 77 VMGSDRTTYLWRPTALTDDTDIYS-----SASRGAQQIYWRGIKERAEVGTSSGAA 128

Query 139 WVRRRIVFSTIGLASQFDPGLLYNNDETRGWGRSLFNIRSGTQQASILEGVVFRYLFQGT 198
W WRRIVE L GL+ N + + G+ R++ + SG A+ L + +LFQG+
Sbjct 129 WRWRRIVFEAKSLNIN---GLISNVETSAGFNAMVEL-SGDPPAN-LRNSLESFLFQGS 183
```

Query 199 EGVGDWANVMTATTAPRHVRVYSDDRRVQSPNSEGGTIRLSKHYDGINRSIIYDDVEIGA 258
 GVDW V A R+ SD+ R +QS N G + + +N++++Y+D E+G
 Sbjct 184 VGVDDWNTVFNAKVDNNRCRLSDKVRHLQSGNDRGRFYTF-RQWIPLNKTVMVYNDDEVG- 241

Query 259 GAKATNAIASGTPYGNMGDLFVLDFSSASDDETSASGASFPHGKFWHE 307
 K +N AS G MGD++V D F + D + +F P YWHE
 Sbjct 242 DDKQSNTFASPGRKG-MGDVYVYDMFQCLTSDAINTLNFPQATLYWHE 289

>>>sptrembl:S5RC76_9VIRU S5rc76 SubName: Full=Putative coat protein;. 12/2013
 Length=304

Score = 124 bits (310), Expect = 7e-26, Method: Compositional matrix adjust.
 Identities = 94/295 (31%), Positives = 136/295 (46%), Gaps = 29/295 (9%)

Query 24 RRRSFRNGRSRTTTRYSRPRTNVRVRNLSRKRCDNM-----ICTPIDDTGTPQTPG 76
 RRRS+ R RRY R +T +R+ N++S K RDNM + PI+ G G
 Sbjct 27 RRRSYPTTR----RRYPR-KTLAKRILNISSTKKRDNMMSYTNVTVANPIN--GQTYAAG 79

Query 77 PVSMDASSVYGFI FSP TARTAGTQLESGERVNAPS-SAQRKSKCYVRGYGETVEIQTD 135
 P + Y F + PTAR + + N P A R + CY+RG E + ++T+
 Sbjct 80 PAVLRGDQTYIFAWIPTARDL---IRADGEPNTPVWLAGRTRQTCYMRGLKEAMNVRTNS 136

Query 136 ASPWVWRRIVFSTIG--LASQFDPGLLYNNDETRGRWGRSLFNIRSGTQQASILEGVVFRY 193
 S W WRR I F+ G L G + + G R++ N+ + IL
 Sbjct 137 GSSWQWRRICFTMKGPENLIGYATTGNSNLWQETSNGMVRMTMNNVNTKPLGNEILAQ----- 191

Query 194 LFQGTGEGVDWANVMTATTAPRHVRVYSDDRRVQSPNSEGGTIRLSKHYDGINRSIIYDD 253
 LF+G + DW NV A T + DR I S N E G +R KH+ +N++++Y+D
 Sbjct 192 LFRGAQNQDWLVNFAQTDTTKFTIKYDRTISIASGN-EQGLVRYTKHWYPMNKNLVYND 250

Query 254 VEIGAGAKATNAIASGTPYGNMGDLFVLDFSSASD-DETSASGASFPHGKFWHE 307
 E G + +G P MGD +V+DFF A D + + P YWHE
 Sbjct 251 DENADGVLTSYTSTNGKP--GMGDYVYVDFQPADGADASDTLALLPEATLYWHE 303

>>>sptrembl:K0A169_9VIRU K0a169 SubName: Full=Putative coat protein;. 5/2013
 Length=289

Score = 123 bits (309), Expect = 1e-25, Method: Compositional matrix adjust.
 Identities = 89/267 (33%), Positives = 131/267 (49%), Gaps = 27/267 (10%)

Query 47 RRVRLASRKRCDNMICTPIDDTG-TPQTPGPVSM DASSVYGFI FSP TARTAGTQLESGE 105
 RR+ N++S+K RDNMI TG P G ++ A+ ++ TAR +
 Sbjct 43 RRILNVSSKKRDNMINVSFSP TGINPAVRGFIATGAAPTM-LVWCATAR-----D 92

Query 106 RV----NAPSSAQRKSKCYVRGYGETVEIQTDASPVWVRRIVFSTIGLASQFDPGLLY 161
 RV N + + R CY+RG E + I T++ + W+WRR I F+ GL S G
 Sbjct 93 RVSNASNPTADSVRESQTCYMRGLKEHIRINTNNPTS WLWRRICFTAKGLFSGL--GTSA 150

Query 162 NNDET-RGWGRSLFNIRSGTQQASILEGVVFRYLFQGTGEGVDWANVMTATTAPRHVRVYS 220
 ++ ET GW R+L + ++GT + L + LF GT GVDW + TA T +V++
 Sbjct 151 DSVETSNGWVRLMD-QTGTSGTSLTNL----LFAGTNGVDWNDTFTANTDSTNVKIMY 205

Query 221 DRRRVQSPNSEGGTIRLSKHYDGINRSIIYDDVEIGAGAKATNAIASGTPYGNMGDLFV 280
 D+ I S S G IR K + G+N++I+Y + E G G G P MGD +V
 Sbjct 206 DKTFITNSGV--GVIRDYKMWGMKNINIVYGEDENGEINNRYSTIGRP--GMGDYV 261

Query 281 LDFSSASDDETSASGASFPHGKFWHE 307
 +D +S S +F YWHE
 Sbjct 262 VDLLASNSSAAADTLTFSTEATLYWHE 288

ProtSweep
A Protein Identification Tool

Results for /home/vir088/rindneu/gemi/HCB18.215.311.pep



Number of residues: 311	Molecular weight: 35.07 kD
Charge: 20	Isoelectric point: 11.40

[more...](#)

Prediction of protein localization:

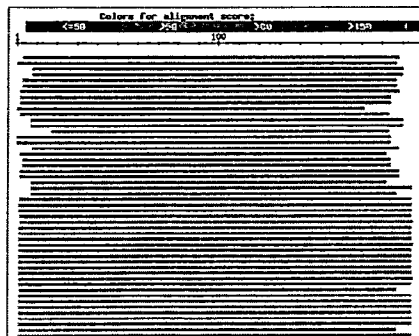
nucl	28 %
mito	3 %



Identified Protein & Transcript

Protein:
id: sptrembl:F6KIE3_9VIRU
valuation: putative
scientific organism: Mosquito VEM
abstract: F6kie3 SubName: Full-Coat protein; 5/2013Length=327
isoform:

ORF 196aa:



```
>>>sptrembl:TLYSG3_9VIRU Tlysg3 SubName: Full=Replication as... 199 5e-49
>>>sptrembl:TLYSF1_9VIRU Tlysf1 SubName: Full=Replication as... 189 5e-46
>>>sptrembl:TLYQU3_9VIRU Tlyqu3 SubName: Full=Replication as... 189 6e-46
```

```
>>>>sptrembl:TLYSG3_9VIRU Tlysg3 SubName: Full=Replication associated protein
Gemycircularvirus 4);. 2/2014
```

Length=338

Score = 199 bits (506), Expect = 5e-49, Method: Compositional matrix adjust.
Identities = 101/186 (54%), Positives = 120/186 (64%), Gaps = 0/186 (0%)

```
Query 4 FRFYARYALTYAQCGGLDPFTIVSHIGDLGGECIIGREAHSDGGTHLHAFVDFGRRFQS 63
      F+F ARY LLTYAQCG LD + + H+ LGGECIIGRE H+DGGTHLHAF DF R+F+S
Sbjct 6 FQFQARYVLLTYAQCGDLDAWAVNDHLAFLGGECIIGREHADGGTHLHAFCDFSRKFRS 65

Query 64 RRAAVFDVGGFHPNINISITKRDPQVHYDYAIKGDGVVAGGLERPAADASAGVENKHHQIVL 123
      RR VFDVGGFHPNI ++ P+ YDYAIKGDGVVAGGL RP + IV
Sbjct 66 RRPDVFVGGFHPNIEASRGRPEGGYDYAIKGDGVVAGGLSRPGRGVYEDVSSWSIIVS 125

Query 124 APTRAEFFEAVRDLDPRLLLTSPFSLKYADWHYRVDPPYRHDPEYEFDTSQFPELNQW 183
      + +F+E V LDPR L T++ SL YA+W YR P Y H EF+ PEL W
Sbjct 126 QESEGDFWECVARLDPRSLCTNYSNLRAYANWKYRPTPVQYVHPAGVEFELGMVPELAVW 185

Query 184 AEDYIG 189
      E +G
Sbjct 186 REIALG 191
```

>>>>sptrembl:TlYSF1_9VIRU TlYSf1 SubName: Full=Replication associated protein;. 4/2014
Length=325

Score = 189 bits (480), Expect = 5e-46, Method: Compositional matrix adjust.
Identities = 100/191 (52%), Positives = 118/191 (61%), Gaps = 3/191 (1%)

```
Query 1  MSTFRFYARYALLTYAQCGLDPFTIVSHIGDLGGECIIGREAHSDGGTHLHAFVDFGRR 60
          MS+FRF ARYALLTYAQCGLDPF + H+ L ECIIGRE H+D G HLHAFVDFG +
Sbjct 1  MSSFRFQARYALLTYAQCGLDPFAVSDHLSGLRAECIIGRETHADEGIHLHAFVDFGVK 60

Query 61 FQSRRAAVFDVGGFHPNISITKRDPQVHYDYAIKGDVAVAGGLERPAADASAGVEN---K 117
          F +R VFDV G HPN+ ++ P YDYAIKGDVAVAGGL+RPA D GV K
Sbjct 61 FSTRNVRVFDVQGHHPNVEPSRGRPAGGYDYAIKGDVAVAGGLDRPADDGGVGVSKTTQK 120

Query 118 HHQIVLAPTRAEFFEAVRDLDPRLLLSFSPSLEKYADWHYRVDPPYRHDPEYEFDTSQF 177
          IV A TR EF+ + +LDP ++ SF YAD +R P+ Y DT
Sbjct 121 WATIVGAQTRDEFWALLAELDPGAMVRSFTQCRAYADHRFRESPEYVTPESIIIDTGHV 180

Query 178 PELNQWAEDYI 188
          LN+WA I
Sbjct 181 DGLNEWALSNI 191
```

>>>>sptrembl:TlYQU3_9VIRU Tlyqu3 SubName: Full=Replication associated protein;. 2/2014
Length=340

Score = 189 bits (479), Expect = 6e-46, Method: Compositional matrix adjust.
Identities = 97/183 (53%), Positives = 119/183 (65%), Gaps = 0/183 (0%)

```
Query 9  RYALLTYAQCGLDPFTIVSHIGDLGGECIIGREAHSDGGTHLHAFVDFGRRFQSRRAAV 68
          RYALLTYAQCGLD + + H+ LGGECIIGRE H+DGGTHLHAFVDF R+F+SRR+ V
Sbjct 13 RYALLTYAQCGLDAWAVNDHMAFLGGECIIGRELHADGGTHLHAFVDFSRKFRSRSDV 72

Query 69 FDVGGFHPNISITKRDPQVHYDYAIKGDVAVAGGLERPAADASAGVENKHHQIVLAPTRA 128
          FDVGG HPN+ + P+ YDYAIKGDVAVAGGL RP+ + IV +
Sbjct 73 FDVGGHRPNLVPYSGSPGGYDYAIKGDVAVAGGLSRPSGRGVYEDVSSWSIIVSQSEGE 132

Query 129 EFFEAVRDLDPRLLLSFSPSLEKYADWHYRVDPPYRHDPEYEFDTSQFPELNQWAEDYI 188
          +F+E V LDPR L T++ SL YA+W Y+ P Y H +F+ PEL W E +
Sbjct 133 DFWEVARLDPRSLCTNYSNLRAYANWKYQPLPVRYEHPAGVKFELGMVPELAVWREVAL 192

Query 189 GWD 191
          G D
Sbjct 193 GVD 195
```

ProtSweep
A Protein Identification Tool

Results for /home/vir088/rindneu/gem/HCBI8.215.196r.pep

Number of residues: 196

Molecular weight: 22.17 kD

Charge: -11

Isoelectric point: 5.31

[more...](#)

Prediction of protein localization:

mito	15 %
cyto nucl	9.5 %
cyto	9 %

Identified Protein & Transcript:

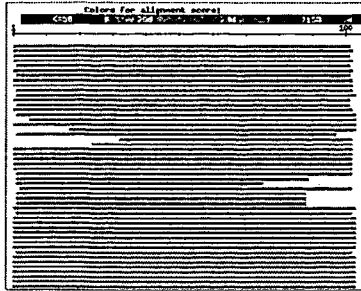
Protein:

id: sptrembl:Q06FD4_9GEMI

valuation: putative

scientific organism: Tomato mosaic

abstract: Q06fd4 SubName: Full=Replication protein; Flags: Fragment;. 4/2014Length=220

ORF 104aa:

```
>>>sptrembl:T1YRW0_9VIRU T1yrw0 SubName: Full=Replication as... 139 4e-31
>>>sptrembl:K0A1L7_9VIRU K0a1l7 SubName: Full=Spliced replic... 137 2e-30
>>>sptrembl:T1YSF1_9VIRU T1ysf1 SubName: Full=Replication as... 127 2e-27
```

```
>>>>sptrembl:T1YRW0_9VIRU T1yrw0 SubName: Full=Replication associated protein
(Gemycircularvirus 8);. 2/2014
Length=320
```

```
Score = 139 bits (351), Expect = 4e-31, Method: Composition-based stats.
Identities = 65/103 (63%), Positives = 77/103 (74%), Gaps = 1/103 (0%)
```

```
Query 1 MWARSLAKHAYFGGLFSLEESLEDVKYAVFDDIAGGLKFFPQYKSWLGNQLDFYSTDKYK 60
+WARSL HAYFGGLFSLEE LE V YA+FDD+ GGLKFF YK WLG Q F++TDKY+
Sbjct 212 LWARSLGNHAYFGGLFSLEEELEGVDYAI FDDMQGGLKFFHSYKFWLGAQSQFWATDKYR 271

Query 61 KKKAIKWGKPCIWISNEYIPIT-EGVDQDWMKNCIFVEVTTPL 102
KK I WG+P I+I N+ P+ EGVD DW+ NC FVE+T L
Sbjct 272 GKGLIHWGRPSIYICNQNLCDGVDHDWLGNCDFVEITESL 314
```

```
>>>>sptrembl:K0A1L7_9VIRU K0a1l7 SubName: Full=Spliced replication-associated protein;.
4/2014
Length=318
```

```
Score = 137 bits (345), Expect = 2e-30, Method: Composition-based stats.
Identities = 61/104 (58%), Positives = 78/104 (75%), Gaps = 1/104 (0%)
```

```
Query 1 MWARSLAKHAYFGGLFSLEESLEDVKYAVFDDIAGGLKFFPQYKSWLGNQLDFYSTDKYK 60
+WARSL +HAYFGGLF L+ESLEDV YAVFDD+ GGL+FF YK WLG Q FY+TDKY+
Sbjct 210 LWARSLGRHAYFGGLFCLDESLEDVDYAVFDDMQGGLFFHAYKFWLGAQKQFYATDKYR 269

Query 61 KKKAIKWGKPCIWISNEYIPITE-GVDQDWMKNCIFVEVTTPLY 103
K+ + WG+P I+ISN P+ + G D DW+ NC+ V V P++
Sbjct 270 GKQLVNWGRPSIYISNTNPLADKGADVDWLMGNCVIVHVD RPIF 313
```

```
>>>>sptrembl:T1YSF1_9VIRU T1ysf1 SubName: Full=Replication associated protein;. 4/2014
Length=325
```

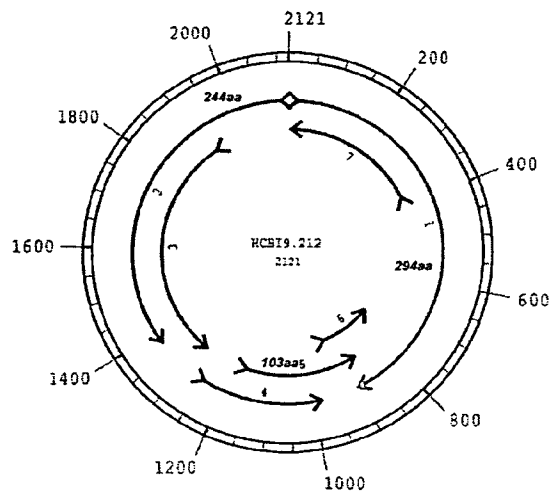
```
Score = 127 bits (319), Expect = 2e-27, Method: Composition-based stats.
Identities = 59/106 (55%), Positives = 76/106 (71%), Gaps = 5/106 (4%)
```

```
Query 1 MWARSLAKHAYFGGLFSLEESLEDVKYAVFDDIAGGLKFFPQYKSWLGNQLDFYSTDKYK 60
+WARSL HAYFGGLFS++E + YAVFDD GGLKF P YK W+G+Q FY TDKYK
Sbjct 215 VWARSLGDHAYFGGLFSMEDISSATYAVFDDF-GGLKFLPSYKFWMGHQQFYVTDKYK 273

Query 61 KKKAIKWGKPCIWISNEYIPITE-GVDQD---WMDKNCIFVEVTTPL 102
KK ++WG+P IWI+NE P E GV D W+++NCIF+ + + +
Sbjct 274 GKKLVEWGRPSIWINNEDPREEHGVRCDIEWLEQNCIFIRLDSTI 319
```

HCBI9.212 (2121bp)
(*rkg.S1.8B.1.2*)

ATTTTTCGACCAGTAAAAAATGGCTTACGGATTCCGCCGCCGCTTTGGTCGCTCCCGTCGCTCGTACCGACGAAG
 GCGCGTTCCGCCGTATTCCCGAAGGGGTAGAAGCGGGTATGGTCGTGGCGGTTACAACCGTCGGCGTTATCGTCG
 TCGTCGCCGCTCTTGGATTAGTCCTTACAACAAAAAGCATGACGTGATTAAGGGGGCGCATGATTCTGAATGATAA
 TTCGTTCCGCCCTATAGGTGCCCCAACACCTTTTTTGGGTTCTGCCCAACCTACATGCCCAACGCCGTTGCTGG
 GGATACTGGTCACAATCCTCAAAAGGTTCCGACCAGTAAGAATGTTGGGTTCAGTGGATATCAGGAACGTGTTCA
 TATTGCTACGGCTGAATCTTTGATTGCGCTCGCATGGTTATTTGGTCATATAACCGGCTTGAGAATACGGCCGG
 TCCGACGAAGACGGATGATCAGGGTACTTCGTACACGACACGTGAGCTTACCCCGCTCCAATTAAACGAAGGGGT
 CCGTGCCTTTTTGTTCCGAGGTACTCAGGGTATCGATTACACTGAGAACACCTTGCATGAGGCTCCGCTGAATTC
 CGATCACTTTACCGTTGCGTATGATCGTACACGTGTGTTACAGCCTCGTCGTCTACTGGTGACGAATACGGTCA
 TATTTTAAATATGAAGTTCTGGAACCTGGTGAAGGATTATTTACTCGGATGATGAGGATGGTTTACAGAAGGG
 TCCTCTTGTGAACGGATGGTCTTCTCTTGGACGTGTGTCTAAGGGAACATGTATATTTTGGATGTCTTTTCTAC
 GGGGTATCTTTGAGTAGTAATACTCCTTCTATTGCTAGATTGCTCCGACTGGGACTCGTTACTGGACGGAGGG
 TTAGATAACAATACAAATGGATCCACCACATTGATTATGTGACAGTTGTTTATAAGCCAATCGTAGTCCACCCCT
 TTAGCGAAGTGTGGTCTCGTTTCATGCACATGATTGAGGGTTTGCCCCAGTAGATCGTTCGCTTCTTTTCGGTAC
 TTGTCCGTGATGACGAATTGTTTTTGTGCGCCTAGCCACCCCTTGTATGAGTGGAAGTATTCGAATCCTCCCTGG
 ATGTCGTGGAAGACGGCGTATTGACAATCGTCGGAGAATTTCGTCCATGTTGAATTGTAGGTTGAAGTACGCATGG
 TGTCCCAACGATCTGGCGTATAGCGTCTTTCCGGTTCTTGTGGTCCCATAAAATGAGTGATTTAACTCTGCCC
 AAGTAAAGATCAGTAATGGGGTTCGGGGCGAAGCCCTGAAGCGGACAAGATTTGATCGTGCGAGGACGCTACGTC
 GGCCGCAGCGCCAGCAGACTTACCTCTCGGGTCTACAGTCCCAATGGCTGCCTGATTCTCCCACTCGGAGATCT
 CAGGATAGCGTTCTCTCTCCACAATGATGGGCGGCGCCTCGTATTCGGGTTTCTCGTCGGGGTAATAAAAAATTGG
 CGTATTCGATGATGCGCTGATTGGTAGCAACCAATCTCTTGGAGCTTGTGACGGACAGCCGACAGAAATGATTC
 TTTAGTGTGCGAGCTGATGATATCAGCCCACTTATTTGAATTATCCCCGTCACATGAGCGAGCATCTGCTGGGGG
 CTCGCCGTGCTCAAGTCGGATGTCTCCATCCTTGCCGACATAATCGAAGACCTTGTGTGGAGTTCGACGAATCGA
 TTTAATATTGCCGTGAGCTCCAAAGTAGTCGAACGCGAGTTGCCGAGTTGACAGTGACAATCTGTCTGAAGCCAAG
 ATATATGTGGTAATGAATTCCTCCATCCGAATGGCGCTCTCGTCCGAGTCGACAGCATAAAGGATTTCCAAAATT
 CGCTGTGACACGTTCAAGGAAAGGAGTTCCTGCGTCGAATTCATGTTGGAGGTGTTCCGGGACTTGTGGAAACGT
 AATGATAAGATTTCTTGCTCTGACTTGAATCGTCGAGGCATTTTAAGGAGTCGTTACGTGACGTGTGCAAATA
 GAGAGAGACTTGCTAATATTATCTCTCTCTATTTGCAGTTTGCACATTGCACATGGCTGCCTATAAATACCCCT
 CATCCCCCTGCGGTGAGTC



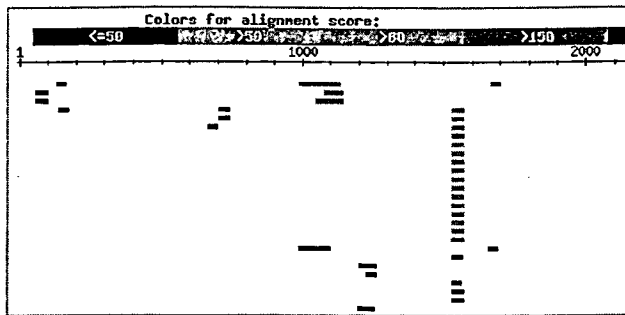
71% nucleotide similarity to *Meles meles* fecal virus

ORFs: 294aa – capsid protein 26% similarity to *Gemycircularvirus 8*

244aa – replication protein 34% similarity to *Meles meles* fecal virus

103aa – replication-associated protein 59% similarity to *Dragonfly-associated virus 1*

Figure 3(b)

Blastn - Geall:

```
>>>nrnuc:GI_365735143 Gi|365735143|gb|JN704610.1| Meles mele... 84.2 5e-12
>>>nrnuc:GI_527290455 Gi|527290455|gb|KF413620.1| Hypericum ... 60.8 6e-05
>>>nrnuc:GI_537367975 Gi|537367975|gb|KF371632.1| Gemycircul... 53.6 0.009
```

```
>>>>nrnuc:GI_365735143 Gi|365735143|gb|JN704610.1| Meles mele fecal virus isolate VS4700006,
complete genome. 0/0
Length=2198.
```

```
Score = 84.2 bits (92), Expect = 5e-12
Identities = 105/146 (71%), Gaps = 0/146 (0%)
Strand=Plus/Plus
```

```
Query 988 GGGTCCTCGTTCATGCACATGATTGAGGGTTTGCCCCAGTAGATCGTTCGCTTCTTTCGG 1047
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 1012 GGGTCCTCGTTCATAAGCCAGATAGTGGGCGCCCCACTGGATGGTTGTCTTCCCTCGG 1071

Query 1048 TACTTGTCGGTGATGACGAATTGTTTTTGTGCGCCTAGCCACCCCTTGTATGAGTGGAAG 1107
          || ||||| ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Sbjct 1072 TAYTTGTCGGTGACGGTGAATGTTTGTGTGACCCYAGCCACCCYTTGTAGGAAGGAAAG 1131

Query 1108 TATTGGAATCCTCCCTGGATGTCGTC 1133
          ||| ||| ||| ||| |||||
Sbjct 1132 AATTGGAAGTTTCCTTGCATATCGTC 1157
```

```
>>>>nrnuc:GI_527290455 Gi|527290455|gb|KF413620.1| Hypericum japonicum associated circular
DNA virus isolate VNHJ1W, complete genome. 0/0
Length=2200
```

```
Score = 60.8 bits (66), Expect = 6e-05
Identities = 56/71 (78%), Gaps = 0/71 (0%)
Strand=Plus/Plus
```

```
Query 1075 TGTGCGCCTAGCCACCCCTTGTATGAGTGGAAGTATTGGAATCCTCCCTGGATGTCGTCG 1134
          || || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 1184 TGAGCACCAGGAGCCACTCCTGTATCCGTGAAAGAATCCAAAACCACTCTGTATGTCGTCG 1243

Query 1135 AAGACGGCGTA 1145
          ||||| |||||
Sbjct 1244 AAGACGGCGTA 1254
```

```
>>>>nrnuc:GI_537367975 Gi|537367975|gb|KF371632.1| Gemycircularvirus 10 isolate P14,
complete genome. 0/0
Length=2229
```

```
Score = 53.6 bits (58), Expect = 0.009
Identities = 75/103 (72%), Gaps = 2/103 (1%)
Strand=Plus/Plus
```

```
Query 1044 TCGGTACTTGTCCGTGA-TGACGAATTGTTTTTGTGCGCCTAGCCACCCCTTGTATGAGT 1102
          || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 1280 TCTGTACTTGTCCGTAAGTAACTTGGACT-GCGCCCCCAACCACCCCTTGTATGACG 1338

Query 1103 GGAAGTATTGGAATCCTCCCTGGATGTCGTCGAAGACGGCGTA 1145
          |||| || || ||||| ||||| || || || || || || || || || ||
Sbjct 1339 GGAAATACTTGAATCCGCCCTGCATATCATCAATACTGCATA 1381
```

Query	1	MAYGFRRRFGRSRRSRRRRPVRRYSRRGRSGYGRGGYNRRYRRRRRSWISPYNK-----	55
Sbjct	1	MAY RRRF SR R+P+RR R G+ RR R+R S + NK	52
Query	56	KHDVIKGAHDSNDNSFA-----PIGA--PNTF-FGCPTYMPNAVAGDTGHN	100
Sbjct	53	RDNMLSFSTNTAEDPFSTDYTIGPAVMRRPVGADLPAEFIVVWNTGRPELS--TGFRG	110
Query	101	QKV----RTSKNVGFSGYQERVHIATAESLIWR-RMVIWSYNRLNTAGPTKTD--DQGT	153
Sbjct	111	K+ RT+ + G+ER+ + T S WR R + ++ A P ++D + +	170
Query	154	SYTTRQLTPLQLNEGVRFLFRGTQGIDYTENTLHEAPLNSDHTVAYDRTRVLQPRRPT	213
Sbjct	171	NGMVRLLRAQSQAEEINDDVFDGQRNVDWLSAIT--APINRRHVSVHYDRTRTIQ----S	224
Query	214	GDEYGHIFNMKFVNPGRG-IIYSDDEDGLQKGPLVNGWSSSLGRVSKGNMYIFDVFSTGYS	272
Sbjct	225	GNQSGVTRTYRQWHPMRKNVYADEQAG--EGMVDSGVSVTKGVGMGNYVMDIF-TKHG	281
Query	273	LSSNTPSIARFAPTGTTRYWTE	293
Sbjct	282	LNDOSTLS-YEPESTFFWHE	301

>>>>sptrembl:C7FFR0_9VIRU C7ffr0 SubName: Full=Coat protein;. 5/2013
Length=312

Score = 67.8 bits (164), Expect = 5e-09, Method: Compositional matrix adjust.
Identities = 86/319 (26%), Positives = 145/319 (45%), Gaps = 50/319 (15%)

```
Query 11  RSRRSYRRRPVRRYSRRGRSGYGRGGYN--RRRYRRR----RRSWISPYNKK--HDVIKG 62
          RS R  RRP++  R G S  R Y+ +R YR++  +RS ++ ++K + ++
Sbjct 7   RSSRFRARRPIKTSRRGGVSKPRRRSYSAKKRTYRKKASMSKRSILNTTSRKKRNGMLTF 66

Query 63  AHDSNDN-SFAPIGAP---NTFFG----FCPTYMPNAVAGDTGHNPKVRTSKNVGFSGY 114
          ++ +N  S AP AP  N  G  +CPT M N VA  N  T+  G
Sbjct 67  SNTTNSGASTAPAAAPVYVNAVLGGTFIWCPTAM-NLVAQSLISN-MSSHTATTCYMRGL 124

Query 115 QERVHIATAESLIW-RRMVIWSYNRL-----ENTAGPTKTDDQGTSY--TTRQLTPLQLN 166
          E + I T+ + W R V +++ L  T PT +D  Y ++ + L N
Sbjct 125 SEHIRIQTS SGIPWFHRRVCFTWKG LGPFNV MATESPT-SDFPWL PYQDSSNGVQRLFFN 183

Query 167 EGV-----RAFLFRGTQGIDYTENTLHEAPLNSDHFTVAYDRTRVLQPRRPTGDE 216
          E V  R LF+G + +D+T+ + APL++ TV +D+T +Q  +G+
Sbjct 184 ENVNMPQTINRQRNILFKGAENV DWTDP II--APLDTGRITVKFDKTWTMQ----SGNT 237

Query 217 YGHIFNMKEWNP GGR-IIYSDD EDG-LQKGPLVNGWSSLGRVSKGNMYIFDV FSTGYSLS 274
          G +  K W+P  ++Y DDE G ++ G  + +S+ +  G+ Y+ D+  G  +
Sbjct 238 NGIVRERKLWHPMNHNLVYDDDEYGDIESG---SWFSTDSKAGMGDYVMDIIQGGQGGT 294

Query 275 SNTPSIARFAPTGTTRYWTE 293
          ++  + R  T YW E
Sbjct 295 AS--DLLRMTANSTMYWHE 311
```

ProtSweep
A Protein Identification Tool

Results for /home/vir088/rindneu/gemi/HCB19.212.294.pep

Number of residues: 294

Molecular weight: 33.86 kD

Charge: 25

Isoelectric point: 11.38

[more...](#)

Prediction of protein localization:

nucl 30 %

Identified Protein & Transcript:

Protein:

id: [sptrembl:T1YTBI_9VIRU](#)

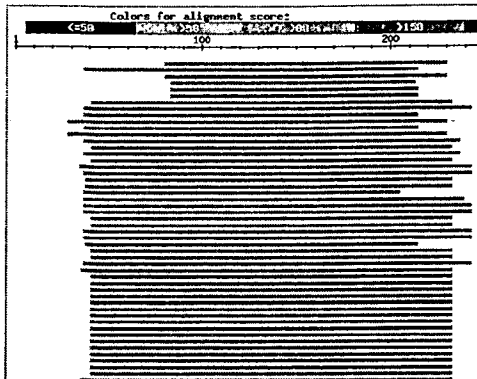
valuation: putative

scientific organism: Gemycircularvirus 8

abstract: T1ytb1 SubName: Full=Capsid protein;. 2/2014Length=302

isoform:

ORF 244aa:



```
>>>sptrembl:G9ITS8_9VIRU G9its8 SubName: Full=Replication-as... 90.5 6e-16
>>>sptrembl:TLYRX6_9VIRU Tlyrx6 SubName: Full=Replication as... 89.4 1e-15
>>>sptrembl:TLYQS5_9VIRU Tlyqs5 SubName: Full=Replication as... 85.5 2e-14
```

```
>>>>sptrembl:G9ITS8_9VIRU G9its8 SubName: Full=Replication-associated protein (Meles meles
fecal virus);. 4/2014
```

Length=237

Score = 90.5 bits (223), Expect = 6e-16, Method: Compositional matrix adjust.
Identities = 52/152 (34%), Positives = 76/152 (50%), Gaps = 3/152 (1%)

```
Query 81 PLCCRLGRERHSDGGIHYHYLGFDDKIVTVNSATAFDYFGAHGNIKSIRRTPHKVFDDYVG 140
P C +GRE H+DGGIH H ++ F K V + FD G H NI RTP K+ DY
Sbjct 36 PAECIIGRENHADGGIHLHAFVDFGKRVDIRDNRFDVEGFHPNIAPCGRTPQKMLDYAI 95

Query 141 KGDGDIRLEHGEPPADARSCDGDNSNKWADIISCDTKESFLSAVRQQAPRDWLLSNQRIIE 200
KGDG+ P D D + + W I + + + F + R+ APR L ++Q +
Sbjct 96 KGDGVVAGGLNPILDLP--DAETDSIWHRIANAPSVDEFWAITRELAPRALLCNHQLSLRA 153

Query 201 YANFYYPDEKPEYEAP-PIIVERERYPEISEW 231
YA ++Y YE P + + + P++ EW
Sbjct 154 YAEWHYRPVATRYEHPHGLSLRTDHPVQLDEW 185
```

```
>>>>sptrembl:TLYRX6_9VIRU Tlyrx6 SubName: Full=Replication associated protein
(Gemycircularvirus 7);. 4/2014
```

Length=332

Score = 89.4 bits (220), Expect = 1e-15, Method: Compositional matrix adjust.
Identities = 59/190 (31%), Positives = 91/190 (47%), Gaps = 17/190 (8%)

```
Query 38 MPRRFQVR-ARNLIITFPQVPEHLQHEFDAGTFFL---ERV TANFGNPLCCRLGRERHSD 93
M R+F ++ A+ ++T+PQ+PE +EF L ERV AN+ +GRE H+D
Sbjct 1 MSRQFCLKDARYCLLTYPQIPETEAYEFPELASELFSSSERVDANY-----VIGRELHAD 54

Query 94 GGIHYHYIYLGFDKIVTVNSATAFDYFGAHGNIKSIRRTPHKVFDDYVGKGDGDIRLEHGE-- 151
GG H +L F + + FD G H NI+ + RTP ++Y KD D+ GE
Sbjct 55 GGYTSHCFLDGFRKFSSRDTRIFDIQGHHPNIERVGRTPRTAYNYTIKDNDVVARSGEYS 114

Query 152 -PPAD----ARSCDGDNSNKWADIISCDTKESFLSAVRQQAPRDWLLSNQRIIEYANFY 206
PP + R D S+ WA I+ ++++ F + PR S + YA++ Y
Sbjct 115 GPPENNNRERQSGNDTSSDWATILCAESRDEFFDLCKSLQPRSLACSLSLTRYADWRY 174

Query 207 PDEKPEYEAP 216
Y+ P
Sbjct 175 RPVPTYQHP 184
```

```
>>>>sptrembl:TLYQS5_9VIRU Tlyqs5 SubName: Full=Replication associated protein
(Gemycircularvirus 10);. 4/2014
```

Length=324

Score = 85.5 bits (210), Expect = 2e-14, Method: Compositional matrix adjust.
Identities = 50/153 (32%), Positives = 75/153 (49%), Gaps = 2/153 (1%)

```
Query 81 PLCCRLGRERHSDGGIHYHYIYLGFDKIVTVNSATAFDYFGAHGNIKSIRRTPHKVFDDYVG 140
P C +GRE H+DGG H H + F + FD G H N++ RTP+K +DY
Sbjct 35 PAECIIGREAHADGGTHLHAFCDGGERFYTADPRRFDVDGYHPNVQPFGRTPYKGDYAI 94
```

```
Query 141 KDGDIRLEHGEPPADARSCDGDNS-NKWADIISCDTKESFLSAVRQQAPRDWLLSNQRII 199
          KDGD+   E P       D   + ++W I+ +T + F   +R+ APR L +   +
Sbjct 95  KDGDVVCGLERPGTGP GTDVS GAGSQWDRIVGAETPDEFWRLIRELAPRTLLTNFNSVR 154

Query 200 EYANFYYPDEKPEYEAPP-IIIVERERYPEISEW 231
          YA+++Y   YE+PP I +       E+ EW
Sbjct 155 AYADWHYRPVVAPYESPPEIEFDLSGVEELREW 187
```

ProtSweep
A Protein Identification Tool

Results for /home/vir088/rindneu/gemi/HCB19.212.244r.pep

Number of residues: 244	Molecular weight: 28.34 kD
Charge: -2	Isoelectric point: 6.97

[more...](#)

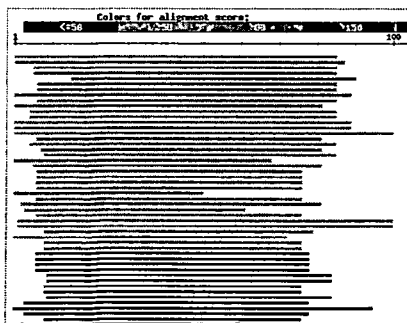
Prediction of protein localization:

cyto_nucl	12.5 %
mito	10 %
cyto	9 %

Identified Protein & Transcript:

Protein:
id: [sptrembl:O8OPT6_9GEMI](#)
valuation: putative
scientific organism: Tomato mottle
abstract: Q8qpt6 SubName: Full=Replication-associated protein; Flags: Fragment; 4/2014Length=232
isoform:

ORF 103aa:



```
>>>sptrembl:K0A1L7_9VIRU K0a117 SubName: Full=Spliced replic... 115 7e-24
>>>sptrembl:Tlyrw0_9VIRU Tlyrw0 SubName: Full=Replication as... 106 4e-21
>>>sptrembl:J0YHY6_9PSED J0yhy6 SubName: Full=Uncharacterize... 106 5e-21
```

```
>>>>sptrembl:K0A1L7_9VIRU K0a117 SubName: Full=Spliced replication-associated protein
(Dragonfly-associated virus 1);.
4/2014
Length=318
Score = 115 bits (288), Expect = 7e-24, Method: Composition-based stats.
Identities = 51/86 (59%), Positives = 62/86 (72%), Gaps = 0/86 (0%)
```

```
Query 1 MDEFSDDCQYAVFDDIQGGFEYFHSYKWLGAQKQFVITDKYRKKRTIYWGKPSIMCMNE 60
      +DE +D YAVFDD+QGG E+FH+YK WLGAQKQF TDYR K+ + WG+PSI N
Sbjct 227 LDESLEDVDYAVFDDMQGGLEFFHAYKFWLGAQKQFYATDKYRGKQLVNWGRPSIYISNT 286

Query 61 DPHFAKGVDDYDWMNNCHIINVVDPI 86
      +P KG D DWLM NC I++V PI
Sbjct 287 NPLADKGAADV DWMGNCVIVHVD RPI 312
```

>>>>sptrembl:T1YRW0_9VIRU T1yrw0 SubName: Full=Replication associated protein
(*Gemycircularvirus 8*);. 2/2014
Length=320

Score = 106 bits (264), Expect = 4e-21, Method: Composition-based stats.
Identities = 43/88 (48%), Positives = 63/88 (71%), Gaps = 0/88 (0%)

```
Query 1  MDEFSDDCQYAVFDDIQGGFEYFHSYKWLGAQKQFVITDKYRKKRTIYWGKPSIMCMNE 60
      ++E  +   YA+FDD+QGG ++FHSYK WLGAQ QF  TDKYR K+ I+WG+PSI  N+
Sbjct 229 LEEELEGVGYAIFDDMQGGLKFFHSYKFWLGAQSQFWATDKYRGKKLIHWGRPSIYICNQ 288

Query 61  DPHFAKGVVDYDWMNNCHIINVVDPICI 88
      +P  +GVD+DWL+ NC  + + + + +
Sbjct 289 NPLCDEGVDDHDLIGNCDFVEITESLLV 316
```

>>>>sptrembl:J0YHY6_9PSED J0yhy6 SubName: Full=Uncharacterized protein;. 4/2013
Length=103

Score = 106 bits (264), Expect = 5e-21, Method: Compositional matrix adjust.
Identities = 46/81 (56%), Positives = 58/81 (71%), Gaps = 0/81 (0%)

```
Query 6  DDCQYAVFDDIQGGFEYFHSYKWLGAQKQFVITDKYRKKRTIYWGKPSIMCMNEDPHFA 65
      DD QYAV DDIQG F++F +YKWLGAQ+ F +TDKYR K TI WG+P+I  MNEDP
Sbjct 21  DDAQYAVLDDIQGNFQFFPAYKWLGAQESFTVTDKYRGKTTITWGRPTIWLNMNEDPEEV 80

Query 66  KGVDDYDWMNNCHIINVVDPI 86
      VD +WL  NC I+++  +
Sbjct 81  GHVDLNLWLRGNCITIVHLTQSL 101
```

ProtSweep
A Protein Identification Tool

Results for /home/vir088/rindneu/gemi/HCB19.212.103r.pep

Number of residues: 103

Charge: -6

Molecular weight: 12.03 kD

Isoelectric point: 4.65

[more...](#)

Prediction of protein localization:

cyto	18.5%
cyto_nucl	11%
extr	10%

Identified Protein & Transcript:

Protein:

id: [sptrembl:H6WAT6_9GEMI](#)

valuation: putative

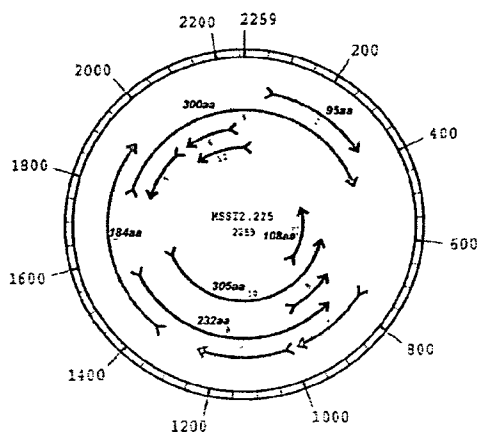
scientific organism: Papaya leaf

abstract: H6wat6 SubName: Full=Replication-associated protein; Flags: Fragment;. 6/2012Length=123

isoform:

MSSI2.225 (2259bp)**(MSSI = Multiple sclerosis serum isolate)****(ms24.RCA.1700.3)**

TTCCTAATCAGCTGTTTCGATCAGACCACCGAAGGTGGATGTCGACGCGTCATCGGCCACTTCTCGGGGTCACTA
 ATGCCCGAGATGAGCTGCAATCTTACGTCTTCCGGGGTCAGGAAGATTCTGATTGGGCAGATCAGTTTACTGCCC
 CCCTTGATACCTCGTCGCATTACGGTTAAGTCCGACAAACTCCGCGTCATCCGCCCGGTAATGAAACAGGTGCTT
 CGCGCTTGTTAAGTTTTGGTACCCCATTCGTCTGATACCATCTCTTACGAAGATGACCTTGAGTCCGACGTTGTCG
 GTGATCGGCCCTTTTCTACCGCAGGTTTACGAGGTGTTGGGGATATGTACGTCATGGATATAATGGGTATTACGA
 ATTTGGTGAATGATGCACCTCCAACGTCTACAGGTTTCGGTCTGAGGGCAGCTTTTATTGGCATGAGCGGTAAA
 TTAGGAGACTATGGGGCTATCTAGGTACACGAAATCGCAGTTGGCATTAAAGCCACTCGGTGTCAACACCAAGCTC
 GTCAGTGGGTCCGAGTTAGATAGCCAAATTGAGGGCCGAGCCAGTGTACCAACTTTTTCCCTTGTTACTTATC
 CGTGACGTAGAACTGCTTTTGATGACCTAACCAGAATTTGTAGCTGGGAAGGAATTTAATTCCTCCGAAGTCGTC
 AAAGATGGCGTACTCAACCCCATCAAGGTTTTTCGTCCAAGCTAAAGAGGCCTCCGAAGTAAGCATGGTTTCCAG
 GCTTCTAGCCCAAATGGTTTTTCCCATTCTGGAAGGCCGCTATACCACAAGTGATTTTCTTCGTTCTGTATATCG
 AAGTCAGCATCCAGATTGTTGATGCGCACTCGAGCGAGGCGCTGCGCGAGGCCCGCCGAGCATGGAGGCGCG
 CGTAGCCGCCGACTGCAGCCGCTGCGACGTGCATGATAACTTACCTCCAGTTTCACCTCGTCCAAGATTTTGTTG
 TACCCATTGATTGAGTTTCAGCCACCCAAGACGTATCGATGTTAACTCCCTCCGGTGTCTCATACGGGACTCTGAT
 TGGTGGGAATTTCCAGGCTGCGTATGCCCTGAGCTGGGTGAAGGAGGTGACCAATGAGCGTGGAGCCAGGCGTTT
 GCATAGGTCCCAAACTCTTGCTCACTCTCAGCATTGATGATTTACGCCACACCCCACTAGTTGCATCCACTCT
 GCCTGCAACAGGTCGCTCGAGTCCCCCAGCAACAACATCTCCATCTTTGATTGCATAATCAAACCCCTCTTCTGG
 TGTTCCGCGAGAAGGCGATACATTTGGGTGGCATCCTTCAACATCGAATGCACGCGCGTTCCGGGTCTTATATTT
 GACTCCGAAGTCGACAAAAGCGTGCAGATGAAGACCCCATCTGCGTGATTCTCTCGGCCGATGATGCATTACAGC
 TCCAAGTCCCGCAAGATGATTGACCACTGCAAAAGGTCGAGGTCTCCACATTGTGGATAGGTGAGCAAGGCGTA
 CCGTGCTTGAAATCGAAAAGAGGACATAGTTTCGTTGTTGCACATTTTCTTGGGTCCAAGTCTGGGCAGTTAATG
 TTATATGCCCAGACGGACCCAGACCCGGGTCTTTCGAATAATAATACCCCTCGCTCGTCCCAACCTCCTCGTT
 TTTTGTATGCGCGATTATCATGTCTGCCTTGATCCACCCCGTTGCCTGCAATCGGACTTGCCAAGACTACGTC
 AACCACGCTAGACACGAATGTTGTTTAGTGTGTAACCCCTCCGCCACCCCTTCATCTCTACTGTGCGACATTGT
 TCAGAATGTCGGTCCGCCGTCGATCTGGTTCGAAGAACTCGCAAAGCTTTCGCCGCGGTGGTAGAGCTACAAGAT
 CTCGTGGCCGAAGTACACGCCGCTCTTCTAGGCGAAGAGTCAGACGGCGAACTATGACTACCCGACGTGTCTCA
 ACATAACCACTCGTAAGAAGGTGGATAACATGTTGCCTGTTGTTGTTGCCGAGGATAACACGACTACTGTGGGGC
 CTTACACGTCCGTGTACCCCTCCTTTCCCTTTTCGTTCCCAATGCCCGCACGACCCGAACCCCTGTTACCAATC
 CTGCCGTCCGAAATTCCTCGGATATTTTCGCGGTCCGATATAGGGAGCGCGTCCGCATTGATCTTATGGGTGGTG
 GAACCTTTATGTGGCGCCGCATCGTCTTCATGTTGAAGGGATCCGATTTGAGAGTTGCTATGAATCCAGTGATT
 CAGGCAACA



75% nucleotide similarity to *Gemycircularvirus 9* isolate P22.

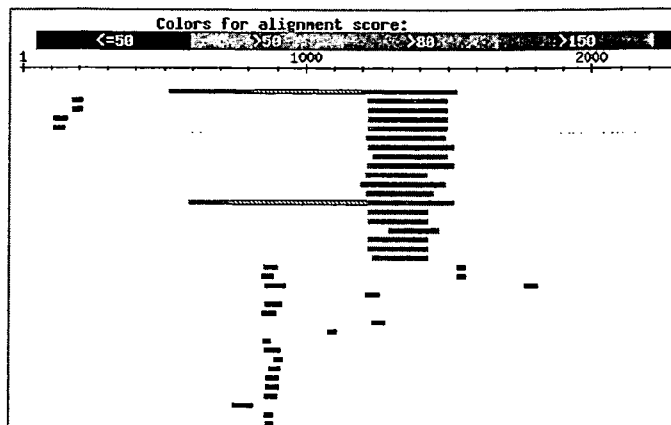
ORFs:

300aa – coat protein 33-36% similarity to *Gemycircularvirus 8* and *Meles meles* fecal virus

232aa – replication-associated protein 61% similarity to *Gemycircularvirus*

108aa – replication-associated protein 71% similarity to *Gemycircularvirus*

Figure 3(c)

Blastn - Geall:

```
>>>nrnuc:GI_537367978 Gi|537367978|gb|KF371633.1| Gemycircul... 230 6e-56
>>>nrnuc:GI_537367981 Gi|537367981|gb|KF371634.1| Gemycircul... 141 2e-29
>>>nrnuc:GI_520689312 Gi|520689312|gb|KF268026.1| Sclerotini... 123 6e-24
```

```
>>>nrnuc:GI_537367978 Gi|537367978|gb|KF371633.1| Gemycircularvirus 9 isolate P22,
complete genome. 0/0
Length=2130
Score = 230 bits (254), Expect = 6e-56
Identities = 254/338 (75%), Gaps = 3/338 (0%)
Strand=Plus/Plus
```

```
Query 1194 CCACCTCTGCCTGCAACAG---GTCGCTCGAGTCCCCCAGCAACAACATCTCCATCTTTGA 1250
          ||||| ||| || ||| ||| || ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 1768 CCACCTCCGCCATCATCAGCGGGTCTGTCCAGTCTCCAGCAACAACATCCCCATCTTTGA 1827

Query 1251 TTGCATAATCAAACCCCTCTTCTGGTGTTCGCGAGAAGGCGATACATTGGGTGGCATC 1310
          ||||| || || || || || || || || || || || || || || || || || || ||
Sbjct 1828 TTGCATAGTCGTATCCGCGAGCTGGGCGTCCCGAGATGGTTCAACATTCGGGTGGTGTC 1887

Query 1311 CTTCAACATCGAATGCACGCGCTTCCGGGTCTATATTTGACTCCGAAGTCGACAAAAG 1370
          ||| ||||| || || ||||| || || ||||| ||||| ||||| ||||| |||||
Sbjct 1888 CTTGCACATCGAAGACGCGTACGTTCCGCGTACTAAATTTGACTCCGAAGTCGACGAAAAG 1947

Query 1371 CGTGCAGATGAAGACCCCATCTGCGTGATTCTCTCGGCCGATGATGCATTAGCTCCAA 1430
          | || ||||| || || || || || || || || || || || || || || || || ||
Sbjct 1948 CATGGAGATGAATTCTTCATCAGCGTGAGTCTCTCGTCCAATGATGCACTCTGCTCTAA 2007

Query 1431 GTCCCGCAAGATGATTGACCACTGCAAAAGGGTCGAGGTCTCCACATTGTGGATAGGTGA 1490
          | || || || || || || || || || || || || || || || || || || || ||
Sbjct 2008 GGCCTGATAGGTGGTCAGACACAGCAAAACGGGTCGAGGTCTCCACACTGGGCGTAAGTGA 2067

Query 1491 GCAAGGCGTACCGTGCTTGAAATCGAAAAGAGGACATA 1528
          | ||||| || || || || || || || || || || || || || || || || || ||
Sbjct 2068 GTAAGGCATATCTGGCCTGGAATCTGAATGATGACATA 2105
```

```
Score = 138 bits (152), Expect = 3e-28
Identities = 219/309 (70%), Gaps = 4/309 (1%)
Strand=Plus/Plus
```

```
Query 515 ACACCAAGCTCGTCACGTGGGTGCGAGTTAGATAGCCAAATTGAGGGCCGAGCCCAGTGT 574
          || ||| |||| || ||||| || || || || || || || || || || || || ||
Sbjct 1092 ACCCCATGCTCCTCGCGTGGGTCCCTCATTGTTGATCCATATTGATGGTCTGCCCATTTCA 1151

Query 575 ACCAACTTTTCCCCTTGTAATCTATCCGTGACGTAGAACTGCTTTTGATGACCTAACCAG 634
          ||||| ||||| ||||| || ||||| ||||| || ||||| ||||| || || || || ||
Sbjct 1152 ACCAACTTTTCCCCTTGTAATCTGCTGTGACGTAAACTGCTTCTGGTGTCCCATCCAA 1211

Query 635 AATTTGT-AGCTGGGAAGGAATTTAATTCCTCCGAAGTCGTCAAAGATGGCGTACTCAAC 693
          ||||| || ||||| ||||| || ||||| || ||||| || ||||| || || || || ||
Sbjct 1212 AATTTGTAAGATGGG-AGGAATTTTAAACCTCCAAATCATCAACACCCGATA-TGTGG 1269

Query 694 CCCATC-AAGGTTTTCGTCCAAGCTAAAGAGGCCTCCGAAGTAAGCATGGTTTCCAGGC 752
          || || || || || ||||| || || ||||| || ||||| || ||||| || || ||
Sbjct 1270 CCGAGCTTATGTCCTCGTCCATTGAGAAAAGCCCTCCGAAATAGGCATGGTCTCCTAGTG 1329
```

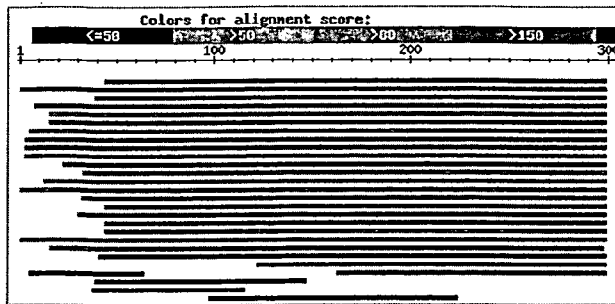
Score = 141 bits (156), Expect = 2e-29
Identities = 200/279 (71%), Gaps = 4/279 (1%)
Strand=Plus/Plus

Query	1217	TCGAGTCCCCAGCAACAACATCTCCATCTTTGATTGCATAATCAAAACCCCTCTTCTGGT	1276
Sbjct	1828	TCGAGTCTCTCCAGCAACAACATCTCCGTCCTTGATCGCATAAGTTGCACCCTTTTCCGGT	1887
Query	1277	GTTCCGCGAGAAGCGCAT--ACATTTGGGTGGCATCCTTCAACATCGAATGCACGCGCGT	1334
Sbjct	1888	GTT--GAGTAACCGGTACGACATTTGGGTGGCGTCCGTCCACATCGAATACGCGGACAT	1945
Query	1335	TCCGGGTCTTATATTGACTCCGAAGTCGACAAAAGCGTGCAGATGAAGACCCCATCTG	1394
Sbjct	1946	TTCTTGACTCAAAC TTCCGTTCGAACATGAAGAAAGCGTGCAGGTGAAGTCTCTCCATCAG	2005
Query	1395	CGTGATTCTCTCGGCCGATGATGCATTCAGCTCCAAGTCCCGCAAGATGATTGACCACTG	1454
Sbjct	2006	CGTGCTCTTCTCGGCCAATGATGCATCAGCTCCCAGGCGTCCAAGATGGTCGTTAACTT	2065
Query	1455	CAAAAGGGTCGAGGTCTCCACATTGTGGATAGGTGAGCA	1493
Sbjct	2066	CCCAGGGGTCCAGGTCTCCGACTGGGGGAACGTGAGCA	2104

Score = 123 bits (136), Expect = 6e-24
Identities = 193/276 (69%), Gaps = 0/276 (0%)
Strand=Plus/Plus

Query	1220	AGTCCCCCAGCAACAACATCTCCATCTTTGATTGCATAATCAAACCCCTCTTCTGGTGTT	1279
Sbjct	1824	AGGCCCCCACAATGACATCTCCATCTTTGATCGCGTAGTCATACCCCTTCTCCGGAGTC	1883
Query	1280	CCGCGAGAAGGCGATACATTGGGTGGCATCCTTCAACATCGAATGCACGCGCGTTCCGG	1339
Sbjct	1884	CCTCGAGACTTTGTAATGTTTGGGTGGTGACCATCGACATCGAATACATCAGCCTTTCTA	1943
Query	1340	GTCTTATATTTGACTCCGAAGTCGACAAAAGCGTGCAGATGAAGACCCCATCTGCGTGA	1399
Sbjct	1944	CTGCGAAACTTCCGTCCGAACCTCTGCAAAGCAGTGTAGGTGAATGCCTCCATCCTCGTGA	2003
Query	1400	TTCTCTCGGCCGATGATGCATTACAGTCCAAGTCCCGCAAGATGATTGACCACTGCAAAA	1459
Sbjct	2004	TGCTCTCTTCCAATGATGCACTCAGTCCCAGAAGTGAAAGTTTATCCATAACTCGGAAA	2063
Query	1460	GGGTCGAGGTCTCCACATTGTGGATAGGTGAGCAAG	1495
Sbjct	2064	GCATCGAGCTCTCCACATTGTGCGTATGTGAGCAAG	2099

ORF300aa:



>>>sptrembl:G9ITS7_9VIRU G9its7 SubName: Full=Coat protein;.... 159 2e-36
 >>>sptrembl:TlytB1_9VIRU TlytB1 SubName: Full=Capsid protein... 151 4e-34
 >>>sptrembl:IOAY94_9VIRU IOay94 SubName: Full=Capsid protein... 137 8e-30

>>>sptrembl:G9ITS7_9VIRU G9its7 SubName: Full=Coat protein (*Meles meles*);. 5/2013
 Length=310

Score = 159 bits (401), Expect = 2e-36, Method: Compositional matrix adjust.
 Identities = 100/271 (36%), Positives = 141/271 (52%), Gaps = 21/271 (7%)

```

Query  45  RRVLNITTRKKVDNMLPVVVAEDNTT-TVGPYTSVSPLLSLFVPNARTTRTPVTN--PAV 101
          RR+LN+ + KK DNML +V D T+GP T + SL++P+AR P + +
Sbjct  46  RRILNVASIKKRDNMLSMVQTSQAPFTIGPITVGTGFASLYMPSARPL-GPANDVGEST 104

Query  102  RNSSDIFAVGYRERVRIDLMMGGGTFMWRRI VFMLKGS DLRVAMNSSDSGNIPNQLFDQTT 161
          R D F+VGY+ERV++D+ GGG + WRR+VF KG L N + Q F+
Sbjct  105  RQRRDTFSVGYKERVQVDIHGGGVWKWRRVVFVTKXGDQL---YNKGEFIGDWTQPFESKA 161

Query  162  EGGCRRVIGPLLGV-----TNARDELQSYVFRGQEDSDWADQFTAPLDTRRITVK 211
          PL GV T + ++ ++ G E DW +FTA + T RITV
Sbjct  162  VKPSPXDPNPL-GVDMVRLIGQPTTEQHNAIRDLDWDGHEGIDWTSEFTAKVSTTRITVL 220

Query  212  SDKLRVIRPGNETGASRLFKFWYPIRRTISYEDDLES DVVGDR--PFSTAGLRGVGDMYV 269
          SDK PGNE+G SR F+ WYP R+ + Y+DD V R P S +G+GD+ +
Sbjct  221  SDKTYHFNPNGNESGMSRTFRLWYPTKRRLIYDDDERGGA VF SRGSPVSVTSRQGMGLXI 280

Query  270  MDIMGITNLVNDAPPTS YRFGPEG SFYWHER 300
          DI + + P S + PEG++YWHER
Sbjct  281  YDI-AYAVVPSAGNPGSMTWAPEGTYWHER 310
  
```

>>>sptrembl:TlytB1_9VIRU TlytB1 SubName: Full=Capsid protein (*Gemycircularvirus 8*);. 2/2014
 Length=302

Score = 151 bits (381), Expect = 4e-34, Method: Compositional matrix adjust.
 Identities = 107/321 (33%), Positives = 161/321 (50%), Gaps = 40/321 (12%)

```

Query  1  MSVARR--SGRRTKAFRRGGRATRSRGRSTRSSRRRVRRRTMTTRVLNITTRKKVDN 58
          M+ ARR + R RK RR R RRS RR R+R M+TR +LN T++KK DN
Sbjct  1  MAYARRRFTSSARKPIRRSRRPA-----GKRRSYRRSTRKRGMSRTRLLNKTSQKKRDN 55

Query  59  MLPV--VVAED---NTTIVGPYTSVSP----LLSLFV-----PNARTT--RTPVTNP 99
          ML AED T+GP P L + F+ P+ +T R +
Sbjct  56  MLSFSNTTAEDPFSTDYITIGPAVMRRPVGADLPAEFIVVW NATGRPHELSTGFRGSKIDT 115

Query  100  AVRNSSDIFAVGYRERVRIDLMMGGGTFMWRRI VFMLKGS DLRVAMNSSDSGNIPNQLFDQ 159
          ++R ++ IFAVG +ER++++ + WRR I F K + SD F +
Sbjct  116  SLRTNTTIFAVGLKERIKLETNNSAPWRWRRICFTSKNDYGEADPDESD-----YFRR 168

Query  160  TTEGGCRRVIGPLLGVTNARDELQSYVFRGQEDSDWADQFTAPLDTRRITVKS DKL RVIR 219
          T+ G R LL + +E+ VF GQ + DW TAP++ R ++V D+ R I+
Sbjct  169  TSNMVR-----LLRAQSQAEEINDVFDGQRNVDWLSAITAPINRRHVS VHYDRTRTIQ 223

Query  220  PGNETGASRLFKFWYPIRRTISYEDDLES DVVGDRPFSTAGLRGVGDMYVMDIMGITNLV 279
          GN++G +R ++ W+P+R+ + Y D+ + + D S G G+G+ YVMDI L
Sbjct  224  SGNQSGVTRTYRQWHPMRKNVVYADEQAGEGMVD SGVSVTGKVGMGNYVYVMDIFTKHGLN 283

Query  280  NDAPPTS YRFGPEG SFYWHER 300
          +D SY PE +F+WHE+
Sbjct  284  DDQSTLSYE--PESTFFWHEK 302
  
```

>>>>sptrembl:I0AY94_9VIRU I0ay94 SubName: Full=Capsid protein (*Cassava associated ssDNA virus*);. 7/2012
Length=315
Score = 137 bits (344), Expect = 8e-30, Method: Compositional matrix adjust.
Identities = 94/288 (32%), Positives = 133/288 (46%), Gaps = 37/288 (12%)

Query 40 RTMTTRRVLNITTRKKVDNMLPVVVAEDNTTTPVGPYTSVSPLLS-----LFVP 87
R M+ RR+LNIT+ KK D M + T G S P + L+
Sbjct 38 RKMSKRRLNITSTKKKDTMTGYTNSTPANQTGGTVYSTDP AIVTGGLNANQCNAFLWCA 97

Query 88 NARTTRTPVTNP-----AVRNSSDIFAVGYRERVRIIDLMGGGTFMWRRI VFM 134
ART + V NP A R S+ + VG E++ I G + W RRI +
Sbjct 98 TARTNQ--VINPTQNTLVPGNKFYTACRTDSNPYVGLSEKIEIQNSGMPWQWRRICWT 155

Query 135 LKGSDLRVAMNSSDSGNIPNQLFDQTTEGGCRRVIGPLLGV TNARDE--LQSYVFRGQED 192
KG A ++S + L Q G RV+ + G D+ L ++FRGQ
Sbjct 156 QKGGQGLPANSTSFN-----LSQQNAITGWNRMVQVAGNPGTGDQYNLFHHIFRGQVG 209

Query 193 SDWADQFTAPLDTRRITVKSDKLRVIRPGNETGASRLFKFWYPIRRTISYEDDLES DVVG 252
SDW D TAPLD + ITVK DK+ + GNE G R ++ W+P+++T+SY D+ +V
Sbjct 210 SDWVDVMTAPLDPQHITVKYDKVVTLASGNEEGFIRAYRRWHPMKKTL SYADEEYGNVEN 269

Query 253 DRPFSTAGLRGVGDMYVMDIMGITNLVNDAPPTS YRFGPEGSFYWHER 300
FS G +G+GD YV+D+ S R P + YWHE+
Sbjct 270 ADAFSQVGKKGMDYYVLDLFRARQGAQTNDISISIR--PCATLYWHEK 315

ProtSweep
A Protein Identification Tool

Results for /home/vir088/rindneu/gemi/MSSI2.225.300.pep

Number of residues: 300	Molecular weight: 34.10 kD
Charge: 23	Isoelectric point: 11.80

[more...](#)

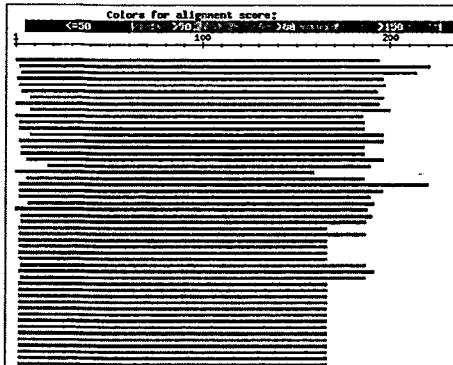
Prediction of protein localization:

nucl	24 %
mito	7 %

Identified Protein & Transcript:

Protein:
id: sptrembl:T1YTB1_9VIRU
valuation: putative
scientific organism: Geinycircularvirus 8
abstract: T1ytl SubName: Full=Capsid protein;. 2/2014Length=302
isoform:

ORF 232aa:



```
>>>sptrembl:Tlysfl_9VIRU Tlysfl SubName: Full=Replication as... 240 3e-61
>>>sptrembl:G9its8_9VIRU G9its8 SubName: Full=Replication-as... 225 9e-57
>>>sptrembl:K0a109_9VIRU K0a109 SubName: Full=Replication-as... 218 2e-54
```

```
>>>>sptrembl:Tlysfl_9VIRU Tlysfl SubName: Full=Replication associated protein;. 4/2014
Length=325
```

Score = 240 bits (613), Expect = 3e-61, Method: Compositional matrix adjust.
Identities = 122/198 (61%), Positives = 139/198 (70%), Gaps = 3/198 (1%)

```
Query 1 MSSFRFQARYALLTYPQCGDLDPFAVNVNHLAAGLGAECIIGRENHADGGLHLHAFVDFGVK 60
MSSFRFQARYALLTY QCGDLDPFAV +HL+GL AECIIGRE HAD G+HLHAFVDFGVK
Sbjct 1 MSSFRFQARYALLTYAQCGLDLPFAVSDHLSGLRAECIIGRETHADEGIHLHAFVDFGVK 60

Query 61 YRTRNARAFDVEGCHPNVSPSRGTPEEGFDYA IKDGDVAVAGGLERPVG---RVDATSGV 117
+ TRN R FDV+G HPNV PSRG P G+DYAIKGDVAVAGGL+RP V T+
Sbjct 61 FSTRNVRVFDVQGHHPNVESRGRPAGGYDYAIKGDVAVAGGLDRPADGGVGVSKTTQK 120

Query 118 WAEIINAESQEFWDLCKRLAPRSLVTSFTQLRAYAAWKFPPIRVPIYETPEGVNIDTSWV 177
WA I+ A++ EFW L L P ++V SFTQ RAYA +F Y TPE + IDT V
Sbjct 121 WATIVGAQTRDEFWALLAELDFGAMVRSFTQCRAYADHRFRESPEYVTPESIIIDTGHV 180

Query 178 AELNQWVQONLGRGETGG 195
LN+W N+ GG
Sbjct 181 DGLNEWALSNIWGHRRGG 198
```

```
>>>>sptrembl:G9its8_9VIRU G9its8 SubName: Full=Replication-associated protein;. 4/2014
Length=237
```

Score = 225 bits (574), Expect = 9e-57, Method: Compositional matrix adjust.
Identities = 118/223 (52%), Positives = 148/223 (66%), Gaps = 10/223 (4%)

```
Query 3 SFRFQARYALLTYPQCGDLDPFAVNVNHLAAGLGAECIIGRENHADGGLHLHAFVDFGVK 62
+FR Q+R LLTY QCGDLDP+ V NH+ AECIIGRENHADGG+HLHAFVDFG +
Sbjct 5 TFRLQSRVLLTYAQCGLDLPFWLVNHNHVAFFAECIIGRENHADGGIHLHAFVDFGKRV 64

Query 63 TRNARAFDVEGCHPNVSPSRGTPEEGFDYA IKDGDVAVAGGLERPVGVRDA-TSGVWAEI 121
R+ R FDVEG HPN++P TP++ DYAIKGDVAVAGGL P+ DA T +W I
Sbjct 65 IRDNRRFDVEGFHPNIAPCGRTPQKMLDYAIKGDVAVAGGLN-PILDLPDAETDSIWHRI 123

Query 122 INAESQEFWDLCKRLAPRSLVTSFTQLRAYAAWKFPPIRVPIYETPEGVNIDTSWVAELN 181
NA S EFW + + LAPR+L+ + LRAYA W + P+ YE P G+++ T V +L+
Sbjct 124 ANAPSDEFWAIITRELAPRALLCNHQLRAYAEWHYRPVATRYEHPHGLSLRTDHPVQLD 183

Query 182 QWVQONLGRGETGGKLSCTSQRLQSAATRAMLGGG--LAQRL 222
+WV+ NL +G S T R SA+ RAS+ GG L Q L
Sbjct 184 EWVRDNL----SGSSKSYTYTR--SASGRASLPSSGGPSLVQSL 220
```

```
>>>>sptrembl:K0a109_9VIRU K0a109 SubName: Full=Replication-associated protein;. 4/2014
Length=249
```

Score = 218 bits (554), Expect = 2e-54, Method: Compositional matrix adjust.
Identities = 114/213 (53%), Positives = 139/213 (65%), Gaps = 2/213 (0%)

```
Query 4 FRFQARYALLTYPQCGDLDPFAVNVNHLAAGLGAECIIGRENHADGGLHLHAFVDFGVK 63
FRF A+Y LLTYPQCGDLDP+AV + L LGAEC+IGRENH GG+HLHAF F K+ +
Sbjct 3 FRFAAKYGLLTYPQCGDLDPWAVNDMLGRLGAEVIGRENHQAAGGVHLHAFMFERKFES 62

Query 64 RNARAFDVEGCHPNVSPSRGTPEEGFDYA IKDGDVAVAGGLERPVA-GRVDATSGVWAEII 122
```

Sbjct 63 RN R FDV+G HPN+ TPE+G+ YA KDGDVVAG LE P V +S WA I
 RNVRVFDVDGHHENIVRGYSTPEDGWAYATKDGDVVAGALECPRPRTEVSESSKWARAI 122
 Query 123 NAESEQEFDLCKRLAPRSLVTSFTQLRAYAAWKFPPIRVPYETPEGVNIDTSWVAELNQ 182
 AE+ +EF+ + L PR+L SF LRAYA WK+ P R PY TPEG++ DTS + EL+
 Sbjct 123 LAETREEFFAIVAEIDPRALCVSFGSLRAYADWKYRPARDPYITPEGISFDTSELPELDL 182
 Query 183 WVQQNLGRGETGGKLSCTSQRQLQSAATRASMIG 215
 WVQQ L G + GK S L+ + S G
 Sbjct 183 WVQQTLS-GLSTGKQPPPSAPLRGVPEQISARG 214

ProtSweep
A Protein Identification Tool

Results for /home/vir088/rindneu/gemi/MSSI2.225.232r.pep

Number of residues: 232	Molecular weight: 25.37 kD
Isoelectric point: 7.34	

[more...](#)

Prediction of protein localization:

cyto_nucl	12.1667 %
cyto	11 %
nucl	9 %

Identified Protein & Transcript:

Protein:
id: sptrembl:G9ITS8_9VIRU
valuation: weakly similar
scientific organism: Meles meles
abstract: G9its8 SubName: Full=Replication-associated protein; 4/2014Length=237
isoform:

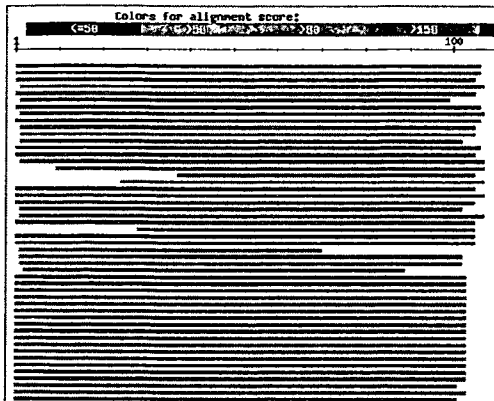
GO INFORMATION:

ACCESSION	GROUP	NAME	EVIDENCE
GO:0016888	molecular function	endodeoxyribonuclease activity, producing 5'-phosphomonoesters	IEA:InterPro
GO:0006260	biological process	DNA replication	IEA:InterPro

INTERPRO INFORMATION:

ID	SHORT NAME
IPR022690	Gemini_ALI_REP_cat-dom
IPR022692	Gemini_ALI_REP_central

ORF 108aa:



```
>>>sptrembl:TIYSF1_9VIRU Tlysf1 SubName: Full=Replication as... 174 1e-41
>>>sptrembl:TIYRW0_9VIRU Tlyrw0 SubName: Full=Replication as... 151 1e-34
>>>sptrembl:K0A1L7_9VIRU K0a117 SubName: Full=Spliced replic... 149 5e-34
```

```
>>>sptrembl:TIYSF1_9VIRU Tlysf1 SubName: Full=Replication associated protein
(Gemyrcircularvirus 9);. 4/2014
Length=325
Score = 174 bits (441), Expect = 1e-41, Method: Compositional matrix adjust.
Identities = 79/110 (71%), Positives = 92/110 (83%), Gaps = 3/110 (2%)
```

```
Query 1 MGKTIWARS LGNHAYFGGLFSLDENLDGVEY AIFDDFGGIKFLPSYKFWLGHQKQFYVTD 60
      GKT+WARSLG+HAYFGGLFS+DE++ YA+FD DFGG+KFLPSYKFW+GHQKQFYVTD
Sbjct 211 TGKTVWARS LGDHAYFGGLFSMDEDISSATYAVFDDFGGLKFLPSYKFWMGHQKQFYVTD 270

Query 61 KYKGKKLVHWARPSIWLSNSDPRDELGV---DTEWLNANCDFVYLDSPIV 107
      KYKGKKLV W RPSIW++N DPR+E GV + EWL NC F+ LDS IV
Sbjct 271 KYKGKKLV EWGRPSIWINNEDPREEHGVR CDEIEWLEQNCIFIRLDSTIV 320
```

```
>>>sptrembl:TIYRW0_9VIRU Tlyrw0 SubName: Full=Replication associated protein;. 2/2014
Length=320
```

```
Score = 151 bits (381), Expect = 1e-34, Method: Compositional matrix adjust.
Identities = 70/108 (64%), Positives = 87/108 (80%), Gaps = 1/108 (0%)
```

```
Query 1 MGKTIWARS LGNHAYFGGLFSLDENLDGVEY AIFDDF-GGIKFLPSYKFWLGHQKQFYVT 59
      +GKT+WARSLGNHAYFGGLFSL+E L+GV+YAIFDD GG+KF SYKFWLG Q QF+ T
Sbjct 208 LGKTLWARS LGNHAYFGGLFSL EEELEGV DYAIFDDMQGGLKFFHSYKFWLGAQSQFWAT 267

Query 60 DKYKGKKLVHWARPSIWLSNSDPRDELGV DTEWLNANCDFVYLDSPIV 107
      DKY+GKKL+HW RPSI++ N +P + GVD +WL NCDFV + ++
Sbjct 268 DKYRGKKLIHWGRPSIYICNQNPLCDEGV DHDWLIGNCDFVEITESLL 315
```

```
>>>sptrembl:K0A1L7_9VIRU K0a117 SubName: Full=Spliced replication-associated protein;.
4/2014
Length=318
```

```
Score = 149 bits (376), Expect = 5e-34, Method: Compositional matrix adjust.
Identities = 67/107 (62%), Positives = 87/107 (81%), Gaps = 1/107 (0%)
```

```
Query 1 MGKTIWARS LGNHAYFGGLFSLDENLDGVEY AIFDDF-GGIKFLPSYKFWLGHQKQFYVT 59
      +GKT+WARSLG HAYFGGLF LDE+L+ V+YA+FD D GG++F +YKFWLG QKQFY T
Sbjct 206 LGKTLWARS LGRHAYFGGLFCLDESLEDVDYAVFDDMQGGLFFHAYKFWLGAQKQFYAT 265

Query 60 DKYKGKKLVHWARPSIWLSNSDPRDELGV DTEWLNANCDFVYLDSP I 106
      DKY+GK+LV+W RPSI++SN++P + G D +WL NC V++D PI
Sbjct 266 DKYRGKQLVNWGRPSIYISNTNPLADKGADVDWLMGNCVIVHVD RPI 312
```

ProtSweep

A Protein Identification Tool

Results for /home/vir088/rindneu/gemi/MSSI2.225.108r.pep

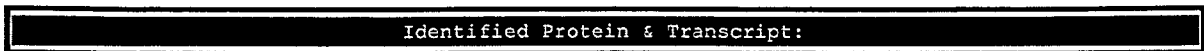


Number of residues: 108	Molecular weight: 12.46 kD
Charge: -3	Isoelectric point: 5.63

[more...](#)

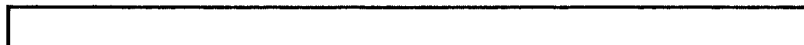
Prediction of protein localization:

mito	11 %
cyto	10.5 %
cyto_nucl	10 %



Identified Protein & Transcript:

Protein:
id: <u>sptrembl:S6DDT1 9GEMI</u>
valuation: putative
scientific organism: <u>Euphorbia caput-medusae</u>
abstract: <u>S6ddt1 SubName: Full-Replication associated protein, . 12/2013Length=130</u>
isoform:

ORF 95aa:

>>>sptrembl:V8NBT2 OPHHA V8nbt2 SubName: Full=Putative E3 ub... 35.4 8.8

>>>sptrembl:V8NBT2_OPHHA V8nbt2 SubName: Full=Putative E3 ubiquitin-protein ligase HERC6;
 Flags: Fragment;. 4/2014
 Length=735

Score = 35.4 bits (80), Expect = 8.8, Method: Composition-based stats.
 Identities = 31/95 (32%), Positives = 47/95 (49%), Gaps = 12/95 (12%)

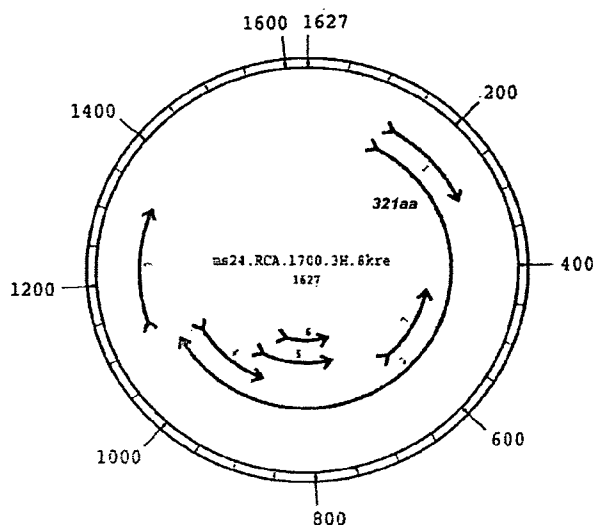
Query 9 TSSGVRKILIGQISLLPPLILVALRLSPNTNSASSAPVMKQVLRACLSFGTTPFVPSL--- 65
 TSSGV+ +L SLL L L R SP A + + V C++ G +PSL
 Sbjct 302 TSSGVKIVLTIDSSLLDHLFLALPRSSPHQEALA--IFLLVPECCVALGA-CRIPSLALR 358

Query 66 -----TKMTLSPTLSVIGPFLPQVYEVLGICTSWI 95
 T ++ + +++GPFL Q+ ++L C S I
 Sbjct 359 FAEAVTGLSKGSS-NILGPFLEQIVQMLKTCVSAI 392

*MSSI1.162 (1627bp)**(MSSI = Multiple sclerosis serum isolate)**(ms24.RCA.1700.3)*

GACTTCTGATTGATTGATGCGCTGCCGACTTTTGATAATTTTTGATAGATGATTTTCTTGACTGAAATAAAAAACCCCTTA
 TTGTTGTAAGTGCTTGGTTCACAACCTTAATAAGAGGTTTAATTATGATGTCACATGATGGCACACCTGTCAATTTTTGTCC
 GATATTCAACCCAGTGAGAAAAGCTGGGACACTCACCGAGCCGAAAGCGTCCGATTGCTTTACAGTCTTTCGACTGA
 GTTCACCAAATACGCTAGCCGTATTTATGACTGTTCCCAAATCCTAAAATTTGCCCTACCCCTGATAAGCTGGTACTTAAAC
 ACGCCTTTTTCTGTCGTGTTAGGTATTGCCCTGTTTGCCAGTGGCGTAGGTGTTTATTATGGCGGGCGGTGATGTTTCAGCAG
 TTACCAGCAATCAAAGAGAAGTACCCTAGTTATCGTTGGGTGTTTCTTACCCTTACTGTCAAAAATCCCCAGTTACTGAATT
 AAGGGACACCTTAAAGGCTATGAATTCAGCTTGGCAACGCTAGCACAACTAAACGCTTTAAAGGGTCTGTTAAAGGCTTTA
 TTCGCACAACCGAAGTTACAAGGGGTAAAGATGGCGACATGATGGCTCACCCCATTTCCATGCGCTGTTATTAGTTCAAAGT
 AATTATTTCACTACAAATTATATTAACAAAATGATTGGGTGCGAATGTGGCAAAAGGCGTTAAGGGTCTGATTATGCCCCATC
 GGTCAACGTAAAAGCCGTTAAACCCCTAAAAAGGGTGAAAAGGATAATCTCGATAAGGCGATATGTGAAACGCTCAAATATA
 GCGTTAAACCGTCTGATATTGCCAAGATGACGATGGGGGCGAATGGTTGCATGAAATGACCAGACAAACCTAAATATGCGT
 TTTATTGCCACTGGTGGCATTTTGAAGGGTGTACTTAAACCCGATGAGCAAGTAACACAACAAGAAATGCTTACCCCACTGG
 TGAAGATGAAGCACCAACCGAACAACAAACGGATTGGATTAGATTTTACCCCATCATGGGCGTTATGTTTTTTCGCCAGCTA
 TACGAATTTTGAAGTGTGTTTTTGCAGTGTGTAAGTTCGCCCTTATAGCAATGTGTTGTGCTGTAAGTGTGTAAGGCA
 TATCCAGCGTTAGCAATAAACGGCTTTATCGTTTATGCGTTAGCGTCTTTATATCCCAACCTTTCCCCCTTTGAAGTCCCA
 GCTTTTTTTGTGGCTGTCTATTGTTGCTCTTTATTGACCTTACGCCACATAAAAAAGTGGGTACATATCGGTTGAATTT
 GGGTTGCAATAGTCTAGGGACATTCAAAAATAATCCTTCTTGGTAGTATTACGAAATCTCATGTAATCCGATTGGTGGCGGT
 TTACCCAAAAACAACAAAGAAAAATTCGTCTTGCGAATTTTTATTGCTTTTTGCTTTTTGGGGTATAAAACCTATAGAAA
 AACTATAAAAAACCTATAGAATTATAGGTGCTTACTGGGGGTTGCCTTGATTATTATTAAAGGGATAACTCTGATTTTATT
 TTATATGATTTTTTATACTGGCAAACCTGTTGAATACCGCTTAAATACT .

No repeat

66% nucleotide similarity to *Psychrobacter* sp. PRwf-1 plasmid pRWF102,ORF – 321aa – replication protein 62% similarity to *Acetivobacter* generi MTCC9824.

This isolate was generated by restriction enzyme EcoRI digestion of the DNA sample after rolling circle amplification.

Figure 4

Colors for alignment score:

1000

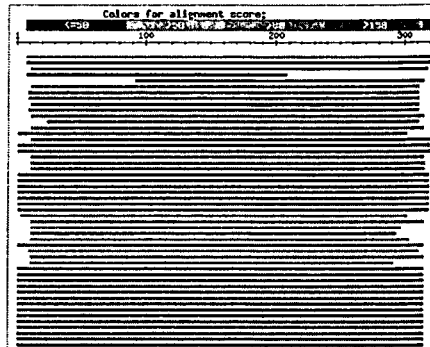
1

```

Query    680    AATTATATTAAACAAAATGATTGGGTCGA-AATGTGGCAAAAGGCGTTAAGGGTCGATTA    738
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Sbjct    1248    AATTATATTAATCAAGAAGAAT-GGACGAGACTGTGGCGAAAGGCGTTAAAAGTCGATTA    1306

Query    739    TGCCCCATCGGTC AACGTAAAAGCCGTTAA    768
          || ||| ||| ||| ||| ||| ||| |||
Sbjct    1307    TGAGCCAGTGGTGCATGTGCAAGTCGTTAA    1336

```

ORF 321aa**BlastP – Uniprot:**

>>>sptrembl:S7XX00_9GAMM S7xx00 SubName: Full=Replication pr... 425 1e-116
>>>sptrembl:N9KWE3_ACIBA N9kwe3 SubName: Full=Uncharacterize... 425 1e-116

>>>>sptrembl:S7XX00_9GAMM S7xx00 SubName: Full=Replication protein;. 12/2013
Length=328

Score = 425 bits (1093), Expect = 1e-116, Method: Compositional matrix adjust.
Identities = 197/316 (62%), Positives = 246/316 (77%), Gaps = 4/316 (1%)

Query 9 VNFLSDIQPSEKSWDTHRAEAESVRLLYSLSTEFTRYASRIYDCSQILKFAPTPD--KLV 66
+NFL++I P E SWD HR++AESV++LY+ S E KYA RI CS ILKF +P+ KLV
Sbjct 12 INFLTEISPHEVSWDEHRSDAESVKILYNYSVELAKYADRINGCSGILKFGVSPNQGLV 71

Query 67 LKHAFFCRVRYPVCQWRRSLLWRAVMFQQLPAIKEKYPSYRWVFLTLTVKNPPVTELRD 126
LK AFFCRVR+CPVCQWRRSLLWRAVMFQQLP I++ YP++RWVFLT+TV+NPP+ ELRD
Sbjct 72 LKQAFFCRVRHCPVCQWRRSLLWRAVMFQQLPKIQKLYPTHRWWFLTMTVRNPPIELRD 131

Query 127 TLKAMNSAWQRLAQTKRFGVVGKGFIRTEVTRGKDGDMMAHPHFHALLLVQSNYFTTNY 186
TLK MN +WQRL QTK FK V GFIRTEVTRGKD +MM+HPH+HALLLV+ YF+ Y
Sbjct 132 TLKHMNESWQRLIQTKAFKSGVAGFIRTEVTRGKDAEMMSHPHYHALLLVKGPFYSKY 191

Query 187 IKQNDWVEMWQKALRVDYAPSVNVKAVKPP--KKGEKDNLDKAICETLKYSVKPSDIAKDD 245
I Q++WVEMW KALR DY PSNVKAA++ + K L+K ICETLKYSVKPSD+A +
Sbjct 192 INQSEWVEMWAKALRADYLPVNVKAIKAGNNEKSKKALEKQICETLKYSVKPSDLAVER 251

Query 246 DGGEWLHEMTRQTLNMRFIATGGILKGVLPDEQVTQQEMLTPTGEDEAPTEQKRIGFRF 305
D G WLHEMT+Q MRF+ATGG+LKG+LKP++++T +EM++ + E E E R+ F+F
Sbjct 252 DKGAWLHEMTKQVHKMRFVATGGVLKGLKPDDEITNEEMISSSEETEDVGE-GRVAFQF 310

Query 306 YPHHGRYVFSPAHTNF 321
+ RYV++P +
Sbjct 311 NSEYRRYVYAPKFNEY 326

>>>>sptrembl:N9KWE3_ACIBA N9kwe3 SubName: Full=Uncharacterized protein;. 12/2013
Length=328

Score = 425 bits (1093), Expect = 1e-116, Method: Compositional matrix adjust.
Identities = 197/316 (62%), Positives = 246/316 (77%), Gaps = 4/316 (1%)

Query 9 VNFLSDIQPSEKSWDTHRAEAESVRLLYSLSTEFTRYASRIYDCSQILKFAPTPD--KLV 66
+NFL++I P E SWD HR++AESV++LY+ S E KYA RI CS ILKF +P+ KLV
Sbjct 12 INFLTEISPHEVSWDEHRSDAESVKILYNYSVELAKYADRINGCSGILKFGVSPNQGLV 71

Query 67 LKHAFFCRVRYPVCQWRRSLLWRAVMFQQLPAIKEKYPSYRWVFLTLTVKNPPVTELRD 126
LK AFFCRVR+CPVCQWRRSLLWRAVMFQQLP I++ YP++RWVFLT+TV+NPP+ ELRD
Sbjct 72 LKQAFFCRVRHCPVCQWRRSLLWRAVMFQQLPKIQKLYPTHRWWFLTMTVRNPPIELRD 131

Query 127 TLKAMNSAWQRLAQTKRFGVVGKGFIRTEVTRGKDGDMMAHPHFHALLLVQSNYFTTNY 186
TLK MN +WQRL QTK FK V GFIRTEVTRGKD +MM+HPH+HALLLV+ YF+ Y
Sbjct 132 TLKHMNESWQRLIQTKAFKSGVAGFIRTEVTRGKDAEMMSHPHYHALLLVKGPFYSKY 191

Query 187 IKQNDWVEMWQKALRVDYAPSVNVKAVKPP--KKGEKDNLDKAICETLKYSVKPSDIAKDD 245
I Q++WVEMW KALR DY PSNVKAA++ + K L+K ICETLKYSVKPSD+A +
Sbjct 192 INQSEWVEMWAKALRADYLPVNVKAIKAGNNEKSKKALEKQICETLKYSVKPSDLAVER 251

Query 246 DGGEWLHEMTRQTLNMRFIATGGILKGVLPDEQVTQQEMLTPTGEDEAPTEQKRIGFRF 305
D G WLHEMT+Q MRF+ATGG+LKG+LKP++++T +EM++ + E E E R+ F+F

Sbjct 252 DKGAWLHEMTRQVHKMRFVATGGVLKGILKPDDEITNEEMISSSEETEDVGE-GRVAFQF 310
Query 306 YPHHGRYVFSPAHTNF 321
+ RYV++P +
Sbjct 311 NSEYRRYVYAPKFNEY 326

>>>sptrembl:A5WI54_PSYWF A5wi54 SubName: Full=Uncharacterized protein;. 3/2014
Length=377

Score = 311 bits (796), Expect = 3e-82, Method: Compositional matrix adjust.
Identities = 151/331 (45%), Positives = 216/331 (65%), Gaps = 23/331 (6%)

ProtSweep

A Protein Identification Tool

Results for /home/vir088/rindneu/tsekompl/MSSI1.162.321.pep

Number of residues: 321 Molecular weight: 37.27 kD
Charge: 12 Isoelectric point: 10.10
more...
Prediction of protein localization:
cyto 24 %
cyto nucl 16.5 %
nucl 5 %

Identified Protein & Transcript:

Protein:
id: sptrembl:H4VAM4 ECOLX
valuation: weakly similar
scientific organism: Escherichia coli
abstract: H4vam4 SubName: Full=Replication family protein;. 4/2013Length=322
isoform:

GO information:

accession	group	name	evidence
GO:0005727	cellular component	extrachromosomal circular DNA	IEA:InterPro
GO:0003677	molecular function	DNA binding	IEA:InterPro
GO:0006260	biological process	DNA replication	IEA:InterPro

Interpro information:

id	short name
IPR000989	Rep

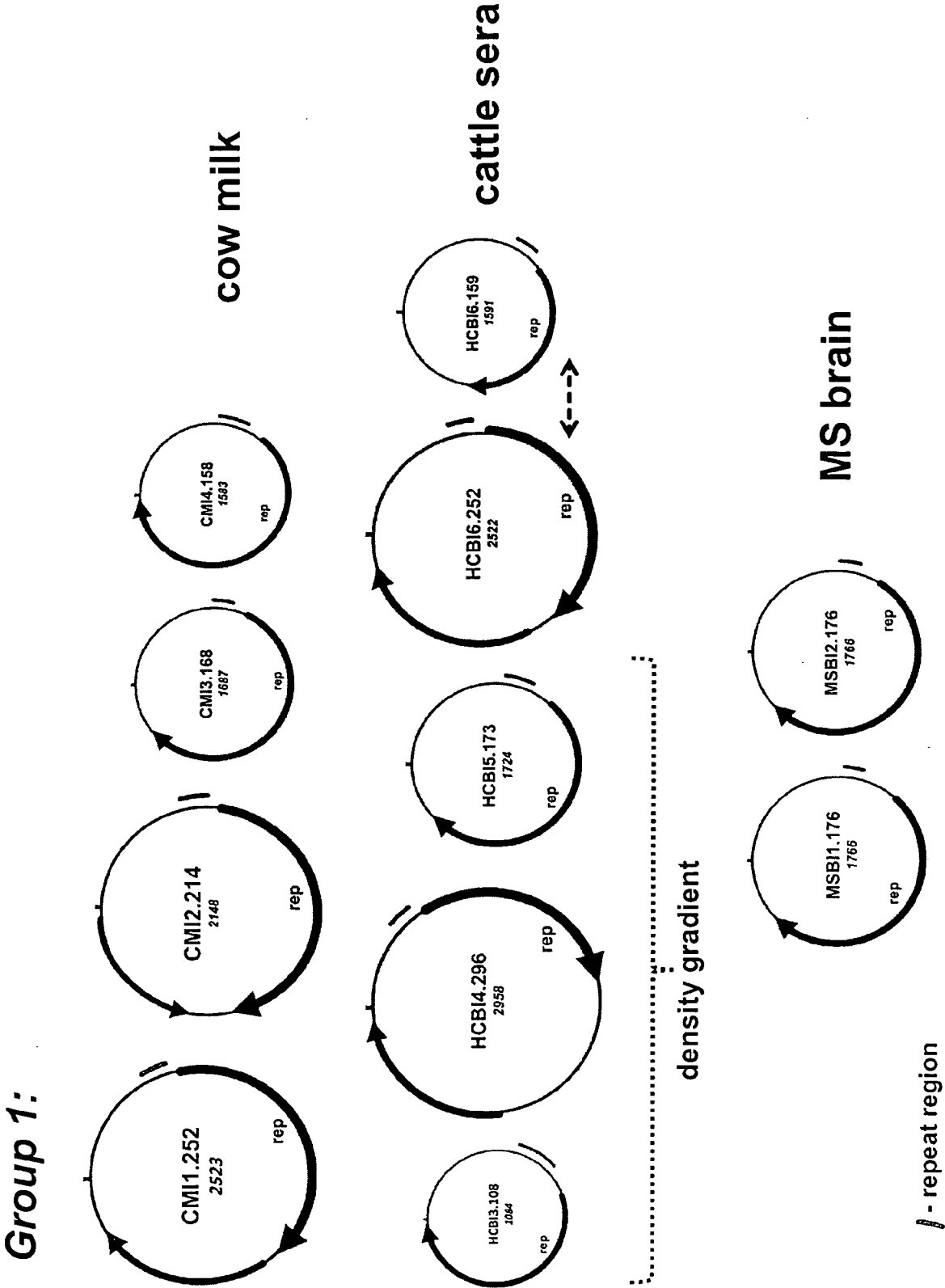
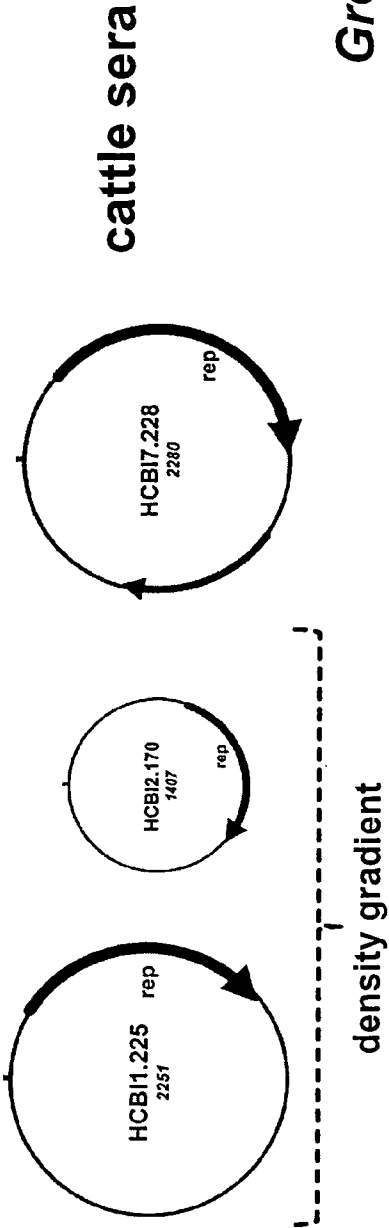
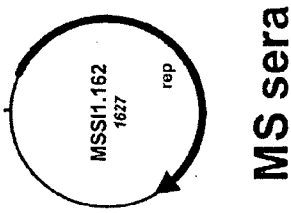


Figure 5

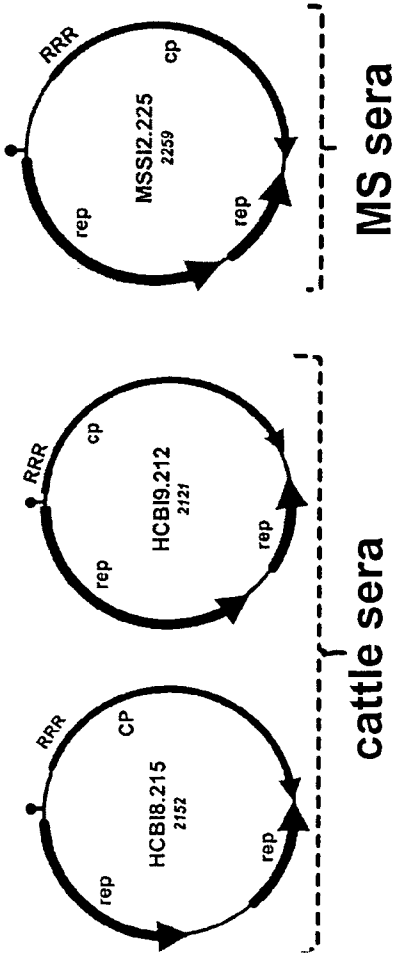
Group 2:



Group 4:



Group 3:



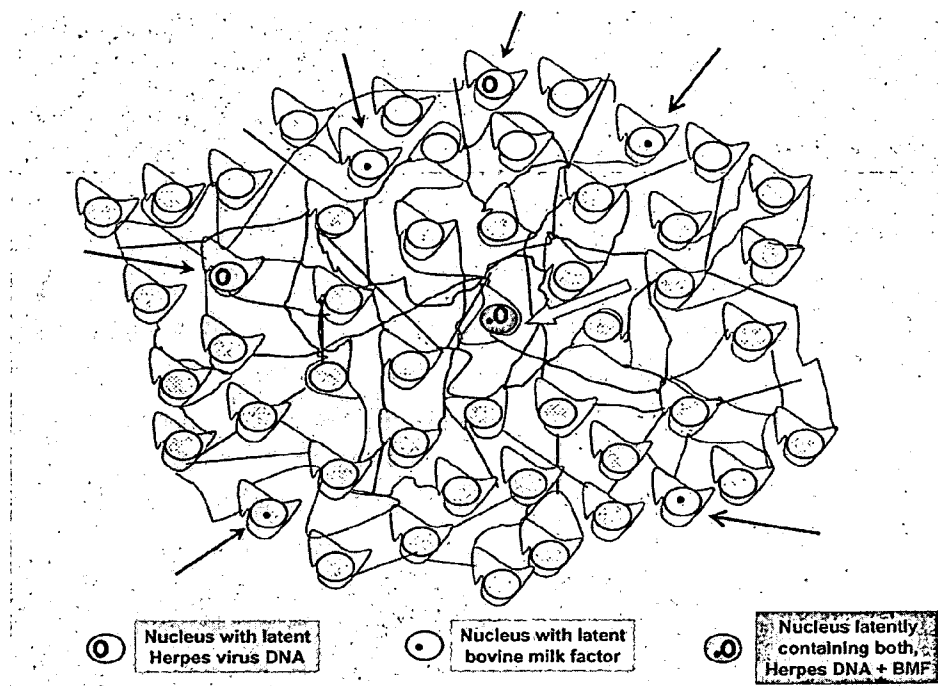


Figure 7

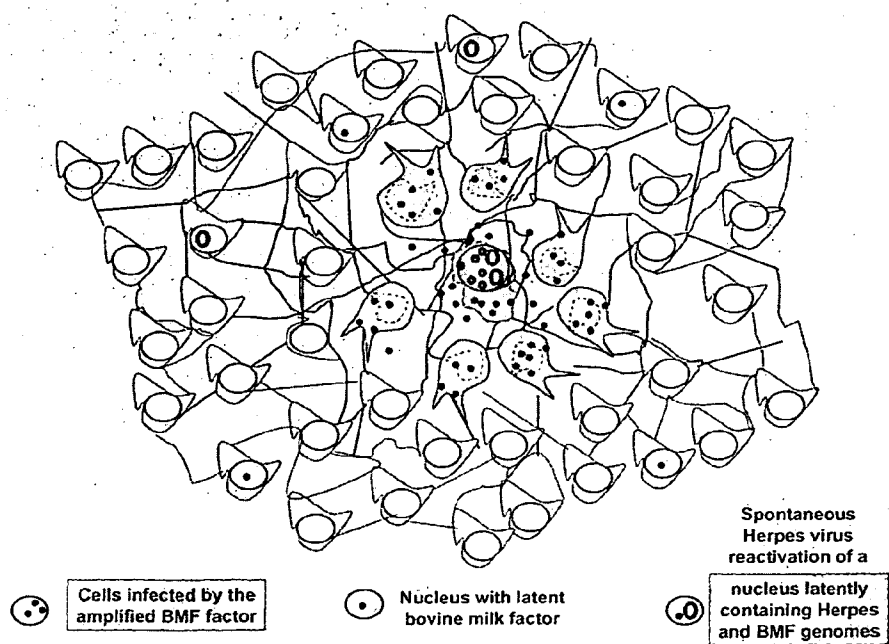


Figure 8

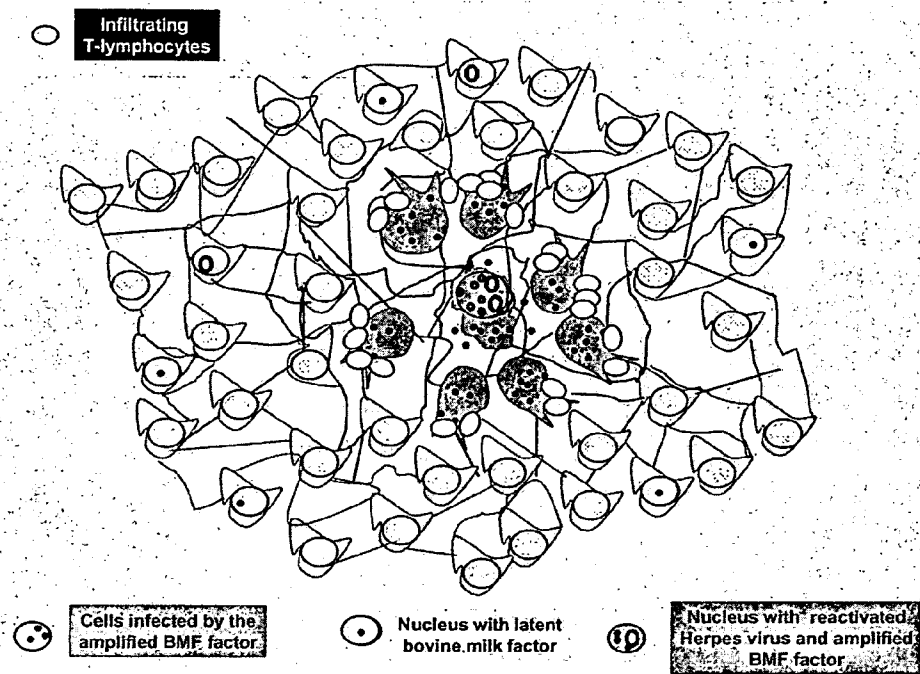


Figure 9

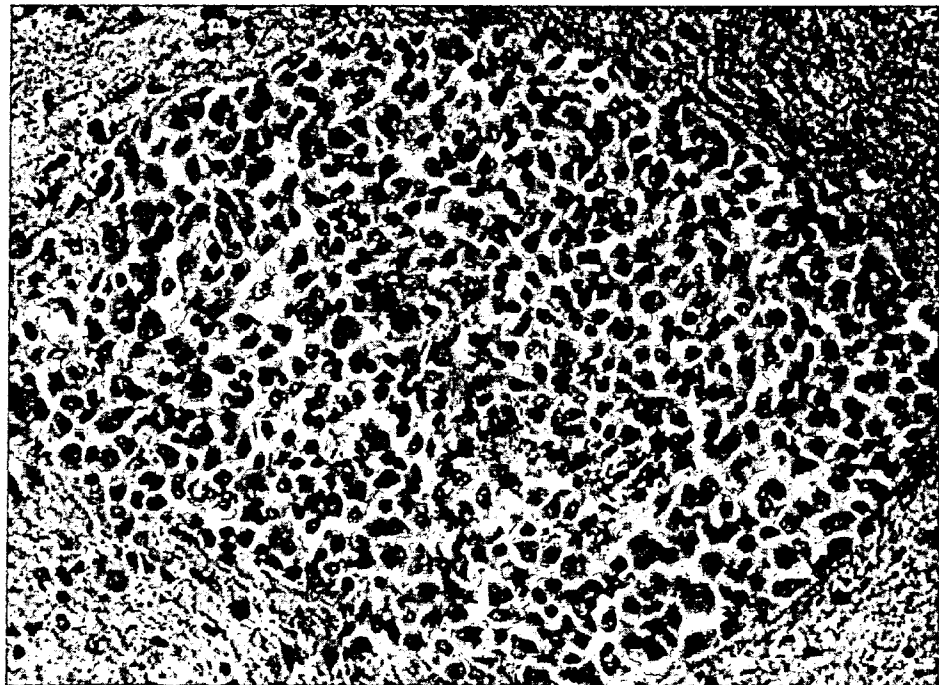


Figure 10

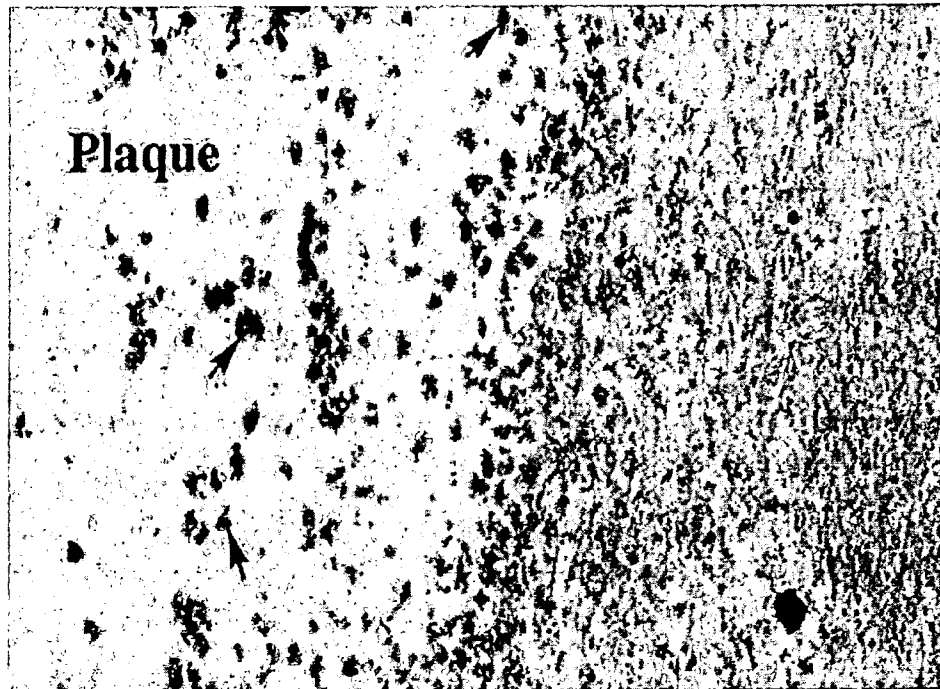
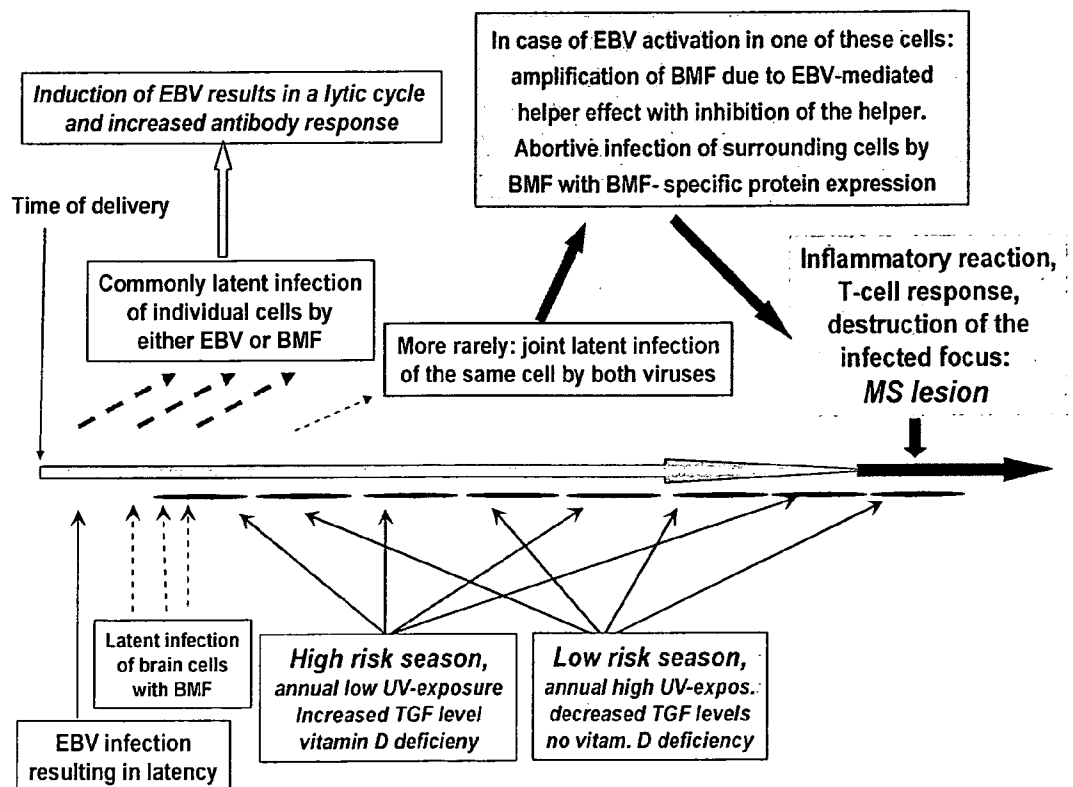


Figure 11



Vitamin D deficiency results in TGF β increase and higher rates of spontaneous EBV induction, seasonably pronounced in early spring (reduced UV-exposure)

Figure 12

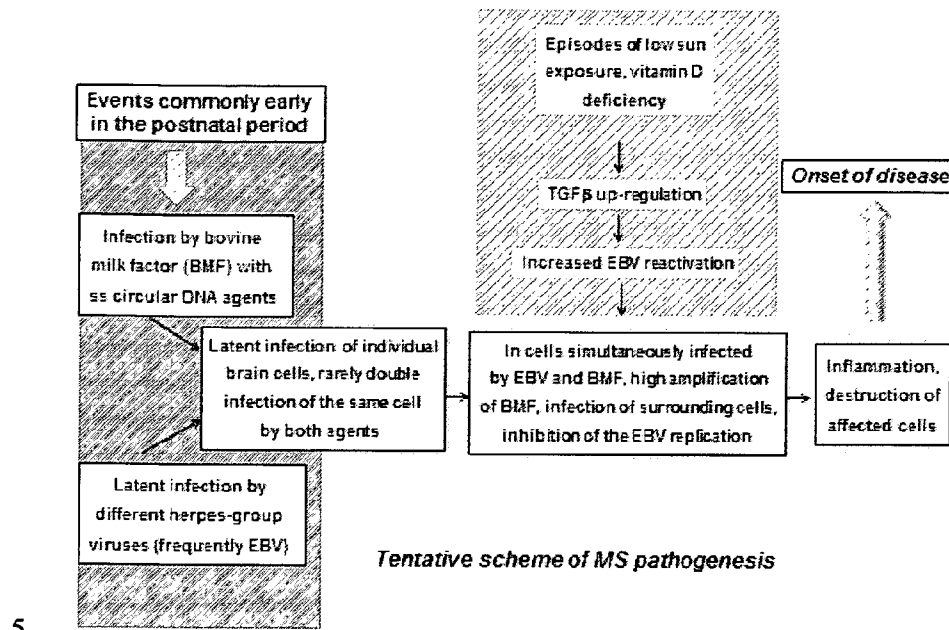


Fig. 13