



(22) Date de dépôt/Filing Date: 2014/10/30
(41) Mise à la disp. pub./Open to Public Insp.: 2015/05/07
(45) Date de délivrance/Issue Date: 2023/03/14
(62) Demande originale/Original Application: 2 922 872
(30) Priorité/Priority: 2013/10/30 (EP13005138.6)

(51) Cl.Int./Int.Cl. *C12N 15/12* (2006.01),
C07H 21/00 (2006.01), *C07K 14/47* (2006.01),
C07K 16/18 (2006.01), *C12N 15/11* (2006.01),
C12Q 1/6876 (2018.01), *C12Q 1/6883* (2018.01),
C12Q 1/6886 (2018.01), *G01N 33/564* (2006.01)

(72) Inventeurs/Inventors:
DE VILLIERS, ETHEL-MICHELE, DE;
GUNST, KARIN, DE;
FUNK, MATHIS, FR;
PEREZ, IRANZU LAMBERTO, ES
(73) Propriétaire/Owner:
DEUTSCHES KREBSFORSCHUNGSZENTRUM, DE
(74) Agent: BENOIT & COTE INC.

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

(57) Abrégé/Abstract:

Described are HCBI (Healthy Cattle Blood 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 as a target for cancer treatment and prevention.

ABSTRACT

Described are HCBI (Healthy Cattle Blood 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 as a target for cancer treatment and prevention.

**HCBI sequences as an early marker for the future development of
cancer and diseases of the CNS and as a target for cancer
treatment and prevention**

5

The present invention relates to HCBI (Healthy Cattle Blood Isolate) nucleotide sequences as well as probes and primers comprising part of said nucleotide sequences and antibodies against polypeptides encoded by said nucleotide sequences.

10 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 cancer treatment and prevention.

15 Several epidemiological analyses conducted in recent decades indicate that the long-term consumption of "red" meat processed by different ways (including smoked meat and meat as component of sausages) can be regarded as a risk factor for colon cancer (World Cancer Report 2013, zur Hausen 2009, zur Hausen 2012).

20 "Red" meat is regarded as comprising beef, pork, mutton, lamb and goat meat, in contrast to "white" meat (poultry meat/fish).

Thus far, chemical carcinogenic substances being produced during roasting, grilling, barbecuing, smoking and air-drying
25 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
30 role as regards the development of colon cancer. Since, in 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
35 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 zebu, water buffalo or yak might be of importance as a main risk factor.

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

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 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 five new viral nucleic acid components of a could be isolated. The DNA sequences and open reading frames of these components showed a recognizable relationship to 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 and diseases of the CNS.

Brief description of the drawings

Figure 1: Nucleotide sequence of HCBI1.225 and putative open reading frames

Also shown are alignments to nucleotide sequences found in data bases and alignments of peptides derived from open reading frames of HCBI1.225 to corresponding "Sphinx" amino acid sequences.

Figure 2: Nucleotide sequence of HCBI2.170 and putative open reading frames

Also shown are alignments to nucleotide sequences found in data bases and alignments of peptides derived from open reading frames of HCBI2.170 to corresponding "Sphinx" amino acid sequences.

Figure 3: Nucleotide sequence of HCBI3.108 and putative open reading frames

Also shown are alignments to nucleotide sequences found in data bases and alignments of peptides derived from open reading frames of HCBI3.108 to corresponding "Sphinx" amino acid sequences.

Figure 4: Nucleotide sequence of HCBI4.296 and putative open reading frames

Also shown are alignments to nucleotide sequences found in data bases and alignments of peptides derived from open reading frames of HCBI4.296 to corresponding "Sphinx" amino acid sequences.

Figure 5: Nucleotide sequence of HCBI5.173 and putative open reading frames

Also shown are alignments to nucleotide sequences found in data bases and alignments of peptides derived from open reading
5 frames of HCBI5.173 to corresponding "Sphinx" amino acid sequences.

Accordingly, the present invention relates to an HCBI
10 polynucleic acid comprising:

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

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

15 (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-
20 given nucleotide sequences.

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
25 analogues or modified nucleotides or may have been adapted for diagnostic or therapeutic purposes. A polynucleic acid may also comprise a double stranded cDNA clone which can be used, for example, for cloning purposes.

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

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 which show at least 90%, 95% or 98% 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 5 can be characterized and isolated according to any of the 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 etc.

The present invention also provides fragments of the nucleotide sequences of the present invention described above that are, preferably, capable of replicating autonomously as described in Example 1. An autonomous replicating nucleotide sequence comprises a nucleotide sequence of the replication gene or a fragment thereof which is capable of inducing autonomous replication. 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 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 5 or which fragments are still capable of replicating 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.

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

10 The HCBI polynucleic acids of the invention 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.

15 The present invention also relates to an oligonucleotide primer 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 polynucleic acid of the invention.

20 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 allow priming the synthesis of the extension products.
25 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.

30 The fact that amplification primers do not have to match exactly with corresponding template sequence to warrant proper amplification is amply documented in the literature. The amplification method used can be , for example, polymerase chain reaction (PCR), ligase chain reaction (LCR), nucleic acid
35 sequence-based amplification (NASBA), transcription-based

amplification system (TAS), strand displacement amplification (SDA) or amplification by means of Q β replicase or any other suitable method to amplify nucleic acid molecules using primer extension. During amplification, the amplified products can be
5 conveniently labelled either using labelled primers or by incorporating labelled nucleotides.

Labels may be isotopic (³²P, ³⁵S, etc.) or non-isotopic (biotin, digoxigenin, etc.). The amplification reaction is repeated
10 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 determine the orf by translating the sequence of the sample into
15 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
20 (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.

25 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.

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

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
5 the target sequence of a HCBI polynucleic acid to be detected.

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.

10

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
15 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
20 homopolymer tailing, coupling with different reactive groups such as aliphatic groups, NH₂ groups, SH groups, carboxylic groups, or coupling with biotin or haptens.

The oligonucleotides according to the present invention, used as
25 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
30 obtain the required specificity and sensitivity. However, the eventual results will be essentially the same as those obtained with the unmodified oligonucleotides.

The introduction of these modifications may be advantageous in
35 order to positively influence characteristics such as

hybridization kinetics, reversibility of the hybrid-formation, biological stability of the oligonucleotide molecules, etc.

5 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.

10 The present invention also relates to a recombinant expression vector comprising an HCBI 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.

15 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 recombinant molecules, BCG or adenoviral vectors, as well as avipox recombinant viruses.

20 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.

25

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.

30

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.

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.

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 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 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.

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.

The segment of the HCBI 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 source, but particularly preferred constructs according to the present invention contain signal sequences appearing in the HCBI 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 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 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. 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 man skilled in the art for 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 amino acid sequence encoded by an HBCI 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 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 like. Included within the definition are, for example, polypeptides containing one or more analogues of an amino acid
5 (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 occurring and non-naturally occurring.

10 The polypeptides according to the present invention contain preferably at least 3, preferably 4 or 5 contiguous HCBI amino acids, 6 or 7 preferably however at least 8 contiguous HCBI amino acids, at least 10 or at least 15.

15 The polypeptides of the invention, and particularly the 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.

20 The present invention also relates to a method for production of 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
25 the appropriate regulatory elements, (b) culturing said 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
30 immunization with at least one polypeptide as defined above, 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
35 antibodies with different epitopic specificities, as well as

distinct monoclonal antibody preparations. Monoclonal antibodies are made from an antigen containing, e.g., a polypeptide encoded by an HCBI polynucleic acid of the invention or a fragment thereof by methods well known to those skilled in the art. As
5 used herein, the term „antibody“(Ab) or „monoclonal 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
10 more rapidly from the circulation, and may have 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
15 chimerical, single chain, and humanized antibodies.

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
20 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.

25

The present invention also relates to a diagnostic kit for use in determining the presence of a HCBI polynucleic acid or polypeptide of the invention, said kit comprising a primer, a probe, and/or an antibody of the invention.

30

The present invention also relates to a method for the detection of an HCBI polynucleic acid according to the invention present in a biological sample, comprising: (a) optionally extracting sample polynucleic acid, (b) amplifying the polynucleic acid as
35 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
5 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
10 during the polymerase step of the amplification or labelled 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
15 of an HCBI polynucleic acid according to the 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.

20 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
25 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,
30 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
35 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 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 detect hybrids formed between the oligonucleotide probes and the HCBI 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 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.

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 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.

The present invention also relates to a method for detecting a polypeptide encoded by an HCBI polynucleic acid of the present invention or an antibody against said polypeptide present in a biological sample, comprising: (a) 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.

5 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 antigens of
10 the present invention may be employed in virtually any assay format that employs a known antigen to detect antibodies. Of course, a format that denatures the HCBI 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
15 suspected of containing HCBI 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
20 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
25 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
30 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
35 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 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 antibodies will bind due to complex formation. In a competitive format, the amount of HCBI antibodies in the sample is deduced by monitoring the competitive effect on the binding of a known amount of labeled antibody (or other competing ligand) in the complex.

Complexes formed comprising anti-HCBI 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 antibodies in the complex may be detected using a conjugate of anti-xenogeneic Ig complexed with a label (e. g. an enzyme label).

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

There currently exist three specific types of particle
10 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
15 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
20 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,
25 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
30 beta-galactosidase. Examples of enhancer compounds include 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
35 developing diseases like cancer or an autoimmune disease due to

the deleterious effects of the presence of a subgenomic HCBI 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.

5

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

10 The generation of suitable antisense oligonucleotides or iRNAs includes determination of a site or sites within the HCBI 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 intragenic site is (a) the
15 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 chosen which are
20 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
25 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,
30 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
35 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
5 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 polynucleic acid of the present
10 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
15 useful for gene therapy. Preferred vectors useful for gene therapy are viral vectors, e.g. adenovirus, adeno-associated virus, herpes simplex virus, vaccinia, or, more preferably, an RNA virus such as a retrovirus. Even more preferably, the retroviral vector is a derivative of a murine or avian
20 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,
25 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
30 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
35 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
5 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 polynucleotide sequences of the invention may also serve as a suitable vector itself, either composed solely of rearranged
10 HCBI sequences or of chimeric HCBI 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
15 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.

20 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
25 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
30 result in increased stability are known to the person skilled in the art, preferably such modification is a phosphorothioate linkage.

A preferred oligonucleotide mimetic is an oligonucleotide
35 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 aminoethylglycine backbone. The nucleobases are retained and are
5 bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone.

Modified oligonucleotides may also contain one or more substituted or modified sugar moieties. Preferred
10 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 C₂ to C₁₀ alkenyl and alkynyl. A particularly preferred modified
15 sugar moiety is a 2'-O-methoxyethyl sugar moiety.

Antisense-oligonucleotides of the invention may also include nucleobase modifications or substitutions. Modified nucleobases include other synthetic and natural nucleobases such as 5-
20 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 and 2-thiocytosine etc., with 5-methylcytosine substitutions
25 being preferred since these modifications have been shown to increase nucleic acid duplex stability.

Another modification of the oligonucleotides of the invention involves chemically linking to the oligonucleotide one or more
30 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
35 polyamine or a polyethylene glycol chain, or adamantane acetic

acid, a palmityl moiety, or an octadecylamine or hexylamino-carbonyl-oxycholesterol moiety.

The present invention also includes antisense compounds which
5 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
10 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.
15 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,
20 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.
25 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.

30

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.
Preferably, in a pharmaceutical composition, such compound as
35 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 toxic to the host to which it is administered. Examples of
5 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 conventional methods and the active compound can be administered
10 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 course and the severity of the disease, leading to the reduction
15 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.

Administration of the suitable compositions may be effected by
20 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 pharmaceutical composition. The dosage regimen will be
25 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 administered, time and route of administration, the kind of
30 therapy, general health and other drugs being administered concurrently.

In a preferred embodiment of the present invention, the disease that can be prevented/treated is cancer, preferably colorectal,
35 breast, prostate and pancreatic cancer, or a disease of the CNS,

preferably Alzheimer's disease or multiple sclerosis (MS). The terms "cancer" and "disease of the CNS" also comprise early stages of said diseases.

Due to a compilation of global epidemiological data published by the inventor (zur Hausen, 2012), evidence was mounting for putative infectious factors, playing a role in cancer, in particular of the colon, and potentially also in other human cancers, transmitted from cattle to humans either by beef or dairy product consumption. For these reasons the inventors initiated the analysis of 130 sera from healthy cows for the presence of as yet unknown infectious agents. After isolating in an initial series the five novel DNAs shown in Fig. 1-5, the inventors started to look for the presence of their DNA in cells derived from cancer, in particular colon cancer. Concomitantly materials from patients with multiple sclerosis were analyzed. Since multiple sclerosis represents an important human disease with often disastrous consequences for the affected patients, the role of isolates according to the present invention in neurological disorders, specifically intensively on MS, has been tested. The results show that the sequences are only found in diseased patients but, thus far, have not been found in healthy subjects. This applies to both the cancer patients as well as those suffering under CNS diseases.

The present invention also relates to a vaccine for immunizing a mammal against an HCBI infection, comprising at least one polypeptide or HCBI polynucleic acid as defined above, in a pharmaceutically acceptable carrier.

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

The term "therapeutic" refers to a composition capable of

treating HCBI 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
5 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
10 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
15 appropriate effective amount can be readily determined using only routine experimentation. Preferred ranges of proteins for prophylaxis of HCBI 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
20 subsequent protection against an HCBI infection and an HCBI linked disease, respectively.

Pharmaceutically acceptable carriers include any carrier that does not itself induce the production of antibodies harmful to
25 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.

30

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-
35 isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-

isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl sn-glycero-3-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 -100µg/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

Bovine sera were separated on Optiprep (iodixanol) step gradients as previously described (Buck et al., 2004).

25 Fractions were collected and DNA extracted. Rolling circle amplification was performed (de Villiers et al., 2011) on each fraction before digestion with the restriction enzymes BamH1 and EcoR1. Resulting fragments were cloned into pUC18 or pUC19 prior to sequencing.

30

(B) In vitro replication (de Villiers et al., 2011) of HCB11.225

Linear full-length HCBI1.225 was transfected into 293TT cells (Lipofectamine). Cultures were passaged every 2-3 days. DNA and RNA were extracted from a fraction of the cells upon passaging. Input DNA was removed by DpnI digestion. DNA and RNA of
5 HCBI1.225 were demonstrated up to day 17 in cell culture.

Example 2

Characterization of the HCBI sequences

10

120 sera from healthy cows obtained from the "Veterinärinstitut der Universität Leipzig" (Professor Müller) were screened and after purification of viral particles, DNA extraction and "rolling circle" amplification (de Villiers et
15 al., 2011) possibly interacting circular molecules of viral DNA differing in size were isolated. The five DNA sequences and open reading frames of these components showed a recognizable relationship to sequences which were already described for transmissible spongiform encephalopathies (TSE)
20 for TSE-diseases of sheep, cattle and humans (Manuelidis, 2011). These agents were not further characterized and designated as "Sphinx virus". Since the DNA sequences (and open reading frames) obtained in the present invention from healthy cows show homologies to the corresponding sequences
25 described by Manuelidis but are not identical the designation HCBI (Healthy Cattle Blood Isolate) is proposed.

In this context, diseases of the CNS (e.g., Multiple Sclerosis, Alzheimer's disease) are also highly interesting
30 since the similar sequences described by Manuelidis are primarily found in the CNS.

The "Sphinx" sequences (Manuelidis, 2011) show high homologies to plasmid sequences of the bacterium *Acinetobacter* (Valle et al., 2008; Longkummer et al., 2013). The sequences obtained in the present invention also exhibit striking homologies to the corresponding plasmid sequences. Thus, at first glance it is tempting to speculate that the results are due to contamination of the samples with *Acinetobacter* viruses. Although meanwhile a plethora of *Acinetobacter* plasmids has been sequenced none of these sequences corresponds to the DNA sequences of the present invention that were obtained from three different batches of sera. Since initially the isolation of these DNAs was achieved by purification of the viruses via density gradient centrifugation the apparently common presence of these viruses in blood of cow sera is not necessarily attributable to viral contamination derived from bacteria. Moreover, the presence of DNA sequences of these viruses in human cancer cells is incommensurate with such interpretation.

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* (*Acinetobacter calcoaceticus*). Unequivocal results were obtained by the group of Ebringer for three strains of *Acinetobacter* (*Acinetobacter lwoffii*, *A. radioresistens* 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.

5

References

10 Buck CB, Pastrana DV, Lowy DR, Schiller JT. Efficient intracellular assembly of papillomaviral vectors. J. Virol. 2004; 78:751-757.

Chapman MD, Hughes LE, Wilson CD, Namnyak S, Thompson EJ, Giovannoni G. No evidence for production of intrathecal
15 immunoglobulin G against Acinetobacter or Pseudomonas in multiple sclerosis. Eur Neurol. 2005; 53(1):27-31.

de Villiers EM, Borkosky SS, Kimmel R, Gunst K, and Fei JW. (2011) The diversity of Torque teno viruses: In vitro
20 replication leads to the formation of additional replication-competent subviral Molecules. J Virol 2011; 85(14):7284-7295

Ebringer A, Hughes L, Rashid T, Wilson C. Acinetobacter Immune Responses in Multiple Sclerosis: Etiopathogenetic Role and Its
25 Possible Use as a Diagnostic Marker. Arch Neurol. 2005; 62:33-36.

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.

Hughes, L.E., Bonell, S., Natt, R.S., Wilson, C., Tiwana, H.,
5 Ebringer, A., Cunningham, P., Chamoun, V., Thompson, E.J.,
Crocker, 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-
10 1188.

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

15 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.*
20 *Review Biochem.* 1997; 66:61-92

Vallenet D, Nordmann P, Barbe V, Poirel 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
25 C, Weissenbach J, Cruveiller S. Comparative analysis of
*Acinetobacter*s: 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,
30 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 et al., The Search for infectious causes of human cancer: Where and Why, Virology 2009, Vol. 392, No. 1: 1-10

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

zur Hausen H. World Cancer Report, IARC, Lyon, 2013.

10 Whitley C, Gunst K, Müller H, Funk M, zur Hausen H, de Villiers
EM.: Novel replication-competent circular DNA molecules from
healthy cattle serum and milk and multiple sclerosis-affected
human brain tissue. Genome Announc. 2014 Aug 28;2(4). pii:
15 e00849-14. doi: 10.1128/genomeA.00849-14.

Claims

1. A Healthy Cattle Blood Isolate (HCBI) polynucleic acid comprising:
 - (a) a nucleotide sequence of SEQ ID NO: 72 encoding an antigenic HCBI polypeptide;
 - (b) a nucleotide sequence having at least 95% identity to the full length of the nucleotide sequence of (a);
 - (c) a fragment of a nucleotide sequence of (a) or (b) that induces autonomous replication;
 - (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.
2. An oligonucleotide primer comprising part of an HCBI polynucleic acid of claim 1, said primer acting as primer for specifically sequencing or specifically amplifying the nucleic acid of a certain HCBI isolate containing a nucleotide sequence of claim 1.
3. An oligonucleotide probe comprising part of an HCBI polynucleic acid of claim 1, said probe acting as a hybridization probe for specific detection of the nucleic acid of a certain HCBI isolate containing a nucleotide sequence of claim 1.
4. An expression vector comprising an HCBI polynucleic acid of any one of claims 1 to 3 operably linked to prokaryotic, eukaryotic or viral transcription and translation control elements.
5. A host cell transformed with an expression vector according to claim 4.
6. A polypeptide being encoded by an HCBI polynucleic acid of claim 1.
7. An antibody or antigen binding fragment thereof specifically binding to a polypeptide of claim 6.

8. A method for the detection of an HCBI polynucleic acid according to claim 1 in a biological sample, comprising: (a) extracting sample polynucleic acid, (b) amplifying the polynucleic acid with at least one primer according to claim 2, and (c) detecting the amplified polynucleic acid.

9. The method according to claim 8, wherein the at least one primer according to claim 2 is a labelled primer.

10. A method for the detection of an HCBI polynucleic acid according to claim 1 in a biological sample, comprising: (a) extracting sample polynucleic acid, (b) hybridizing the polynucleic acid with at least one probe according to claim 3, and (c) detecting the hybridized polynucleic acid.

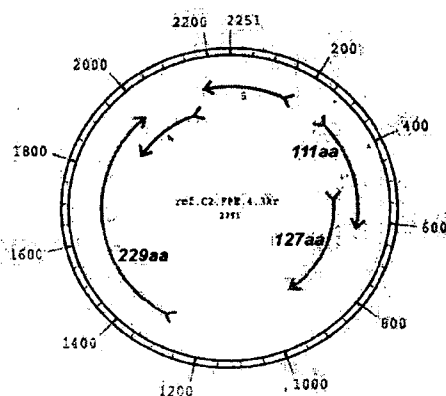
11. The method according to claim 10, wherein the at least one probe according to claim 3 is a labelled probe.

12. A method of detecting a polypeptide of claim 6 present in a biological sample, comprising: (a) contacting the biological sample with an antibody of claim 7, and (b) detecting the immunological complex formed between said antibody and said polypeptide.

13. A method of detecting an antibody of claim 7 present in a biological sample, comprising: (a) contacting the biological sample with a polypeptide of claim 6, and (b) detecting the immunological complex formed between said antibody and said polypeptide.

HCBI1.225**rmf.C2.F8E.4.3kr (2251)**

5 CCTTTTCCCT GATTACAAGC TCCCATCTGC TCCAGATCGG GAGCTTTTAG CCATCCCGAA TGGGTGGCGA CACAAATTAC
 AGGCGAAAAA AAAGACCATA TTGGCCTTCG GTTTAGATTG GCTGCCGTAT CAGCCAATTT GATTGAGTGC CGCATCACTC
 AATCTATCAA GTGCCGCATC ACCTGATTAC TCAAACTCTAC ATGAAATTTT CTTTCTTGCA AACGGTTTCTG AAAAGCATT
 TAAATATCCC ACCAATGAAT CATAAATGAA TCACTATAGA TCAGGAAACA ACTATGAAAC TCCATAATCC AAATCCAAAT
 10 GAGCCTACTA ACCTACAAAT GCTTGTGCA GAAGTTAAAA AATCCGCTTC AAGCTCTTAT CACGGTGGTT ATATTCAAGT
 TCCTTTCCGT GTTGAGTTTG CATCATATAC ACGCCTTGAG GCACCTTGTTA AACATACTGG CTCAAGTCGC AATAAGATCA
 TGAATGATCT GTTAAGAATC GGTATTGAAA CTCTAGCTGC CTCTTTGGAT GATGAAACAA TTAAACTCT TTTTGAAATT
 GAAACCTCAA TCACTGCGGA TCTCTATGCT TCAGGAAAAA TTAAATCAGG GGATCAATCA GATGATTAAT TTACAAGGAA
 CTCTAATTAA TGCTTTTCGT GTGGATGGTG GTAAAGGTAA AGACGGCAAA GAATATGAAG CGCGTGACAA GGTGCAAAAT
 CTTGGTTCGT TGGAGTTACC CAATGGGGAA ATTAACACAG AACTTGTGTA CCTAACAGTT GAGGATTCTC GACTTTTGA
 15 ACCATTCAAA AATCAGGTTA TTAGCATCTC ATGTGGTGCT ATGGCTGTCG GTCGTAATGT TATTTTTTAT GTTCGCAAAG
 GCGCAAAACC AGTTTTAGCA GAACAAATGT GAAACCCGAA AAAAAAACA GTGAGTTAGC GAAGCAACTC GATACTATTG
 ATTAAGTAGG GGACAAAAAT TGACTATAAA AGACAATAAA AAAGCCCATTA ATTCGGATGC TTTGGCGGGC GATGAATTAC
 AGGCTTTGCA ATCTGCTAAG GCAGATCAAC ACAGGGATAG AATAACACGT TTTGGAATTT TGAACATAG AGCGAAGCTA
 CAAGAACAAT ATTTGTGGAC GCAAGTTGAT TTTAAATCCG AAGGTGAAAA TGATCTATCT AATAAGGCTC TTAAGGCTGC
 20 AACCAAACTC AAGGGATGCG GTCAATTTCT CTTATTCCAT AATTACTACA CAATAGACCA AGTTAACTT GCTAAGGCCT
 ATTACTGTTC TCAGCATTTG CTATGTCTTA TGTGTGCTGG TGTAAGGGCT GCTAAGTCAA TGAGTCGTTA TATTCAGCGT
 TTTGAAGAAC TAATGCGCCA GAATCGCCAT TTAAAGCCCG TCCTGATCAC TTTGACGGTA AAGAATGGCC CTGACCTACA
 AGAACGCTTC AAACACCTTA GAGCATCATT TAGAACGCTT TTAGATCGTT ATAACGATTA CAAGAAAAA GGTCTGGTT
 TTAATCAGTT TTGTAATAAT GATGGTGCTT TTTATTCAAC TGAATATACA TACAATCCAA AAATAAAGA ATGGCATCCG
 25 CATATTCATA TTTTTCCTT ACTCAATCAA TGGATAGACC AGGAAGAATT GTCCGAACT TGGCAGGATA TTAATCTGGA
 TTCTTATATC GTCGATATTC GTAGAGTTAA AAAAACCAAA GAACACGGCT ATAGCAAAGC TGTGTCTGAG GTATGTAAAT
 ACGCCCTTAA GTTTAGTGAT CTCTCACTGG AGAACACTTG GGAGGCATAT CTTTCTTTGA AAGGTAACAG GCTTACTGGC
 TGTTTTGGTT CTATGTATGG TGTCAAATTG CCTGAAAAAC TTACTGATGA TCTACCCCTT GATGATCTTC CATATTGGA
 GCTTTTATAC CGTTTTGTTT TTGGTAAAAA ATCTTATTAC AACCTAGAAA TAACAAAAGA CGTAAAGCCA CAAAAGTAGG
 30 ACTACAACGA TGAGGTGAGG GCGACGCGCG ACCGTGCGCG CTACTTCGTG CATGTACGCG CGCGCCTTTG CCTCTTTGTT
 GCGAGGTGTG GACGAAAAA GCAAGGATAC ATACAGACCC CTGCATGACC TTGTAAGGGG TTCGACCCTT TAGACCCCAA
 AGGGCGCACT TATGCAAACT CTTGAGTTT GCCAGTGCT CCACAGTAA CAGAGGGCGC GGAGTGCGCC CGAACTGACG
 CTATAGAATT C



40

45 86% nucleotide similarity to Sphinx2.36 (2364 bp) in 845nt overlap

50 **Putative ORFs:**

229 amino acids 82% similarity to Sphinx2.36 (232 aa)
 127 amino acids 91% similarity to Sphinx2.36 (96 aa)
 111 amino acids 97% similarity to Sphinx2.36 (124 aa)

Replication protein
 Function unknown
 Function unknown

229 amino acids 82% similarity to Sphinx2.36 (232 aa) Replication protein

		1				50
	SPHINX2.36.232.PEP	MQEYLWTQV	DFKVGSETSI	KALKAATKLK	GCGQFLFRN	YTTIDQIKLE
5	HCBI1.225.229.pep	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
		51				100
	SPHINX2.36.232.PEP	KFHVCGQHLL	CPMCAGIRAA	RSMNRYIQRI	EEIMRQNRKL	KPVLITLTVK
	HCBI1.225.229.pep	~~~~~	~~MCAGVRAA	KSMSRYIQRI	EELMRQNRHL	KPVLITLTVK
		101				150
10	SPHINX2.36.232.PEP	NGEDLQERFE	HLTGSFKTLL	QRYRDFKKKG	RGFNQFCKID	GGFYTTEITY
	HCBI1.225.229.pep	NGFDLQERFA	HLRASFTLL	DRYNDYKKKG	RGFNQFCKID	GAIFYSTEITY
		151				200
	SPHINX2.36.232.PEP	NETTQOWHPH	IHIFALVTDR	IDQEELAETW	HDITLDSYIV	DIRRVKKTKE
	HCBI1.225.229.pep	NPKTKEWHPH	IHIFALLNQW	IDQEELSETW	HDITLDSYIV	DIRRVKKTKE
15		201				250
	SPHINX2.36.232.PEP	HGYAKAAEV	CKYALKFSDL	STEKTQFAFF	DP~~~~~	~~~~~
	HCBI1.225.229.pep	HGYSKAAEV	CKYALKFSDL	SLENTWEAYL	SLKGNRLTGC	FGSMYGVKLP
		251				291
	SPHINX2.36.232.PEP	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
20	HCBI1.225.229.pep	EKLTDLPLD	DLPYLELLYR	FVFGKKSYYN	LEITKDVKPQ	N

96 amino acids 91% similarity to Sphinx2.36 (96 aa) Function unknown

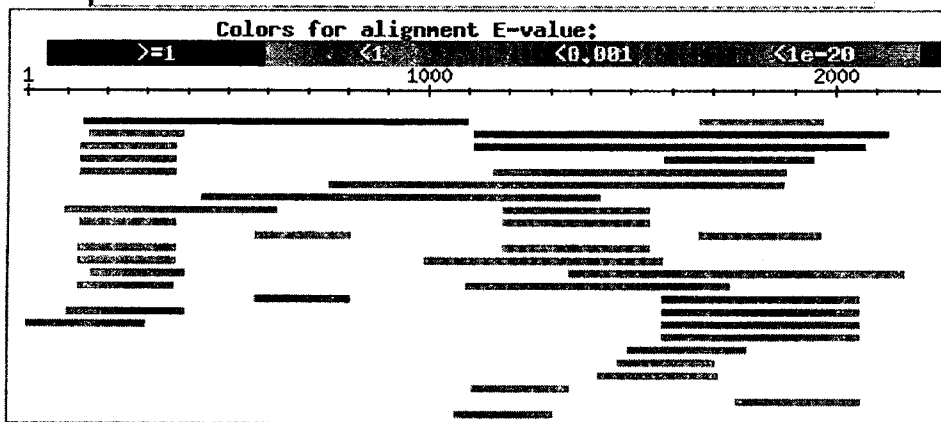
		1				50
	SPHINX2.36.96.PEP	~~~~~	~~~~~	~~~~~	~MINLQGTLI	NAFRLDGGKG
25	HCBI1.225.127.pep	MMKQLKFLK	LKPQSLRISM	LQEKLNQGIN	QMINLQGTLI	NAFRVDGGKG
		51				100
	SPHINX2.36.96.PEP	KDGKEYEARD	KVQILGSLEL	PNGEIKHELV	DLTVEDARIF	EPFKHKVISI
	HCBI1.225.127.pep	KDGKEYEARD	KVQILGSLEL	PNGEIKHELV	DLTVEDSRLF	EPFKNQVISI
		101		127		
30	SPHINX2.36.96.PEP	SCGAMAIGRN	VVFYVRKGAK	PVLADVM		
	HCBI1.225.127.pep	SCGAMAVGRN	VIFYVRKGAK	PVLAEQM		

111 amino acids 97% similarity to Sphinx2.36 (124 aa) Function unknown

		1				50
	SPHINX2.36.124.PEP	MNHKLIAIDQ	ELTMKLHNPN	PNEPTNLQML	VAEIKKSASS	SYHGGYIQVP
35	HCBI1.225.111.pep	~~~~~	~~MKLHNPN	PNEPTNLQML	VAEVKKSASS	SYHGGYIQVP
		51				100
	SPHINX2.36.124.PEP	FRVECASYTR	LEALVKHTGS	SRNKIMNDLL	RIGIETLAAS	LDDETIKTLF
40	HCBI1.225.111.pep	FRVEFASYTR	LEALVKHTGS	SRNKIMNDLL	RIGIETLAAS	LDDETIKTLF
		101		124		
	SPHINX2.36.124.PEP	EIETSITADL	YASGKMKSGD	QSDD		
	HCBI1.225.111.pep	EIETSITADL	YASGKTSGD	QSDD		

FastA *Geall

Mouse-over to show sequence definition. Click to show alignments



EMSTD UNC:HQ444405 Begin: 1414 End: 2256
! Hq444405 TSE-associated circular DN... 2999 3585 3088 2455.8 8.2e-131
EMSTD PRO:CU459139 Begin: 211 End: 1134
! Cu459139 Acinetobacter baumannii st... 1490 1972 1849 1467.2 8.2e-76
EMSTD PRO:JN872565 Begin: 223 End: 1111
! Jn872565 Acinetobacter baumannii st... 1692 2403 1742 1382.9 4.9e-71
EMEST PLN:DN449645 Begin: 1 End: 281

! Dn449645 EST945444 Sequencing ESTs ... 665 665 712 572.5 5.3e-25

EMSTD_UNC:HQ444405

ID HQ444405; SV 1; circular; genomic DNA; STD; UNC; 2364 BP.

AC HQ444405;

DE TSE-associated circular DNA isolate *Sphinx* 2.36, complete sequence.

SCORES z-score: 2455.8 E(): 8.2e-131

>>EMSTD_UNC:HQ444405

(2364 nt)

85.7% identity in 845 nt overlap

(175-1011:1414-2256)

150 160 170 180 190 200
rmf.C2.F8E.4 GAGTGGCGCATCACTCAATCTATCAAGTGGCGCATCACCTGATTACTCA-AATCTACATG
|||||

1390 1400 1410 1420 1430 1440
HQ444405 ATAATTTGCGGTCACTGTGCCGAATCACTACGCAATATTTATTACCCATAATTTAACCA

210 220 230 240 250 260
rmf.C2.F8E.4 AAATTTCTTTCTTGCAAACGGTTCAGAAAAGCATTATAAATATCCCAATGAATCAT
| |||

1450 1460 1470 1480 1490 1500
HQ444405 ATTCTTAATTTCTTGCAAACGGTTCAGAAAAGCATTATAAATATCTCACCATGAATCAT

270 280 290 300 310 320
rmf.C2.F8E.4 AAATGAATCACTATAGATCAGGAAACAACATGAACTCCATAATCCAAATCCAAATGAG
|||

1510 1520 1530 1540 1550 1560
HQ444405 AAATGAATCACTATAGATCAGGAAATTAACATGAACTCCATAATCCAAATCCAAATGAG

330 340 350 360 370 380
rmf.C2.F8E.4 CCTACTAACCTACAAATGCTTGTTCAGAAAGTAAAAAATCCGCTTCAAGCTCTTATCAC
|||||

1570 1580 1590 1600 1610 1620
HQ444405 CCTACTAACCTGCAATGCTTGTTCAGAAATTAAGAAATCCGCTTCAAGCTCTTATCAC

390 400 410 420 430 440
rmf.C2.F8E.4 GGTGGTTATATTCAAGTTCCTTTCCGTTGAGTTTGCATCATATACACGCCTTGAGGCA
|||||

1630 1640 1650 1660 1670 1680
HQ444405 GGTGGCTATATTCAAGTTCCTTTCCGTTGAGTTTGCATCATATACACGCCTTGAAGCA

450 460 470 480 490 500
rmf.C2.F8E.4 CTTGTTAAACATACTGGCTCAAGTCGCAATAAGATCATGAATGATCTGTTAAGAATCGGT
|||||

1690 1700 1710 1720 1730 1740
HQ444405 CTTGTTAAACATACTGGCTCAAGTCGCAATAAGATCATGAATGATCTCTTAAGAATCGGT

510 520 530 540 550 560
rmf.C2.F8E.4 ATTGAAATCTAGCTGCCTCTTTGGATGATGAAACAATTAAGTCTTTTGAATGAA
|||||

1750 1760 1770 1780 1790 1800
HQ444405 ATTGAAATCTAGCTGCCTCTTTGGACGACGAAACAATTAAGTCTTTTGAATGAA

570 580 590 600 610 620
rmf.C2.F8E.4 ACCTCAATCACTGCGGATCTCTATGCTTCAGGAAAAATTAATCAGGGGATCAATCAGAT
|||

1810 1820 1830 1840 1850 1860
HQ444405 ACTTCAATCACTGCGGATCTCTATGCTTCAGGAAAAATGAATCAGGAGATCAGTCAGAT

630 640 650 660 670 680
rmf.C2.F8E.4 GATTAATTTACAAGGAACCTAATTAATGCTTTTCGTGTGGATGGTGGTAAAGGTAAAGA
|||||

1870 1880 1890 1900 1910 1920
HQ444405 GATTAATTTACAAGGGACACTTATAAATGCTTTTCGTCTTGATGGTGGTAAAGGGAAAGA

690 700 710 720 730 740
rmf.C2.F8E.4 CGGCAAAGAATATGAAGCGCGTGACAAGGTGCAAATCTTGGTTCGTTGGAGTTACCCAA
|||||

1930 1940 1950 1960 1970 1980
HQ444405 CGGCAAAGAATATGAAGCACGTGACAAGGTGCAAATCTTGGTTCGCTGGAGTTGCCCAA

750 760 770 780 790 800
rmf.C2.F8E.4 TGGGGAAATTAACACGAACTTGTGACCTAACAGTTGAGGATTCTCGACTTTTGAACC
|||

1990 2000 2010 2020 2030 2040
HQ444405 CGGTGAGATCAACATGAGCTTGTGACTTAACGGTTGAAGATGCTCGCATCTTTGAGCC

810 820 830 840 850 860
rmf.C2.F8E.4 ATTCAAAAATCAGGTTATTAGCATCTCATGTGGTGCTATGGCTGTCGGTCGTAATGTTAT
|||||

CA 3065670 2019-12-19

rmf.C2.F8E.4 GGCATCCGCATATTCATATTTTTGCCTTACTCAATCAATGGATAGACCAGGAAGAATTGT
CU459139 GGCATCCTCATATTCATATGTTTGCACCTTTTGGATGATTGGATAGATCAGGATGAATTGT
5 670 680 690 700 710 720

rmf.C2.F8E.4 1660 1670 1680 1690 1700 1710
CU459139 CCGAAACTTGGCAGCATATTACTCTGGATTCTTATATCGTCGATATTCGTAGAGTTAAAA
10 CTCAATATTGGCAATCCATTACTGGGGACTCTATGGTCGTTGATATTCGTAGAGCCAAAA
730 740 750 760 770 780

rmf.C2.F8E.4 1720 1730 1740 1750 1760 1770
CU459139 AAACCAAGAACACGGCTATAGCAAAGCTGTTGCTGAGGTATGTAAATACGCCCTTAAGT
15 AACAAAAAGACTTAGGCTATTAGGTGCTGCTGCTGAAGTCTGTAAATATGCTCTCAAAT
790 800 810 820 830 840

rmf.C2.F8E.4 1780 1790 1800 1810 1820 1830
CU459139 TTAGTGATCTCTCACTGGAGAACACTTGGGAGGCATATCTTTCTTTGAAAGGTAACAGGC
20 TTGGTGATCTTTCTGTAGAAAAGACTTGGGAAGCTTTCAAAGTTTGAAGGTAAGCGAT
850 860 870 880 890 900

rmf.C2.F8E.4 1840 1850 1860 1870 1880 1890
CU459139 TTACTGGGTGTTTGGTCTATGTATGGTGTCAAATTGCCTGAAAACTTACTGATGATC
25 TAAGTGGGGCTTTTGGATCTCTTTGGGGCGTGAAAAATCCTGAATCATTGATAGATGATC
910 920 930 940 950 960

rmf.C2.F8E.4 1900 1910 1920 1930 1940
CU459139 TACC---CCTTGATGATCTTCCATATTTGGAGCTGTTATACCGTTTTGTTTTGGTAAAA
30 TTCCAGACGATTCTGATTACCTTATTTAGAAATGATTATAAGTTCGTCTTTCTAAGA
970 980 990 1000 1010 1020

rmf.C2.F8E.4 1950 1960 1970 1980 1990 2000
CU459139 AATCTTATTACAACCTAGAAATAACAAAAGACGTAAAGCCACAAAACCTAGGACTACAACG
40 AGTCTTATTACGATTTACAACCTACTCGTCATGTGCAACCTACAGGTAAGGACGACGCCG
1030 1040 1050 1060 1070 1080

rmf.C2.F8E.4 2010 2020 2030 2040 2050 2060
CU459139 ATGAGGTGAG-GGCGACGCGCGACCGTGCAGCTACTTCGTGCATGTACGCGCGCGCCTT
45 ACGAGCTTCGAGGAGAAGAAGGACGCAACCTGTTGGT--GAGCATGGACGGGCGAGGAGC
1090 1100 1110 1120 1130

rmf.C2.F8E.4.3kr.seq
EMSTD_PRO:JN872565
ID JN872565; SV 1; circular; genomic DNA; STD; PRO; 2252 BP.
AC JN872565;
DE *Acinetobacter baumannii* strain DS002 plasmid pTS236, complete sequence.
SCORES z-score: 1382.9 E(): 4.9e-71
>>EMSTD_PRO:JN872565 (2252 nt)
55 66.4% identity in 889 nt overlap
(1143-2030:223-1111)

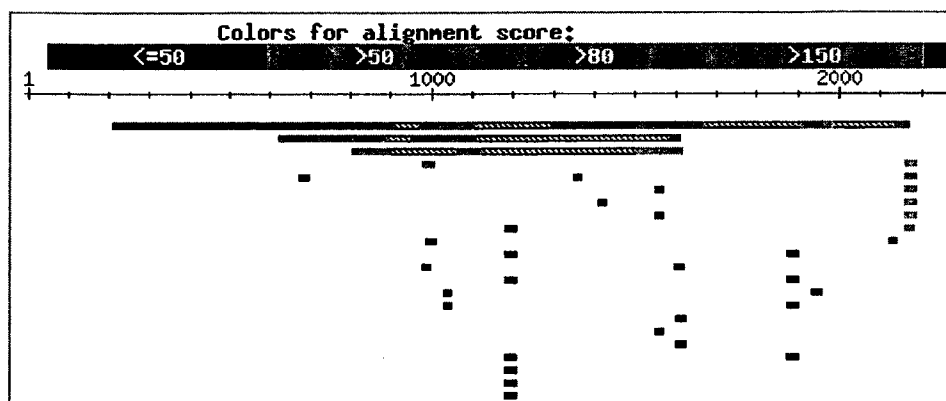
rmf.C2.F8E.4 1120 1130 1140 1150 1160 1170
JN872565 CGAAGCTACAAGAACAATATTTGTGGACGCAAGTTGATTTTAAATCCGAAGGTGAAAATG
60 AAAATTATTTATTTACCCTGGCTAAGTTTAAAGAGAATATGAAAAAGACGTTAAAAACG
200 210 220 230 240 250

rmf.C2.F8E.4 1180 1190 1200 1210 1220 1230
JN872565 ATCTATCTAATAAGGCTCTTAAGGCTGCAACCAAACTCAAGGGATGCGGTCAATTTCTCT
65 AAGAATCTATCAAGGCTCTAAAATCTGCTCAGAAATTGAATGAATGCGGAACTATCTGC
260 270 280 290 300 310

rmf.C2.F8E.4 1240 1250 1260 1270 1280 1290
JN872565 TATTCCATAATTACTACACAATAGACCAAGTTAAACTTGCTAAGGCCTATTACTGTTCTC
70 TATTCAAAAATTTTACACAATAGGCGAAGTTAAACTCTCAAGCTCCGCACCTGCGGAC
320 330 340 350 360 370

rmf.C2.F8E.4 1300 1310 1320 1330 1340 1350
JN872565 AGCATTGCTATGTCTATGTGTGCTGGTGAAGGGCTGCTAAGTCAATGAGTCGTTATA
75 AGCATTGCTTTGCCCTTTCTGTGCTGCCATTCTGCTGCTATCCAAAAATACG
380 390 400 410 420 430

BlastN2:



>>>nrnuc:GI_313507479 Gi|313507479|gb|HQ444405.1| TSE-associated... 926 0.0
>>>nrnuc:GI_377552564 Gi|377552564|gb|JN872565.1| Acinetobacter ... 192 1e-44
>>>nrnuc:GI_169147044 Gi|169147044|emb|CU459139.1| Acinetobacter... 107 5e-19

>>>>nrnuc:GI_313507479 Gi|313507479|gb|HQ444405.1| TSE-associated circular DNA isolate
Sphinx 2.36, complete sequence. 0/0
Length = 2364

Identities = 635/691 (91%)

Strand = Plus / Plus

Query: 212 tttcttgcaaacggttcagaaaaagcattataaaatcccaccaatgaatcataaatgaat 271

Sbjct: 1452 tttcttgcaaacggttcagaaaaagcattataaaatctcaccaatgaatcataaactaat 1511

Query: 272 cactatagatcaggaaacaactatgaaactccataatccaaatccaaatgagcctactaa 331

Sbjct: 1512 cgctatagatcaggaaattaactatgaaactccataatccaaatccaaatgagcctactaa 1571

Query: 332 cctacaatgcttgttgagaaagttaaaaaatccgcttcaagctcttatcacggtgggta 391

Sbjct: 1572 cctgcaatgcttgttgagaaattaaaaaatccgcttcaagctcttatcacggtgggta 1631

Query: 392 tattcaagttcctttccgtgttgagttgcatcatatacacgccttgaggcacttggttaa 451

Sbjct: 1632 tattcaagttcctttccgtgttgagttgcatcatatacacgccttgaggcacttggttaa 1691

Query: 452 acatactggctcaagtcgcaataagatcatgaatgatctgttaagaatcggtattgaaac 511

Sbjct: 1692 acatactggctcaagtcgcaataagatcatgaatgatctttaaagaatcggtattgaaac 1751

Query: 512 tctagctgcctctttggatgatgaaacaattaaaactcttttgaattgaaacctcaat 571

Sbjct: 1752 tctagctgcctctttggacgacgaaacaattaaaactcttttgaattgaaacctcaat 1811

Query: 572 cactgaggatctctatgcttcaggaaaaattaaatcaggggatcaatcagatgattaatt 631

Sbjct: 1812 cactgaggatctctatgcttcaggaaaaattgaaatcaggagatcagtcagatgattaatt 1871

Query: 632 tacaaggaactctaattaatgcttttcgtgttgatgggtggttaaaggtaaagacggcaaag 691

Sbjct: 1872 tacaagggacacttataaatgcttttcgtgttgatgggtggttaaaggtaaagacggcaaag 1931

Query: 692 aatatgaagcgcgtagaagggtgcaaatcttgggttcgttgagttacccaatggggaaa 751

Sbjct: 1932 aatatgaagcacgtgacaagggtgcaaatcttgggttcgttgagttgccaacgggtgaga 1991

Query: 752 ttaaacacgaacttgttgacctaacagttgaggattctcgacttttgaaccattcaaaa 811

Sbjct: 1992 tcaaacatgagcttgttgacttaacggttgaagatgctcgcatctttgagccattcaaac 2051

Query: 812 atcagggtattagcatctcatgttggtgctatgggtgctcggtcgtaagtatttttatg 871

Sbjct: 2052 acaaggttaattagcatttcatgcggtgctatgggtgctcggtcgtaagtatttttatg 2111

Query: 872 ttcgcaaaggcgcaaaaccagtttttagcaga 902

Sbjct: 2112 ttcgaaaagggtgcaaaagcctgttttagcaga 2142

Score = 214 bits (108), Expect = 3e-51

Identities = 126/132 (95%)

Strand = Plus / Plus

Query: 976 aaaattgactataaaaagacaataaaaaagcccataattcggtatgctttggcgggcatga 1035

Sbjct: 2233 aaaattgactaaaaagacaataaaaaagcccataattctgatgctttggcggcgacga 2292

Query: 1036 attacaggctttgcaatctgctaaggcagatcaacacagggatagaataacacggttttg 1095
|||||

5 Sbjct: 2293 attacaggctttgcaatctgctaaggcagatcaacacagggatagaataatcacggttttg 2352

Query: 1096 aattttgaaaca 1107
| |||||

10 Sbjct: 2353 acttttgaaaca 2364

Score = 143 bits (72), Expect = 9e-30
Identities = 303/380 (79%)
Strand = Plus / Plus

15 Query: 1292 cagcatttgctatgtcctatgtgtgctggtgtaagggtgctaagtcaatgagtcgttat 1351
|||||

Sbjct: 180 cagcatttgctatgtccaatgtgtgctggtattcgtgctgccgttcaatgaatcggtat 239

Query: 1352 attcagcgtattgaagaactaatgctgcagaatcgccatttaaagcccgtcctgatcact 1411
|||||

20 Sbjct: 240 attcaacgcatcgaagaaataatgctgcagaatcgcaagctaaagcccgtattgatcact 299

Query: 1412 ttgacggtaagaatggccctgacctacaagaacgcttcaaacaccttagagcatcattt 1471
|||||

25 Sbjct: 300 ttgacggtaagaacgggtgaagacctacaggaacgctttgaacacctcacaggctcattt 359

Query: 1472 agaacgcttttagatcggtataacgattacaagaaaaaggctcggttttaacagttt 1531
| |||||

Sbjct: 360 aagacgcttttacagcggtaccgtgattttaagaaaaagggtcgagggttttaacattt 419

30 Query: 1532 tgtaaaattgatggtgctttttattcaactgaatatacatatacctcaaaaaactaaagaa 1591
|| |||||

Sbjct: 420 tgcaaaattgatggtgctttttatacagaccgaatacacctacaacgaacaacccaacaa 479

35 Query: 1592 tggcatccgcataattcatatttttgccttactcaatgaatggatagaccaggaagaattg 1651
|||||

Sbjct: 480 tggcatccgcataattcatatttttgcgttagtgactgaccggttgaccaggaggaacta 539

Query: 1652 tccgaaacttggcagcatat 1671
| |||||

40 Sbjct: 540 gcagaaacttggcagcatat 559

Score = 69.9 bits (35), Expect = 1e-07
Identities = 41/43 (95%)
Strand = Plus / Plus

45 Query: 2130 cttgtaaggggttcgaccccttagaccccaaggggcgactta 2172
|||||

Sbjct: 1063 cttgtaaggggttcgaccccttagaccccaaggggcgactta 1105

Score = 61.9 bits (31), Expect = 3e-05
Identities = 70/83 (84%)
Strand = Plus / Plus

50 Query: 1907 cttccatatttgagctgtttataccgtttttggttaaaaaatcttattacaaccta 1966
|||||

55 Sbjct: 796 cttccatatttgagctgtctatcgcttttgggtgaaagggtcttattacaaccta 855

Query: 1967 gaaataacaaaagacgtaaagcc 1989
|| |||||

60 Sbjct: 856 gagttaactaaggacgtaaagcc 878

>>>>nrnuc:GI_377552564 Gi|377552564|gb|JN872565.1| Acinetobacter baumannii strain DS002
plasmid pTS236, complete sequence. 0/0
Length = 2252

65 Identities = 226/269 (84%)
Strand = Plus / Plus

Query: 622 atgattaatttacaaggaactctaattaatgcttttctggtggatgggtgtaagggtaaa 681
|||||

70 Sbjct: 1925 atgattaatttacaaggaactctaattaatgcttttccgtatggacggtgtaagggtaaa 1984

Query: 682 gacggcaagaatataagcgcgtgacaagggtgcaaatcttgggttcgttgaggttacc 741
|| |||||

75 Sbjct: 1985 gatgggaaagagtacgaagcgcgtgacaagggtacaaatacttgggttcgttggaactacc 2044

Query: 742 aatggggaaattaacacgaacttgttgacctaacagttgaggattctcgactttttgaa 801
|||||

Sbjct: 2045 aatggagagattaacacgaacttgttgacctcacagttgatgacaggtgtctaccag 2104

Query: 802 ccattcaaaaatcaggttattagcatctcatgtggtgctatggctgtcggtcgtaattgt 861
||| |||||

Sbjct: 2105 ccactaaaaataaagtaatttctatttctcggtgctatggctgtaggtcgtaatgtt 2164

Query: 862 attttttatgttcgcaaaggcgcaaaacc 890
|||||

5 Sbjct: 2165 gttttttatgttcgaaaaggtgcaaaacc 2193

Score = 182 bits (92), Expect = 1e-41
Identities = 143/160 (89%)
Strand = Plus / Plus

10 Query: 951 gatactattgattaactaggggacaaaaattgactataaaagacaataaaaaagcccata 1010
|||||

Sbjct: 19 gatactattgattaagtgaggacaaaaattgcttaaaaaagacaataaaaaagcccata 78

15 Query: 1011 attcggatgcttttggcggcgcatgaattacaggcctttgcaatctgctaaggcagatcaac 1070
||| |||

Sbjct: 79 ttccagaagcttttggcggcggagaaatacaggcctttgcaatctgcaaacgcagatcaac 138

20 Query: 1071 acagggatagaataaacacgttttggaattttgaaacatag 1110
|||||

Sbjct: 139 acagggatagaataaacacgttttgcgactttgaaacatag 178

Score = 46.1 bits (23), Expect = 1.6
Identities = 23/23 (100%)
Strand = Plus / Plus

25 Query: 1589 gaatggcatccgcattcatat 1611
|||||

Sbjct: 669 gaatggcatccgcattcatat 691

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

35 Score = 107 bits (54), Expect = 5e-19
Identities = 87/98 (88%)
Strand = Plus / Plus

Query: 805 ttcaaaaatcaggttattagcatctcatgttggtgctatggctgtcggtcgtaatgttatt 864
|||||

40 Sbjct: 2570 ttcaaaaataagcttattagcatcagttgttggtgctatggctgttggtcgtaacgttatt 2629

Query: 865 ttttatgttcgcaaaggcgcaaaaccagtttttagcaga 902
|||||

Sbjct: 2630 ttttatgttcgaaaaggtgcgaacactgttttagcaga 2667

45 Score = 95.6 bits (48), Expect = 2e-15
Identities = 60/64 (93%)
Strand = Plus / Plus

Query: 1063 agatcaacacagggatagaataaacacgttttggaattttgaaacatagagcgaagctaca 1122
|||||

50 Sbjct: 119 agatcaacatagggatagaataaacacgttttggcattttgaaacatagatcgaagcaaca 178

Query: 1123 agaa 1126
||||

55 Sbjct: 179 agaa 182

Score = 73.8 bits (37), Expect = 7e-09
Identities = 94/113 (83%)
Strand = Plus / Plus

60 Query: 1505 aaaaaaggtcggtgttttaaatcagttttgtaaaattgatgggtgctttttattcaactgaa 1564
|||||

Sbjct: 573 aaaaaagggcggtgctttaatgaattttgcaaaattaatgggtgcgatgtattcatatgag 632

65 Query: 1565 tatacatacaatccaaaaactaaagaatggcatccgcattcatatttttgc 1617
|||||

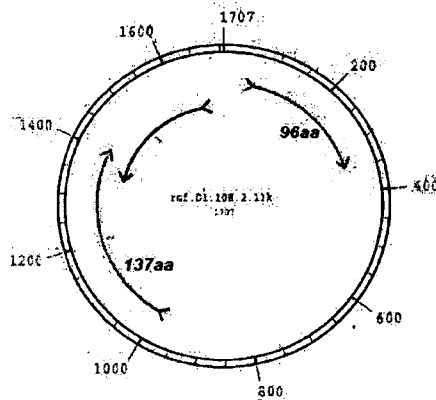
Sbjct: 633 aatacttacaatgaaaaaactaatgaatggcatcctcatattcatatgtttgc 685

70

Figure 1

HCBI2.170 (rmf.D1.10E.2.11k) (1707bp)

GAATTCATAA TGTTCGCTTC TAACCATTAC AACGACTTTA CTGGTTCTGG AGATTATATCA GATGATTAAA TTAGAAGGAA
 TCGTTTTAAA CGTCTTCACT CAGCAAGGTG GACAAAACAA AAAAGGCGAA TCATTGTATG ATCGTGACAA GGTACAAATT
 5 TTAGTGTGTA TGGATCTGCC CAATGGTGAT GTAAAAATG AGCTTTTAC GTTATCTGTA GATGATTATC GGGACTTTAA
 AGACTTCCTA AATCGAAAAA TTTGTATTGC TGTTGGTGCA ATGGCAAGTG GCCGTAATGT TATTTTTAT GTTGCTAAAG
 GTGCAAAAGCC TATATCAGCA GAATTTGCGT GAAACGCTAA AAAAAACAGT GAGTTAGCCT TGGCTTCTCG CTGTTTTTTC
 TTTCTCTTGA TACTATTGAT TAACTAGGGG ACAAAAATTG ACTAAAAAAG ACAATAAAAA AGCCCATAAAT TCGGACACTT
 TGGCGGGTGA TGAATTACAG GCTTTGCAAT CTGCTAAGGC AGATCAACAC AGGGATAGAA TATCACGTTT TGGACTTTTG
 10 AAACATAGAG CGAAGCTACA AGAACAATAT TTGTGGACGC AGGTTGATTT CAAATCCGAA GGTGAAAATG AGACATCCAA
 TAAGGCTCTT AAGGCTGCAA CCAAATTAAA GGGATGCGGT CAATTTTTCG TATTCCATAA CTACTACACA ATTGACCAAG
 TTAAGCTTGC TAAGGCCCAT TATTGTTCTC AGCATTGTGCT TTGCCCTATG TGTGCTGGTG TAAGGGCTGC TAAGTCAATG
 AGTAGATATG TTCAACGTAT TGAAGAATTG ATGCGTCAGA ATCGCAAATT AAAGCCCGTA TTGATCACTT TAACGGTTAA
 GAATGGGGAA GACCTAGAAG AACGCTTTAA ACACCTTAGAC GCTCATTTAG GACGCTTTTA GATCGTTATA ACGATTACAA
 AAAGAAAGGT CGTGGTTTTA ATCAATTCTG CAAGATTGAT GGTGCTTTTT TATTCCACTG AATATACCTA CAATTCAAAA
 15 ACAAAGAGT GGCATCCCCA TATCCATATT TTCGCTTTAC TCAATGAATG GATAGACCAG GAAGAATTGG CCGAGACCTG
 GCATGACATT ACCCTGGATT CTTATATCGT AGATATTCTG AGAGTTAAAA GGACCAAAGA ACACGGCTAT AGCAAAGCTG
 TTGCAGAGGT TTGTAAATAT GCTCTTAAGT TTAGTGATTT GTCACTTGAG AATACGTGGG AAGCTTATCT TTCTTTAAAA
 GGTAAATAGC TTAAGTGGTG TTTTGGTTCT ATGTATGGTG TCAAGTTGCC TGAAAACTC ACAGATGATT TACCCCTTGA
 TGATCTTCCA TATATGGAGC TGCTATACCG TTTTGTCTTT GGTAAAAAAT AACATTACGG CCACTTGCCA TTGCACCAAC
 20 AGCAATACAA ATTTTTCGAT TTAGGAAGTC TTTAAAGTCC CGATAATCAT CTACAGATAA CGTAAAAAGC TCATTTTTTA
 CATCACCATT GGGCAGATCC ATAGCACCTA AAATTTGTAC CTGTGCACGA TCATCAAATG ATTGCGCTTT TTTGTTTTGT
 CCACCTTGCT GAGTGAAGAC GTTTAAACG ATTCCTCTTA ATTTAATCAT CTGATAAATC TCCAGAACCA GTAAAGTCGT
 TGTAAATGTT AGAAGCGAAC ATATGAA



74% nucleotide similarity to Sphinx2.36 (2364bp)

ORFs:

137 amino acids 81% similarity to Sphinx2.36 (232aa) replication protein
 96 amino acids 53% similarity to Sphinx2.36 (96aa) function unknown

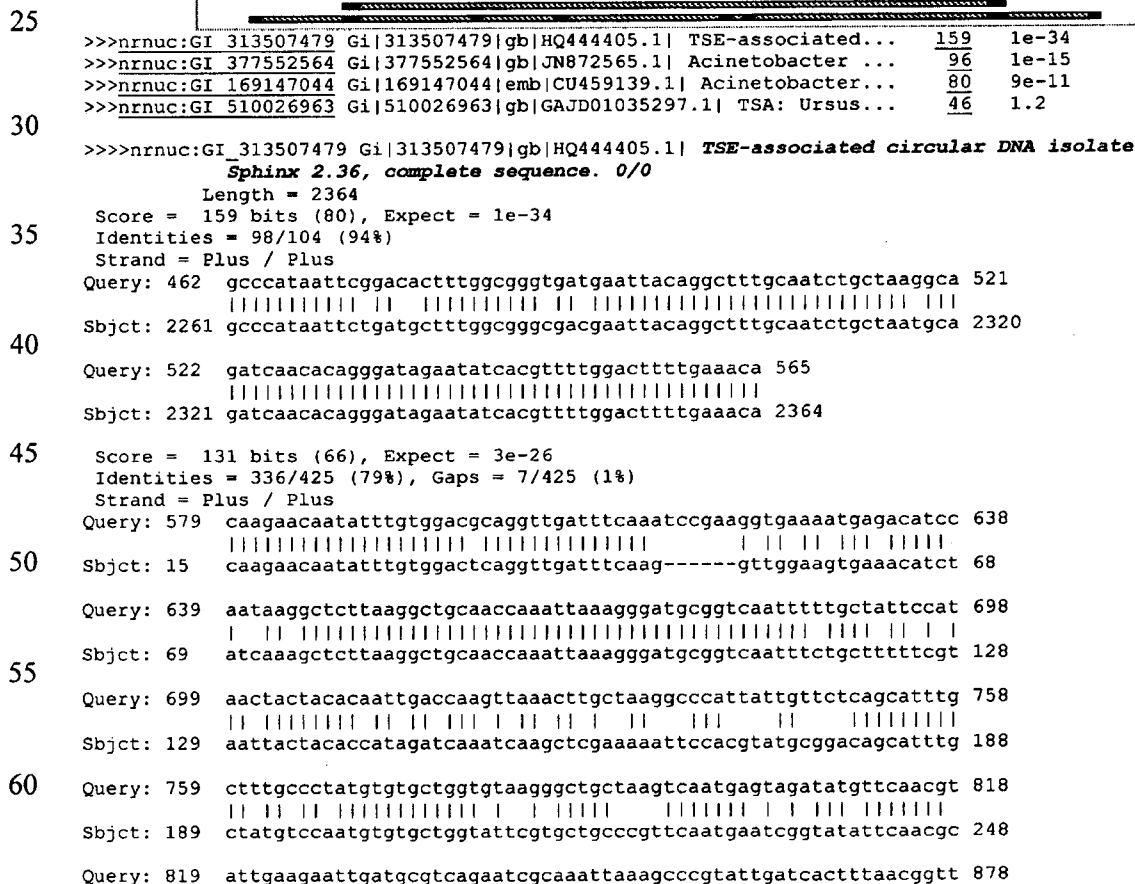
137 amino acids 81% similarity to Sphinx2.36 (232aa) replication protein

SPHINX2.36.96.PEP	1	MINLQGTLIN	AFRLDGGK GK	DGKEYEARDK	VQILGSLELP	NGEIKHELVD	50
HCBI2.170.137.pep		MIKLEGI VLN	VTTQGGQNK	KGESFDDRDK	VQILGAMDLP	NGDVKNELFT	
SPHINX2.36.96.PEP	51	LTVEDARIFE	PFKHKVISIS	CGAMAIGRNV	VFYVRKGAKP	VLADVM	96
HCBI2.170.137.pep		LSVDDYRDFK	DFLNRRICIA	VGAMASGRNV	IFYVAKGAKP	ISAEFA	

96 amino acids 53% similarity to Sphinx2.36 (96 aa) function unknown

SPHINX2.36.232.PEP	1	MQEQYLWTQV	DFKVGSETSI	KALKAAATKLK	GCGQFLLFRN	YYTIDQIKLE	50
HCBI2.170.96.pep		~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	

BlastN2:



Sbjct: 249 atcgaagaaataatgcgtcagaatcgcaagctaaagcccgctattgatcactttgaccggtt 308

5 Query: 879 aagaatggggaagacctagaagaacgctttaaaca-cttagacgctcatttaggacgctt 937
Sbjct: 309 aagaacggtgaagacctacaggaacgctttgaacacctcacaggctcatttaagacgctt 368

10 Query: 938 ttagatcggtataacgattacaaaaagaaaggctcggtgttttaataatcaattctgcaagatt 997
Sbjct: 369 ttacagcggtaccgtgattttaagaaaaagggtcgagggtttaataatcaattttgcaaaatt 428

15 Query: 998 gatgg 1002
Sbjct: 429 gatgg 433

20 Score = 54.0 bits (27), Expect = 0.005
Identities = 45/51 (88%)
Strand = Plus / Plus
Query: 367 cagtgagttagccttggtctctcgctgtttttctttctcttgatactatt 417
Sbjct: 2167 cagtgagttaggttgcgcgactcgctgtttttcttttacttgatactatt 2217

25 >>>>nrnuc:GI_377552564 Gi|377552564|gb|JN872565.1| Acinetobacter baumannii strain DS002
plasmid pTS236, complete sequence. 0/0
Length = 2252
Score = 95.6 bits (48), Expect = 1e-15
Identities = 83/92 (90%), Gaps = 2/92 (2%)
Strand = Plus / Plus

30 Query: 478 ctttggcgggtgatgaattacaggctttgcaatctgctaaggcagatcaacacagggata 537
Sbjct: 88 ctttggcggcgagaaatacaggctttgcaatctgcaaacgcagatcaacacagggata 147

35 Query: 538 gaatatcacgttttg-gacttttgaaacatag 568
Sbjct: 148 gaataacacgttttgcgac-tttgaaacatag 178

40 Score = 50.1 bits (25), Expect = 0.078
Identities = 31/33 (93%)
Strand = Plus / Plus
Query: 391 ctgttttttctttctcttgatactattgattaa 423
Sbjct: 1 ctgttttttcttttacttgatactattgattaa 33

45 >>>>nrnuc:GI_169147044 Gi|169147044|emb|CU459139.1| Acinetobacter baumannii str. AYE
plasmid p4ABAYE, complete genome. 0/0
Length = 2726
Score = 79.8 bits (40), Expect = 9e-11
Identities = 58/64 (90%)
Strand = Plus / Plus

50 Query: 521 agatcaacacagggatagaatatcacgttttggacttttgaaacatagagcgaagctaca 580
Sbjct: 119 agatcaacatagggatagaataacacgttttggcattttgaaacatagatcgaagcaaca 178

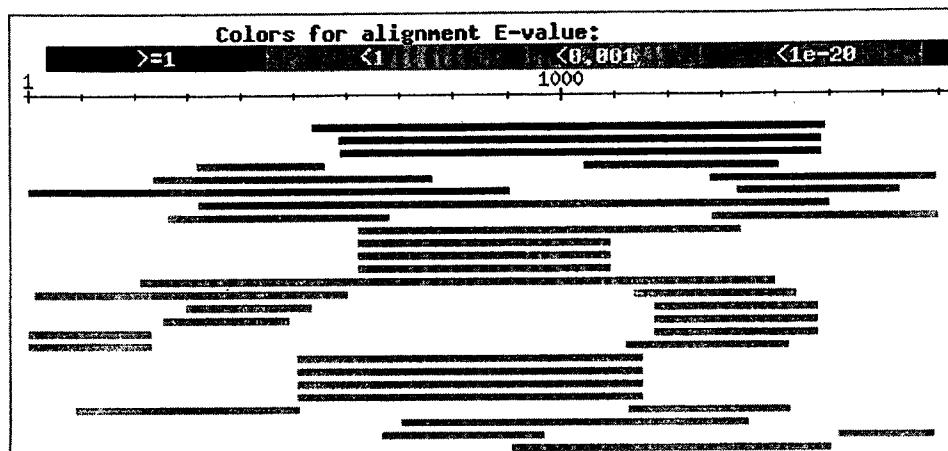
55 Query: 581 agaa 584
Sbjct: 179 agaa 182

60 Score = 50.1 bits (25), Expect = 0.078
Identities = 43/49 (87%)
Strand = Plus / Plus
Query: 849 ttaaagcccgctattgatcactttaacggttaagaatggggaagacctag 897

65 Sbjct: 459 ttaaagcccgcttatgatcacgtttacggttaaaatggggtagacctag 507

FastA *Geall

70



EMSTD UNC:HQ444405 Begin: 1 End: 885
! Hq444405 TSE-associated circular DN... 1438 1868 2309 1706.3 5.3e-89
EMSTD PRO:CU459139 Begin: 226 End: 1032
! Cu459139 Acinetobacter baumannii st... 1194 1691 1655 1222.6 4e-62
EMSTD PRO:JN872565 Begin: 240 End: 1072
! Jn872565 Acinetobacter baumannii st... 1234 1990 1624 1200.2 8.6e-61
EMSTD UNC:HQ444405 Begin: 1846 End: 2175 Strand: -
! Hq444405 TSE-associated circular DN... 513 580 520 384.2 2.3e-15

EMSTD UNC:HQ444405
ID HQ444405; SV 1; circular; genomic DNA; STD; UNC; 2364 BP.
AC HQ444405;
DE TSE-associated circular DNA isolate Sphinx 2.36, complete sequence.
SCORES E(): 5.3e-89
>>EMSTD UNC:HQ444405 (2364 nt)
74.1% identity in 892 nt overlap
(566-1451:1-885)

	540	550	560	570	580	590
HCBI2.170	TAGAATATCACGTTTGGACTTTTGAAACATAGAGCGAAGC-TACAAGAACAATATTGT					
HQ444405	TAGATCTAAACATGCAAGAACAATATTGT					
				10	20	30
	600	610	620	630	640	650
HCBI2.170	GGAGCGAGGTTGATTTCAAATCCGAGGTGAAATGAGACATCCAATAAGGCTCTTAAGG					
HQ444405	GGACTCAGGTTGATTTC-----AGGTTGGAAGTGAAACATCTATCAAAGCTCTTAAGG					
	40	50	60	70	80	
	660	670	680	690	700	710
HCBI2.170	CTGCAACCAAATTAAGGGATGCGGTCAATTTTGCTATTCCATAACTACTACCAATTG					
HQ444405	CTGCAACCAAATTAAGGGATGCGGTCAATTTCTGCTTTTCGTAATTACTACCATAG					
	90	100	110	120	130	140
	720	730	740	750	760	770
HCBI2.170	ACCAAGTTAAACTTGCTAAGGCCCATTTATTGTTCTCAGCATTGCTTGGCCCTATGTGTG					
HQ444405	ATCAAATCAAGCTCGAAAAATTCACGTATGCGGACAGCATTGCTATGTCCAATGTGTG					
	150	160	170	180	190	200
	780	790	800	810	820	830
HCBI2.170	CTGGTGTAAGGCTGCTAAGTCAATGAGTAGATATGTTCAACGTATTGAAGAATTGATGC					
HQ444405	CTGGTATTCGTGCTGCCCGTTCAATGAATCGGTATATTCAACGCATCGAAGAAATAATGC					
	210	220	230	240	250	260
	840	850	860	870	880	890
HCBI2.170	GTCAGAATCGCAAATTAAGCCCGTATTGATCACTTTAACGGTTAAGAATGGGGAAGACC					
HQ444405	GTCAGAATCGCAAGCTAAAGCCCGTATTGATCACTTTGACCGTTAAGAACGGTGAAGACC					
	270	280	290	300	310	320
	900	910	920	930	940	950
HCBI2.170	TAGAAGACGCTTTAAACA-CTTAGACGCTCATTTAGGACGCTTTTAGATCGTTATAACG					
HQ444405	TACAGGAACGCTTTGAACACCTCACAGGCTCATTTAAGACGCTTTTACAGCGTTACCGTG					

```
EMSTD_PRO:CU459139
ID   CU459139; SV 1; circular; genomic DNA; STD; PRO; 2726 BP.
AC   CU459139;
DE   Acinetobacter baumannii str. AYE plasmid p4ABAYE, complete genome. . .
SCORES   z-score: 1222.6 E(): 4e-62
>>EMSTD_PRO:CU459139                               (2726 nt)
    67.7% identity in 808 nt overlap
    (616-1419:226-1032)
```

EMSTD_PRO:JN872565
ID JN872565; SV 1; circular; genomic DNA; STD; PRO; 2252 BP.
AC JN872565;
DE *Acinetobacter baumannii* strain DS002 plasmid pTS236, complete sequence.

SCORES z-score: 1200.2 E(): 8.6e-61

>>EMSTD_PRO:JN872565

(2252 nt)

67.0% identity in 836 nt overlap

(618-1447:240-1072)

EMSTD_UNC:HQ444405

ID HQ444405; SV 1; circular; genomic DNA; STD; UNC; 2364 BP.

AC HQ444405;

DT 10-DEC-2010 (Rel. 107, Created)

DT 10-DEC-2010 (Rel. 107, Last updated, Version 1)

DE TSE-associated circular DNA isolate *Sphinx* 2.36, complete sequence.

SCORES z-score: 384.2 E(): 2.3e-15

>>EMSTD_UNC:HQ444405

(2364 nt)

62.7% identity in 332 nt overlap

(1666-1336:1846-2175)

1690 1680 1670 1660 1650 1640
HCB12.170 GCTTCTAACCATACACGACTTTACTGGTCTGGAGATTTATCAGATGATTAAATTAGA
HQ444405 GCGGATCTCTATGCTTCAGGAAAAATGAAATCAGGAGATCAGTCAGATGATTAAATTACA
1820 1830 1840 1850 1860 1870

1630 1620 1610 1600 1590 1580
HCB12.170 AGGAATCGTTTTAAACGCTCTTCACTCAGCAAGGTGGACAAAAAAGGCGAATCATT
HQ444405 AGGGACACTTATAAATGCTTTTCGCTTGTATGGTGGTAAAGGAAAGACGGCAAAGAATA
1880 1890 1900 1910 1920 1930

1570 1560 1550 1540 1530 1520
HCB12.170 TGATGATCGTGACAAGGTACAAATTTAGGTGCTATGGATCTGCCCAATGGTGATGTAAA
HQ444405 TGAAGCACGTGACAAGGTGCAAAATCTTGGTTCGCTGGAGTGGCCCAACGGTGAGATCAA
1940 1950 1960 1970 1980 1990

1510 1500 1490 1480 1470 1460
HCB12.170 AAATGAGCTTTTACGTTATCTGTAGATGATTATCGGGACTTTAAAGACTTCCTAAATCG
HQ444405 ACATGAGCTTGTGACTTAACGGTTGAAGATGCTCGCATCTTTGAGCCATTCAAACACAA
2000 2010 2020 2030 2040 2050

1450 1440 1430 1420 1410 1400
rmf.D1.10E.2 AAAAATTTGTATTGCTGTTGGTGCAATGGCAAGTGGCCGTAATGTTATTTTACCAAAG
HQ444405 GGTAAATTAGCATTTTCATGCGGTGCTATGGCTATCGGTCGAAATGTTGTTTTTATGTTTCG
2060 2070 2080 2090 2100 2110

1390 1380 1370 1360 1350 1340
HCB12.170 ACAAACGGTATAGCAGCTCCA-TATATGGAAGATCATCAAGGGGTAATCATCTGTGAG
HQ444405 AAAAGGTGCAAAGCCTGTTTGTAGCAGATGTAATGTAATTCA--TGAAAAAACAGTGAG
2120 2130 2140 2150 2160 2170

1330 1320 1310 1300 1290 1280
HCB12.170 TTTTTCAGGCAACTTGACACCATACATAGAACCAAAACAGCCAGTAAGCCTATTACCTTT
HQ444405 TTAGGCTTGCCGACTCGCTGTTTTTTCTTTTACTTGATACTATTAACATAAGTGGGGACA
2180 2190 2200 2210 2220 2230

EMSTD_PRO:CU459139

ID CU459139; SV 1; circular; genomic DNA; STD; PRO; 2726 BP.

AC CU459139;

PR Project:PRJNA28921;

DT 26-FEB-2008 (Rel. 94, Created)

DT 05-NOV-2010 (Rel. 106, Last updated, Version 2)

DE *Acinetobacter baumannii* str. AYE plasmid p4ABAYE, complete genome. . . .

SCORES z-score: 339.5 E(): 6.2e-13

>>EMSTD_PRO:CU459139

(2726 nt)

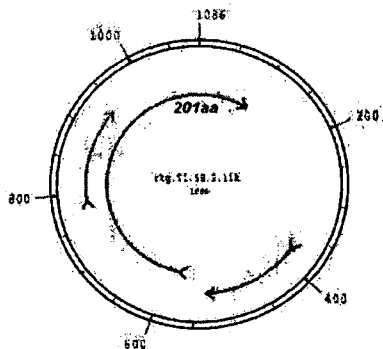
66.9% identity in 257 nt overlap

(1656-1403:2380-2634)

Figure 2

HCBI3.108 (rkg.T1.5E.2.11k) (1082bp)

5 GCGAGTGTCT ACGAGCGAAG TTATGAAAGT TCGATTCTTC CCCCTCTGA AAAACCGCTT TAAAAATAT TGGCTGCTAG
 ATGTTTTTA CTATCGTGAG CTTTGTCTT TAAAAACA CGAGCGAAGC GAGTTCATAG TTGCTTTTGG GGGTTTCGGG
 10 GGGCTTGCCC CCTGAACAAG ATCACGGAGT GGGAAATTTAT CACGGTAGTG AAAAAGTACC CTCTGTGTAT CCTTGCTTAT
 TTCTTTTTAA ACCTTTGAAC TTTTCCCGT AATTTGAAGA AATTGCCCT CGACTAAGCT TGCTTAGTCA AAAAAGTTTG
 AGCAAAGCGA AAACATAGGG CAATTTTCAT GATGAAAATG GGCTTTTAAG GCTTTTAAAT GCTTTTAAGG CTTTATAGACA
 TGCTGAAACG CAAGCCTAGC AAGGCATACA GAGGGCATT AACACCGTTT ACCTACCAAT ACCCCACCGT TTACCTACCA
 15 ATACCCACC GTTTACCTAC CAATACCCCA CCGTTTACCT ACCTTAATAC ACAATAATAT TTTTATGTGG TATAACGTAA
 TAAATATAT AGGTGGTTTA TGAGTGATTT AATAGTAAAA GATAACGCC TAATGAACGC TAGTTATAAC TTAGACTTGG
 TTGAACAGCG GTTAATCTT TTAGCTATCC TTGAAGCTAG AGAATCAGGC AAAGGAATTA ACGCAAATGA CCCTCTTACA
 GTCCATGCAG AGAGTTATAT CAATCAATTT GGTGTAGCTA GACAGACTGC TTATCAAGCC CTAAAGATG CCTGCAAGA
 TTTATTTGCC CGTCAATTCA GCTATCAAGA AAAGCGTGAA CGTGGACGAG CTAATATTAC AAGTCGTTGG GTCAGCCAAA
 20 TTGCTTACAT TGATGAACT GCAACGGTTG AGGTATTTT CGCCCTGCG GTTGTCCAC TGATCACAAG GTTAGAGGAA
 CAATTCTCGA AGTAGATAT TGAACAAAT AGTAGTCTAT CGAGTGCCTA TGCAGTTCGC TTATATGAGT TATTGATCTG
 CTGAGAACA ACAGGAAAGA CACCAGTTAT TGACTTAACA GAATTC

**ORF 201 amino acids – putative replication protein, DNA-binding domain**

20 Similarity to Spinx1.76 (324aa) = 88,2% in 169aa

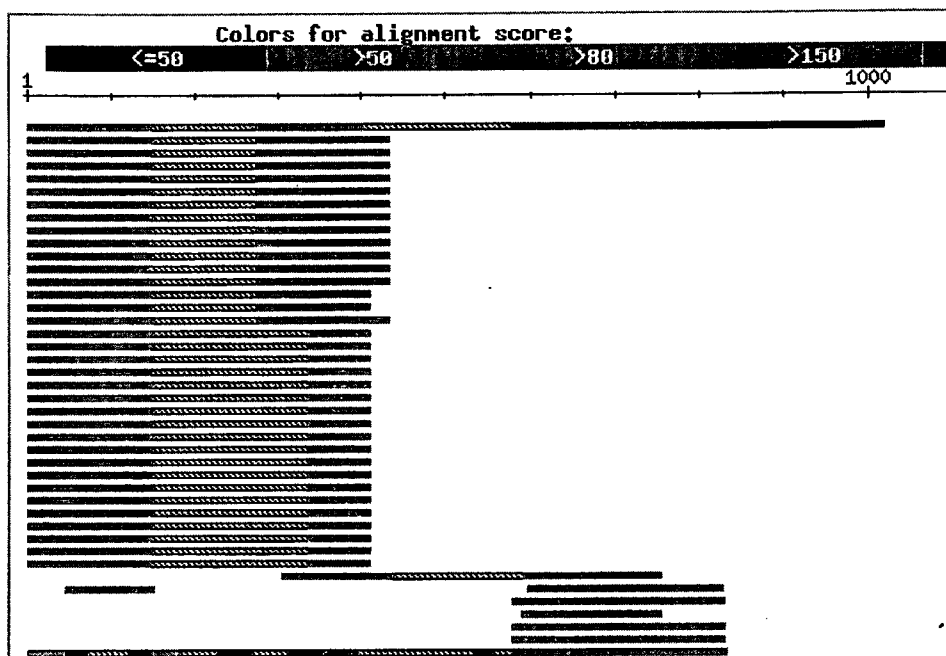
	1	50
SPHINX1.76.324.PEP	MSDLIVKDNA LMNASYNLAL VEQRLILLAI IEARETGKGI NANDPLTVHA	
HCBI3.108.201.pep	MSDLIVKDNA LMNASYNLDL VEQRLILLAI LEARESGKGI NANDPLTVHA	
	51	100
25 SPHINX1.76.324.PEP	GSYINQFNQ RHTAYQALKD ACKDLFARQF SYQEKREGR INITSRWVSQ	
HCBI3.108.201.pep	ESYINQFGVA RQTAYQALKD ACKDLFARQF SYQEKREGR ANITSRWVSQ	
	101	150
SPHINX1.76.324.PEP	IGYMDDTATV EIIFAPAVVP LITRLEEQFT QYDIEQISGL SSAYAVRMYE	
HCBI3.108.201.pep	IAYIDETATV EVIFAPAVVP LITRLEEQFS KYDIEQISSL SSAYAVRLYE	
30	151	200
SPHINX1.76.324.PEP	LLICWRSTGK TPIELDEFR KRIGVLDTEY TRTDNLKMQV IELALKQINE	
HCBI3.108.201.pep	LLICWRRTGK TPVIDLTEFA S.....VY ER.....	
	201	250
SPHINX1.76.324.PEP	HTDITASIEQ HKKGRVITGF SFMFKHKKQN SDKTPDNTAS SPRIVKHSQI	
HCBI3.108.201.pep	SYES S.....	
35	251	300
SPHINX1.76.324.PEP	PTNIVKQEN AKMSDLEHRA SRVTGEIMRN RLSDRFKQGD ESAIDMMKRI	
HCBI3.108.201.pep	...ILPPSEK PLLKILAARW FLLS~~~~~	
	301	324
40 SPHINX1.76.324.PEP	QSEIITDAIA DQWESKLEEF GVVVF	
HCBI3.108.201.pep	~~~~~	

45

BLASTN2:

50





>>>nrnuc:GI_313507477 Gi|313507477|gb|HQ444404.1| TSE-associated... 321 9e-84
>>>nrnuc:GI_347602264 Gi|347602264|gb|JN541297.1| Uncultured bac... 281 7e-72

5 >>>>nrnuc:GI_313507477 Gi|313507477|gb|HQ444404.1| TSE-associated circular DNA isolate
Sphinx 1.76, complete sequence. 0/0
Length = 1758
Score = 321 bits (162), Expect = 9e-84
Identities = 372/442 (84%)
Strand = Plus / Plus

10 Query: 578 ttatgagtgatttaataagataaacgccctaatgaacgctagttataacttagact 637
Sbjct: 604 ttatgagcgatttaataagataaacgccctaatgaatgctagttataacttagctt 663

15 Query: 638 tgggtgaacagcggttaattcttttagctatccttgaagctagagaatcaggcaaaggaa 697
Sbjct: 664 tgggtgaacagaggttaattctattagcaatcatagaagcgagagaacaggcaaaggga 723

20 Query: 698 ttaacgcaaatagacctcttacagtcctgcagagagttatatcaatcaatttggtgtag 757
Sbjct: 724 ttaatgccaatgatcctcttacggttcagcaggtagctatatcaatcaatttaacgtac 783

25 Query: 758 ctagacagactgcttatcaagccctaaagatgcctgcaaagatttattgcccgtcaat 817
Sbjct: 784 aaaggcatacggcatatcaagccctcaagatgcttgtaaagactgtttgcccgtcaat 843

30 Query: 818 tcagctatcaagaaaagcggtgaacgtggagagctaatattacaagtcggttggtcagcc 877
Sbjct: 844 tcagttaccaagaaaagcgagaaacgaggaacgaattaattacaagtcgatgggtttcgc 903

35 Query: 878 aaattgcttacattgatgaaactgcaacggttgaggttatcttcgcccctgcggttggtc 937
Sbjct: 904 aaattggctatatggacgatacagcaaccggttgagattatcttcgcccctgcggttggtc 963

40 Query: 938 cactgatcacaaggttagaggaacaattctcgaagtacgatattgaacaaattagtagtc 997
Sbjct: 964 ctctgattacacggctagaggaacagttcaccagtagcatattgagcaaattagcggtt 1023

45 Query: 998 tatcgagtgccctatgcagttcg 1019
Sbjct: 1024 tatcgagtgcatatgctgttcg 1045

50 Score = 170 bits (86), Expect = 2e-38
Identities = 121/132 (91%), Gaps = 3/132 (2%)
Strand = Plus / Plus
Query: 273 ttggaagaaattgcccctcgactaagcttgcttagtcaaaaaagtttgagcaaagcgaaa 332
Sbjct: 305 ttgacagaaattgcccctcga---agcttgcttagtcaaaaaagtttgagcaaagcgaaa 361

Query: 333 acatagggcaattttcatgatgaaaatgggcttttaaggcttttaaatgcttttaaggct 392
|||||
Sbjct: 362 acatagggcaattttcatgaagaaatgggcttttaaggcttttaaatgcttttaaatgct 421

5
Query: 393 tttagacatgct 404
|||||
Sbjct: 422 tttagacatgct 433

10
Score = 167 bits (84), Expect = 3e-37
Identities = 132/149 (88%), Gaps = 1/149 (0%)
Strand = Plus / Plus

15
Query: 1 gcgagtgtctacgagcgaagttatgaaagttcgattcttccccctctgaaaaaccgctt 60
|||||
Sbjct: 26 gcgagtgtctacgagcgactcaatgaaagttcgattattcccc-tctggaaaaccgctt 84

20
Query: 61 ttaaaaaatattggctgctagatgggttttactatcgtgagcttttgcttttnnnnnnca 120
|||||
Sbjct: 85 ttaaaaaatattggctgctagatgggttttactatagtgagggttttgcttttaaaaaaaca 144

Query: 121 cgagcgaagcgagttcatagttgcttttg 149
|||||
25 Sbjct: 145 cgagcaaagcgagttcatagttgcttttg 173

>>>nrnuc:GI_347602264 Gi|347602264|gb|JN541297.1| Uncultured bacterium plasmid clone
S6GIO-28 genomic sequence. 0/0
Length = 1214
Score = 281 bits (142), Expect = 7e-72
Identities = 157/162 (96%)
Strand = Plus / Plus

30
Query: 273 tttgaagaaattgccctcgactaagcttgcttagtcaaaaaagtttgagcaaagcgaaa 332
|||||
Sbjct: 869 tttgcagaaattgccctcgactaagcttgcttagtcaaaaaagtttgagcaaagcgaaa 928

35
Query: 333 acatagggcaattttcatgatgaaaatgggcttttaaggcttttaaatgcttttaaggct 392
|||||
Sbjct: 929 acatagggcaattttcgtgatgaaaatgggcttttaaatgcttttaaatgcttttaaatgct 988

40
Query: 393 tttagacatgctgaaacgcaagcctagcaaggcatacagagg 434
|||||
45 Sbjct: 989 tttagacatgctgaaacgcaagcctagcaaggcatacagagg 1030

Score = 167 bits (84), Expect = 3e-37
Identities = 132/149 (88%), Gaps = 1/149 (0%)
Strand = Plus / Plus

50
Query: 1 gcgagtgtctacgagcgaagttatgaaagttcgattcttccccctctgaaaaaccgctt 60
|||||
Sbjct: 590 gcgagtgtctacgagcgactcaatgaaagttcgattattcccc-tctggaaaaccgctt 648

55
Query: 61 ttaaaaaatattggctgctagatgggttttactatcgtgagcttttgcttttnnnnnnca 120
|||||
Sbjct: 649 ttaaaaaatattggctgctagatgggttttactatagtgagggttttgcttttaaaaaaaca 708

60
Query: 121 cgagcgaagcgagttcatagttgcttttg 149
|||||
Sbjct: 709 cgagcaaagcgagttcatagttgcttttg 737

Figure 3

65

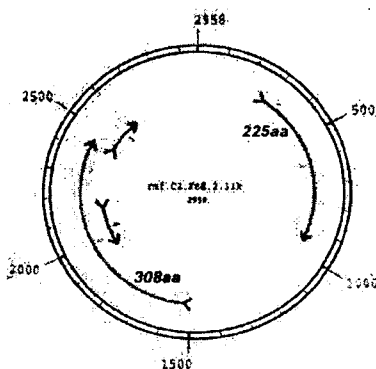
HCB14.296

rmf.C2.F8E.2.31k (2958bp)

```

5  CTGAGAAGCG AAGAAAAGAT GCGGTTTTGT GTATTGAACA TCTTGTGACT GCATCACCAG AATGGGACGG CTGGGGAAGT
   GAAAAAGAAA CTGCATTTTT TGAACAGTCA AGGAAATGGC TTGAAAGCAA ATATGGTAAA AAAAATGTGG TCAGTACAAC
   GATTCATAGA GATGAACAAA CTCACACATT AGTTGCGTAT GTTGTTCCTG TTGACGAAGA AACGGGACGT TTAAATGCTA
   AAAAAATTTAT TGGTGGATCT CGACATACAC TTTCACAGAT GCAGACTGAT TTTGCAGTTG AAGTAAAGGA TTAGGATTA
   GATCGCGGGG TACAGGGGAG CAAAGCAAAA CATACGTCTA TTCAAGAATA TTATGAAAAA TTGAACAATT ATGAGAATGA
10 ACCAGGTATT GAAAAAGGAC TCACCTATGA AGTGCCTGAA CCTGAGTTT TTGAATCTAA AAATGTTTAC GGTGAGAGAG
   TCGCAGAAGC TGTGGCTGCT CAGATAATTG ATCAAATTGC ACCTCGATTG GACAATGCTA ATTTATTGGC TAGTCAACA
   AAAAAATTTAA AAAAAAGAACT GTTAAACACT AGAAAAACGC TTGATGAAGT ACAGAAACGA GCAAAACCTT ATTTGGATAT
   AATCAACGAA TATAATCATC CAAATCTTGA GAAAGAATTT AATAAGCAAG TTGCTAAATT AAAAGATAAT TTTGATTGAG
   CACTTGAGCA TCATAGATTG CTAAAAAGAC AGGAAGAAC AAGAAAGATT AACCAACAAC GTGAACCTCG CAATCAATTA
   CACTTAGAGC AAGAGCAAAA AAAACAACGT GTTGAGCAAG AAAGGCAAGA AAAAGAACGT TTAGCATTCT TAAGACGTCA
15 AGAATTAGAA AATCAGCGGA AAAATGAGCC TAAAAAACCT GATAATGGCA ATAATAACGA CTACTCACC TCATAATACC
   CGTTTAAACG CAAAAAAGAC GGGGGTTTTA GCCCTGTTTG GCCTTTTATG CATAAAGTA TAATTGAAAT AAAAAAATGC
   TCTTAGAACG CAAATGATGA GCATTTAGCG AGTGTCTACG AGCGACACAA TGAAATTCG CCTATTCCCC CTCTGAAAAA
   CTGCTTTTGC TCTTTTGTG TTCTGGAGAG ACTTGTTAGC GAGTGTCTAC GAGCGAAGTA TTGATGCTTT TGCTCTTAAA
   AAAGCATGAG CATAGCGAAT GCATTATCTA TGCTTTTGAC TTTGATTGCT CTCTTGAAAC GACACGAGCA ACGCGAGTGC
20 CATAGCCTTT GATTTTGCTT TTTTGGGCT TTTAATGTTT TTAATGCTT TTAATGCTT TTAGATAGCC TGAAAGCATT
   GCTATACATA TGTTCAGAG CTTATAAGA TACAGATTCC TTGCTATAAA GATACAGATT CCTTGCTATA AAGATACAGA
   TTCCTTGCTA TAAAGATACA GATTCCCTGC ATTAAGGTAC ATAGTCATAT AATGTATCTT TTAATACATG ATTGTATGAA
   ATAATAATA TGAAAAATGG GTTAGTAGTG AAAGATAATG CGTTAATGAA TGCCAGCTAT AATTTAGAAG TAACAGAACA
   GCGCTTAATA CTCTAGCAA TCATTAGTGC AAGAGAAACA GGGCAAGGGA TTACGTCAGA TAGCAAATTA GAAATACATG
25 CTAGTGACTA TGCAATTCAG TTCAGTGTG CAAAAGAAAC AGCTTATGAG GCACATAAAA GTGCTGTAAA TAATTTATTT
   GAGCGTCAAT TTTCAATTTAG AGAAGAAACA AAAAAAGGCA CTGGCATTGT ACGGTCACGA TGGGTTAGCA GAATTTAAATA
   CATTGATGAC GCAGCAATAC TTGAAATCAC TTTTGCCTT GACGTTGTAC CATTAAATCAC TAGACTTGAA GAACACTTCA
   CAAGTTATCA AATCAAGCAA ATAGCACAGC TTACAAGTAA GTACGCTATC CGTTTATATG AACTTCTTAT TGCTTGGCGA
   ACTACAGGCA AAGTCCCTGA GCTTGAAC TAAGAAATTA GAAATAGATT AGGCATAGCT AGTAATGAAT ACACAGCAAT
30 GAACAACCTT AAAAGCCGTG TATTAGAGCC GTCTATTAAG CAGATCAATG AACACACAGA CATTACTGTG ACGTATGAAC
   AGCATAAAAA AGGGCGGACA ATTACAGGCT TTTCATTGAG ATTTAAGCAG AAGCAACAAG CAAAAAAAT AGAACTAAC
   AGAGATCCAA ACACACCTGA CTTTTTTATC AAAATGACCG ATGCTCAACG GCATTATTTC GCTAACAAAA TGTCTGAAAT
   GCCCGAAATG GGGAAATACT CACAAGGTAC AGAAAGCTAC CAACAATTTG CTATTCGCAT CGCCGACATG CTTTTACAAC
   CTGAAAAATT CAGAGAGCTT TATCCAATTT TAGAAAAAGC CGGATTCAAA GAATGATTGA AAAAGAGATA ACAAATTTG
35 AGAAAGAGAT TTTGCTACAA GACAAAATCT CTCAGCTCGA AAATGAAC TAAGAAATTT CTGATCTTCA AAAAAAAGCA
   TATAGCGAAC GGCTTCAAAA AAGTATCGTG GGTTTAGAAA ATAGAATCTA TCGAATCAAG AAAATGCTTT ATACAACCTG
   AAAGGTTGTC AGGGGTTAAG GGGAGGATTC CCCTTACTCA TGAACCTTCA ATGTACCTTT AGGTGCTTTG AAAGTGAGGT
   GAGTCCACTT CGCTATCGCA AAGCTCAAAA ACCTCTGCTA TCGTGTGTTT TTTCGATATC GGATTTATCA AATGTCATTT
40 GCTATTTTGC GTATTCAAAA ATTGAAATCG TTTGCAGACG TTGGCGGTAG TCTTTGCGAT AATTATCGCA ATCGAGAAAC
   GTTAAATGCA GATGATGCTC GTACTCATT GAATGAACAT ACGCTAGATA CAAACGAAAA ATGTATGTCC GCAATCAG

```



69% nucleotide similarity to *Acinetobacter baumannii* SDF plasmid p3ABSDF (24922bp) in 1361nt overlap

70% nucleotide similarity to *Acinetobacter baumannii* ATCC 17978 plasmid pAB1 (13408bp) in 964bp overlap

ORFs:

308 amino acids

58% similarity to *Sphinx1.76* (324 aa)

replication protein

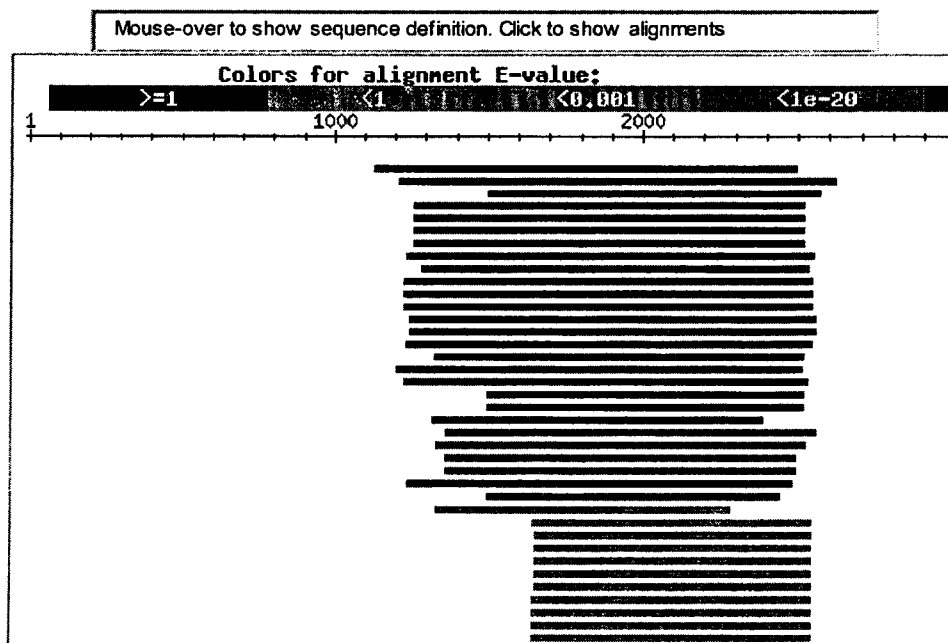
225 amino acids

mobilization protein (mob)

308 amino acids 58% similarity to *Sphinx1.76 (324 aa)* replication protein

5	SPHINX1.76.324.PEP	1	50
	HCBI4.296.308.pep	~MSDLIVKDN ALMNASYNLA LVEQRLILLA IIEARETGKG INANDPLTVH MKNGLVVKDN ALMNASYNLE VTEQRLILLA IISARETGQG ITSDSKLEIH	
10	SPHINX1.76.324.PEP	51	100
	HCBI4.296.308.pep	AGSYINQFNV QRHTAYQALK DACKDLFARQ FSYQEKRRER RINITSRWVS ASDYAIQFSV AKETAYEALK SAVNNLFEHQ FSFREETKKG TGIVRSRWVS	
15	SPHINX1.76.324.PEP	101	150
	HCBI4.296.308.pep	QIGYMDTAT VEIIFAPAVV PLITRLEEQF TQYDIEQISG LSSAYAVRMY RIKYIDDAI LEITFAPDVV PLITRLEEHE TSYQIKQIAQ LTSKYAIRLY	
20	SPHINX1.76.324.PEP	151	200
	HCBI4.296.308.pep	ELLICWRSTG KTPIIELDEF RKRIGVLDTE YTRTDNLKMQ VIELALKQIN ELLIAWRTG KVEPELESEF RNRLGIASNE YTAMNFKSR VLEPSIKQIN	
25	SPHINX1.76.324.PEP	201	250
	HCBI4.296.308.pep	EHTDITASYE QHKKGRVITG FSFMFKHKKQ NSDKTPDTNA SSPR.IVKHS EHTDITVTE QHKKGRITG FSFRFKQKQQ AKKIETNRDP NTPDFIKMT	
30	SPHINX1.76.324.PEP	251	300
	HCBI4.296.308.pep	QIPTNIVKQP ENAKMSDLEH RASRVTEIM RNRLSDRFKQ GDESAIDMMK DAQRHLFAN. ...KMSMP MG..... KYSQ GTESYQQFAI	
35	SPHINX1.76.324.PEP	301	326
	HCBI4.296.308.pep	RIQSEIITDA IADQWESKLE EFGVVF RIADMLLQPE KFRELYPILE KAGFKE	

35 FastA *Geall



40	EMHTG HUM:AL391558	Begin: 42149	End: 43446	Strand: -	
	! A1391558 Human DNA sequence *** SEQ...	1068	2897	2878	2016.7 2.3e-108
45	EMSTD PRO:CU468233	Begin: 275	End: 1601		
	! Cu468233 Acinetobacter baumannii SD...	1182	3120	2725	1914.3 1.1e-101
45	EMSTD PRO:CP000522	Begin: 1	End: 957		
	! Cp000522 Acinetobacter baumannii AT...	946	1947	2123	1492.1 6.9e-78
45	EMSTD PRO:EF138630	Begin: 121	End: 1303		
	! Ef138630 Acinetobacter baumannii is...	570	1739	2096	1473.3 8.4e-77
45	EMSTD PRO:JX069966	Begin: 90	End: 1272		

! Jx069966 Acinetobacter baumannii pl... 570 1739 2096 1473.6 9.1e-77
EMSTD PRO:CP002523 Begin: 953 End: 2135
! Cp002523 Acinetobacter baumannii TC... 570 1876 2096 1474.1 1.1e-76

5 EMHTG HUM:AL391558
ID AL391558; SV 6; linear; genomic DNA; HTG; HUM; 237855 BP.
AC AL391558;
DE Human DNA sequence *** SEQUENCING CANCELLED *** from clone RP11-192D3
KW HTG; HTGS_CANCELLED; HTGS_PHASE1. . . .
10 SCORES z-score: 2016.7 E(): 2.3e-108
>>EMHTG_HUM:AL391558 (237855 nt)
70.4% identity in 1310 nt overlap
(2463-1172:42149-43446)

15 2490 2480 2470 2460 2450 2440
rmf.C2.F8E.2 AAAATCTCTTTCTCAAATTTTGTATCTCTTTTCAATCATTCTTGAATCCGGCTTTT
AL391558 TTCTAAAGAAAAGGTCATTTTAAGTGAATGTTTAAATTTATGTTGAAAACCAACTTTT
20 42120 42130 42140 42150 42160 42170
2430 2420 2410 2400 2390 2380
rmf.C2.F8E.2 CTAAATTTGGATAAAGCTCTCTGAATTTTTCAGGTTGTAAAAGCATGTCGGCGATGCGAA
AL391558 CTAAACAGGATACAGTTCTCTAAATTTTTCAGGCTCTAACACATATTTGCGATACGTA
25 42180 42190 42200 42210 42220 42230
2370 2360 2350 2340 2330 2320
rmf.C2.F8E.2 TAGCAAATGTTGGTAGCTTTCTGTACCTTGTGAGTATTCCCATTTTCGGGCATTTCAG
AL391558 TGGCAAATGTTGATAGCTTTCCGTACCTGTGAATACTTCCCATTTTCGGGCATTTCAG
30 42240 42250 42260 42270 42280 42290
2310 2300 2290 2280 2270 2260
rmf.C2.F8E.2 ACATTTTGTAGCGAATAAATGCCGTTGAGCATCGGTCATTTTGATAAAAAAGTCAGGTG
AL391558 ACATTTTATTTGCAAATAAATGACGTTGTGCATCGGTCATTTTGACAAAAAATCAGGGG
35 42300 42310 42320 42330 42340 42350
2250 2240 2230 2220 2210 2200
rmf.C2.F8E.2 TGTTTGGATCTCTGTTAGTTTCTATTTTTTTTGTCTGTTGCTTCTGCTTAAATCTGAATG
AL391558 TATTTGGATCACGTTTAAATTCAA---CTTTCGGCTGTTGTTTATGCTTAAATTTGAATG
40 42360 42370 42380 42390 42400 42410
2190 2180 2170 2160 2150 2140
rmf.C2.F8E.2 AAAAGCCTGTAATTTGCCGCCCTTTTTATGCTGTTTCATACGTCACAGTAATGCTGTGT
AL391558 AAAACCCTGTGATGGTACGCCCTTGTATTGCTGTTTCATAAGTTGCTTTTATGCTGTAT
45 42420 42430 42440 42450 42460 42470
2130 2120 2110 2100 2090 2080
rmf.C2.F8E.2 GTTCATTGATCTGCTTAATAGACGGCTCTAATACACGGCTTTTAAAGTTGTTTCATTGCTG
AL391558 GCTCATTGATTGAGTAATAGCTGTTTCTAGTACACGGCTTTTAAATGGTGCATACGTT
50 42480 42490 42500 42510 42520 42530
2070 2060 2050 2040 2030 2020
rmf.C2.F8E.2 TGTATTCACTAGCTATGCCTAATCTATTTCTAAATCTGATAGTTCAAGCTCAGGGA
AL391558 GATACTCATCATCGTCCACACCTAAGTTTTCAGTAGCTGTTGCAGTTCAAAGGTTGGCG
55 42540 42550 42560 42570 42580 42590
2010 2000 1990 1980 1970 1960
rmf.C2.F8E.2 CTTTGCCTGTAGTTCCGCAAGCAATAAGAAGTTCATATAACGGATAGCGTACTTACTTG
AL391558 TTTTCCAACTTCACGCCAAGCGATTAATAATTCATAAAGTCTAGTGGCGTACTTACTCG
60 42600 42610 42620 42630 42640 42650
1950 1940 1930 1920 1910 1900
rmf.C2.F8E.2 TAAGCTGTGCTATTTGCTTGATTGATAACTTGTGAAGTGTCTTCAAGTCTAGTGATTA
AL391558 TTAGGTGTGCAACTTGTTCGCTTCATATTTGTAAAATGCTCTTCCAATCGAGTGATTA
65 42660 42670 42680 42690 42700 42710
1890 1880 1870 1860 1850 1840
rmf.C2.F8E.2 ATGGTACAACGTCAGGCGCAAAAGTGATTTCAAGTATTGCTGCGTCATCAATGATTTAA
AL391558 AAGGTACTACATCAGGTGCGAAAGTAATTTCTAGTAATGCTAAGTCATCTACATAAAAAA
70 42720 42730 42740 42750 42760 42770

```
75  EMSTD_PRO:CU468233
    ID   CU468233; SV 1; circular; genomic DNA; STD; PRO; 24922 BP.
    AC   CU468233;
    DE   Acinetobacter baumannii SDF plasmid p3ABSDF, complete genome. . .
    SCORES  Init1: 1182  Initn: 3120  Opt: 2725  z-score: 1914.3  E(): 1.1e-101
```

>>EMSTD_PRO:CU468233 (24922 nt)
initn: 3120 initl: 1182 opt: 2725 Z-score: 1914.3 expect(): 1.1e-101
69.1% identity in 1361 nt overlap
(1239-2569:275-1601)

EMSTD_PRO:CP000522
ID CP000522; SV 1; circular; genomic DNA; STD; PRO; 13408 BP.
AC CP000522;
DE *Acinetobacter baumannii* ATCC 17978 plasmid pAB1, complete sequence. . . .
SCORES z-score: 1492.1 E(): 6.9e-78
>>EMSTD_PRO:CP000522 (13408 nt)
69.6% identity in 964 nt overlap
(1530-2484:1-957)

EMSTD_PRO:EF138630
ID EF138630; SV 1; linear; genomic DNA; STD; PRO; 12200 BP.
AC EF138630;
DT 12-MAY-2007 (Rel. 91, Created)
DT 27-JUN-2007 (Rel. 92, Last updated, Version 3)
DE *Acinetobacter baumannii* isolate 186 plasmid pOUR, partial sequence.
KW
initn: 1739 initl: 570 opt: 2096 Z-score: 1473.3 expect(): 8.4e-77
65.1% identity in 1198 nt overlap
(1292-2484:121-1303)

EMSTD_PRO:JX069966
ID JX069966; SV 1; circular; genomic DNA; STD; PRO; 10879 BP.
AC JX069966;
DE *Acinetobacter baumannii* plasmid pAB120, complete sequence.
KW
SCORES z-score: 1473.6 E(): 9.1e-77
>>EMSTD_PRO:JX069966 (10879 nt)
65.1% identity in 1198 nt overlap
(1292-2484:90-1272)

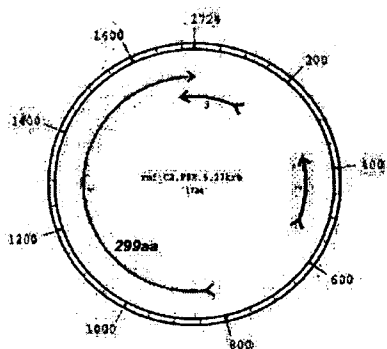
Figure 4

HCBI5.173 *rmf.C2.F8E.5.23kr (1724 bp)*

```

5  CCATTAGCTT TATAGTGCCA AGTAATATCT AGTAAACGGA TATGTTTTAT ATGATCTTTT CTTGAAAAT GAAGAATCAT
   ATCTGAGGTC GAATAGTTGT TATATTTACT TTTTACGTAA TCAGGCATAC CGATATGATA ATGCCAAAGA TTATTCGAAT
   ATGTGAAATC ATAGATTGGA TCAGTTTTGT CTATTCCACG CCATGAAGGT GCAATTTTAC CTTGGTATTT AGAAAAATCC
   TTTAAACCAA ACTCTTCATA GATATCAGTG AAATCTAAAA TTTTATCTTG TTGATCTTCA GGATAACGCC CAAATGAAC
   TGCAAATTGT TTTCCAAATT CCCAAGAGTA AGACAATTAC TTTTCCCTG CTCTAATAT TAATCTTCTT TTCTCTTCTC
10  TGCTTAAACC TTGAGGCATA ACAAAGGTTT CTGAATCAAG AGCTGATTTC ATTCTATCAA GATTAATACG AACAGGTGCT
   GTTTTTTGAT TTGATGTTTG GGTAAATAACG AAGCCCATTA CATTCAATTG CCACCTCGAC AAATAAAAAAT TTCTTTTAAA
   ATTTTAAACA AAAAAATAGC GTAGCGGGTT TTTAATTGTT TTTAAATACT TTTAATTGTT TTTAGACCTG CTGAAACCCG
   CTATACAAAA GGCTTTTACG CCATATAAAA AGACATTTAC CCTGCCATAA AAAGACATTT ACCCTGCCAT AAAAAGACAT
   TTACCCCTGCC ATAAAAAGAC ATTTACCCCTG CTTAAAAAGA CATTTTACAA AATGTCTTTT TAATGGTAGC TTTAGGTTGT
   TTTTATATAA AAAAATTGGC GTTAAAAAAA TATGGATTGA GTCGTAAAG ATAATAATTT GATCAATGCT AGCTACTCCC
15  TAGGTTTGGT TGAGCAACGT CTTGTATTGC TAGCTATCAT TGAAGCTAGA GAATCTAGTA AGGGGATAGA TTCAGAAACA
   TTNTTAGAAA TACATGCTCA ACACATATGCA GATAGATTG ATGTAAATGT AAAAAATACG TATGCCATGC TTTAGAGGC
   GGCACAGACT CTTTTTAATC GTCAAGTCAC TTACATGATG GTTGATGAAA AAAGGAATAA ACCTGAAAAG CGTGTAATTC
   GCTGGGTTAG TGGTATTTCA TATGTAGAAG GGGCTGGAGT TCTTAAACTT CGTTTTTCTC CAGAAATTGT CCCTCTAATA
   ACTAGGCTAG AGCAAAATTT TACGAGTTAT GAACAGAAC AAGTTAAAAG CTTAAACGTA TATGCAACTC GACTATATGA
20  ATTATTAGTC TGTGCGCTA GTACAGGTAA AACTCCCATT ATTGAAATAG AAGATTTTCG TTCAAAAATA GGAGTCTTAC
   CTACTGAATA TAAATTAATG AGTGATTCA AAAAGCGAGT TTTTGAGCCA GCAATACAGC AAATTAATAA AAACACAGAT
   TTGACGGTTA ATTATGATCA ACATAAATCA GGCCGTACAA TTACAGGCTT CTCATTAAA TTTAAGCAAA AGAAACGAA
   ATCAGAAAAG GTCGTAACCG CTAACGAGA CCCAATACA CCTGACTTTT TTGTAAAAAT GACCGATTCT CAACGTCATT
   TATTGCGAAC TAAACTTTCT GAAATGCCAG AGATGTCTAA ATATTCTCAA GGTACAGAGT CGTATCAACA GTTTGCTATC
25  CGTATTGCCG ATATGCTTTT AGAACCTGAA AAGTTAAAGA ATTC

```

*Nucleotide sequence similarities to:*

Dendroctonus ponderosae genomic DNA (GI_462465081) 70% in 605bp
 Acinetobacter baumannii SDF plasmid p3ABSDF (GI_169150781) 74% in 367bp, 72% in 403
 and 77% in 71bp

299 amino acids 49% similarity to Sphinx1.76 (324aa) replication protein

ORF:

299 amino acids 49% similarity to Sphinx1.76 (324aa) replication protein

SPHINX1.76.324.PEP	MSDLIVKDNA	LMNASYNLAL	VEQRLILLAI	IEARETGKGI	NANDPLTVHA
HCBI5.173.299.pep	~MDLVVKDNN	LINASYSLGL	VEQRLVLLAI	IEARESSKGI	DSETXLEIHA

SPHINX1.76.324.PEP	GSYINQFNVO	RHTAYQALKD	ACKDLFARQF	SYQEKREGR	..INITSRWV
HCBI5.173.299.pep	QHYADRFQVN	VKNITYAMLSE	AAQTLFNRQV	TYMMVDEKRN	KPEKRVIRWV

SPHINX1.76.324.PEP	SQIGYMDDTA	TVEIIFAPAV	VPLITRLEEQ	FTQYDIEQIS	GLSSAYAVRM
HCBI5.173.299.pep	SGISYVEGAG	VKLRESPEI	VPLITRLEQN	FTSYELEQVK	SLN.VYATRL

SPHINX1.76.324.PEP	YELLICWRST	GKTPIIELDE	FRKRIGVLDT	EYTRDNLKM	QVIELALKQI
HCBI5.173.299.pep	YELLVCWRST	GKTPIIIEIED	FRSKIGVLPT	EYKLMSDFKK	RVFEPAIQOI

201

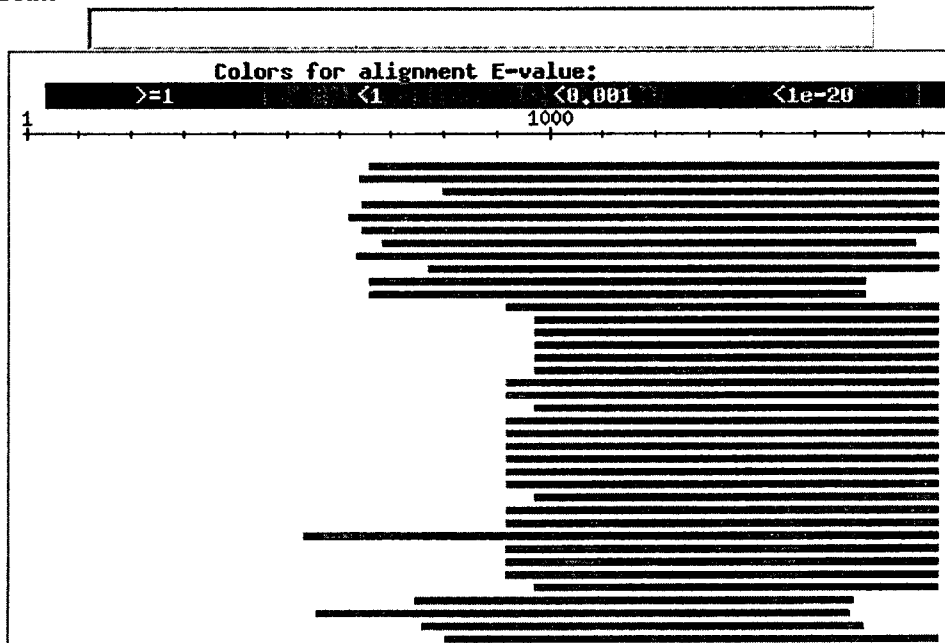
250

SPHINX1.76.324.PEP NEHTDITASY EQHKKGRVIT GFSFMFKHKK QNSDKTPDTN ASSPRIVKHS
 HCB15.296.310.pep NKMTDLTVNY DQHKSGRTIT GFSFKFKQKK TKSEKVVTA KDP.....

5 SPHINX1.76.324.PEP 251 300
 HCB15.296.310.pep QIPTNIVKQP ENAKMSDLEH RASRVTEIM RNRLSDRFKQ GDESAIDMMK
 NTPDFFVKMT D....SQRHL FATKLS.... EMPEMSKYSQ GTESYQQFAI

10 SPHINX1.76.324.PEP 301 326
 HCB15.173.299.pep RIQSEIITDA IADQWESKLE EFGVVVF
 RIADMLE... ..PEKLLK NSH~~~

FastA *Geall



15 EMSTD PRO:GU978996 Begin: 1 End: 1025
 ! Gu978996 Acinetobacter baumannii pl... 459 1221 1633 994.9 4.3e-49
 EMSTD PRO:GU978999 Begin: 30 End: 1085
 ! Gu978999 Acinetobacter baumannii pl... 554 1233 1613 982.7 1.9e-48
 20 EMSTD PRO:CP000522 Begin: 4 End: 893
 ! Cp000522 Acinetobacter baumannii AT... 531 1254 1504 908.7 2.1e-45
 EMSTD PRO:DQ278485 Begin: 1 End: 1049
 ! Dq278485 Acinetobacter venetianus s... 377 683 1445 873.8 2.3e-43
 25 EMSTD PRO:GU979001 Begin: 8 End: 1085
 ! Gu979001 Acinetobacter baumannii pl... 402 733 1453 886.1 4.7e-43
 EMSTD ENV:JN098514 Begin: 1 End: 1049
 ! Jn098514 Uncultured bacterium plasm... 377 683 1445 877.8 4.7e-43
 EMSTD PRO:GQ861437 Begin: 7 End: 910 Strand: -
 ! Gq861437 Acinetobacter baumannii st... 525 1189 1441 874.7 5.7e-43
 30 EMSTD PRO:CU468231 Begin: 323 End: 1369
 ! Cu468231 Acinetobacter baumannii SD... 484 1069 1395 845.5 1.6e-41
 EMSTD PRO:AY541809 Begin: 19 End: 944

35 EMSTD PRO:GU978996
 ID GU978996; SV 1; linear; genomic DNA; STD; PRO; 1065 BP.
 AC GU978996;
 DE **Acinetobacter baumannii plasmid p736 plasmid replication protein Aci 7**
 DE (repA) gene, complete cds. . . .

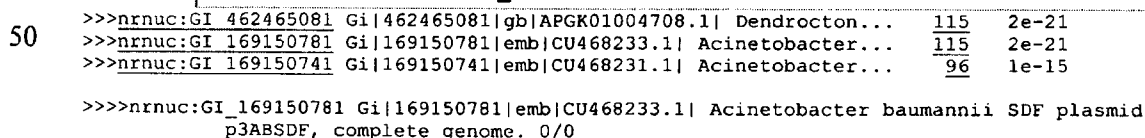
40 SCORES z-score: 994.9 E(): 4.3e-49
 >>EMSTD PRO:GU978996 (1065 nt)
63.2% identity in 1034 nt overlap
 (688-1709:1-1025)

45 rmf.C2.F8E.5 AGGCCATATAAAAGACATTTACCTGCCATAAAAGACATTTACCTGCCATAAAAGA
 ||||| || ||||| ||| ||||| ||
 GU978996 TAAACGAGGTTTACCTTGCAATAAACGA
 10 20 30

50 rmf.C2.F8E.5 720 730 740 750 760 770
 CATTACCTGCCATAAAAGACATTTACCTGC-TTAAAGACATT----TTCAAAA

27/29

45 ***BLASTN2 :***



70

29/29