

(19) **DANMARK**



Patent- og
Varemærkestyrelsen

(10) **DK/EP 3447149 T3**

(12) **Oversættelse af
europæisk patentskrift**

-
- (51) Int.Cl.: **C 12 Q 1/68 (2018.01)** **C 12 Q 1/6883 (2018.01)** **C 12 Q 1/6886 (2018.01)**
C 12 Q 1/70 (2006.01)
- (45) Oversættelsen bekendtgjort den: **2021-05-17**
- (80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2021-02-17**
- (86) Europæisk ansøgning nr.: **18192671.8**
- (86) Europæisk indleveringsdag: **2014-10-30**
- (87) Den europæiske ansøgnings publiceringsdag: **2019-02-27**
- (30) Prioritet: **2013-10-30 EP 13005138**
- (62) Stamansøgningsnr: **14806534.5**
- (84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**
- (73) Patenthaver: **Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, 69120 Heidelberg, Tyskland**
- (72) Opfinder: **ZUR HAUSEN, Harald, Eichenstrasse 1, 69483 Wald-Michelbach, Tyskland**
DE VILLIERS, Ethel-Michele, Eichenstrasse 1, 69483 Wald-Michelbach, Tyskland
GUNST, Karin, Obergasse 2, 69493 Hirschberg, Tyskland
Funk, Mathis, rue du Muguet 17, 67400 Illkirch, Frankrig
Lamperto Perez, Iranzu, C/ Avenida Bayona n°35 - 2D, 31011 Pamplona, Spanien
- (74) Fuldmægtig i Danmark: **Patrade A/S, Ceresbyen 75, 8000 Århus C, Danmark**
- (54) Benævnelse: **HCBI-SEKVENSER SOM EN TIDLIG MARKØR FOR FREMTIDIG UDVIKLING AF CANCER OG SYGDOMME I CNS OG SOM ET MÅL FOR BEHANDLING OG FOREBYGGELSE AF CANCER**
- (56) Fremdragne publikationer:
BRASSARD ET AL.: "Molecular detection of bovine and porcine Torque teno virus in plasma and feces", VETERINARY MICROBIOLOGY, ELSEVIER BV, NL, vol. 126, no. 1-3, 19 November 2007 (2007-11-19), pages 271-276, XP022351086, ISSN: 0378-1135, DOI: 10.1016/J.VETMIC.2007.07.014
SCHUURMAN R ET AL.: "Bovine polyomavirus, a frequent contaminant of calf serum", BIOLOGICALS, vol. 19, 1991, pages 265-270, XP002723221,
LEARY T P ET AL.: "Improved detection systems for TT virus reveal high prevalence in humans, non-human primates and farm animals", JOURNAL OF GENERAL VIROLOGY, vol. 80, 1999, pages 2115-2120, XP002723222,
FUCHS M ET AL.: "Detection of bovine herpesvirus 1 in blood from naturally infected cattle by using a sensitive PCR that discriminates between wild-type virus and virus lacking glycoprotein E", JOURNAL OF CLINICAL MICROBIOLOGY, vol. 37, no. 8, August 1999 (1999-08), pages 2498-2507, XP002723223,
SPRADBROW P B ET AL.: "Skin cancer and papillomaviruses in cattle", JOURNAL OF COMPARATIVE PATHOLOGY, ACADEMIC PRESS, LONDON, GB, vol. 97, no. 4, 1 July 1987 (1987-07-01), pages 469-479, XP022995403, ISSN: 0021-9975, DOI: 10.1016/0021-9975(87)90025-9 [retrieved on 1987-07-01]
GIOVANNY M ET AL.: "Bovine leukemia virus gene segment detected in human breast tissue", OPEN

Fortsættes ...

JOURNAL OF MEDICAL MICROBIOLOGY, vol. 3, March 2013 (2013-03), pages 84-90, XP002723224,
ZUR HAUSEN H: "Red meat consumption and cancer: reasons to suspect involvement of bovine infectious factors in colorectal cancer", INTERNATIONAL JOURNAL OF CANCER, vol. 130, 2012, pages 2475-2483, XP002723225,
ZUR HAUSEN ET AL: "The search for infectious causes of human cancers: Where and why", VIROLOGY, ELSEVIER, AMSTERDAM, NL, vol. 392, no. 1, 15 September 2009 (2009-09-15), pages 1-10, XP026541903, ISSN: 0042-6822, DOI: 10.1016/J.VIROL.2009.06.001 [retrieved on 2009-08-29]
"Viren in Rindfleisch können Darmkrebs verursachen", ONLINE FOCUS , 4 March 2012 (2012-03-04), XP002723226, Retrieved from the Internet: URL:http://www.focus.de/gesundheit/news/darmkrebs-forscher-vermutet-krebserreger-in-rindfleisch_aid_720424.html [retrieved on 2014-04-10]

DESCRIPTION

[0001] 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. 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.

[0002] 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). "Red" meat is regarded as comprising beef, pork, mutton, lamb and goat meat, in contrast to "white" meat (poultry meat/fish).

[0003] Thus far, chemical carcinogenic substances being produced during roasting, grilling, barbecuing, smoking and air-drying were blamed as risk factors for cancer. However, often the fact was disregarded that the same substances are also produced in comparable concentrations during analogous ways of preparation of poultry meat/fish. Accordingly, this does not support the assumption that these chemical substances play an exclusive role as regards the development of colon cancer. Since, in 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 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.

[0004] 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.

[0005] 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 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.

[0006] 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. Brassard et al., *Veterinary Microbiology*, Vol. 126, No. 1-3, pp. 271-276 (2007) describe the molecular detection of bovine and porcine Torque teno virus in plasma and feces. Schuurman et al., *Biologicals*, Vol. 19, pp. 265-270 (1991) describe that bovine polyomavirus is a frequent contamination of calf serum. Leary et al., *J. of General Virology*, Vol. 80, pp. 2115-2120 (1999) describe improved systems for TT virus detection. Fuchs et al., *J. of Clinical Microbiology*, Vol. 37, No. 8, pp. 2498-2507 (1999) describe the detection of bovine herpesvirus 1 in blood from naturally infected cattle by using sensitive PCR that discriminates between wild-type virus and virus lacking glycoprotein E. Spradbrow et al., *J. of Comparative Pathology*, Vol. 97, No. 4, pp. 469-479 (1987) describe a connection between skin cancer and papillomaviruses in cattle. Giovanni et al., *Open J. of Medical Microbiology*, Vol. 3, pp. 84-90 (2013) describe that a bovine leukemia virus gene segment has been detected in human breast tissue.

Brief description of the drawings

[0007]

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

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(Comparative Figure)

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(Comparative Figure)

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 (Comparative Figure)

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 frames of HCBI5.173 to corresponding "Sphinx" amino acid sequences.

[0008] Accordingly, the present invention relates to an HCBI polynucleic acid comprising:

1. (a) a nucleotide sequence depicted in Figure 5;
2. (b) a nucleotide sequence having at least 90% identity to a nucleotide sequence of (a);
3. (c) a fragment of a nucleotide sequence of (a) or (b);
4. (d) a nucleotide sequence being complementary to a nucleotide sequence of (a), (b) or (c); or
5. (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.

[0009] The term "polynucleic acid" refers to a single-stranded or double-stranded nucleic acid sequence. A polynucleic acid may consist of deoxyribonucleotides or ribonucleotides, nucleotide analogues or modified nucleotides or may have been adapted for diagnostic or therapeutic purposes. A polynucleic acid may also comprise a double stranded cDNA clone which can be used, for example, for cloning purposes.

[0010] The HCBI5.173 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.

[0011] 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 sequence as shown in Figure 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.

[0012] 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.

[0013] The person skilled in the art can easily determine which nucleic acid sequences are related to a nucleotide sequence of Figure 5 or which fragments are still capable of replicating autonomously by using standard assays.

[0014] 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. These variant polynucleic acid sequences will thus encode the same amino acid sequence as the polynucleic acids they are derived from.

[0015] The HCBI5.173 polynucleic acid 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.

[0016] 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 HCBI5.173 polynucleic acid of the invention.

[0017] 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. 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.

[0018] 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 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 conveniently labelled either using labelled primers or by incorporating labelled nucleotides.

[0019] Labels may be isotopic (^{32}P , ^{35}S , etc.) or non-isotopic (biotin, digoxigenin, etc.). The amplification reaction is repeated between 20 and 70 times, advantageously between 25 and 45 times.

[0020] 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 the corresponding amino acid sequence. Exemplary sequencing reactions include

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

[0021] 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.

[0022] The present invention also relates to an oligonucleotide probe comprising or consisting of part of a HCBI5.173 polynucleic acid as defined above, with said probe being able to act as a hybridization probe for specific detection of a HCBI5.173 polynucleic acid according to the invention.

[0023] The probe can be labelled or attached to a solid support.

[0024] The term "probe" refers to single stranded sequence-specific oligonucleotides which have a sequence which is complementary to the target sequence of a HCBI5.173 polynucleic acid to be detected.

[0025] 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.

[0026] 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 substrate will be a microtiter plate, a membrane (e.g. nylon or nitrocellulose) or a microsphere (bead). Prior to application to the membrane or fixation it may be convenient to modify the nucleic acid probe in order to facilitate fixation or improve the hybridization efficiency. Such modifications may encompass homopolymer tailing, coupling with different reactive groups such as aliphatic groups, NH₂ groups, SH groups, carboxylic groups, or coupling with biotin or haptens.

[0027] The oligonucleotides according to the present invention, used as primers or probes may also contain or consist of nucleotide analogues such as phosphorothioates, alkylphosphoriates or peptide nucleic acids or may contain intercalating agents. These modifications will necessitate adaptations with respect to the conditions under which the oligonucleotide should be used to obtain the required specificity and sensitivity. However, the eventual results will be essentially the same as those obtained with the unmodified oligonucleotides.

[0028] The introduction of these modifications may be advantageous in order to positively influence characteristics such as hybridization kinetics, reversibility of the hybrid-formation, biological stability of the oligonucleotide molecules, etc.

[0029] 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.

[0030] The present invention also relates to a recombinant expression vector comprising an HCBI5.173 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.

[0031] 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 HCBI5.173 recombinant molecules, BCG or adenoviral vectors, as well as avipox recombinant viruses.

[0032] 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.

[0033] 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.

[0034] 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.

[0035] 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.

[0036] 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 insect cell 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.

[0037] 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.

[0038] The segment of the HCB15.173 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-HCB15.173 source, but particularly preferred constructs according to the present invention contain signal sequences appearing in the HCB1 genome before the respective start points of the proteins.

[0039] 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.

[0040] 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 post translational modification recognized by insect cells are known in the art.

[0041] The present invention also relates to a polypeptide having an amino acid sequence encoded by an HCB1 polynucleic acid as defined above, or a part or an analogue thereof being substantially similar and biologically equivalent.

[0042] 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 (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.

[0043] The polypeptides according to the present invention contain preferably at least 3, preferably 4 or 5 contiguous HCB15.173 amino acids, 6 or 7 preferably however at least 8 contiguous HCB15.173 amino acids, at least 10 or at least 15.

[0044] 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. 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 the appropriate regulatory elements, (b) culturing said transformed cellular host under conditions enabling the expression of said insert, and (c) harvesting said polypeptide.

[0045] The present invention also relates to an antibody raised upon 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 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 HCB15.173 polynucleic acid of the invention or a fragment thereof by methods well known to those skilled in the art. As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')₂ fragments) which are capable of specifically binding to protein. Fab and F(ab')₂ fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies useful for the purposes of the present invention include chimerical, single chain, and humanized antibodies.

[0046] Preferably, the antibody or antigen binding fragment thereof carries a detectable label. The antibody/fragment can be directly or indirectly detectably labeled, for example, with a radioisotope, a fluorescent compound, a bioluminescent compound, a chemiluminescent compound, a metal chelator or an enzyme. Those of ordinary skill in the art will know of other suitable labels for binding to the antibody, or will be able to ascertain such, using routine experimentation.

[0047] The present invention also relates to a diagnostic kit for use in determining the presence of a HCB15.173 polynucleic acid or polypeptide of the invention, said kit comprising a primer, a probe, and/or an antibody of the invention.

[0048] The present invention also relates to a method for the detection of an HCB15.173 polynucleic acid according to the invention present in a biological sample, comprising: (a) optionally extracting sample polynucleic acid, (b) amplifying the polynucleic acid as described above with at least one primer as defined above, optionally a labelled primer, and (c) detecting the amplified polynucleic acids.

[0049] The term "polynucleic acid" can also be referred to as analyte strand and corresponds to a single- or double-stranded polynucleic acid molecule.

[0050] The term "labelled" refers to the use of labelled nucleic acids. This may include the use of labelled nucleotides incorporated during the polymerase step of the amplification or labelled primers, or by any other method known to the person skilled in the art.

[0051] The present invention also relates to a method for the detection of an HCBI5.173 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.

[0052] 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 to attain sufficient specificity.

[0053] 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, 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 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 TMAI solutions.

[0054] Suitable assay methods for purposes of the present invention to detect hybrids formed between the oligonucleotide probes and the HCBI5.173 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.

[0055] 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.

[0056] The present invention also relates to a method for detecting a polypeptide encoded by an HCBI5.173 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.

[0057] 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 antigen of 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 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 detection of immune complexes comprised of the antigen.

[0058] Design of the immunoassays is subject to a great deal of variation, and many formats are known in the art. Protocols may, for example, use solid supports, or immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, enzymatic, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the immune complex are also known; examples of which are assays which utilize biotin and avidin or streptavidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

[0059] The immunoassay may be in a heterogeneous or in a homogeneous format, and of a standard or competitive type. In a heterogeneous format, the polypeptide is typically bound to a solid matrix or support to facilitate separation of the sample from the polypeptide after incubation. Examples of solid supports that can be used are nitrocellulose (e. g., in membrane or microtiter well form), polyvinyl chloride (e. g., in sheets or microtiter wells), polystyrene latex (e. g., in beads or microtiter plates, polyvinylidene fluoride (known as Immulon), diazotized paper, nylon membranes, activated beads, and Protein A beads. The solid support containing the antigenic polypeptides is typically washed after separating it from the test sample, and prior to detection of bound antibodies. Both standard and competitive formats are known in the art.

[0060] 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.

[0061] 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.

[0062] 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).

[0063] 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 forms a visible layer or film of precipitate. If no anti-HCBI antibody is present in the test specimen, no visible precipitate is formed.

[0064] There currently exist three specific types of particle 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 specific antigen/antibodies present in the body component, if any, causes the RBCs coated with the purified antigen to agglutinate.

[0065] To eliminate potential non-specific reactions in the hemagglutination assay, two artificial carriers may be used instead of RBC in the PA. The most common of these are latex particles.

[0066] The solid phase selected can include polymeric or glass beads, 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 beta-galactosidase. Examples of enhancer compounds include biotin, anti-biotin and avidin. Examples of enhancer compounds binding members include biotin, anti-biotin and avidin.

[0067] The above methods are useful for evaluating the risk of 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.

[0068] Thus, the present invention also relates to an antisense oligonucleotide or iRNA specific for the HCBI5.173 virus polynucleic acid of the invention.

[0069] The generation of suitable antisense oligonucleotides or iRNAs includes determination

of a site or sites within the HCB15.173 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 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 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 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, then the oligonucleotide and the DNA or RNA are considered to be complementary to each other at that position. The oligonucleotide and the DNA or RNA are complementary to each other when a sufficient number of corresponding positions in each molecule are occupied by nucleotides which can hydrogen bond with each other. Thus, "specifically hybridizable" and "complementary" are terms which are used to indicate a sufficient degree of complementarity or precise pairing such that stable and specific binding occurs between the oligonucleotide and the DNA or RNA target. It is understood in the art that the sequence of an antisense compound does not need to be 100% complementary to that of its target nucleic acid to be specifically hybridizable. An antisense compound is specifically hybridizable when binding of the compound to the target DNA or RNA molecule interferes with the normal function of the target DNA or RNA to cause a loss of utility, and there is a sufficient degree of complementarity to avoid non-specific binding of the antisense compound to non-target sequences under conditions in which specific binding is desired, i.e., in the case of therapeutic treatment.

[0070] "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.

[0071] Antisense compounds include ribozymes, external guide sequences (EGS), oligonucleotides (oligozymes), and other short catalytic RNAs or catalytic oligonucleotides

which hybridize to the target nucleic acid and inhibit its expression. The antisense compounds also include an iRNA comprising a sense sequence and an antisense sequence, wherein the sense and antisense sequences form an RNA duplex and wherein the antisense sequence comprises a nucleotide sequence sufficiently complementary to the nucleotide sequence of an HCBI5.173 polynucleic acid of the present invention.

[0072] Alternatively, the invention provides a vector allowing to transcribe an antisense oligonucleotide of the invention, e.g., in a mammalian host. Preferably, such a vector is a vector useful for gene therapy. Preferred vectors useful for gene therapy are viral vectors, e.g. adenovirus, adeno-associated virus, herpes simplex virus, vaccinia, or, more preferably, an RNA virus such as a retrovirus. Even more preferably, the retroviral vector is a derivative of a murine or avian retrovirus. Examples of such retroviral vectors which can be used in the present invention are: Moloney murine leukemia virus (MoMuLV), Harvey murine sarcoma virus (HaMuSV), murine mammary tumor virus (MuMTV) and Rous sarcoma virus (RSV). Most preferably, a non-human primate retroviral vector is employed, such as the gibbon ape leukemia virus (GaLV), providing a broader host range compared to murine vectors. Since recombinant retroviruses are defective, assistance is required in order to produce infectious particles. Such assistance can be provided, e.g., by using helper cell lines that contain plasmids encoding all of the structural genes of the retrovirus under the control of regulatory sequences within the LTR. Suitable helper cell lines are well known to those skilled in the art. Said vectors can additionally contain a gene encoding a selectable marker so that the transduced cells can be identified. Moreover, the retroviral vectors can be modified in such a way that they become target specific. This can be achieved, e.g., by inserting a polynucleotide encoding a sugar, a glycolipid, or a protein, preferably an antibody. Those skilled in the art know additional methods for generating target specific vectors. Further suitable vectors and methods for in vitro- or in vivo-gene therapy are described in the literature and are known to the persons skilled in the art; see, e.g., WO 94/29469 or WO 97/00957. The HCBI polynucleotide sequences of the invention may also serve as a suitable vector itself, either composed solely of rearranged 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.

[0073] In order to achieve expression only in the target organ, the DNA sequences for transcription of the antisense oligonucleotides can be linked to a tissue specific promoter and used for gene therapy. Such promoters are well known to those skilled in the art.

[0074] Within an oligonucleotide structure, the phosphate groups are commonly referred to as forming the internucleoside backbone of the oligonucleotide. The normal linkage or backbone of RNA and DNA is a 3' to 5' phosphodiester linkage. Specific examples of preferred antisense compounds useful in the present invention include oligonucleotides containing modified backbones or non-natural internucleoside linkages. Oligonucleotides having modified backbones include those that retain a phosphorus atom in the backbone and those that do not have a phosphorus atom in the backbone. Modified oligonucleotide backbones which can result in increased stability are known to the person skilled in the art, preferably such

modification is a phosphorothioate linkage.

[0075] A preferred oligonucleotide mimetic is an oligonucleotide mimetic that has been shown to have excellent hybridization properties, and is referred to as a peptide nucleic acid (PNA). In PNA compounds, the sugar-backbone of an oligonucleotide is replaced with an amide containing backbone, in particular an aminoethylglycine backbone. The nucleobases are retained and are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone.

[0076] Modified oligonucleotides may also contain one or more substituted or modified sugar moieties. Preferred oligonucleotides comprise one of the following at the 2' position: OH; F; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; O-, S- or N-alkynyl; or O-alkyl-O-alkyl, wherein the alkyl, alkenyl and alkynyl may be substituted or unsubstituted C₁ to C₁₀ alkyl or C₂ to C₁₀ alkenyl and alkynyl. A particularly preferred modified sugar moiety is a 2'-O-methoxyethyl sugar moiety.

[0077] Antisense-oligonucleotides of the invention may also include nucleobase modifications or substitutions. Modified nucleobases include other synthetic and natural nucleobases such as 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine etc., with 5-methylcytosine substitutions being preferred since these modifications have been shown to increase nucleic acid duplex stability.

[0078] Another modification of the oligonucleotides of the invention involves chemically linking to the oligonucleotide one or more moieties or conjugates which enhance the activity, cellular distribution or cellular uptake of the oligonucleotide. Such moieties include lipid moieties such as a cholesterol moiety, cholic acid, a thioether, a thiocholesterol, an aliphatic chain, e.g., dodecandiol or undecyl residues, a phospholipid, a polyamine or a polyethylene glycol chain, or adamantane acetic acid, a palmityl moiety, or an octadecylamine or hexylamino-carbonyl-oxycholesterol moiety.

[0079] The present invention also includes antisense compounds which are chimeric compounds. "Chimeric" antisense compounds or "chimeras," in the context of this invention, are antisense compounds, particularly oligonucleotides, which contain two or more chemically distinct regions, each made up of at least one monomer unit, i.e., a nucleotide in the case of an oligonucleotide compound. These oligonucleotides typically contain at least one region wherein the oligonucleotide is modified so as to confer upon the oligonucleotide increased resistance to nuclease degradation, increased cellular uptake, and/or increased binding affinity for the target nucleic acid. An additional region of the oligonucleotide may serve as a substrate for enzymes capable of cleaving RNA:DNA or RNA:RNA hybrids. By way of example, RNase H is a cellular endonuclease which cleaves the RNA strand of an RNA:DNA duplex. Activation of RNase H, therefore, results in cleavage of the RNA target, thereby greatly enhancing the efficiency of oligonucleotide inhibition of gene expression. Consequently, comparable results can often be obtained with shorter oligonucleotides when chimeric oligonucleotides are used, compared to

phosphorothioate deoxyoligonucleotides hybridizing to the same target region. Chimeric antisense compounds of the invention may be formed as composite structures of two or more oligonucleotides, modified oligonucleotides, oligonucleosides and/or oligonucleotide mimetics as described above. Such compounds have also been referred to in the art as hybrids or gapmers.

[0080] 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 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 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 to the subject at an effective dose.

[0081] 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 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.

[0082] Administration of the suitable compositions may be effected by different ways, e.g. by intravenous, intraperitoneal, oral, subcutaneous, intramuscular, topical or intradermal administration. The route of administration, of course, depends on the kind of therapy and the kind of compound contained in the pharmaceutical composition. The dosage regimen will be determined by the attending physician and other clinical factors. As is well known in the medical arts, dosages for any one patient depends on many factors, including the patient's size, body surface area, age, sex, the particular compound to be administered, time and route of administration, the kind of therapy, general health and other drugs being administered concurrently.

[0083] In a preferred embodiment of the present invention, the disease that can be prevented/treated is cancer, preferably colorectal, 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.

[0084] 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.

[0085] 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.

[0086] 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.

[0087] 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 detectably immunoreact in its intended system (e. g., immunoassay). Preferably, the effective amount is sufficient to effect treatment, as defined above. The exact amount necessary will vary according to the application. For vaccine applications or for the generation of polyclonal antiserum/antibodies, for example, the effective amount may vary depending on the species, age, and general condition of the individual, the severity of the condition being treated, the particular polypeptide selected and its mode of administration, etc. Effective amounts will be found within a relatively large, non-critical range. An appropriate effective amount can be readily determined using 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 subsequent protection against an HCBI infection and an HCBI linked disease, respectively.

[0088] Pharmaceutically acceptable carriers include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the vaccine. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, and amino acid copolymers. Such carriers are well known to those of ordinary skill in the art.

[0089] Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: aluminium hydroxide (alum), N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP) as found in U.S. Patent No. 4,606,918, N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-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 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 Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA) may be used for non-human applications and research purposes.

[0090] The immunogenic compositions typically will contain pharmaceutically acceptable vehicles, such as water, saline, 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.

[0091] Typically, the immunogenic compositions are prepared as 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 incorporated into Immune Stimulating Complexes together with saponins, for example Quil A (ISCOMS).

[0092] Immunogenic compositions used as vaccines comprise a "sufficient amount" or "an immunologically effective amount" of the proteins 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 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. 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.

Example 1

Material and Methods

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

[0093] Bovine sera were separated on Optiprep (iodixanol) step gradients as previously described (Buck et al., 2004). 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.

(B) In vitro replication (de Villiers et al., 2011) of HCBI1.225 (Comparative Example)

[0094] 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 Dpn1 digestion. DNA and RNA of HCBI1.225 were demonstrated up to day 17 in cell culture.

Example 2

Characterization of the HCBI sequences

[0095] 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 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) 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 described by Manuelidis but are not identical the designation HCBI (Healthy Cattle Blood Isolate) is proposed.

[0096] In this context, diseases of the CNS (e.g., Multiple Sclerosis, Alzheimer's disease) are also highly interesting since the similar sequences described by Manuelidis are primarily found in the CNS.

[0097] The "Sphinx" sequences (Manuelidis, 2011) show high homologies to plasmid sequences of the bacterium *Acinetobacter* (Vallenet 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.

[0098] 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 1woffii*, *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.

References

[0099]

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 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 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 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., Ebringer, A., Cunningham, P., Chamoun, V., Thompson, E.J., Croker, J., and Vowles, J. Antibody responses to *Acinetobacter* spp. and *Pseudomonas aeruginosa* in multiple sclerosis: prospects for diagnosis using the myelin-*Acinetobacter*-neurofilament antibody index. *Clin. Diagn. Laboratory Immunol.* 2001; 8: 1181-1188.

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

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

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

Vallenet D, Nordmann P, Barbe V, 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 C, Weissenbach J, Cruveiller S. Comparative analysis of Acinetobacters: three genomes for three lifestyles. PLoS One 2008; 3(3):e1805-e1805.

Xu B, Zhi N, Hu G, Wan Z, Zheng X, Liu X, Wong S, Kajigaya S, Zhao K, Mao Q, Young NS. Hybrid DNA virus in Chinese patients with seronegative hepatitis discovered by deep sequencing. Proc Natl Acad Sci USA. 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 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.

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: e00849-14. doi: 10.1128/genomeA.00849-14.

SEQUENCE LISTING

[0100]

<110> Deutsches Krebsforschungszentrum

<120> HCBI sequences

<130> K 3506EP

<160> 83

<170> BiSSAP 1.0

<210> 1

<211> 2251

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..2251

<223> /mol_type="DNA" /note="HCBI 1.225" /organism="artificial sequences"

<400> 1

```

ccttttccct gattacaagc tcccatctgc tccagatcgg gagcttttag ccatcccgaa      60
tggggtggcga cacaaattac aggcgaaaaa aaagaccata ttggccttcg gtttagattg      120
gctgccgtat cagccaattt gattgagtgc cgcactcactc aatctatcaa gtgccgcctc      180
acctgattac tcaaactctac atgaaatttt ctttcttgca aacggttcag aaaagcatta      240
taaatatccc accaatgaat cataaatgaa tcactataga tcaggaaaca actatgaaac      300
tccataatcc aaatccaaat gagcctacta acctacaaat gcttggtgca gaagttaaaa      360
aatccgcttc aagctcttat cacgggtggtt atattcaagt tcctttccgt gttgagtttg      420
catcatatac acgccttgag gcacttggtt aacatactgg ctcaagtcgc aataagatca      480
tgaatgatct gttaagaatc ggtattgaaa ctctagctgc ctctttggat gatgaaacaa      540
ttaaaactct ttttgaaatt gaaacctcaa tcactgcgga tctctatgct tcaggaaaaa      600
ttaaatcagg ggatcaatca gatgattaat ttacaaggaa ctctaattaa tgcttttcgt      660
gtggatggtg gtaaaggtaa agacggcaaa gaatatgaag cgcgtgacaa ggtgcaaatt      720
cttggttcgt tggagttacc caatggggaa attaaacacg aacttggtga cctaacagtt      780
gaggattctc gactttttga accattcaaa aatcagggtta ttagcatctc atgtggtgct      840
atggctgtcg gtcgtaatgt tattttttat gttcgcaaag gcgcaaaacc agtttttagca      900
gaacaaatgt gaaacccgaa aaaaaaaca gtgagtttagc gaagcaactc gatactattg      960
attaactagg ggacaaaaat tgactataaa agacaataaa aaagcccata attcggatgc     1020
tttggcgggc gatgaattac aggccttgca atctgctaag gcagatcaac acagggatag     1080
aataacacgt tttggaattt tgaaacatag agcgaagcta caagaacaat atttgtggac     1140
gcaagttgat tttaaatccg aagggtgaaaa tgatctatct aataaggctc ttaaggctgc     1200
aaccaaaactc aagggatgcg gtcaatttct ctatttccat aattactaca caatagacca     1260
agttaaactt gctaaggcct attactgttc tcagcatttg ctatgtccta tgtgtgctgg     1320
tgtaagggct gctaagtcaa tgagtcgtta tattcagcgt attgaagaac taatgcgcca     1380
gaatcgccat ttaaagcccg tcctgatcac tttgacggta aagaatggcc ctgacctaca     1440
agaacgcttc aaacacctta gagcatcatt tagaacgctt ttagatcggt ataacgatta     1500
caagaaaaaa ggtcgtgggt ttaatcagtt ttgtaaaatt gatggtgctt tttattcaac     1560
tgaatataca tacaatccaa aaactaaaga atggcatccg catattcata tttttgcctt     1620
actcaatcaa tggatagacc aggaagaatt gtccgaaact tggcacgata ttactctgga     1680
ttcttatatc gtcgatattc gtagagttaa aaaaacccaa gaacacggct atagcaaagc     1740

```

```

tgttgctgag gtatgtaaat acgcccttaa gtttagtgat ctctcactgg agaacacttg      1800
ggaggcatat ctttctttga aaggtaacag gcttactggc tgttttggtt ctatgtatgg      1860
tgtcaaattg cctgaaaaac ttactgatga tctacccctt gatgatcttc catatttgga      1920
gctgttatac cgttttggtt ttggtaaaaa atcttattac aacctagaaa taacaaaaga      1980
cgtaaagcca caaaactagg actacaacga tgaggtgagg gcgacgcgcg accgtgcgcg      2040
ctacttcgtg catgtacgcg cgcgcctttg cctctttggt gcgaggtgtg gacgaaaaaa      2100
gcaaggatac atacagaccc ctgcatgacc ttgtaagggg ttcgaccctt tagaccccaa      2160
agggcgcact tatgcaaact cttcgagttc gccagtgtc ccaccagtaa cagagggcgc      2220
ggagtgcgcc cgaactgacg ctatagaatt c                                     2251

```

<210> 2

<211> 232

<212> PRT

<213> artificial sequences

<220>

<221> SOURCE

<222> 1..232

<223> /mol_type="protein" /note="Sphinx2.36 ORF 229 amino acids" /organism="artificial sequences"

<400> 2

```

Met Gln Glu Gln Tyr Leu Trp Thr Gln Val Asp Phe Lys Val Gly Ser
1          5          10          15
Glu Thr Ser Ile Lys Ala Leu Lys Ala Ala Thr Lys Leu Lys Gly Cys
          20          25          30
Gly Gln Phe Leu Leu Phe Arg Asn Tyr Tyr Thr Ile Asp Gln Ile Lys
          35          40          45
Leu Glu Lys Phe His Val Cys Gly Gln His Leu Leu Cys Pro Met Cys
          50          55          60
Ala Gly Ile Arg Ala Ala Arg Ser Met Asn Arg Tyr Ile Gln Arg Ile
65          70          75          80
Glu Glu Ile Met Arg Gln Asn Arg Lys Leu Lys Pro Val Leu Ile Thr

          85          90          95
Leu Thr Val Lys Asn Gly Glu Asp Leu Gln Glu Arg Phe Glu His Leu
          100          105          110
Thr Gly Ser Phe Lys Thr Leu Leu Gln Arg Tyr Arg Asp Phe Lys Lys
          115          120          125
Lys Gly Arg Gly Phe Asn Gln Phe Cys Lys Ile Asp Gly Gly Phe Tyr
          130          135          140
Thr Thr Glu Tyr Thr Tyr Asn Glu Thr Thr Gln Gln Trp His Pro His
145          150          155          160
Ile His Ile Phe Ala Leu Val Thr Asp Arg Ile Asp Gln Glu Glu Leu
          165          170          175
Ala Glu Thr Trp His Asp Ile Thr Leu Asp Ser Tyr Ile Val Asp Ile
          180          185          190
Arg Arg Val Lys Lys Thr Lys Glu His Gly Tyr Ala Lys Ala Val Ala
          195          200          205
Glu Val Cys Lys Tyr Ala Leu Lys Phe Ser Asp Leu Ser Thr Glu Lys
210          215          220
Thr Phe Gln Ala Phe Phe Asp Pro
225          230

```

<210> 3

<211> 229

<212> PRT

<213> artificial sequences

<220>

<221> SOURCE

<222> 1..229

<223> /mol_type="protein" /note="HCBI ORF 229aa" /organism="artificial sequences"

<400> 3

```

Met Cys Ala Gly Val Arg Ala Ala Lys Ser Met Ser Arg Tyr Ile Gln
1      5      10      15
Arg Ile Glu Glu Leu Met Arg Gln Asn Arg His Leu Lys Pro Val Leu
20     25     30
Ile Thr Leu Thr Val Lys Asn Gly Pro Asp Leu Gln Glu Arg Phe Lys
35     40     45
His Leu Arg Ala Ser Phe Arg Thr Leu Leu Asp Arg Tyr Asn Asp Tyr
50     55     60
Lys Lys Lys Gly Arg Gly Phe Asn Gln Phe Cys Lys Ile Asp Gly Ala
65     70     75     80
Phe Tyr Ser Thr Glu Tyr Thr Tyr Asn Pro Lys Thr Lys Glu Trp His
85     90     95
Pro His Ile His Ile Phe Ala Leu Leu Asn Gln Trp Ile Asp Gln Glu
100    105    110
Glu Leu Ser Glu Thr Trp His Asp Ile Thr Leu Asp Ser Tyr Ile Val
115    120    125
Asp Ile Arg Arg Val Lys Lys Thr Lys Glu His Gly Tyr Ser Lys Ala
130    135    140
Val Ala Glu Val Cys Lys Tyr Ala Leu Lys Phe Ser Asp Leu Ser Leu
145    150    155    160
Glu Asn Thr Trp Glu Ala Tyr Leu Ser Leu Lys Gly Asn Arg Leu Thr
165    170    175
Gly Cys Phe Gly Ser Met Tyr Gly Val Lys Leu Pro Glu Lys Leu Thr
180    185    190
Asp Asp Leu Pro Leu Asp Asp Leu Pro Tyr Leu Glu Leu Leu Tyr Arg
195    200    205
Phe Val Phe Gly Lys Lys Ser Tyr Tyr Asn Leu Glu Ile Thr Lys Asp
210    215    220
Val Lys Pro Gln Asn
225

```

<210> 4

<211> 96

<212> PRT

<213> artificial sequences

<220>

<221> SOURCE

<222> 1..96

<223> /mol_type="protein" /note="Sphinx 96aa" /organism="artificial sequences"

<400> 4

```

Met Ile Asn Leu Gln Gly Thr Leu Ile Asn Ala Phe Arg Leu Asp Gly
1      5      10      15
Gly Lys Gly Lys Asp Gly Lys Glu Tyr Glu Ala Arg Asp Lys Val Gln
20     25     30
Ile Leu Gly Ser Leu Glu Leu Pro Asn Gly Glu Ile Lys His Glu Leu
35     40     45
Val Asp Leu Thr Val Glu Asp Ala Arg Ile Phe Glu Pro Phe Lys His
50     55     60
Lys Val Ile Ser Ile Ser Cys Gly Ala Met Ala Ile Gly Arg Asn Val
65     70     75     80
Val Phe Tyr Val Arg Lys Gly Ala Lys Pro Val Leu Ala Asp Val Met
      OF      ON      OF

```

<210> 5

<211> 127

<212> PRT

<213> artificial sequences

<220>

<221> SOURCE

<222> 1..127

<223> /mol_type="protein" /note="HCBI 96aa" /organism="artificial sequences"

<400> 5

```

Met Met Lys Gln Leu Lys Leu Phe Leu Lys Leu Lys Pro Gln Ser Leu
1      5      10      15
Arg Ile Ser Met Leu Gln Glu Lys Leu Asn Gln Gly Ile Asn Gln Met
20     25     30
Ile Asn Leu Gln Gly Thr Leu Ile Asn Ala Phe Arg Val Asp Gly Gly
35     40     45
Lys Gly Lys Asp Gly Lys Glu Tyr Glu Ala Arg Asp Lys Val Gln Ile
50     55     60
Leu Gly Ser Leu Glu Leu Pro Asn Gly Glu Ile Lys His Glu Leu Val
65     70     75     80
Asp Leu Thr Val Glu Asp Ser Arg Leu Phe Glu Pro Phe Lys Asn Gln
85     90     95
Val Ile Ser Ile Ser Cys Gly Ala Met Ala Val Gly Arg Asn Val Ile
100    105    110
Phe Tyr Val Arg Lys Gly Ala Lys Pro Val Leu Ala Glu Gln Met
115    120    125

```

<210> 6

<211> 124

<212> PRT

<213> artificial sequences

<220>

<221> SOURCE

<222> 1..124

<223> /mol_type="protein" /note="Sphinx 111aa" /organism="artificial sequences"

<400> 6

```

Met Asn His Lys Leu Ile Ala Ile Asp Gln Glu Leu Thr Met Lys Leu
1      5      10      15
His Asn Pro Asn Pro Asn Glu Pro Thr Asn Leu Gln Met Leu Val Ala
20     25     30
Glu Ile Lys Lys Ser Ala Ser Ser Ser Tyr His Gly Gly Tyr Ile Gln
35     40     45
Val Pro Phe Arg Val Glu Cys Ala Ser Tyr Thr Arg Leu Glu Ala Leu
50     55     60
Val Lys His Thr Gly Ser Ser Arg Asn Lys Ile Met Asn Asp Leu Leu
65     70     75     80
Arg Ile Gly Ile Glu Thr Leu Ala Ala Ser Leu Asp Asp Glu Thr Ile
85     90     95
Lys Thr Leu Phe Glu Ile Glu Thr Ser Ile Thr Ala Asp Leu Tyr Ala
100    105    110
Ser Gly Lys Met Lys Ser Gly Asp Gln Ser Asp Asp
115    120

```

<210> 7

<211> 111

<212> PRT

<213> artificial sequences

<220>

<221> SOURCE

<222> 1..111

<223> /mol_type="protein" /note="HCBI 111aa" /organism="artificial sequences"

<400> 7

Met	Lys	Leu	His	Asn	Pro	Asn	Pro	Asn	Glu	Pro	Thr	Asn	Leu	Gln	Met
1				5					10					15	
Leu	Val	Ala	Glu	Val	Lys	Lys	Ser	Ala	Ser	Ser	Ser	Tyr	His	Gly	Gly
			20					25					30		
Tyr	Ile	Gln	Val	Pro	Phe	Arg	Val	Glu	Phe	Ala	Ser	Tyr	Thr	Arg	Leu
		35				40					45				
Glu	Ala	Leu	Val	Lys	His	Thr	Gly	Ser	Ser	Arg	Asn	Lys	Ile	Met	Asn
	50					55				60					
Asp	Leu	Leu	Arg	Ile	Gly	Ile	Glu	Thr	Leu	Ala	Ala	Ser	Leu	Asp	Asp
65					70				75					80	
Glu	Thr	Ile	Lys	Thr	Leu	Phe	Glu	Ile	Glu	Thr	Ser	Ile	Thr	Ala	Asp
			85				90						95		
Leu	Tyr	Ala	Ser	Gly	Lys	Ile	Lys	Ser	Gly	Asp	Gln	Ser	Asp	Asp	
			100				105						110		

<210> 8

<211> 898

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..898

<223> /mol_type="DNA" /note="HQ444405" /organism="artificial sequences"

<400> 8

ataatttgcg gtcactgtgc cgaatcacta cgcaatat	ttt tattacccat aattttaacca	60
attcttaatt tcttgcaaac gggttcaaaaa agcattataa	atatctcacc aatgaatcat	120
aaactaatcg ctatagatca ggaattaact atgaaactcc	ataatccaaa tccaaatgag	180
cctactaacc tgcaaagtgt tgttgacagaa attaaaaaat	ccgcttcaag ctcttatcac	240
gggtggctata ttcaagttcc tttccgtgtt gagtgtgcat	catatacacg ccttgaagca	300
cttggttaaac atactggctc aagtcgcaat aagatcatga	atgatctctt aagaatcggc	360
attgaaactc tagctgcctc tttggacgac gaaacaatta	aaactctttt tgaaattgaa	420
acttcaatca ctgcggatct ctatgcttca ggaaaaatga	aatcaggaga tcagtcagat	480
gattaattta caaggacac ttataaatgc ttttcgtctt	gatggtggtg aagggaaaga	540
cggcaaagaa tatgaagcac gtgacaaggt gcaaattctt	gggttcgctgg agttgcccaa	600
cggtgagatc aaacatgagc ttgttgactt aacggttgaa	gatgctcgca tctttgagcc	660
attcaaacac aaggtaatta gcatttcacg cggtgctatg	gctatcggtc gaaatgttgt	720
tttttatgtt cgaaaagggtg caaagcctgt ttttagcagat	gtaatgtaat tcatgaaaaa	780
aaacagtgag ttaggcttgc cgactcgctg ttttttcttt	tacttgatac tattaactaa	840

agtggggaca aaattgacta aaaaagacaa taaaaaagcc cataattctg atgctttg 898

<210> 9

<211> 958

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..958

<223> /mol_type="DNA" /note="CU459139" /organism="artificial sequences"

<400> 9

aaaactat	ttttcg	ttag	ctaa	gatta	aagaaa	atta	tcat	gccg	at	gtaaaa	acg	60
atgaat	ctat	tgcg	ccat	g	aaaact	gccc	aaaa	attaaa	tggg	tgcg	gt	120
tattca	aaaa	tttt	ta	cacc	atta	at	caaa	ttaa	actc	gc	caag	180
ttccaa	caaa	gctt	gtag	tg	agcatt	tgt	atgt	ccgt	tt	tgtg	ctgg	240
ttag	agct	tc	taag	gca	att	caaaa	ata	ct	ctgag	cgtg	t	300
tgat	caag	tc	ttat	ctg	aaa	atc	ctc	gt	tt	aaag	ccc	360
gt	atgat	cac	gt	ttac	ggt	tta	aaat	ggg	gta	gac	ctag	420
ggg	gact	ata	ttaaaa	aag	gcgt	gg	ct	tt	aat	ga	at	480
ttt	ta	atg	g	g	g	g	g	g	g	g	g	540
g	g	g	g	g	g	g	g	g	g	g	g	600
g	g	g	g	g	g	g	g	g	g	g	g	660
g	g	g	g	g	g	g	g	g	g	g	g	720
g	g	g	g	g	g	g	g	g	g	g	g	780
g	g	g	g	g	g	g	g	g	g	g	g	840
g	g	g	g	g	g	g	g	g	g	g	g	900
g	g	g	g	g	g	g	g	g	g	g	g	958

<210> 10

<211> 959

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..959

<223> /mol_type="DNA" /note="JN872565" /organism="artificial sequences"

<400> 10

aaaatt	at	ttacc	ctg	gcta	ag	ttta	aagaga	acta	tga	aaa	agac	gtta	aaa	acg	60
aaga	at	ctat	caag	gct	cta	aat	ctg	ctc	agaa	att	gaa	tga	atg	cgga	120

tattcaaaaa	tttttacaca	atagggcgaag	ttaaactctc	caagctccgc	acctgcggac	180
agcatttgct	ttgccctttc	tgtgctgcc	ttcgtgcttc	tcgtgctatc	caaaaatacg	240
ttgaacgtat	tgatcaagtc	ctgcaagaaa	atcgcaagct	caagcccgtt	ctaatacgc	300
tcaccgttaa	aaacggctct	gacctagcag	aacgctccga	acatcttatg	aagtccttta	360
gaacgctcct	agagcgctcg	agggactatg	aaaagaaagg	tcgaggtttt	aatgagttct	420
gtaaggttca	aggggctatg	tactcctatg	aaaatacatt	caatgaaaaa	acaggcgaat	480
ggcatccgca	tattcatatg	ttcgctttgg	ttgatcaatg	gattgatcag	caagagtttt	540
cagaatattg	gcatagcctt	actggggact	cgatggttgt	cgatgtccgc	agggcaagaa	600
aagaaaaagg	ttacggttat	agcaaagcgg	ctgccgaagt	ctgtaagtat	gctttgaagt	660
ttggtgatct	gtccgttgaa	aagacttggg	aagcatttaa	ggttcttaag	ggaaagcgtt	720
taacaggttc	ttttggtctg	ctatgggggtg	tcaaaatccc	tgactcaatg	acagacgatt	780
gccatcagaa	gacttgccat	atctcgaaat	gctgtacaag	tttgccctaca	gtaaaaagtc	840
ttactacgac	ttactaatca	caaggcacgt	agagccacaa	ccgcatgagg	acgagcgtgc	900
gaggagcttc	gacgagtgcg	attgtattta	tgcggtggag	gctcagacgt	ttgactgtg	959

<210> 11

<211> 690

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..690

<223> /mol_type="DNA" /note="Blast N2 Query1" /organism="artificial sequences"

<400> 11

tttcttgcaa	acggttcaga	aaagcattat	aaatatccca	ccaatgaatc	ataaatgaat	60
cactatagat	caggaaacaa	ctatgaaact	ccataatcca	aatccaaatg	agcctactaa	120
cctacaaatg	cttgttgcag	aagttaaaaa	atccgcttca	agctcttatc	acggtggtta	180
tattcaagtt	cctttccgtg	ttgagtttgc	atcatataca	cgccttgagg	cacttggtta	240
acatactggc	tcaagtcgca	ataagatcat	gaatgatctg	ttaagaatcg	gtattgaaac	300
tctagctgcc	tctttggatg	atgaaacaat	taaaactctt	tttgaaattg	aaacctcaat	360
cactgcggat	ctctatgctt	caggaaaaat	taaatcaggg	gatcaatcag	atgattaatt	420
tacaaggaac	tctaattaat	gcttttcgtg	tggatggtgg	taaaggtaaa	gacggcaaag	480
aatatgaagc	gcgtgacaag	gtgcaaattc	ttggttcgtt	ggagttaccc	aatggggaat	540
tcgcaaaggc	gcaaaaccag	ttttagcaga	atcaggttat	tagcatctca	tgtggtgcta	600
tggctgtcgg	tcgtaatgtt	attttttatg	ttaaacacga	acttgttgac	ctaacagttg	660
aggattctcg	actttttgaa	ccattcaaaa				690

<210> 12

<211> 692

<212> DNA
<213> artificial sequences

<220>
<221> source
<222> 1..692
<223> /mol_type="DNA" /note="BlastN2 Query1" /organism="artificial sequences"

<400> 12
tttcttgcaa acggttcaaa aaagcattat aaatatctca ccaatgaatc ataaactaat 60
cgctatagat caggaattaa ctatgaaact ccataatcca aatccaaatg agcctactaa 120
cctgcaaatg cttgttgacg aaattaaaaa atccgcttca agctcttatac acggtggcta 180
tattcaagtt cctttccgtg ttgagtgtgc atcatataca cgccttgaag cacttggttaa 240
acatactggc tcaagtcgca ataagatcat gaatgatctc ttaagaatcg gtattgaaac 300
tctagctgcc tctttggacg acgaaacaat taaaactctt tttgaaattg aaacttcaat 360
cactgcggat ctctatgctt caggaaaaat gaaatcagga gatcagtcag atgattaatt 420

tacaagggac acttataaat gcttttcgtc ttgatgggtg taaagggaaa gacggcaaag 480
aatatgaagc acgtgacaag gtgcaaattc ttggttcgct ggagttgccc aacggtgaga 540
ttcgaaaagg tgcaaagcct gtttttagcag astcaaacat gagcttgttg acttaacggt 600
tgaagatgct cgcactcttg agccattcaa acacaaggta attagcattt catgcggtgc 660
tatggctatc ggtcgaaatg ttgtttttta tg 692

<210> 13
<211> 132
<212> DNA
<213> artificial sequences

<220>
<221> source
<222> 1..132
<223> /mol_type="DNA" /note="BlastN2 Query2" /organism="artificial sequences"

<400> 13
aaaattgact ataaaagaca ataaaaaagc ccataattcg gatgcttttg cgggcgatga 60
attacaggct ttgcaatctg ctaaggcaga tcaacacagg gatagaataa cacgttttgg 120
aattttgaaa ca 132

<210> 14
<211> 132
<212> DNA
<213> artificial sequences

<220>
<221> source

<222> 1..132

<223> /mol_type="DNA" /note="BlastN2 Subject2" /organism="artificial sequences"

<400> 14

actttttgaaa caattacagg ctttgcaatc tgctaattgca gatcaacaca gggatagaat	60
atcacgtttt ggaaaattga ctaaaaaaga caataaaaaa gcccataatt ctgatgcttt	120
ggcgggac ga	132

<210> 15

<211> 380

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..380

<223> /mol_type="DNA" /note="lastN2 Query3" /organism="artificial sequences"

<400> 15

cagcatttgc tatgtcctat gtgtgctggt gtaagggctg ctaagtcaat gagtcgttat	60
attcagcgta ttgaagaact aatgcgccag aatcgccatt taaagcccgt cctgatcact	120
ttgacggtaa agaatggccc tgacctaca gaacgcttca aacaccttag agcatcattt	180
agaacgcttt tagatcgtaa taacgattac aagaaaaaag gtcgtggttt taatcagttt	240
tgtaaaattg atgggtgcttt ttattcaact gaatatacat acaatccaaa aactaaagaa	300
tggcatccgc atattcatat ttttgcctta ctcaatcaat ggatagacca ggaagaattg	360
tccgaaactt ggcacgatat	380

<210> 16

<211> 380

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..380

<223> /mol_type="DNA" /note="BlastN2 Subject3" /organism="artificial sequences"

<400> 16

cagcatttgc tatgtccaat gtgtgctggt attcgtgctg cccgttcaat gaatcggtat	60
attcaacgca tcgaagaaat aatgcgtcag aatcgcaagc taaagcccgt attgatcact	120
ttgaccgtaa agaacggtga agacctacag gaacgctttg aacacctcac aggctcattt	180
aagacgcttt tacagcgtaa ccgtgatttt aagaaaaagg gtcgagggtt taatcaattt	240
tgcaaaattg atggcggttt ttatacgacc gaatacacct acaacgaaac aaccaacaa	300
tggcatccgc atattcatat ttttgcgtaa gtgactgacc ggattgacca ggaggaaacta	360
acagaaactt accacgatat	380

```

-----
<210> 17
<211> 43
<212> DNA
<213> artificial sequences

<220>
<221> source
<222> 1..43
<223> /mol_type="DNA" /note="BlastN2 Query4" /organism="artificial sequences"

<400> 17
cttctaaggg gttcgacccc ttagacccca aagggcgcac tta      43

<210> 18
<211> 43
<212> DNA
<213> artificial sequences

<220>
<221> source
<222> 1..43
<223> /mol_type="DNA" /note="BlastN2 Subject4" /organism="artificial sequences"

<400> 18
cttctaaggg ggtcggcccc ttagacccca aagggcgcac tta      43

<210> 19
<211> 83
<212> DNA
<213> artificial sequences

<220>
<221> source
<222> 1..83
<223> /mol_type="DNA" /note="BlastN2 Query5" /organism="artificial sequences"

<400> 19
cttccatatt tggagctggt ataccgtttt gtttttggtt aaaaatctta ttacaacctt      60
gaaataacaa aagacgtaaa gcc                                83

<210> 20
<211> 83
<212> DNA
<213> artificial sequences

<220>
<221> source

```

<222> 1..83

<223> /mol_type="DNA" /note="BlastN2 Subject5" /organism="artificial sequences"

<400> 20

cttccatatac ttgagctgct ctatcgtttc gtttttggtg aaaggtctta ttacaaccta 60

gagttaacta aggacgtaaa gcc 83

<210> 21

<211> 269

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..269

<223> /mol_type="DNA" /note="BlastN2 Query6" /organism="artificial sequences"

<400> 21

atgattaatt tacaaggaac tctaattaat gcttttcgtg tggatggtgg taaaggtaaa 60

gacggcaaag aatatgaagc gcgtgacaag gtgcaaattc ttggttcggt ggagttaccc 120

atTTTTtatg ttcgcaaagg cgcaaaaccc cattcaaaaa tcaggttatt agcatctcat 180

gtggtgctat ggctgtcggc cgtaatgtta atggggaaat taaacacgaa cttgttgacc 240

taacagttga ggattctcga ctttttgaa 269

<210> 22

<211> 269

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..269

<223> /mol_type="DNA" /note="BlastN2 Subject6" /organism="artificial sequences"

<400> 22

atgattaatt tacaaggaac tctaataat gctttccgta tggacggtgg taagggtaaa 60

gttttttatg ttcgaaaagg tgcaaaacca atggagagat taaacacgaa cttgttgacc 120

tcacagttga tgatgccagt gtctaccagc cactaaaaaa taaagtaatt tctatttcct 180

gcggtgctat ggctgtaggt cgtaatgttg atgggaaaga gtacgaagcg cgtgacaagg 240

tacaaatact tggttcgctg gaactaccc 269

<210> 23

<211> 160

<212> DNA

<213> artificial sequences

<220>
<221> source
<222> 1..160
<223> /mol_type="DNA" /note="BlastN2 Query7" /organism="artificial sequences"

<400> 23
acagggatag aataacacgt tttggaattt tgaacatag attcggatgc tttggcgggc 60
gatgaattac aggctttgca atctgctaag gcagatcaac gatactattg attaactagg 120
ggacaaaaat tgactataaa agacaataaa aaagcccata 160

<210> 24
<211> 160
<212> DNA
<213> artificial sequences

<220>
<221> source
<222> 1..160
<223> /mol_type="DNA" /note="BlastN2 Subject7" /organism="artificial sequences"

<400> 24
acagggatag aataacacgt tttgcgactt tgaacatag tttcagaagc tttggcgggc 60
ggagaaatac aggctttgca atctgcaaac gcagatcaac gatactattg attaaagtgg 120
ggacaaaaat tgcttaaaaa agacaataaa aaagcccata 160

<210> 25
<211> 23
<212> DNA
<213> artificial sequences

<220>
<221> source
<222> 1..23
<223> /mol_type="DNA" /note="BlastN2 Query8" /organism="artificial sequences"

<400> 25
gaatggcatc cgcatattca tat 23

<210> 26
<211> 23
<212> DNA
<213> artificial sequences

<220>
<221> source
<222> 1..23
<223> /mol_type="DNA" /note="BlastN2 Subject8" /organism="artificial sequences"

<400> 26

gaatggcatc cgcatattca tat 23

<210> 27

<211> 98

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..98

<223> /mol_type="DNA" /note="BlastN2 Query9" /organism="artificial sequences"

<400> 27

ttcaaaaatc aggttattag catctcatgt ggtgctatgg ctgtcggtcg taatgttatt 60

ttttatgttc gcaaaggcgc aaaaccagtt ttagcaga 98

<210> 28

<211> 98

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..98

<223> /mol_type="DNA" /note="BlastN2 Subject9" /organism="artificial sequences"

<400> 28

ttcaaaaata agcttattag catcagttgt ggtgctatgg ctggtggtcg taacgttatt 60

ttttatgttc gaaaagggtgc gaaacctgtt ttagcaga 98

<210> 29

<211> 64

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..64

<223> /mol_type="DNA" /note="BlastN2 Query10" /organism="artificial sequences"

<400> 29

agaaagatca acacagggat agaataacac gttttggaat tttgaaacat agagcgaagc 60

taca 64

<210> 30

<211> 64

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..64

<223> /mol_type="DNA" /note="BlastN2 Subject10" /organism="artificial sequences"

<400> 30

agaaagatca acatagggat agaataacac gttttggcat tttgaaacat agatcgaagc 60

aaca 64

<210> 31

<211> 113

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..113

<223> /mol_type="DNA" /note="BlastN2 Query11" /organism="artificial sequences"

<400> 31

tatacataca atccaaaaac taaagaatgg catccgcata ttcatatattt tgcaaaaaag 60

gtcgtgggtt taatcagttt tgtaaaattg atgggtgctt ttattcaact gaa 113

<210> 32

<211> 113

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..113

<223> /mol_type="DNA" /note="BlastN2 Subject11" /organism="artificial sequences"

<400> 32

aatacttaca atgaaaaaac taatgaatgg catcctcata ttcatatggt tgcaaaaaag 60

ggcgtggctt taatgaattt tgcaaaatta atgggtgcgat gtattcatat gag 113

<210> 33

<211> 1707

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..1707

<223> /mol_type="DNA" /note="HCBI2.170" /organism="artificial sequences"

<400> 33

gaattcaata tgttcgcttc taaccattac aacgacttta ctggttctgg agatttatca	60
gatgattaaa ttagaaggaa tcgtttttaa cgtcttcact cagcaagggtg gacaaaacaa	120
aaaaggcgaa tcatttgatg atcgtgacaa ggtacaaatt ttaggtgcta tggatctgcc	180
caatggtgat gtaaaaaatg agctttttac gttatctgta gatgattatc gggactttaa	240
agacttccta aatcgaaaaa tttgtattgc tgttggtgca atggcaagtg gccgtaatgt	300
tattttttat gttgctaaag gtgcaaagcc tatatcagca gaatttgctg gaaacgctaa	360
aaaaaacagt gagttagcct tggcttctcg ctgttttttc tttctcttga tactattgat	420
taactagggg acaaaaattg actaaaaaag acaataaaaa agcccataat tcggacactt	480
tggcgggtga tgaattacag gctttgcaat ctgctaaggc agatcaacac agggatagaa	540
tatcacgttt tggacttttg aaacatagag cgaagctaca agaacaatat ttgtggacgc	600
aggttgattt caaatccgaa ggtgaaaatg agacatccaa taaggctctt aaggctgcaa	660
ccaaattaaa gggatgcggt caatttttgc tattccataa ctactacaca attgaccaag	720
ttaaacttgc taaggcccat tattgttctc agcatttgct ttgccctatg tgtgctggtg	780
taagggtgct taagtcaatg agtagatatg ttcaacgtat tgaagaattg atgcgtcaga	840
atcgcaaatt aaagcccgta ttgatcactt taacggttaa gaatggggaa gacctagaag	900
aacgctttta acacttagac gctcatttag gacgctttta gatcggtata acgattacaa	960
aaagaaaggc cgtggtttta atcaattctg caagattgat ggtgcttttt tattccactg	1020
aatataccta caattcaaaa acaaaagagt ggcaccccca tatccatatt ttcgctttac	1080
tcaatgaatg gatagaccag gaagaattgg ccgagacctg gcatgacatt accctggatt	1140
cttatatcgt agatatcgt agagttaaaa ggaccaaaga acacggctat agcaaagctg	1200
ttgcagaggt ttgtaaatat gctcttaagt ttagtgattt gtcacttgag aatacgtggg	1260
aagcttatct ttctttaaaa ggtaataggc ttactggctg ttttggttct atgtatggtg	1320
tcaagttgcc tgaaaaactc acagatgatt tacccttga tgatcttcca tatatggagc	1380
tgctataccg ttttgtcttt ggtaaaaaat aacattacgg ccacttgcca ttgcaccaac	1440
agcaatacaa atttttcgat ttaggaagtc tttaaagtcc cgataatcat ctacagataa	1500
cgtaaaaagc tcatttttta catcaccatt gggcagatcc atagcaccta aaatttgtac	1560
cttgtcacga tcatcaaatg attcgccttt tttgttttgt ccaccttgct gagtgaagac	1620
gtttaaaacg attccttcta atttaatcat ctgataaatc tccagaacca gtaaagtcgt	1680
tgtaatggtt agaagcgaac atatgaa	1707

<210> 34

<211> 96

<212> PRT

<213> artificial sequences

<220>

<221> SOURCE

<222> 1..96

<223> /mol_type="protein" /note="Sphinx 2.36.96" /organism="artificial sequences"

<400> 34

```

Met Ile Asn Leu Gln Gly Thr Leu Ile Asn Ala Phe Arg Leu Asp Gly
1           5           10           15
Gly Lys Gly Lys Asp Gly Lys Glu Tyr Glu Ala Arg Asp Lys Val Gln
20           25           30
Ile Leu Gly Ser Leu Glu Leu Pro Asn Gly Glu Ile Lys His Glu Leu
35           40           45
Val Asp Leu Thr Val Glu Asp Ala Arg Ile Phe Glu Pro Phe Lys His
50           55           60
Lys Val Ile Ser Ile Ser Cys Gly Ala Met Ala Ile Gly Arg Asn Val
65           70           75           80
Val Phe Tyr Val Arg Lys Gly Ala Lys Pro Val Leu Ala Asp Val Met
85           90           95

```

<210> 35

<211> 96

<212> PRT

<213> artificial sequences

<220>

<221> SOURCE

<222> 1..96

<223> /mol_type="protein" /note="HCB12.170.137" /organism="artificial sequences"

<400> 35

```

Leu Ser Val Asp Asp Tyr Arg Asp Phe Lys Asp Phe Leu Asn Arg Lys
1           5           10           15
Ile Cys Ile Ala Val Gly Ala Met Ala Ser Gly Arg Asn Val Ile Phe
20           25           30
Tyr Val Ala Lys Gly Ala Lys Pro Ile Ser Ala Glu Phe Ala Met Ile
35           40           45
Lys Leu Glu Gly Ile Val Leu Asn Val Phe Thr Gln Gln Gly Gly Gln
50           55           60
Asn Lys Lys Gly Glu Ser Phe Asp Asp Arg Asp Lys Val Gln Ile Leu
65           70           75           80
Gly Ala Met Asp Leu Pro Asn Gly Asp Val Lys Asn Glu Leu Phe Thr
85           90           95

```

<210> 36

<211> 232

<212> PRT

<213> artificial sequences

<220>

<221> SOURCE

<222> 1..232

<223> /mol_type="protein" /note="Sphinx 2.36.232" /organism="artificial sequences"

<400> 36

```

Met Gln Glu Gln Tyr Leu Trp Thr Gln Val Asp Phe Lys Val Gly Ser
1           5           10           15
Glu Thr Ser Ile Lys Ala Leu Lys Ala Ala Thr Lys Leu Lys Gly Cys
20           25           30
Gly Gln Phe Leu Leu Phe Arg Asn Tyr Tyr Thr Ile Asp Gln Ile Lys
35           40           45

```

```

Leu Glu Lys Phe His Val Cys Gly Gln His Leu Leu Cys Pro Met Cys
 50          55          60
Ala Gly Ile Arg Ala Ala Arg Ser Met Asn Arg Tyr Ile Gln Arg Ile
 65          70          75          80
Glu Glu Ile Met Arg Gln Asn Arg Lys Leu Lys Pro Val Leu Ile Thr
      85          90          95
Leu Thr Val Lys Asn Gly Glu Asp Leu Gln Glu Arg Phe Glu His Leu
      100          105          110
Thr Gly Ser Phe Lys Thr Leu Leu Gln Arg Tyr Arg Asp Phe Lys Lys
      115          120          125
Lys Gly Arg Gly Phe Asn Gln Phe Cys Lys Ile Asp Gly Gly Phe Tyr
      130          135          140
Thr Thr Glu Tyr Thr Tyr His Gly Tyr Ala Lys Ala Val Ala Glu Val
      145          150          155          160
Cys Lys Tyr Ala Leu Lys Phe Ser Asp Leu Ser Thr Glu Lys Thr Phe
      165          170          175
Gln Ala Phe Phe Asp Pro Asn Glu Thr Thr Gln Gln Trp His Pro His
      180          185          190
Ile His Ile Phe Ala Leu Val Thr Asp Arg Ile Asp Gln Glu Glu Leu
      195          200          205

Ala Glu Thr Trp His Asp Ile Thr Leu Asp Ser Tyr Ile Val Asp Ile
      210          215          220
Arg Arg Val Lys Lys Thr Lys Glu
      225          230

```

<210> 37

<211> 137

<212> PRT

<213> artificial sequences

<220>

<221> SOURCE

<222> 1..137

<223> /mol_type="protein" /note="HCB12.170.96.PEP" /organism="artificial sequences"

<400> 37

```

Met Val Leu Phe Tyr Ser Thr Glu Tyr Thr Tyr Asn Ser Lys Thr Lys
 1          5          10          15
Glu Trp His Pro His Ile His Ile Phe Ala Leu Leu Asn Glu Trp Ile
      20          25          30
Asp Gln Glu Glu Leu Ala Glu Thr Trp His Asp Ile Thr Leu Asp Ser
      35          40          45
Tyr Ile Val Asp Ile Arg Arg Val Lys Arg Thr Lys Glu Glu Lys Leu
      50          55          60
Thr Asp Asp Leu Pro Leu Asp Asp Leu Pro Tyr Met Glu Leu Leu Tyr
      65          70          75          80
Arg Phe Val Phe Gly Lys Lys His Gly Tyr Ser Lys Ala Val Ala Glu
      85          90          95
Val Cys Lys Tyr Ala Leu Lys Phe Ser Asp Leu Ser Leu Glu Asn Thr
      100          105          110
Trp Glu Ala Tyr Leu Ser Leu Lys Gly Asn Arg Leu Thr Gly Cys Phe
      115          120          125
Gly Ser Met Tyr Gly Val Lys Leu Pro
      130          135

```

<210> 38

<211> 104

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..104

<223> /mol_type="DNA" /note="BlastN2 Query12" /organism="artificial sequences"

<400> 38

gatcaacaca gggatagaat atcacgtttt ggacttttga aacagcccat aattcggaca 60

ctttggcggg tgatgaatta caggctttgc aatctgctaa ggca 104

<210> 39

<211> 104

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..104

<223> /mol_type="DNA" /note="BlastN2 Subject1212" /organism="artificial sequences"

<400> 39

gatcaacaca gggatagaat atcacgtttt ggacttttga aacagcccat aattctgatg 60

ctttggcggg cgacgaatta caggctttgc aatctgctaa tgca 104

<210> 40

<211> 424

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..424

<223> /mol_type="DNA" /note="lastN2 Query13" /organism="artificial sequences"

<400> 40

caagaacaat atttgtggac gcaggttgat ttcaaattccg aaggtgaaaa tgagacatcc 60

aataaggctc ttaaggctgc aaccaaatta aagggatgcg gtcaattttt gctattccat 120

aactactaca caattgacca agttaaactt gctaaggccc attattgttc tcagcatttg 180

ctttgcccta tgtgtgctgg tgtaagggtc gctaagtcaa tgagtagata tgttcaacgt 240

attgaagaat tgatgcgtca gaatcgcaa ttaaagcccg tattgatcac tttaacggtt 300

gatggttaga tcgttataac gattacaaaa agaaagggtcg tggttttaat caattctgca 360

agattaagaa tggggaagac ctagaagaac gctttaaaca cttagacgct catttaggac 420

gctt 424

<210> 41

<211> 419

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..419

<223> /mol_type="DNA" /note="BlastN2 Subject13" /organism="artificial sequences"

<400> 41

caagaacaat atttgtggac tcaggttgat ttcaagggtg gaagtgaaac atctatcaaa	60
gctcttaagg ctgcaaccaa attaaaggga tgcggtcaat ttctgctttt tcgtaattac	120
tacaccatag atcaaataca gctcgaaaaa ttccacgtat gcggacagca tttgctatgt	180
ccaatgtgtg ctggtattcg tgctgcccgt tcaatgaatc ggtatattca acgcatcgaa	240
gaaataatgc gtcagaatcg caagctaaag cccgtattga tcactttgac cgtaagaac	300
ggtgaagacc tacaggaacg ctttgaacac ctcacaggct catttaagac gcttgatggt	360
tacagcgtta ccgtgatttt aagaaaaagg gtcgagggtt taatcaattt tgcaaaatt	419

<210> 42

<211> 51

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..51

<223> /mol_type="DNA" /note="BlastN2 Query14" /organism="artificial sequences"

<400> 42

cagtgagtta gccttggctt ctcgctgttt ttctttctc ttgatactat t	51
---	----

<210> 43

<211> 51

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..51

<223> /mol_type="DNA" /note="BlastN2 Subject14" /organism="artificial sequences"

<400> 43

cagtgagtta ggcttgccga ctcgctgttt ttcttttac ttgatactat t	51
---	----

<210> 44

<211> 91

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..91

<223> /mol_type="DNA" /note="BlastN2 Query15" /organism="artificial sequences"

<400> 44

gaatatcacg ttttggactt ttgaaacata gctttggcgg gtgatgaatt acaggctttg 60

caatctgcta aggcagatca acacagggat a 91

<210> 45

<211> 91

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..91

<223> /mol_type="DNA" /note="BlastN2 Subject15" /organism="artificial sequences"

<400> 45

gaataacacg ttttgcgact ttgaaacata gctttggcgg ggggagaaat acaggctttg 60

caatctgcaa acgcagatca acacagggat a 91

<210> 46

<211> 33

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..33

<223> /mol_type="DNA" /note="BlastN2 Query16" /organism="artificial sequences"

<400> 46

ctgttttttc ttctcttga tactattgat taa 33

<210> 47

<211> 33

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..33

<223> /mol_type="DNA" /note="BlastN2 Subject16" /organism="artificial sequences"

<400> 47

ctgttttttc ttctacttga tactattgat taa 33

<210> 48
<211> 60
<212> DNA
<213> artificial sequences

<220>
<221> source
<222> 1..60
<223> /mol_type="DNA" /note="BlastN2 Query17" /organism="artificial sequences"

<400> 48
agatcaacac agggatagaa taccacgttt tggacttttg aaacatagag cgaagctaca 60

<210> 49
<211> 60
<212> DNA
<213> artificial sequences

<220>
<221> source
<222> 1..60
<223> /mol_type="DNA" /note="BlastN2 Subject17" /organism="artificial sequences"

<400> 49
agatcaacat agggatagaa taacacgttt tggcattttg aaacatagat cgaagcaaca 60

<210> 50
<211> 49
<212> DNA
<213> artificial sequences

<220>
<221> source
<222> 1..49
<223> /mol_type="DNA" /note="BlastN2 Query18" /organism="artificial sequences"

<400> 50
ttaaagcccg tattgatcac tttaacgggtt aagaatgggg aagacctag 49

<210> 51
<211> 49
<212> DNA
<213> artificial sequences

<220>
<221> source
<222> 1..49

<223> /mol_type="DNA" /note="BlastN2 Subject18" /organism="artificial sequences"

<400> 51

ttaaagcccg ttatgatcac gtttacgggt aaaaatgggg tagacctag 49

<210> 52

<211> 923

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..923

<223> /mol_type="DNA" /note="HQ444405" /organism="artificial sequences"

<400> 52

tagatctaaa catgcaagaa caatatttgt ggactcaggt tgatttcaag gttggaagtg	60
aaacatctat caaagctctt aaggctgcaa ccaaattaaa gggatgcggt caatttctgc	120
tttttcgtaa ttactacacc atagatcaaa tcaagctcga aaaattccac gtatgcggac	180
agcatttgct atgtccaatg tgtgctggta ttcgtgctgc ccgttcaatg aatcggtata	240
ttcaacgcat cgaagaaata atgcgtcaga atcgcaagct aaagcccgtg ttgatcactt	300
tgaccgttaa gaacggtgaa gacctacagg aacgctttga acacctcaca ggctcattta	360
agacgctttt acagcgttac cgtgtaagga cgtaaagccg aacaaaagga atgaggaaag	420
atgaacgagg agcttcgacg aggcacgacc gaatacacct acaacgaaac aaccaacaa	480
tggcatccgc atattcatat ttttgcgtta gtgactgacc ggattgacca ggaggaacta	540
gcagaaactt ggcacgatat aacgcttgat tcatacattg tggacatccg caggggtcaaa	600
aaaactaaag aacacggata tgcaaaggct gttgccgaag tctgcaaata cgctcttaag	660
tttagcgatc tatccactga gaaaaccttt caagcatttt ttgaccctta agggcaaaag	720
gcttacagggt tcattcgggt ctatgcatgg tgtaaaaatt cctgaaagcg gaccgatga	780
aatgcctaaa gaggaacttc catatcttga gctgctctat cgtttcgttt ttggtgaaag	840
gtcttattac aacctagagt taacatttta agaaaaaggg tcgagggttt aatcaatttt	900
gcaaaattga tggcggtttt tat	923

<210> 53

<211> 899

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..899

<223> /mol_type="DNA" /note="CU459139" /organism="artificial sequences"

<400> 53

cgttagctaa gattaaagaa aattatcatg ccgatgtaaa aaacgatgaa tctattcgcg	60
ccatgaaaac tgcccaaaaa ttaaattgggt gcggtaatth tcttctattc aaaaattttt	120
acaccattaa tcaaattaaa ctgcgcaagt tccaagcttg tagtgagcat ttgttatgtc	180
cgttttgtgc tggatttaga gcttctaagg caattcaaaa atactctgag cgtgttgatc	240
aagtcttattc tgaaaatcct cgtttaaaagc ccgttatgat cacgtttacg gttaaaaatg	300
gggtagacct aggggaacgg ttcacccatc ttataaaatc gtttagaacg cttatagagc	360
gtcgtaggga ctatatataa aaagggcgtg gctttaatga attttgcaaa attaatgggtg	420
cgatgtattc atatgagaat acttacaatg aaaaaactaa tgaatggcat cctcatattc	480
atatgtttgc acttttggat gattggatag atcaggatga attgtctcaa tattggcaat	540
ccattactgg ggactctatg gtcgttgata ttcgtagagc caaaaaacaa aaagacttag	600
gctattcagg tgctgctgct gaagtctgta aatatgctct caaatttggt gatctttctg	660
tagaaaagac ttgggaagct ttcaaagttt tgaaaggtaa gcgattaagt ggggcttttg	720
gatctctttg gggcgtgaaa attcctgaat cattgataga tgatcttcca gacgattctt	780
tacaacttac tcgtcatgtc gaacctacag gtaaggacga cgccgacgag cttcgaggag	840
attacctta tttagaaatg atttataagt tcgtcttttc taagaagtct tattacgat	899

<210> 54

<211> 418

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..418

<223> /mol_type="DNA" /note="444405" /organism="artificial sequences"

<400> 54

gcggatctct atgcttcagg aaaaatgaaa tcaggagatc agtcagatga ttaatttaca	60
agggacactt ataaatgctt ttcgtcttga tgggtggtaaa gggaaagacg gcaaagaata	120
tgaagcacgt gacaagggtc aaattcttgg ttcgctggag ttgccaacg gtgagatcaa	180
acatgagctt gttgacttaa cggttgaaga tgctcgcac tttgagccat tcaaacacaa	240
ggtaattagc atttcatgcg gtgctatggc tatcggtcga aatgttgttt tttatgttcg	300
ttaggcttgc cgactcgctg ttttttcttt tacttgatac tattaactaa agtggggaca	360
aaaagggtgca aagcctgttt tagcagatgt aatgtaattc atgaaaaaaa acagtgag	418

<210> 55

<211> 1086

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..1086

<223> /mol_type="DNA" /note="3.108" /organism="artificial sequences"

<400> 55

```

gcgagtgtct acgagcgaag ttatgaaagt tcgattcttc cccctctga aaaaccgctt      60
ttaaaaaatat tggctgctag atggttttta ctatcgtgag cttttgcttt taaaaaaca      120
cgagcgaagc gagttcatag ttgcttttgg gggtttcggg gggcttgccc cctgaacaag      180
atcacggagt ggggaatttat cacggtagtg aaaaagtacc ctctgtgtat ccttgcttat      240
ttcttttttaa acctttgaac tttttcccg t aatttgaaga aattgcccct cgactaagct      300
tgcttagtca aaaaagtttg agcaaagcga aaacataggg caattttcat gatgaaaatg      360
ggctttttaag gcttttaaat gcttttaagg ctttttagaca tgctgaaacg caagcctagc      420
aaggcataca gagggcattt aacaccgttt acctaccaat accccaccgt ttacctacca      480
ataccccacc gtttacctac caatacccca ccgtttacct accttaatac acaataatat      540
ttttatgtgg tataacgtaa taaaatatat aggtggttta tgagtgattt aatagtaaaa      600
gataacgccc taatgaacgc tagttataac ttagacttgg ttgaacagcg gttaattctt      660
ttagctatcc ttgaagctag agaatcaggc aaaggaatta acgcaaatac ccctcttaca      720
gtccatgcag agagttatat caatcaattt ggtgtagcta gacagactgc ttatcaagcc      780
ctaaaagatg cctgcaaaga tttatttgcc cgtcaattca gctatcaaga aaagcgtgaa      840
cgtggacgag ctaatattac aagtcgttgg gtcagccaaa ttgcttacat tgatgaaact      900
gcaacggttg aggttatttt cggccctgcg gttgttccac tgatcacaag gttagaggaa      960
caattctcga agtacgatat tgaacaaatt agtagtctat cgagtgccta tgcagttcgc     1020
ttatatgagt tattgatctg ctggagaaca acaggaaaga caccagttat tgacttaaca     1080
gaattc                                     1086

```

<210> 56

<211> 324

<212> PRT

<213> artificial sequences

<220>

<221> SOURCE

<222> 1..324

<223> /mol_type="protein" /note="Sphinx1.76.324.PEP" /organism="artificial sequences"

<400> 56

```

Gln Ser Glu Ile Ile Thr Asp Ala Ile Ala Asp Gln Trp Glu Ser Lys
1           5           10           15
Leu Glu Glu Phe Gly Val Val Phe Gly Ser Tyr Ile Asn Gln Phe Asn
20           25           30
Val Gln Arg His Thr Ala Tyr Gln Ala Leu Lys Asp Ala Cys Lys Asp
35           40           45
Leu Phe Ala Arg Gln Phe Ser Tyr Gln Glu Lys Arg Glu Arg Gly Arg
50           55           60
Ile Asn Ile Thr Ser Arg Trp Val Ser Gln Ile Gly Tyr Met Asp Asp
65           70           75           80

```

```

Thr Ala Thr Val Glu Ile Ile Phe Ala Pro Ala Val Val Pro Leu Ile
      85              90              95
Thr Arg Leu Glu Glu Gln Phe Thr Gln Tyr Asp Ile Glu Gln Ile Ser
      100            105            110
Gly Leu Ser Ser Ala Tyr Ala Val Arg Met Tyr Glu Leu Leu Ile Cys
      115            120            125
Trp Arg Ser Thr Gly Lys Thr Pro Ile Ile Glu Leu Asp Glu Phe Arg
      130            135            140
Lys Arg Ile Gly Val Leu Asp Thr Glu Tyr Thr Arg Thr Asp Asn Leu
      145            150            155            160
Lys Met Gln Val Ile Glu Leu Ala Leu Lys Gln Ile Asn Glu His Thr
      165            170            175
Asp Ile Thr Ala Ser Tyr Glu Gln His Lys Lys Gly Arg Val Ile Thr
      180            185            190
Gly Phe Ser Phe Met Phe Lys His Lys Lys Gln Asn Ser Asp Lys Thr
      195            200            205

Pro Asp Thr Asn Ala Ser Ser Pro Arg Ile Val Lys His Ser Gln Ile
      210            215            220
Pro Thr Asn Ile Val Lys Gln Pro Glu Asn Ala Lys Met Ser Asp Leu
      225            230            235            240
Glu His Arg Ala Ser Arg Val Thr Gly Glu Ile Met Arg Asn Arg Leu
      245            250            255
Ser Asp Arg Phe Lys Gln Gly Asp Glu Ser Ala Ile Asp Met Met Lys
      260            265            270
Arg Ile Met Ser Asp Leu Ile Val Lys Asp Asn Ala Leu Met Asn Ala
      275            280            285
Ser Tyr Asn Leu Ala Leu Val Glu Gln Arg Leu Ile Leu Leu Ala Ile
      290            295            300
Ile Glu Ala Arg Glu Thr Gly Lys Gly Ile Asn Ala Asn Asp Pro Leu
      305            310            315            320
Thr Val His Ala

```

<210> 57

<211> 201

<212> PRT

<213> artificial sequences

<220>

<221> SOURCE

<222> 1..201

<223> /mol_type="protein" /note="HCB13.108.201.PEP" /organism="artificial sequences"

<400> 57

```

Met Ser Asp Leu Ile Val Lys Asp Asn Ala Leu Met Asn Ala Ser Tyr
1      5      10      15
Asn Leu Asp Leu Val Glu Gln Arg Leu Ile Leu Leu Ala Ile Leu Glu
20     25     30
Ala Arg Glu Ser Gly Lys Gly Ile Asn Ala Asn Asp Pro Leu Thr Val
35     40     45
His Ala Ile Leu Pro Pro Ser Glu Lys Pro Leu Leu Lys Ile Leu Ala
50     55     60
Ala Arg Trp Phe Leu Leu Ser Ile Ala Tyr Ile Asp Glu Thr Ala Thr
65     70     75     80
Val Glu Val Ile Phe Ala Pro Ala Val Val Pro Leu Ile Thr Arg Leu
85     90     95
Glu Glu Gln Phe Ser Lys Tyr Asp Ile Glu Gln Ile Ser Ser Leu Ser
100    105    110
Ser Ala Tyr Ala Val Arg Leu Tyr Glu Leu Leu Ile Cys Trp Arg Thr
115    120    125
Thr Gly Lys Thr Pro Val Ile Asp Leu Thr Glu Phe Ala Ser Val Tyr
130    135    140
Glu Arg Ser Tyr Glu Ser Ser Glu Ser Tyr Ile Asn Gln Phe Gly Val
145    150    155    160
Ala Arg Gln Thr Ala Tyr Gln Ala Leu Lys Asp Ala Cys Lys Asp Leu
165    170    175
Phe Ala Arg Glu Phe Ser Thr Gln Gln Tyr Asp Glu Ser Gln Asp Ala
180    185    190    195

```

Phe Ala Arg Gln Phe Ser Tyr Gln Glu Lys Arg Glu Arg Gly Arg Ala
180 185 190
Asn Ile Thr Ser Arg Trp Val Ser Gln
195 200

<210> 58
<211> 442
<212> DNA
<213> artificial sequences

<220>
<221> source
<222> 1..442
<223> /mol_type="DNA" /note="BlastN2Query19" /organism="artificial sequences"

<400> 58
ttatgagtga tttaatagta aaagataacg ccctaatagaa cgctagttat aacttagact 60
tggttgaaca gcggttaatt ctttttagcta tccttgaagc tagagaatca ggcaaaggaa 120
ttaacgcaaa tgaccctctt acagtccatg cagagagtta tatcaatcaa tttggtgtag 180
tatcgagtgc ctatgcagtt cgctagacag actgcttatc aagccctaaa agatgcctgc 240
aaagatttat ttgcccgtca attcagctat caagaaaagc gtgaacgtgg acgagctaata 300
attacaagtc gttgggtcag ccaaattgct tacattgatg aaactgcaac ggttgagggtt 360
attttcgccc ctgcggttgt tccactgatc acaaggttag aggaacaatt ctcgaagtac 420
gatattgaac aaattagtag tc 442

<210> 59
<211> 442
<212> DNA
<213> artificial sequences

<220>
<221> source
<222> 1..442
<223> /mol_type="DNA" /note="BlastN2 Subject19" /organism="artificial sequences"

<400> 59
tatcgagtgc atatgctggt cgtgggttgaa cagagggttaa ttctattagc aatcatagaa 60
gcgagagaaa caggcaaagg gattaatgcc aatgatcctc ttacgggttca tgcaggtagc 120
tatatcaatc aatttaacgt acaaaggcat acggcatatc aagccctcaa agatgcttgt 180
aaagacttgt ttgcccgtca attcagttac caagaaaagc gagaacgagg acgaattaat 240
attacaagtc gatgggtttc gcaaattggc tatatggacg atacagcaac cgttgagatt 300
atttttgccc ctgcggttgt tcctctgatt acacgggtag aggaacagtt caccagtag 360
gatattgagc aaattagcgg ttttatgagc gatttaatag taaaagataa cgccctaata 420
aatgctagtt ataacttagc tt 442

<210> 60

<211> 132
<212> DNA
<213> artificial sequences

<220>
<221> source
<222> 1..132
<223> /mol_type="DNA" /note="BlastN2 Query20" /organism="artificial sequences"

<400> 60
tttgaagaaa ttgcccctcg actaagcttg cttagtcaaa aaagtttgag caaagcgaaa 60
acatagggca attttcatga tgaaaatggg cttttaaggc ttttaaagtc ttttaaggct 120
tttagacatg ct 132

<210> 61
<211> 130
<212> DNA
<213> artificial sequences

<220>
<221> source
<222> 1..130
<223> /mol_type="DNA" /note="BlastN2 Subject20" /organism="artificial sequences"

<400> 61
tttgcagaaa ttgcccctcg aagcttgctt agtcaaaaaa gtttgagcaa agcgaaattt 60
agacatgcts acatagggca attttcatga agaaattggg cttttaaaagt ttttaaagtg 120
tttaaagtgct 130

<210> 62
<211> 149
<212> DNA
<213> artificial sequences

<220>
<221> source
<222> 1..149
<223> /mol_type="DNA" /note="BlastN2 Query21" /organism="artificial sequences"

<400> 62
cgagcgaagc gagttcatag ttgcttttgt taaaaatatt ggctgctaga tggtttttac 60
tatcgtgagc ttttgctttt nnnnnnncag cgagtgtcta cgagcgaagt tatgaaagtt 120
cgattcttcc cccctctgaa aaaccgctt 149

<210> 63
<211> 148
<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..148

<223> /mol_type="DNA" /note="BlastN2 Subject21" /organism="artificial sequences"

<400> 63

gcgagtgtct acgagcgact caatgaaagt tcgattattc cccctctgga aaaccgcttc	60
gagcaaagcg agttcatagt tgcttttggt aaaaatattg gctgctagat ggtttttact	120
atagtgaggt tttgctttta aaaaaaca	148

<210> 64

<211> 162

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..162

<223> /mol_type="DNA" /note="BlastN2 Query22" /organism="artificial sequences"

<400> 64

tttagacatg ctgaaacgca agcctagcaa ggcatacaga ggacataggg caattttcat	60
gatgaaaatg ggcttttaag gcttttaaat gcttttaagg cttttgaaga aattgcccct	120
cgactaagct tgcttagtca aaaaagtttg agcaaagcga aa	162

<210> 65

<211> 162

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..162

<223> /mol_type="DNA" /note="BlastN2 Subject22" /organism="artificial sequences"

<400> 65

tttgcagaaa ttgcccctcg actaagcttg cttagtcaaa aaagtttgag caaagcgaaa	60
acatagggca attttcgtga tgaaaatggg cttttaatgt ttttaaatgc ttttaatgct	120
tttagacatg ctgaaacgca agcctagcaa ggcatacaga gg	162

<210> 66

<211> 149

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..149

<223> /mol_type="DNA" /note="BlastN2 Query23" /organism="artificial sequences"

<400> 66

cgagcgaagc gagttcatag ttgcttttgt taaaaatatt ggctgctaga tggtttttac	60
tatcgtgagc ttttgctttt nnnnnnncag cgagtgtcta cgagcgaagt tatgaaagtt	120
cgattcttcc cccctctgaa aaaccgctt	149

<210> 67

<211> 148

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..148

<223> /mol_type="DNA" /note="BlastN2 Subject23" /organism="artificial sequences"

<400> 67

cgagcaaagc gagttcatag ttgcttttgt taaaaatatt ggctgctaga tggtttttac	60
tatagtgagg ttttgctttt aaaaaaacag cgagtgtcta cgagcgactc aatgaaagtt	120
cgattattcc ccctctggaa aaccgctt	148

<210> 68

<211> 2958

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..2958

<223> /mol_type="DNA" /note="HCBI4.296" /organism="artificial sequences"

<400> 68

ctgagaagcg aagaaaagat gcggttttgt gtattgaaca tcttgtagct gcatcaccag	60
aatgggacgg ctggggaact gaaaaagaaa ctgcattttt tgaacagtca aggaaatggc	120
ttgaaagcaa atatggtaaa aaaaatgtgg tcagtacaac gattcataga gatgaaacaa	180
ctccacattt agttgcgtat gttgttcccg ttgacgaaga aacgggacgt ttaaattgcta	240
aaaaatttat tggtaggatct cgacatacac tttcacagat gcagactgat tttgcagttg	300
aagtaaagga tttaggatta gatcgcgggg tacaggggag caaagcaaaa catacgtcta	360
ttcaagaata ttatgaaaaa ttgaacaatt atgagaatga accaggtatt gaaaaaggac	420
tcacctatga agtgcctgaa cctgagtttt ttgaatctaa aaatgtttac ggtgagagag	480
tcacaaagac tataactact cacataattc atcaaatttc acctcaattc cacaaactca	540

```

-----
atttattggc tagtcaaaca aaaaaattaa aaaaagaact gttaaactact agaaaaacgc 600
ttgatgaagt acagaaacga gcaaaaccct atttggatat aatcaacgaa tataatcatc 660
caaatcttga gaaagaattt aataagcaag ttgctaaatt aaaagataat tttgattcag 720
cacttgagca tcatagattc ctaaaaagac aggaagaaca agaaagattt aaccaacaac 780

gtgaacttcg caatcaatta cacttagagc aagagcaaaa aaaacaactg gttgagcaag 840
aaaggcaaga aaaagaacgt ttagcacttt taagacgtca agaattagaa aatcagcgga 900
aaaatgagcc taaaaaacct gataatggca ataataacga ctactcacc ccataatacc 960
cgtttaaacg caaaaaaac ggggggttta gcctgtttg gccttttatg cataaaagta 1020
taattgaaat aaaaaaatgc tcttagaacg caaatgatga gcatttagcg agtgtctacg 1080
agcgacacaa tgaaaattcg cctattcccc ctctgaaaaa ctgcttttgc tcttttttgc 1140
ttctggagag acttgttagc gagtgtctac gagcgaagta ttgatgcttt tgctcttaaa 1200
aaagcatgag catagcgaat gcattatcta tgcttttgac tttgattttg ctcttgaaac 1260
gacacgagca acgcgagtgc catagccttt gattttgcct ttttgggct tttaatgttt 1320
ttaaatgctt ttaaatgctt ttagatagcc tgaaagcatt gctatacata tgtttcagag 1380
cttataaaga tacagattcc ttgctataaa gatacagatt ccttgctata aagatacaga 1440
ttccttgcta taaagataca gattccttgc attaaggtac atagtcatat aatgtatctt 1500
ttaatacatg attgtatgaa ataactaata tgaaaaatgg gttagtagtg aaagataatg 1560
cgttaatgaa tgccagctat aatttagaag taacagaaca gcgcttaata cttctagcaa 1620
tcattagtgc aagagaaaca gggcaaggga ttacgtcaga tagcaaatta gaaatacatg 1680
ctagtgacta tgcaattcag ttcagtgttg caaaagaaac agcttatgag gcactaaaaa 1740
gtgctgtaaa taatttattt gagcgtcaat tttcatttag agaagaaaca aaaaaaggca 1800
ctggcattgt acggtcacga tgggttagca gaattaaata cattgatgac gcagcaatac 1860
ttgaaatcac ttttgccct gacgttgtag cattaatcac tagacttgaa gaacacttca 1920
caagttatca aatcaagcaa atagcacagc ttacaagtaa gtacgctatc cgtttatatg 1980
aacttcttat tgcttggcga actacaggca aagtccttga gcttgaacta tcagaattta 2040
gaaatagatt aggcatagct agtaatgaat acacagcaat gaacaacttt aaaagccgtg 2100
tattagagcc gtctattaag cagatcaatg aacacacaga cattactgtg acgtatgaac 2160
agcataaaaa agggcggaca attacaggct tttcattcag atttaagcag aagcaacaag 2220
caaaaaaat agaaactaac agagatccaa acacacctga cttttttatc aaaatgaccg 2280
atgctcaacg gcatttattc gctaacaaaa tgtctgaaat gcccgaaatg gggaaatact 2340
cacaaggtac agaaagctac caacaatttg ctattcgcac cgccgacatg cttttacaac 2400
ctgaaaaatt cagagagctt tatccaattt tagaaaaagc cggattcaaa gaatgattga 2460
aaaagagata acaaaatttg agaaagagat tttgctacaa gacaaaatct ctacgctcga 2520
aatgaacta aaagaatttt ctgatcttca aaaaaaagca tatagcgaac ggcttcaaaa 2580
aagtatcgtg ggtttagaaa atagaatcta tcgaatcaag aaaatgcttt atacaacctg 2640
aaaggttgtc aggggttaag gggaggattc cccttactca tgaactttca atgtaccttt 2700

```

```

aggtgctttg aaagtggagt gaggccactt cgctatcgca aagctcaaaa acctctgcta 2760
tcgtgtgttt tttcgggtatc ggatttatca aatgtcattt gctattttgc gtattcaaaa 2820
attgaaatcg tttgcagacg ttggcggtag tctttcgcac aattatcgca atcgagaaac 2880
gttaaataca gatgatgctc gtactcattt gaatgaacat acgctagata caaacgaaaa 2940
atgtatgtcc gcaatcag 2958

```

<210> 69

<211> 324

<212> PRT

<213> artificial sequences

<220>

<221> SOURCE

<222> 1..324

<223> /mol_type="protein" /note="sphinx1.76.324.pep" /organism="artificial sequences"

<400> 69

```

Met Ser Asp Leu Ile Val Lys Asp Asn Ala Leu Met Asn Ala Ser Tyr
1      5      10      15
Asn Leu Ala Leu Val Glu Gln Arg Leu Ile Leu Leu Ala Ile Ile Glu
20     25     30
Ala Arg Glu Thr Gly Lys Gly Ile Asn Ala Asn Asp Pro Leu Thr Val
35     40     45
His Ala Gly Ser Tyr Ile Asn Gln Phe Asn Val Gln Arg His Thr Ala
50     55     60
Tyr Gln Ala Leu Lys Asp Ala Cys Lys Asp Leu Phe Ala Arg Gln Phe
65     70     75     80
Ser Tyr Gln Glu Lys Arg Glu Arg Gly Arg Ile Asn Ile Thr Ser Arg
85     90     95
Trp Val Ser Gln Ile Gly Tyr Met Asp Asp Thr Ala Thr Val Glu Ile
100    105    110
Ile Phe Ala Pro Ala Val Val Pro Leu Ile Thr Arg Leu Glu Glu Gln
115    120    125
Phe Thr Gln Tyr Asp Ile Glu Gln Ile Ser Gly Leu Ser Ser Ala Tyr
130    135    140
Ala Val Arg Met Tyr Glu Leu Leu Ile Cys Trp Arg Ser Thr Gly Lys
145    150    155    160
Thr Pro Ile Ile Glu Leu Asp Glu Phe Arg Lys Arg Ile Gly Val Leu
165    170    175
Asp Thr Glu Tyr Thr Arg Thr Asp Asn Leu Lys Met Gln Val Ile Glu
180    185    190
Leu Ala Leu Lys Gln Ile Asn Glu His Thr Asp Ile Thr Ala Ser Tyr
195    200    205
Glu Gln His Lys Lys Gly Arg Val Ile Thr Gly Phe Ser Phe Met Phe
210    215    220
Lys His Lys Lys Gln Asn Ser Asp Lys Thr Pro Asp Thr Asn Ala Ser
225    230    235    240
Ser Pro Arg Ile Val Lys His Ser Gln Ile Pro Thr Asn Ile Val Lys
245    250    255
Gln Pro Glu Asn Ala Lys Met Ser Asp Leu Glu His Arg Ala Ser Arg
260    265    270
Val Thr Gly Glu Ile Met Arg Asn Arg Leu Ser Asp Arg Phe Lys Gln
275    280    285
Gly Asp Glu Ser Ala Ile Asp Met Met Lys Arg Ile Gln Ser Glu Ile

290      295      300
Ile Thr Asp Ala Ile Ala Asp Gln Trp Glu Ser Lys Leu Glu Glu Phe
305      310      315      320
Gly Val Val Phe

```

<210> 70

<211> 308

<212> PRT

<213> artificial sequences

<220>

<221> SOURCE

<222> 1..308

<223> /mol_type="protein" /note="HCBI4.296.308.PEP" /organism="artificial sequences"

<400> 70

```

Arg Ile Ala Asp Met Leu Leu Gln Pro Glu Lys Phe Arg Glu Leu Tyr
1      5      10      15
Pro Ile Leu Glu Lys Ala Gly Phe Lys Glu Ala Ser Asp Tyr Ala Ile
20     25     30
Gln Phe Ser Val Ala Lys Glu Thr Ala Tyr Glu Ala Leu Lys Ser Ala
35     40     45
Val Asn Asn Leu Phe Glu Arg Gln Phe Ser Phe Arg Glu Glu Thr Lys
50     55     60
Lys Gly Thr Gly Ile Val Arg Ser Arg Trp Val Ser Arg Ile Lys Tyr
65     70     75     80
Ile Asp Asp Ala Ala Ile Leu Glu Ile Thr Phe Ala Pro Asp Val Val
85     90     95
Pro Leu Ile Thr Arg Leu Glu Glu His Phe Thr Ser Tyr Gln Ile Lys
100    105    110
Gln Ile Ala Gln Leu Thr Ser Lys Tyr Ala Ile Arg Leu Tyr Glu Leu
115    120    125
Leu Ile Ala Trp Arg Thr Thr Gly Lys Val Pro Glu Leu Glu Leu Ser
130    135    140
Glu Phe Arg Asn Arg Leu Gly Ile Ala Ser Asn Glu Tyr Thr Ala Met
145    150    155    160
Asn Asn Phe Lys Ser Arg Val Leu Glu Pro Ser Ile Lys Gln Ile Asn
165    170    175
Glu His Thr Asp Ile Thr Val Thr Tyr Glu Gln His Lys Lys Gly Arg
180    185    190
Thr Ile Thr Gly Phe Ser Phe Arg Phe Lys Gln Lys Gln Gln Ala Lys
195    200    205
Lys Ile Glu Thr Asn Arg Asp Pro Asn Thr Pro Asp Phe Phe Ile Lys
210    215    220
Met Thr Asp Ala Gln Arg His Leu Phe Ala Asn Lys Met Ser Glu Met
225    230    235    240
Pro Glu Met Gly Lys Tyr Ser Gln Gly Thr Glu Ser Tyr Gln Gln Phe
245    250    255
Ala Ile Met Lys Asn Gly Leu Val Val Lys Asp Asn Ala Leu Met Asn
260    265    270
Ala Ser Tyr Asn Leu Glu Val Thr Glu Gln Arg Leu Ile Leu Leu Ala
275    280    285
Ile Ile Ser Ala Arg Glu Thr Gly Gln Gly Ile Thr Ser Asp Ser Lys
290    295    300
Leu Glu Ile His
305

```

<210> 71

<211> 1368

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..1368

<223> /mol_type="DNA" /note="AL391558" /organism="artificial sequences"

<400> 71

```

ttctaaagaa aaggtcattt taactgaatg ttttaattta tggttgaaaa ccaacttttt      60
ctaaaacagg atacagttct ctaaattttt caggctctaa caacatattt gcgatacgta      120
tggcaaattg ttgatagctt tccgtaccct gtgaatactt ccccatctcg ggcatttcag      180
acattttatt tgcaaataaa tgacgttggt catcggtcat tttgacaaaa aaatcagggg      240
tatttggatc acgtttaatt tcaactttcg gctggtgttt atgcttaaatt ttgaatgaaa      300
accctgtgat ggtacgccct tgtttatgct gttcataagt tgcttttatg tctgtatgct      360
cattgatttg agtaatagct gtttctagta cacggctttt aaaatgggtc atacgttgat      420
actcatcatc gtccacacct aagtttttac gtagctgttg cagttcaaag gttggcgttt      480
ttccaacttc acgccaagcg attaataatt cataaagtct agtggcgtag ttactcgta      540
ggtgtgcaac ttgtttcgct tcatattttg taaaatgctc ttccaatcga gtgattaaag      600
gtactacatc aggtgcgaaa gtaatttcta gtaatgctaa gtcactaca taaaaaatac      660
ggctaacca acgagagcga acaataccta ctttccccgt tcttttatat tcagccgtat      720
agctaaactg tctattaaac agattgttta cagcttcttt tagggcttta taggatgcat      780
ctggcgacac attaaatagc tttgcataat cactagcatg tatttctagt ttgctatctg      840
ccgtgattcc ctgccctaatt tcccttgcat taattattgc cagcataatt aaacgctgtt      900
ctgttagttc taaattataa cttgcgttta ttaatgcatt atctttcaca actaaaccat      960
tcttcatatt aaatacacgt tattattaat gtgtatttac aactacata taatatgtgt     1020
gttaatcaag agaatttgtg tgcttataca agagaatttg tgtgcttata caagagaatt     1080
tgtgtgctta tacaagagaa tttgtgtgct tatataggct gaaacatact gatatcaatg     1140
ctttcagcta gtctaaaagc atttaaaagc tttaaaaata attaaaagcc taacggcagg     1200
ggagtacaac tgccctaggt actcgctaaa gctcgctact tattgctgtt cgctctgctc     1260
acagcgaggg gcagttttgg ttaattctat ttgtgaaatt taaaagcaaa agcaattatg     1320
aatcgctac gctcataaac cttttttact cgcttacgct cgatcatg                     1368

```

<210> 72

<211> 1724

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..1724

<223> /mol_type="DNA" /note="HCB15.173" /organism="artificial sequences"

<400> 72

```

ccattagctt tatagtgcc aagtaatatct agtaaacgga tatgttttat atgatctttt      60
ctttgaaaat gaagaatcat atctgaggtc gaatagttgt tatatttact ttttacgtaa      120
tcaggcatac cgatatgata atgccaaaga ttattcgaat atgtgaaatc atagattgga      180
tcagttttgt ctattccacg ccatgaaggt gcaattttac cttgggtattt agaaaaatcc      240
tttaaaccba actcttcata gatatcagtg aaatctaaaa ttttatcttg ttgatcttca      300

```

ggataacgcc caaattgaac tgcaaattgt tttccaaatt cccaagagta agacaattac 360
tttttccctg cttctaatat taatcttctt ttctcttctc tgcttaaacc ttgaggcata 420
acaaagggtt ctgaatcaag agctgatttc attctatcaa gattaatacg aacagggtgct 480
gttttttgat ttgatgtttg ggtaataacg aagcccatta cattcatttg ccacctcgac 540
aaataaaaat ttctttttaa attttaacaa aaaaaatagc gtagcgggtt ttttaattgtt 600
tttaaatact ttttaattgtt ttttagacctg ctgaaacccg ctatacaaaa ggcttttcagg 660
ccatataaaa agacattttac cctgccataa aaagacattt accctgccat aaaaagacat 720
ttacctgcc ataaaaagac atttacctg cttaaaaaga cattttacaa aatgtctttt 780
taatggtagc tttaggttgt tttttataaa aaaaattggc gttaaaaaaa tatggattta 840
gtcgttaaag ataataattt gatcaatgct agctactccc taggtttggt tgagcaacgt 900
cttgtattgc tagctatcat tgaagctaga gaatctagta aggggataga ttcagaaaca 960
ttnttagaaa tacatgctca acactatgca gatagatttg atgtaaatgt aaaaaatagc 1020
tatgccatgc tttcagaggc ggcacagact ctttttaatc gtcaagtcac ttacatgatg 1080
gttgatgaaa aaaggaataa acctgaaaag cgtgtaattc gctgggtag tggtatttca 1140
tatgtagaag gggctggagt tcttaaaactt cgtttttctc cagaaattgt ccctctaata 1200
actaggctag agcaaaattt tacgagttat gaactagaac aagttaaaag cttaaacgta 1260
tatgcaactc gactatatga attattagtc tgttggcgta gtacaggtaa aactcccatt 1320
attgaaatag aagattttcg ttcaaaaata ggagtcttac ctactgaata taaattaatg 1380
agtgatttca aaaagcgagt ttttgagcca gcaatacagc aaattaataa aaacacagat 1440
ttgacggtta attatgatca acataaatca ggccgtacaa ttacaggctt ctcattttaa 1500
tttaagcaaa agaaaacgaa atcagaaaag gtcgtaaccg ctaaacgaga cccaaataca 1560
cctgactttt ttgtaaaaat gaccgattct caacgtcatt tattcgcaac taaactttct 1620
gaaatgccag agatgtctaa atattctcaa ggtacagagt cgtatcaaca gtttgctatc 1680
cgtattgccg atatgctttt agaacctgaa aagttaaaga attc 1724

<210> 73
<211> 324
<212> PRT
<213> artificial sequences

<220>
<221> SOURCE
<222> 1..324
<223> /mol_type="protein" /note="Sphinx1.76.324.PEP" /organism="artificial sequences"

<400> 73
Arg Ile Gln Ser Glu Ile Ile Thr Asp Ala Ile Ala Asp Gln Trp Glu
1 5 10 15
Ser Lys Leu Glu Glu Phe Gly Val Val Phe Gly Ser Tyr Ile Asn Gln
20 25 30
Phe Asn Val Gln Arg His Thr Ala Tyr Gln Ala Leu Lys Asp Ala Cys
35 40 45
Lys Asp Leu Phe Ala Arg Gln Phe Ser Tyr Gln Glu Lys Arg Glu Arg
50 55 60

```

Gly Arg Ile Asn Ile Thr Ser Arg Trp Val Ser Gln Ile Gly Tyr Met
65          70          75          80
Asp Asp Thr Ala Thr Val Glu Ile Ile Phe Ala Pro Ala Val Val Pro
85          90          95
Leu Ile Thr Arg Leu Glu Glu Gln Phe Thr Gln Tyr Asp Ile Glu Gln
100        105        110
Ile Ser Gly Leu Ser Ser Ala Tyr Ala Val Arg Met Tyr Glu Leu Leu
115        120        125
Ile Cys Trp Arg Ser Thr Gly Lys Thr Pro Ile Ile Glu Leu Asp Glu
130        135        140
Phe Arg Lys Arg Ile Gly Val Leu Asp Thr Glu Tyr Thr Arg Thr Asp
145        150        155        160
Asn Leu Lys Met Gln Val Ile Glu Leu Ala Leu Lys Gln Ile Asn Glu
165        170        175
His Thr Asp Ile Thr Ala Ser Tyr Glu Gln His Lys Lys Gly Arg Val
180        185        190
Ile Thr Gly Phe Ser Phe Met Phe Lys His Lys Lys Gln Asn Ser Asp
195        200        205
Lys Thr Pro Asp Thr Asn Ala Ser Ser Pro Arg Ile Val Lys His Ser
210        215        220
Gln Ile Pro Thr Asn Ile Val Lys Gln Pro Glu Asn Ala Lys Met Ser
225        230        235        240
Asp Leu Glu His Arg Ala Ser Arg Val Thr Gly Glu Ile Met Arg Asn
245        250        255
Arg Leu Ser Asp Arg Phe Lys Gln Gly Asp Glu Ser Ala Ile Asp Met
260        265        270
Met Lys Met Ser Asp Leu Ile Val Lys Asp Asn Ala Leu Met Asn Ala
275        280        285
Ser Tyr Asn Leu Ala Leu Val Glu Gln Arg Leu Ile Leu Leu Ala Ile
290        295        300
Ile Glu Ala Arg Glu Thr Gly Lys Gly Ile Asn Ala Asn Asp Pro Leu
305        310        315        320
Thr Val His Ala

```

<210> 74

<211> 299

<212> PRT

<213> artificial sequences

<220>

<221> SOURCE

<222> 1..299

<223> /mol_type="protein" /note="HCB15.173.299.PEP" /organism="artificial sequences"

<400> 74

```

Met Asp Leu Val Val Lys Asp Asn Asn Leu Ile Asn Ala Ser Tyr Ser
1          5          10          15
Leu Gly Leu Val Glu Gln Arg Leu Val Leu Leu Ala Ile Ile Glu Ala
20        25        30
Arg Glu Ser Ser Lys Gly Ile Asp Ser Glu Thr Xaa Leu Glu Ile His
35        40        45
Ala Gln His Tyr Ala Asp Arg Phe Asp Val Asn Val Lys Asn Thr Tyr
50        55        60
Ala Met Leu Ser Glu Ala Ala Gln Thr Leu Phe Asn Arg Gln Val Thr
65        70        75        80
Tyr Met Met Val Asp Glu Lys Arg Asn Lys Pro Glu Lys Arg Val Ile
85        90        95
Arg Trp Val Ser Gly Ile Ser Tyr Val Glu Gly Ala Gly Val Leu Lys
100       105       110
Leu Arg Phe Ser Pro Glu Ile Val Pro Leu Ile Thr Arg Leu Glu Gln
115       120       125
Asn Phe Thr Ser Tyr Glu Leu Glu Gln Val Lys Ser Leu Asn Val Tyr
130       135       140
Ala Thr Arg Leu Arg Ile Ala Asp Met Leu Leu Glu Pro Glu Lys Leu
145       150       155       160
Lys Asn Ser His Asn Lys Asn Thr Asp Leu Thr Val Asn Tyr Asp Gln
165       170       175

```

```

His Lys Ser Gly Arg Thr Ile Thr Gly Phe Ser Phe Lys Phe Lys Gln
      180              185              190
Lys Lys Thr Lys Ser Glu Lys Val Val Thr Ala Lys Arg Asp Pro Asn
      195              200              205
Thr Pro Asp Phe Phe Val Lys Met Thr Asp Ser Gln Arg His Leu Phe
      210              215              220
Ala Thr Lys Leu Ser Glu Met Pro Glu Met Ser Lys Tyr Ser Gln Gly
225              230              235              240
Thr Glu Ser Tyr Gln Gln Phe Ala Ile Tyr Glu Leu Leu Val Cys Trp
      245              250              255
Arg Ser Thr Gly Lys Thr Pro Ile Ile Glu Ile Glu Asp Phe Arg Ser
      260              265              270
Lys Ile Gly Val Leu Pro Thr Glu Tyr Lys Leu Met Ser Asp Phe Lys
      275              280              285
Lys Arg Val Phe Glu Pro Ala Ile Gln Gln Ile
      290              295

```

<210> 75

<211> 1041

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..1041

<223> /mol_type="DNA" /note="GU978996" /organism="artificial sequences"

<400> 75

```

taaaacgagg ttaccttgc attaaaacga ggtttacctt gcattaaaac gaggtttacc      60

ttgcattaaa acgaggttta ccttgcatta agcgagttaa taatataacc tcgtcttatt      120
aaattaaaaa cataatcttt tcatatgaaa acagaactaa tagttaaaga taatgcctta      180
attaatgcca gttataacct tgatctagtg gagcaacggt taattctttt agcgatcctt      240
gaagcaaggg aatcgggtaa aggaataaat gctaatagat ctttaacagt tcatgctgaa      300
agttatatca atcaatttgg tgttcacgga aatacggcct atcaagcatt aaaagatgct      360
tgtgatgatc tattegtaag acaatttagt tatcaaagcc ttagtgaaaa aggaaatggt      420
attaatcaca aatcaagatg ggtgagttag gttgcttata ttgataacga ggctgtcgtt      480
agacttatct ttgctccgcg tattgtgcct ttaattacta gactagaaga acaatttaca      540
aagtatgaaa tacaacaaat aagcaattta acaagtgcct atgccgttcg tttatatgaa      600
atattaattg catggcgtag taccggaaaa acgcctctca taaccctgta cgacttcaga      660
caaaaaatag gtgtactcga tactgaatac aaacgaatgt atgattttta aaaatatgtc      720
ttggacattg cattaaaaca agtcaatgaa cataccgata ttactgtcaa agttgaacag      780
cataagacgg gcagatcaat tactggcttt tcatttagct ttaaacaaaa aaagtcagct      840
actcagtctg tcggatctaa aagagatcca aatacattgg accctttttc aacaatgaca      900
gataaacaac gtcactctatt ccagcagttt gctgtacgta tcgctggcat gctgcaagat      960
acagagcgat ttagggaaat tgctagtaaa ctctccgagc ttcttgagat gagtaaatat     1020
tcacaaggta cggaaagcta t                                     1041

```

<210> 76

<211> 158
 <212> DNA
 <213> artificial sequences

<220>
 <221> source
 <222> 1..158
 <223> /mol_type="DNA" /note="BlastN2 Query25" /organism="artificial sequences"

<400> 76
 gctatccgta ttgccgatat gcttttagaa cctgaaaact ttctgaaatg ccagagatgt 60
 ctaaatatcc tcaaggatca gagtcgtatc aacagtttaa tacacctgac ttttttgtaa 120
 aaatgaccga ttctcaacgt catttatccg caactaaa 158

<210> 77
 <211> 158
 <212> DNA
 <213> artificial sequences

<220>
 <221> source
 <222> 1..158
 <223> /mol_type="DNA" /note="BlastN2 Subject25" /organism="artificial sequences"

<400> 77
 gctattcgta tcgctgatat gcttttagag cctgaaaaat gtctgaaatg cctgatatga 60
 gtaaatatcc gcaaggaaca gaaagctacc aacaatttaa tacacctgac ttttttgtca 120
 aaatgactga tgcccaacgt catttatctg ccaataaa 158

<210> 78
 <211> 161
 <212> DNA
 <213> artificial sequences

<220>
 <221> source
 <222> 1..161
 <223> /mol_type="DNA" /note="BlastN2 Query26" /organism="artificial sequences"

<400> 78
 cagtttgcta tccgtattgc cgatatgctt ttagaacctg aactaaactt tctgaaatgc 60
 cagagatgtc taaatatctt caaggatcac agtcgtatca agaccctaat acacctgact 120
 tttttgtaaa aatgaccgat tctcaacgtc atttatccgc a 161

<210> 79
 <211> 161
 <212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..161

<223> /mol_type="DNA" /note="BlastN2 Subject26" /organism="artificial sequences"

<400> 79

```
cagttttacaa ttcgcattgc tgatatgctt ttacaacctg aaataaaatg tctgaaatgc      60
ctgaaatgat taaatattct caaggcacag aaagctatca agacccaaac acacccgact      120
tctttataaaa aatgactgat gcacaacgcc atctattcgc c                          161
```

<210> 80

<211> 152

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..152

<223> /mol_type="DNA" /note="BlastN2 Query27" /organism="artificial sequences"

<400> 80

```
attgccgata tgctttttaga acctgaaaag ttatgccaga gatgtctaaa tattctcaag      60
gtacagagtc gtatcaacag tttgctatcc gtgacttttt tgtaaaaatg accgattctc      120
aacgtcattt attcgcaact aaactttctg aa                          152
```

<210> 81

<211> 152

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..152

<223> /mol_type="DNA" /note="BlastN2 Subject27" /organism="artificial sequences"

<400> 81

```
atcgctgaca tgctttttaga gcctgaaaag ttatgcctga aatgagcaaa tattcacaag      60
gcacagaaag ctatcaacag tttgctatcc gtgacttctt tgtcaaaatg accgatgcac      120
aacgccatct attcgccaat aaaatgtctg ag                          152
```

<210> 82

<211> 153

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..153

<223> /mol_type="DNA" /note="BlastN2 Query28" /organism="artificial sequences"

<400> 82

tattgccgat atgcttttag aacctgaaaa gttaatgccg gagatgtcta aatattctca	60
aggtacagag tcgtatcaac agtttgctat ccgtgacttt tttgtaaaaa tgaccgattc	120
tcaacgtcat ttattcgcaa ctaaactttc tga	153

<210> 83

<211> 153

<212> DNA

<213> artificial sequences

<220>

<221> source

<222> 1..153

<223> /mol_type="DNA" /note="BlastN2 Subject28" /organism="artificial sequences"

<400> 83

tatcgctgac atgcttttag agcctgaaaa gttgatgcct gaaatgagca aatattcaca	60
aggcacagaa agctatcaac agttttctat ccgtgacttc tttgtcaaaa tgaccgatgc	120
acaacgccat ctattcgcca ataaaatgtc tga	153

REFERENCES CITED IN THE DESCRIPTION

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

Patent documents cited in the description

- [WO9416101A \[0020\]](#)
- [WO9429469A \[0072\]](#)
- [WO9700957A \[0072\]](#)
- [US4606918A \[0089\]](#)

Non-patent literature cited in the description

- **CNS. BRASSARD et al.** *Veterinary Microbiology*, 2007, vol. 126, 1-3271-276 [0006]
- **SCHUURMAN et al.** *Biologicals*, 1991, vol. 19, 265-270 [0006]
- **LEARY et al.** *J. of General Virology*, 1999, vol. 80, 2115-2120 [0006]
- **FUCHS et al.** *J. of Clinical Microbiology*, 1999, vol. 37, 82498-2507 [0006]
- **E. SPRARDBROW et al.** *J. of Comparative Pathology*, 1987, vol. 97, 4469-479 [0006]
- **GIOVANNY et al.** *Open J. of Medical Microbiology*, 2013, vol. 3, 84-90 [0006]
- **BUCK CBPASTRANA DVLOWY DRSCHILLER JT.** Efficient intracellular assembly of papillomaviral vectors *J. Virol.*, 2004, vol. 78, 751-757 [0099]
- **CHAPMAN MDHUGHES LEWILSON CDNAMNYAK STHOMPSON E** **GIOVANNONI** G No evidence for production of intrathecal immunoglobulin G against *Acinetobacter* or *Pseudomonas* in multiple sclerosis *Eur Neurol.*, 2005, vol. 53, 127-31 [0099]
- **VILLIERS EMBORKOSKY SSKIMMEL RGUNST KFEI JW** The diversity of Torque teno viruses: In vitro replication leads to the formation of additional replication-competent subviral Molecules *J Virol*, 2011, vol. 85, 147284-7295 [0099]
- **EBRINGER AHUGHES LRASHID TWILSON C** *Acinetobacter* Immune Responses in Multiple Sclerosis: Etiopathogenetic Role and Its Possible Use as a Diagnostic Marker *Arch Neurol.*, 2005, vol. 62, 33-36 [0099]
- **EBRINGER ARASHID TWILSON C** The role of *Acinetobacter* in the pathogenesis of multiple sclerosis examined by using Popper sequences *Med Hypotheses*, 2012, vol. 78, 6763-769 [0099]
- **HUGHES, L.E. BONELL, S. NATT, R.S. WILSON, C. TIWANA, H. EBRINGER, A. CUNNINGHAM, P. CHAMOUN, V. THOMPSON, E. J. CROKER, 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, vol. 8, 1181-1188 [0099]
- **LONGKUMER TKAMIREDDY SMUTHYALA VRAKBARPASHA SPITCHIKA GKKODETHAM GAYALURU MSIDDAVATTAM D** *Scientific Reports*, 2013, vol. 3, 2240- [0099]
- **MANUELIDIS L** Nuclease resistant circular DNAs copurify with infectivity in scrapie and CJD *J. Neurovirol.*, 2011, vol. 17, 131-145 [0099]
- **WOLD, MS** Replication protein A: heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism *Ann. Review Biochem.*, 1997, vol. 66, 61-92 [0099]
- **VALLENET DNORDMANN P BARBE V POIREL L MANGENOT S BATAILLE E DOSSAT CGAS SKREIMEYER A LENOBLE P** Comparative analysis of *Acinetobacter*s: three genomes for three lifestyles *PLoS One*, 2008, vol. 3, 3e1805-e1805 [0099]
- **XU BZHI NHU GWAN ZZHENG XLIU XWONG SKAJIGAYA SZHAO KMAO Q** Hybrid DNA virus in Chinese patients with seronegative hepatitis discovered by deep sequencing *Proc Natl Acad Sci USA*, 2013, vol. 110, 10264-9 [0099]

- **ZUR HAUSEN et al.**The Search for infectious causes of human cancer: Where and WhyVirology, 2009, vol. 392, 11-10 [0099]
- **HAUSEN H**Red meat consumption and cancer: Reasons to suspect involvement of bovine infectious factors in colorectal cancerInt J Cancer, 2012, vol. 130, 2475-2483 [0099]
- **HAUSEN H**World Cancer ReportIARC20130000 [0099]
- **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 tissueGenome Announc., 2014, vol. 2, 4e00849-14 [0099]

Patentkrav

1. HCBI-polynukleinsyre, der omfatter:

(a) en nukleotidsekvens som vist i figur 5 (SEQ ID NO: 72);

5 (b) en nukleotidsekvens, der har mindst 90 % identitet med en nukleotidsekvens ifølge (a);

(c) en nukleotidsekvens, der er komplementær til en nukleotidsekvens ifølge (a) eller (b); eller

10 (d) en nukleotidsekvens, der er i overflod som følge af den degenererede genetiske kode sammenlignet med en hvilken som helst af ovennævnte nukleotidsekvenser.

2. Oligonukleotidprimer, der omfatter en del af en HCBI-polynukleinsyre ifølge krav 1, hvilken primer er i stand til at virke som primer til specifik sekventering eller specifik amplificering af nukleinsyren i et bestemt HCBI-isolat, der indeholder en nukleotidsekvens ifølge krav 1.

15

3. Oligonukleotidprobe, der omfatter en del af en HCBI-polynukleinsyre ifølge krav 1, hvilken probe er i stand til at virke som en hybridiseringsprobe til specifik påvisning af nukleinsyren i et bestemt HCBI-isolat, der indeholder en nukleotidsekvens ifølge krav 1.

20

4. Ekspressionsvektor, der omfatter en HCBI-polynukleinsyre ifølge krav 1, der er operabelt koblet til prokaryote, eukaryote eller virale transkriptions- og translationskontrolelementer.

25

5. Værtscelle, der er transformeret med en ekspressionsvektor ifølge krav 4.

6. Polypeptid, der kodes af en HCBI-polynukleinsyre ifølge krav 1.

30

7. Antistof eller antigenbindende fragment deraf, der binder specifikt til et polypeptid ifølge krav 6.

8. Fremgangsmåde til påvisning af en HCBI-polynukleinsyre ifølge krav 1 i en biologisk prøve, som omfatter: (a) eventuel ekstraktion af prøvepolynukleinsyre, (b) amplificering af polynukleinsyren som beskrevet

35

ovenfor med mindst én primer ifølge krav 2, eventuelt en mærket primer og (c) påvisning af den amplificerede polynukleinsyre.

5 9. Fremgangsmåde til påvisning af en HCBI-polynukleinsyre ifølge krav 1 i en biologisk prøve, som omfatter: (a) eventuel ekstraktion af prøvepolynukleinsyre, (b) hybridisering af polynukleinsyren som beskrevet ovenfor med mindst én probe ifølge krav 3, eventuelt en mærket probe og (c) påvisning af den hybridiserede polynukleinsyre.

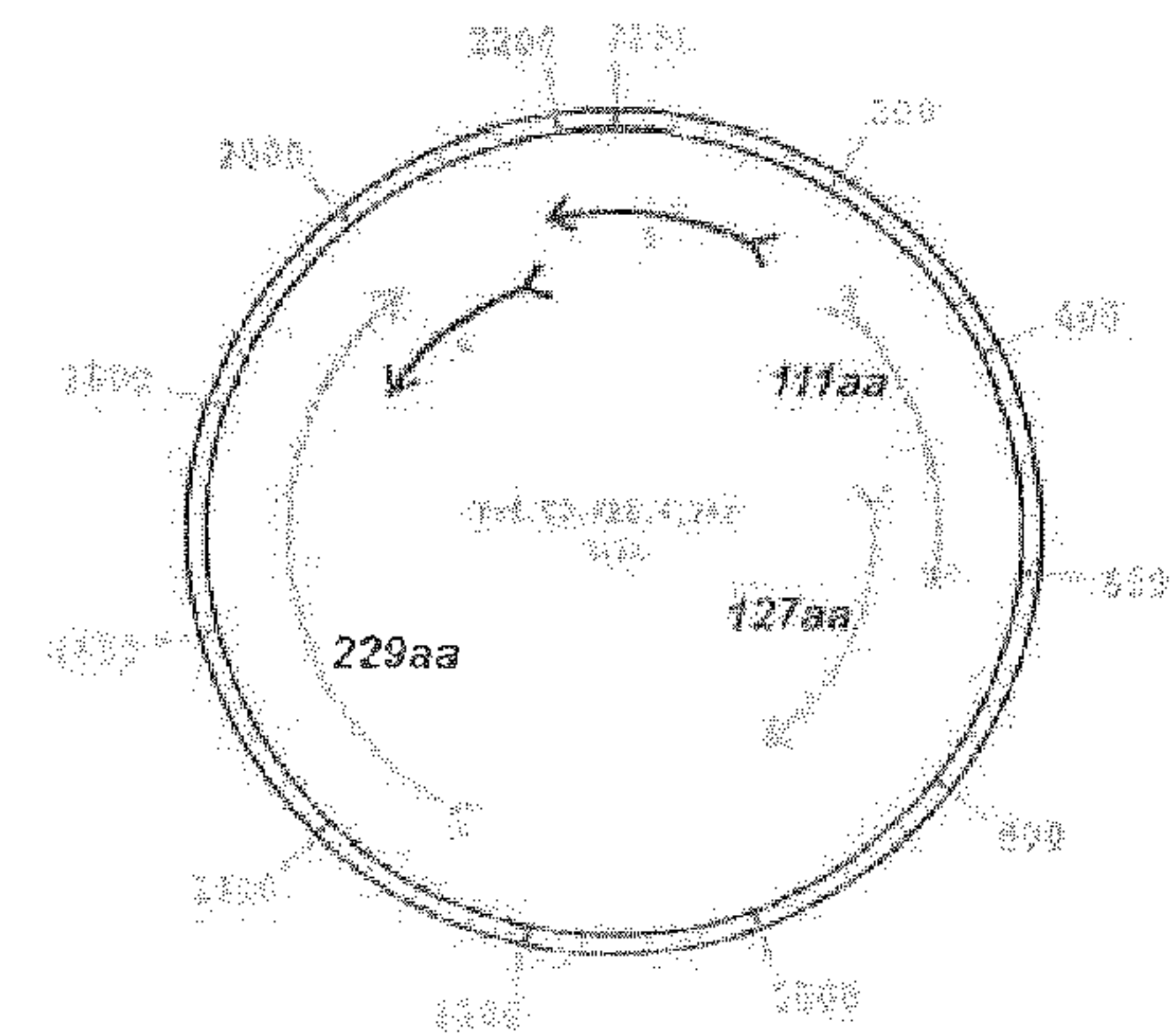
10 10. Fremgangsmåde til påvisning af et polypeptid ifølge krav 6 eller et antistof ifølge krav 7, der er til stede i en biologisk prøve, som omfatter: (a) etablering af kontakt med den biologiske prøve for tilstedeværelse af et sådant polypeptid eller antistof som defineret ovenfor og (b) påvisning af det immunologiske kompleks, der er dannet mellem antistoffet og polypeptidet.

15 11. Antisense-oligonukleotid, der reducerer eller inhiberer expressionen af en HCBI-polynukleinsyre ifølge krav 1, eller vektor, der indeholder antisense-oligonukleotidet.

20

DRAWINGS

CCTTTTCCCT GATTACAAGC TCCCATCTGC TCCAGATCGG GAGCTTTTAG CCATCCCGAA TGGGTGGCGA CACAAATTAC
AGGCGAAAAA AAAGACCATA TTGGCCTTCG GTTTAGATTG GCTGCCGTAT CAGCCAATTT GATTGAGTGC CGCATCACTC
AATCTATCAA GTGCCGCATC ACCTGATTAC TCAAATCTAC ATGAAATTTT CTTTCTTGCA AACCCTTCAG AAAAGCATT
TAAATATCCC ACCAATGAAT CATAAATGAA TCACTATAGA TCAGGAAACA ACTATGAAC TCCATAATCC AAATCCAAAT
GAGCCTACTA ACCTACAAAT GCTTGTTGCA GAAGTTAAAA AATCCGCTTC AAGCTCTTAT CACGCTGGTT ATATTCAAST
TCCTTTCCGT GTTGAGTTTG CATCATATAC ACGCCTTGAG GCACTTGTTA AACATACTGG CTCAAGTCGC AATAAGATCA
TGAATGATCT GTTAAGAATC GGTATTGAAA CTCTAGCTGC CTCTTTGGAT GATGAAACAA TTAAACTCT TTTGAAATT
GAAACCTCAA TCACTGCGGA TCTCTATGCT TCAGGAAAAA TTAAATCAGG GGATCAATCA GATGATTAAT TTACAAGGAA
CTCTAATTAA TGCTTTTCGT GTGGATGGTG GTAAAGGTAA AGACGGCAAA GAATATGAAG CGCGTGACAA GGTGCAAATT
CTTGTTCCGT TGGAGTTACC CAATGGGGAA ATTAAACACG AACTTGTGTA CCTAACAGTT GAGGATTCTC GACTTTTTGA
ACCATTCAAA AATCAGGTTA TTAGCATCTC ATGTGGTGCT ATGGCTGTCG GTCGTAATGT TATTTTTTAT GTTCGCAAAG
GCGCAAAACC AGTTTTAGCA GAACAAATGT GAAACCCGAA AAAAAAACA GTGAGTTAGC GAAGCAACTC GATACTATTG
ATTAAGTAGG GGACAAAAAT TGAATATAAA AGACAATAAA AAAGCCGATA ATTCCGATGC TTTGGCGGGC GATGAATTAC
AGGCTTTGCA ATCTGCTAAG GCAGATCAAC ACAGGGATAG AATAACACGT TTTGGAATTT TGAAACATAG ACCGAAGCTA
CAAGAACAAT ATTTGTGGAC GCAAGTTGAT TTAAATCCG AAGGTGAAAA TGATCTATCT AATAAGGCTC TTAAGGCTGC
AACCAACTC AAGGGATGCG GTCAATTTCT CTTATTCCAT AATTACTACA CAATAGACCA AGTTAACTT GCTAAGGCCT
ATTACTGTT TCAGCATTTG CTATGTCCTA TGTGTGCTGG TGTAAGGGCT GCTAAGTCAA TGAGTCGTTA TATTAGCGT
ATTGAAGAAC TAATGCGCCA GAATCGCCAT TTAAAGCCCG TCCTGATCAC TTTGACGGTA AAGAATGGCC CTGACCTACA
AGAACGCTC AAACACCTTA GAGCATCAT TAGAACGCTT TTAGATCGTT ATAACGATTA CAAGAAAAAA GGTGCTGGTT
TTAATCAGTT TTGTAAATTT GATGGTGCTT TTTATTCAAC TGAATATACA TACAATCCAA AAATAAAGA ATGGCATCCG
CATATTGATA TTTTGGCCTT ACTCAATCAA TGGATAGACC AGGAACAATT GTCCGAACT TGGCACGATA TTAATCTGGA
TTCTTATATC GTCGATATC GTAGAGTTAA AAAAACCAAA GAACACGGCT ATAGCAAAGC TGTTGCTGAG GTATGTAAAT
ACGCCCCTAA GTTTAGTGAT CTCTCACTGG AGAACACTTG GGAGGCATAT CTTTCTTTGA AAGGTAACAG GCTTACTGGC
TGTTTTGGTT CTATGTAIGG TGTCAAATTG CCTGAAAAAC TTACTGATGA TCTACCCCTT GATGATCTTC CATATTTGGA
GCTGTTATAC CGTTTTGTTT TTGGTAAAAA ATCTTATTAC AACCTAGAAA TAACAAAAGA CGTAAAGCCA CAAAAC TAGG
ACTACAACGA TGAGGTGAGG GCGACGCGCG ACCGTGCGCG CTACTTCGTG CATGTACGCG CGCGCCTTTG CCTCTTTGTT
GCGAGGTGTG GACGAAAAAA GCAAGGATAC ATACAGACCC CTGCATGACC TTGTAAGGGG TTCGACCCCT TAGACCCCAA
AGGGCGCACT TATGCAAACT CTTGAGTTT GCCAGTGCTC CCACCAGTAA CAGAGGGCGC GGAGTGCGCC CGAACTGACC
CTATAGAATT C



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

Putative ORFs:

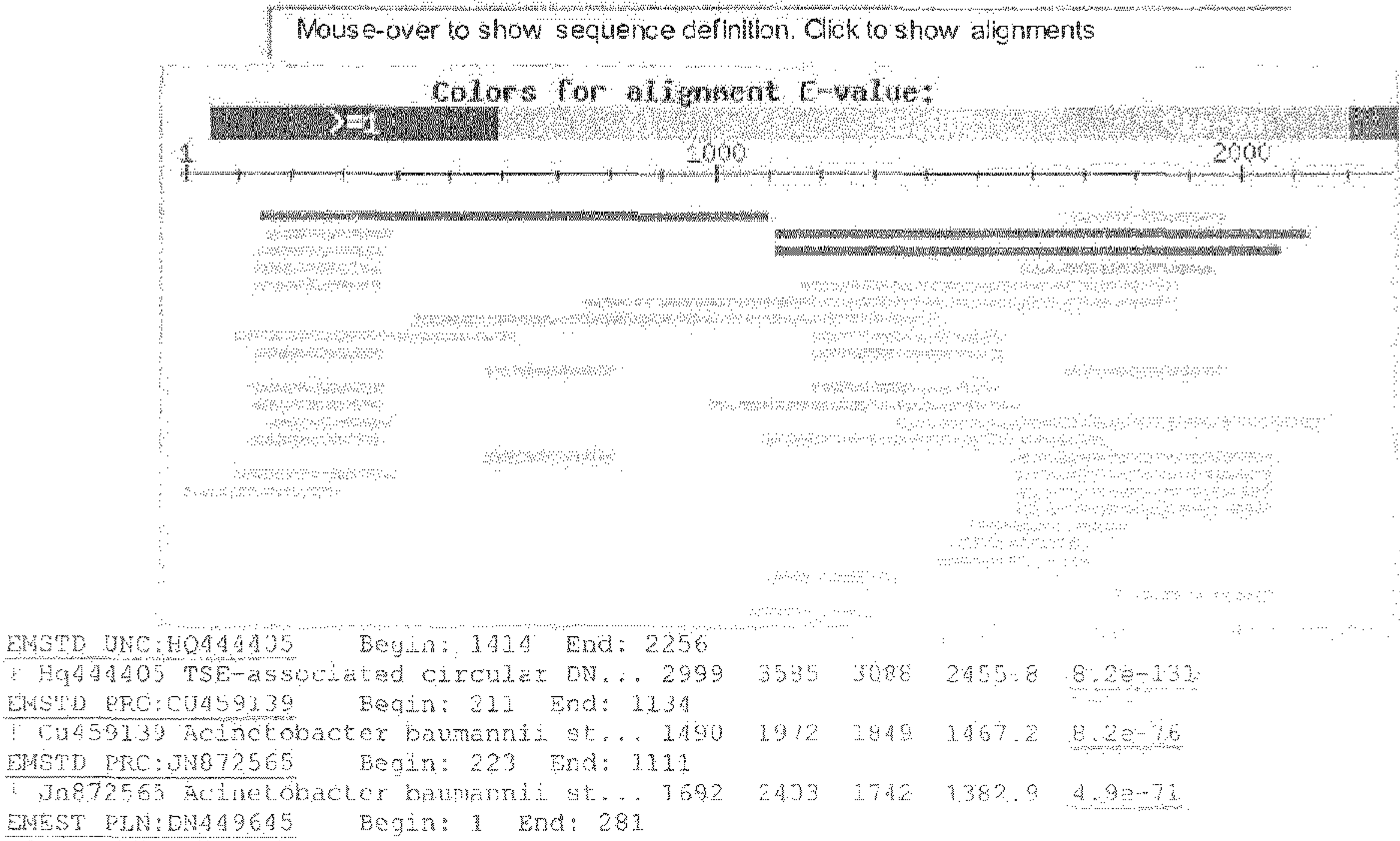
229 amino acids	82% similarity to Sphinx2.36 (232 aa)	Replication protein
127amino acids	91% similarity to Sphinx2.36 (96 aa)	Function unknown
111 amino acids	97% similarity to Sphinx2.36 (124 aa)	Function unknown

229 amino acids	82% similarity to Sphinx2.36 (232 aa)	Replication protein
	1	50
SPHINX2.36.232.PEP	MQEQLWTQV DFKVGSETSI KALKAAATKTK GCGQELLFRN YYTIDQIKLE	
HCBI1.225.229.pep		
	51	100
SPHINX2.36.232.PEP	KFHVCGQHLL CPMCAGIRAA RSMNRYIQRI EEIMRONRKL KPVLIITLVK	
HCBI1.225.229.pep	~~~~~MCAGVRAA ASMSRYIQRI EEIMRONRAH KPVLIITLVK	
	101	150
SPHINX2.36.232.PEP	NGEDLQERFE HLTGSFKTLL QRYRDEKRRG RGFNOFCRIG GGFYTTEYTY	
HCBI1.225.229.pep	NGPDLOERFK HLRASFETLL DRYNDYKKKG RGFNOFCRIG GAFYSTETTY	
	151	200
SPHINX2.36.232.PEP	NETTQQWHPH IHIFALVTDR IDQEELAEFW HDITLDSYIV DIRMVKKTK	
HCBI1.225.229.pep	NPKTKLWHPH IHIFALLNQW IDQEELSEFW HDITLDSYIV DIRMVKKTK	
	201	250
SPHINX2.36.232.PEP	HGYAKAVAEV CKYALRPSDL STEKTFQAFF DP~~~~~	
HCBI1.225.229.pep	HGYSKAVAEV CKYALRPSDL SLENTWEAYL SLRGNALTCG FGSMYGVKLP	
	251	291
SPHINX2.36.232.PEP	~~~~~	
HCBI1.225.229.pep	EKLTDDELFLD DLPYLELLYR FVEGKKSYYN DEITRDVKEQ N	

96 amino acids	91% similarity to Sphinx2.36 (96 aa)	Function unknown
	1	50
SPHINX2.36.96.PEP	~~~~~ -MINLQGTII NAERLDGGKG	
HCBI1.225.127.pep	MMKQLKFLK IKPOSLRISM LOEKLNOGIN QMINLQGTII NAERVDGGKG	
	51	100
SPHINX2.36.96.PEP	KDGKEYEARD KVQILGSLEL PNGEIKHELV DLIVEDARIF EEFKHKVISI	
HCBI1.225.127.pep	KDGKEYEARD KVQILGSLEL PNGEIKHELV DLTVEDSRIF EEFKNQVISI	
	101	127
SPHINX2.36.96.PEP	SCGAMAIGRN VVEYVRKGAK PVLADVM	
HCBI1.225.127.pep	SCGAMAVGRN VVEYVRKGAK PVLAEQM	

111 amino acids	97% similarity to Sphinx2.36 (124 aa)	Function unknown
	1	50
SPHINX2.36.124.PEP	MNHKLIADQ ELTMKLNPN PNEPTNLQML VAEIKKSASS SYHGGYIQVP	
HCBI1.225.111.pep	~~~~~ ~~~~MKLNPN PNEPTNLQML VAEVKKSSASS SYHGGYIQVP	
	51	100
SPHINX2.36.124.PEP	FRVECASYTR LEALVKHTGS SRNKIMNDLL RIGIETLAAS LEDETIKTLF	
HCBI1.225.111.pep	FRVEFASYTR LEALVKHTGS SRNKIMNDLL RIGIETLAAS LEDETIKTLF	
	101	124
SPHINX2.36.124.PEP	EIETSITADL YASGKMKSGD QSDQ	
HCBI1.225.111.pep	EIETSITADL YASCKIKSGD QSDQ	

FastA *Geall



! 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	GAGTSCCGCATCACTCAATCTATCAAGTGC	CGCATCACCTGATTACTCA-ATCTACATG				
HQ444405	ATAAATTGCGGTCACTGTGCGGAATCACTACGCAATATTTATTACCCATAATTTAACCA					
	1390	1400	1410	1420	1430	1440
	210	220	230	240	250	260
rmf.C2.F8E.4	AAATTTTCTTTCTTGCAAACGGTTTCAGAAAAGCATTATAAAATATCCACCAATGAATCAT					
HQ444405	ATTCTTAATTTCTTGCAAACGGTTTCAGAAAAGCATTATAAAATATCTCACCATGAATCAT					
	1450	1460	1470	1480	1490	1500
	270	280	290	300	310	320
rmf.C2.F8E.4	AAATGAATCACTATAGATCAGGAAACCACTATGAAACTCCATAATCCAAATCCAAATGAG					
HQ444405	AAACTAATCGCTATAGATCAGGAAATTAACATGAAACTCCATAATCCAAATCCAAATGAG					
	1510	1520	1530	1540	1550	1560
	330	340	350	360	370	380
rmf.C2.F8E.4	CCTACTAACCTACAAATGCTTGTTCAGAAAGTTAAAAAATCCGCTTCAAGCTCTTATCAC					
HQ444405	CCTACTAACCTGCAATGCTTGTTCAGAAATTAATAAATCCGCTTCAAGCTCTTATCAC					
	1570	1580	1590	1600	1610	1620
	390	400	410	420	430	440
rmf.C2.F8E.4	GCTGGTTATATTCAAGTTCTTTCCGTTGAGTTTGCATCATATACACGCCTTGAGGCA					
HQ444405	GGTGGCTATATTCAAGTTCTTTCCGTTGAGTTTGCATCATATACACGCCTTGAGGCA					
	1630	1640	1650	1660	1670	1680
	450	460	470	480	490	500
rmf.C2.F8E.4	CTTGTTAAACATACTGGCTCAAGTCCGAATAAGATCATGAATGATCTCTTAACAATCGGT					
HQ444405	CTTGTTAAACATACTGGCTCAAGTCCGAATAAGATCATGAATGATCTCTTAACAATCGGT					
	1690	1700	1710	1720	1730	1740
	510	520	530	540	550	560
rmf.C2.F8E.4	ATTGAAACTCTAGCTCCCTCTTTGGATGATGAAACAATTAAACTCTTTTGAATTGAA					
HQ444405	ATTGAAACTCTAGCTCCCTCTTTGGATGATGAAACAATTAAACTCTTTTGAATTGAA					
	1750	1760	1770	1780	1790	1800
	570	580	590	600	610	620
rmf.C2.F8E.4	ACCTCAATCACTGCGGATCTCTATGCTTCAGGAAAATTAAATCAGGGGATCAATCAGAT					
HQ444405	ACTTCAATCACTGCGGATCTCTATGCTTCAGGAAAATTGAAATCAGGAGATCAATCAGAT					
	1810	1820	1830	1840	1850	1860
	630	640	650	660	670	680
rmf.C2.F8E.4	GATTAAATTACAAAGCAACTCTAATTAATGCTTTTCGTTGGATGGTGGTAAAGGTAAAGA					
HQ444405	GATTAAATTACAAAGCAACTCTAATTAATGCTTTTCGTTGGATGGTGGTAAAGGTAAAGA					
	1870	1880	1890	1900	1910	1920
	690	700	710	720	730	740
rmf.C2.F8E.4	CGGCAAGAATATGAAGGCGGTGACAAAGGTGCAAAATTCCTTGGTTGGTGGTAAAGGTAAAGA					
HQ444405	CGGCAAGAATATGAAGGCGGTGACAAAGGTGCAAAATTCCTTGGTTGGTGGTAAAGGTAAAGA					
	1930	1940	1950	1960	1970	1980
	750	760	770	780	790	800
rmf.C2.F8E.4	TGGGGAAATTAACACCAACTTGTTCACCTAACAGTTGAGGATCTTCGACTTTTGAACC					
HQ444405	CGGTGAGATCAACATGAGCTTGTTCACCTAACAGTTGAGGATCTTCGACTTTTGAACC					
	1990	2000	2010	2020	2030	2040
	810	820	830	840	850	860
rmf.C2.F8E.4	ATTCAAAAATCAGGTTATTAGCATCTCATGTGGTGGTATGGCTCTCGGTGCTAATGTAT					

```
HQ444405      ATTCAAACACAAGGTAATTAGCATTTCATGCGGTGCTATGGCTATCGGTCGAAATGTTTGT
                2050      2060      2070      2080      2090      2100

rmf.C2.F8E.4      870      880      890      900      910      920
TTTTTATGTTTCGCAAAGGCGCAAAACCAAGTTTTCAGCAGAACAAATGTGAAACCCGAAATAA
||||| ||||| ||||| || ||||| ||||| ||||| || || || |||
HQ444405      TTTTATGTTTCGAAAAGGTGCAAAGCCTGTTTTCAGCAGATGTAATGT--AATTCATGAAA
                2110      2120      2130      2140      2150      2160

rmf.C2.F8E.4      930      940      950      960      970
AAAAACAGTGAGTTA-GCGAAGCAACTCGATACTATTGATTAACTAGGGGGACAA-----
||||| ||||| ||||| || ||||| || ||||| || || || |||
HQ444405      AAAAACAGTGAGTTAGGCTTGCCGACTCGCTGTTTTTCTTTTACTTGATACTATTAACT
                2170      2180      2190      2200      2210      2220

rmf.C2.F8E.4      980      990      1000      1010      1020      1030
AAATTGACTATAAAAGACAATAAAAAACCCCATTAATTCGATGCTTTGGCGGGCCGATGAA
||| || ||||| || ||||| || ||||| || ||||| || |||||
HQ444405      AAAGTGGGGACAAAATTGACTAAAAAAGACAATAAAAAACCCCATTAATTCGATGCTTTG
                2230      2240      2250      2260      2270      2280

rmf.C2.F8E.4.3kr.seq
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: 1467.2 E(): 8.2e-76
>>EMSTD_PRO:CU459139 (2726 nt)
67.1% identity in 926 nt overlap
(1143-2064:211-1134)

rmf.C2.F8E.4      1120      1130      1140      1150      1160      1170
CGAAGCTACAAGAACAAATATTTGTGGACGCAAGTTGATTTTAAATCCGAAGGTGAAAAATG
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CU459139      AAAACTATTTATTTTCGTTAGCTAAGATTAAAGAAAATTATCATGCCGATGTAAAAACG
                190      200      210      220      230      240

rmf.C2.F8E.4      1180      1190      1200      1210      1220      1230
ATCTATCTAATAAGGCTCTTAAGGCTGCAACCAAACTCAAGGGATGCGGTCAATTCTCT
|| ||||| || || ||||| ||||| ||||| ||||| ||||| |||||
CU459139      ATGAATCTATTTCGCGCCATGAAAACCTGCCCAAAATTAATGGGTGCGGTAAATTTCTTC
                250      260      270      280      290      300

rmf.C2.F8E.4      1240      1250      1260      1270      1280      1290
TATTCGATAATTACTACACAATAGACCAAGTTAACTTGCTAAGGCCTATTACTGTTCTC
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CU459139      TATTCAAAAATTTTACACCATTAAATCAAATTAAGCTCGCCAGTCCAGCTTGTAGTG
                310      320      330      340      350      360

rmf.C2.F8E.4      1300      1310      1320      1330      1340      1350
AGCATTTGCTATGTCCTATGTGTGCTGCTGTAAGGGCTGCTAAGTCAATGAGTCGTTATA
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CU459139      AGCATTTGTTATGTCCTGTTGTGCTGCTATTAGAGCTTCTAAGGCAATTCAAAAATACT
                370      380      390      400      410      420

rmf.C2.F8E.4      1360      1370      1380      1390      1400      1410
TTCAGCGTATTGAAGAACTAATGCGCCAGAAATCGCCATTAAAGCCCGTCCTCATCACTT
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CU459139      CTGAGCGTGTGATCAAGTCTTATCTGAAAATCCTCGTTTAAAGCCCGTTATGATCACGT
                430      440      450      460      470      480

rmf.C2.F8E.4      1420      1430      1440      1450      1460      1470
TGACCGTTAAAGAATGGCCCTGACCTACAAGAACGCTTCAAACACCTTAGAGCATCATTTA
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CU459139      TTACCGTTAAAAATGGGGTAGACCTAGCGGGAACGGTTCACCCATCTTATAAAATCGTTTA
                490      500      510      520      530      540

rmf.C2.F8E.4      1480      1490      1500      1510      1520      1530
GAACGCTTTTATGATCGTTATTAACGATTACAAGAAAAAAGGTCGTGGTTTAAATCAGTTT
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CU459139      GAACGCTTATAGAGCGTCTAGGGACTATATTAAGAAAGGGCGTGGCTTTAATGAATTTT
                550      560      570      580      590      600

rmf.C2.F8E.4      1540      1550      1560      1570      1580      1590
GTAAAAATGATGGTGGCTTTTATTCACTGAATATACATACAATCCAAAAACTAAAGAAT
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CU459139      GCAAAATTAATGGTGGGATGTATTCATATGAGAATACTTACAATGAAAAAACTAATGAAT
                610      620      630      640      650      660

rmf.C2.F8E.4      1600      1610      1620      1630      1640      1650
```

```
rmf.C2.F8E.4 GGCATCCGCATATTCATATTTTGGCCTTACTCAATCAATGGATAGACCAGGAAGAATTGT
CU459139 GGCATCCGCATATTCATATGTTTGCACCTTTGGATGATTGGATAGATCAGGATGAATTGT
670 680 690 700 710 720

1660 1670 1680 1690 1700 1710
rmf.C2.F8E.4 CCGAAACTTGGCAGCATATTACTCTGGATTCTTATATCGTCGATATTCGTAGACTTAAAA
CU459139 CTCAATATTGGCAATCCATTACTGGGCACTCTATGCTCGTTGATATTCGTAGACCCAAA
730 740 750 760 770 780

1720 1730 1740 1750 1760 1770
rmf.C2.F8E.4 AAACCAAGAAGACCGGCTATAGCAAGCTGTTGCTGAGGTATGTAAATACGCCCTTAAGT
CU459139 AACAAAAAGACTTACGCTATTACAGGTGCTGCTGCTCAAGTCTGTAAATATCCTCTCAAT
790 800 810 820 830 840

1780 1790 1800 1810 1820 1830
rmf.C2.F8E.4 TTAGTGATCTCTCAGCTGGAGAACACTTGGGAGGCATATCTTTCTTTGAAAGGTAAACAGGC
CU459139 TTGCTGATCTTCTGTAGAAAAGACTTGGGAAGCTTTCAAAGCTTTGAAAGGTAAACGAT
850 860 870 880 890 900

1840 1850 1860 1870 1880 1890
rmf.C2.F8E.4 TTAGTGATCTCTTGGCTTCTATGTATGCTGCTCAATTCCTGAAAACTTACTGATGATC
CU459139 TAAGTGGGGCTTTTGGATCTCTTGGGGCTGAAAAATTCCTGAATCATTCATAGATGATC
910 920 930 940 950 960

1900 1910 1920 1930 1940
rmf.C2.F8E.4 TACC---CCTTGATGATCTTCCATATTTGGAGCTGTTATACCGTTTCTTTTGGTAAAA
CU459139 TTCCAGAGGATTTCTGATTTACCTTATTAGAAATGATTTATAAGTTCGCTTTTCTAAGA
970 980 990 1000 1010 1020

1950 1960 1970 1980 1990 2000
rmf.C2.F8E.4 AATCTTATTACAACCTAGAAATAACAAAAGACGTAAAGCCACAAACTAGGACTACACCG
CU459139 AGTCTTATTACGATTTACAACCTTACTCGTCAIGTCGAACCTACAGGTAAGGACGACGCCG
1030 1040 1050 1060 1070 1080

2010 2020 2030 2040 2050 2060
rmf.C2.F8E.4 ATGAGGTGAG-GGCGACGCGCGACCGTGGCGGCTACTTCGTGCAITGACGCGCGCGCCTT
CU459139 ACGAGCTTCGAGGAGAGAAGGACGCAACCTGTTGGT---GAGCATGACGGGCGAGGAGC
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-scores: 1382.9 E(): 4.9e-71
>>EMSTD_PRO:JN872565 (2252 nt)
66.4% identity in 889 nt overlap
(1143-2030:223-1111)

1120 1130 1140 1150 1160 1170
rmf.C2.F8E.4 CGAAGCTACAAGAACAAATATTTGTGGACGCAAGTTGATTTTAAATCCGAAGGTGAAAATC
JN872565 AAAATTATTTATTTACCCTGCGCTAAGTTTAAAGAGAACTATGAAAAGACGTTAAAAACG
200 210 220 230 240 250

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

1240 1250 1260 1270 1280 1290
rmf.C2.F8E.4 TATTCCATAATTACTACACAATAGACCAAGTTAAACCTGCTAAGGCTTATTACTGTTCTC
JN872565 TATTCAAAAATTTTACACAATAGGCGAAGTTAAACCTGCTCAAGCTCGGCACTGCGGAC
320 330 340 350 360 370

1300 1310 1320 1330 1340 1350
rmf.C2.F8E.4 AGCATTTGCTATGTCCTATGTGTGCTGGTGTAAAGGCTGCTAAGTCAATGAGTGGTTATA
JN872565 AGCATTTGCTTTGCCCTTCTGTGCTGCAATTCCTGCTTCTGCTGCTATGCCAAAATACG
380 390 400 410 420 430
```

1360 1370 1380 1390 1400 1410
rmf.C2.F8E.4 TTCAGCGTATTGAAGAACTAATGCGCCAGAATCGCCATTTAAAGCCCGTCCTGATCACTT
JN872565 TTGAACCTATTGATCAAGTCCTGCAACAAAATCGCAAGCTCAACCCCGTTCTAATCACGC
440 450 460 470 480 490

1420 1430 1440 1450 1460 1470
rmf.C2.F8E.4 TGACGGTAAAGAATGCGCCTGACCTACAAGAACGCTTCAAACACCTTAGAGCATCATTTA
JN872565 TCACCGTTAAAAACGGCTCTGACCTAGCAGAACGCTCCGAACATCTTATGAAGTCCTTTA
500 510 520 530 540 550

1480 1490 1500 1510 1520 1530
rmf.C2.F8E.4 GAACGCTTTTAGATCGTTATAACGATTACAAGAAAAAGGTCGTGGTTTTAATCAGTTT
JN872565 GAACGCTCCTAGAGCGTCGTAGGGACTATGAAAAGAAAGGTCGAGGTTTAAATGAGTTCT
560 570 580 590 600 610

1540 1550 1560 1570 1580 1590
rmf.C2.F8E.4 GTAAATTTGATGGTGCTTTTTATTCAACTGAATATACATACAATCCAAAAACTAAAGAAT
JN872565 GTAAGGTTCAAGGGGCTATGTACTCCTATGAAAATACATTCAATGAAAAAACAGGCGAAT
620 630 640 650 660 670

1600 1610 1620 1630 1640 1650
rmf.C2.F8E.4 GGCATCCGCATATTTCATATTTTTGCCTTACTCAATCAATGGATAGACCAGGAAGAATGT
JN872565 GGCATCCGCATATTTCATATGTTCCGCTTTGGTTGATCAATGGATTGATCAGCAAGAGTTT
680 690 700 710 720 730

1660 1670 1680 1690 1700 1710
rmf.C2.F8E.4 CCCAAACTTGGCACGATATTACTCTGGATICTTATATCGTCGATATTTCGTAGAGTTAAAA
JN872565 CAGAATAATTGGCATAGCCTTACTGGGGACICGATGGTTGTCCGATGTCGCGAGGGCAAGAA
740 750 760 770 780 790

1720 1730 1740 1750 1760 1770
rmf.C2.F8E.4 AAACCAAAGAACACGGCTATAGCAAAGCTGTTGCTGAGGTATGTAATACGCCCTTAAGT
JN872565 AAGAAAAAGGTTACGCTTATAGCAAAGCGGCTCCCGAAGTCTGTAAGTATGCTTTGAACT
800 810 820 830 840 850

1780 1790 1800 1810 1820 1830
rmf.C2.F8E.4 TTAGTGATCTCTCACTGGAGAACACTTGGGAGGCAATATCTTCTTTGAAAGGTATACAGGC
JN872565 TTGGTGATCTGTCCGTTGAAAGACTTGGGAAGCATTTAAGCTTCTTAAGGCAAGCGTT
860 870 880 890 900 910

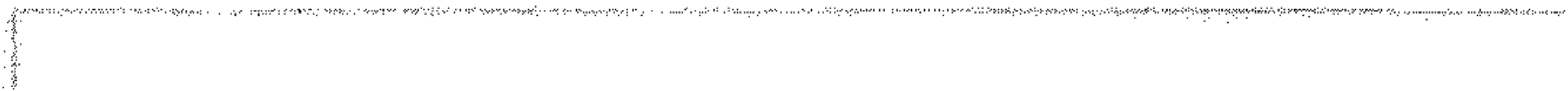
1840 1850 1860 1870 1880 1890
rmf.C2.F8E.4 TTACTGGCTGTTTGGTTCTATGTATGGTGTCAAATTGCGCTGAAAAACTTACTGATGATC
JN872565 TAACAGGTTCTTTTGGTCTGCTATGGGCTGTCAAATCCCTGACTCAATGACAGACGATA
920 930 940 950 960 970

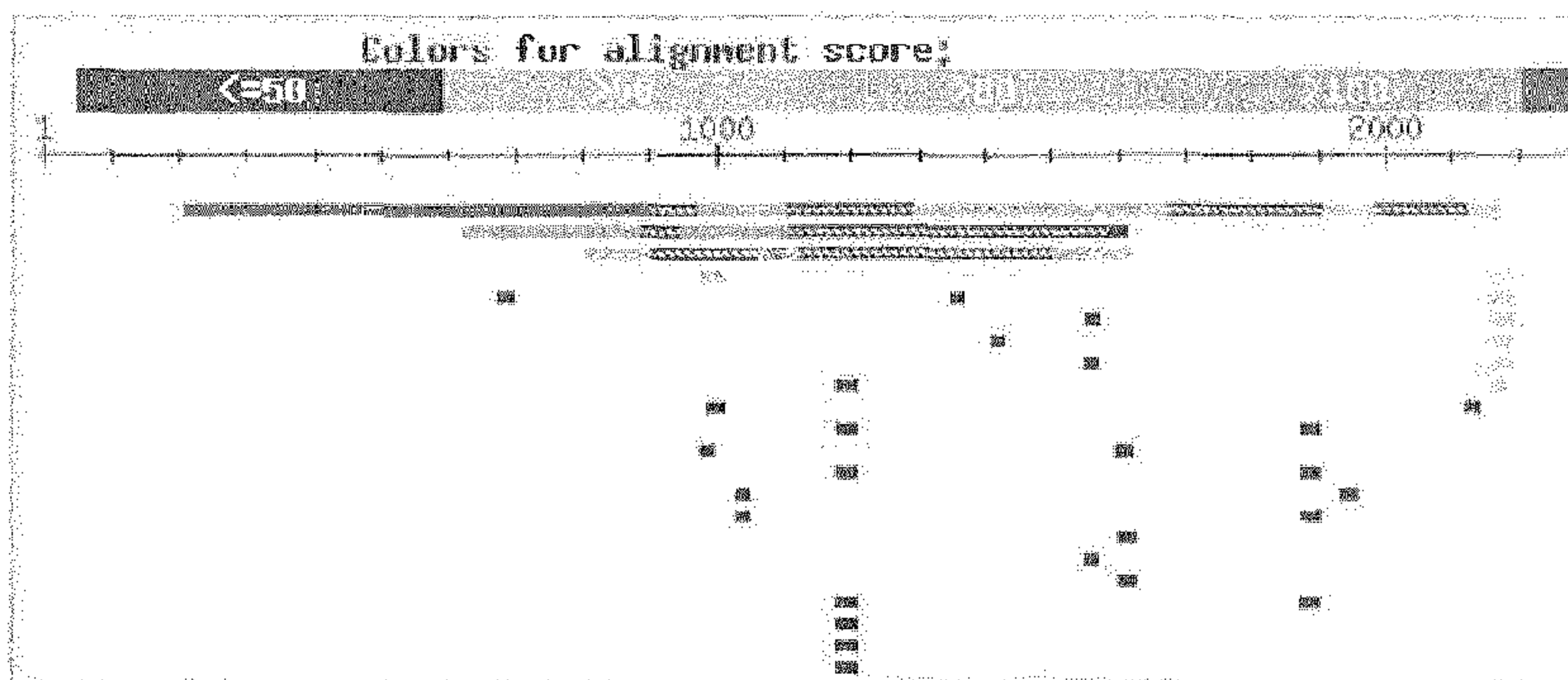
1900 1910 1920 1930 1940 1950
rmf.C2.F8E.4 TACCCCTTGATGATCTTCCATATTTGGAGCTGTTATACCGTTTGT'TTTTGGTAAAAAAT
JN872565 TGCCATCAGAAGACTTGCCATATCTCGAAATGCTGTACAAGTTTGCTTACAGTAAAAAGT
980 990 1000 1010 1020 1030

1960 1970 1980 1990 2000 2010
rmf.C2.F8E.4 CTTATTACAACCTAGAAATAACAAAAGACGTAAAGCCACAAAACCTAGGACTACAACGATG
JN872565 CTTACTACGACTTACIAATCACAAAGGCACGTAGAGCCACAACCGCATGAGGACGAGCGTG
1040 1050 1060 1070 1080 1090

2020 2030 2040 2050 2060 2070
rmf.C2.F8E.4 AGGTGAG-GCCGACGCGCGACCGTCCCGCTACTTCCGTGCATGTACGCGCGCGCCTTTGC
JN872565 CGAGGAGCTTCGACGAGTGCATTTGTATTTATGCGGTGGAGGCTCAGACGTTTGAAGTGTG
1100 1110 1120 1130 1140 1150

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 tttcttgcaaacgggttcagaaaagcattataaatacccacacatgaatcataaatgaat 271

|||||

Sbjct: 1452 tttcttgcaaacgggttcagaaaagcattataaatacccacacatgaatcataaatgaat 1511

Query: 272 caclatagatcaggaaacaactatgaaacccataatccaaatccaaatgagcctactaa 331

|||||

Sbjct: 1512 cgtatagatcaggaaattaactatgaaacccataatccaaatccaaatgagcctactaa 1571

Query: 332 cctacaaatgcttgttgacagaagttaaaaaatccgcttcaagctcttatcacgggtggtta 391

|||||

Sbjct: 1572 cctgcaaatgcttgttgacagaagttaaaaaatccgcttcaagctcttatcacgggtggtta 1631

Query: 392 tattcaagtccctttccgtgttgagtttgcacatatacacgccttgaggcacttggttaa 451

|||||

Sbjct: 1632 tattcaagtccctttccgtgttgagtttgcacatatacacgccttgaggcacttggttaa 1691

Query: 452 acatactggctcaagtcgcaataagatcatgaatgatctgttaagaatcgggtattgaaac 511

|||||

Sbjct: 1692 acatactggctcaagtcgcaataagatcatgaatgatctgttaagaatcgggtattgaaac 1751

Query: 512 tctagctgcctcttttgatgatgaaacaattaaaactctttttgaaattgaaacctcaat 571

|||||

Sbjct: 1752 tctagctgcctcttttgatgatgaaacaattaaaactctttttgaaattgaaacctcaat 1811

Query: 572 cactgaggatctctatgcttcaggaaaaattaaatcaggggatcaatcagatgattaatt 631

|||||

Sbjct: 1812 cactgaggatctctatgcttcaggaaaaattaaatcaggaggatcagtcagatgattaatt 1871

Query: 632 tacaagggaactctaatatgcttttctgtgttgatgggtggttaagggttaaggacggcnaag 691

|||||

Sbjct: 1872 tacaagggaactctaatatgcttttctgtgttgatgggtggttaagggttaaggacggcnaag 1931

Query: 692 aatatgaagcgcggtgacaagggtgcaaatctttggttcggttgagttacccaatggggaaa 751

|||||

Sbjct: 1932 aatatgaagcgcggtgacaagggtgcaaatctttggttcggttgagttacccaatggggaaa 1991

Query: 752 ttaaaccacgacattgttgacctaacagttgaggattctcgactttttgaaccattcaaaa 811

|||||

Sbjct: 1992 ttaaaccacgacattgttgacctaacagttgaggattctcgactttttgaaccattcaaaa 2051

Query: 812 atcagggttattagcatctcatgtgtgctatggctgtcggtcgtaaatgttatttttatg 871

|||||

Sbjct: 2052 acaaggtaattagcatctcatgtgtgctatggctgtcggtcgtaaatgttatttttatg 2111

Query: 872 ttgcgaaaggcgcaaaaccagtttttagcaga 902

|||||

Sbjct: 2112 ttgcgaaagggtgcaaaaccagtttttagcaga 2142

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

Identities = 126/132 (95%)

Strand = Plus / Plus

Query: 976 aaaattgactataaagagaaataaaaagccataaattcggatgctttggcggggcgatga 1035

|||||

Sbjct: 2233 aaaallgactaaaagacaataaaaaagccataatlulgalgctttggyggyggaoga 2292
Query: 1036 attacaggcctttgcaatctgctaaggcagatcaacacagggatagaataacacgcttttgg 1095
|||||
Sbjct: 2293 attacaggcctttgcaatctgctaaggcagatcaacacagggatagaataacacgcttttgg 2352
Query: 1096 aattttgaaaca 1107
|
Sbjct: 2353 acttttgaaaca 2364

Score = 143 bits (72), Expect = 9e-30
Identities = 303/380 (79%)
Strand = Plus / Plus
Query: 1292 cagcatttgcctatgtccatgtgtgctggtgtaagggtgctaagtcaatgagtcgttat 1351
|||||
Sbjct: 180 cagcatttgcctatgtccaatgtgtgctggttattcgtgctgccogttcaatgaatcggtat 239
Query: 1352 attcagcgtattgaagaactaatgcgccagaatcgccatttaagcccgctcctgatcact 1411
|||||
Sbjct: 240 attcaacgcctcgaagaataatgcgctcagaatcgcaagctaaagcccgctattgatcact 299
Query: 1412 ttgacggtaaaagaatggccctgacctacaagaacgcttcaaacacottagagcatcattt 1471
|||||
Sbjct: 300 ttgacggttaagaacggtgaagacctacaggaacgctttgaacacotcagaggtcattt 359
Query: 1472 agaacgcttttagatcgctataacgattacaagaaaaagggtcggtggttttaatcagttt 1531
|
Sbjct: 360 aagaacgcttttagacgcttaccgtgattttaagaaaaagggtcgaggtttaatcaattt 419
Query: 1532 tgtaaaattgatgggtgctttttatttcaactgaatatacatatacaatccaaaactaaagaa 1591
||
Sbjct: 420 tgcaaaattgatggcggtttttatagacogaatacacctacaacgaacaacccaacaa 479
Query: 1592 tggcatcgcataatccatatttttgcttaactcaatcaatggatagaccaggaagaattg 1651
|||||
Sbjct: 480 tggcatcgcataatccatatttttgcttagtgactgaccggattgaccaggaggaaacta 539
Query: 1652 tccgaaaactggcacgatat 1671
|
Sbjct: 540 gcagaaaactggcacgatat 559

Score = 69.9 bits (35), Expect = 1e-07
Identities = 41/43 (95%)
Strand = Plus / Plus
Query: 2130 cttgtaagggttcgaccccttagaccccaaaggggcgcaactta 2172
|||||
Sbjct: 1063 cttgtaagggttcgaccccttagaccccaaaggggcgcaactta 1105
Score = 61.9 bits (31), Expect = 3e-05
Identities = 70/83 (84%)
Strand = Plus / Plus
Query: 1907 cttccatatttggagctgtatataccgcttttgbtttttggtaaaaaatcttattacaaccta 1966
|||||
Sbjct: 796 cttccatatttggagctgtctatocgctttcgttttqgtgaaaqgtcttattacaaccta 855
Query: 1967 gaaataacaaaagacgtaaagcc 1989
||
Sbjct: 856 gagttaactaaggacgtaaagcc 878

>>>>nrruc:GI_377552564 GI|377552564|gb|JN872565.1| Acinetobacter baumannii strain DS002
plasmid pCS236, complete sequence. 0/0
Length = 2252
Identities = 226/269 (84%)
Strand = Plus / Plus
Query: 622 atgattaatttacaaggaaactctaattaatgcttttgcgtglggatgggtgtaagggtaaa 681
|||||
Sbjct: 1925 atgattaatttacaaggaaactctaattaatgcttttgcgtatggacgggtgtaagggtaaa 1984
Query: 682 gacggcaagaatatgaagcgcgtgacgaggglgcaaatctcttggttcyltggagttaccc 741
||
Sbjct: 1985 gatgggaagagatcgaagcgcgtgacgagggtacaaatacttggttcgclggaactaccc 2044
Query: 742 aatggggaaattaaacacgaacttgttgacctaacagttgaggattcgcactttttgaa 801
|||||
Sbjct: 2045 aatggagagactaaacacgaacttgttgacctaacagttgatgatgacagttgtctaccag 2104
Query: 802 cccttcaaaaatcaggttatttagcatctcatgtggtgctatgggtgctcggtcgttaattgtt 861
|||

Sbjct: 2105 ccactaaaaaataaaagtaattttctatfttctgcggtgctatggctgtagggtcgtaatgtt 2164

Query: 862 attttttatgttcgcaaaggcgcaaaacc 890
|||||
Sbjct: 2165 gttttttatgttcgaaaaggtgcaaaacc 2193

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

Query: 951 gatactattgattaactaggggacaaaaattgactataaaagacaataaaaaagccata 1010
|||||
Sbjct: 19 gatactattgattaaagtggggacaaaaattgcttaaaaaagacaataaaaaagccata 78

Query: 1011 attcggatgcttttggcggcgcatgaattacaggctttgcaatctgctaaggcagatcaac 1070
||| |||
Sbjct: 79 ttccagaagctttgcgggcgagaaaatacaggctttgcaatctgcaaacgcagatcaac 138

Query: 1071 acagggatagaataaacacgttttggaattttgaaacatag 1110
|||||
Sbjct: 139 acagggatagaataaacacgttttgcgactttgaaacatag 178

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

Query: 1589 gaatggcatccgcataattcatat 1611
|||||
Sbjct: 669 gaatggcatccgcataattcatat 691

>>>>nrmuc:GI_169147044 GI[169147044|emb|CU459139.1| Acinetobacter baumannii str. AYE plasmid
p4ABAYE, complete genome. 0/0
Length = 2726
Score = 107 bits (54), Expect = 5e-19
Identities = 87/98 (88%)
Strand = Plus / Plus

Query: 805 ttcaaaaatcaggttattagcatctcatgttggtgctatggctgtcggtcgtaatgttatt 864
|||||
Sbjct: 2570 ttcaaaaataagcttattagcatcagttgttggtgctatggctgttggtcgtaacgttatt 2629

Query: 865 ttttatgttcgcaaaggcgcaaaaccagtttttagcaga 902
|||||
Sbjct: 2630 ttttatgttcgaaaaggtgcgaaacctgttttagcaga 2667

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

Query: 1063 agatcaacacagggatagaataaacacgttttggaattttgaaacatagagcgaagctaca 1122
|||||
Sbjct: 119 agatcaacatagggatagaataaacacgttttggcattttgaaacatagatcgaagcaaca 178

Query: 1123 agaa 1126
|||
Sbjct: 179 agaa 182

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

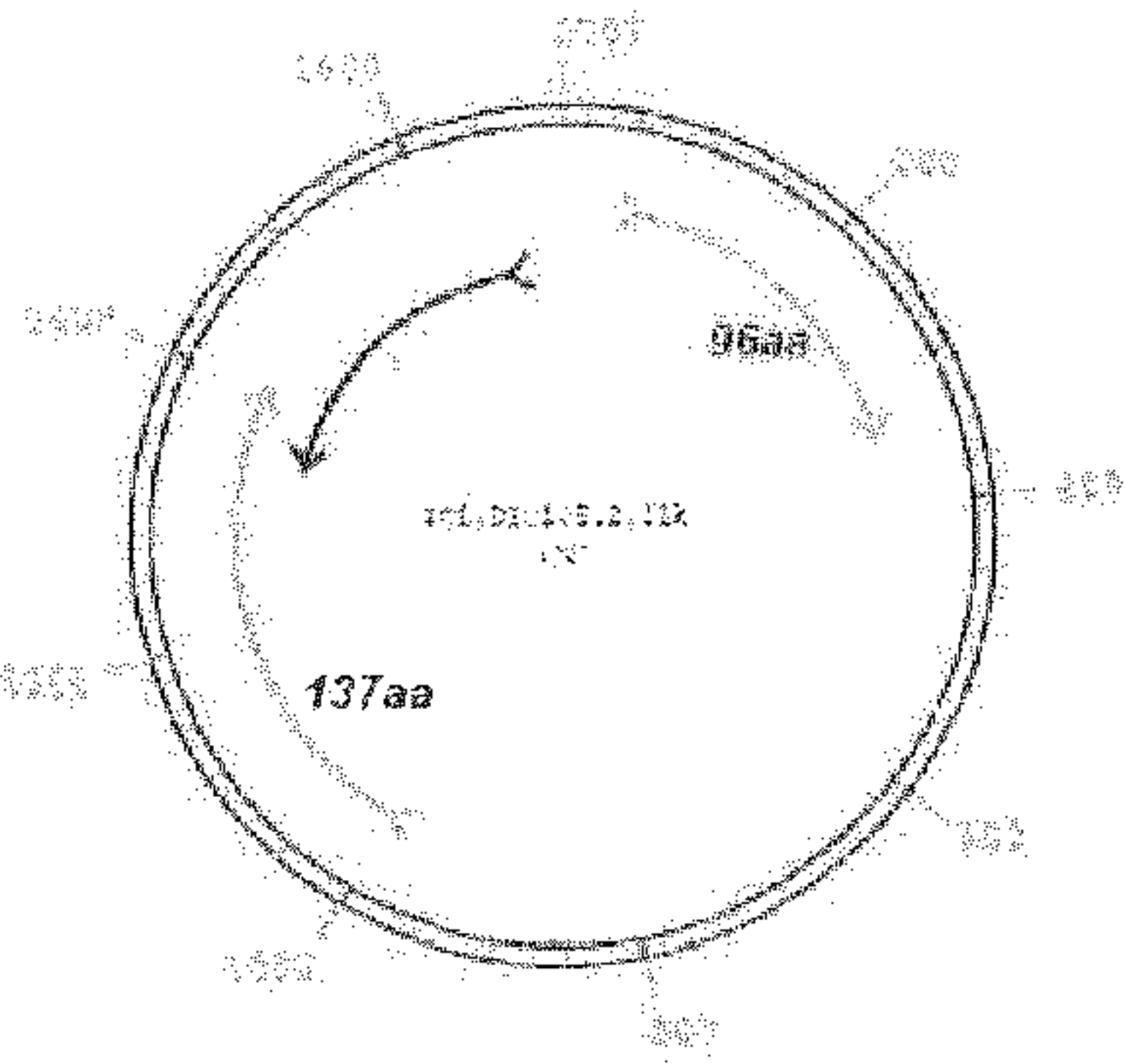
Query: 1505 aaaaaaggtcgtggttttaatcagttttgtaaaattgatggtgctttttattcaactgaa 1564
|||||
Sbjct: 573 aaaaaagggcgtggcttttaatgaaattttgcaaaattaatggtgcgatgtattcaalalgag 632

Query: 1565 tatacatacaatccaaaaactaaagaatggcatccgcataattcatatttttgc 1617
||| |||
Sbjct: 633 aataacttacaatgaaaaaactaatgaatggcatcccatattcatatggtttgc 685

Figure 1

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

GAATTCRATA TGTTCGCTTC TAAACCAATAC AACGACTTTA CTGGTTCTGG AGATTTATCA GATGATTBAA TTAGAAGGAR
TCGTTTTTAAA CGTCTTCACT CAGCAAGCTG GACAAAACAA AAAAGCGGAA TCATTTGATG ATCGTGACAA GGTACAAATT
TTAGGTGCTA TGGATCTGCC CAATGGTGAT GTAAAAAATG AGCTTTTAC GTTATCTGTA GATGATTATC GGGACTTAA
AGACTTCCTA AATCGAAAAA TTTGTATIGC TGTTCGTSCA ATGCCAAGTG CCGTAATCT TATTTTTTAT GTTGCTAAAG
GTCGAAAGCC TATATCACCA GAATTTCCGT CAAACGCTAA AAAAAACAGT GAGTTAGCCT TGGCTTCTCG CTGTTTITTC
TTTCTCTTGA TACTATIGAT TAACTAGGGG AAAAAAATTG ACTAAAAAG ACAATAAATA AGCCCATANT TCGGACACTT
TGGCGGCTGA TGAATTACAG GCTTTGCAAT CTGCTAAGGC AGATCAACAC AGGGATAGAA TATACCTTTT TGGACTTTTG
AAACATAGAG CGAAGCTACA AGAACAAAT TTTGTGGACG AGGTTGATTT CAAATCCGAA GGTGAAAATG AGACATCCAA
TAAGGCTCTT AAGGCTGCAA CCAAATTAAA GGGATGCGGT CAATTTTTCG TATTCCATAA CTACTACACA ATTGACCAAG
TTAAACTTGC TAAGGCCCAT TATTGTTCTC AGCATTITGT TTGCCCTATG TGTGCTGGTG TAAGGGCTGC TAAGTCAATG
AGTAGATATG TTCAACGTAT TGAAGAATTG ATGCGTCAGA ATCCGAAATT AAAGCCCGTA TTGATCACTT TAACGGTTAA
GAATGGGGAA GACCTAGAAG AACGCTTAA ACACTTAGAC GCTCATTTAG GACGCTTTTA SATCGTTATA ACGATTACAA
AAAGAAAGGT CGTGGTTTTA ATCAATTCTG CAGATTGAT GGTGCTTTTT TATTCCACTG AATATACCTA CAATTCAAAA
ACAAAAGAGT CGCATCCCCA TATCCATATT TTCGCTTTC TCAATGAATG GATAGACCAG GAAGGATIGG CCGAGACCTG
GCATGACATT ACCCTGGATT CTTATATCGT AGATATTCGT AGAGTTAAAA GGACCAAGA ACACGGCTAT AGCAAAGCTG
TTGCAGAGGT TTGTAATAT GCTCTTAAGT TTAGTGATTI GTCACITGAG AATACSTGGG AAGCTTATCT TCTTTAAAA
GGTAATAGGC TTACTGGCTG TTTTGGTCTT AAGTATGGTG TCAAGTTGCC TGA AAAACCTC ACAGATGATT TACCCCTTGA
TGATCTTCCA TATATGGAGC TGCTATACCG TTTTGTCTTT GGTAAAAAAT AACATTACGG CCACTTGCCA TTGCACCAAC
AGCAATACAA ATTTTTCGAT TTAGGAAGTC TTTAAAGTCC CGATAATCAT CTACAGATAA CGTAAAAAGC TCAATTTTTA
CATCACCATT GGGCAGATCC ATAGCACCTA AAATTTGTAC CTTGTCACGA TCATCAATG ATTGGCTTT TTTGTTTTGT
CCACCTTGCT GAGTGAGAC GTTAAAAAGC ATTCCTTCTA ATTTAATCAT CTGATAAATC TCCAGAACCA GTAAAGTCGT
TGTAATGGTT AGAAGCGAC 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 MINIQCTLIN AFRLDGGKKG DKEYEARDK VQILGSLELF NGEIKHELVD	50
HCBI2.170.137.pep	MIKLECIVLN VETQGGQNK AGESFDDRDK VQILGAMDLP NGDVKNELFT	
SPHINX2.36.96.PEP	51 LTVEDARIFE PFKHKVISIS CGAMAIGRNV VFYVRKGAKP VLADVM	96
HCBI2.170.137.pep	LSVDYRDFK DELNRKICIA VGAMASGRNV IFYVAKGAKP ISAEFA	

96 amino acids	53% similarity to Sphinx2.36 (96 aa)	function unknown
SPHINX2.36.232.PEP	1 MQEQYLWTQV DFKVGSETSI KALKAATRLK GCGQFLLEFN YYTIDQIKLE	50
HCBI2.170.96.pep	MQEQYLWTQV DFKVGSETSI KALKAATRLK GCGQFLLEFN YYTIDQIKLE	

```

51                                     100
SPHINX2.36.232.PEP KFHVCGQHLL CPMCAGIRAA RSMNRYIQRI EEIMRQNRKL KPVLTITLVK
HCB12.170.96.pep ~~~~~MVL...

101                                     150
SPHINX2.36.232.PEP NGEDLQERFE HLTGSFKILL QRYRDFKKKG RGFNQFCKID GGFYTTTEYTY
HCB12.170.96.pep .....FYSTTEYTY

151                                     200
SPHINX2.36.232.PEP NETTQOWHPH IHIFALVTER IDQEELAEIW HDITLDSYIV DIRRVKKTKE
HCB12.170.96.pep NSKTKAEWHPH IHIFALLNEW IDQEELAEIW HDITLDSYIV DIRRVKKTKE

201                                     250
SPHINX2.36.232.PEP HGYAKAVAEV CKYALKFSDL STEKTFQAF DP~~~~~
HCB12.170.96.pep HGYSKAVAEV CKYALKFSDL SLENTWEAYL SLKGNRLTGC FGSMYGVKLP

251                                     276
SPHINX2.36.232.PEP ~~~~~
HCB12.170.96.pep EKLTDLPLD DLPYMELLYR FVFGKK
```

BlastN2:

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

Colors for alignment score:

<=50

1000

2000

>>>nrnuc:GI_313507479 GI|313507479|gb|H0444405.1| TSE-associated...

159 1e-34

>>>nrnuc:GI_377552564 GI|377552564|gb|JN872565.1| Acinetobacter...

96 1e-15

>>>nrnuc:GI_169147044 GI|169147044|emb|CU459139.1| Acinetobacter...

80 9e-11

>>>nrnuc:GI_510026963 GI|510026963|gb|GAJD01035297.1| TSA: Ursus...

46 1.2

>>>>nrnuc:GI_313507479 GI|313507479|gb|H0444405.1| TSE-associated circular DNA isolate

Sphinx 2.36, complete sequence. 0/0

Length = 2364

Score = 159 bits (80), Expect = 1e-34

Identities = 98/104 (94%)

Strand = Plus / Plus

Query: 462 gccataattccgacactttggcgggtgattgaattacaggctttgcaatctgctaaggca 521

|||||

Subject: 2261 gccataattctgatgctttggcggggaagaattacaggctttgcaatctgctaaggca 2320

Query: 522 gatcaacacagggaatagaatatcacgttttgacttttgaaaca 565

|||||

Subject: 2321 gatcaacacagggaatagaatatcacgttttgacttttgaaaca 2364

Score = 131 bits (66), Expect = 3e-26

Identities = 336/425 (79%), Gaps = 7/425 (1%)

Strand = Plus / Plus

Query: 579 caagaacaatatttgggacgcaggttgatttcaaatccgaaggtgaaaatgagacatcc 638

|||||

Subject: 15 caagaacaatatttgggactcaggttgatttcaag-----gllggaagtgaacatct 68

Query: 639 aataaggctcttaaggctgcaaccaaattaaagggtatgcggtaatttttgcattccat 698

|||||

Subject: 69 atcaaagctcttaaggctgcaaccaaattaaagggtatgcggtaatttttgcattccat 128

Query: 699 aactactacacaattgaccaaagttaaacttgctaaggccattattgttctcagcatttg 758

|||||

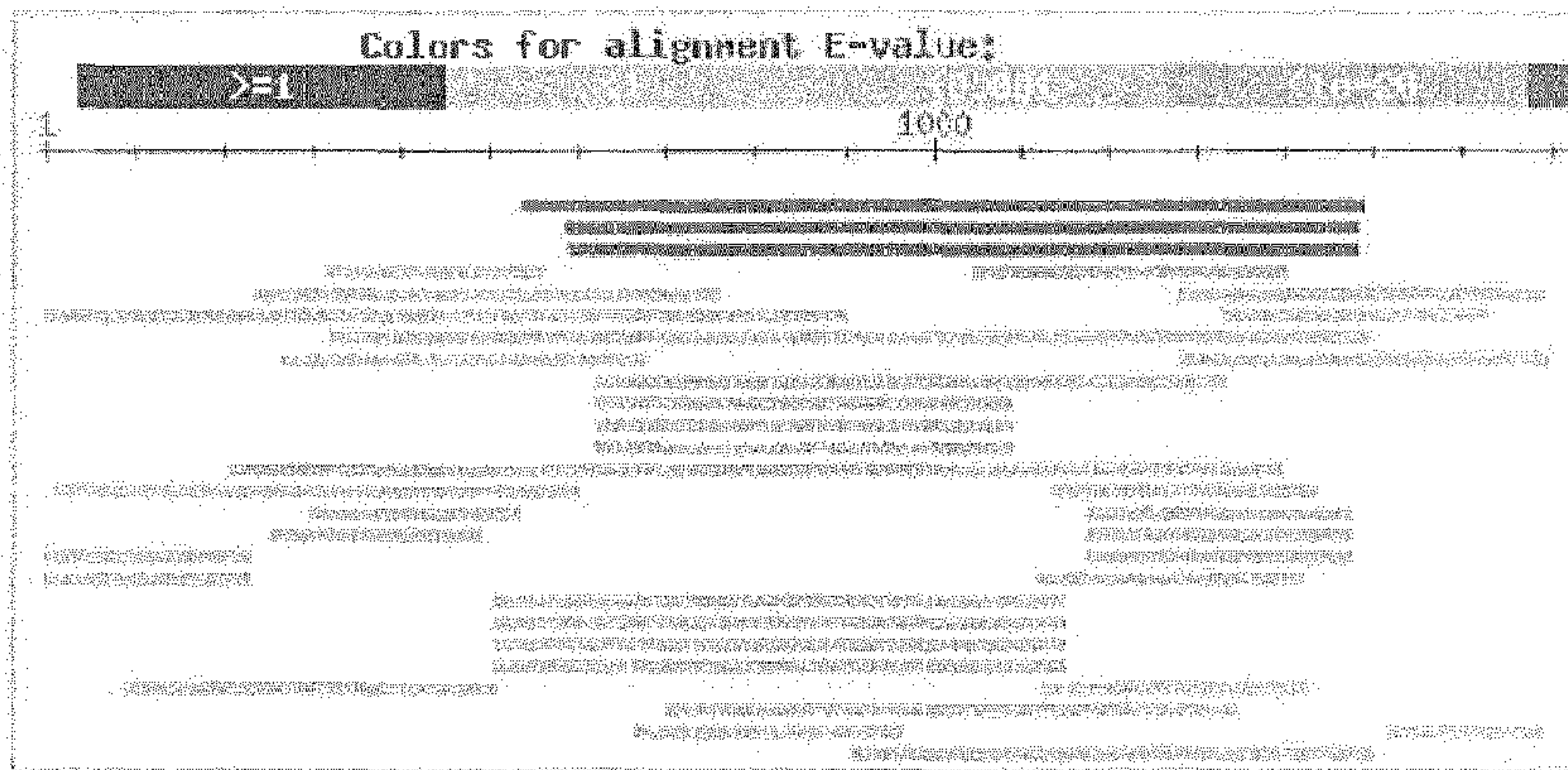
Subject: 129 aactactacacatagatcaaataagctcgaaaaattccacgtatgcggacagcatttg 188

Query: 753 ctttgccctatgtgtgctggtgtaagggtgctaagtcaatgagtagatatgttcaagct 818

|||||

Subject: 189 ctatgtccaatgtgtgctggtgtaagggtgctaagtcaatgagtagatatgttcaagct 248

Query: 819 attgaaguattgatgcgtcagaatcgcaaataaagcccgattgatcactttaacgggt 878



```
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
HCB12.170	TAGAATATCACGTTT	TTGGACTTTTGAAACATAGAGCGAAGC-	TACAAGAACATATTGT			
HQ444405						
				TAGATCTAACATGC	AAGAACAATATTGT	
				10	20	30
	600	610	620	630	640	650
HCB12.170	GGACGCAGGTTGATTTC	AAATCCGAAGGTGAAAATGAGACATCCAATAAGGCTCTTTAAGG				
HQ444405						
	GGACTCAGGTTGATTTC	-----AGGTTGGAAGTGAAACATCTATCAAAGCTCTTTAAGG				
	40	50	60	70	80	
	660	670	680	690	700	710
HCB12.170	CTGCCAACCAAAATTA	AGGGATGCGGTC	CAATTTTTTGCTATTCCATAACTACTACACAATTG			
HQ444405						
	CTGCCAACCAAAATTA	AGGGATGCGGTC	CAATTTCTGCTTTTTCGTAATTACTACACCATAG			
	90	100	110	120	130	140
	720	730	740	750	760	770
HCB12.170	ACCAAAGTTAAACTTG	CTAAGGCCCATTTATCTTCTCAGCATTTGCTTTGCCCTATGTGTG				
HQ444405						
	ATCAAATCAAGCTCG	AAAAAATTCACGATCGGACAGCATTGCTATGTC	CAATGTGTG			
	150	160	170	180	190	200
	780	790	800	810	820	830
HCB12.170	CTGGTGTAAGGGCTG	CTAAGTCAATGAGTAGATATGTTCAACGTATTG	PAGAAATTGATGC			
HQ444405						
	CTGGTATTCGTGCTG	CCCGTTCAATGAATCGGTATATTCAACGCATCG	PAGAAATAATGC			
	210	220	230	240	250	260
	840	850	860	870	880	890
HCB12.170	GTCAGAATCGCAAATTA	AGCCCSTATTGATCACTTTAACGGTTAAGAAATGGGGAAGACC				
HQ444405						
	GTCAGAATCGCAAGCT	AAAGCCCSTATTGATCACTTTGACCGTTAAGAACGGTGAAGACC				
	270	280	290	300	310	320
	900	910	920	930	940	950
HCB12.170	TAGAAGAACCCTTTA	ACA-CTTAGACCGCTCATTTAGCACCGCTTTT	TAGATCGTTATAACG			
HQ444405						
	TACAGGAACGCTTTG	AACNCTCACAGGCTCATTTAAGACGCTTTT	TACAGCGTTACCGGT			

	330	340	350	360	370	380
	960	970	980	990	1000	1010
HCBI2.170	ATTACAAAAAGAAAGGTCGTGGTTTTAAATCAATTCTGCAAGATTGATGGTGCTTTTAT					
HQ444405	ATTTTAAGAAAAAGGTCGAGGTTTAAATCAATTCTGCAAAATTGATGGCG-GTTTTAT					
	390	400	410	420	430	440
	1020	1030	1040	1050	1060	1070
HCBI2.170	TCCACTGAATATACCTACAATTCAAAAACAAAAGAGTGGCATCCCCATATCCATAATTTT					
HQ444405	ACGACCGAATACACCTACAACGAAACAACCCCAACAATGGCATCCGCATATTCAATATTTT					
	450	460	470	480	490	500
	1080	1090	1100	1110	1120	1130
HCBI2.170	GCTTTACTCAATGANTGGATAGACCAGGAAGAATTGGCCGAGACCTGGCATGACATTACC					
HQ444405	GCGTTAGTGACTGACCGGATTGACCAGGAGGAAGTACGAGAACTTGGCACCATATAACG					
	510	520	530	540	550	560
	1140	1150	1160	1170	1180	1190
HCBI2.170	CTGGATTCTTATATCGTAGATATTCGTAGAGTTAAAAGGACCAAGAACACGGCTATAGC					
HQ444405	CTTGATTCAATATCTGTGGACATCCGCGGGTCAAAAAAATAAAGAACACGGATATGCA					
	570	580	590	600	610	620
	1200	1210	1220	1230	1240	1250
HCBI2.170	AAAGCTGTTCAGAGGTTTGTAAATATGCTCTTAAGTTTAGTGATTGTGACCTGACAAT					
HQ444405	AAGGCTGTTCGCGAAGTCTGCAATATACGCTCTTAAGTTTAGCGATCTATCCACTGAGAAA					
	630	640	650	660	670	680
	1260	1270	1280	1290	1300	1310
HCBI2.170	ACGTGGGAAGC-TTATCTTCTTTTAAAGGTAATAGGCTTACTGGGTGTTTGGTTCTAT					
HQ444405	ACCTTTCAAGCATTTTTGAACCTTAAGGGCAAAAGGCTTACAGGTTCAATCGGCTCTAT					
	690	700	710	720	730	740
	1320	1330	1340	1350	1360	1370
HCBI2.170	GTATGGTGTCAAGTTGCCTGAAAAACTCACAGATGATTTACCCCTTGATGATCTTCCATA					
HQ444405	GCATGCTCTAAAAATTCTGAAACCGGACCCGATGAAATGCCTAAAGACCAACTTCCATA					
	750	760	770	780	790	800
	1380	1390	1400	1410	1420	
HCBI2.170	TATGGAGCTGCTATACCGTTTGTCTTTGGTAAAAAATAACATTAGGGCC--ACTTGGC					
HQ444405	TCTTGAGCTCCTCATCGTTTCGTTTGTGTAAGGTCTTATTACAACTAGAGTTAAC					
	810	820	830	840	850	860
	1430	1440	1450	1460	1470	1480
HCBI2.170	ATTGCACCAACAGCAATAACAATTTTTCGATTTAGGAAGTCTTTAAAGTCCCGATAATCA					
HQ444405	TAAGGACGTAAAGCCGAACAAAGGAATGAGGAAAGATGACGAGGAGCTTCGACGAGGC					
	870	880	890	900	910	920

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)

	590	600	610	620	630	640
HCBI2.170	AATATTGTG3ACGCAGGTTGATTTCAAATCCGAAGGTGAAAATGAGACATCCAATAAGG					
CU459139	CGTTAGCTAAGATTAAAGAAAATTATCATGCCGATGTAAAAACGATGAATCTATTGGG					
	200	210	220	230	240	250
	650	660	670	680	690	700
HCBI2.170	CTCTTAAGGCTGCAACCAATTAAAGGGATCCGGTCAATTTTGGTATTCCATAACTACT					
CU459139	CCATGAAACTGCCCAAAATTAATGGGTGGGTAAATTTCTCTATTCATAAAATTTT					
	260	270	280	290	300	310
	710	720	730	740	750	760
HCBI2.170	ACACATTEACCAAGTTAAACTTGCTAAGGCCCATTTATTGTTCTCAGCATTTGCTTTGCC					

CU459139	ACACCATTAAATCAAATTA	AAACTGCGCAAACTTCC	AAAGCTTGTASTGAGCA	TTTGTATTGTC		
	320	330	340	350	360	370
	770	780	790	800	810	820
HCBI2.170	CTATGTGTGCTGGTGTAA	AGGGGTGCTAAGTCAAT	AGTAGATATGTTCAACG	TATTGAAG		
CU459139	CGTTTTGCTGCTGCTATT	ACAGCTTCTAAGCGCAAT	TCAAAATACCTCTCAGC	GTGTTGATC		
	380	390	400	410	420	430
	830	840	850	860	870	880
HCBI2.170	AATTCAATGCGCTCAGAA	TGCGAAATTAAGGCCGT	TATTGATCACTTTAACG	GTTAAGAAATG		
CU459139	AAGTCTTATCTGAAAATC	CTCGTTTAAAGGCCGT	TATTGATCACGTTTACG	GTTAATAATG		
	440	450	460	470	480	490
	890	900	910	920	930	940
HCBI2.170	GGGAAGACCTAGAAAGAC	GCTTTAAACA-CTTAGAC	GCTCATTTAGCACGCTT	TTAGATC		
CU459139	GGGTAGACCTAGGGGAAC	GCTTCACCCATCTTATA	AAATCGTTTAGAACGCT	TATAGAC		
	500	510	520	530	540	550
	950	960	970	980	990	1000
HCBI2.170	GTTATAACGATTACAAAA	AGAAAGGTCGTGGTTTT	TAATCAATTCGTCAAG	ATTGATGGTG		
CU459139	GTCGTAGGGACTATATTA	AAAAAGGCGTGGCTTAA	TGAATTTTGCAAAATT	TAATGGTG		
	560	570	580	590	600	610
	1010	1020	1030	1040	1050	1060
HCBI2.170	CTTTTTTATTCCTGAA	TATACCTACAAATTC	AAAAACAAAGAGTGGC	ATCCCCATATC		
CU459139	C-GATGTATTCAATATG	AGAACTACTTACAATG	AAAAAACTAATGAATG	GCATCCTCATATT		
	620	630	640	650	660	670
	1070	1080	1090	1100	1110	1120
HCBI2.170	CATATTTTCGCTTTACT	CAATCGATACACCAGCA	GAATTTGGCCGACACCT	CGCAT		
CU459139	CATATGTTTGCATTTTG	GATGATTGGATAGATC	AGGATGAATTTGTCTC	CAATATTGGCAA		
	680	690	700	710	720	730
	1130	1140	1150	1160	1170	1180
HCBI2.170	GACATTACCGCTGGATT	CTTATATCGTAGATATT	CGTAGASTTAAAGG	ACCAAGAACAC		
CU459139	TCCATTACTGGGCACTC	TATGGTCGTTGATATT	CGTAGAGCCAAAAAC	AAAAAGACITAT		
	740	750	760	770	780	790
	1190	1200	1210	1220	1230	1240
HCBI2.170	GGCTATAGCAAGCTGTT	GCAGAGGTTGTAAATAT	GCTCTTAAGTTTAGTG	ATTGTCA		
CU459139	GGCTATTACAGTCTGCT	CGCTCAACTCTGTAATAT	CTCTCAAATTCCTCAT	CTTTCT		
	800	810	820	830	840	850
	1250	1260	1270	1280	1290	1300
HCBI2.170	CTTGAGAATACTGGGAAG	CTTATCTTTCTTAAAG	SGTAATAGGCTTAC	TGGCTGTTTT		
CU459139	GTAGAAAAAGACTTGGG	AAGCTTTCAAAGTTTT	GAAAGGTAAGCGATT	AACTGGGGCTTT		
	860	870	880	890	900	910
	1310	1320	1330	1340	1350	1360
HCBI2.170	GTTTCTATGTAGGTGTCA	AGTTGCTGAAAAACTC	ACAGATGATTTAGC	---CCTTGAT		
CU459139	GGATCTCTTTGCGGCGT	GAAATTCCTCAATCAT	TACATGATCTTCCAG	ACGATTCT		
	920	930	940	950	960	970
	1370	1380	1390	1400	1410	1420
HCBI2.170	GATCTTCCATATATGG	AGCTGCTATACCGTTT	TGCTTTGGTAAAAA	ATAACATTACGGC		
CU459139	GATTTACCTTATTTAG	AAATGATTATAAGT	TCGTCTTTCTAAGA	AGTCTATTACGAT		
	980	990	1000	1010	1020	1030
	1430	1440	1450	1460	1470	1480
HCBI2.170	CACTTGCCATTGCACCA	AGCAATACAAATTTT	TCATTTAGGAAGTCT	TTAAGTCCC		

```
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(): 8.5e-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)

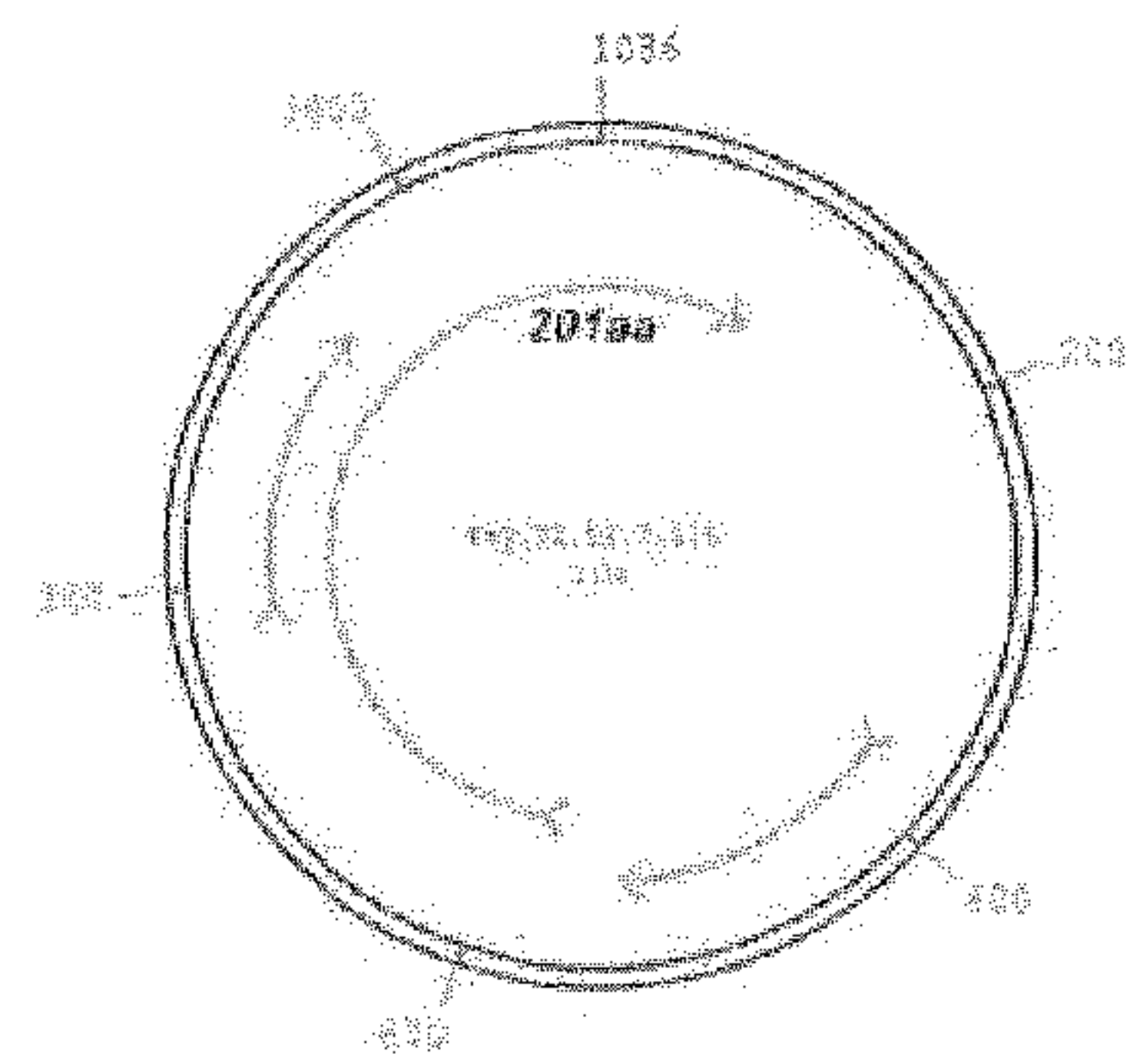
	1690	1680	1670	1660	1650	1640
HCBI2.170	GCTTCTAACCATTACAACGAC	TTTACTGGTCTCGGAGATTT	TATCAGATGATTA	AAATFAGA		
HQ444405	SCGGATCTCTATGCTTCAGG	AAAAATGAAATCAGGAGAT	CAGTCAGATGATTA	ATTTACA		
	1820	1830	1840	1850	1860	1870
	1630	1620	1610	1600	1590	1580
HCBI2.170	AGGAATCGTTTTTAAACGT	CTTCACTCAGCAAGGTGG	ACAAAACAAAAAGGCG	AAATCAIT		
HQ444405	AGGGACACTTATAAATGCT	TTTCCTCTTGATSGTGG	TAAAGGGAAGACGGCA	AAGAATA		
	1880	1890	1900	1910	1920	1930
	1570	1560	1550	1540	1530	1520
HCBI2.170	TGATGATCGTGACAAGGT	TACAAATTTTAGGTGCT	TATGGATCTGCCCAAT	GGTGATGTAA		
HQ444405	TGAAGCACGCGACAAGGT	GCAAATCTTGCTCGCT	GGAGTTGCCCAACGGT	GAGATCAA		
	1940	1950	1960	1970	1980	1990
	1510	1500	1490	1480	1470	1460
HCBI2.170	AAATCAGCTTTTACGTT	TATCTCTACATGATAT	CGCCACTTTAAACACT	TCCTAAATCG		
HQ444405	ACATCAGCTTGTTGACT	TTAACGCTTGAAGATG	CTCGCATCTTTGAGCC	ATTCABACAA		
	2000	2010	2020	2030	2040	2050
	1450	1440	1430	1420	1410	1400
ymf.D1.10E.2	AAAAATTTGTATTGCT	GTGTGGTGCAATGGCA	AGTGGCCGTAATGT	TATTTTTTACCAA	AG	
HQ444405	GGTAATTAGCATTTCA	TGCGGTGCTATCGGT	CGAAATGTTCTTTT	TATGTTG		
	2060	2070	2080	2090	2100	2110
	1390	1380	1370	1360	1350	1340
HCBI2.170	ACAAAACGGTATAGCAG	CTCCA-TATATGGAAG	ATCATCAAGGGGTAA	ATCATCTGTGAG		
HQ444405	AAAAGGTGCAAAGCCT	GTTTTAGCAGATGTA	ATGTAATTCA--TG	AAAAAAACACTGAG		
	2120	2130	2140	2150	2160	2170
	1330	1320	1310	1300	1290	1280
HCBI2.170	TTTTTCAGGCAACTTG	ACACCATACATAGA	ACCAAPACAGCCAG	TAAAGCCTATTAC	CTTT	
HQ444405	TTAGCCTTGCCGACT	CGCTGTTTTTCTTT	TACTTGATACTACT	AAAGTGGGAC		
	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)

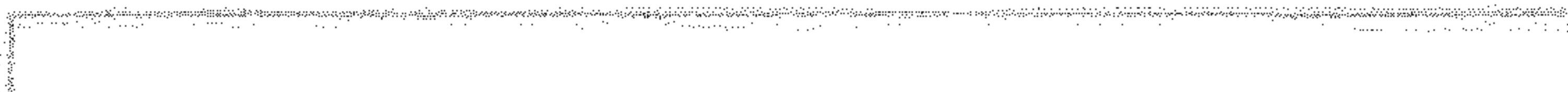
GCAGTGTCT ACGAGCGAAG TTATGAAAGT TCGATTCTTC CCCCCTCTGA AAAACCGCTT TAAAAAATAT TGGCTGCTAG
ATGGTTTTTA CTATCGTGAG CTTTTCGCTT TAAAAAACA CGAGCGAAGC GAGTTCATAG TTGCTTTTGG GGGTTTCGGG
GGGCTTGCCC CCTGAACAAG ATCACGGAGT GGGAAATTAT CACGGTAGTG AAAAAGTACC CTCTGTGTAT CCTTGCTTAT
TTCTTTTAA ACCTTTGAAC TTTTTCCTGT AATTTGAAGA AATTGCCCTT CGACTAAGCT TGCTTAGTCA AAAAAGTTTG
AGCAAAGCGA AAACATAGGG CAATTTTCAT GATGAAAATG GGCTTTTAAG GCTTTTAAAT GCTTTTAAGG CTTTLAGACA
TGCTGAAACG CAAGCCTAGC AAGGCATACA GAGGGCATT AACACCGTTT ACCTACCAAT ACCCCACCGT TTACSTACCA
ATACCCACCC GTTTACCTAC CAATACCCCA CCGTTTACCT ACCTTAATAC ACAATAATAT TTTTATGTGG TATAACGTAA
TAAATATAT AGGTGGTTTA TGAGTGATTT AATAGTAAAA GATAACGCC TAATGAACGC TAGTTATAAC TTAGACTTGG
TTGAACAGCG GTTAATTCTT TTAGCTATCC TTGAAGCTAG AGAATCAGGC AAAGGAATTA ACGCAAATGA CCTCTTACA
GTCCATGCAG AGAGTTATAT CAATCAATTT GGTGTAGCTA GACAGACTGC TTATCAAGCC CTAAAAGATG CCTGCAAAGA
TTTATTGCCC CGTCAATTCA GCTATCAAGA AAAGCGTGAA CGTGGACGAG CTAATATTAC AAGTCGTTGG GTCAGCCAAA
TTGCTTACAT TGATGAAACT GCAACGGTTG AGGTTATTTT CGCCCTGCG GTTGTTCCAC TGATCACAAG GTTAGAGGAA
CAATTCTCGA AGTACGATAT TGAACAAAT AGTAGTCTAT CGAGTGCCTA TGCAGTTCGC TTATATGAGT TATTGATCTG
CTGGAGAACA ACAGGAAAGA CACCAATTAT TGACTTAACA GAATTC



ORF 201 amino acids – putative replication protein, DNA-binding domain
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 LMNASYNLAL VEQRLILLAI LEARESGKGI NANDPLTVHA	
	51	100
SPHINX1.76.324.PEP	GSYINQFNVQ RHTAYQALKD ACKDLFARQF SYQEKREGR INITSRWVSQ	
HCBI3.108.201.pep	ESYINQFVVA RHTAYQALKD ACKDLFARQF SYQEKREGR ANITSRWVSQ	
	101	150
SPHINX1.76.324.PEP	IGYMDDTATV EIIFAPAVVP LITRLEEQFT QYDIEQISGL SSAYAVRMYE	
HCBI3.108.201.pep	IAYIDETATV EVIFAPAVVP LITRLEEQFS KYDIEQISGL SSAYAVRLYE	
	151	200
SPHINX1.76.324.PEP	LLICWRSTGK TPIELDEFER KRIGVLDTEY TRTDNLKMQV IELALKQINE	
HCBI3.108.201.pep	LLICWRSTGK TPVIDLTEFA S.....VY ER.....	
	201	250
SPHINX1.76.324.PEP	HTDITASIEQ HKKGRVITGF SFMFRHKKQN SDKTPDTNAS SPRIVKHSQI	
HCBI3.108.201.pep	SYES S.....	
	251	300
SPHINX1.76.324.PEP	PTNIVKQPEN AKMSDLEHRA SRVTGEIMRN RLSDRFRQGD ESAIDMMKRI	
HCBI3.108.201.pep	...ILPPSEK PLLKILAAEW FLLS.....	
	301	324
SPHINX1.76.324.PEP	QSEIITDAIA QWESKLEEF GVVF	
HCBI3.108.201.pep		

BLASTN2:





```
>>>nrnuc:GI_313507477 GI|313507477|gb|HQ444404.1| TSE-associated, 321 9e-84
>>>nrnuc:GI_347602264 GI|347602264|gb|LN541297.1| Uncultured bac., 291 7e-72
```

```
>>>>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
```

```
Query: 578 ttatgagtgatttaatatagtaaaagataacgcctaatgaacgctagttataacttagact 637
          ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 604 ttatgagcgatttaatatagtaaaagataacgcctaatgaatgctagttataacttagctt 663

Query: 638 tgggtgaacagcggttaattcttttagctatccttgaagctagagaaatcaggcaaaaggaa 697
          ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 664 tgggtgaacagaggttaattctattagcaatcatagaagcgagagaaacaggcaaaaggga 723

Query: 698 ttaacgcaaatgacctctttacagtcctatgcagagagtttatatcaatcaatttggtgtag 757
          ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 724 ttaatgccaatgatctctttacagttctatgcaggttagcttatatcaatcaatttaacgtac 783

Query: 758 cttagacagactgcttatcaagccctaaaagatgcttgcgaagatttatttgcccgtaaat 817
          || || || || ||||| ||||| ||||| || ||||| || ||||| |||||
Sbjct: 784 aaaagcatacggcatatcaagccctcaaaagatgcttgtaaagacttggttgcccgtaaat 843

Query: 818 tcagctatcaagaaaagcggtgaacgtggacgagetaatattacaagtcgttgggtcagac 877
          |||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 844 tcagttaccaagaaaagcgagaaagaggaacgaattaattacaagtcgatgggttttcgc 903

Query: 878 aaattggcttaccatlgatgaaacgcgaacgggtgaggttatttttgcacctgcgggttgttc 937
          ||||| || || || || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 904 aaattggcttatatggacgatacagcaaccgggtgagactatttttgcacctgcgggttgttc 963

Query: 938 cactgatcacaaagggttagagggaacaattctcgaagtacgatattgaacaaatagtagtc 997
          || ||||| || || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 964 ctctgattacaaggcttagagggaacagttcaccacagtaacgataltgagcaaatagcgggtt 1023

Query: 998 tatogagtgcttatgcagtttcg 1019
          ||||| || || |||||
Sbjct: 1024 tatogagtgcatatgctgttcg 1045
```

```
Score = 170 bits (86), Expect = 2e-38
```

```
Identities = 121/132 (91%), Gaps = 3/132 (2%)
```

```
Strand = Plus / Plus
```

```
Query: 273 ttggaagaatttgccctcgaactaagcttgcttagtcaaaaaagtttgagcaaaagcgaaa 332
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 305 ttgcaagaatttgccctcga---agcttgcttagtcaaaaaagtttgagcaaaagcgaaa 361
```

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

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

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

Query: 1 gcgagtgtctacgagcggaagttatgaaagttcgattcttccccctctgaaaaaccgctt 60
|||||
Sbjct: 26 gcgagtgtctacgagcgactcaatgaaagttcgattattccccctctgaaaaaccgctt 84

Query: 61 ttaaaaatattggctgctagatgggttttactatcgtgagcttttgcttttnnnnnnnca 120
|||||
Sbjct: 85 ttaaaaatattggctgctagatgggttttactatagtgagggttttgcttttaaaaaaaca 144

Query: 121 cgagcgaagcgagttcatagttgcttttg 149
|||||
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

Query: 273 tttgaagaaattgccctcgaactaagcttgcttagtcaaaaaagtttgagcaaagcgaaa 332
|||||
Sbjct: 869 tttgcagaaattgccctcgaactaagcttgcttagtcaaaaaagtttgagcaaagcgaaa 928

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

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

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

Query: 1 gcgagtgtctacgagcggaagttatgaaagttcgattcttccccctctgaaaaaccgctt 60
|||||
Sbjct: 590 gcgagtgtctacgagcgactcaatgaaagttcgattattccccctctgaaaaaccgctt 648

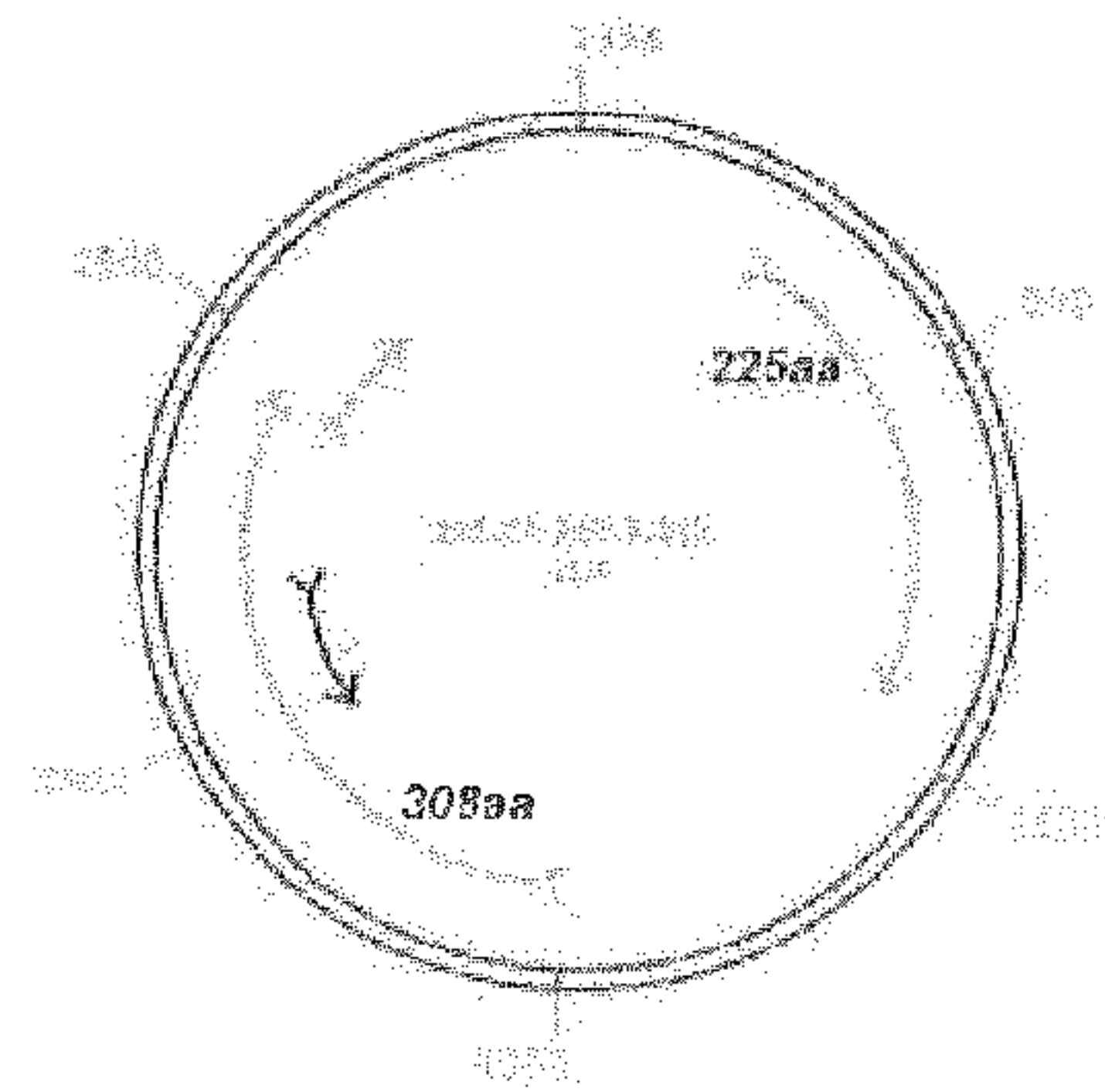
Query: 61 ttaaaaatattggctgctagatgggttttactatcgtgagcttttgcttttnnnnnnnca 120
|||||
Sbjct: 649 ttaaaaatattggctgctagatgggttttactatagtgagggttttgcttttaaaaaaaca 708

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

Figure 3

HCBI4.296 *rmf.C2.F8E.2.31k (2958bp)*

CTGAGAAGCG AAGAAAAGAT GCGGTTTTGT GTATTGAACA TCTTGTGACT GCATCACCAG AATGGGACGG CTGGGGAAC
GAAAAAGAAA CTGCATTTTT TGAACAGTCA AGGAAATGCG TTGAAAGCAA ATATGGTAAA AAAAAATGTG TCAGTACAAC
GATTCATAGA GAIGAAACAA CTCCACATTT AGTTGCGTAT GTTGTTCOCG TTGACGAAGA AACGGGACGT TTAAATGCTA
AAAAATTTAT TGCTGGATCT CGACATACAC TTTACAGAT GCAGATCGAT TTTGCAGTTG AAGTAAAGCA TTTAGGATTA
GATCGCGGGG TACAGGGGAG CAAAGCAAAA CATACGTCTA TTCAAGAATA TTATGAAAAA TTGAACAATT ATGAGAATGA
ACCAGGATAT GAAAAAGGAC TCACCTATGA AGTGCCTGAA CCTGAGTTTT TTGAATCTAA AAATGTTTTAC GGTGAGAGAG
TCGCAGAAGC TGTGGCTGCT CAGATAATTG ATCAAATTGC ACCTCGATTG GACAAFGCTA ATTTATTGGC TAGTCAAACA
AAAAAATTAA AABAAGAACT GTTAAACACT AGAAAAACGC TTGATGAAGT ACAGAAACGA GCAAAACCCT ATTTGGATAT
AATCAACGAA TATAATCATC CAAATCTTGA GAAAGAATTT AATAAGCAAG TTGCTAAATF ABAAGATAAT TTTGATTCAG
CACTTGAGCA TCATAGATTG CTAAAAAGAC AGCAAGAACA AGAAACATTT AACCAACAAC GTGAACCTCG CAATCAATTA
CACTTAGAGC AACAGCAAAA AAAACAACCTG GTTGAGCAAG AAAGGCAAGA AAAAGAACGT TTACCACTTT TAAGACGTCA
AGAATTAGAA AATCAGCGGA AAAATGAGCC TAAAAAACCT GATAATGGCA ATAATAACGA CTACTCACCC TCATAATACC
CGTTTAAACG CAAAAAAAC GCGGGTTTTA CCCCTGTTTG GCCTTTTATG CATAAAAGTA TAATTGAAAT AAAAAATGC
TCTTAGAACG CAAATGATGA GCATTTAGCG AGTGTCTACG AGCGACACAA TGAAAATTCT CCTATTCCCC CTCTGAAAAA
CTGCTTTTGC TCTTTTTTGC TTCTGGAGAG ACTTGTTAGC GAGTGTCTAC GAGCGAAGTA TTGATGCTTT TGCTCTTAA
AAAGCATGAG CATAGCGAAT GCATTATCTA TGCTTTTGAC TTTGATTTTG CTCTTGAAC GACACGAGCA ACGCGAGTGC
CATAGCCTTT GATTTTGCTT TTTTGGGCT TTTAATGTTT TTAATGCTT TTAATGCTT TTAGATAGCC TCAAGCATT
GCTATACATA TGTTCAGAG CTTATAAGA TACAGATTCC TTGCTATAAA GATACAGATT CCTTGCTATA AAGATACAGA
TTCCTTGCTA TAAAGATACA GATTCCCTGC ATTAACGTAC ATAGTCATAT AATGTATGTT TTAATACATC ATTGTATGAA
ATAACTAATA TGAAAATGG GTTAGTAGTG AAAGATAATG CGTTAATGAA TGCCAGCTAT AATTTAGAAG TAACAGAACA
GCGCTTAATA CTCTAGCAA TCATTAGTGC AAGAGAAACA GGGCAAGGGA TTACGTCAGA TAGCAAATTA GAAATACATG
CTAGTGACTA TGCAATTCAG TTCAGTGTG CAAAAGAAC AGCTTATGAG GCACTAAAAA GTGCTGTAAA TAATTTATTT
GAGCGTCAAT TTTCATTTAG AGAAGAAACA AAAAAAGGCA CTGGCATTTG ACGGTCACGA TGGGTTAGCA GAATTAATA
CATTGATGAC GCAGCAATAC TTGAAATCAC TTTTGGCCT GACGTTCTAC CATTAATCAC TAGACTTGAA GAACACTTCA
CAAGTTATCA AATCAAGCAA ATAGCACAGC TTACAAGTAA CTACGCTATC CGTTTATATG AACTTCTTAT TGCTTGGCGA
ACTACAGGCA AAGTCCCTGA GCTTGAACTA TCAGAATTTA GAAATAGATT AGGCATAGCT AGTAATGAAT ACACAGCAAT
GAACAACCTT AAAAGCCGTG TATTAGAGCC GTCTATTAAG CAGATCAATG AACACACAGA CATTACTGTG ACGTATGAAC
AGCATAAAA AGGGCGGACA ATTACAGGCT TTTTATTCAG ATTAAAGCAG AAGCAACAAG CAAAAAAAT AGAACTAAC
AGAGATCCAA ACACACCTGA CTTTTTATC AAAATGACCG ATGCTCAACG GCATTTATTG GCTAACAAAA TGCTGAAAT
GCCCCAATG GGGAAATACT CACAAGGTAC AGAAAGCTAC CAACAATTTG CTATTCGCAT CGCCGACATG CTTTTACAAC
CTGAAAAATT CAGAGAGCTT TATCCAATTT TAGAAAAAGC CGGATTCAAA GAATGATTGA AAAAGAGATA ACAAAATTTG
AGAAAGAGAT TTTGCTACAA GACAAAATCT CTCAGCTCCA AATGAACTA AAAGAATTTT CTGATCTTCA AAAAAAGCA
TATAGCGAAC GGCTTCAAAA AAGTATCGTG GGTTTAGAAA ATAGAATCTA TCGAATCAAG AAAATGCTTT ATACAACCTG
AAAGGTTGTC AGGGGTTAAG GGGAGGATC CCCTTACTCA TGAACCTTCA ATGTACCTTT AGGTGCTTTG AAAGTGGAGT
GAGTCCACTT CGCTATCGCA AAGCTCAAAA ACCTCTGCTA TCGTGTGTTT TTTCGGTATC GGATTTATCA AATGTCAATTT
GCTATTTTGC GTATTCAAAA ATTGAAATCG TTTGCAGACG TTGCGGTAG TCTTCGCAT AATTATCGCA ATCGAGAAC
GTTAAATGCA GATGATGCTC GTACTCATTT GAATGAACAT ACGCTAGATA CAAACGAAAA ATGTATGTCC GCAATCAG



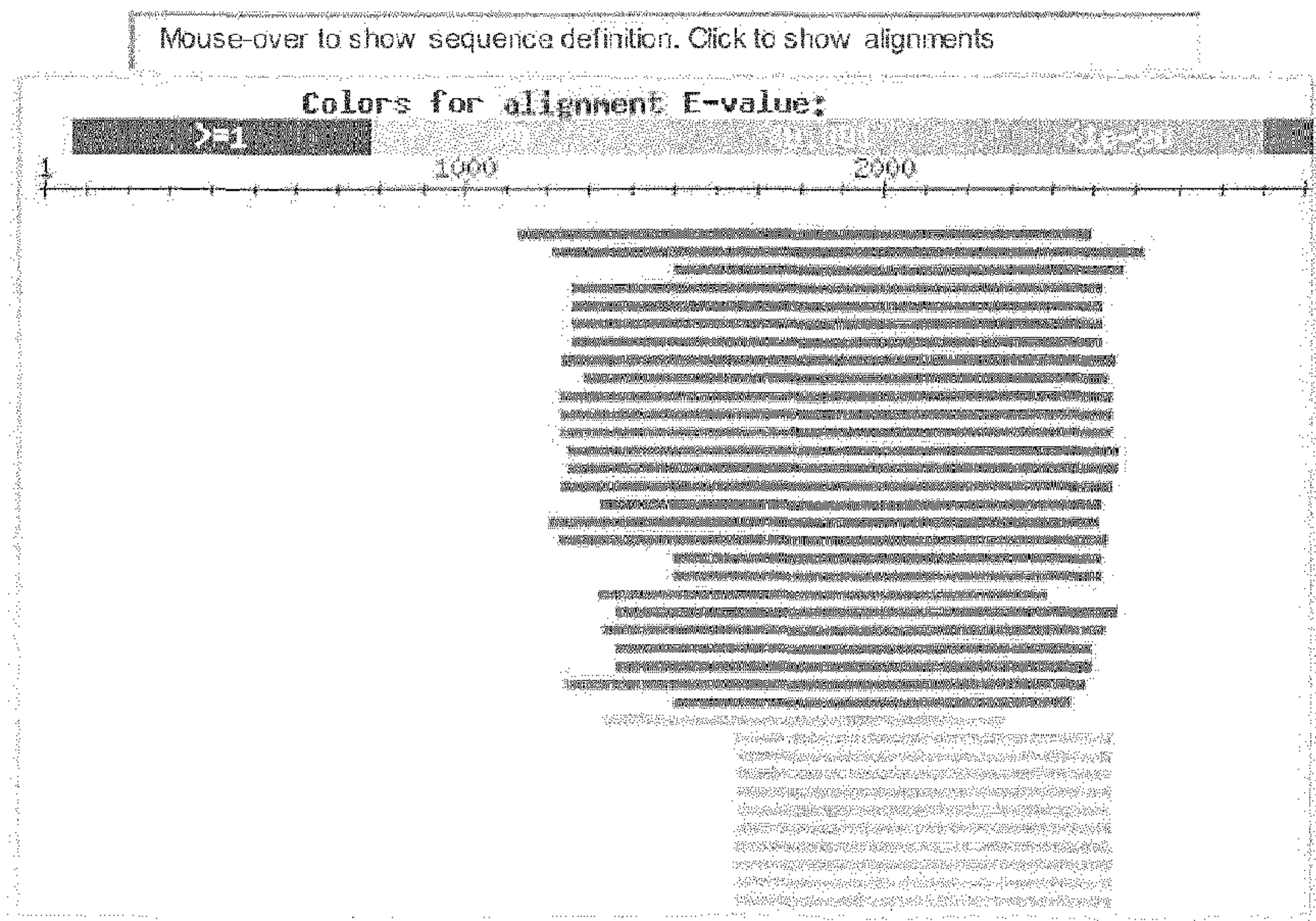
69% nucleotide similarity to *Acinetobacter baumannii* SDF plasmid p3ABSDF (24922bp) in 1361nt overlap
70% nucleotide similarity to *Acinetobacter baumannii* ATCC 17978 plasmid pAB1 (13408bp) in 964bp overlap

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

	1	50
SPHINX1.76.324.PEP	-MSDLIVKDN ALMNASYNLA IVEQRLILLA IIEARETCRG INANDPLTVH	
HCB14.296.308.pep	MKNGLVVKDN ALMNASYNLE VTEQRLILLA IISARETGCG ITSDSKLEIH	
	51	100
SPHINX1.76.324.PEP	AGSYINQFNV QRHTAYQALK DACKDLFARQ FSYQEKRRER RINITSRWVS	
HCB14.296.308.pep	ASDYAIQFSV AKETAYEALK SAVNNLFERO FS FREETKKG TGI VRSRWVS	
	101	150
SPHINX1.76.324.PEP	QICYMDDTAT VEIIFAPAVV PLITRLEEQF TQYDIEQISG LSSAYAVPMY	
HCB14.296.308.pep	RIKYIDDAAI LEITFAPDVV PLITRLEEHF TSYQIKQIAQ LISKYAIRLY	
	151	200
SPHINX1.76.324.PEP	ELLICWRSTG KPIIELDEF RKRIGVLDTE YTRTDNLKMQ VIELALKQIN	
HCB14.296.308.pep	ELLIAWRTPG KYEELSEF RNLGIASNE YTAMNFKSR VLEPSIKQIN	
	201	250
SPHINX1.76.324.PEP	ENTCITASYE QHKRGVITG FEFMFKKKQ NSDKTPDTNA SSPR.IVKHS	
HCB14.296.308.pep	ENTCITVITY QHKGRITIG FEFREKQKQ AKKIETNRDP NTFDFFIKMT	
	251	300
SPHINX1.76.324.PEP	QIPTNIVKQP ENAKMSDLEH RASRVTEIM RNRLSDRFKQ GDESAIDMMK	
HCB14.296.308.pep	DAQRHLFAN. ...KNSEMPE MG.....KYSQ GTESYQQFAI	
	301	326
SPHINX1.76.324.PEP	RIQSEIITDA IADQWESKLE EFGVVF	
HCB14.296.308.pep	RJAEMLLQPE KERETIYFTE KAGFKE	

FastA *Geall



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

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. . . .
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)

2490 2480 2470 2460 2450 2440
rmf.C2.F8E.2 AAAATCTCTTTCTCAAAATTTGTTATCTCTTTTCAATCATTCTTTGAATCCGGCTTTT
AL391558 TTCTAAGAGAAAGGTCATTTTAACTGAATGTTTAAATCTATGGTTGAAAACCAACTTTT
42120 42130 42140 42150 42160 42170

2430 2420 2410 2400 2390 2380
rmf.C2.F8E.2 CTAAAAATTGGATAAAGCTCTCTGAATTTTTCAGGTTGCTAAAAGCATGTCGGCGATGCGAA
AL391558 CTAAAAACAGGATACAGTTCTCTAAATTTTTCAGGCTCTTAAACATATTTGCGATACGTA
42180 42190 42200 42210 42220 42230

2370 2360 2350 2340 2330 2320
rmf.C2.F8E.2 TAGCAAATTGTTGGTAGCTTCTGTACCTTGTGAGTATTTCCCCATTTCGGCCATTTCAG
AL391558 TGGCAAATTGTTGATAGCTTCCGTACCTGTGATATCTCCCCATTTCGGGCATTTCAG
42240 42250 42260 42270 42280 42290

2310 2300 2290 2280 2270 2260
rmf.C2.F8E.2 ACATTTTGTAGCGAATAAATGCCGTTGAGCATCGGTCTATTTTGATAAAAAAGTCAGGTC
AL391558 ACATTTTATTTGCAAATAAATGACGTTCTGTCATCGGTCTATTTTGACAAAAAATCAGGGG
42300 42310 42320 42330 42340 42350

2250 2240 2230 2220 2210 2200
rmf.C2.F8E.2 TGTTTGGATCTCTGTTAGTTCTATTTTGTCTGCTTCTGCTTAAATCTGAATG
AL391558 TATTTGGATCAGCTTTAATTCAA---CTTTCCGCTGCTGTTTATGCTTAAATTTCAATG
42360 42370 42380 42390 42400 42410

2190 2180 2170 2160 2150 2140
rmf.C2.F8E.2 AAAAGCCTGTAATTGTCCGCCCTTTTATGCTGTTTATACGTCACAGTAACGTCGTCT
AL391558 AAAACCCGTGTGATGGTACGCCCTTGTATAGCTGTTTATAGTTGCTTTTATGCTGTAT
42420 42430 42440 42450 42460 42470

2130 2120 2110 2100 2090 2080
rmf.C2.F8E.2 GTTCATTGATCTGCTTAATAGACGGCTCTAATACACGGCTTTTAAAGTTGTTTATTGCTG
AL391558 GCTCATTTGATTTGAGTAATAGCTGTTTCTAGTACACGGCTTTTAAATGGTGCATACGTT
42480 42490 42500 42510 42520 42530

2070 2060 2050 2040 2030 2020
rmf.C2.F8E.2 TGATATTCATTACTAGCTATGCCATAATCTATTTCTAAATCTGATAGTTCAAGCTCAGGGA
AL391558 CATACTCATCATCGTCCACACCTAAGTTTTTACGTAGCTGTTGCAGTTCAAAGGTTGGCG
42540 42550 42560 42570 42580 42590

2010 2000 1990 1980 1970 1960
rmf.C2.F8E.2 CTTTGGCTGTAGTTCCGCCAAGCAATAACAAGTTCATATAAAGGGATAGCGTACTTACTTG
AL391558 TTTTTCGAACCTTCACGCCAAGCGATTAAATAATTCATAAAGTCTAGTGGCGTACTTACTCG
42600 42610 42620 42630 42640 42650

1950 1940 1930 1920 1910 1900
rmf.C2.F8E.2 TAAGCTGTGCTATTTGCTTGATTTGATTAACCTTGTGAAGTGTCTTCAAGTCTAGTGATTA
AL391558 TTAGGTGTGCAACTGTTTTCGCTCATATTTTGTAAATGCTCTTCCAATCCAGTGATTA
42660 42670 42680 42690 42700 42710

1890 1880 1870 1860 1850 1840
rmf.C2.F8E.2 ATGGTACAACGTCAGGCGCAAAAGTGATTTCAAGTATTTGCTGGGTCATCAATGATTTAA
AL391558 AAGGTACTACATCAGGTGCGAAAGTAATTTCTAGTAATGCTAAGTCATCTACATAAAAAA
42720 42730 42740 42750 42760 42770

	1830	1820	1810	1800	1790
rmf.C2.F8E.2	TTCTGCTAACCCATCGTGACCGGTACAATGCC-ACT----	CCCTTTT	TTTGT	ITCTTCTC	
AL391558	TACGGCTAACCAACGAGAGCGGAACAATACCTACTITCCCGTTC	TTTTATATT	CAGCGG		
	42780	42790	42800	42810	42820 42830
	1770	1760	1750	1740	1730
rmf.C2.F8E.2	TAAATGAAAATTGACGCTCAAATAAATTATTTACAGC	ACTTTT	TAGTGCC	TCATAAG	TG
AL391558	TATAGCTAAACTGTCTATTAAACAGATTGTTTACAG	CTTCTTT	TAGGCTT	TATAGG	ATG
	42840	42850	42860	42870	42880 42890
	1710	1700	1690	1680	1670
rmf.C2.F8E.2	TTTCTTTTGCAACACCTGA	ACTGAATTG	CATAGTCA	CTAGCATG	TATTTCT
AL391558	CATCTGCCGACACATTAA	ATAGCTTTG	CATAATCA	CTAGCATG	TATTTCT
	42900	42910	42920	42930	42940 42950
	1650	1640	1630	1620	1610
rmf.C2.F8E.2	CTGACGTAATCCCTTG	CCCTGTTTCT	CTTGCAC	TAATGATTG	CTAGAA
AL391558	CTGCCGTGATTCCCTG	CCCTAATTC	CTTGCAT	TAAATTAT	TGCCAGC
	42960	42970	42980	42990	43000 43010
	1590	1580	1570	1560	1550
rmf.C2.F8E.2	GTCTGTITACTTCTAA	ATTATAGCT	GGCATTCA	TAAACGC	ATTATCT
AL391558	GTCTGTITAGTTCTAA	ATTATAACT	TGCCTTTA	TATGCA	CTTATCT
	43020	43030	43040	43050	43060 43070
	1530	1520	1510	1500	1490
rmf.C2.F8E.2	CATTTTTCATATTAG	TAA----TTT	CATACAAT	CATGTATT	TAAAGA-
AL391558	CATTCTTCATATTAA	TACACGTT	ATTATTAT	TGTGTATT	TACACACT
	43080	43090	43100	43110	43120 43130
	1480	1470	1460	1450	1440 1430
rmf.C2.F8E.2	TATGTACCTTAATG	CAAGGAAT	CTGTATCT	TTATAG	CAAGGAAT
AL391558	TATGTGIGTTAAT	CAAGACAAT	TGTGTGCT	TATACA	AGAGAATT
	43140	43150	43160	43170	43180 43190
	1420	1410	1400	1390	1380 1370
rmf.C2.F8E.2	AGGAATCTGTATCT	TTATAG	CAAGGAAT	CTGTATCT	TTATAAG
AL391558	GAGCAATTTGTG	TGCTTAT	ACAAGGAAT	TTGTGTG	CTTATAT
	43200	43210	43220	43230	43240 43250
	1360	1350	1340	1330	1320 1310
rmf.C2.F8E.2	AGCAATGCTTT	CAGGCTAT	CTAAAAG	CATTTAA	AGCATTT
AL391558	ATCAATGCTTT	CAGCTAG	CTAAAAG	CATTTAA	AGCATTT
	43260	43270	43280	43290	43300 43310
	1300	1290	1280	1270	1260
rmf.C2.F8E.2	-AAAAGGC	CAAAATCA	AAAGGCTAT	-GGCACT	CCG-GT
AL391558	CCCCAGCGG	ACTACA	ACCGCCCT	AGCTACT	CCCTAA
	43320	43330	43340	43350	43360 43370
	1250	1240	1230	1220	1210 1200
rmf.C2.F8E.2	-AAAATCA	AAAGTC	CAAAAG	CATAGTA	ATGCATT
AL391558	TCTGCTC	CACAGCG	AGGGG	CAGTTT	TGGTTA
	43380	43390	43400	43410	43420
	1190	1180	1170	1160	1150 1140
rmf.C2.F8E.2	AAAGC-AT	CAATAC	TTCGCT	CGTAGA	CACCTC
AL391558	AAAGCAAT	TATGAAT	TTCGCT	CACCTC	ATAAACC
	43430	43440	43450	43460	43470 43480

```
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: 1919.3 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)
```

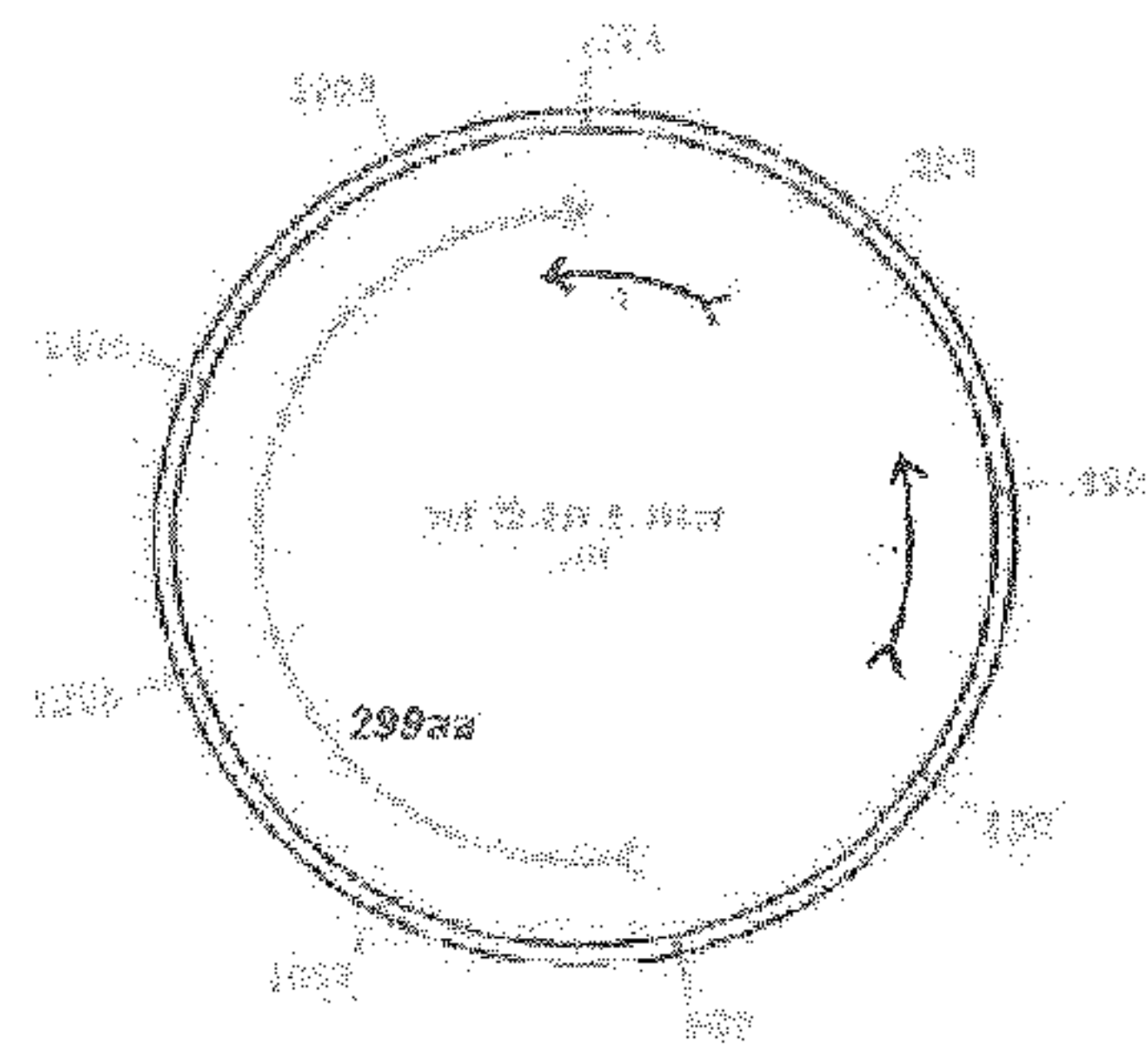
```
EMSTD_PRO:EF138630
ID   EF138630; SV 1; linear; genomic DNA; STD; PRO; 12200 BP.
AC   EF138630;
DT   12-MAY-2007 (Rel. 91, Created)
DT   27-JUN-2007 (Rel. 92, Last updated, Version 3)
DE   Acinetobacter baumannii isolate 186 plasmid pOUR, partial sequence.
KW   . . . . .
initn: 1739 initl: 570 opt: 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 pAE120, 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

HCB15.173 rmf.C2.F8E.5.23kr (1724 bp)

CCATTAGCTT TATAGTGCCA AGTAATAICT AGTAAACGGA TATGTTTTAT ATGATCTTTT CTTTGAAAT GAAGAATCAT
ATCTGAGGTC GAATAGCTGT TATATTTACT TTTTACGTAA TCAGGCATAC CGATATGATA ATGCCAAGA TTATTGCAAT
ATCTGAAATC ATAGATTGGA TCAGTTTGT CTATTCCACG CCATGAAGCT GCAATTTTAC CTTGGTATTT AGAAAAATCC
TTTAAACCAA ACTCTTCATA GATATCAGTG AAATCTAAAA TTTTATCTTG TTGATCTTCA GGATAACGCC CAAATTGAAC
TGCAAAATTGT TTTCCAAATT CCCAAGAGTA AGACAATTAC TTTTCCCTG CTTCTAATAT TAATCTTCTT TTCTCTTCTC
TGCTTAAACC TTGAGGCATA ACAAAGGTTT CTGAAICAAG AGCTGATTTC ATTCTATCAA GATTAAIACG AACAGGTGCI
GTTTTTTCAT TTGATGTTTG GGTAAATAAG AAGCCCATTA CATTCATTTC CCACCTCGAC AAATAAAAAAT TTCTTTTAAA
ATTTTAACAA AAAAAATAGC GTAGCGGGTT TTTAATTGTT TTTAAATACT TTTAATTGTT TTTAGACCTG CTGARACCCS
CTATACAAAA GGCTTTCAGG CCATAIAAAA AGACATTAC CCTGCCATAA AAAGACATTT ACCCTGCCAT AAAAAGACAT
TTACCTCGCC ATAAAAAGAC ATTACCCCTG CTTAAAAAGA CATTTTACAA AATGTCTTTT TAATGGTAGC TTTAGGTTGT
TTTTTATAAA AAAAAATGGC CTAAAAAAA TATGGATTTA GTCCTTAAAG ATAATAATTT CATCAATCCT AGCTACTCCC
TAGGTTTGGT TGAGCNAAGT CTGCTATTGC TAGCTATCAT TCAAGCTAGA GATPCTAGTA AGGGGATAGA TTCAGAAACA
TTNTTAGAAA TACATGCTCA ACACATATGCA GATAGATTTC ATGTAAATGT AAAAAATACG TATGCCATGC TTTCAGAGGC
GGCACAGACT CTTTFTAATC GTCAAETCAC TTACAIGATG GTTGATGAAA AAAGGAATAA ACCTGAAAAG CGTGTAATTC
GCTGGCTTAG TGGTATTTCA TATGTAGAAG GCGCTGAGT TCTTAAACTT CGTTTTTCTC CAGAAATTGT CCCTCTAATA
ACTAGGCTAG AGCAAATTT TACGAGTTAT GAACTAGAAC AAGTTAAAAG CTTAAACGTA TATGCAACTC GACTATATGA
ATTATTAGTC TGTGGCGTA GTACAGGTAA AACTCCCAT ATTGAAATAG AAGATTTTCG TTCAAAAATA GGAGTCTTAC
CTACTGAATA TAAATTAAAG AGTGATTTC AAAAAAGAGT TTTTGAGCCA GCATACAGC AAATTAATAA AAACACAGAT
TTGACGGTTA ATTATGATCA ACATAAATCA GGCCGTACAA TTACAGGCTT CTCATTTAAA TTTAAGCAAA AGAAAAACGAA
ATCAGAAAAG GTCSTAACCG CTAAACGAGA CCCAAATACA CCTGACTTTT TTGTAAAAAT GACCGATTCT CAACGYCAT
TATTTCGAAC TAAACTTTCT GAAATGCCAG AGATGTCTAA ATATTCTCAA GGTACAGAGT CGTATCAACA GTTTGCTATC
CCTATTGCCG 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

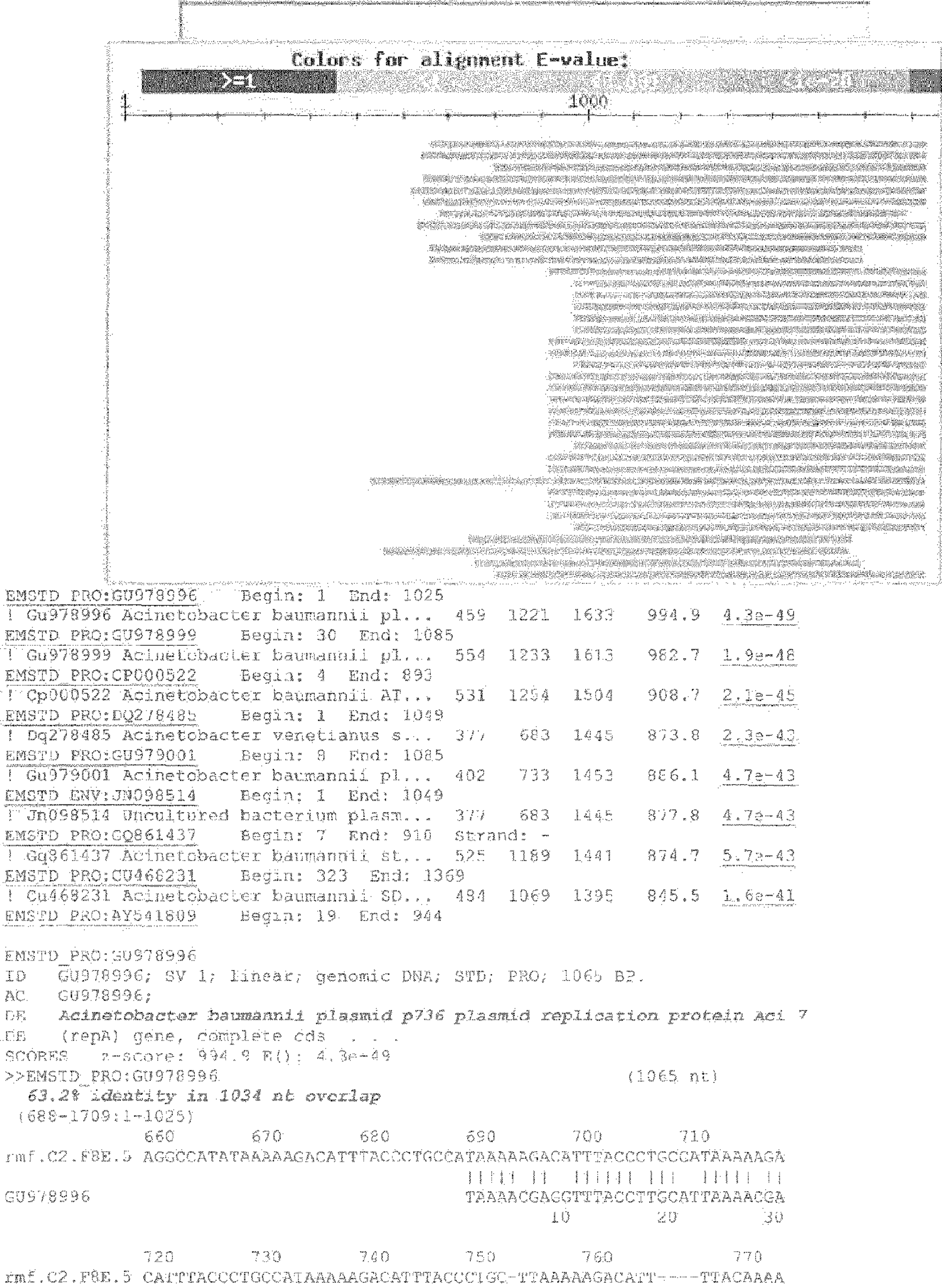
299 amino acids 49% similarity to Sphinx1.76 (324aa) replication protein

ORF:

299 amino acids	49% similarity to Sphinx1.76 (324aa)	replication protein
SPHINX1.76.324.PEP	MSDLIVKDNA LMNASYNLAL VEQRLILLAI IEARETGKGI NANDELEVIA	1
HCB15.173.299.pep	MDLVVKDNN LNASYSLGL VEQRLVILAI IEARESSKGI DSETXLEVIA	50
SPHINX1.76.324.PEP	GSYINQENVQ RHTAVQALKD ACKOLFARQF SYQEKRRER . . INITSRWV	51
HCB15.173.299.pep	QHYADRFQVN VKNTYAMLSE AAGTFENRQV TYMMVDEKRN RPEKRVIRWV	100
SPHINX1.76.324.PEP	SGISYMDDTA TVEIIFAPAV VFLITRLEEQ FTQYDIEQIS GLSSAYAVRM	101
HCB15.173.299.pep	SGISYVEGAG VLKLRFSPEI VFLITRLEQN FTSYFLEOVK SLN.VYATRL	150
SPHINX1.76.324.PEP	YELLICWRST GKTPPIELDE FRKRIGVLEF EYTRTDNLKM QVIELALKOI	151
HCB15.173.299.pep	YELLICWRST GKTPPIELDE FRKRIGVLEF EYKLMSEKSK KYLEPAIQOI	200
		201
		250

SPHINX1.76.324.PEP NEHTDITASY EQHKKGRVIT GFSEFMFKHKK QNSDKTPDTN ASSPRIVKHS
HCB15.296.310.pep NKWTDITVNY DQHKSGRTIT GFSEFKKQKE TKSEKWTAK RDP.....
251 300
SPHINX1.76.324.PEP QIPTNIVKQP ENAKMSDLEH RASRVTEGIM RNRLSDRFKQ GDESALDMMK
HCB15.296.310.pep NTPDFFVKMT D...SQRLH FATKLS... EMPKSKYSQ GTESVQOFAI
301 325
SPHINX1.76.324.PEP RIQSEIITDA IADQWESKLE EFGVVVF
HCB15.173.299.pep RTAIDMLEE...PEKIK MSH---

Fasta *Geall



```

      ||||| ||| ||||| || ||||| || ||||| || | || ||| |
GU978996 GGTTTACCTTGCATTAAAACGAGGTTTACCTTGCATTAAAACGAGGTTTACCTTGCATTA
      40      50      60      70      80      90

      780      790      800      810      820      830
rmf.C2.F8E.5 TGTCTTTTAAATGCTAGCITTAGCTTGTCTTTTATATAAAAAAATTGGCGTT--AAAAAA
      || ||||| || || || || ||||| || || || || || ||
GU978996 AG-CGACTTAATAATATAACCTCGTCTTATTAAATTAAAAACATAATCTTTTCATATGAA
      100     110     120     130     140

      840      850      860      870      880      890
rmf.C2.F8E.5 TATGGATTTAGTCGTTAAAGATAATAATTTGATCAATGCTAGCTACTCCCTAGGTTTGGT
      | || ||| || ||||| || || || || || || || || || ||
GU978996 AACAGAACTAAATAGTTAAAGATAATGCCCTTAATTAATGCCAGTTATAACCTTGATCTAGT
      150     160     170     180     190     200

      900      910      920      930      940      950
rmf.C2.F8E.5 TGAGCAACGTCITCTATTGCTAGCTATCATTTGAAGCTAGAGAAATCTAGTAAGGGGATAGA
      ||||| || || || || ||||| || || || || || || || || ||
GU978996 GGAGCAACGGTTAATTCTTTTAGCGATCCTTGAAGCAAGGGAATCGGGTAAGGAATATA
      210     220     230     240     250     260

      960      970      980      990     1000     1010
rmf.C2.F8E.5 TTCAGAAACATTNTTGAAGAAACATGCTCAACACTATGCAGATAGATTGATGTAAATCT
      || | || || || ||||| || || || || || || || || ||
GU978996 TGCTAATGATCCTTTAACAGTTCATGCTGAAAGTTATATCAATCAATTTGGTGTTCATCG
      270     280     290     300     310     320

      1020     1030     1040     1050     1060
rmf.C2.F8E.5 AAAAAATACGTATGCCATGCTTT--CAGAGGGGGCACAGACTCTTTTAAATCGTCAAGTC
      ||| | || || || || || || || || || || || || || ||
GU978996 AAATACGGCTTAT--CAAGCAATTAAAGATGCTTGTGATGATCTATTCGTAAGACAATTT
      330     340     350     360     370     380

      1070     1080     1090     1100     1110     1120
rmf.C2.F8E.5 ACTTACATCATGCTTGATGAAAAAGGAATAAACCTGAAAAGCGTGTAAATCGCTGGGTT
      |||| | || ||||| || || || || || || || || || ||
GU978996 AGTTATCAAAGCCTTAGTGAAGAAAGGAAAT--GTATTAAATCACAAATCAAGATGGGGT
      390     400     410     420     430     440

      1130     1140     1150     1160     1170     1180
rmf.C2.F8E.5 AGTGGTACTTCATATCTAGAAAGGGCTGGAGTTCTTAACTTCGTTTTCCTCCAGAAATT
      |||| | || || || || || || || || || || || || || ||
GU978996 AGTGAGGTGCTTATATTGATAACGAGGCTGTCTTAGACITATCTTGTCTCCCGCTATT
      450     460     470     480     490     500

      1190     1200     1210     1220     1230     1240
rmf.C2.F8E.5 GTCCCTCTAATAACTAGGCTAGAGCAAAATTTTACGAGTTATGAAGTAGAACAAGTTAAA
      || || || || || || || || || || || || || || || || ||
GU978996 GTGCCTTTAATTACTAGACTAGAAGACAATTACAAAGTATGAAATACAACATATAAGC
      510     520     530     540     550     560

      1250     1260     1270     1280     1290     1300
rmf.C2.F8E.5 AGCTTAA-ACGT--ATATGCCAACTCGACTATATGAATTATTAGTCTGTTGGCGTAGTACA
      | |||| | || ||||| || ||||| ||||| || ||||| |||||
GU978996 AATTTAACAAGTGTATGCGCTTATATGAAATATTAAATGTCATGGCGTAGTACC
      570     580     590     600     610     620

      1310     1320     1330     1340     1350     1360
rmf.C2.F8E.5 GGTAAAACTCCCATTTATGAAATAGAAGATTTTCGTTCAAAAATAGGAGTCTTACCTACT
      || ||||| || || || || || || || || || || || || || ||
GU978996 CCAAAAACCCCTCTCATAACCTGTACCACTTCACACAAAATAAGCTGTACTCCGATACT
      630     640     650     660     670     680

      1370     1380     1390     1400     1410     1420
rmf.C2.F8E.5 GAATATAAAATTAATGAGTGAATTTCAAAAAGCGAGTTTGTGAGCCAGCAATACACCAAAAT
      |||| | || ||||| ||||| || || || || || || || || ||
GU978996 GAATACAAACGAATGTATCATTTTAAAAAATATGTCTTCGACATTGCATTAAACAAGTC
      690     700     710     720     730     740

      1430     1440     1450     1460     1470     1480
rmf.C2.F8E.5 AATAAAAACACAGATTTGACGGTTAATZATGATCAACATAAATCAGGCCGTACAATTACA
      ||| || || || || || || || || || || || || || || || ||
GU978996 AATGAACATACCGATATTACTGTCAAAAGTTCGAACAGCATAAAGACGGCCAGATCAATTACT
      750     760     770     780     790     800

      1490     1500     1510     1520     1530     1540
rmf.C2.F8E.5 GGCCTCTCATTTAAATTTTAAACAAAAGAAAACGAAATCAGAAAAGGTCGTAACCGCTAAA
      |||| | ||||| ||||| || || || || || || || || || ||

```

```

GU978996      GGCTTTTCATTTAGCTTTAAACAAAAAAATCAGCTACTCAGTCTGTCCGA---TCTAAA
                810          820          830          840          850          860

                1550          1560          1570          1580          1590          1600
rmf.C2.F8E.5  CGAGACCCCAAATACACCTGACTTTTTTGTAAAAATGACCGATTCTCAACGTCATTTTATTC
                ||| ||||| ||| ||| || ||||| ||| ||||| ||||| ||||
GU978996      AGAGATCCAAATACATTGGACCCCTTTTCAACAATGACAGATAAACACGTCATCTATTC
                870          880          890          900          910          920

                1610          1620          1630          1640          1650          1660
rmf.C2.F8E.5  GCAACTAAACTTTCTGAAATGCCAGAGATGTCTAAATATTCTCAAGGTACAGAGTCGTAT
                ||| ||||| ||| ||| || ||||| ||||| ||||| ||| |||
GU978996      GCTAGTAAACTCTCCGAGCTTCCTGAGATGAGTAAATATTCACAAGGTACGGAAAGCTAT
                930          940          950          960          970          980

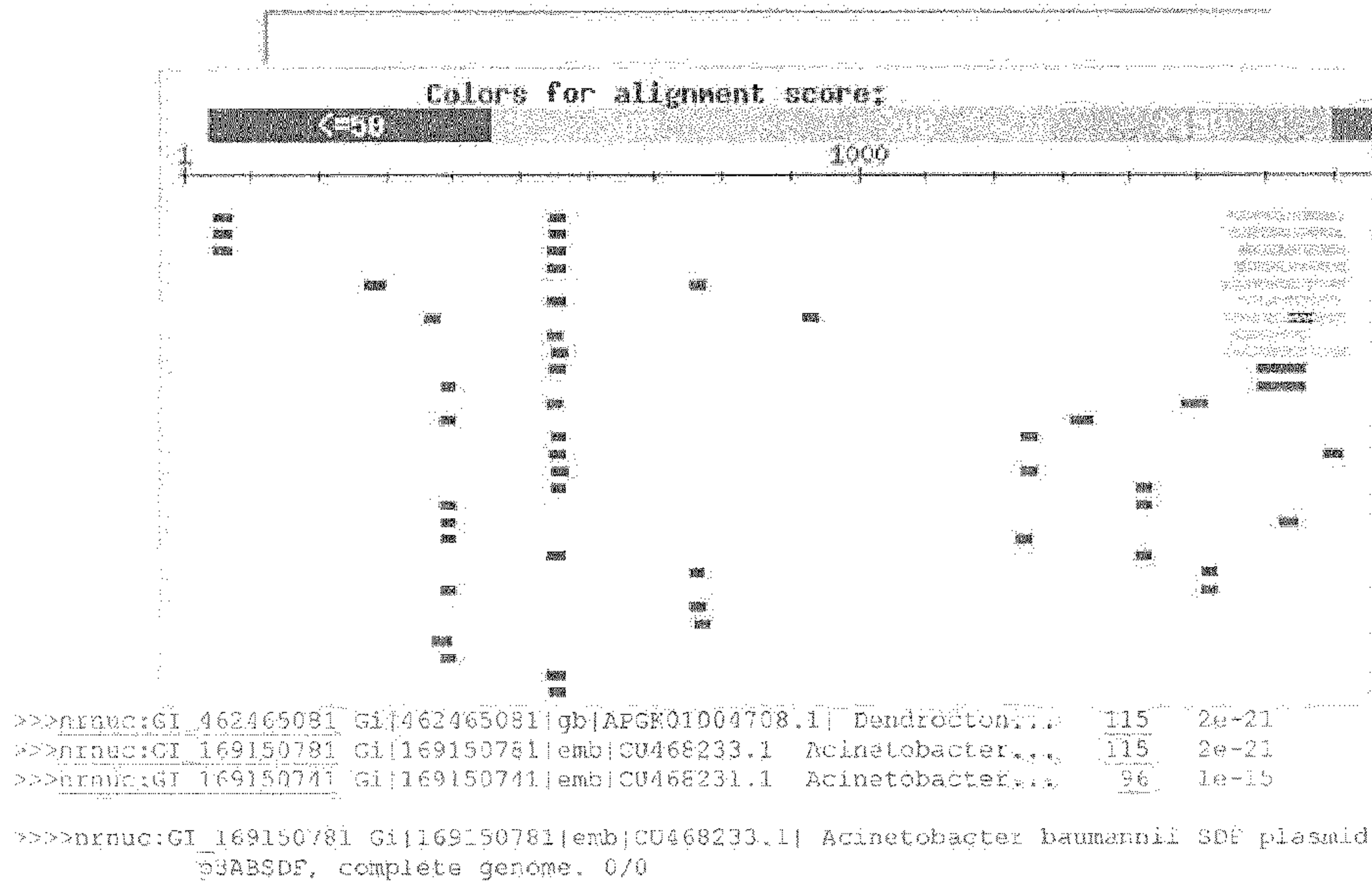
                1670          1680          1690          1700          1710          1720
rmf.C2.F8E.5  CAACAGTTTGTCTATCCGTATTGCCGATATGCTTTTAGAACCTGAAAACCTAAAGAAATTCG
                || ||||| ||||| ||| ||| ||| ||| |||
GU978996      CAGCAGITTTGCTGTACGTATCGCTGGCATGCTGCAAGATACACAGCGATTTAGGGAAATT
                990          1000          1010          1020          1030          1040

```

```
EMSTD_PRO:GU978999
ID   GU978999; SV 1; linear; genomic DNA; STD; PRO; 1125 BP.
AC   GU978999;
DE   Acinetobacter baumannii plasmid p537 plasmid replication protein Aci 5
DE   (repA) gene, complete cds. . . .
SCORES   z-score: 982.7 E(): 1.9e-48
>>EMSTD_PRO:GU978999                                     (1125 nt)
      63.4% identity in 1069 nt overlap
      (673-1723:30-1085)
```

```
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: 908.7 E(): 2.1e-45
>>EMSTD_PRO:CP000522                                     (13408 nt)
      64.6% identity in 905 nt overlap
      (829-1723:4-893)
```

BLASTN2 :



```

      Length = 24922
      Score = 115 bits (58), Expect = 2e-21
      Identities = 133/158 (84%)
      Strand = Plus / Plus
Query: 1555 aatacscctgacttttttgtaaaaatgaccgattctcaacgctcatttattcgcaactaaa 1614
      ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 1283 aatacacctgacttttttgtcaaaatgactgatgcccaacgctcatttatttgccaataaa 1342

Query: 1615 cttttctgaaatgccagagatgtctaaatattctcaagggtacagagtcgtatcaacagttt 1674
      | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 1343 atgtctgaaatgacctgatgatgagtaaatattcgcaagggaacagaaagctaccaacaattt 1402

Query: 1675 gctatccgtattgcccagatagctttttagaacctgaaaa 1712
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 1403 gctattcgctatcgctgatagcttttagagcctgaaaa 1440

      Score = 97.6 bits (49), Expect = 4e-16
      Identities = 133/161 (82%)
      Strand = Plus / Plus
Query: 1549 gacccaaatcacacctgacttttttgtaaaaatgaccgattctcaacgctcatttattcgca 1608
      ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 7634 gacccaaacacacccgactttctttataaaaaatgactgatgcacaacgccatctattcgcc 7693

Query: 1609 actaaactttctgaaatgccagagatgtctaaatattctcaagggtacagagtcgtatcaa 1668
      | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 7694 aataaaatgtctgaaatgacctgaaatgattaaatattctcaagggtacagaaagctatcaa 7753

Query: 1669 cagttttgctatccgtattgcccagatagctttttagaacctga 1709
      |||||| | || || ||||| ||||| ||||| ||||| |||||
Sbjct: 7754 cagttttacaattcgctattgctgatagcttttacaacctga 7794

>>>>nrmuc:GI_169150741 GI|169150741|emb|CU468231.1| Acinetobacter baumannii SDF plasmid
      p1ABSDF, complete genome. 0/0
      Length = 6106
      Score = 95.6 bits (48), Expect = 1e-15
      Identities = 126/152 (82%)
      Strand = Plus / Plus
Query: 1564 gacttttttgtaaaaatgaccgattctcaacgctcatttattcgcaactaaactttctgaa 1623
      |||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 1216 gacttttttgtcaaaatgaccgatgcacaacgccatctattcgccaataaaaatgtctgag 1275

Query: 1624 atgccagagatgtctaaatattctcaagggtacagagtcgtatcaacagtttgcctatccgt 1683
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 1276 atgcctgaaatgagcaaatattcacaacggcacagaaagctatcaacagtttgcctatccgt 1335

Query: 1684 attgcccagatagctttttagaacctgaaaagtt 1715
      || || || ||||| ||||| ||||| |||||
Sbjct: 1336 atcgctgacatgcttttagagcctgaaaagtt 1367

>>>>nrmuc:GI_169150750 GI|169150750|emb|CU468232.1| Acinetobacter baumannii SDF plasmid
      p2ABSDF, complete genome. 0/0
      Length = 25014
      Score = 89.7 bits (45), Expect = 9e-14
      Identities = 126/153 (82%)
      Strand = Plus / Plus
Query: 1563 tgacttttttgtaaaaatgaccgattctcaacgctcatttattcgcaactaaactttctga 1622
      |||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 11463 tgacttttttgtcaaaatgaccgatgcacaacgccatctattcgccaataaaaatgtctga 11522

Query: 1623 aatgccagagatgtctaaatattctcaagggtacagagtcgtatcaacagtttgcctatccg 1682
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 11523 gatgcctgaaatgagcaaatattcacaagggcacagaaagctatcaacagttttctatccg 11582

Query: 1683 tattgcccagatagctttttagaacctgaaaagtt 1715
      || || || ||||| ||||| ||||| |||||
Sbjct: 11583 tatcgctgacatgcttttagagcctgaaaagtt 11615

```

Figure 5