

(19) **DANMARK**

(10) **DK/EP 2336362 T3**



Patent- og
Varemærkestyrelsen

(12) Oversættelse af
europæisk patentskrift

-
- (51) Int.Cl.: **A 23 C 9/12 (2006.01)** **C 12 N 1/21 (2006.01)** **C 12 N 15/74 (2006.01)**
- (45) Oversættelsen bekendtgjort den: **2019-01-21**
- (80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2018-09-19**
- (86) Europæisk ansøgning nr.: **10169298.6**
- (86) Europæisk indleveringsdag: **2006-08-25**
- (87) Den europæiske ansøgnings publiceringsdag: **2011-06-22**
- (30) Prioritet: **2005-08-26 US 711396 P** **2006-05-19 US 747683 P**
- (62) Stamansøgningsnr: **06802307.6**
- (84) Designerede stater: **AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU LV MC NL PL PT RO SE SI SK TR**
- (73) Patenthaver: **DUPONT NUTRITION BIOSCIENCES ApS, Langebrogade 1, 1411 København K, Danmark**
- (72) Opfinder: **BARRANGOU, Rodolphe, 4818 Valor Way, Madison, WI 53718, USA**
ROMERO, Dennis, 126 Amanda Court, Oregon, WI 53575, USA
Horvath, Philippe, 8, Rue Des Sabions, 86140 Scorbe-Clairvaux, Frankrig
FREMAUX, Christophe, 19 rue des Joncs, 86000 Poitiers, Frankrig
Boyaval, Patrick, Les Ruisseles, 35520 La Meziere, Frankrig
- (74) Fuldmægtig i Danmark: **RWS Group, Europa House, Chiltern Park, Chiltern Hill, Chalfont St Peter, Bucks SL9 9FG, Storbritannien**
- (54) Benævnelse: **ANVENDELSE AF CRISPR-ASSOCIEREDE GENER (CAS)**
- (56) Fremdragne publikationer:
VISCARDI M ET AL: "Selection of bacteriophage-resistant mutants of Streptococcus thermophilus.", JOURNAL OF MICROBIOLOGICAL METHODS, vol. 55, no. 1, October 2003 (2003-10), pages 109-119, XP002428070, ISSN: 0167-7012
COFFEY AIDAN ET AL: "Bacteriophage-resistance systems in dairy starter strains: Molecular analysis to application", ANTONIE VAN LEEUWENHOEK, vol. 82, no. 1-2, August 2002 (2002-08), pages 303-321, XP002424878, ISSN: 0003-6072
BOLOTIN ALEXANDER ET AL: "Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin.", MICROBIOLOGY (READING, ENGLAND) AUG 2005, vol. 151, no. Pt 8, 1 August 2005 (2005-08-01), pages 2551-2561, XP002424872, ISSN: 1350-0872
BARRANGOU RODOLPHE ET AL: "CRISPR provides acquired resistance against viruses in prokaryotes.", SCIENCE (NEW YORK, N.Y.) 23 MAR 2007, vol. 315, no. 5819, 23 March 2007 (2007-03-23), pages 1709-1712, XP002428071, ISSN: 1095-9203
SHIRAZ A. SHAH ET AL: "Protospacer recognition motifs", RNA BIOLOGY, vol. 10, no. 5, 1 May 2013 (2013-05-01), pages 891-899, XP055153088, ISSN: 1547-6286, DOI: 10.4161/rna.23764

DESCRIPTION

FIELD OF INVENTION

[0001] The present invention relates to *inter alia* modulating the resistance of a cell against a target nucleic acid or a transcription product thereof. In particular, the present invention relates, in one aspect, to a method for conferring resistance to a bacterial cell using a functional combination of one or more *cas* genes or proteins and two or more CRISPR repeats from the same CRISPR locus within a genome or plasmid against a target nucleic acid or a transcription product thereof.

BACKGROUND TO THE INVENTION

[0002] Cultures - such as starter cultures - are used extensively in the food industry in the manufacture of fermented products including milk products (such as yoghurt, butter and cheese), meat products, bakery products, wine and vegetable products. The preparation of cultures is labour intensive, occupying much space and equipment, and there is a considerable risk of contamination with spoilage bacteria and/or phages during the step of propagation. The failure of bacterial cultures by bacteriophage (phage) infection and multiplication is a major problem with the industrial use of bacterial cultures. There are many different types of phages with varying mechanisms to attack bacteria. Moreover, new strains of bacteriophages appear.

[0003] Strategies used in industry to minimise bacteriophage infection, and thus failure of a bacterial culture, include the use of: (i) mixed starter cultures; and (ii) the alternate use of strains having different phage susceptibility profiles (strain rotation).

1. (i) Traditionally, starter cultures in the dairy industry are mixtures of lactic acid bacterial strains. The complex composition of mixed starter cultures ensures that a certain level of resistance to phage attack is present. However, repeated sub-culturing of mixed strain cultures leads to unpredictable changes in the distribution of individual strains and eventually undesired strain dominance. This in turn may lead to increased susceptibility to phage attack and risk of fermentation failures.
2. (ii) The rotation of selected bacterial strains which are sensitive to different phages is another approach to limit phage development. However, it is difficult and cumbersome to identify and select a sufficient number of strains having different phage type profiles to provide an efficient and reliable rotation program. In addition, the continuous use of strains requires careful monitoring for new infectious phages and the need to quickly substitute a strain which is infected by the new bacteriophage by a resistant strain. In manufacturing plants where large quantities of bulk starter cultures are made ahead of time, such a quick response is usually not possible.

[0004] Several attempts have been made to improve the resistance of cultures for use in industry.

[0005] Pedersen et al (7th symposium on lactic acid bacteria: genetics, metabolism and applications, September 1-5, 2002, Egmond aan Zee, The Netherlands) teach a phage resistant *Lactococcus lactis* strain, which has no thymidylate synthase activity and which requires thymidine for DNA replication.

[0006] WO 01/14520 discloses a lactic acid bacterium which have a reduced susceptibility towards attack by at least one type of bacteriophage. Said lactic acid bacteria comprise a mutated gene involved in pyrimidine metabolism, namely pyrG which results in a defect in CTP-synthetase.

[0007] Kosuge et al (1998 - Appl. Environ. Microbiol., Volume: 64, Issue: 11, Page(s): 4328-4332) and Kosuge et al (1994 - FEMS Microbiology Letters, 123 (1/2) 55-62) teach a *Thermus thermophilus* HB27 bacterium which is mutated in the proB gene and is unable to utilise proline for growth.

[0008] However, there is a continuing need to improve cultures for use in industry.

SUMMARY OF THE INVENTION

[0009] There is described herein the use of CRISPR loci or a component thereof for modulating the resistance of a cell against a target nucleic acid or a transcription product thereof.

[0010] CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats) (also known as SPIDRs - SPacer Interspersed Direct Repeats) constitute a family of recently described DNA loci. CRISPR loci consist of short and highly conserved DNA repeats (typically 24 to 40 bp, repeated from 1 to 140 times) which are partially palindromic. The repeated sequences (usually specific to a species) are interspaced by variable sequences of constant length (typically 20 to 58 bp depending on the CRISPR). Up to 20 distinct CRISPR loci have been found within a single chromosome.

[0011] Although the biological function of CRISPR loci is unknown some hypotheses have been proposed. For example, it has been proposed that they may be involved in the attachment of the chromosome to a cellular structure, or in the chromosome replication and replicon partitioning (Jansen *et al.*, 2002; Pourcel *et al.*, 2005). Moreover, Mojica *et al.* 2005 hypothesis that CRISPR could be involved in conferring specific immunity against foreign DNA and Pourcel *et al.* (2005) hypothesise that CRISPRs are structures that are able to take up pieces of foreign DNA as part of a defence mechanism. Bolotin *et al.* (2005) suggest that the CRISPR spacer elements are the traces of past invasions by extrachromosomal elements, and hypothesise

that they provide a cell with immunity against phage infection, and more generally foreign DNA expression, by coding an anti-sense RNA. Bolotin *et al.* (2005) also suggest that *cas* genes are necessary for CRISPR formation.

[0012] In contrast to the teachings of the prior art which hypothesise that CRISPR or CRISPR spacers could be involved in conferring specific immunity, the present invention is based, in part, on the surprising finding that *cas* genes or proteins are required for immunity against a target nucleic acid or a transcription product thereof.

[0013] Even more surprisingly, the inventors have discovered that one or more *cas* genes or proteins are associated with two or more CRISPR repeats within CRISPR loci. In other words, *cas* genes or proteins seem to be specific for a given DNA CRISPR repeat, meaning that *cas* genes or proteins and the repeated sequence form a functional pair. Accordingly, one or more CRISPR spacers may be used together with one or more of these functional pairs (i.e. CRISPR repeats and *cas* genes) in order to modulate the resistance of a cell against a target nucleic acid or a transcription product thereof.

[0014] In one embodiment, for one or more CRISPR spacers to confer immunity to the cell, the CRISPR repeat(s) and the *cas* gene(s) or proteins form a functional combination *ie.* the CRISPR repeat(s) and the *cas* gene(s) or proteins are compatible.

[0015] Accordingly, we suggest here for the first time that a *cas* gene or protein influences resistance - such as the resistance of a bacteria to one or more bacteriophages. In particular, the knowledge of two or more CRISPR repeats and/or one or more *cas* genes or proteins for a given cell will be an advantage to predict, determine and modify its resistance, for example, its lysotype, which defines the resistance/sensitivity of a given bacterium to various bacteriophages. Consequently, identification and detection of CRISPR loci in, for example, cells and bacteriophages could help to determine, predict and modify the resistance profile of a cell or phage-host interactions.

[0016] Advantageously, the application of one or more CRISPR loci, two or more CRISPR repeats, one or more *cas* genes or proteins and/or one or more CRISPR spacers in genetic engineering could lead to resistant or sensitive variants of cells for use within a wide variety of applications in the biotechnology industry.

SUMMARY ASPECTS OF THE PRESENT INVENTION

[0017] In one aspect there is provided a method for conferring resistance to a bacterial cell comprising one or more *cas* genes or proteins and two or more CRISPR repeats against a target nucleic acid or a transcription product thereof; wherein said one or more *cas* genes or proteins and two or more CRISPR repeats form a functional combination; further wherein the at least one or more *cas* genes or proteins and the two or more CRISPR repeats are from the same CRISPR locus within a genome or plasmid; comprising the steps of: (i) identifying one or

more CRISPR spacer(s) in an organism resistant to the target nucleic acid or transcription product thereof; (ii) inserting the sequence of the one or more CRISPR spacer(s) into the bacterial cell; such that the cell becomes substantially resistant to the target nucleic acid or transcription product thereof; and (iii) determining if the cell modulates resistance against the target nucleic acid or transcription product thereof, wherein modulation of the resistance of the cell against the target nucleic acid or transcription product thereof is indicative that the CRISPR spacer modulates the resistance of the cell.

[0018] In a second aspect there is provided a method for conferring resistance to a bacterial cell against a target nucleic acid or a transcription product thereof, comprising the steps of: (i) identifying one or more CRISPR spacer(s) in an organism resistant to the target nucleic acid or transcription product thereof; (ii) preparing a recombinant nucleic acid comprising at least one *cas* gene and at least two CRISPR repeats, together with said identified one or more CRISPR spacer(s) wherein the at least one *cas* gene and the at least two or more CRISPR repeats form a functional combination; further wherein the at least one *cas* gene and the at least two or more CRISPR repeats are from the same CRISPR locus within a genome or plasmid; (iii) introducing said recombinant nucleic acid into the cell such that the cell becomes substantially resistant to the target nucleic acid or transcription product thereof; and (iv) determining if the cell modulates resistance against the target nucleic acid or transcription product thereof, wherein modulation of the resistance of the cell against the target nucleic acid or transcription product thereof is indicative that the CRISPR spacer modulates the resistance of the cell.

[0019] The use of one or more *cas* genes or proteins for modulating resistance in a cell against a target nucleic acid or a transcription product thereof is disclosed herein.

[0020] The use of a recombinant nucleic acid sequence comprising at least one *cas* gene and at least two CRISPR repeats together with at least one CRISPR spacer, wherein at least one CRISPR spacer is heterologous to at least one *cas* gene and/or at least two CRISPR repeats to modulate resistance against a target nucleic acid or transcription product thereof is disclosed herein.

[0021] A nucleic acid sequence consisting essentially of at least one *cas* gene is disclosed herein.

[0022] A nucleic acid sequence consisting essentially of at least one *cas* gene and at least two CRISPR repeats is disclosed herein.

[0023] A nucleic acid sequence consisting essentially of at least one *cas* gene and at least one CRISPR spacer is disclosed herein.

[0024] A nucleic acid sequence consisting essentially of at least one *cas* gene, at least one CRISPR spacer and at least two CRISPR repeats is disclosed herein.

[0025] A recombinant nucleic acid sequence comprising at least one *cas* gene and at least two

CRISPR repeats together with at least one CRISPR spacer, wherein the CRISPR spacer is heterologous to the at least one *cas* gene and/or the at least two CRISPR repeats is disclosed herein.

[0026] A construct comprising one or more of the nucleic acid sequences described herein is disclosed herein.

[0027] A vector comprising one or more of the nucleic acid sequences or one or more of the constructs described herein is disclosed herein.

[0028] A cell comprising the nucleic acid sequence or the construct or the vector described herein is disclosed herein.

[0029] A method for modulating (e.g. conferring or increasing) the resistance of a cell against a target nucleic acid or a transcription product thereof comprising the steps of: (i) identifying a sequence (eg. a conserved sequence) in an organism (preferably, a sequence essential to the function or survival of the organism); (ii) preparing a CRISPR spacer which is homologous to the identified sequence; (iii) preparing a nucleic acid (eg. a recombinant nucleic acid) comprising at least one *cas* gene and at least two CRISPR repeats together with the CRISPR spacer; and (iv) introducing said nucleic acid into a cell thus to render the cell resistant to said target nucleic acid or transcription product thereof is disclosed herein.

[0030] A method for modulating (eg. conferring or increasing) the resistance of a cell against a target nucleic acid or a transcription product thereof comprising the steps of: (i) identifying one or more CRISPR spacers or pseudo CRISPR spacers in an organism resistant to the target nucleic acid or transcription product thereof; (ii) preparing a recombinant nucleic acid comprising at least one *cas* gene or protein and at least two CRISPR repeats together with said identified one or more spacers; and (iii) introducing said recombinant nucleic acid into a cell thus to render the cell resistant to said target nucleic acid or transcription product thereof is disclosed herein.

[0031] A method for modulating (eg. conferring or increasing) the resistance of a cell comprising at least one or more *cas* genes or proteins and two or more CRISPR repeats against a target nucleic acid or a transcription product thereof comprising the steps of: (i) identifying one or more CRISPR spacers in an organism resistant to the target nucleic acid or transcription product thereof; and (ii) modifying the sequence of one or more CRISPR spacer(s) in the cell such that the CRISPR spacer(s) has homology to the CRISPR spacer(s) in the organism is disclosed herein.

[0032] A method for modulating (eg. reducing or decreasing) the resistance of a cell comprising at least one or more *cas* genes or proteins and two or more CRISPR repeats against a target nucleic acid or a transcription product thereof comprising the steps of: (i) identifying one or more CRISPR spacers in an organism that is substantially resistant to the target nucleic acid or transcription product thereof; and (ii) modifying the sequence of at least

one or more CRISPR spacer(s) in the cell such that the CRISPR spacer(s) has a reduced degree of homology to the spacer(s) in the organism is disclosed herein.

[0033] A method for modulating (eg. reducing or decreasing) the resistance of a cell comprising at least one or more *cas* genes or proteins and two or more CRISPR repeats against a target nucleic acid or a transcription product thereof comprising modifying the one or more *cas* genes or proteins and/or two or more CRISPR repeats in the cell is disclosed herein.

[0034] A method for identifying a CRISPR spacer or pseudo CRISPR spacer for use in modulating the resistance of a cell against a target nucleic acid or a transcription product thereof comprising the steps of: (i) preparing a cell comprising at least two CRISPR repeats and at least one *cas* gene or protein; (ii) identifying at least one CRISPR spacer or pseudo CRISPR spacers in an organism that is substantially resistant to the target nucleic acid or transcription product thereof; (iii) modifying the sequence of the CRISPR spacer in the cell such that the CRISPR spacer has homology to the spacer of the organism; and (iv) determining if the cell modulates resistance against the target nucleic acid or transcription product thereof, wherein modulation of the resistance of the cell against the target nucleic acid or transcription product thereof is indicative that the CRISPR spacer modulates the resistance of the cell is disclosed herein.

[0035] A method for identifying a *cas* gene for use in modulating the resistance of a cell against a target nucleic acid or transcription product thereof comprising the steps of: (i) preparing a cell comprising at least one CRISPR spacer and at least two CRISPR repeats; (ii) engineering the cell such that it comprises at least one *cas* gene; and (iii) determining if the cell modulates resistance against the target nucleic acid or transcription product thereof, wherein modulation of the resistance of the cell against the target nucleic acid or transcription product thereof is indicative that the *cas* gene can be used to modulate the resistance of the cell is disclosed herein.

[0036] A method for identifying a CRISPR repeat for use in modulating the resistance of a cell against a target nucleic acid or transcription product thereof comprising the steps of: (i) preparing a cell comprising at least one CRISPR spacer and at least one *cas* gene; (ii) engineering the cell such that it contains the CRISPR repeat; and (iii) determining if the cell modulates resistance against the target nucleic acid or transcription product thereof, wherein modulation of the resistance of the cell against the target nucleic acid or transcription product thereof is indicative that the CRISPR repeat can be used to modulate resistance is disclosed herein.

[0037] A method for identifying a functional combination of a *cas* gene and a CRISPR repeat comprising the steps of: (a) determining the sequences of the *cas* gene and the CRISPR repeat; (b) identifying one or more clusters of *cas* genes as determined by sequence comparison analysis; (c) identifying one or more clusters of CRISPR repeats; and (d) combining those *cas* gene and CRISPR repeat sequences that fall within the same cluster, wherein the combination of the *cas* gene and CRISPR repeat sequences within the same

cluster is indicative that the combination is a functional combination is disclosed herein,

[0038] A method for modulating the lysotype of a bacterial cell comprising one or more *cas* genes or proteins and two or more CRISPR repeats comprising the steps of: (i) identifying one or more pseudo CRISPR spacers in the genomic sequence of a bacteriophage against which resistance is to be modulated; and (ii) modifying the sequence of one or more CRISPR spacers of the bacterial cell such that the CRISPR spacer(s) of the bacterial cell has homology to the pseudo CRISPR spacer(s) of the bacteriophage against which resistance is to be modulated is disclosed herein.

[0039] A method for modulating (eg. conferring or increasing) the resistance of a bacterial cell against a bacteriophage comprising the steps of: (i) identifying a sequence (eg, a conserved sequence) in a bacteriophage (preferably, a sequence essential to the function or survival of the bacteriophage); (ii) preparing a CRISPR spacer which is homologous to the identified sequence; (iii) preparing a nucleic acid comprising at least one *cas* gene and at least two CRISPR repeats together with the CRISPR spacer; and (iv) introducing said nucleic acid into the bacterial cell thus to render the bacterial cell resistant to said target nucleic acid or transcription product thereof is disclosed herein.

[0040] A method for modulating (eg. conferring or increasing) the resistance of a bacterial cell against a target nucleic acid or transcription product in a bacteriophage thereof comprising the steps of: (i) identifying one or more pseudo CRISPR spacers in a bacteriophage genome that is capable of providing resistance to the target nucleic acid or transcription product thereof; (ii) preparing a recombinant nucleic acid comprising at least one *cas* gene and at least two CRISPR repeats together with said identified one or more pseudo CRISPR spacers; and (iii) introducing said recombinant nucleic acid into said bacterial cell thus to render the bacterial cell resistant to said target nucleic acid or transcription product thereof is disclosed herein.

[0041] A method for modulating the resistance of a bacterial cell comprising one or more *cas* genes or proteins and two or more CRISPR repeats against a target nucleic acid or transcription product thereof in a bacteriophage comprising the steps of: (i) identifying one or more pseudo CRISPR spacers in a bacteriophage that is capable of providing resistance to a target nucleic acid or transcription product thereof; (ii) identifying one or more CRISPR spacers in a bacterial cell in which resistance is to be modulated; and (iii) modifying the sequence of the CRISPR spacer(s) in the bacterial cell in which resistance is to be modulated such that the CRISPR spacer(s) has a higher degree of homology to the pseudo CRISPR spacer(s) of the bacteriophage against which resistance is to be modulated is disclosed herein.

[0042] A method for determining the resistance of a cell against a target nucleic acid or a transcription product thereof comprising identifying one or more functional CRISPR repeat-*cas* combinations and one or more CRISPR spacers in the cell is disclosed herein.

[0043] A cell obtained or obtainable by the method(s) described herein is disclosed herein.

[0044] A CRISPR spacer or pseudo CRISPR spacer obtained or obtainable by the method(s) described herein is disclosed herein.

[0045] A *cas* gene obtained or obtainable by the method(s) described herein is disclosed herein,

[0046] A CRISPR repeat obtained or obtainable by the method(s) described herein is disclosed herein.

[0047] A functional combination obtained or obtainable by the method(s) described herein is disclosed herein.

[0048] A recombinant CRISPR locus comprising a CRISPR spacer or pseudo CRISPR spacer, and/or a *cas* gene, and/or a CRISPR repeat and/or a functional combination is disclosed herein.

[0049] The use of a cell, a CRISPR spacer or pseudo CRISPR spacer, a *cas* gene, a CRISPR repeat or a functional combination for modulating the resistance of a cell against a target nucleic acid or a transcription product thereof is disclosed herein.

[0050] A cell culture comprising a cell, a CRISPR spacer or pseudo CRISPR spacer, a *cas* gene, a CRISPR repeat or a functional combination for modulating the resistance of a cell against a target nucleic acid or a transcription product thereof is disclosed herein.

[0051] A food product or feed comprising the culture described herein is disclosed herein.

[0052] A process for preparing a food product or feed comprising the use of the culture described herein is disclosed herein.

[0053] A food product of feed obtained or obtainable by the process described herein is disclosed herein.

[0054] The use of the culture described herein for preparing a food product is disclosed herein.

[0055] A nucleotide sequence comprising or consisting of the sequence set forth in any of SEQ ID Nos. 7-10 and SEQ ID Nos. 359-405 or a variant, fragment, homologue or derivative thereof is disclosed herein.

[0056] An amino acid sequence encoded by the nucleotide sequence described herein is disclosed herein.

[0057] A construct or vector comprising one or more of the nucleotide sequences described herein is disclosed herein.

[0058] A host cell into which has been incorporated one or more of the nucleotide sequences described herein or the construct or vector described herein is disclosed herein,

PREFERRED EMBODIMENTS

[0059] In some embodiments, the one or more *cas* genes or proteins are used in combination with two or more CRISPR repeats.

[0060] The one or more *cas* genes or proteins and the two or more CRISPR repeats are or are derivable (preferably, derived) from the same cell.

[0061] In some embodiments, the one or more *cas* genes or proteins and the two or more CRISPR repeats naturally co-occur in the same cell.

[0062] In some embodiments, the one or more *cas* genes or proteins are used in combination with one or more CRISPR spacers.

[0063] In some embodiments, the CRISPR spacer(s) is or is derivable (preferably, derived) from an organism that is different to the cell from which the one or more *cas* genes or proteins and/or the two or more CRISPR repeats are or are derivable (preferably, derived).

[0064] The spacer is obtained from a cell which is resistant to a target nucleic acid.

[0065] In some embodiments, the CRISPR spacer is a synthetic nucleic acid sequence.

[0066] In some embodiments, the CRISPR spacer(s) have homology to the target nucleic acid.

[0067] In some embodiments, the CRISPR spacer(s) have 100% identity to the target nucleic acid over at least the length of the CRISPR spacer core.

[0068] In some embodiments, the one or more *cas* genes or proteins are used in combination with at least one or more CRISPR spacers and at least two or more CRISPR repeats.

[0069] In some embodiments, the target nucleic acid or transcription product thereof is or is derivable (preferably, derived) from bacteriophage DNA.

[0070] In some embodiments, the target nucleic acid or transcription product thereof is or is derivable (preferably, derived) from plasmid DNA.

[0071] In some embodiments, the target nucleic acid or transcription product thereof is or is derivable (preferably, derived) from a mobile genetic element.

[0072] In some embodiments, the target nucleic acid or transcription product thereof is or is derivable (preferably, derived) from a transposable element or an insertion sequence.

[0073] In some embodiments, the target nucleic acid or transcription product thereof is or is derivable (preferably, derived) from an antibiotic resistance gene.

[0074] In some embodiments, the target nucleic acid or transcription product thereof is or is derivable (preferably, derived) from a nucleic acid encoding a virulence factor.

[0075] In some embodiments, the virulence factor is selected from the group consisting of a toxin-, an internalin- and a hemolysin-encoding nucleic acid.

[0076] The one or more *cas* genes and the two or more CRISPR repeats are or are derivable (preferably, derived) from the same cell.

[0077] In some embodiments, the one or more *cas* genes and the two or more CRISPR repeats naturally co-occur in the same cell.

[0078] In some embodiments, the CRISPR spacers are or are derivable (preferably, derived) from an organism that is different to the cell from which the one or more *cas* genes and/or the two or more CRISPR repeats are or are derivable (preferably, derived).

[0079] In some embodiments, the cell is a recipient cell or a host cell.

[0080] The one or more *cas* genes or proteins and the two or more CRISPR repeats are or are derivable (preferably, derived) from the same cell.

[0081] In some embodiments, the spacers are or are derivable (preferably, derived) from an organism that is different to the cell comprising the one or more *cas* genes or proteins and/or the two or more CRISPR repeats.

[0082] In some embodiments, the one or more *cas* genes or proteins and the two or more CRISPR repeats naturally co-occur in the same cell.

[0083] In some embodiments, said modification comprises inserting one or more CRISPR spacers and/or pseudo CRISPR spacers into the cell.

[0084] In some embodiments, the spacer of the cell has 100% homology to the CRISPR spacer or pseudo CRISPR spacer of the organism.

[0085] In some embodiments, said modification comprises genetically engineering the CRISPR spacer of the cell.

[0086] In some embodiments, all or part of the spacer in the cell is modified.

[0087] In some embodiments, said modification comprises the modification of a recombinant spacer.

[0088] In some embodiments, said modification occurs through spontaneous mutation or mutagenesis.

[0089] In some embodiments, the nucleotide sequences of the *cas* gene and the CRISPR repeat are or are derivable (preferably, derived) from the same strains.

[0090] In some embodiments, the nucleotide sequences of the *cas* gene and the CRISPR repeat are or are derivable (preferably, derived) from the same species.

[0091] In some embodiments, the nucleotide sequences of the *cas* gene and the CRISPR repeat are or are derivable (preferably, derived) from the same genera.

[0092] In some embodiments, the nucleotide sequences of the *cas* gene and the CRISPR repeat are or are derivable (preferably, derived) from the same organisms.

[0093] In some embodiments, the target nucleic acid in the bacteriophage is a highly conserved nucleic acid sequence.

[0094] In some embodiments, the target nucleic acid in the bacteriophage encodes a host specificity protein.

[0095] In some embodiments, the target nucleic acid in the bacteriophage encodes a protein that is essential for survival, replication or growth of the bacteriophage.

[0096] In some embodiments, the target nucleic acid in the bacteriophage encodes a helicase, a primase, a head or tail structural protein, a protein with a conserved domain (eg. holin, lysin, and others) or a conserved sequences amongst important phage genes.

[0097] In some embodiments, the method for determining the resistance of a cell to a target nucleic acid or a transcription product thereof comprises the additional step of comparing the sequence of the one or more CRISPR spacers in the cell with the sequence of the target nucleic acid .

[0098] In some embodiments, the method for determining the resistance of a cell to a target nucleic acid or a transcription product thereof comprises the additional step of determining the resistance profile of the cell.

[0099] In some embodiments, said culture is a starter culture or a probiotic culture.

FIGURES

[0100]

Figure 1 Schematic representation of CRISPR1 of *S. thermophilus* CNRZ1066 (42 repeats).

Figure 2 Dotplot analysis of Cas protein sequences (A) and CRISPR locus sequences (B). Organism names (genus, species, strain) are indicated on the right side of each dotplot (for example Sth_LMG18311 = *S. thermophilus* strain LMG18311).

Figure 3 Spacer sequences of *S. thermophilus* CNRZ1066 CRISPR locus were blasted (short nearly exact sequence searches using BlastN at the NCBI website) against the viruses database, and aligned with the subsequent matches in *S. thermophilus* bacteriophages. (A) The table indicates the spacer sequences of CNRZ1066 CRISPR presenting significant sequence identities with phage sequences (dark cells). (B) Alignment of the sequence of interspacing sequence #29 with eight phage sequences. (Remark: spacer #20 shows similarity to a number of host specificity proteins).

Figure 4 Putative stem-loop secondary structure of a CRISPR repeat sequence of *S. thermophilus*. Only one DNA strand is shown.

Figure 5 Integration of a CRISPR spacer into the CRISPR locus of *Streptococcus thermophilus* provides resistance against a bacteriophage that the CRISPR spacer shows identity to. The parent DGCC7710 is phage sensitive, and the BIM DGCC7710RH1 is phage resistant. The BIM DGCC7710RH1 has a new spacer (Sn) in the CRISPR locus, which shows 100% identity to phage sequence. In step (b) the strain is challenged with phage 858 and a phage resistant mutant is selected. In step (c) the CRISPR I locus of the mutant has an additional spacer which shares 100% identity with region 31.921-31.950bp of the phage.

Figure 6 Integration of a CRISPR spacer into the CRISPR locus of *Streptococcus thermophilus* provides resistance against a bacteriophage that the CRISPR spacer shows identity to. The parent DGCC7710 is phage sensitive, and the BIM DGCC7710RH2 is phage resistant. The BIM DGCC7710RH2 has a new spacer (Sn) in the CRISPR locus, which shows 100% identity to phage sequence. In step (b) the strain is challenged with phage 858 and a phage resistant mutant is selected. In step (c) The experiment was independently repeated and another mutant was selected. The CRISPR I locus of the mutant has an additional spacer (different from that in RH1) which shares 100% identity with region 17.125-17.244bp of the phage.

Figure 7 Spacer arrangement of CRISPR I in various *Streptococcus thermophilus* strains. Numbers indicate the position of the spacer. Strain names are listed on the left. Letters indicate CRISPR spacer type, with identical spacers described with a similar 2-letter code. Spacers with single nucleotide polymorphisms are labeled with identical letter combination, complemented with a "prime" label. Unique spacers are not described by a letter combination, and are left blank.

Figure 8 Homology of CRISPR spacers with known sequences, including bacterial chromosomal sequences (shaded in gray), plasmid DNA sequences (black squares) and

phage genomic sequences (shaded in black).

Figure 9 A graphical representation of the plasmid system used to genetically engineer a number of constructs in *Streptococcus thermophilus* as described by Russell, M. W., and T. R. Klaenhammer (2001) Efficient system for directed integration into the *Lactobacillus acidophilus* and *Lactobacillus gasseri* chromosomes via homologous recombination. *Applied and Environmental Microbiology* 67:4361-4364.

Figure 10 A graphical representation of the plasmid used to subclone PCR products of the various constructs described herein (*cas1* KO, *cas4* KO, RT and S1S2). The plasmid is available commercially from Invitrogen in the TOPO TA Cloning® kit.

Figure 11 A graphical representation of the plasmid used for homologous recombination in one embodiment of the present invention.

Figure 12 A graphical representation illustrating the preparation of the CAS1KO construct in which the *cas1* gene is disrupted by homologous recombination.

Figure 13 A graphical representation illustrating the preparation of the CAS4KO construct in which the *cas4* gene is disrupted by homologous recombination.

Figure 14 A graphical representation illustrating the S1S2 construct engineering using specific primers and iterative PCR reactions. The first panel illustrates all primers used and the set up for the first two PCR reactions (reaction #1 with primers P1 and P2 and reaction #2 with primers P2 and P3). The second panel shows the PCR products obtained from the first two PCR reactions, with the product from reaction #1 on the left and the product from reaction #2 on the right. The third panel shows the third PCR reaction, using a combination of the products from the first two PCRs as the template for the third PCR reaction, and primer P1 from the first reaction along with primer P4 from the second reaction. The fourth panel shows the product of PCR#3, which technically generates the S1 S2 construct.

Figure 15 A graphical representation of the details for primer design for primers 2 and 3, which contain key sequences for the experiment, derived from spacers identical to phage sequences (the PCR products derived from these PCR primers will generate the spacers that will ultimately provide resistance to the phages).

Figure 16 A graphical representation of the integration of the S1S2 construct.

Figure 17 A graphical representation of the preparation of the RT construct using a restriction enzyme to generate the RT construct from the S1S2 construct. There are *Bgl*I restriction sites within the repeats allow the "middle" part of the construct to be cut. Following enzymatic digestion, a ligase is used to patch together the two end pieces, thus generating a new construct that has RT, but no spacers.

Figure 18 A graphical representation of the integration of the RT construct.

Figure 19 A graphical representation of the RT' construct.

Figure 20 A graphical representation of the RT' construct.

DETAILED DESCRIPTION OF THE INVENTION

CRISPR LOCUS

[0101] CRISPR loci are a distinct class of interspersed short sequence repeats (SSRs) that were first recognized in *E. coli* (Ishino et al. (1987) J. Bacteriol. 169:5429-5433; Nakata et al. (1989) J. Bacteriol. 171:3553-3556). Similar interspersed SSRs have been identified in *Haloferax mediterranei*, *Streptococcus pyogenes*, *Anabaena*, and *Mycobacterium tuberculosis* (Groenen et al. (1993) Mol. Microbiol. 10:1057-1065; Hoe et al. (1999) Emerg. Infect. Dis. 5:254-263; Masepohl et al. (1996) Biochim. Biophys. Acta 1307:26-30; Mojica et al. (1995) Mol. Microbiol. 17:85-93). The CRISPR loci differ from other SSRs by the structure of the repeats, which have been termed short regularly spaced repeats (SRSRs) (Janssen et al. (2002) OMICS J. Integ. Biol. 6:23-33; Mojica et al. (2000) Mol. Microbiol. 36:244-246). The repeats are short elements that occur in clusters, that are always regularly spaced by unique intervening sequences with a constant length (Mojica et al. (2000) Mol. Microbiol. 36:244-246). Although the repeat sequences are highly conserved between strains, the number of interspersed repeats and the sequences of the spacer regions differ from strain to strain (van Embden et al. (2000) J. Bacteriol. 182:2393-2401).

[0102] The common structural characteristics of CRISPR loci are described in Jansen *et al.* (2002) as (i) the presence of multiple short direct repeats, which show no or very little sequence variation within a given locus; (ii) the presence of non-repetitive spacer sequences between the repeats of similar size; (iii) the presence of a common leader sequence of a few hundred basepairs in most species harbouring multiple CRISPR loci; (iv) the absence of long open reading frames within the locus; and (v) the presence of one or more *cas* genes.

[0103] CRISPRs are typically short partially palindromic sequences of 24-40bp containing inner and terminal inverted repeats of up to 11 bp. Although isolated elements have been detected, they are generally arranged in clusters (up to about 20 or more per genome) of repeated units spaced by unique intervening 20-58bp sequences. CRISPRs are generally homogenous within a given genome with most of them being identical. However, there are examples of heterogeneity in, for example, the Archaea (Mojica *et al.* 2000).

[0104] By way of example, the genome of *Streptococcus thermophilus* LMG18311 contains 3 CRISPR loci; the 36-bp repeated sequences are different in CRISPR1 (34 repeats), CRISPR2 (5 repeats), and CRISPR3 (a single sequence). Nevertheless, they are perfectly conserved within each locus. CRISPR1 and CRISPR2 repeats are respectively interspaced by 33 and 4 sequences of 30 bp in length. All these interspacing sequences are different from each other.

They are also different from those found in strain CNRZ1066 (41 interspacing sequences within CRISPR1) and in strain LMD-9 (16 within CRISPR1 and 8 within CRISPR3), which both are *S. thermophilus*. Figure 1 describes one of the CRISPRs identified in *S. thermophilus*.

[0105] Various methods for identifying CRISPR loci are already known in the art. By way of example, Jensen *et al.* (2002) describe a computer based approach in which nucleotide sequences are searched for CRISPR motifs using the PATSCAN program at the server of the Mathematics and Computer Science Division at the Argonne National Laboratory, Argonne, IL, USA. The algorithm that was used for identifying CRISPR motifs was $p1 = a...bc...dp1c...dp1c...dp1$, where a and b are the lower and upper size limit of the repeat and $p1$ and c and d are the lower and upper size limit of the spacer sequences. The values of a , b , c and d may be varied from about 15 to about 70 bp at increments of about 5 bp.

[0106] CRISPR loci may be identified using dotplots (using, for example, a computer program called Dotter).

[0107] Sequence similarity analysis may be performed using various methods that are well known in the art. By way of example, analysis may be performed using NCBI BLAST with a microbial genomes database (<http://www.ncbi.nlm.nih.gov>) and GenBank.

[0108] The amplification of CRISPR loci has been described in, for example, Mojica *et al.* (2005) and Pourcel *et al.* (2005). Amplification of the desired region of DNA may be achieved by any method known in the art, including polymerase chain reaction (PCR). By "amplification" we mean the production of additional copies of a nucleic acid sequence. This is generally carried out using PCR technologies well known in the art (Dieffenbach and Dveksler (1995) PCR Primer, a Laboratory Manual (Cold Spring Harbor Press, Plainview, New York).

[0109] By "polymerase chain reaction" or "PCR" we mean a method such as that disclosed in US4,683,195 and US4,683,202, which describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. The length of the amplified segment of the desired target sequence is determined by the relative positions of two oligonucleotide primers . with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as "PCR". Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified."

[0110] In the PCR approach, oligonucleotide primers can be designed for use in PCR reactions to amplify all or part of a CRISPR locus. By "primer" we mean an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced (i.e., in the presence of nucleotides and an inducing agent - such as DNA polymerase and at a suitable temperature and pH). The primer may be single stranded for maximum efficiency in

amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. The primer may be an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer, and the use of the method. PCR primers are typically at least about 10 nucleotides in length, and most typically at least about 20 nucleotides in length.

[0111] Methods for designing PCR primers and PCR cloning are generally known in the art and are disclosed in Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New York). See also Innis et al., eds. (1990) *PCR Protocols: A Guide to Methods and Applications* (Academic Press, New York); Innis and Gelfand, eds. (1995) *PCR Strategies* (Academic Press, New York); and Innis and Gelfand, eds. (1999) *PCR Methods Manual* (Academic Press, New York). Known methods of PCR include, but are not limited to, methods using paired primers, nested primers, single specific primers, degenerate primers, gene-specific primers, vector-specific primers, partially mismatched primers, and the like.

[0112] The CRISPR loci may comprise, consist or consist essentially of DNA or RNA of genomic, synthetic or recombinant origin.

[0113] The CRISPR loci may be double-stranded or single-stranded whether representing the sense or antisense strand or combinations thereof.

[0114] The CRISPR loci may be prepared by use of recombinant DNA techniques (e.g. recombinant DNA), as described herein.

[0115] Nucleotide sequences described herein may be obtained from databases - such as GenBank or the JGI website at http://genome.jgi-psf.org/mic_home.html.

CRISPR ORIENTATION

[0116] For the avoidance of doubt, in the context of the present invention the CRISPR locus is orientated as follows.

[0117] The CRISPR leader is a conserved DNA segment of defined size. For example, the leader sequence of *S. thermophilus* CRISPR1 is the DNA segment starting immediately after the stop codon of gene str0660, and ending just before the first repeat. The CRISPR leader is located at the 5' end of the CRISPR locus. The CRISPR leader is located immediately upstream of the first CRISPR repeat of the CRISPR locus.

[0118] The CRISPR trailer is a conserved DNA segment of defined size. For example, the trailer sequence of *S. thermophilus* CRISPR1 is the DNA segment starting immediately after

occur in a recipient cell and one or more heterologous spacers may be integrated or inserted adjacent to the one or more of the *cas* genes or proteins.

[0128] In some embodiments, suitably one or more of the *cas* genes and/or proteins may be heterologous to the recipient cell and one or more of the spacers may be homologous or heterologous. In this instance, the spacers may be integrated or inserted adjacent to the one or more of the *cas* gene or proteins.

[0129] The use of one or more *cas* genes or proteins and at least two CRISPR repeats for modulating resistance in a cell (eg. a recipient cell) against a target nucleic acid or a transcription product thereof is disclosed herein.

[0130] The use of one or more *cas* genes or proteins, at least two CRISPR repeats and at least one CRISPR spacer for modulating resistance in a cell (eg. a recipient cell) against a target nucleic acid or a transcription product thereof is disclosed herein.

[0131] CRISPR structures are typically found in the vicinity of four genes named *cas1* to *cas4*. The most common arrangement of these genes is *cas3-cas4-cas1-cas2*. The Cas3 protein appears to be a helicase, whereas Cas4 resembles the RecB family of exonucleases and contains a cysteine-rich motif, suggestive of DNA binding. Cas1 is generally highly basic and is the only Cas protein found consistently in all species that contain CRISPR loci, Cas2 remains to be characterized, *cas1-4* are typically characterized by their close proximity to the CRISPR loci and their broad distribution across bacterial and archaeal species. Although not all *cas1-4* genes associate with all CRISPR loci, they are all found in multiple subtypes.

[0132] Bolotin *et al.* (2005) have recently reported another cluster of three genes associated with CRISPR structures in many bacterial species, named here as *cas1B*, *cas5* and *cas6*.

[0133] The *cas* gene may be *cas1*, *cas2*, *cas3*, *cas4*, *cas1B*, *cas5* and/or *cas6*. In one embodiment, the *cas* gene is *cas1*.

[0134] The *cas* gene may be *cas1*, *cas2*, *cas3*, *cas4*, *cas1B*, *cas5* and/or *cas6* or a fragment, variant, homologue or derivative thereof.

[0135] The *cas* genes may be *cas1*, *cas2*, *cas3*, *cas4*, *cas1B*, *cas5* and/or *cas6* or a plurality thereof or a combination thereof - such as *cas1* and *cas2*; *cas1* and *cas3*; *cas1* and *cas4*; *cas1* and *cas1B*; *cas1* and *cas5*; *cas1* and *cas6*; *cas2* and *cas3*; *cas2* and *cas4*; *cas2* and *cas1B*; *cas2* and *cas5*; *cas2* and *cas6*; *cas3* and *cas4*; *cas3* and *cas1B*; *cas3* and *cas5*; *cas3* and *cas6*; *cas4* and *cas1B*; *cas4* and *cas5*; *cas4* and *cas6*; *cas1B* and *cas5*; *cas1B* and *cas6*; *cas1*, *cas2* and *cas3*; *cas1*, *cas2* and *cas4*; *cas1*, *cas2* and *cas1B*; *cas1*, *cas2*, *cas3* and *cas4*; *cas1*, *cas2*, *cas3* and *cas1B*; *cas1*, *cas2*, *cas3* and *cas5*; *cas1*, *cas2*, *cas3* and *cas6*; *cas1*, *cas2*, *cas3*, *cas4* and *cas1B*; *cas1*, *cas2*, *cas3*, *cas4* and *cas5*; *cas1*, *cas2*, *cas3*, *cas4*, *cas1B* and *cas6*; *cas1*, *cas2*, *cas3*, *cas4*, *cas1B*, *cas5*; *cas1*, *cas2*, *cas3*, *cas4*, *cas1B* and *cas6*; *cas1*, *cas2*, *cas3*, *cas4*, *cas1B*, *cas5* and *cas6*; *cas2*, *cas3* and *cas4*; *cas2*, *cas3* and *cas1B*; *cas2*,

cas3 and *cas5*; *cas2*, *cas3* and *cas6*; *cas2*, *cas3*, *cas4* and *cas1B*; *cas2*, *cas3*, *cas4*, and *cas5*; *cas2*, *cas3*, *cas4* and *cas6*; *cas2*, *cas3*, *cas4*, *cas1B* and *cas5*; *cas2*, *cas3*, *cas4*, *cas1B* and *cas6*; *cas2*, *cas3*, *cas4*, *cas1B*, *cas5* and *cas6*; *cas3*, *cas4* and *cas1B*; *cas3*, *cas4* and *cas5*; *cas3*, *cas4* and *cas6*; *cas3*, *cas4*, *cas1B* and *cas5*; *cas3*, *cas4*, *cas1B* and *cas6*; *cas3*, *cas4*, *cas1B*, *cas5* and *cas6*; *cas4*, *cas1B* and *cas5*; *cas4*, *cas1B* and *cas6*; *cas4*, *cas1B*, *cas5* and *cas6*; *cas5* and *cas6* or combinations thereof.

[0136] The *cas* genes may be *cas1* and *cas2*; *cas1* and *cas3*; *cas1* and *cas4*; *cas1* and *cas1B*; *cas1* and *cas5*; *cas1* and *cas6*; *cas2* and *cas3*; *cas2* and *cas4*; *cas2* and *cas1B*; *cas2* and *cas5*; *cas2* and *cas6*; *cas3* and *cas4*; *cas3* and *cas1B*; *cas3* and *cas5*; *cas3* and *cas6*; *cas4* and *cas1B*; *cas4* and *cas5*; *cas4* and *cas6*; *cas1B* and *cas5* or *cas1B* and *cas6* or combinations thereof.

[0137] The *cas* genes may be a *cas1*, *cas2* and *cas3*; *cas1*, *cas2* and *cas4*; *cas1*, *cas2* and *cas1B*; *cas1*, *cas2*, *cas3* and *cas4*; *cas1*, *cas2*, *cas3* and *cas1B*; *cas1*, *cas2*, *cas3* and *cas5*; *cas1*, *cas2*, *cas3* and *cas6*; *cas1*, *cas2*, *cas3*, *cas4* and *cas1B*; *cas1*, *cas2*, *cas3*, *cas4* and *cas5*; *cas1*, *cas2*, *cas3*, *cas4*, *cas1B* and *cas6*; *cas1*, *cas2*, *cas3*, *cas4*, *cas1B* and *cas5*; *cas1*, *cas2*, *cas3*, *cas4*, *cas1B* and *cas6*; *cas1*, *cas2*, *cas3*, *cas4*, *cas1B*, *cas5* and *cas6* or combinations thereof.

[0138] The *cas* genes may be *cas2*, *cas3* and *cas4*; *cas2*, *cas3* and *cas1B*; *cas2*, *cas3* and *cas5*; *cas2*, *cas3* and *cas6*; *cas2*, *cas3*, *cas4* and *cas1B*; *cas2*, *cas3*, *cas4*, and *cas5*; *cas2*, *cas3*, *cas4* and *cas6*; *cas2*, *cas3*, *cas4*, *cas1B* and *cas5*; *cas2*, *cas3*, *cas4*, *cas1B* and *cas6*; *cas2*, *cas3*, *cas4*, *cas1B*, *cas5* and *cas6* or combinations thereof.

[0139] The *cas* genes may be *cas3*, *cas4* and *cas1B*; *cas3*, *cas4* and *cas5*; *cas3*, *cas4* and *cas6*; *cas3*, *cas4*, *cas1B* and *cas5*; *cas3*, *cas4*, *cas1B* and *cas6*; *cas3*, *cas4*, *cas1B*, *cas5* and *cas6*; *cas4*, *cas1B* and *cas5*; *cas4*, *cas1B* and *cas6*; *cas4*, *cas1B*, *cas5* and *cas6*; *cas5* and *cas6* or combinations thereof.

[0140] The *cas* gene maybe one or more of *cas1*, *cas2*, *cas3*, *cas4*, *cas1B*, *cas5* and/or *cas6* or a plurality thereof - such as a plurality of any 2 *cas* genes, any 3 *cas* genes, any 4 *cas* genes, any 5 *cas* genes, any 6 *cas* genes, or any 7 *cas* genes.

[0141] The plurality of *cas* genes may comprise, consist or consist essentially of a plurality of the same *cas* genes - such as 2 *cas* genes, 3 *cas* genes, 4 *cas* genes, 5 *cas* genes, 6 *cas* genes, 7 *cas* genes, 8 *cas* genes, 9 *cas* genes, 10 *cas* genes, 15 *cas* genes, 20 *cas* genes, 25 *cas* genes, 30 *cas* genes, 35 *cas* genes, 40 *cas* genes or even 50 or more *cas* genes.

[0142] The plurality of *cas* genes may comprise, consist or consist essentially of a plurality of different *cas* genes-such as 2 different *cas* genes, 3 different *cas* genes, 4 different *cas* genes, 5 different *cas* genes, 6 different *cas* genes, 7 different *cas* genes, 8 different *cas* genes, 9 different *cas* genes, 10 different *cas* genes, 15 different *cas* genes, 20 different *cas* genes, 25 different *cas* genes, 30 different *cas* genes, 35 different *cas* genes, 40 different *cas* genes or

even 50 or more different *cas* genes.

[0143] The plurality of *cas* genes may comprise, consist or consist essentially of a plurality of the same and/or different *cas* genes - such as 2 different *cas* genes, 3 different *cas* genes, 4 different *cas* genes, 5 different *cas* genes, 6 different *cas* genes, 7 different *cas* genes, 8 different *cas* genes, 9 different *cas* genes, 10 different *cas* genes, 15 different *cas* genes, 20 different *cas* genes, 25 different *cas* genes, 30 different *cas* genes, 35 different *cas* genes, 40 different *cas* genes or even 50 or more different *cas* genes. The same *cas* gene may be duplicated a plurality of times.

[0144] Suitably, the term "*cas* gene" refers to a plurality of *cas* genes - such as between 2 and 12 *cas* genes, more preferably, between 3 and 11 *cas* genes, more preferably, between 4 and 10 *cas* genes, more preferably, between 4 and 9 *cas* genes, more preferably, between 4 and 8 *cas* genes, more preferably, between 4 and 7 *cas* genes - such as 4, 5, 6, or 7 *cas* genes.

[0145] The *cas* gene(s) may comprise, consist or consist essentially of DNA or RNA of genomic, synthetic or recombinant origin.

[0146] The *cas* gene(s) may be double-stranded or single-stranded whether representing the sense or antisense strand or combinations thereof.

[0147] The *cas* gene(s) may be prepared by use of recombinant DNA techniques (e.g. recombinant DNA), as described herein.

[0148] As described herein below, the *cas* gene may be a fragment of a *cas* gene, thereby indicating that the *cas* gene comprises a fraction of a wild-type sequence. Suitably, the sequence comprises at least 30%, at least 40%, at least 50%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% of the wild-type sequence.

[0149] For some embodiments it is preferred that the *cas* gene is the *cas* gene that is closest to the leader sequence or the first CRISPR repeat at the 5' end of the CRISPR locus- such as *cas4* or *cas6*.

[0150] The Cas protein may be Cas1, Cas2, Cas3, Cas4, Cas1B, Cas5 and/or Cas6.

[0151] The Cas protein may be Cas1, Cas2, Cas3, Cas4, Cas1B, Cas5 and/or Cas6 or a fragment, variant, homologue or derivative thereof.

[0152] The Cas protein may be Cas1, Cas2, Cas3, Cas4, Cas1B, Cas5 and/or Cas6 or a combination thereof - such as Cas1 and Cas2 ; Cas1 and Cas3 ; Cas1 and Cas4; Cas1 and Cas1B; Cas1 and Cas5 ; Cas1 and Cas6 ; Cas2 and Cas3 ; Cas2 and Cas4 ; Cas2 and Cas1B; Cas2 and Cas5 ; Cas2 and Cas6 ; Cas3 and Cas4; Cas3 and Cas1B ; Cas3 and Cas5 ; Cas3 and Cas6 ; Cas4 and Cas1B ; Cas4 and Cas5; Cas4 and Cas6 ; Cas1B and Cas5 ; Cas1B and

Cas6 ; Cas1, Cas2 and Cas3 ; Cas1, Cas2 and Cas4 ; Cas1, Cas2 and Cas1B ; Cas1, Cas2, Cas3 and Cas4 ; Cas1, Cas2, Cas3 and Cas1B ; Cas1, Cas2, Cas3 and Cas5; Cas1, Cas2, Cas3 and Cas6; Cas1, Cas2, Cas3, Cas4 and Cas1B ; Cas1, Cas2, Cas3, Cas4 and Cas5 ; Cas1, Cas2, Cas3, Cas4, Cas1B and Cas6; Cas1, Cas2, Cas3, Cas4, Cas1B and Cas5 ; Cas1, Cas2, Cas3, Cas4, Cas1B and Cas6; Cas1, Cas2, Cas3, Cas4, Cas1B, Cas5 and Cas6; Cas2, Cas3 and Cas4; Cas2, Cas3 and Cas1B; Cas2, Cas3 and Cas5 ; Cas2, Cas3 and Cas6 ; Cas2, Cas3, Cas4 and Cas1B ; Cas2, Cas3, Cas4, and Cas5; Cas2, Cas3, Cas4 and Cas6; Cas2, Cas3, Cas4, Cas1B and Cas5; Cas2, Cas3, Cas4, Cas1B and Cas6; Cas2, Cas3, Cas4, Cas1B, Cas5 and Cas6; Cas3, Cas4 and Cas1B; Cas3, Cas4 and Cas5; Cas3, Cas4 and Cas6; Cas3, Cas4, Cas1B and Cas5; Cas3, Cas4, Cas1B and Cas6; Cas3, Cas4, Cas1B, Cas5 and Cas6; Cas4, Cas1B and Cas5; Cas4, Cas1B and Cas6 or Cas4, Cas1B, Cas5 and Cas6, Cas5 and Cas6.

[0153] The Cas protein may be Cas1 and Cas2; Cas1 and Cas3 ; Cas1 and Cas4; Cas1 and Cas1B; Cas1 and Cas5; Cas1 and Cas6; Cas2 and Cas3 ; Cas2 and Cas4 ; Cas2 and Cas1B; Cas2 and Cas5 ; Cas2 and Cas6; Cas3 and Cas4 ; Cas3 and Cas1B; Cas3 and Cas5 ; Cas3 and Cas6; Cas4 and Cas1B; Cas4 and Cas5 ; Cas4 and Cas6; Cas1B and Cas5 or Cas1B and Cas6 or combinations thereof.

[0154] The Cas protein may be Cas1, Cas2 and Cas3; Cas1, Cas2 and Cas4; Cas1, Cas2 and Cas1B; Cas1, Cas2, Cas3 and Cas4; Cas1, Cas2, Cas3 and Cas1B; Cas1, Cas2, Cas3 and Cas5; Cas1, Cas2, Cas3 and Cas6; Cas1, Cas2, Cas3, Cas4 and Cas1B; Cas1, Cas2, Cas3, Cas4 and Cas5 ; Cas1, Cas2, Cas3, Cas4, Cas1B and Cas6 ; Cas1, Cas2, Cas3, Cas4, Cas1B and Cas5 ; Cas1, Cas2, Cas3, Cas4, Cas1B and Cas6; Cas1, Cas2, Cas3, Cas4, Cas1B, Cas5 and Cas6 or combinations thereof.

[0155] The Cas protein may be Cas2, Cas3 and Cas4 ; Cas2, Cas3 and Cas1B ; Cas2, Cas3 and Cas5; Cas2, Cas3 and Cas6; Cas2, Cas3, Cas4 and Cas1B; Cas2, Cas3, Cas4, and Cas5 ; Cas2, Cas3, Cas4 and Cas6; Cas2, Cas3, Cas4, Cas1B and Cas5; Cas2, Cas3, Cas4, Cas1B and Cas6; Cas2, Cas3, Cas4, Cas1B, Cas5 and Cas6 or combinations thereof.

[0156] The Cas protein may be Cas3, Cas4 and Cas1B ; Cas3, Cas4 and Cas5 ; Cas3, Cas4 and Cas6 ; Cas3, Cas4, Cas1B and Cas5 ; Cas3, Cas4, Cas1B and Cas6 ; Cas3, Cas4, Cas1B, Cas5 and Cas6 ; Cas4, Cas1B and Cas5 ; Cas4, Cas1B and Cas6; Cas4, Cas1B, Cas5 and Cas6; Cas5 and Cas6 or combinations thereof.

[0157] The Cas protein may be one or more of Cas1, Cas2, Cas3, Cas4, Cas1B, Cas5 and/or Cas6 or a plurality thereof - such as a plurality of any 2 Cas genes, any 3 Cas genes, any 4 Cas genes, any 5 Cas genes, any 6 Cas genes, or any 7 Cas genes.

[0158] The plurality of Cas proteins may comprise, consist or consist essentially of a plurality of the same Cas proteins - such as 2 Cas proteins, 3 Cas proteins, 4 Cas proteins, 5 Cas proteins, 6 Cas proteins, 7 Cas proteins, 8 Cas proteins, 9 Cas proteins, 10 Cas proteins, 15 Cas proteins, 20 Cas proteins, 25 Cas proteins, 30 Cas proteins, 35 Cas proteins, 40 Cas

proteins or even 50 or more Cas proteins.

[0159] The plurality of Cas proteins may comprise, consist or consist essentially of a plurality of different Cas proteins - such as 2 different Cas proteins, 3 different Cas proteins, 4 different Cas proteins, 5 different Cas proteins, 6 different Cas proteins, 7 different Cas proteins, 8 different Cas proteins, 9 different Cas proteins, 10 different Cas proteins, 15 different Cas proteins, 20 different Cas proteins, 25 different Cas proteins, 30 different Cas proteins, 35 different Cas proteins, 40 different Cas proteins or even 50 or more different Cas proteins.

[0160] The plurality of Cas proteins may comprise, consist or consist essentially of a plurality of the same and/or different Cas proteins - such as 2 different Cas proteins, 3 different Cas proteins, 4 different Cas proteins, 5 different Cas proteins, 6 different Cas proteins, 7 different Cas proteins, 8 different Cas proteins, 9 different Cas proteins, 10 different Cas proteins, 15 different Cas proteins, 20 different Cas proteins, 25 different Cas proteins, 30 different Cas proteins, 35 different Cas proteins, 40 different Cas proteins or even 50 or more different Cas proteins. The same Cas proteins may be duplicated a plurality of times.

[0161] Suitably, the term "Cas protein" refers to a plurality of Cas proteins - such as between 2 and 12 Cas proteins, more preferably, between 3 and 11 Cas proteins, more preferably, between 4 and 10 Cas proteins, more preferably, between 4 and 9 Cas proteins, more preferably, between 4 and 8 Cas proteins, more preferably, between 4 and 7 proteins genes - such as 4, 5, 6, or 7 Cas proteins.

[0162] The Cas protein(s) may be encoded by a *cas* gene which may comprise DNA or RNA of genomic, synthetic or recombinant origin.

[0163] The Cas protein(s) may be encoded by a *cas* gene which may be double-stranded or single-stranded whether representing the sense or antisense strand or combinations thereof.

[0164] The Cas protein(s) may be prepared by use of recombinant DNA techniques (e.g. recombinant DNA), as described herein.

[0165] A method for identifying a *cas* gene for use in modulating the resistance of a cell against a target nucleic acid or transcription product thereof comprising the steps of: (i) preparing a cell comprising at least one CRISPR spacer and at least two CRISPR repeats; (ii) engineering the cell such that it comprises at least one *cas* gene; and (iii) determining if the cell modulates resistance against the target nucleic acid or transcription product thereof, wherein modulation of the resistance of the cell against the target nucleic acid or transcription product thereof is indicative that the *cas* gene can be used to modulate the resistance of the cell is disclosed herein.

[0166] One or more of the *cas* genes may be used to engineer a cell - such as a recipient cell. In particular, one or more *cas* genes may be used to engineer a cell-such as a recipient cell - that in combination with two or more CRISPR repeats and one or more CRISPR spacers can

be used to modulate the resistance of a cell against a target nucleic acid or a transcription product thereof. By way of example, the *cas* gene(s) may be inserted into the DNA of a cell (eg. a recipient cell) - such as plasmid DNA or genomic DNA of a cell - using various methods that are well known in the art. By way of further example, the *cas* genes may be used as a template upon which to modify (eg. mutate) the DNA of a cell (eg. a recipient cell) - such as plasmid DNA or genomic DNA- such that *cas* genes are created or formed in the DNA of the cell. By way of further example, the *cas* genes may be cloned into a construct, a plasmid or a vector and the like which is then transformed into the cell, using methods such as those described herein.

[0167] The *cas* genes may comprise or consist of a *cas* cluster selected from the group consisting of any one or more of SEQ ID No. 461, SEQ ID No. 466, SEQ ID No. 473, SEQ ID No. 478, SEQ ID No. 488, SEQ ID No. 493, SEQ ID No. 498, SEQ ID No. 504, SEQ ID No. 509, SEQ ID No. 517

[0168] The *cas* genes may comprise or consist of any one or more of SEQ ID Nos. 462-465, 467-472, 474-477, 479-487, 489-492, 494-497, 499-503, 505-508, 510-516 and/or 517-521.

[0169] Suitably, the one or more *cas* genes or proteins are used together with or in combination with two or more CRISPR repeats and optionally one or more CRISPR spacers.

CRISPR REPEAT

[0170] As used herein, the term "CRISPR repeat" has the conventional meaning as used in the art *ie.* multiple short direct repeats, which show no or very little sequence variation within a given CRISPR locus.

[0171] As used herein, the term "CRISPR" is synonymous with the term "CRISPR repeat".

[0172] The number of nucleotides in a repeat is generally about 20 to about 40 base pairs, but may be about 20 to about 39 base pairs, about 20 to about 37 base pairs, about 20 to about 35 base pairs, about 20 to about 33 base pairs, about 20 to about 30 base pairs, about 21 to about 40 base pairs, about 21 to about 39 base pairs, about 21 to about 37 base pairs, about 23 to about 40 base pairs, about 23 to about 39 base pairs, about 23 to about 37 base pairs, about 25 to about 40 base pairs, about 25 to about 39 base pairs, about 25 to about 37 base pairs, about 25 to about 35 base pairs, or about 28 or 29 base pairs. The number of repeats may range from about 1 to about 140, from about 1 to about 100, from about 2 to about 100, from about 5 to about 100, from about 10 to about 100, from about 15 to about 100, from about 20 to about 100, from about 25 to about 100, from about 30 to about 100, from about 35 to about 100, from about 40 to about 100, from about 45 to about 100, from about 50 to about 100, from about 1 to about 135, from about 1 to about 130, from about 1 to about 125, from about 1 to about 120, from about 1 to about 115, from about 1 to about 110, from about 1 to about 105, from about 1 to about 100, from about 1 to about 95, from about 1 to about 90,

from about 1 to about 80, from about 1 to about 70, from about 1 to about 60, from about 1 to about 50, from about 10 to about 140, from about 10 to about 130, from about 10 to about 120, from about 10 to about 110, from about 10 to about 95, from about 10 to about 90, from about 20 to about 80, from about 30 to about 70, from about 30 to about 60, from about 30 to about 50, from about 30 to about 40, or about 32.

[0173] Suitably, the number of nucleotides in a repeat is generally about 20 to about 40 base pairs, but may be about 20 to about 39 base pairs, about 20 to about 37 base pairs, about 20 to about 35 base pairs, about 20 to about 33 base pairs, about 20 to about 30 base pairs, about 21 to about 40 base pairs, about 21 to about 39 base pairs, about 21 to about 37 base pairs, about 23 to about 40 base pairs, about 23 to about 39 base pairs, about 23 to about 37 base pairs, about 25 to about 40 base pairs, about 25 to about 39 base pairs, about 25 to about 37 base pairs, about 25 to about 35 base pairs, or about 28 or 29 base pairs.

[0174] Suitably, the number of repeats may range from about 2 to about 140, from about 2 to about 100, from about 2 to about 100, from about 5 to about 100, from about 10 to about 100, from about 15 to about 100, from about 20 to about 100, from about 25 to about 100, from about 30 to about 100, from about 35 to about 100, from about 40 to about 100, from about 45 to about 100, from about 50 to about 100.

[0175] Suitably, the number of repeats may range from about 2 to about 135, from about 2 to about 130, from about 2 to about 125, from about 2 to about 120, from about 2 to about 115, from about 2 to about 110, from about 2 to about 105, from about 2 to about 100, from about 2 to about 95, from about 2 to about 90, from about 2 to about 80, from about 2 to about 70, from about 2 to about 60, from about 2 to about 50, from about 2 to about 40, from about 2 to about 30, from about 2 to about 20, from about 2 to about 10, from about 2 to about 9, from about 2 to about 8, from about 2 to about 7, from about 2 to about 6, from about 2 to about 5, from about 2 to about 4, or from about 2 to about 3.

[0176] The CRISPR repeat(s) may comprise, consist or consist essentially of DNA or RNA of genomic, synthetic or recombinant origin.

[0177] The CRISPR repeat(s) may be double-stranded or single-stranded whether representing the sense or antisense strand or combinations thereof.

[0178] The CRISPR repeat(s) may be prepared by use of recombinant DNA techniques (e.g. recombinant DNA), as described herein.

[0179] One or more of the CRISPR repeats may be used to engineer a cell - such as a recipient cell. In particular, one or more, preferably, two or more CRISPR repeats may be used to engineer a cell - such as a recipient cell - that in combination with one or more cas genes or proteins and one or more CRISPR spacers can be used to modulate the resistance of a cell against a target nucleic acid or a transcription product thereof. By way of example, the CRISPR repeat(s) may be inserted into the DNA of a cell (eg. a recipient cell) - such as plasmid DNA or

genomic DNA of a cell - using various methods that are well known in the art. By way of further example, the CRISPR repeat(s) may be used as a template upon which to modify (eg. mutate) the DNA of a cell (eg. a recipient cell) - such as plasmid DNA or genomic DNA- such that CRISPR repeat(s) are created or engineered in the DNA of the cell. By way of further example, CRISPR repeat(s) may be cloned into a construct, a plasmid or a vector and the like which is then transformed into the cell, using methods such as those described herein.

[0180] A method for identifying a CRISPR repeat for use in modulating the resistance of a cell against a target nucleic acid or transcription product thereof comprising the steps of: (i) preparing a cell comprising at least one CRISPR spacer and at least one *cas* gene; (ii) engineering the cell such that it contains a CRISPR repeat; and (iii) determining if the cell modulates resistance against the target nucleic acid or transcription product thereof, wherein modulation of the resistance of the cell against the target nucleic acid or transcription product thereof is indicative that the CRISPR repeat can be used to modulate resistance is disclosed herein.

[0181] Suitably, one or more *cas* genes or proteins are used together with or in combination with two or more CRISPR repeats and optionally one or more CRISPR spacers. The *cas* gene(s) or protein(s) and CRISPR repeat(s) form a functional combination as described below.

[0182] A CRISPR spacer is flanked by two CRISPR repeats. In other words, a CRISPR spacer has at least one CRISPR repeat on each side.

[0183] The CRISPR repeats may comprise or consist of the nucleotide sequence set forth in any one or more of SEQ ID Nos. 1-22.

FUNCTIONAL COMBINATION

[0184] As mentioned above, surprisingly, the inventors have discovered that a given set of *cas* genes or proteins is always associated with a given repeated sequence within a particular CRISPR locus. In other words, *cas* genes or proteins seem to be specific for a given DNA repeat, meaning that *cas* genes or proteins and the repeated sequence form a functional pair.

[0185] Accordingly, particular combinations of one or more *cas* genes or proteins and two or more CRISPR repeats are used in order for a CRISPR spacer to confer resistance against a target nucleic acid or transcription product thereof in a cell (eg. a recipient cell). Accordingly, it has been surprisingly found that it is not possible to merely use any *cas* genes or proteins or any CRISPR repeat. Instead it is a feature of the present invention that the combination is functional.

[0186] In the context of the CRISPR repeat-*cas* gene or protein combination described herein, the term "functional" means that the combination is able to confer resistance to a target nucleic acid or a transcription product thereof when used together with a CRISPR spacer which aligns

with or is homologous to a target nucleic acid or transcription product thereof.

[0187] As used herein the term "functional CRISPR repeat-cas combination" and "functional CRISPR repeat-cas gene combination" includes a functional combination in which *cas* is a *cas* gene or a Cas protein.

[0188] The one or more *cas* genes or proteins and the two or more CRISPR repeats are or are derivable (preferably, derived) from the same cell (eg. the same recipient cell).

[0189] In one embodiment, the term "derivable" is synonymous with the term "obtainable".

[0190] In one embodiment, the term "derived" is synonymous with the term "obtained".

[0191] The one or more *cas* genes or proteins and the two or more CRISPR repeats are derivable (preferably, derived) from the same CRISPR locus within a genome or plasmid, preferably a genome or plasmid of the same strain, species or genera.

[0192] Suitably, the one or more *cas* genes or proteins and the two or more CRISPR repeats are derivable (preferably, derived) from the same CRISPR locus within a single genome or plasmid, preferably a single genome or plasmid of the same strain, species or genera.

[0193] Suitably, the one or more *cas* genes or proteins and the two or more CRISPR repeats naturally co-occur.

[0194] Suitably, the one or more *cas* genes or proteins and the two or more CRISPR repeats naturally co-occur in the same cell (eg. recipient cell).

[0195] Suitably, the one or more *cas* genes or proteins and the two or more CRISPR repeats naturally co-occur in the same genome of a cell (eg. recipient cell).

[0196] Suitably, the one or more *cas* genes or proteins and the two or more CRISPR repeats naturally co-occur in the same genome of a strain, species or genera.

[0197] A combination or nucleic acid consisting essentially of at least two CRISPR repeats and at least one *cas* gene or protein is disclosed herein.

[0198] The term "consists essentially of" may refer to a combination of at least two CRISPR repeats and at least one *cas* gene or protein and excluding at least one further component of a CRISPR locus - such as the absence of one or more CRISPR spacer(s) and/or the absence of

one or more common leader sequence(s) of a CRISPR locus.

[0199] Alternatively, the term "consists essentially of" may refer to a combination of at least two CRISPR repeats and at least one *cas* gene or protein only and excluding all other components of a CRISPR locus - such as a naturally occurring CRISPR locus.

[0200] Alternatively, the term "consists essentially of" may refer to a combination of at least two CRISPR repeats and at least one *cas* gene or protein only and excluding at least one further component of a CRISPR locus - preferably excluding at least one further component of a naturally occurring CRISPR locus.

[0201] Alternatively, the term "consists essentially of" may refer to a combination of at least two CRISPR repeats and at least one *cas* gene or protein with the proviso that at least one further component of the natural CRISPR locus is absent (eg. substantially absent).

[0202] A combination of at least two CRISPR repeats and at least one *cas* gene or protein with the proviso that all other components of the CRISPR locus are absent (eg. substantially absent), preferably that all other components of the CRISPR locus of the natural combination of CRISPR repeat(s) and *cas* gene(s) are absent is disclosed herein.

[0203] Suitably, the one or more *cas* genes or proteins are used in combination or together with one or more CRISPR spacers.

[0204] Suitably, the one or more *cas* genes or proteins are used in combination or together with at least one or more CRISPR spacers and at least two or more CRISPR repeats.

[0205] Various arrangements of CRISPR repeats(s) and *cas* gene(s) or protein(s) - such as functional CRISPR repeat-*cas* combinations - are contemplated.

[0206] The combination may comprise, consist or consist essentially of at least any of 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 CRISPR repeat(s) in combination with any of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 *cas* genes or proteins - such as 16 CRISPR repeat and 12 *cas* genes or proteins or 18 CRISPR repeats and 20 *cas* genes or proteins or any other combinations thereof.

[0207] The CRISPR repeat(s) and *cas* gene(s) may be arranged in various ways.

[0208] The combination may be *cas1*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats), *cas2*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats), *cas3*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats), *cas4*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats), *cas1B*-repeat (wherein the repeat is at least two repeats,

preferably, with at least one spacer in between the repeats), *cas5*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats), and/or *cas6*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats),.

[0209] The *cas* gene may be *cas1*, *cas2*, *cas3*, *cas4*, *cas1B*, *cas5* and/or *cas6* or a fragment, variant, homologue or derivative thereof.

[0210] The *cas* genes may be *cas1*, *cas2*, *cas3*, *cas4*, *cas1B*, *cas5* and/or *cas6* or a plurality thereof or a combination thereof - such as *cas1* and *cas2*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas1* and *cas3*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas1* and *cas4* -repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas1* and *cas1B*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas1* and *cas5*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) ; *cas1* and *cas6*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas2* and *cas3* -repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas2* and *cas4*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) ; *cas2* and *cas1b*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) ; *cas2* and *cas5*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas2* and *cas6*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) ; *cas3* and *cas4* -repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas3* and *cas1B*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) ; *cas3* and *cas5*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) ; *cas3* and *cas6*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas4* and *cas1B*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas4* and *cas5*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) ; *cas4* and *cas6*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) ; *cas1B* and *cas5* -repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas1B* and *cas6*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas1*, *cas2* and *cas3*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas1*, *cas2* and *cas4*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas1*, *cas2* and *cas1B*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas1*, *cas2*, *cas3* and *cas4*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas1*, *cas2*, *cas3* and *cas1B*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats);

[illegible]

cas3, *cas4*, *cas1B* and *cas6*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) ; *cas1*, *cas2*, *cas3*, *cas4*, *cas1B* and *cas5*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas1*, *cas2*, *cas3*, *cas4*, *cas1B* and *cas6*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas1*, *cas2*, *cas3*, *cas4*, *cas1B*, *cas5* and *cas6* -repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) or combinations thereof.

[0213] The *cas* genes may be *cas2*, *cas3* and *cas4* -repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas2*, *cas3* and *cas1B*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas2*, *cas3* and *cas5*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas2*, *cas3* and *cas6*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas2*, *cas3*, *cas4* and *cas1B*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) ; *cas2*, *cas3*, *cas4*, and *cas5*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas2*, *cas3*, *cas4* and *cas6*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) ; *cas2*, *cas3*, *cas4*, *cas1B* and *cas5*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) ; *cas2*, *cas3*, *cas4*, *cas1B* and *cas6*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) ; *cas2*, *cas3*, *cas4*, *cas1B*, *cas5* and *cas6* -repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) or combinations thereof.

[0214] The *cas* genes may be *cas3*, *cas4* and *cas1B*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas3*, *cas4* and *cas5*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas3*, *cas4* and *cas6*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) ; *cas3*, *cas4*, *cas1B* and *cas5*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas3*, *cas4*, *cas1B* and *cas6*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas3*, *cas4*, *cas1B*, *cas5* and *cas6* -repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas4*, *cas1B* and *cas5*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas4*, *cas1B* and *cas6*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) ; *cas4*, *cas1B*, *cas5* and *cas6* -repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas5* and *cas6* -repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) or combinations thereof.

[0215] Where the combination of a *cas* gene and a CRISPR repeat comprises more than one *cas* gene, it will be understood that the CRISPR repeat may be inserted at the 3' end of the *cas*

genes, the 5' end of the *cas* genes, or even in between the *cas* genes, provided that at least one of the *cas* genes remains functional.

[0216] In one embodiment, a first CRISPR repeat-*cas* gene or protein combination (comprising at least one *cas* gene or protein and at least two CRISPR repeats, wherein both are derivable (preferably, derived) from the same CRISPR locus within a genome) may be used in combination with a second CRISPR repeat-*cas* gene or protein combination (comprising at least one *cas* gene or protein and at least two CRISPR repeats, wherein both are derivable (preferably, derived) from the same or a different CRISPR locus within a genome). Accordingly, in this embodiment of the invention, the first and second combination are derivable (preferably, derived) from the same or different CRISPR loci within a genome.

[0217] Thus the first and second CRISPR repeat-*cas* gene or protein combinations may even be from different genomes - such as different genomes within the same cluster, as described in further detail herein.

[0218] In a still further embodiment of the present invention, a first and/or a second CRISPR repeat-*cas* gene or protein combination (comprising at least one *cas* gene and at least two CRISPR repeats derivable (preferably, derived) from the same CRISPR locus within a genome) may be used in combination with 3, 4, 5, 6, 7, 8, 9 or 10 or more CRISPR repeat-*cas* gene or protein combinations (each comprising at least one *cas* gene or protein and at least two CRISPR repeats derivable (preferably, derived) from the same or a different CRISPR loci within a genome). Accordingly, in this embodiment of the invention, the combinations are derivable (preferably, derived) from the same or different CRISPR loci within a genome.

[0219] In a further embodiment of the invention, the combinations may even be from different genomes - such as different genomes within the same cluster, as described in further detail herein.

[0220] In other words, for the CRISPR-repeat-*cas* gene or protein combination to confer resistance, in some embodiments, the CRISPR-repeat(s) and *cas* gene(s) or protein(s) naturally co-occur within a given CRISPR locus of a genome. In some embodiments, the CRISPR-repeat(s) and *cas* gene(s) or protein(s) naturally co-occur within the same CRISPR locus of a genome. These functional combinations together may confer resistance against a target nucleic acid or a transcription product thereof.

[0221] A method for identifying a functional combination of a *cas* gene or protein and a CRISPR repeat comprising the steps of: (i) analysing the sequences (eg. nucleic acid or protein sequences) of the *cas* gene or protein and the CRISPR repeat; (ii) identifying one or more clusters of *cas* genes or proteins; (iii) identifying one or more clusters of CRISPR repeats; and (iv) combining those *cas* gene or protein and CRISPR repeat sequences that fall within the same cluster is disclosed herein.

[0222] A method for identifying a functional combination of a *cas* gene or protein and a

CRISPR repeat for use in modulating the resistance of a cell against a target nucleic acid or a transcription product thereof comprising the steps of: (i) preparing a cell comprising a combination of one or more

cas genes or proteins and one or more, preferably, two or more CRISPR repeats; (ii) engineering the cell such that it contains one or more CRISPR spacers; and (iii) determining if the cell modulates resistance against a target nucleic acid, wherein modulation of the resistance of the cell against the target nucleic acid or a transcription product thereof is indicative that the combination can be used to modulate the resistance of the cell against the target nucleic acid is disclosed herein.

[0223] The sequences of the *cas* gene or protein and the CRISPR repeat are or are derivable (preferably, derived) from the same strains.

[0224] The sequences of the *cas* gene or protein and the CRISPR repeat are or are derivable (preferably, derived) from the same species.

[0225] The sequences of the *cas* gene or protein and the CRISPR repeat are or are derivable (preferably, derived) from the same genera.

[0226] The sequences of the *cas* gene or protein and the CRISPR repeat are or are derivable (preferably, derived) from the same organism.

[0227] Suitably, the analysis is performed using dotplot analysis.

[0228] The combination may comprise, consist or consist essentially of DNA or RNA of genomic, synthetic or recombinant origin.

[0229] The combination may comprise, consist or consist essentially of DNA and RNA of genomic, synthetic or recombinant origin.

[0230] The combination may comprise, consist or consist essentially of a DNA CRISPR repeat of genomic, synthetic or recombinant origin.

[0231] The combination may comprise, consist or consist essentially of a RNA CRISPR repeat of genomic, synthetic or recombinant origin.

[0232] The combination may comprise, consist or consist essentially of a DNA *cas* gene repeat of genomic, synthetic or recombinant origin.

[0233] The combination may comprise, consist or consist essentially of a RNA *cas* gene of genomic, synthetic or recombinant origin.

[0234] The combination may comprise, consist or consist essentially of a DNA CRISPR repeat and DNA *cas* gene of genomic, synthetic or recombinant origin.

[0235] The combination may comprise, consist or consist essentially of a DNA CRISPR repeat and RNA *cas* gene of genomic, synthetic or recombinant origin.

[0236] The combination may comprise, consist or consist essentially of a RNA CRISPR repeat and DNA *cas* gene of genomic, synthetic or recombinant origin.

[0237] The combination may comprise, consist or consist essentially of a RNA CRISPR repeat and RNA *cas* gene of genomic, synthetic or recombinant origin.

[0238] The CRISPR repeat may be double-stranded or single-stranded whether representing the sense or antisense strand or combinations thereof.

[0239] The *cas* gene may be double-stranded or single-stranded whether representing the sense or antisense strand or combinations thereof.

[0240] The CRISPR repeat may be double-stranded or single-stranded whether representing the sense or antisense strand or combinations thereof and the *cas* gene may be double-stranded or single-stranded whether representing the sense or antisense strand or combinations thereof.

[0241] The CRISPR repeat may be double-stranded whether representing the sense or antisense strand or combinations thereof and the *cas* gene may be double-stranded whether representing the sense or antisense strand or combinations thereof.

[0242] The CRISPR repeat may be double-stranded whether representing the sense or antisense strand or combinations thereof and the *cas* gene may be single-stranded whether representing the sense or antisense strand or combinations thereof.

[0243] The CRISPR repeat may be single-stranded whether representing the sense or antisense strand or combinations thereof and the *cas* gene may be double-stranded whether representing the sense or antisense strand or combinations thereof.

[0244] The CRISPR repeat may be single-stranded whether representing the sense or antisense strand or combinations thereof and the *cas* gene may be single-stranded whether representing the sense or antisense strand or combinations thereof.

[0245] One or more of the functional combinations as described above may be used to engineer a cell - such as a recipient cell. In particular, one or more functional combinations may be used to engineer a cell - such as a recipient cell - that in combination with one or more CRISPR spacers can be used to modulate the resistance of a cell against a target nucleic acid or a transcription product thereof. By way of example, the functional combinations may be inserted into the DNA of a cell (eg. a recipient cell) - such as plasmid DNA or genomic DNA of a cell - using various methods that are well known in the art. By way of further example, the

functional combinations may be used as a template upon which to modify (*eg.* mutate) the DNA of a cell (*eg.* a recipient cell) - such as plasmid DNA or genomic DNA- such that functional combinations are created in the DNA of the cell. By way of further example, functional combinations may be cloned into a construct, a plasmid or a vector and the like which are then transformed into the cell, using methods such as those described herein.

[0246] Described herein the functional combination is obtained or obtainable by a method comprising the steps of: (a) analysing the sequences of a *cas* gene and a CRISPR repeat; (b) identifying one or more clusters of *cas* genes; (c) identifying one or more clusters of CRISPR repeats; and (d) combining those *cas* gene and CRISPR repeat sequences that fall within the same cluster, wherein the combination of the *cas* gene and CRISPR repeat sequences within the same cluster is indicative that the combination is a functional combination.

[0247] Clusters are described in further detail below.

CRISPR SPACER

[0248] As used herein, the term "CRISPR spacer" has the conventional meaning as used in the art and refers to the non-repetitive spacer sequences that are found between multiple short direct repeats (*i.e.* CRISPR repeats) of CRISPR loci. In other words, a CRISPR spacer is found between two CRISPR repeats.

[0249] It has been found that CRISPR spacer sequences in prokaryotes often have significant similarities to a variety of DNA molecules - such as genetic elements (including, but not limited to, chromosomes, bacteriophages, and conjugative plasmids). Interestingly, cells carrying these CRISPR spacers are unable to be infected by DNA molecules containing sequences homologous to the spacers (Mojica *et al.* 2005).

[0250] Typically, the CRISPR spacer is naturally present in between two identical multiple short direct repeats that are palindromic.

[0251] Suitably, the CRISPR spacer is homologous to the target nucleic acid or a transcription product thereof or an identified sequence. Although homology can also be considered in terms of similarity, in the context of the present invention it is preferred to express homology in terms of sequence identity. A homologous sequence is taken to include a CRISPR spacer, which may be at least 70, 75, 80, 85 or 90% identical, or at least 91, 92, 93, 94, 95, 96, 97, 98 or 99 % identical to the target nucleic acid sequence or a transcription product thereof or an identified sequence.

[0252] In some embodiments,, the CRISPR spacer is 100% identical to the target nucleic acid sequence.

[0253] The number of CRISPR spacers at a given CRISPR loci or locus can vary between

species.

[0254] Suitably, the number of spacers may range from about 1 to about 140, from about 1 to about 100, from about 2 to about 100, from about 5 to about 100, from about 10 to about 100, from about 15 to about 100, from about 20 to about 100, from about 25 to about 100, from about 30 to about 100, from about 35 to about 100, from about 40 to about 100, from about 45 to about 100, from about 50 to about 100.

[0255] Suitably, the number of spacers may range from about 1 to about 135, from about 1 to about 130, from about 1 to about 125, from about 1 to about 120, from about 1 to about 115, from about 1 to about 110, from about 1 to about 105, from about 1 to about 100, from about 1 to about 95, from about 1 to about 90, from about 1 to about 80, from about 1 to about 70, from about 1 to about 60, from about 1 to about 50, from about 1 to about 40, from about 1 to about 30, from about 1 to about 20, from about 1 to about 10, from about 1 to about 9, from about 1 to about 8, from about 1 to about 7, from about 1 to about 6, from about 1 to about 5, from about 1 to about 4, from about 1 to about 3, from about 1 to about 2.

[0256] Typically, CRISPR spacers are identified by sequence analysis as the DNA stretches located in between two repeats.

[0257] As described herein, the inventors have surprisingly discovered that the use of one or more *cas* genes or proteins in combination with one or more, preferably, two or more CRISPR repeats (preferably, functional combination(s) thereof) provides for the specificity of immunity to be conferred by one or more CRISPR spacers in a cell - such as a recipient cell.

[0258] As used herein, the term "specificity of immunity" means that immunity can be conferred against a specific nucleic acid sequence or transcription product thereof using a specific CRISPR spacer (or pseudo CRISPR spacer) sequence. Accordingly, a given CRISPR spacer does not confer resistance against any nucleic acid sequence or transcription product thereof but only to those sequences against which the CRISPR spacer (or pseudo CRISPR spacer) is homologous - such as 100% identical.

[0259] The CRISPR spacer(s) may be or may be derivable (preferably, derived) from an organism - such as a donor organism - that is different to the cell - such as the recipient cell or even a further donor organism - from which the one or more *cas* genes or proteins and/or the one or more, preferably, two or more CRISPR repeats are or are derivable (preferably, derived). The CRISPR spacers may be or may be derivable (preferably, derived) from an organism - such as a donor organism - that is heterologous to the cell - such as the recipient cell or even a further donor organism - from which the one or more *cas* genes or proteins and/or the one or more, preferably, two or more CRISPR repeats are or are derivable (preferably, derived). The one or more *cas* genes or proteins and/or the one or more, preferably, two or more CRISPR repeats may be or may be derivable (preferably, derived) from a homologous (*ie.* the same) cell - such as a homologous recipient cell.

[0260] For the avoidance of doubt, the CRISPR spacer(s) may be designed and produced synthetically (eg. using recombinant DNA techniques).

[0261] In one embodiment, the CRISPR spacers are heterologous (*ie.* different) to the cell - such as the recipient cell - from which the one or more *cas* genes or proteins and/or the one or more, preferably, two or more CRISPR repeats are or are derivable (preferably, derived) and the one or more *cas* genes or proteins and/or the one or more, preferably, two or more CRISPR repeats are or are derivable (preferably, derived) from a homologous cell - such as a homologous recipient cell.

[0262] In another embodiment, the CRISPR spacers are heterologous (*ie.* different) to the cell - such as the recipient cell - from which the one or more *cas* genes or proteins are or are derivable (preferably, derived).

[0263] In another embodiment, the CRISPR spacers are heterologous to the cell - such as the recipient cell and the one or more *cas* genes or proteins and/or the one or more, preferably, two or more CRISPR repeats are or are derivable (preferably, derived) from a homologous cell - such as a homologous recipient cell.

[0264] In another embodiment, the CRISPR spacers are heterologous to the cell - such as the recipient cell - whereas the one or more *cas* genes or proteins and/or the one or more, preferably, two or more CRISPR repeats is/are homologous to the cell - such as the recipient cell.

[0265] In another embodiment, the CRISPR spacers are heterologous to the recipient cell, whereas the recipient cell is homologous for the one or more *cas* genes or proteins and/or the one or more, preferably, two or more CRISPR repeats.

[0266] In another embodiment, the CRISPR spacer used in accordance with the present invention is one which is not naturally associated with the CRISPR repeat and/or *cas* genes and/or functional CRISPR repeat-*cas* gene combination. In other words, the CRISPR spacer in the recombinant CRISPR locus according to the present invention is heterologous to the CRISPR repeat and/or *cas* genes of the CRISPR locus.

[0267] One or more of CRISPR spacers may be used to engineer a cell - such as a recipient cell. In particular, one or more CRISPR spacers may be used to engineer a cell - such as a recipient cell - that in combination with one or more *cas* genes or proteins and/or one or more, preferably, two or more CRISPR repeats (preferably, one or more functional combination thereof) can be used to modulate the resistance of a cell against a target nucleic acid or a transcription product thereof.

[0268] Suitably one or more of CRISPR spacers may be used to engineer a cell - such as a recipient cell. In particular, one or more CRISPR spacers are used to engineer a cell - such as a recipient cell - that in combination with one or more *cas* genes or proteins can be used to

modulate the resistance of a cell against a target nucleic acid or a transcription product thereof.

[0269] By way of example, the CRISPR spacers may be inserted into the DNA of a cell (eg. a recipient cell) - such as plasmid DNA or genomic DNA of a cell - using various methods that are well known in the art.

[0270] By way of further example, the CRISPR spacers may be used as a template upon which to modify (eg. mutate) the DNA of a cell (eg. a recipient cell) - such as plasmid DNA or genomic DNA- such that CRISPR spacers are created in the DNA of the cell.

[0271] By way of further example, CRISPR spacers may be cloned into a construct, a plasmid or a vector and the like which are then transformed into the cell, using methods such as those described herein.

[0272] A method for identifying a CRISPR spacer for use in modulating the resistance of a cell against a target nucleic acid or a transcription product thereof comprising the steps of: (i) preparing a cell comprising at least two CRISPR repeats and at least one *cas* gene or protein; (ii) identifying at least one CRISPR spacer in an organism - such as a donor organism; (iii) modifying the sequence of the CRISPR spacer of the cell such that it has homology to the CRISPR spacer of the donor organism comprising the target nucleic acid; and (iv) determining if the cell modulates resistance against the target nucleic acid, wherein modulation of the resistance of the cell against the target nucleic acid or transcription product thereof is indicative that the CRISPR spacer modulates the resistance of the cell against the target nucleic acid is disclosed herein,

[0273] The CRISPR spacers may comprise or consist of the nucleotide sequence set forth any one or more of in any of SEQ ID Nos. 23-460 and/or SEQ ID Nos. 522-665.

[0274] A CRISPR spacer is flanked by two CRISPR repeats. In other words, a CRISPR spacer has at least one CRISPR repeat on each side.

[0275] Without wishing to be bound by any particular theory, the further a given CRISPR spacer is from the 5' end of the CRISPR locus (comprising the *cas* gene(s) and/or the leader sequence), the lower the resistance conferred by that CRISPR spacer may be. Accordingly, in one embodiment of the present invention it is preferred that one or more of the first 100 CRISPR spacers from the 5' end of the CRISPR locus (comprising the *cas* genes and/or the leader sequence) are modified, more preferably, that one or more of the first 50 CRISPR spacers from the 5' end of the CRISPR locus (comprising the *cas* genes and/or the leader sequence) are modified, more preferably, that one or more of the first 40 CRISPR spacers from the 5' end of the CRISPR locus (comprising the *cas* genes and/or the leader sequence) are modified, more preferably, that one or more of the first 30 CRISPR spacers from the 5' end of the CRISPR locus (comprising the *cas* genes and/or the leader sequence) are modified, more preferably, that one or more of the first 20 CRISPR spacers from the 5' end of the

CRISPR locus (comprising the cas genes and/or the leader sequence) are modified, more preferably, that one or more of the first 15 CRISPR spacers from the 5' end of the CRISPR locus (comprising the cas genes and/or the leader sequence) are modified, more preferably, that one or more of the first 10 CRISPR spacers from the 5' end of the CRISPR locus (comprising the cas genes and/or the leader sequence) are modified.

[0276] As will be appreciated by the skilled person, different bacteria have different numbers of CRISPR spacers.

CRISPR SPACER CORE

[0277] For a specific CRISPR type within a microbial species, the CRISPR spacer is typically represented by a defined predominant length, but the size may vary. CRISPR types described to date have been found to contain a predominant spacer length of between about 20 bp and about 58 bp.

[0278] As used herein, the term "CRISPR spacer core" means the length of the shortest observed spacer within a CRISPR type. Thus, by way of example, within *Streptococcus thermophilus* CRISPR Type 1, the dominant spacer length is 30 bp with a minority of spacers between 28 bp and 32 bp in size. So in this particular example (*S. thermophilus* CRISPR Type 1), the CRISPR spacer core is defined as a continuous stretch of 28 bp.

[0279] Suitably, the CRISPR spacer core is homologous to the target nucleic acid or a transcription product thereof or an identified sequence over the length of the core sequence. Although homology can also be considered in terms of similarity, in the context of the present invention it is preferred to express homology in terms of sequence identity. A homologous sequence is taken to include a CRISPR spacer core, which may be at least 90% identical or at least 91, 92, 93, 94, 95, 96, 97, 98 or 99 % identical to the target nucleic acid sequence or a transcription product thereof or an identified sequence over the length of the core sequence.

[0280] Suitably, the CRISPR spacer core is 100% identical to the target nucleic acid sequence or a transcription product thereof or an identified sequence over the length of the core sequence.

PSEUDO-CRISPR SPACER

[0281] As used herein, the term "pseudo-CRISPR spacer" refers to a nucleic acid sequence present in an organism (eg. a donor organism) - such as a bacteriophage - which is preferably essential for function and/or survival and/or replication and/or infectivity and the like, and which forms a CRISPR spacer sequence; and/or can be used to form or prepare a CRISPR spacer sequence which is complementary to or homologous to the pseudo-CRISPR spacer; and/or

can be used to modulate resistance.

[0282] One or more of pseudo CRISPR spacers or CRISPR spacer(s) which is/are complementary or homologous to the one or more pseudo CRISPR spacer(s) may be used to engineer a cell - such as a recipient cell. In particular, one or more pseudo CRISPR spacers or CRISPR spacer(s) which is/are complementary or homologous to the one or more pseudo CRISPR spacer(s) may be used to engineer a cell - such as a recipient cell - that in combination with one or more *cas* genes or proteins and/or one or more CRISPR repeats (eg, one or more functional combinations thereof) can be used to modulate the resistance of a cell against a target nucleic acid or a transcription product thereof.

[0283] One or more pseudo CRISPR spacers or CRISPR spacer(s) which is/are complementary or homologous to the one or more pseudo CRISPR spacer(s) may be used to engineer a cell - such as a recipient cell - that in combination with one or more *cas* genes or proteins can be used to modulate the resistance of a cell against a target nucleic acid or a transcription product thereof.

[0284] By way of example, the pseudo CRISPR spacers or CRISPR spacer(s) which is/are complementary or homologous to the one or more pseudo CRISPR spacer(s) may be inserted into the DNA of a cell (eg. a recipient cell) - such as plasmid DNA or genomic DNA of a cell - using various methods that are well known in the art.

[0285] By way of further example, the pseudo CRISPR spacers may be used as a template upon which to modify (eg. mutate) the DNA of a cell (eg. a recipient cell) - such as plasmid DNA or genomic DNA- such that CRISPR spacers are created in the DNA of the cell. By way of further example, pseudo CRISPR spacers or CRISPR spacer(s) which is/are complementary or homologous to the one or more pseudo CRISPR spacer(s) may be cloned into a construct, a plasmid or a vector and the like which are then transformed into the cell, using methods such as those described herein.

NUCLEIC ACID SEQUENCE

[0286] A nucleic acid sequence (eg. a recombinant or an isolated nucleic acid sequence) consisting essentially of at least one *cas* gene or protein is disclosed herein.

[0287] The nucleic acid sequence may be DNA or RNA of genomic, synthetic or recombinant origin e.g. cDNA. The nucleotide sequence may be double-stranded or single-stranded whether representing the sense or antisense strand or combinations thereof. Recombinant nucleic acid sequences may be prepared by use of recombinant DNA techniques, as described herein. The target nucleic acid sequence may be or may be derived from a gene.

[0288] As used herein, the term "consisting essentially of" in the context of the nucleic acid sequence refers to a nucleic acid sequence comprising one or more *cas* genes and excluding

at least one further component of a CRISPR locus - such as the CRISPR repeats, the CRISPR spacers and/or the common leader sequence.

[0289] A nucleic acid sequence consisting essentially of at least one *cas* gene and at least two CRISPR repeats is disclosed herein.

[0290] A nucleic acid sequence consisting essentially of at least one *cas* gene and at least one CRISPR spacer is disclosed herein.

[0291] A nucleic acid sequence consisting essentially of at least one *cas* gene, at least one CRISPR spacer and at least two CRISPR repeats is disclosed herein.

[0292] A nucleic acid sequence comprising at least one *cas* gene with the proviso that at least one further component of a CRISPR locus is absent, suitably, with the proviso that at least one further component of a natural CRISPR locus is absent (eg. substantially absent) is disclosed herein.

[0293] A nucleic acid sequence comprising at least one *cas* gene with the proviso that the CRISPR spacers of the CRISPR locus are absent, suitably, with the proviso that CRISPR spacers of a natural CRISPR locus are absent (eg. substantially absent) is disclosed herein.

[0294] A nucleic acid sequence comprising at least one *cas* gene with the proviso that the CRISPR repeats of the CRISPR locus are absent, suitably, with the proviso that the CRISPR repeats of a natural CRISPR locus are absent is disclosed herein.

[0295] A nucleic acid sequence comprising at least one *cas* gene with the proviso that the common leader sequences of the CRISPR locus are absent, suitably, with the proviso that the common leader sequences of the natural CRISPR locus are absent is disclosed herein.

[0296] A nucleic acid sequence comprising at least one *cas* gene with the proviso that the CRISPR spacers and the CRISPR repeats of the CRISPR locus are absent, suitably, with the proviso that the CRISPR spacers and the CRISPR repeats of the natural CRISPR locus are absent is disclosed herein.

[0297] A nucleic acid sequence comprising at least one *cas* gene with the proviso that the CRISPR spacers and the CRISPR repeats of the CRISPR locus are absent, suitably, with the proviso that the CRISPR spacers and the CRISPR repeats of the natural CRISPR locus are absent is disclosed herein.

[0298] A nucleic acid sequence comprising at least one *cas* gene with the proviso that the CRISPR spacers and the common leader sequences of the CRISPR locus are absent, suitably, with the proviso that the CRISPR spacers and the common leader sequences of the natural CRISPR locus are absent is disclosed herein.

[0299] A nucleic acid sequence comprising at least one *cas* gene with the proviso that the CRISPR repeats and the common leader sequences of the CRISPR locus are absent, suitably, with the proviso that the CRISPR repeats and the common leader sequences of the natural CRISPR locus are absent is disclosed herein.

[0300] A nucleic acid sequence comprising at least one *cas* gene with the proviso that the CRISPR repeats, the CRISPR spacers and the common leader sequences of the CRISPR locus are absent, suitably, with the proviso that the CRISPR repeats, the CRISPR spacers and the common leader sequences of the natural CRISPR locus are absent is disclosed herein.

[0301] The nucleic acid sequence and the nucleic acids may be isolated or substantially purified. By "isolated" or "substantially purified" is intended that the nucleic acid molecules, or biologically active fragments or variants, homologues, or derivatives thereof are substantially or essentially free from components normally found in association with the nucleic acid in its natural state. Such components include other cellular material, culture media from recombinant production, and various chemicals used in chemically synthesising the nucleic acids.

[0302] An "isolated" nucleic acid sequence or nucleic acid is typically free of nucleic acid sequences that flank the nucleic acid of interest in the genomic DNA of the organism from which the nucleic acid was derived (such as coding sequences present at the 5' or 3' ends). However, the molecule may include some additional bases or moieties that do not deleteriously affect the basic characteristics of the composition.

[0303] The nucleic acid sequence(s) may be used in the engineering of a cell - such as a recipient cell. By way of example, the nucleic acid sequence may be inserted into the DNA - such as plasmid DNA or genomic DNA - of a recipient cell, using methods - such as homologous recombination. By way of further example, the nucleic acid sequence(s) may be used as a template upon which to modify (eg. mutate) the DNA of a cell (eg. a recipient cell) - such as plasmid DNA or genomic DNA- such that the nucleic acid sequence(s) are created in the DNA of the cell. By way of further example, the nucleic acid sequence(s) may be cloned into a construct, a plasmid or a vector and the like which are then transformed into the cell, using methods such as those described herein.

[0304] A CRISPR spacer is flanked by two CRISPR repeats. In other words, a CRISPR spacer has at least one CRISPR repeat on each side.

TARGET NUCLEIC ACID SEQUENCE

[0305] As used herein, the term "target nucleic acid sequence" refers to any nucleic acid sequence or transcription product thereof, against which resistance in a cell - such as a recipient cell - is to be modulated.

[0306] The resistance may be against the target nucleic acid sequence *per se*.

Advantageously, this confers resistance to a cell against a donor organism from which the target nucleic acid(s) is derivable (preferably, derived). Thus, by way of example, the insertion of a pseudo CRISPR spacer derivable (preferably, derived) from a bacteriophage or a CRISPR spacer(s) which is/are complementary or homologous to the one or more pseudo CRISPR spacer(s) into a cell - such as a recipient cell - may confer resistance to the bacteriophage. Thus, by way of further example, insertion between two CRISPR repeats of a pseudo CRISPR spacer derivable (preferably, derived) from a bacteriophage or CRISPR spacer(s) which is/are complementary or homologous to the one or more pseudo CRISPR spacer(s) into a cell - such as a recipient cell - may confer resistance to the bacteriophage.

[0307] The resistance may be against a transcription product of the target nucleic acid sequence - such as a transcript of the target nucleic acid sequence (eg. an RNA (eg. mRNA) transcript (eg. a sense or an antisense RNA transcript) or even a polypeptide transcription product. Advantageously, this confers resistance to a cell against a donor organism from which the transcription product is derivable (preferably, derived).

[0308] The target nucleotide sequence may be DNA or RNA of genomic, synthetic or recombinant origin.

[0309] The nucleotide sequence may be double-stranded or single-stranded whether representing the sense or antisense strand or combinations thereof.

[0310] The nucleotide sequence may be prepared by use of recombinant DNA techniques (e.g. recombinant DNA).

[0311] The nucleotide sequence may be the same as a naturally occurring form, or may be derivable (preferably, derived) therefrom.

[0312] The target nucleic acid sequence may be or may be derivable (preferably, derived) from a gene.

[0313] The target nucleic acid sequence may be or may be derivable (preferably, derived) from a variant, homologue, fragment or derivative of a gene.

[0314] In one embodiment, the target nucleic sequence is or is derivable (preferably, derived) from bacteriophage.

[0315] In one embodiment, the target nucleic sequence is or is derivable (preferably, derived) from plasmid DNA.

[0316] In one embodiment, the target nucleic sequence is or is derivable (preferably, derived) from a mobile genetic element.

[0317] In one embodiment, the target nucleic sequence is or is derivable (preferably, derived)

from a transposable element or an insertion sequence.

[0318] In one embodiment, the target nucleic sequence is or is derivable (preferably, derived) from a gene that confers resistance.

[0319] In one embodiment, the target nucleic sequence is or is derivable (preferably, derived) from a gene that confers resistance to an antibiotic.

[0320] In one embodiment, the target nucleic sequence is or is derivable (preferably, derived) from a virulence factor.

[0321] In one embodiment, the target nucleic sequence is or is derivable (preferably, derived) from a toxin, an internalin or a hemolysin.

MODULATING RESISTANCE

[0322] In a further aspect, there is provided a method for modulating the resistance of a cell - such as a recipient cell - against a target nucleic acid or a transcription product thereof.

[0323] As used herein, the term "modulating resistance" may refer to suppressing, reducing, decreasing, inducing, conferring, restoring, elevating, increasing or otherwise affecting the resistance of a cell to a target nucleic acid.

[0324] As used herein, the term "resistance" is not meant to imply that a cell is 100% resistant to a target nucleic acid or a transcription product thereof, but includes cells that are tolerant of the target nucleic acid or a transcription product thereof.

[0325] As used herein the term "resistance to target nucleic acid or transcription product thereof" means that resistance is conferred against a cell or an organism - such as a phage - that comprises or produces the target nucleic acid or transcription product thereof.

[0326] Without being bound by any particular theory, we believe that resistance or immunity is not linked to the "entry" of foreign DNA into a cell (ie. penetration through the cell membrane). Immunity or resistance would rather correspond to an obstruction, hurdle, impediment, barrier or avoidance to persistency, maintenance or survival of the incoming nucleic acid (either, for example, in a free linear form, or integrated within the bacterial chromosome, outside from a CRISPR locus or within a circular molecule - such as a plasmid), or to a obstruction, hurdle, impediment, barrier or avoidance to its replication and/or transcription and/or expression.

[0327] In one embodiment, the minimal components conferring immunity or resistance against a target nucleic acid or expression product thereof is at least one *cas* gene (or one Cas protein) and at least two CRISPR repeats flanking a spacer.

[0328] In one embodiment, it is preferred that "modulating resistance" means inducing, conferring, elevating or increasing the resistance of a cell to a target nucleic acid.

[0329] A method for modulating (e.g. conferring or increasing) the resistance of a cell against a target nucleic acid or a transcription product thereof comprising the steps of: (i) identifying a sequence (eg. a conserved sequence) in an organism (preferably, a sequence essential to the function or survival of the organism); (ii) preparing a CRISPR spacer which is a sequence homologous, (suitably 100% identical), to the identified sequence; (iii) preparing a nucleic acid comprising at least one *cas* gene and at least two CRISPR repeats together with the CRISPR spacer; and (iv) transforming a cell with said nucleic acid thus to render the cell resistant to said target nucleic acid or transcription product thereof is disclosed herein.

[0330] As used herein, the term "conserved sequence" in the context of identifying a conserved sequence in an organism does not necessarily have to be conserved in its strictest sense since the knowledge of one sequence from a given organism will be enough. Furthermore the sequence does not need to be part of an essential entity, since we believe that a spacer inspired from an essential gene would be more efficient in conferring immunity or resistance.

[0331] In one embodiment, the conserved sequence is a sequence that is essential for function and/or survival and/or replication and/or infectivity and the like of an organism or a cell. By way of example, the conserved sequence may be a helicase, a primase a head or tail structural protein, a protein with a conserved domain (eg. holing, lysine, and others) or a conserved sequences amongst important phage genes.

[0332] A method for modulating (eg. conferring or increasing) the resistance of a cell against a target nucleic acid or a transcription product thereof comprising the steps of: (i) identifying one or more CRISPR spacers in an organism resistant to the target nucleic acid or transcription product thereof; (ii) preparing a recombinant nucleic acid comprising at least one *cas* gene or protein and at least two CRISPR repeats together with said identified one or more spacers; and (iii) transforming a cell with said recombinant nucleic acid thus to render the recipient cell resistant to said target nucleic acid or transcription product thereof is disclosed herein.

[0333] A method for modulating (eg. conferring or increasing) the resistance of a cell comprising at least one or more *cas* genes or proteins and one or more, preferably, two or more CRISPR repeats against a target nucleic acid or a transcription product thereof comprising the steps of: (i) identifying one or more CRISPR spacers in an organism resistant to the target nucleic acid or transcription product thereof, and (ii) modifying the sequence of one or more CRISPR spacer(s) in the cell such that the CRISPR spacer(s) has homology to the CRISPR spacer(s) in the organism is disclosed herein.

[0334] Described herein one or more CRISPR spacers in a cell - such as a recipient cell - are modified (eg. genetically engineered) such that the CRISPR spacer(s) have homology to one or more CRISPR spacer(s) in an organism - such as a donor organism - that is substantially

resistant to a target nucleic acid or a transcription product thereof in order to render the cell resistant to the target nucleic acid.

[0335] Suitably, the one or more *cas* genes or proteins and two or more CRISPR repeats in the cell are a functional combination as described herein.

[0336] The genetic engineering may include, but is not limited to, adding (eg. inserting), deleting (eg. removing) or modifying (eg. mutating) the sequence of the one or more CRISPR spacers or in a cell such that the CRISPR spacer has homology (eg. increased homology after the genetic engineering) to one or more CRISPR spacers of a donor organism. This engineering step will result in a cell that was substantially sensitive to a target nucleic acid or a transcription product thereof being substantially resistant to the target nucleic acid or a transcription product thereof.

[0337] The genetic engineering may even include, but is not limited to, adding (eg. inserting) or deleting (eg. removing) the sequence of the one or more pseudo CRISPR spacers in to a cell. This engineering step will result in a cell that was substantially sensitive to a target nucleic acid or a transcription product thereof being substantially resistant to the target nucleic acid or a transcription product thereof.

[0338] Described herein "modulating resistance" means suppressing, reducing or decreasing the resistance of a cell to a target nucleic acid.

[0339] A method for decreasing or reducing the resistance of a cell - such as a recipient cell - comprising at least one or more *cas* genes or proteins and one or more, preferably, two or more CRISPR repeats against a target nucleic acid or a transcription product thereof is disclosed herein.

[0340] Described herein the method comprises the steps of: (i) identifying one or more CRISPR spacers in an organism that is substantially resistant to the target nucleic acid or a transcription product thereof; and (ii) modifying the sequence of one or more CRISPR spacer(s) in the cell such that the CRISPR spacer(s) has a reduced degree of homology to the CRISPR spacer(s) in the organism.

[0341] A method for modulating (eg. decreasing) the resistance of a cell comprising one or more *cas* genes or proteins and one or more, preferably, two or more CRISPR repeats against a target nucleic acid or transcription product thereof comprising the steps of: (i) identifying a CRISPR spacer or a pseudo CRISPR spacer in an organism comprising a target nucleic acid or transcription product thereof against which resistance is to be modulated; and (ii) identifying the CRISPR spacer in the organism in which resistance is to be modulated; and (iii) adapting the sequence of the CRISPR spacer in the organism in which resistance is to be modulated such that the CRISPR spacer has a lower degree of homology to the CRISPR spacer or pseudo CRISPR spacer of the organism comprising the target nucleic acid or transcription product thereof against which resistance is to be modulated is disclosed herein.

[0342] One or more CRISPR spacers in a substantially resistant cell are engineered in order to render the cell sensitive to a target nucleic acid. The genetic engineering may include, but is not limited to, the addition (eg. insertion), deletion (eg. removal) or modification of one or more functional CRISPR repeat-cas combinations or portions or fragments thereof in the substantially resistant cell and/or the addition (eg. insertion), deletion (eg. removal) or modification of one or more CRISPR spacers or portions or fragments thereof in the substantially resistant cell.

[0343] This engineering step will then result in a cell that was substantially resistant to a target nucleic acid or a transcription product thereof becoming substantially sensitive to a target nucleic acid or a transcription product thereof.

[0344] Typically, in order to confer sensitivity to a cell, it is expected that one or more CRISPR spacers, one or more cas genes or proteins, one or more, preferably, two or more CRISPR repeats or one or more functional CRISPR repeat-cas combinations from a substantially resistant cell will be removed, deleted or modified such that resistance is no longer conferred.

[0345] Advantageously, cells that are sensitive to a target nucleic acid or a transcription product thereof may be prepared such that their levels within a given culture - such as a starter culture - may be modulated (eg. decreased) as desired. Thus, by way of example, a starter culture comprising two or more bacterial strains may be developed such that all members of the culture are sensitive to the same agent (eg. bacteriophage). Accordingly, at a time when it is no longer desired for the culture to be alive, the culture may be contacted with the same single agent in order to kill all members of the culture.

[0346] Moreover, it may even be possible to modulate the sensitivity of a cell to one or more agents (eg. bacteriophage) such that the agent kills only a certain proportion of the cells in a given culture -such as 10, 20, 30, 40, 50, 60, 70, 80, 90 or 95% of the cells in a given culture.

[0347] In one aspect, a cell - such as a recipient cell - may be engineered such that it comprises a CRISPR spacer or a sequence corresponding to a pseudo CRISPR spacer thereby rendering the cell resistant to a target nucleic acid or transcription product thereof. Suitably, the cell is engineered such that the CRISPR spacer or sequence corresponding to the pseudo CRISPR spacer is used together with a functional cas gene-CRISPR repeat combination, as described herein.

[0348] In one aspect, a cell that is resistant to a target nucleic acid or transcription product thereof is engineered such that the CRISPR spacer conferring the immunity against the target nucleic acid or transcription product thereof is inserted into a cell that comprises a functional cas gene-CRISPR repeat combination, thereby rendering the cell resistant to the target nucleic acid or transcription product thereof.

[0349] In one aspect, the sequence of one or more CRISPR spacers or pseudo CRISPR

spacers of a cell that is resistant to a target nucleic acid or transcription product thereof is determined. A cell - such as a recipient cell - is then engineered such that it comprises the sequence of the CRISPR spacer and a functional *cas* gene-CRISPR repeat combination, thereby rendering the cell resistant to the target nucleic acid or transcription product thereof.

[0350] In one aspect, a CRISPR spacer from a cell - such as a recipient cell - and a functional *cas* gene-CRISPR repeat combination from the same or different cell - such as the same or different recipient cell - are prepared. A further cell - such as a recipient cell - is then engineered such that it comprises the CRISPR spacer sequence and functional *cas* gene-CRISPR repeat combination thereby rendering the cell resistant to the target nucleic acid or transcription product thereof.

[0351] A CRISPR spacer is flanked by two CRISPR repeats. In other words, a CRISPR spacer has at least one CRISPR repeat on each side.

BACTERIOPHAGE

[0352] In a particularly preferred aspect of the present invention, the resistance of a cell against a bacteriophage is modulated.

[0353] The bacteriophage is virulent to the cell.

[0354] The bacteriophage may be a virulent or a temperate bacteriophage.

[0355] As used herein, the term "bacteriophage" has its conventional meaning as understood in the art *ie.* a virus that selectively infects prokaryotes - such as bacteria. Many bacteriophages are specific to a particular genus or species or strain of cell.

[0356] The bacteriophage may be a lytic bacteriophage or a lysogenic bacteriophage.

[0357] A lytic bacteriophage is one that follows the lytic pathway through completion of the lytic cycle, rather than entering the lysogenic pathway. A lytic bacteriophage undergoes viral replication leading to lysis of the cell membrane, destruction of the cell, and release of progeny bacteriophage particles capable of infecting other cells.

[0358] A lysogenic bacteriophage is one capable of entering the lysogenic pathway, in which the bacteriophage becomes a dormant, passive part of the cell's genome through prior to completion of its lytic cycle.

[0359] The term "bacteriophage" is synonymous with the term "phage".

[0360] Whilst resistance against any bacteriophage (including wild type, naturally occurring, isolated or recombinant bacteriophage) may be employed, bacteriophage active against

bacteria are preferred. More suitably, bacteriophage active against bacteria that are pathogenic to plants and/or animals (including humans) are of particular interest.

[0361] By way of example, the bacteriophage include, but are not limited to, those bacteriophage capable of infecting a bacterium that naturally comprises one or more CRISPR loci. CRISPR loci have been identified in more than 40 prokaryotes (Jansen *et al.* 2002b; Mojica *et al.*, 2005; Haft *et al.*, 2005) including, but not limited to *Aeropyrum*, *Pyrobaculum*, *Sulfolobus*, *Archaeoglobus*, *Halocarcula*, *Methanobacterium*, *Methanococcus*, *Methanosarcina*, *Methanopyrus*, *Pyrococcus*, *Picrophilus*, *Thermoplasma*, *Corynebacterium*, *Mycobacterium*, *Streptomyces*, *Aquifex*, *Porphyromonas*, *Chlorobium*, *Thermus*, *Bacillus*, *Listeria*, *Staphylococcus*, *Clostridium*, *Thermoanaerobacter*, *Mycoplasma*, *Fusobacterium*, *Azarcus*, *Chromobacterium*, *Neisseria*, *Nitrosomonas*, *Desulfovibrio*, *Geobacter*, *Myxococcus*, *Campylobacter*, *Wolinella*, *Acinetobacter*, *Erwinia*, *Escherichia*, *Legionella*, *Methylococcus*, *Pasteurella*, *Photobacterium*, *Salmonella*, *Xanthomonas*, *Yersinia*, *Treponema* and *Thermotoga*.

[0362] By way of example, the bacteriophage include, but are not limited to, those bacteriophage capable of infecting bacteria belonging to the following genera: *Escherichia*, *Shigella*, *Salmonella*, *Erwinia*, *Yersinia*, *Bacillus*, *Vibrio*, *Legionella*, *Pseudomonas*, *Neisseria*, *Bordetella*, *Helicobacter*, *Listeria*, *Agrobacterium*, *Staphylococcus*, *Streptococcus*, *Enterococcus*, *Clostridium*, *Corynebacterium*, *Mycobacterium*, *Treponema*, *Borrelia*, *Francisella*, *Brucella* and *Xanthomonas*.

[0363] By way of further example, the bacteriophage include, but are not limited to, those bacteriophage capable of infecting (or transducing) lactic acid bacteria species, a *Bifidobacterium* species, a *Brevibacterium* species, a *Propionibacterium* species, a *Lactococcus* species, a *Streptococcus* species, a *Lactobacillus* species including the *Lactobacillus acidophilus*, *Enterococcus* species, *Pediococcus* species, a *Leuconostoc* species and *Oenococcus* species.

[0364] By way of further example, the bacteriophage include, but are not limited to, those bacteriophage capable of infecting *Lactococcus lactis*, including *Lactococcus lactis* subsp. *lactis* and *Lactococcus lactis* subsp. *cremoris*, *Lactococcus lactis* subsp. *lactis biovar diacetylactis*, *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus*, *Lactobacillus helveticus*, *Bifidobacterium lactis*, *Lactobacillus acidophilus*, *Lactobacillus casei*, *Bifidobacterium infantis*, *Lactobacillus paracasei*, *Lactobacillus salivarius*, *Lactobacillus plantarum*, *Lactobacillus reuteri*, *Lactobacillus gasseri*, *Lactobacillus johnsonii* or *Bifidobacterium longum*.

[0365] By way of further example, the bacteriophages include, but are not limited to, those bacteriophage capable of infecting any fermentative bacteria susceptible to disruption by bacteriophage infection, including but not limited to processes for the production of antibiotics, amino acids, and solvents. Products produced by fermentation which are known to have encountered bacteriophage infection, and the corresponding infected fermentation bacteria,

include Cheddar and cottage cheese (*Lactococcus lactis* subsp. *lactis*, *Lactococcus lactis* subsp. *cremoris*), Yogurt (*Lactobacillus delbrueckii* subsp. *bulgaricus*, *Streptococcus thermophilus*), Swiss cheese (*S. thermophilus*, *Lactobacillus lactis*, *Lactobacillus helveticus*), Blue cheese (*Leuconostoc cremoris*), Italian cheese (*L. bulgaricus*, *S. thermophilus*), Viili (*Lactococcus lactis* subsp. *cremoris*, *Lactococcus lactis* subsp. *lactis* biovar *diacetylactis*, *Leuconostoc cremoris*), Yakult (*Lactobacillus casei*), casein (*Lactococcus lactis* subsp. *cremoris*), Natto (*Bacillus subtilis* var. *natto*), Wine (*Leuconostoc oenos*), Sake (*Leuconostoc mesenteroides*), Polymyxin (*Bacillus polymyxa*), Colistin (*Bacillus colistrium*), Bacitracin (*Bacillus licheniformis*), L-Glutamic acid (*Brevibacterium lactofermentum*, *Microbacterium ammoniaphilum*), and acetone and butanol (*Colstridium acetobutylicum*, *Clostridium saccharoperbutylacetonicum*).

[0366] Preferred bacteria are *S. thermophilus*, *L. delbrueckii* subsp. *bulgaricus* and/or *L. acidophilus*.

[0367] By way of further example, the bacteriophages include, but are not limited to, those bacteriophage capable of infecting a bacterium that comprises one or more heterologous CRISPR loci. The bacterium may comprise one or more heterologous CRISPR loci, one or more heterologous *cas* genes, one or more heterologous CRISPR repeats and/or one or more heterologous CRISPR spacers.

[0368] Bacteriophages may include, but are not limited to, bacteriophages that belong to any of the following virus families: Corticoviridae, Cystoviridae, Inoviridae, Leviviridae, Microviridae, Myoviridae, Podoviridae, Siphoviridae, or Tectiviridae.

[0369] To cause bacteriophage infection of cells, it "infects" a cell when it injects or transfers its nucleic acid into the cell, with the phage nucleic acid existing independently of the cell's genome. Infection may lead to expression (transcription and translation) of the bacteriophage nucleic acid within the cell and continuation of the bacteriophage life cycle. In the case of recombinant bacteriophage, recombinant sequences within the phage genome, such as reporter nucleic acids, may be expressed as well.

[0370] It has been found that CRISPR spacer sequences in prokaryotes often have significant similarities to a variety of DNA molecules - such as genetic elements (including, but not limited to, chromosomes, bacteriophages, conjugative plasmids). Interestingly, cells carrying these CRISPR spacers are unable to be infected by DNA molecules containing sequences homologous to the spacers (Mojica *et al.* 2005).

[0371] In the context of the present invention, one or more particular pseudo-spacers derivable or derived from bacteriophage DNA or CRISPR spacer(s) which is/are complementary or homologous to the one or more pseudo-CRISPR spacer(s) can be added within a CRISPR locus of a cell - such as a recipient cell - in order to modulate (eg. provide) resistance against a particular bacteriophage, thus substantially preventing phage attack.

[0372] Typically, particular regions within the phage genome may be targeted to prepare the pseudo-spacers - such as genes coding for host specificity proteins - that provide particular phage-host recognition - such as helicases, primase, head or tail structural proteins, proteins with a conserved domain (eg. holing, lysine, and others) or conserved sequences amongst important phage genes.

[0373] Any nucleic acid originating from the phage genome may confer immunity against the phage when inserted, for example, between two repeats in an active CRISPR locus. Immunity may be more "efficient" if the CRISPR spacer corresponds to an internal sequence of a phage gene, and even more "efficient" when this gene encodes "essential" proteins (eg. the antireceptor).

[0374] A method for conferring resistance to a cell (suitably, a bacterial cell) against a bacteriophage comprising the steps of: (a) providing one or more pseudo CRISPR. spacers from at least one bacteriophage; (b) identifying one or more functional CRISPR repeat-cas combinations in at least one cell that is substantially sensitive to the bacteriophage; and (c) engineering the one or more CRISPR loci in the substantially sensitive cell such that they comprise one or more pseudo CRISPR spacers from a bacteriophage or one or more CRISPR spacer(s) which is/are complementary or homologous to the one or more pseudo CRISPR spacer(s) to render the cell resistant is disclosed herein.

[0375] A method for conferring resistance to a cell (suitably, a bacterial cell) against a bacteriophage comprising the steps of: (a) providing one or more pseudo CRISPR spacers from at least one bacteriophage; (b) identifying one or more functional CRISPR repeat-cas combinations in at least one cell that is substantially sensitive to the bacteriophage; and (c) inserting one or more pseudo CRISPR spacers from the bacteriophage or one or more CRISPR spacer(s) which is/are complementary or homologous to the one or more pseudo CRISPR spacer(s) into the substantially sensitive cell such that the cell is rendered substantially resistant to the bacteriophage is disclosed herein.

[0376] A method for modulating the lysotype of a bacterial cell comprising the steps of: (a) providing one or more pseudo CRISPR spacers from at least one bacteriophage; (b) identifying one or more functional CRISPR repeat-cas combinations in at least one cell that is substantially sensitive to the bacteriophage; and (c) engineering the one or more CRISPR loci in the substantially sensitive cell such that they comprise one or more pseudo CRISPR spacers from a bacteriophage or one or more CRISPR spacer(s) which is/are complementary or homologous to the one or more pseudo CRISPR spacer(s) is disclosed herein.

[0377] A method for modulating the lysotype of a bacterial cell comprising the steps of: (a) providing one or more pseudo CRISPR spacers from at least one bacteriophage; (b) identifying one or more functional CRISPR repeat-cas combinations in at least one cell that is substantially sensitive to the bacteriophage; and (c) inserting one or more one or more pseudo CRISPR spacers from the bacteriophage or one or more CRISPR spacer(s) which is/are complementary or homologous to the one or more pseudo CRISPR spacer(s) into the

substantially sensitive cell is disclosed herein.

[0378] A method for conferring resistance to a cell (suitably, a bacterial cell) against a bacteriophage comprising the steps of:

1. (i) identifying a pseudo CRISPR spacer in a bacteriophage comprising a target nucleic acid or a transcription product thereof against which resistance is to be modulated; and
2. (ii) modifying the sequence of the CRISPR spacer of the cell such that the CRISPR spacer of the cell has homology to the pseudo CRISPR spacer of the bacteriophage comprising the target nucleic acid is disclosed herein.

[0379] A method for conferring resistance to a cell (suitably, a bacterial cell) against a bacteriophage comprising the steps of: (i) identifying a pseudo CRISPR spacer in a bacteriophage comprising a target nucleic acid or a transcription product thereof against which resistance is to be modulated; and (ii) modifying the sequence of the CRISPR spacer of the cell such that the CRISPR spacer of the cell has 100% homology or identity to the pseudo CRISPR spacer of the bacteriophage comprising the target nucleic acid is disclosed herein.

[0380] A method for modulating the lysotype of a bacterial cell comprising the steps of: comprising the steps of: (i) identifying a pseudo CRISPR spacer in a bacteriophage comprising a target nucleic acid or a transcription product thereof against which resistance is to be modulated; and (ii) modifying the sequence of the CRISPR spacer of the cell such that the CRISPR spacer of the cell has homology to the pseudo CRISPR spacer of the bacteriophage comprising the target nucleic acid is disclosed herein.

[0381] A method for modulating the lysotype of a bacterial cell comprising the steps of: (i) identifying a pseudo CRISPR spacer in a bacteriophage comprising a target nucleic acid or a transcription product thereof against which resistance is to be modulated; and (ii) modifying the sequence of the CRISPR spacer of the cell such that the CRISPR spacer of the cell has 100% homology or identity to the pseudo-CRISPR spacer of the bacteriophage comprising the target nucleic acid is disclosed herein.

[0382] Suitably, the CRISPR spacer of the bacterial cell will have 100% homology or identity to a sequence - such as a pseudo CRISPR spacer - in the bacteriophage comprising the target nucleic acid.

[0383] Suitably, the CRISPR spacer of the bacterial cell will form a component part of a CRISPR locus comprising a functional CRISPR repeat-cas combination as described herein.

[0384] Suitably, the target nucleic acid or a transcription product thereof in the bacteriophage is a highly conserved nucleic acid sequence.

[0385] Suitably, the target nucleic acid or transcription product thereof in the bacteriophage is

a gene coding for a host specificity protein.

[0386] Suitably, the target nucleic acid or transcription product thereof in the bacteriophage encodes an enzyme that is essential for survival, replication or growth of the bacteriophage.

[0387] Suitably, the target nucleic acid or transcription product thereof in the bacteriophage encodes a helicase, a primase, a head or tail structural protein, or a protein with a conserved domain (eg. holing, lysine, and others).

[0388] Advantageously, bacterial cells may be prepared according to the present invention that have a "reduced susceptibility to bacteriophage multiplication or infection". As used herein, this term refers to the bacterium as having a low or no susceptibility to bacteriophage multiplication or infection when compared to the wild-type bacterium when cultured, in for example, a dairy medium.

[0389] In one embodiment, the term "low susceptibility to bacteriophage multiplication" refers to the level of bacteriophage multiplication in a bacterium being below a level, which would cause a deleterious effect to a culture in a given period of time. Such deleterious effects on a culture include, but are not limited to, no coagulation of milk during production of fermented milk products (such as yoghurt or cheese), inadequate or slow lowering of the pH during production of fermented milk products (such as yoghurt or cheese), slow ripening of cheese and deterioration of a food's texture to the point where it is unappetising or unsuitable for human consumption.

[0390] For an equivalent set of culture conditions the susceptibility towards a bacteriophage of a bacterium of the present invention is, in comparison to the wild-type bacterium, 100 times lower (efficiency of plaquing [EOP] = 10^{-2}), preferably 1000 times lower (EOP = 10^{-3}), preferably 10 000 times lower (EOP = 10^{-4}), more preferably 100 000 times lower (EOP = 10^{-5}). Preferably, the level of bacteriophage multiplication in a culture is measured after about 14 hours incubation of the culture, more preferably after about 12 hours, more preferably after about 7 hours, more preferably after about 6 hours, more preferably after about 5 hours and more preferably after about 4 hours.

[0391] A method for conferring sensitivity to a cell (preferably, a bacterial cell) against a bacteriophage comprising the steps of: (a) providing a pseudo CRISPR spacer from at least one bacteriophage; (b) identifying one or more functional CRISPR repeat-cas combinations in a cell that is substantially resistant to the bacteriophage; and (c) engineering the one or more CRISPR loci in the substantially sensitive cell such that they comprise one or more pseudo CRISPR spacers or one or more CRISPR spacer(s) which is/are complementary or homologous to the one or more pseudo CRISPR spacer(s) that have a reduced degree of homology as compared to the one or more CRISPR loci in the substantially resistant cell is disclosed herein.

[0392] A method for modulating (eg. reducing) the lysotype of a cell (preferably a bacterial

cell), comprising one or more *cas* genes or proteins and one or more, preferably, two or more CRISPR repeats comprising the steps of: (i) identifying a pseudo CRISPR spacer in a bacteriophage against which resistance is to be modulated; and (ii) modifying the sequence of the CRISPR spacer of the cell such that the CRISPR spacer of the cell has a reduced degree of homology to the pseudo CRISPR spacer of the bacteriophage comprising the target nucleic acid is disclosed herein.

[0393] A method for modulating (eg. reducing or decreasing) the resistance of a bacterial cell comprising one or more *cas* genes or proteins and one or more, preferably, two or more CRISPR repeats against a bacteriophage comprising the steps of: (i) identifying one or more pseudo CRISPR spacers in a bacteriophage against which resistance is to be modulated; (ii) identifying a CRISPR spacer in the bacterial cell in which resistance is to be modulated that is homologous to the pseudo CRISPR spacer(s); and (iii) modifying the sequence of the CRISPR spacer in the bacterial cell in which resistance is to be modulated such that the CRISPR spacer has a lower degree of homology to the pseudo CRISPR spacer(s) of the bacteriophage against which resistance is to be modulated is disclosed herein.

[0394] Suitably, the CRISPR spacer of the cell will have a reduced degree of homology - such as a 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90 or 95% reduction in homology as compared to the pseudo CRISPR spacer(s) of the bacteriophage against which resistance is to be modulated.

[0395] Bacterial cells may therefore be prepared according to the present description that have an "increased susceptibility to bacteriophage multiplication". As used herein, this term refers to the bacterium as having an increased or high susceptibility to bacteriophage multiplication when compared to the wild-type bacterium when cultured, in for example, a dairy medium.

[0396] In one embodiment, the term "high susceptibility to bacteriophage multiplication" refers to the level of bacteriophage multiplication in a bacterium being above a level, which would cause a deleterious effect to a culture in a given period of time. Such deleterious effects on a culture include, but are not limited to, no coagulation of milk during production of fermented milk products (such as yoghurt or cheese), inadequate or slow lowering of the pH during production of fermented milk products (such as yoghurt or cheese), slow ripening of cheese and deterioration of a food's texture to the point where it is unappetising or unsuitable for human consumption. For an equivalent set of culture conditions the susceptibility towards a bacteriophage of a bacterium of the present invention is, in comparison to the wild-type bacterium, 100 times higher, 1000 times higher, 10 000 times higher, or 100 000 times higher ($EOP=10^{-5}$). The level of bacteriophage multiplication in a culture is measured after about 14 hours incubation of the culture, more preferably after about 12 hours, more preferably after about 7 hours, more preferably after about 6 hours, more preferably after about 5 hours and in a highly preferred embodiment after about 4 hours.

[0397] A CRISPR spacer is flanked by two CRISPR repeats. In other words, a CRISPR spacer

has at least one CRISPR repeat on each side.

BACTERIA

[0398] In a further embodiment, the target nucleic sequence or a transcription product thereof may be or may be derivable (preferably, derived) from one or more bacteria. Accordingly, resistance of a cell, eg. a bacterial cell, against bacteria or a component thereof may be modulated.

[0399] The target nucleotide sequence may be or may be derived from a gene that is or is associated with resistance to plasmid transfer in bacteria. According to this embodiment of the present invention, one or more CRISPR spacers in the cell are modified such that the CRISPR spacer of the cell has homology to the CRISPR spacer and/or pseudo CRISPR spacer contained in the plasmid DNA of the bacterial cell so as to provide resistance against the particular plasmid(s), thus preventing transfer of foreign DNA into the cell. Specifically, particular regions within the plasmid DNA can be targeted as to provide immunity against plasmid DNA, such as sequences within the plasmids origin of replication or sequences within genes coding for replication proteins.

[0400] Described herein the method comprises the steps of: (i) identifying a CRISPR spacer and/or pseudo CRISPR spacer derivable (preferably, derived) from the plasmid DNA of a bacterial cell against which resistance is to be modulated; and (ii) modifying the sequence of a CRISPR spacer in the cell in which resistance is to be modulated such that the CRISPR spacer of the cell has homology to the CRISPR spacer and/or pseudo CRISPR spacer contained in the plasmid DNA of the bacterial cell.

[0401] A further method for conferring resistance to a cell against plasmid transfer comprising the steps of: (a) identifying a CRISPR spacer and/or pseudo CRISPR spacer derivable (preferably, derived) from plasmid DNA; (b) identifying one or more functional CRISPR repeat-cas gene combinations in a cell that is substantially sensitive to the plasmid; and (c) engineering the one or more CRISPR loci in the substantially sensitive cell such that they comprise one or more CRISPR spacers and/or pseudo CRISPR spacers from the plasmid to render the cell resistant is disclosed herein.

[0402] The target nucleotide sequence may be or may be derived from a gene that is or is associated with resistance to one or more mobile genetic elements. Particular CRISPR spacers and/or pseudo CRISPR spacers derivable (preferably, derived) from one or more mobile genetic elements can be added within a CRISPR locus of a cell so as to provide resistance against mobile genetic elements - such as transposable elements and insertion sequences, thus preventing transfer of foreign DNA and genetic drift. Specifically, particular regions within transposons and insertion sequences can be targeted so as to provide immunity against mobile genetic elements. For example, targets can include conjugative transposons (Tn916), class II transposons (Tn501), insertion sequences (IS26) or transposase genes.

[0403] Described herein the method comprises the steps of: (i) identifying a CRISPR spacer and/or pseudo CRISPR spacer derivable (preferably, derived) from one or more mobile genetic elements of a cell against which resistance is to be modulated; and (ii) modifying the sequence of a CRISPR spacer in a cell in which resistance is to be modulated such that the CRISPR spacer and/or pseudo CRISPR spacer of the cell has homology to the CRISPR spacer contained in the mobile genetic element(s) of the cell.

[0404] A further method for conferring resistance to a cell against one or more mobile genetic elements comprising the steps of: (a) identifying a CRISPR spacer and/or pseudo CRISPR spacer derivable (preferably, derived) from one or more mobile genetic elements; (b) identifying one or more functional CRISPR repeat-cas combinations in a cell that is substantially sensitive to the one or more mobile genetic elements; and (c) engineering the one or more CRISPR loci in the substantially sensitive cell such that they comprise or have homology to one or more CRISPR spacers and/or pseudo CRISPR spacers from the one or more mobile genetic elements to render the cell resistant is taught herein.

[0405] The target nucleotide sequence may be or may be derived from a gene that is or is associated with resistance to antibiotics. By "antibiotic" is understood a chemical composition or moiety which decreases the viability or which inhibits the growth or reproduction of microbes. Antibiotic resistance genes include, but are not limited to *bla_{tem}*, *bla_{rob}*, *bla_{shv}*, *aadB*, *aacC1*, *aacC2*, *aacC3*, *aacA4*, *mecA*, *vanA*, *vanH*, *vanX*, *satA*, *aacA-aphH*, *vat*, *vga*, *msrA* *sul*, and/or *int*. The antibiotic resistance genes include those that are or are derivable (preferably, derived) from bacterial species that include but are not limited to the genera *Escherichia*, *Klebsiella*, *Pseudomonas*, *Proteus*, *Streptococcus*, *Staphylococcus*, *Enterococcus*, *Haemophilus* and *Moraxella*. The antibiotic resistance genes also include those that are or are derivable (preferably, derived) from bacterial species that include but are not limited to *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Proteus mirabilis*, *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Enterococcus faecalis*, *Staphylococcus saprophyticus*, *Streptococcus pyogenes*, *Haemophilus influenzae*, and *Moraxella catarrhalis*.

[0406] Particular CRISPR spacers and/or pseudo CRISPR spacers derivable (preferably, derived) from antibiotic resistance encoding genes can be added within a CRISPR locus of a cell - such as a recipient cell - so as to prevent transfer of genes conferring resistance to antibiotics into the cell, thus reducing the risk of acquiring antibiotic resistance markers. By way of example, targets can also include *vanR*, (a gene conferring resistance to vancomycin), or *tetR*, a gene conferring resistance to tetracycline, or targeting beta-lactamase inhibitors.

[0407] Described herein the method comprises the steps of: (i) identifying one or more CRISPR spacers and/or pseudo CRISPR spacers derivable (preferably, derived) from a cell that comprises one or more antibiotic resistance genes or markers; and (ii) modifying the sequence of the CRISPR spacer in a cell that does not comprise or does not express the antibiotic resistance genes or markers such that the CRISPR spacer of the cell has homology

to the one or more CRISPR spacers and/or pseudo CRISPR spacers contained in the cell that comprises one or more antibiotic resistance genes or markers.

[0408] A method for modulating the acquisition of antibiotic resistance markers in a cell comprising the steps of: (a) identifying one or more CRISPR spacers and/or pseudo CRISPR spacers derivable (preferably, derived) from a cell that comprises one or more antibiotic resistance genes or markers; (b) identifying one or more CRISPR loci in a cell that does not comprise or does not express the antibiotic resistance genes or markers; and (c) modifying the sequence of the CRISPR spacer in the cell that does not comprise or does not express the antibiotic resistance genes or markers such that the CRISPR spacer and/or pseudo CRISPR spacers has homology to the CRISPR spacer contained in the cell resistant to the transfer of genes conferring resistance to one or more antibiotics is disclosed herein.

[0409] The target nucleotide sequence may be or may be derived from a gene that is or is associated with genes encoding virulence factors. Particular CRISPR spacers and/or pseudo CRISPR spacers derivable (preferably, derived) from genes encoding virulence factors can be added within a bacterium CRISPR locus to provide resistance against the transfer of genes conferring virulence into the bacterium. For example, factors commonly contributing to virulence in microbial pathogens can be targeted, such as toxins, internalins and hemolysins.

[0410] Described herein the method comprises the steps of: (i) identifying one or more CRISPR spacers and/or pseudo CRISPR spacers derivable (preferably, derived) from a cell that comprises one or more virulence factors; and (ii) modifying the sequence of the CRISPR spacer in a cell that does not comprise or does not express the virulence factor(s) or marker(s) such that the CRISPR spacer of the cell has homology to the one or more CRISPR spacers and/or pseudo CRISPR spacers contained in the cell that comprises one or more virulence factors.

[0411] A further method for conferring resistance to a cell against one or more virulence factor(s) or marker(s) comprising the steps of: (a) identifying a CRISPR spacer and/or pseudo CRISPR spacer derivable (preferably, derived) from one or more virulence factor(s) or marker(s); (b) identifying one or more functional CRISPR repeat-cas combinations in a cell that is substantially sensitive to the one or more virulence factor(s) or marker(s); and (c) engineering the one or more CRISPR loci in the substantially sensitive cell such that they comprise one or more CRISPR spacers and/or pseudo CRISPR spacers from the one or more virulence factor(s) or marker(s) to render the cell resistant is disclosed herein,

[0412] A CRISPR spacer is flanked by two CRISPR repeats. In other words, a CRISPR spacer has at least one CRISPR repeat on each side.

MODIFICATION

[0413] Nucleic acid sequences may be modified by genetically engineering nucleic acid

sequences.

[0414] All or part of a nucleic acid sequence may be modified.

[0415] All or part of one or more CRISPR spacers, *cas* genes or proteins, CRISPR repeats or CRISPR loci may be modified.

[0416] Recombinant CRISPR spacers, *cas* genes or proteins, CRISPR repeats or CRISPR loci may be modified.

[0417] Naturally occurring CRISPR spacers, *cas* genes or proteins, CRISPR repeats or CRISPR loci may be modified.

[0418] Naturally co-occurring *cas* genes or proteins and CRISPR repeats may be modified.

[0419] The genetic engineering may be mediated using various methods that are known in the art and will typically include well known methods - such as PCR amplification, cloning and site-directed mutagenesis. Mutations may be introduced using synthetic oligonucleotides. These oligonucleotides contain nucleotide sequences flanking the desired mutation sites. A suitable method is disclosed in Morinaga et al., (Biotechnology (1984) 2, p646-649). Another method of introducing mutations into enzyme-encoding nucleotide sequences is described in Nelson and Long (Analytical Biochemistry (1989), 180, p 147-151). A further method is described in Sarkar and Sommer (Biotechniques (1990), 8, p404-407 - "The megaprimer method of site directed mutagenesis"). Commercially available kits are also now widely available for performing site directed mutagenesis. Genetic engineering methods are described in detail in J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, Molecular Cloning: A Laboratory Manual, Second Edition, Books 1-3, Cold Spring Harbor Laboratory Press; Ausubel, F. M. et al. (1995 and periodic supplements; Current Protocols in Molecular Biology, ch. 9, 13, and 16, John Wiley & Sons, New York, N.Y.); B. Roe, J. Crabtree, and A. Kahn, 1996, DNA Isolation and Sequencing: Essential Techniques, John Wiley & Sons; M. J. Gait (Editor), 1984, Oligonucleotide Synthesis: A Practical Approach, Irl Press; and, D. M. J. Lilley and J. E. Dahlberg, 1992, Methods of Enzymology: DNA Structure Part A: Synthesis and Physical Analysis of DNA Methods in Enzymology, Academic Press.

[0420] The genetic engineering step may even include methods such as homologous recombination which may be particularly useful when, for example, CRISPR spacers are being inserted or deleted.

[0421] The genetic engineering step may even include the activation of one or more nucleic acid sequences - such as one CRISPR loci, CRISPR repeats, CRISPR spacers, *cas* genes or proteins, functional combinations of *cas* genes or proteins and CRISPR repeats or even combinations thereof.

[0422] Suitably, one or more CRISPR spacers or pseudo CRISPR spacers may be inserted

into at least one CRISPR locus.

[0423] In one embodiment, the modification does not interrupt one or more *cas* genes of the at least one CRISPR locus. In another embodiment, the one or more *cas* genes remain intact.

[0424] In one embodiment, the modification does not interrupt one or more CRISPR repeats of the at least one CRISPR locus. In one embodiment, the one or more CRISPR repeats remain intact.

[0425] Suitably, one or more CRISPR spacers or pseudo CRISPR spacers may be inserted into or within at least one CRISPR locus.

[0426] Suitably, one or more CRISPR spacers or pseudo CRISPR spacers may be inserted at the 5' end of at least one CRISPR locus.

[0427] In one embodiment, the modification comprises inserting at least one CRISPR spacer or pseudo CRISPR spacers into a cell - such as a recipient cell. In another embodiment, the modification comprises inserting one or more CRISPR spacers or pseudo CRISPR spacers into (eg. to modify or replace) one or more CRISPR spacers of a cell - such as a recipient cell.

[0428] In one embodiment, the modification comprises inserting at least one CRISPR spacer or pseudo CRISPR spacer from an organism - such as a donor organism - into the cell. In another embodiment, the modification comprises inserting one or more CRISPR spacers or pseudo CRISPR spacers from an organism - such as a donor organism - into (eg. to modify or replace) one or more CRISPR spacers or pseudo CRISPR spacers of a cell.

[0429] In one embodiment, one or more CRISPR spacers or pseudo CRISPR spacers - such as one or more CRISPR spacers or pseudo CRISPR spacers from an organism - such as a donor organism - are inserted into (eg. to modify or replace) one or more CRISPR spacers or pseudo CRISPR spacers of the cell.

[0430] In one embodiment, one or more CRISPR spacers or pseudo CRISPR spacers - such as one or more CRISPR spacers or pseudo CRISPR spacers from an organism- such as a donor organism - are inserted into (eg. to modify or replace) one or more, preferably, two or more CRISPR repeats of the cell. In this embodiment of the invention, it is preferred that at least one functional CRISPR repeat-cas combination remains intact in the cell.

[0431] In one embodiment, one or more CRISPR spacers or pseudo CRISPR spacers - such as one or more CRISPR spacers or pseudo CRISPR spacers from an organism - such as a donor organism - are inserted into (eg. to modify or replace) the same or different CRISPR spacers of the cell.

[0432] In one embodiment, one or more CRISPR spacers or pseudo CRISPR spacers - such as one or more CRISPR spacers or pseudo CRISPR spacers from an organism - such as a

donor organism - are inserted adjacent to (eg. to modify or replace) one or more CRISPR spacers or pseudo CRISPR spacers of the cell.

[0433] In the context of the present invention, the term "adjacent" means "next to" in its broadest sense and includes "directly adjacent". Thus, in one embodiment, one or more CRISPR spacers or pseudo CRISPR spacers from an organism may be inserted "directly adjacent" to one or more CRISPR spacers or pseudo CRISPR spacers of the cell. *ie.* the CRISPR spacer(s) or pseudo CRISPR spacer(s) is inserted such that there are no intervening nucleotides between the spacers.

[0434] In another embodiment, the CRISPR spacer(s) or pseudo CRISPR spacer(s) are inserted such that there are at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 10,000, 100,000 or even 1,000,000 or more intervening nucleotides between the spacers.

[0435] In another embodiment, the intervening nucleotide may be called a leader sequence. These terms are used interchangeably herein. The leader sequence can be of a different length in different bacteria. Suitably the leader sequence is at least 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400 or 500 or more nucleotides in length. Suitably the leader sequence is between the last *cas* gene (at the 3' end) and the first CRISPR repeat (at the 5' end) of the CRISPR locus.

[0436] In one embodiment the leader sequence may be between about 20-500 nucleotides in length.

[0437] In one embodiment, one or more CRISPR spacers or pseudo CRISPR spacers - such as one or more CRISPR spacers or pseudo CRISPR spacers from a donor organism - are inserted adjacent to one or more, preferably, two or more CRISPR repeats of the cell.

[0438] In another embodiment, one or more CRISPR spacers or pseudo CRISPR spacers - such as one or more CRISPR spacers or pseudo CRISPR spacers from a donor organism - are inserted adjacent to one or more *cas* genes of the cell.

[0439] In another embodiment, one or more CRISPR spacers or pseudo CRISPR spacers - such as one or more CRISPR spacers or pseudo CRISPR spacers from a donor organism - are inserted adjacent to the same or different spacers of the recipient cell.

[0440] In another embodiment, one or more CRISPR spacers or pseudo CRISPR spacers - such as one or more CRISPR spacers or pseudo CRISPR spacers from a donor organism - are each inserted adjacent to the same or different CRISPR repeats of the cell.

[0441] In another embodiment, one or more CRISPR spacers or pseudo CRISPR spacers - such as one or more CRISPR spacers or pseudo CRISPR spacers from a donor organism - are each inserted adjacent to the same or different *cas* genes of the recipient cell.

[0442] In another embodiment, two or more CRISPR spacers or pseudo CRISPR spacers - such as two or more CRISPR spacers or pseudo CRISPR spacers from a donor organism - are each inserted adjacent to the same or different CRISPR spacers or pseudo CRISPR spacers and/or CRISPR repeats and/or *cas* genes of the recipient cell.

[0443] In another embodiment, the sequence of the CRISPR spacer - such as one or more CRISPR spacers from a donor organism - of the recipient cell is modified such that the CRISPR spacer has homology to the CRISPR spacer of the donor organism.

[0444] In another embodiment, the sequence of the spacer of the cell is modified such that it has homology to the CRISPR spacer or pseudo CRISPR spacer of the organism.

[0445] In one embodiment, the CRISPR spacer has 100% homology to the CRISPR spacer of the donor organism.

[0446] The CRISPR spacer(s) or pseudo CRISPR spacers may comprise DNA or RNA of genomic, synthetic or recombinant origin.

[0447] The CRISPR spacer (s) or pseudo CRISPR spacers may be double-stranded or single-stranded whether representing the sense or antisense strand or combinations thereof.

[0448] The CRISPR spacer (s) or pseudo CRISPR spacers may be prepared by use of recombinant DNA techniques (e.g. recombinant DNA), as described herein.

[0449] The modification may comprise inserting one or more CRISPR spacers or pseudo CRISPR spacers from an organism - such as a donor organism - that is substantially resistant to a target nucleic acid or a transcription product thereof into one or more CRISPR loci of a substantially sensitive cell.

[0450] The modification may comprise inserting one or more CRISPR spacers or pseudo CRISPR spacers from an organism - such as a donor organism - that is substantially resistant to a target nucleic acid or a transcription product thereof into (eg. between) a functional combination of at least two CRISPR repeats and at least one *cas* gene in a substantially sensitive cell.

[0451] The modification may even comprise modifying (eg. mutating) the DNA of a cell (eg. a recipient cell) - such as plasmid DNA or genomic DNA- such that one or more *cas* genes are created in the DNA of the cell. By way of further example, the *cas* genes may be cloned into a construct, a plasmid or a vector and the like which is then transformed into the cell, using methods such as those described herein.

[0452] The modification may even comprise modifying (eg. mutating) the DNA of a cell (eg. a recipient cell) - such as plasmid DNA or genomic DNA- such that one or more, preferably, two

or more CRISPR repeats are created in the DNA of the cell. By way of further example, the CRISPR repeats may be cloned into a construct, a plasmid or a vector and the like which is then transformed into the cell, using methods such as those described herein.

[0453] The modification may even comprise modifying (eg. mutating) the DNA of a cell (eg. a recipient cell) - such as plasmid DNA or genomic DNA- such that one or more cas-CRISPR repeat functional combinations are created in the DNA of the cell. By way of further example, the cas-CRISPR repeat functional combinations may be cloned into a construct, a plasmid or a vector and the like which is then transformed into the cell, using methods such as those described herein.

[0454] The modification may even comprise modifying (eg. mutating) the DNA of a cell (eg. a recipient cell) - such as plasmid DNA or genomic DNA- such that one or more CRISPR spacers are created in the DNA of the cell. By way of further example, the CRISPR spacers may be cloned into a construct, a plasmid or a vector and the like which is then transformed into the cell, using methods such as those described herein.

[0455] In one embodiment, a CRISPR spacer is flanked by two CRISPR repeats. In other words, a CRISPR spacer has at least one CRISPR repeat on each side.

[0456] Suitably, the modification comprises inserting one or more CRISPR spacers (eg. heterologous CRISPR spacers) in the vicinity of (eg. adjacent to, suitably, directly adjacent to) one or more *cas* genes and/or the leader sequence. Suitably, according to this embodiment of the present invention, the organisation of the naturally occurring CRISPR locus is maintained following insertion of the one or more CRISPR spacers.

CLUSTER

[0457] It has also been surprisingly found that it is not possible to merely exchange CRISPR repeat-*cas* combinations between any cells (eg. any strains, species or genera of cells) since it is believed that this will not necessarily result in functional CRISPR repeat-*cas* combinations.

[0458] Rather, for the CRISPR repeat-*cas* combination(s) to be functional they should be compatible. Accordingly, it is believed that it is not possible to switch *cas* genes or CRISPR repeats between different CRISPR loci unless they are from the same cluster.

[0459] Even more surprising is that the clusters do not follow the "organism" phylogeny. Specifically, within one organism, there may be more than one CRISPR. These CRISPR(s) can belong to different clusters, even though they are present in the same organism. As a result, it is believed that a functional CRISPR repeat-*cas* combination requires that the combination be switched within a cluster as opposed to within an organism.

[0460] For the avoidance of doubt, the term "cluster" as used herein does not refer to a cluster

of genes located at the same locus (typically forming an operon) but to the output from sequence comparison analysis - such as multiple sequence comparison analysis and/or multiple sequence alignments and/or dot plot analysis. Accordingly, cluster analysis of CRISPR loci may be performed using various methods that are known in the art - such as dot-plot analysis as taught herein below for example or multiple alignment followed by dendrogram calculation.

[0461] Advantageously, the use of naturally co-occurring CRISPR repeat-cas combination(s) provides for the interchange of the combination both within and between a given species, thereby making it possible to engineer the resistance of one strain using the combination from a different strain.

[0462] The cluster may be a class, a family or a group of sequences.

DETERMINING RESISTANCE

[0463] A method for determining the resistance profile of a cell against a target nucleic acid. As used herein, the term "resistance profile" means one or more entities against which the cell is sensitive or resistant. Accordingly, the resistance profile of a cell may be that the cell is resistant to a first bacteriophage, sensitive to a second bacteriophage, resistant to a first mobile genetic element and sensitive to a first antibiotic resistance gene *etc* is disclosed herein.

[0464] One or more *cas* genes or proteins, and/or one or more, preferably, two or more CRISPR repeats and/or one or more CRISPR spacers *etc.* within a cell may be detected or sequenced so as to predict/determine the likely resistance profile of a particular cell.

[0465] Suitably, one or more CRISPR spacers within a cell are detected or sequenced so as to predict/determine the likely resistance profile of a particular cell.

[0466] Suitable detection methods may include PCR, DNA-DNA hybridization (or DNA-RNA hybridization *ie.* using DNA or RNA probes that could be synthetic, labelled oligonucleotides, for example). DNA microarrays may also be used.

[0467] One or more *cas* -CRISPR repeat functional combinations and/or one or more CRISPR spacers within a cell may be detected or sequenced so as to predict/determine the likely resistance profile of a particular cell. By way of example, it is possible to predict/determine the likely resistance profile of a particular bacterial cell to one or more bacteriophage which can be used as a lysotype predictor for microbial selection.

[0468] One or more *Cas* genes and/or one or more CRISPR repeats may be sequenced in addition to one or more CRISPR spacers in order to verify the compatibility of the *Cas* gene-CRISPR repeat combination or even to identify new pairs of compatible *cas*/repeats.

RECIPIENT CELL

[0469] As used herein, the term "recipient cell" refers to any bacterial cell in which resistance against a target nucleic acid or a transcription product thereof is modulated or is to be modulated.

[0470] In one embodiment, the recipient cell refers to any bacterial cell comprising the recombinant nucleic acid according to the present description.

[0471] The recipient cell may comprise one or more, preferably, two or more CRISPR repeats and one or more *cas* genes or proteins. Suitably, the CRISPR repeats and the *cas* genes or proteins form a functional combination in the recipient cell, as described herein.

[0472] The recipient cell may comprise one or more modified CRISPR repeats and/or one or more modified *cas* genes or proteins. Suitably, the modified CRISPR repeats and/or the modified *cas* genes or proteins form a functional combination in the recipient cell, as described herein.

[0473] The recipient cell may comprise one or more genetically engineered CRISPR repeats and/or one or more genetically engineered *cas* genes or proteins. Suitably, the genetically engineered CRISPR repeats and/or the genetically engineered *cas* genes or proteins form a functional combination in the recipient cell, as described herein.

[0474] The recipient cell may comprise one or more recombinant CRISPR repeats and/or one or more recombinant *cas* genes or proteins. Suitably, the recombinant CRISPR repeats and/or the recombinant *cas* genes or proteins form a functional combination in the recipient cell, as described herein.

[0475] The recipient cell may comprise one or more naturally occurring CRISPR repeats and one or more naturally occurring *cas* genes or proteins. Suitably, the CRISPR repeats(s) and the *cas* gene(s) or proteins form a functional combination.

[0476] By "naturally occurring" we mean occurring naturally in nature.

[0477] The recipient cell may even comprise combinations of one or more modified, genetically engineered, recombinant or naturally occurring CRISPR repeats and one or more modified, genetically engineered, recombinant or naturally occurring *cas* genes or proteins. Suitably, the one or more modified, genetically engineered, recombinant or naturally occurring CRISPR spacer(s) or the one or more modified, genetically engineered, recombinant or naturally occurring *cas* gene(s) or proteins form a functional combination.

[0478] The recipient cell is a bacterial cell. Suitable bacterial cells are described herein.

[0479] The bacterial cell may be selected from a lactic acid bacteria species, a *Bifidobacterium* species, a *Brevibacterium* species, a *Propionibacterium* species, a *Lactococcus* species, a *Streptococcus* species, a *Lactobacillus* species including the *Enterococcus* species, *Pediococcus* species, a *Leuconostoc* species and *Oenococcus* species.

[0480] Suitable species include, but are not limited to *Lactococcus lactis*, including *Lactococcus lactis* subsp. *lactis* and *Lactococcus lactis* subsp. *cremoris*, *Lactococcus lactis* subsp. *cremoris*, *Leuconostoc* sp., *Lactococcus lactis* subsp. *lactis* biovar, *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Lactobacillus helveticus*, *Bifidobacterium lactis*, *Lactobacillus acidophilus*, *Lactobacillus casei*.

[0481] The cell in which resistance is to be modulated may be a bacterial cell used for the fermentation of meat (including beef, pork, and poultry) including, but not limited to, lactic acid bacteria, *Pediococcus cerevisiae*, *Lactobacillus plantarum*, *Lactobacillus brevis*, *Micrococcus* species, *Lactobacillus sakei*, *Lactobacillus curvatus*, *Pediococcus pentosaceus*, *Staphylococcus xylosus* and *Staphylococcus vitulinus* and mixtures thereof (Food Biotechnology, 538-39 (D. Knorr Ed. 1987); C. Pederson, Microbiology of Fermented Foods, 210-34 (2d ed. 1979); US 2,225,783).

[0482] The cell in which resistance is to be modulated may be a bacterial cell used for the fermentation of vegetables (e.g., carrots, cucumbers, tomatoes, peppers, and cabbage) including, but not limited to, *Lactobacillus plantatum*, *Lactobacillus brevis*, *Leuconostoc mesenteroides*, *Pediococcus pentosaceus*, and mixtures thereof (Food Biotechnology, 540 (D. Knorr Ed. 1987); C. Pederson, Microbiology of Fermented Foods, 153-209 (2d ed. 1979); US 3,024,116; US 3,403,032; US 3,932,674; and US 3,897,307).

[0483] The cell in which resistance is to be modulated may be a bacterial cell used for the fermentation of dough formed from cereals (e.g., wheat, rye, rice, oats, barley, and corn).

[0484] The cell in which resistance is to be modulated may be a bacterial cell used for the production of wine. Typically, this is achieved by the fermentation of fruit juice, typically grape juice.

[0485] The cell in which resistance is to be modulated may be a bacterial cell used for the fermentation of milk to produce cheese - such as *Lactobacillus delbrueckii* subsp. *bulgaricus*, *Lactobacillus helveticus*, *Streptococcus thermophilus*, *Lactococcus lactis* subsp. *lactis*, *Lactococcus lactis* subsp. *cremoris*, *Lactococcus lactis* subsp. *lactis* biovar *diacetylactis*, *Bifidobacteria* and *Enterococci* etc and mixtures thereof (Food Biotechnology, 530 (D. Knorr Ed. 1987); C. Pederson, Microbiology of Fermented Foods, 135-51 (2d ed. 1979)).

[0486] The cell in which resistance is to be modulated may be a bacterial cell used for the fermentation of milk to produce cheese - such as *Lactobacillus bulgaricus*, *Lactobacillus helveticus*, *Streptococcus thermophilus*, *Lactococcus lactis* subsp. *lactis*, *Lactococcus lactis*

subsp. cremoris, *Lactococcus lactis* subsp. *lactis* biovar, *Lactococci*, *Bifidobacteria* and *Enterococci* etc and mixtures thereof (Food Biotechnology, 530 (D. Knorr Ed. 1987); C. Pederson, Microbiology of Fermented Foods, 135-51 (2d ed. 1979)). The cell in which resistance is to be modulated may be a bacterial cell used for the fermentation of egg - such as *Pediococcus pentosaceus*, *Lactobacillus plantarum*, and mixtures thereof (Food Biotechnology, 538-39 (D. Knorr Ed. 1987)).

[0487] The cell in which resistance is to be modulated may be a bacterium that naturally comprises one or more CRISPR loci. CRISPR loci have been identified in more than 40 prokaryotes (Jansen *et al.* 2002b; Mojica *et al.*, 2005; Haft *et al.*, 2005) including, but not limited to *Aeropyrum*, *Pyrobaculum*, *Sulfolobus*, *Archaeoglobus*, *Halocarcula*, *Methanobacterium*, *Methanococcus*, *Methanosarcina*, *Methanopyrus*, *Pyrococcus*, *Picrophilus*, *Thermoplasma*, *Corynebacterium*, *Mycobacterium*, *Streptomyces*, *Aquifex*, *Porphyromonas*, *Chlorobium*, *Thermus*, *Bacillus*, *Listeria*, *Staphylococcus*, *Clostridium*, *Thermoanaerobacter*, *Mycoplasma*, *Fusobacterium*, *Azarcus*, *Chromobacterium*, *Neisseria*, *Nitrosomonas*, *Desulfovibrio*, *Geobacter*, *Myxococcus*, *Campylobacter*, *Wolinella*, *Acinetobacter*, *Erwinia*, *Escherichia*, *Legionella*, *Methylococcus*, *Pasteurella*, *Photobacterium*, *Salmonella*, *Xanthomonas*, *Yersinia*, *Treponema* and *Thermotoga*.

[0488] The cell in which resistance is to be modulated may be a bacterium for use in cosmetic or pharmaceutical compositions. Such compositions may comprise a microbial culture and/or labelled bacterium and/or a cell culture according to the present invention. Thus the microbial culture and/or labelled bacterium and/or a cell culture according to the present invention may be compatible in cosmetics or in pharmacy or in therapy.

DONOR ORGANISM

[0489] In one embodiment, the term "donor organism" refers to an organism or cell from which the CRISPR repeat and/or *cas* gene and/or combination(s) thereof and/or CRISPR spacers are derivable (preferably, derived). These can be the same or different.

[0490] In one embodiment, the term "donor organism" refers to an organism or cell from which the one or more, preferably, two or more CRISPR repeats and/or one or more *cas* gene and/or combination(s) thereof and/or CRISPR spacers are derivable (preferably, derived). These can be the same or different.

[0491] In one embodiment, the CRISPR spacer or pseudo CRISPR spacer is synthetically derived.

[0492] In one embodiment, the donor organism or cell comprises one or more CRISPR spacers, which confers the specific of immunity against a target nucleic acid or transcription product thereof.

[0493] In one embodiment, the donor organism or cell from which the *cas* gene and/or CRISPR repeat and/or combination thereof is derivable (preferably derived) is also the recipient cell/organism for the recombinant CRISPR locus. These can be the same or different.

[0494] In one embodiment, the donor organism or cell from which the CRISPR spacer is derivable (preferably derived) is also the recipient cell/organism for the recombinant CRISPR locus. These can be the same or different.

[0495] When it is the case that the donor organism is a bacterial cell then the donor organism will typically comprise a CRISPR spacer which confers the specific immunity against the target nucleic acid or transcription product thereof.

[0496] The organism may be a bacterial cell or a bacteriophage.

[0497] Suitably, the organism is a bacteriophage.

HOST CELLS

[0498] As used herein, the term "host cell" refers to any cell that comprises the combination, the construct or the vector and the like according to the present invention.

[0499] Host cells may be transformed or transfected with a nucleotide sequence contained in a vector e.g. a cloning vector. Said nucleotide sequence may be carried in a vector for the replication and/or expression of the nucleotide sequence. The cells will be chosen to be compatible with the said vector and may, for example, be prokaryotic (for example bacterial) cells.

[0500] Described herein are host cells comprising the combination, construct or the vector of the present description. The construct or the vector may comprise a nucleotide sequence for replication and expression of the sequence. The cells will be chosen to be compatible with the vector and may, for example, be prokaryotic (for example bacterial) cells.

CONSTRUCT

[0501] A construct comprising one or more of the nucleic acid sequences described herein is disclosed herein.

[0502] The term "construct" - which is synonymous with terms such as "conjugate", "cassette" and "hybrid" - includes a nucleotide sequence directly or indirectly attached to another sequence - such as a regulatory sequence (eg. a promoter). By way of example, the present invention covers a construct comprising a nucleotide sequence operably linked to such a

regulatory sequence. The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A regulatory sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under condition compatible with the control sequences.

[0503] The term "regulatory sequences" includes promoters and enhancers and other expression regulation signals.

[0504] The term "promoter" is used in the normal sense of the art, e.g. an RNA polymerase binding site.

[0505] The construct may even contain or express a marker, which allows for the selection of the nucleotide sequence construct in, for example, a bacterium. Various markers exist which may be used, for example those markers that provide for antibiotic resistance - e.g. resistance to bacterial antibiotics - such as Erythromycin, Ampicillin, Streptomycin and Tetracycline.

VECTOR

[0506] The construct may be or may be included in a vector (eg. a plasmid).

[0507] A vector comprising one or more of the constructs or sequences described herein is disclosed herein.

[0508] The term "vector" includes expression vectors and transformation vectors and shuttle vectors.

[0509] The term "transformation vector" means a construct capable of being transferred from one entity to another entity - which may be of the species or may be of a different species. If the construct is capable of being transferred from one species to another then the transformation vector is sometimes called a "shuttle vector".

[0510] The vectors may be transformed into a suitable cell (eg. a host cell) as described below.

[0511] The vectors may be for example, plasmid or phage vectors provided with an origin of replication, optionally a promoter for the, expression of the said polynucleotide and optionally a regulator of the promoter.

[0512] The vectors may contain one or more selectable marker nucleotide sequences. The most suitable selection systems for industrial micro-organisms are those formed by the group of selection markers which do not require a mutation in the host organism.

[0513] The vectors may be used *in vitro*, for example for the production of RNA or used to

transfect or transform a host cell.

[0514] Thus, polynucleotides may be incorporated into a recombinant vector (typically a replicable vector), for example, a cloning or expression vector. The vector may be used to replicate the nucleic acid in a compatible host cell.

TRANSFECTION

[0515] Introduction of a nucleic acid (eg. a construct or vector) into a cell can be effected by various methods. For example, calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction or infection may be used. Such methods are described in many standard laboratory manuals - such as Sambrook et al., Molecular Cloning: A Laboratory Manual, 2d ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

[0516] Cells containing the nucleic acid (eg. a construct or vector) may be selected by using, for example, Erythromycin for cells transfected with a nucleic acid (eg. a construct or vector) carrying a resistance selectable marker.

TRANSFORMATION

[0517] Teachings on the transformation of cells are well documented in the art, for example see Sambrook et al (Molecular Cloning: A Laboratory Manual, 2nd edition, 1989, Cold Spring Harbor Laboratory Press) and Ausubel et al., Current Protocols in Molecular Biology (1995), John Wiley & Sons, Inc.

[0518] A cell may be transformed with a nucleic acid (eg. a construct or vector). Cells transformed with the nucleotide sequence may be cultured under conditions suitable for the replication or expression of the nucleotide sequence.

INTRODUCING

[0519] In the context of introducing a nucleic acid into a cell, in one embodiment it is preferred that the term "introducing" means one or more of transforming, transfecting, conjugating or transducing.

STARTER CULTURES

[0520] Starter cultures are used extensively in the food industry in the manufacture of

fermented products including milk products - such as yoghurt and cheese, meat products, bakery products, wine and vegetable products.

[0521] Starter cultures used in the manufacture of many fermented milk, cheese and butter products include cultures of bacteria, generally classified as lactic acid bacteria. Such bacterial starter cultures impart specific features to various dairy products by performing a number of functions.

[0522] Commercial non-concentrated cultures of bacteria are referred to in industry as 'mother cultures', and are propagated at the production site, for example a dairy, before being added to an edible starting material, such as milk, for fermentation. The starter culture propagated at the production site for inoculation into an edible starting material is referred to as the 'bulk starter'.

[0523] Suitable starter cultures for use in the present invention may include any organism which is of use in the food, cosmetic or pharmaceutical industry.

[0524] For example, the starter culture may be suitable for use in the dairy industry. When used in the dairy industry the starter culture may be selected from a lactic acid bacteria species, a *Bifidobacterium* species, a *Brevibacterium* species, a *Propionibacterium* species. Suitable starter cultures of the lactic acid bacteria group include commonly used strains of a *Lactococcus* species, a *Streptococcus* species, a *Lactobacillus* species including the *Lactobacillus acidophilus*, *Enterococcus* species, *Pediococcus* species, a *Leuconostoc* species and *Oenococcus* species.

[0525] Cultures of lactic acid bacteria are commonly used in the manufacture of fermented milk products - such as buttermilk, yoghurt or sour cream, and in the manufacture of butter and cheese, for example Brie or Harvati. *Lactococcus* species include the widely used *Lactococcus lactis*, including *Lactococcus lactis* subsp. *lactis* and *Lactococcus lactis* subsp. *cremoris*.

[0526] Other lactic acid bacteria species include *Leuconostoc* sp., *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Lactobacillus helveticus*. In addition, probiotic strains - such as *Lactococcus* species - include the widely used *Lactococcus lactis*, including *Lactococcus lactis* subsp. *lactis* and *Lactococcus lactis* subsp. *cremoris*. Mesophilic cultures of lactic acid bacteria commonly used in the manufacture of fermented milk products such as buttermilk, yoghurt or sour cream, and in the manufacture of butter and cheese, for example Brie or Harvati. Other *Lactococcus* species include *Lactococcus lactis* subsp. *cremoris*, *Lactococcus lactis*, *Leuconostoc* sp., *Lactococcus lactis* subsp. *lactis* biovar, *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Lactobacillus helveticus*. In addition, probiotic strains such as *Bifidobacterium lactis*, *Lactobacillus acidophilus*, *Lactobacillus casei* may be added during said manufacturing to enhance flavour or to promote health.

[0527] Cultures of lactic acid bacteria commonly used in the manufacture of cheddar and Monterey Jack cheeses include *Streptococcus thermophilus*, *Lactococcus lactis* subsp. *lactis*

and *Lactococcus lactis* subsp. *cremoris* or combinations thereof.

[0528] Thermophilic cultures of lactic acid bacteria commonly used in the manufacture of Italian cheeses such as Pasta filata or parmesan, include *Streptococcus thermophilus* and *Lactobacillus delbrueckii* subsp. *bulgaricus*. Other *Lactobacillus* species - such as *Lactobacillus helveticus* - may be added during manufacturing to obtain a desired flavour.

[0529] Advantageously, the starter culture organism may comprise or consist of a genetically modified strain (prepared according to the methods desired herein) of one of the above lactic acid bacteria strains or any other starter culture strain.

[0530] The selection of organisms for the starter culture of the invention will depend on the particular type of products to be prepared and treated. Thus, for example, for cheese and butter manufacturing, mesophilic cultures of *Lactococcus* species, *Leuconostoc* species and *Lactobacillus* species are widely used, whereas for yoghurt and other fermented milk products, thermophilic strains of *Streptococcus* species and of *Lactobacillus* species are typically used.

[0531] The starter culture may even be a dried starter culture.

[0532] The starter culture may be a concentrated starter culture.

[0533] The starter culture may be a concentrated starter culture used in direct inoculation.

[0534] The starter culture may be a frozen starter culture.

[0535] The starter culture may consist of one bacterial strain, ie., a pure culture. In this case, substantially all, or at least a significant portion of the bacterial starter culture would generally comprise the same bacterium.

[0536] In the alternative, the starter culture may comprise several bacterial strains, ie., a defined mixed culture.

LACTIC ACID BACTERIA

[0537] Particularly suitable starter cultures, in particular dried starter cultures, for use in the present invention description comprise lactic acid bacteria.

[0538] As used herein the term "lactic acid bacteria" refers to Gram positive, microaerophilic or anaerobic bacteria which ferment sugar with the production of acids including lactic acid as the predominantly produced acid, acetic acid, formic acid and propionic acid. The industrially most useful lactic acid bacteria are found among *Lactococcus* species, such as *Lactococcus lactis*, *Lactobacillus* species, *Bifidobacterium* species, *Streptococcus* species, *Leuconostoc* species, *Pediococcus* species and *Propionibacterium* species.

[0539] The starter cultures of the present description may comprise one or more lactic acid bacteria species such as, *Lactococcus lactis*, *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Streptococcus thermophilus* or combinations thereof.

[0540] Lactic acid bacteria starter cultures are commonly used in the food industry as mixed strain cultures comprising one or more species. For a number of mixed strain cultures, such as yoghurt starter cultures comprising strains of *Lactobacillus delbrueckii* subsp *bulgaricus* and *Streptococcus thermophilus*, a symbiotic relationship exists between the species wherein the production of lactic acid is greater compared to cultures of single strain lactic acid bacteria (Rajagopal et al. J.Dairy Sci., 73, p.894-899, 1990).

PREPARING STARTER CULTURES

[0541] Starter cultures may be prepared by techniques well known in the art such as those disclosed in US 4,621,058. By way of example, starter cultures may be prepared by the introduction of an inoculum, for example a bacterium, to a growth medium to produce an inoculated medium and ripening the inoculated medium to produce a starter culture.

PREPARING DRIED STARTER CULTURES

[0542] Dried starter cultures may be prepared by techniques well known in the art, such as those discussed in US 4, 423, 079 and US 4,140,800.

[0543] Dried starter cultures for use in the present description may be in the form of solid preparations. Examples of solid preparations include, but are not limited to tablets, pellets, capsules, dusts, granules and powders which may be wettable, spraydried, freeze-dried or lyophilised.

[0544] The dried starter cultures for use in the present description may be in either a deep frozen pellet form or freeze-dried powder form. Dried starter cultures in a deep frozen pellet or freeze-dried powder form may be prepared according to the methods known in the art.

[0545] The starter cultures for use in the present description may be in the form of concentrates which comprise a substantially high concentration of one or more bacteria. Suitably the concentrates may be diluted with water or resuspended in water or other suitable diluents, for example, an appropriate growth medium or mineral or vegetable oils, for use in the present invention. The dried starter cultures of the present description in the form of concentrates may be prepared according to the methods known in the art, for example by centrifugation, filtration or a combination of such techniques.

PRODUCT

[0546] Suitable products for use in the present description include, but are not limited to, a foodstuffs, cosmetic products or pharmaceutical products.

[0547] Any product, which is prepared from, or comprises, a culture is contemplated in accordance with the present description. These include, but are not limited to, fruits, legumes, fodder crops and vegetables including derived products, grain and grain-derived products, dairy foods and dairy food-derived products, meat, poultry, seafood, cosmetic and pharmaceutical products.

[0548] The term "food" is used in a broad sense and includes feeds, foodstuffs, food ingredients, food supplements, and functional foods.

[0549] As used herein the term "food ingredient" includes a formulation, which is or can be added to foods and includes formulations which can be used at low levels in a wide variety of products that require, for example, acidifying or emulsifying.

[0550] As used herein, the term "functional food" means a food which is capable of providing not only a nutritional effect and/or a taste satisfaction, but is also capable of delivering a further beneficial effect to consumer. Although there is no legal definition of a functional food, most of the parties with an interest in this area agree that there are foods marketed as having specific health effects.

[0551] The term "food" covers food for humans as well as food for animals (i.e. a feed). In a preferred aspect, the food is for human consumption.

[0552] The cells described herein may be - or may be added to - a food ingredient, a food supplement, or a functional food.

[0553] The food may be in the form of a solution or as a solid - depending on the use and/or the mode of application and/or the mode of administration.

[0554] The cells described herein can be used in the preparation of food products such as one or more of: confectionery products, dairy products, meat products, poultry products, fish products and bakery products.

[0555] By way of example, the bacterium can be used as ingredients to soft drinks, a fruit juice or a beverage comprising whey protein, health teas, cocoa drinks, milk drinks and lactic acid bacteria drinks, yoghurt, drinking yoghurt and wine.

[0556] A method of preparing a food, the method comprising admixing the cells according to the present description with a food ingredient (such as a starting material for a food) is

disclosed herein.

[0557] Suitably a food as described herein is a dairy product. More preferably a dairy product as described herein is one or more of the following: a yoghurt, a cheese (such as an acid curd cheese, a hard cheese, a semi-hard' cheese, a cottage cheese), a buttermilk, quark, a sour cream, kefir, a fermented whey-based beverage, a koumiss, a milk drink and a yoghurt drink.

[0558] Here, the term "food" is used in a broad sense - and covers food for humans as well as food for animals (i.e. a feed). In a preferred aspect, the food is for human consumption.

[0559] The term feed as used herein includes raw and processed plant material and non plant material. The feed may be any feed suitable for consumption by an animal, including livestock (animal) feed, for example poultry feed, fish feed or crustacean feed for example.

VARIANTS/HOMOLOGUES/DERIVATIVES/FRAGMENTS

[0560] The present invention encompasses the use of variants, homologues, derivatives and fragments thereof, including variants, homologues, derivatives and fragments of CRISPR loci, CRISPR spacers, pseudo CRISPR spacers, cas genes or proteins, CRISPR repeats, functional CRISPR repeat-cas gene combinations and target nucleic acid sequences or transcription products thereof.

[0561] The term "variant" is used to mean a naturally occurring polypeptide or nucleotide sequences which differs from a wild-type sequence.

[0562] The term "fragment" indicates that a polypeptide or nucleotide sequence comprises a fraction of a wild-type sequence. It may comprise one or more large contiguous sections of sequence or a plurality of small sections. The sequence may also comprise other elements of sequence, for example, it may be a fusion protein with another protein. Preferably the sequence comprises at least 50%, more preferably at least 65%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, more preferably at least 96%, more preferably at least 97%, more preferably at least 98%, most preferably at least 99% of the wild-type sequence.

[0563] Preferably, the fragment retains 50%, more preferably 60%, more preferably 70%, more preferably 80%, more preferably 85%, more preferably 90%, more preferably 95%, more preferably 96%, more preferably 97%, more preferably 98%, or most preferably 99% activity of the wild-type polypeptide or nucleotide sequence.

[0564] Preferably, a CRISPR spacer or pseudo CRISPR spacer comprises at least 50%, more preferably at least 65%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, more preferably at least 96%, more preferably at least 97%, more preferably at least 98%, most preferably at least 99% of the wild-

type sequence. Preferably, a CRISPR spacer retains 50%, more preferably 60%, more preferably 70%, more preferably 80%, more preferably 85%, more preferably 90%, more preferably 95%, more preferably 96%, more preferably 97%, more preferably 98%, or most preferably 99% activity of the wild-type polypeptide or nucleotide sequence.

[0565] Preferably, a *cas* gene comprises at least 50%, more preferably at least 65%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, more preferably at least 96%, more preferably at least 97%, more preferably at least 98%, most preferably at least 99% of the wild-type sequence. Preferably, a *cas* gene retains 50%, more preferably 60%, more preferably 70%, more preferably 80%, more preferably 85%, more preferably 90%, more preferably 95%, more preferably 96%, more preferably 97%, more preferably 98%, or most preferably 99% activity of the wild-type polypeptide or nucleotide sequence.

[0566] Preferably, a Cas protein comprises at least 50%, more preferably at least 65%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, more preferably at least 96%, more preferably at least 97%, more preferably at least 98%, most preferably at least 99% of the wild-type sequence. Preferably, a Cas protein retains 50%, more preferably 60%, more preferably 70%, more preferably 80%, more preferably 85%, more preferably 90%, more preferably 95%, more preferably 96%, more preferably 97%, more preferably 98%, or most preferably 99% activity of the wild-type polypeptide or nucleotide sequence.

[0567] Preferably, a CRISPR repeat comprises at least 50%, more preferably at least 65%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, more preferably at least 96%, more preferably at least 97%, more preferably at least 98%, most preferably at least 99% of the wild-type sequence. Preferably, a CRISPR repeat retains 50%, more preferably 60%, more preferably 70%, more preferably 80%, more preferably 85%, more preferably 90%, more preferably 95%, more preferably 96%, more preferably 97%, more preferably 98%, or most preferably 99% activity of the wild-type polypeptide or nucleotide sequence.

[0568] Preferably, a functional CRISPR repeat-*cas* combination comprises at least 50%, more preferably at least 65%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, more preferably at least 96%, more preferably at least 97%, more preferably at least 98%, most preferably at least 99% of the wild-type sequence. Preferably, functional CRISPR repeat-*cas* combination retains 50%, more preferably 60%, more preferably 70%, more preferably 80%, more preferably 85%, more preferably 90%, more preferably 95%, more preferably 96%, more preferably 97%, more preferably 98%, or most preferably 99% activity of the wild-type polypeptide or nucleotide sequence.

[0569] Preferably, a target nucleic acid sequence comprises at least 50%, more preferably at least 65%, more preferably at least 80%, more preferably at least 85%, more preferably at

least 90%, more preferably at least 95%, more preferably at least 96%, more preferably at least 97%, more preferably at least 98%, most preferably at least 99% of the wild-type sequence. Preferably, a target nucleic acid sequence retains 50%, more preferably 60%, more preferably 70%, more preferably 80%, more preferably 85%, more preferably 90%, more preferably 95%, more preferably 96%, more preferably 97%, more preferably 98%, or most preferably 99% activity of the wild-type polypeptide or nucleotide sequence.

[0570] The fragment may be a functional fragment.

[0571] By a "functional fragment" of a molecule is understood a fragment retaining or possessing substantially the same biological activity as the intact molecule. In all instances, a functional fragment of a molecule retains at least 10% and at least about 25%, 50%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% of the biological activity of the intact molecule.

[0572] The term "homologue" means an entity having a certain homology with the subject amino acid sequences and the subject nucleotide sequences. Here, the term "homology" can be equated with "identity".

[0573] In the present context, a homologous sequence is taken to include an amino acid sequence, which may be at least 75, 85 or 90 % identical, preferably at least 95%, 96%, 97%, 98 % or 99% identical to the subject sequence. Although homology can also be considered in terms of similarity (*i.e.* amino acid residues having similar chemical properties/functions), in the context of the present invention it is preferred to express homology in terms of sequence identity.

[0574] In the present context, a homologous sequence is taken to include a nucleotide sequence, which may be at least 75, 85 or 90% identical, preferably at least 95%, 96%, 97%, 98 % or 99% identical to the subject sequence. Although homology can also be considered in terms of similarity (*i.e.* amino acid residues having similar chemical properties/functions), in the context of the present invention it is preferred to express homology in terms of sequence identity.

[0575] Homology comparisons may be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs can calculate % homology between two or more sequences.

[0576] % homology may be calculated over contiguous sequences, *i.e.* one sequence is aligned with the other sequence and each amino acid in one sequence is directly compared with the corresponding amino acid in the other sequence, one residue at a time. This is called an "ungapped" alignment. Typically, such ungapped alignments are performed only over a relatively short number of residues.

[0577] Although this is a very simple and consistent method, it fails to take into consideration that, for example, in an otherwise identical pair of sequences, one insertion or deletion will

cause the following amino acid residues to be put out of alignment, thus potentially resulting in a large reduction in % homology when a global alignment is performed. Consequently, most sequence comparison methods are designed to produce optimal alignments that take into consideration possible insertions and deletions without penalising unduly the overall homology score. This is achieved by inserting "gaps" in the sequence alignment to try to maximise local homology.

[0578] However, these more complex methods assign "gap penalties" to each gap that occurs in the alignment so that, for the same number of identical amino acids, a sequence alignment with as few gaps as possible - reflecting higher relatedness between the two compared sequences - will achieve a higher score than one with many gaps. "Affine gap costs" are typically used that charge a relatively high cost for the existence of a gap and a smaller penalty for each subsequent residue in the gap. This is the most commonly used gap scoring system. High gap penalties will of course produce optimised alignments with fewer gaps. Most alignment programs allow the gap penalties to be modified. However, it is preferred to use the default values when using such software for sequence comparisons. For example, when using the GCG Wisconsin Bestfit package the default gap penalty for amino acid sequences is -12 for a gap and -4 for each extension.

[0579] Calculation of maximum % homology therefore firstly requires the production of an optimal alignment, taking into consideration gap penalties. A suitable computer program for carrying out such an alignment is the GCG Wisconsin Bestfit package (University of Wisconsin, U.S.A.; Devereux et al., 1984, *Nucleic Acids Research* 12:387). Examples of other software than can perform sequence comparisons include, but are not limited to, the BLAST package (see Ausubel et al., 1999 *ibid* - Chapter 18), FASTA (Atschul et al., 1990, *J. Mol. Biol.*, 403-410), the GENWORKS suite of comparison tools and CLUSTAL. Both BLAST and FASTA are available for offline and online searching (see Ausubel et al., 1999 *ibid*, pages 7-58 to 7-60). However, for some applications, it is preferred to use the GCG Bestfit program. A new tool, called BLAST 2 Sequences is also available for comparing protein and nucleotide sequence (see *FEMS Microbiol Lett* 1999 174(2): 247-50; *FEMS Microbiol Lett* 1999 177(1): 187-8).

[0580] Although the final % homology can be measured in terms of identity, the alignment process itself is typically not based on an all-or-nothing pair comparison. Instead, a scaled similarity score matrix is generally used that assigns scores to each pairwise comparison based on chemical similarity or evolutionary distance. An example of such a matrix commonly used is the BLOSUM62 matrix - the default matrix for the BLAST suite of programs. GCG Wisconsin programs generally use either the public default values or a custom symbol comparison table if supplied (see user manual for further details). For some applications, it is preferred to use the public default values for the GCG package, or in the case of other software, the default matrix - such as BLOSUM62.

[0581] Once the software has produced an optimal alignment, it is possible to calculate % homology, preferably % sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result.

[0582] Should Gap Penalties be used when determining sequence identity, then suitably the following parameters are used:

FOR BLAST			
GAP OPEN			0
GAP EXTENSION			0
FOR CLUSTAL	DNA	PROTEIN	
WORD SIZE	2	1	K triple
GAP PENALTY	10	10	
GAP EXTENSION	0.1	0.1	

[0583] For polypeptide sequence comparison the following settings may be used: GAP creation penalty of 3.0 and GAP extension penalty of 0.1. Suitably, the degree of identity with regard to an amino acid sequence is determined over at least 5 contiguous amino acids, determined over at least 10 contiguous amino acids, over at least 15 contiguous amino acids, over at least 20 contiguous amino acids, over at least 30 contiguous amino acids, over at least 40 contiguous amino acids, over at least 50 contiguous amino acids, or over at least 60 contiguous amino acids.

[0584] The sequences may also have deletions, insertions or substitutions of amino acid residues, which produce a silent change and result in a functionally equivalent substance. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the secondary binding activity of the substance is retained. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine, valine, glycine, alanine, asparagine, glutamine, serine, threonine, phenylalanine, and tyrosine.

[0585] Conservative substitutions may be made, for example, according to the Table below. Amino acids in the same block in the second column and suitably in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar - uncharged	C S T M
		N Q
	Polar - charged	D E
		K R
AROMATIC		H F W Y

[0586] The present invention also encompasses homologous substitution (substitution and replacement are both used herein to mean the interchange of an existing amino acid residue, with an alternative residue) may occur i.e. like-for-like substitution - such as basic for basic, acidic for acidic, polar for polar etc. Non-homologous substitution may also occur i.e. from one class of residue to another or alternatively involving the inclusion of unnatural amino acids - such as ornithine (hereinafter referred to as Z), diaminobutyric acid ornithine (hereinafter referred to as B), norleucine omithine (hereinafter referred to as O), pyriylalanine, thienylalanine, naphthylalanine and phenylglycine.

[0587] Replacements may also be made by unnatural amino acids include; α * and α -disubstituted* amino acids, N-alkyl amino acids*, lactic acid*, halide derivatives of natural amino acids - such as trifluorotyrosine*, p-Cl-phenylalanine*, p-Br-phenylalanine*, p-I-phenylalanine*, L-allyl-glycine*, β -alanine*, L- α -amino butyric acid*, L- γ -amino butyric acid*, L- α -amino isobutyric acid*, L- ϵ -amino caproic acid[#], 7-amino heptanoic acid*, L-methionine sulfone^{#*}, L-norleucine*, L-norvaline*, p-nitro-L-phenylalanine*, L-hydroxyproline[#], L-thioprolin*, methyl derivatives of phenylalanine (Phe) - such as 4-methyl-Phe*, pentamethyl-Phe*, L-Phe (4-amino)[#], L-Tyr (methyl)*, L-Phe (4-isopropyl)*, L-Tic (1,2,3,4-tetrahydroisoquinoline-3-carboxyl acid)*, L-diaminopropionic acid[#] and L-Phe (4-benzyl)*. The notation * has been utilised for the purpose of the discussion above (relating to homologous or non-homologous substitution), to indicate the hydrophobic nature of the derivative whereas # has been utilised to indicate the hydrophilic nature of the derivative, #* indicates amphipathic characteristics.

[0588] Variant amino acid sequences may include suitable spacer groups that may be inserted between any two amino acid residues of the sequence including alkyl groups - such as methyl, ethyl or propyl groups - in addition to amino acid spacers - such as glycine or β -alanine residues. A further form of variation involves the presence of one or more amino acid residues in peptoid form will be well understood by those skilled in the art. For the avoidance of doubt, "the peptoid form" is used to refer to variant amino acid residues wherein the α -carbon substituent group is on the residue's nitrogen atom rather than the α -carbon. Processes for preparing peptides in the peptoid form are known in the art, for example, Simon RJ et al., PNAS (1992) 89(20), 9367-9371 and Horwell DC, Trends Biotechnol. (1995) 13(4), 132-134.

[0589] The nucleotide sequences for use in the present invention may include within them synthetic or modified nucleotides. A number of different types of modification to oligonucleotides are known in the art. These include methylphosphonate and phosphorothioate backbones and/or the addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood that the nucleotide sequences may be modified by any method available in the art. Such modifications may be carried out to enhance the *in vivo* activity or life span of nucleotide sequences useful in the present invention.

GENERAL RECOMBINANT DNA METHODOLOGY TECHNIQUES

[0590] The present invention employs, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA and immunology, which are within the capabilities of a person of ordinary skill in the art. Such techniques are explained in the literature. See, for example, J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, *Molecular Cloning: A Laboratory Manual*, Second Edition, Books 1-3, Cold Spring Harbor Laboratory Press; Ausubel, F. M. et al. (1995 and periodic supplements; *Current Protocols in Molecular Biology*, ch. 9, 13, and 16, John Wiley & Sons, New York, N.Y.); B. Roe, J. Crabtree, and A. Kahn, 1996, *DNA Isolation and Sequencing: Essential Techniques*, John Wiley & Sons; M. J. Gait (Editor), 1984, *Oligonucleotide Synthesis: A Practical Approach*, Irl Press; and, D. M. J. Lilley and J. E. Dahlberg, 1992, *Methods of Enzymology: DNA Structure Part A: Synthesis and Physical Analysis of DNA Methods in Enzymology*, Academic Press. Each of these general texts is herein incorporated by reference.

[0591] The invention will now be further described by way of Examples, which are meant to serve to assist one of ordinary skill in the art in carrying out the invention and are not intended in any way to limit the scope of the invention.

EXAMPLES

Example 1

[0592] Insertion of a phage specific spacer into an existing, functional CRISPR to provide resistance to the corresponding phage.

Strain - *Streptococcus thermophilus* ST0089

Phage- 2972

[0593] *Streptococcus thermophilus* ST0089 is an industrially important strain used in the manufacture of yogurt, is genetically amenable to manipulation, and susceptible to virulent phage 2972. The full genome sequence for phage 2972 was recently determined.

[0594] The CRISPR loci is determined in strain ST0089. This is determined preferentially by sequencing the entire genome of ST0089. Alternatively, the CRISPR loci is identified via PCR using primer sets with sequences identical to *S. thermophilus* CRISPR elements previously identified.

[0595] Once identified, the CRISPR loci sequence is determined as well as the proximal regions which should contain the relevant *cas* genes.

[0596] At least one particular CRISPR-cas locus is selected for further manipulation. Functionality of this locus is ascertained through *in silico* analysis of the spacer regions and their homologies to phage DNA sequences (*i.e.* absence and/or presence of spacer sequences and correlation to phage infectivity with strain ST0089). In the absence of this correlation, functionality is assumed based on the presence of all documented elements (*i.e.* repeats, spacers, leader sequences, *cas* genes - putatively encoding full length proteins).

[0597] A suitable spacer sequence(s) is chosen from the genome of phage 2972. The criteria of the selected spacer is based on: 1) length of the spacers within the selected CRISPR locus; 2) about 100% identity to the phage sequence; 3) theoretically any phage sequence may be selected.

[0598] In the simplest example, a CRISPR unit consisting of a phage 2972 spacer sequence, flanked by two repeating elements (identical to the selected CRISPR locus) is chemically synthesized. By definition this synthetic "CRISPR unit" is approximately 100 bp in length and is too short for ensuing integration into the CRISPR locus.

[0599] Therefore, additional flanking DNA is constructed along with the CRISPR unit. A minimum of 500 bp of homologous DNA, identical to the targeted CRISPR locus flanks the synthetic CRISPR unit, to facilitate integration.

[0600] There are at least two approaches. One construct emulates the addition of a new spacer onto the existing CRISPR. Alternatively, the entire CRISPR locus is replaced with the synthetic CRISPR unit.

[0601] The resulting CRISPR integrant is verified through DNA sequencing of the CRISPR locus prior to biological testing.

[0602] Phage sensitivity patterns of the CRISPR integrant against phage 2972 is tested and compared with the parental strain.

[0603] The constructed CRISPR integrant successfully demonstrates the direct correlation between the presence of a specific spacer within the proper context of CRISPR-cas.

Example 2

[0604] A spacer homologous to a phage DNA is inserted into a cell - such as recipient cell. The cell becomes resistant to the phage. In a CRISPR locus within the selected strain, a new CRISPR spacer is designed from phage DNA (with 100% identity to phage DNA) within the anti-receptor gene and inserted into the cell. The anti-receptor gene is targeted because

CRISPR spacers from other strains have been found to show similarity to phage anti-receptor genes. Four strains bearing spacers showing identity to phage anti-receptor genes are resistant to the particular phage. The mutant is exposed to phage and it becomes resistant to it.

Example 3

[0605] A plasmid comprising a CRISPR spacer is prepared, and we show that this plasmid cannot be transferred into a cell that contains the same spacer, whereas the plasmid without the spacer can be transformed into the cell.

Example 4

[0606] A spacer is inserted into an original host, but not in a CRISPR locus, and the resulting mutant retains its sensitivity to the phage, showing that the spacer needs to be in a particular environment within a CRISPR and *cas* genes

Example 5

[0607] A whole CRISPR repeat-*cas* combination is inserted into a cell - such as a recipient cell - to provide immunity against incoming nucleic acid.

Example 6

[0608] For a particular CRISPR repeat-*cas* combination present in two different strains, the "exchange" of spacers modifies their phenotypes (phage sensitivity/resistance).

Example 7

[0609] One or more *cas* genes (from a functional CRISPR-*cas* unit) are deleted. *Cas* genes are necessary for immunity to be provided. *Cas* mutants are still sensitive to the phage, despite the presence of the spacer identical to phage DNA.

Example 8

[0610] The deleted *cas* genes are cloned on a plasmid. It is possible to provide the *cas* genes

in trans to the host. Where the *cas* gene is knocked out, immunity can be restored.

Example 9

[0611] Different *cas*-CRISPR-repeat combinations are prepared. Not only are *cas* genes or proteins required, but also, specific *cas*-CRISPR repeat pairs are required for functionality. When *cas* genes or proteins are provided from another CRISPR locus, the strain remains sensitive to the phage.

Example 10

[0612] When a particular CRISPR spacer is deleted from a naturally occurring CRISPR locus, this removes immunity against a given phage and the host becomes sensitive (loses resistance) to the phage to which the spacer is homologous to.

Example 11

[0613] Integration of a CRISPR spacer into the CRISPR locus of a bacterium provides resistance against a bacteriophage that the CRISPR spacer shows identity to

(A) *Streptococcus thermophilus* strain DGCC7710RH1

Streptococcus thermophilus

[0614] *Streptococcus thermophilus* strain DGCC7710 (deposited at the French "Collection Nationale de Cultures de Microorganismes" under number CNCM I-2423) possesses at least 3 CRISPR loci: CRISPR1, CRISPR2, and CRISPR3. In strains CNRZ1066 and LMG18311 for which the complete genome sequence is known (Bolotin *et al.*, 2004), CRISPR1 is located at the same chromosomal locus: between *str0660* (or *stu0660*) and *str0661* (or *stu0661*).

[0615] In strain DGCC7710, CRISPR1 is also located at the same chromosomal locus, between highly similar genes. CRISPR1 of strain DGCC7710 contains 33 repeats (including the terminal repeat), and thus 32 spacers.

[0616] All these spacers are different from each other. Most of these spacers are new (not yet described within CRISPR loci), but four spacers close to the CRISPR1 trailer are identical to already known CRISPR1 spacers:

- the 28th spacer of DGCC7710 is 100% identical to the 31st CRISPR1 spacer of strain CNRZ1575 (Genbank accession number DQ072991);
- the 30th spacer of DGCC7710 is 100% identical to the 27th CRISPR1 spacer of strain CNRZ703 (Genbank accession number DQ072990);
- the 31st spacer of DGCC7710 is 100% identical to the 28th CRISPR1 spacer of strain CNRZ703 (Genbank accession number DQ072990);
- the 32nd spacer of DGCC7710 is 100% identical to the 30th CRISPR1 spacer of strain CNRZ703 (Genbank accession number DQ072990).

Virulent bacteriophage

[0617] D858 is a bacteriophage belonging to the *Siphoviridae* family of viruses. Its genome sequence has been completely determined but is not published yet. This phage is virulent to *S. thermophilus* strain DGCC7710.

Phage resistant mutant

[0618] *Streptococcus thermophilus* strain DGCC7710RH1 has been isolated as a natural phage resistant mutant using DGCC7710 as the parental strain, and phage D858 as the virulent phage.

[0619] CRISPR1 of strain DGCC7710-RH1 contains 34 repeats (including the terminal repeat), and thus 33 spacers. When compared to the CRISPR1 sequence of *Streptococcus thermophilus* strain DGCC7710, the CRISPR1 sequence of *Streptococcus thermophilus* strain DGCC7710-RH1 possesses one additional new spacer (and of course one additional repeat which flanks the new spacer) at one end of the CRISPR locus (ie. close to the leader, at the 5' end of the CRISPR locus).

[0620] All the other spacers of CRISPR1 locus are unchanged.

[0621] The CRISPR1 sequence (5'-3') of strain DGCC7710-RH1 is:

>CRISPR1_DGCC7710-RH1

```
caaggacagttattgattttataatcactatgtgggtataaaaaacgtcaaaatttcatttgag
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtcaacaattgcaacatcttataacccactt
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgtttgacagcaaatcaagattcgaattgt
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgacgaggagctattggcacaacttaca
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatttgacaatctgtgtgaccactgttatac
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacacttggcaggcttatctactcaacagcga
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACctgttccctgttcttttgttctgtatcttttc
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACttcattcttccgtttttgtttgogaatect
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgctggcgaggaaacgaacaaggcctcaaca
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcctagagtggaaaactagaaacagattcaa
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataatqccqttgaattacacqgcaaaqqtca.....
```

GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgagcgagctcgaaataatcttaattacaag
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgttcgctagcgtcatgtggtaacgtattta
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACggcgctcccaatcctgattaataacttactcg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaacaacagcaagacaaggagatgatgctatg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgacacaagaacgtatgcaagagttcaag
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacaattcttcatccggtaactgctcaagtg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattaagggcatagaaggagacaacatg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatattttaaatacattttcataacttcat
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACGcagtatcagcaagcaagctgttagttact
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataaactatgaaattttataatttttaaga
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaataattttatggtagcttaatatcattg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgcatcgagcagcttcgagtttacggttc
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtctatatcgagggtcaactaacaattatgct
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcggttcaaattctgttttaggtacattt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcaatacgacaagagtttaaattggtctt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcttagctgtccaatccacgaacgtggatg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcaaccaacggtaacagctactttttacagt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataactgaaggataggagcttgtaaagtct
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtaatgctacatctcaaaggatgatcccaga
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaagtagttgatgacctctacaatggtttat
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacctagaagcatttgagcgtatattgattg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattttgccccttctttgccccttgactag
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaccattagcaatcatttgtgccattgagt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAGT
 ttgattcaacataaaaagccagttcaattgaacttggcttt

Legend

Leader sequence:

5' caaggacagttattgattttataatcactatgtgggtataaaaacgtcaaaatttcatttgag 3' Integrated sequence comprising a CRISPR Repeat in upper case and a CRISPR spacer (ie. tagging sequence) in lower case.

CRISPR Repeats

[0622]

Terminal repeat: 5' gttttgtactctcaagatttaagtaactgtacagt 3'

Trailer sequence: 5' ttgattcaacataaaaagccagttcaattgaacttggcttt3'

[0623] The sequence of the new spacer exists within the D858 phage genome and is represented herein as SEQ ID No. 534.

[0624] The sequence of the spacer is found between positions 31921 and 31950 bp (ie. on the plus strand) of D858's genome (and has 100% identity to the D858 genomic sequence over 30 nucleotides):

spacer	1	tcaacaattgcaacatcttataaccactt	30
D858	31921	tcaacaattgcaacatcttataaccactt	31950

[0625] The new spacer that is integrated into the CRISPR1 locus of *Streptococcus thermophilus* strain DGCC7710-RH1 confers to this strain resistance to phage D858, as represented in Figure 5 and Table 1.

(B) *Streptococcus thermophilus* strain DGCC7710RH2

[0626] *Streptococcus thermophilus* strain DGCC7710-RH2 has been isolated as a natural phage resistant mutant using *Streptococcus thermophilus* strain DGCC7710 as the parental strain, and phage D858 as the virulent phage.

[0627] CRISPR1 of *Streptococcus thermophilus* strain DGCC7710-RH2 contains 34 repeats (including the terminal repeat), and thus 33 spacers. When compared to the CRISPR1 sequence of *Streptococcus thermophilus* strain DGCC7710, the CRISPR1 sequence of *Streptococcus thermophilus* strain DGCC7710-RH2 possesses one additional new spacer (and of course one additional repeat which flanks the new spacer) at one end of the CRISPR locus (ie. close to the leader, at the 5' end of the CRISPR locus). All the other spacers of CRISPR1 locus are unchanged.

[0628] The CRISPR1 sequence (5'-3') of strain DGCC7710-RH2 is:

```
>CRISPR1_DGCC7710-RH2
caaggacagttattgattttataatcactatgtgggtataaaaacgtcaaaatttcatttgag
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACttacgtttgaaaagaatatcaaatcaatga
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgtttgacagcaaatcaagattcgaattgt
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgacgaggagctattggcacaacttaca
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatttgacaatctgctgaccactgttate
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacacttggcagggettattactcaacagcga
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACctgttcccttgttcttttgttgatctcttcc
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACttcattcttccggtttttgtttgcgaatcct
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgctggcgaggaaacgaacaaggcctcaaca
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcatagagtggaactagaaacagattcaa
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataatgcccgttgaaattacacggcaagggtca
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgagcgagctcgaaataatcttaattacaag
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgttcgctagcgtcatgttgtaacgtattta
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACggcgtcccaatccttgattaatacttactcgt
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaacaacagcaagacaagaggatgatgctatg
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgacacaagaacgatatgcaagagtccaag
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacaattcttcatccggttaactgctcaagtg
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatataagggtcatagaaaggagacaaatg
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatatattaaaatcattttcataacttcat
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACGcagtatcagcaagcaagctgttagttact
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataaactatgaaattttataatttttaaga
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaataattttatggtatagcttaatatcattg
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgcacgcagcagcttcgagtttacggttc
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtctatatcagaggtcaactaacaattatgct
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcggttcaaattctgttttaggtacattt
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcaatacgacaagagttaaaatggtctt
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcttagctgtccaatccacgaacgtggatg
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcaaccaacggtaacagctactttttacagt
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataactgaaggataggagcttgtaaagtct
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtaatgctacatctcaaaggatgatcccaga
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaagtagttgatgacctctacaatggtttat
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacctagaagcatttgagcgatatattgattg
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattttgccccttcttttggccccttgactag
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaccattagcaatcatttgtgcccattgagtg
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAGT
ttgattcaacataaaaagccaggttcaattgaacttggccttt
```

Legend

[0629] Leader sequence:

5' caaggacagttattgattttataatcactatgtgggtataaaaacgtcaaaatttcatttgag 3'

Integrated sequence comprising a CRISPR Repeat in upper case and a CRISPR spacer (ie. tagging sequence) in lower case.

CRISPR Repeats

[0630]

Terminal repeat: 5' gttttgtactctcaagatttaagtaactgtacagt 3'

Trailer sequence: 5' ttgattcaacataaaaagccagttcaattgaacttggcttt3'

[0631] It has been shown that the sequence of the new spacer exists within the D858 phage genome.

[0632] The sequence of the spacer (represented herein as SEQ ID No. 535) is found between positions 17215 and 17244 bp (ie. on the plus strand) of D858's genome (and has 100% identity to the D858 genomic sequence over 30 nucleotides):

```
spacer      1 ttacgtttgaaaagaatatcaaatcaatga 30
              |||
D858      17215 ttacgtttgaaaagaatatcaaatcaatga 17244
```

[0633] The new spacer that is integrated into the CRISPR1 locus of *Streptococcus thermophilus* strain DGCC7710-RH2 confers to *Streptococcus thermophilus* strain DGCC7710-RH2 a resistance to phage D858, as represented in Figure 6 and Table 1.

Example 12

Construct integration and knockout

Materials and Methods

Strains and plasmids

[0634]

Streptococcus thermophilus DGCC7710 parent strain, sensitive to phages 858 and 2972

Streptococcus thermophilus DGCC7778 CRISPR mutant resistant to 858

Streptococcus thermophilus DGCC7778cas1KO

Streptococcus thermophilus DGCC7778cas4KO

Streptococcus thermophilus DGCC7778RT

Streptococcus thermophilus DGCC7778RT'

Streptococcus thermophilus DGCC7710R2 CRISPR mutant resistant to 2972

Streptococcus thermophilus DGCC7710R2S1S2

Escherichia coli EC1,000 provides pORI28 (Russell and Klaenhammer, 2001)

Escherichia coli pCR2.1TOPO provides pTOPO (Invitrogen catalog #K4500-01)

pTOPO is a plasmid used for sub-cloning of the various constructs

pTOPOcas1ko contains an integral fragment of cas1

pTOPOcas4ko contains an integral fragment of cas4

pTOPOS1S2 contains the S1S2 spacer construct

pTOPO RT contains the RT terminal repeat construct

pORI28 is a plasmid used for integration of the various constructs in the chromosome of *Streptococcus thermophilus* strains.

pORlcas1ko contains an integral fragment of cas1

pORlcas4ko contains an integral fragment of cas4

pORIS1S2 contains the S1S2 spacer construct

purist contains the RT terminal repeat construct

Primers

[0635] Cas1

5'-caaatggatagagaaacgc-3' and 5'-ctgataaggtgttcgtgtcc-3'

Cas4

5'-ggagcagatggaatacaagaaagg-3' and 5'-gagagactaggtgtctcagca-3'

S1S2 and RT

P1 5'-acaacaacagagaagtatctcattg-3'

P2 5'-aacgagtacactcactattgtacg-3'

P3 5'-tccactcacgtacaaatagtgagtgtactcgttttgtattctcaagatttaagtaactgtacagtttgattcaacataaaaag-3'

P4 5'-cttccctcatcctcgctttgggt-3'

[0636] Strains and phages were obtained from the Danisco Culture Collection, or from referenced material (Russell and Klaenhammer, Applied and Environmental Microbiology 2001, 67:43691-4364; Levesque et al., Applied and Environmental Microbiology 2005 71:4057-4068).

[0637] Phage preparation, purification and tests were carried out using methods described previously (Duplessis et al., Virology 2005, 340:192-208; Levesque et al., Applied and Environmental Microbiology 2005 71:4057-4068).

[0638] *Streptococcus thermophilus* strains were grown at 37C or 42C in M17 (Difco Laboratories) supplemented with 0.5 % lactose or sucrose. For phage infection, 10mM CaCl₂ were added to the medium prior to phage infection, as described previously (Duplessis et al., Virology 2005, 340:192-208; Levesque et al., Applied and Environmental Microbiology 2005 71:4057-4068).

[0639] Enzymes used to carry out restriction digests and PCR were purchased from Invitrogen and used according to the manufacturer's instructions. PCRs were carried out on an Eppendorf Mastercycler Gradient thermocycler as described previously (Barrangou et al., 2002 Applied and Environmental Microbiology 68:2877-2884).

[0640] Gene inactivation and site-specific plasmid insertion via homologous recombination in the *Streptococcus thermophilus* chromosome were carried out by sub-cloning into the Invitrogen pCR2.1TOPO system, subsequent cloning in the pORI system using *Escherichia coli* as a host and the constructs were ultimately purified and transformed into *Streptococcus thermophilus* as previously described (Russell and Klaenhammer, Applied and Environmental Microbiology 2001, 67:43691-4364)

(1) RT Construct Integration

[0641] Using the RT Construct engineered as shown in Figure 17, the construct was inserted just after *cas4*, as shown in Figure 18.

[0642] The parent DGCC7778 is resistant to phage 858.

[0643] The parent has two spacers (S1 and D2) which are identical to phage 858 DNA.

[0644] The resulting strain (RT) loses resistance to phage 858, as shown in Table 1. This demonstrates that *cas* genes need to be in the immediate vicinity of the spacer(s) to confer resistance.

(2) *Cas1* knockout

[0645] As shown in Figure 12 the parent DGCC7778 is engineered such that the *cas1* gene is disrupted. As shown in Table 1, this results in a loss of resistance, meaning that *cas1* is needed to confer resistance.

(3) *Cas4* knockout

[0646] As shown in Figure 12 the parent DGCC7778 is engineered such that the *cas4* gene is disrupted.

(4) S1S2 construct integration

[0647] As shown in Figures 14-16 the a S1S2 construct is integrated into the parent DGCC7710.

SUMMARY

[0648] CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats) (a.k.a. SPIDR - SPacer Interspersed Direct Repeats) constitute a family of recently described DNA loci widespread throughout prokaryotic genomes. They are constituted of short and highly conserved DNA palindromic repeats which are regularly interspaced by highly polymorphic sequences of about the same length. Additionally, *cas* genes (CRISPR-associated genes) are usually present in the vicinity of CRISPR sequences. In the literature no clear physiological function has been attributed yet to CRISPR sequences or *cas* genes.

[0649] Here we suggest that CRISPR sequences in combination with *cas* genes may be used to provide resistance against incoming nucleic acid. Particularly, we propose that the spacers within CRISPR loci provide the specificity for immunity against incoming nucleic acid. As a result, we suggest that *cas* genes in association with CRISPR sequences be used to provide cells with resistance against particular nucleic acid sequences - such as bacteriophages, plasmids, transposons, and insertion sequences. Additionally, these elements can be manipulated to generate targeted immunity against particular nucleic acid sequences, such as

phage components, antibiotic resistance genes, virulence factors, novel sequences, undesirable elements and the like. Thus, the simple knowledge of *inter alia* CRISPR spacer sequences for a given bacterial strain would be an advantage to determine its lysotype (the lysotype defines the resistance/sensitivity of a given bacterium to various bacteriophages) and predict its ability to survive exposure to defined nucleic acid sequences. Consequently, characterisation of CRISPR loci in bacteria could help to determine, predict and modify host-phage interaction. Particular application of CRISPR genetic engineering, by addition, deletion or modification of spacer sequences, could lead to phage resistant bacterial variants.

[0650] CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats), also known as SPIDR (SPacer Interspersed Direct Repeats), form a new family of repeated sequences which have been identified in complete genome sequences, in numerous prokaryotes, mostly on chromosomes but also on plasmids (Mojica *et al.*, 2000; Jansen *et al.*, 2002a). CRISPR loci are constituted of short and highly conserved DNA repeats (24 to 40 bp, repeated from 1 to 140 times) which are partially palindromic. While there are certain limits to the base degeneracy between repeats from different loci and species, there is no absolute conserved sequence throughout all observed repeats. Moreover the repeats are seemingly oriented within a particular locus, with regards to the neighbouring genes. These repeated sequences (usually specific to a species) are interspaced by polymorphic sequences of constant length (20 to 58 bp depending on the CRISPR) which are designated as "spacers". Up to 20 different CRISPR loci have been found within a single chromosome. Figure 1 describes one of the CRISPR identified in *Streptococcus thermophilus* CNRZ1066.

[0651] For example, the genome of *S. thermophilus* LMG18311 contains 3 CRISPR loci. The 36-bp repeated sequences are different in CRISPR1 (34 repeats), CRISPR2 (5 repeats), and CRISPR3 (one single sequence); nevertheless, they are perfectly conserved within each locus. CRISPR1 and CRISPR2 repeats are respectively interspaced by 33 and 4 sequences of 30 bp in length. All these spacers are different from each other (apart from minor exceptions: very few spacers may be present twice within a given CRISPR locus). They are also different from those found in other strains - such as CNRZ1066 (41 spacers within CRISPR1) or LMD-9 (16 spacers within CRISPR1 and 8 within CRISPR3), which are *S. thermophilus* strains that have very similar genomes.

[0652] Although the biological function of CRISPR loci is unknown some hypotheses have been proposed. For example, it has been proposed that they may be involved in the attachment of the chromosome to a cellular structure, or in the chromosome replication and replicon partitioning, but no experimental demonstration has been reported to confirm these hypotheses.

[0653] Generally CRISPR loci are immediately adjacent to a group of 4 to 7. genes which have been called *cas* (CRISPR-associated) genes (Jansen *et al.*, 2002b). At the present time no clear physiological role has been attributed to Cas proteins, but for some of them the presence of particular protein motifs suggests that they could act as a DNA gyrase or a DNA polymerase. These clusters of 4 to 7 *cas* genes, either originating from different loci within a given genome

or originating from different microorganisms, can be distinguished and grouped into different types on the basis of sequence similarity. One of our major findings is that a given set of *cas* genes is always associated with a given repeated sequence within a particular CRISPR locus. In other words, *cas* genes [or Cas proteins] seem to be specific for a given DNA repeat, meaning that *cas* genes [or Cas proteins] and the repeated sequence could form a functional pair. Dotplot analyses indicate that the clusters and groups obtained when analyzing Cas protein sequences are similar to those obtained when analyzing CRISPR repeats (as shown in figure 2).

[0654] In *S. thermophilus*, a bacterial species for which several phage genomes have been sequenced, the 30-bp spacers are often identical to phage DNA (Figure 3). This observation has also been made for the spacer sequences of many other bacterial genera and species for which phage DNA sequences are known. Furthermore it has been previously mentioned in at least two recent publications (Pourcel *et al.*, 2005; Mojica *et al.*, 2005). On the other hand the absence of significant sequence similarity for the remaining spacer sequences may be explained by the fact that only a few phage genomes are available at this time. On the basis of very high DNA sequence similarities between some CRISPR spacers and bacteriophage sequences, we propose that the specificity of the CRISPR spacers participates in the determination of the strain lysotype. To support the proposal of an effect of CRISPR spacer sequences on the bacterial immunity against bacteriophages, it was found that a significant proportion of matches for spacers in bacteriophage genome sequences occur within genes likely involved in the host specificity (see figure 3). Another hypothesis could be that the spacer sequences are recognized by the bacterium as foreign DNA. Thus, the bacterium would eliminate the nucleic acid molecule bearing this sequence when entering the cell. One argument supporting this idea is the deduced peculiar structure of CRISPR. Indeed, we propose that the repeat elements provide a structural feature while the spacers containing the sequence providing specific immunity against incoming nucleic acid. The palindromic repeats have the potential to form very stable hairpin (stem-loop) structures (see figure 4), and they are separated by spacers whose size corresponds to roughly 3 turns of the DNA helix (although it can vary between 2 and 5). Thus any CRISPR locus could be highly structured into a series of regularly spaced DNA hairpins.

[0655] Advantageously, the lysotype of a given bacterial strain may be modified either by natural generation of resistant derivatives (Bacteriophage Insensitive Mutants), or by genetic engineering. Specifically, genetic engineering solutions may be designed by, for example, addition, by deletion, or by modification of the spacer sequences or even a complete CRISPR locus.

[0656] Examples of applications of this invention include, but are not limited to:

1. (i) Phage resistance. Particular CRISPR spacers derived from bacteriophage DNA may be added within a bacterial host CRISPR locus as to provide resistance against this particular bacteriophage, thus preventing phage attack. Additionally, particular regions within the phage genome (host specificity proteins) can be targeted that provide

particular phage-host recognition, or that are highly conserved within phage DNA, such as sequences from helicase or primase genes, head and tail structural proteins, or proteins with conserved domains (eg. helicase, holing, lysine, and others) or conserved sequences amongst important phage genes.

2. (ii) Resistance to plasmid transfer. Particular CRISPR spacers derived from plasmid DNA can be added within a bacterium CRISPR locus as to provide resistance against this particular plasmid, thus preventing transfer of foreign DNA into the microbe. Specifically, particular regions within the plasmid DNA can be targeted as to provide immunity against plasmid DNA, such as sequences within the plasmid's origin of replication.
3. (iii) Resistance to mobile genetic elements. Particular CRISPR spacers derived from mobile genetic element DNA can be added within a bacterium CRISPR locus as to provide resistance against mobile genetic elements such as transposable elements and insertion sequences, thus preventing transfer of foreign DNA and genetic drift. Specifically, particular regions within transposons and insertion sequences can be targeted as to provide immunity against mobile genetic elements. For example, targets can include conjugative transposons (Tn916), class II transposons (Tn501), or insertion sequences (IS26).
4. (iv) Resistance to antibiotic resistance genes. Particular CRISPR spacers derived from antibiotic resistance encoding genes can be added within a bacterium CRISPR locus as to prevent transfer of genes conferring resistance to antibiotics into the bacterial host, thus reducing the risk of acquiring antibiotic resistance markers. For example, targets can include *vanR*, a gene conferring resistance to vancomycin, or *tetR*, a gene conferring resistance to tetracycline, or targeting beta-lactamase inhibitors.
5. (v) Resistance to genes encoding virulence factors. Particular CRISPR spacers derived from genes encoding virulence factors can be added within a bacterium CRISPR locus as to provide resistance against the transfer of genes conferring virulence into the bacterium. For example, factors commonly contributing to virulence in microbial pathogens can be targeted, such as toxins, internalins and hemolysins.
6. (vi) Diagnostics. The CRISPR spacers within a particular bacterium may be detected or sequenced as to predict/determine the likely sensitivity of particular microbes to bacteriophage, and thus be used as a lysotype predictor for microbial selection.
7. (vii) Resistance to novel sequences. Novel spacer sequences can be synthesized *de novo*, engineered and integrated into a CRISPR within a selected bacterial host as to provide resistance to a particular identical and novel sequence present into an infecting DNA molecule.

[0657] Since CRISPRs are wide-spread among bacterial species, the aforementioned applications could be used in a large variety of organisms. CRISPR loci have been described in a number of Gram-positive (including lactic acid bacteria) and Gram-negative bacteria. Thus, CRISPR loci in association with *cas* genes can be used to characterize/modify strain lysotype and generate resistance to nucleic acid in a wide range of bacteria. In addition to potential applications for phage resistance, it has been mentioned in the literature that CRISPR

sequences show some homology to mobile genetic elements such as plasmids and transposons (Mojica *et al.*, 2005).

[0658] The use of a combination of a CRISPR locus and one or more *cas* genes to provide resistance against a defined nucleic acid is disclosed herein.

[0659] Suitably, the nucleic acid is DNA.

[0660] Suitably, the nucleic acid is RNA.

[0661] Suitably, the nucleic acid is derivable (preferably, derived) from a phage.

[0662] Suitably, the nucleic acid is derivable (preferably, derived) from a plasmid.

[0663] Suitably, the nucleic acid is derivable (preferably, derived) from a mobile genetic element.

[0664] Suitably, the nucleic acid is derivable (preferably, derived) from a transposon (Tn).

[0665] Suitably, the nucleic acid is derivable (preferably, derived) from an insertion sequence (IS).

[0666] Suitably, the nucleic acid nucleic acid is derivable (preferably, derived) from undesirable targeted genetic elements.

[0667] Suitably, the nucleic acid is derivable (preferably, derived) from an antibiotic resistance gene.

[0668] Suitably, the nucleic acid is derivable (preferably, derived) from a virulence factor.

[0669] Suitably, the nucleic acid is derivable (preferably, derived) from a pathogenicity island.

[0670] Suitably, the nucleic acid nucleic acid is derivable (preferably, derived) from a novel sequence, so as to provide resistance against entities carrying this particular molecule.

[0671] The use of CRISPR for identification and typing is disclosed herein.

[0672] The use of one or more *cas* genes and one or more CRISPR elements (eg. one or more CRISPR repeats and/or CRISPR spacers) for modulating resistance in a cell against a target nucleic acid or a transcription product thereof is disclosed herein.

TABLE 1

Strains	Phage 2972			Phage 858	
	BIM on ¹	Phage sensitivity ²	Spacer-phage homology ³	Phage sensitivity ²	Spacer-phage homology ³
DGCC7710	-	S	Ctrl	S	Ctrl
DGCC7778	858	S	> 10 SNPs	R	100% (2 spacers)
DGCC7710-RH1	858	R	100%	R	100%
DGCC7710-RH2	858	R	100%	R	100%
DGCC7778RT	858	S	> 10 SNPs	S	100% but not next to <i>cas</i>
DGCC7778RT'	858	S	> 10 SNPs	S	No spacers left
DGCC7778 <i>cas1</i>	858	S	> 10 SNPs	S	100% (2 spacers):but <i>cas1</i> KO
DGCC7778 <i>cas4</i>	858	S	> 10 SNPs	R	100% (2 spacers):but <i>cas4</i> KO
DGCC7710-R2	2972	R	100% (1 spacer)	S	5 SNPs
DGCC7710-R2S1S2	2972	S	100% but not next to <i>cas</i>	R	S1S2 are 100% identical to phage 858
1 Phage used to generate Bacteriophage Insensitive Mutants (BIMs) 2 Phage sensitivity of the strain, S = sensitive, R = resistant as determined by spot and plaque assays 3 Homology between the new spacer of the mutant, and the DNA sequence of the phage used to generate the mutant Phages retained the ability to adsorb to the mutants					

REFERENCES

[0673]

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151(8):2551-61.

Groenen PM, Bunschoten AE, van Soolingen D, & JD van Embden (1993). Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain

differentiation by a novel typing method. *Molecular Microbiology* 10:1057-1065.

Hoe N, Nakashima K, Grigsby D, Pan X, Dou SJ, Naidich S, Garcia M, Kahn E, Bergmire-Seat D, & JM Musser (1999). Rapid molecular genetic subtyping of serotype M1 group A *Streptococcus* strains. *Emerging Infectious Diseases* 5:254-263.

Jansen R, Van Embden JDA, Gaastra W, & LM Schouls (2002a). Identification of a novel family of sequence repeats among prokaryotes. *OMICS* 6:23-33.

Jansen R, Van Embden JDA, Gaastra W, & LM Schouls (2002b). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology* 43:1565-1575

Kamerbeek J, Schouls L, Kolk A, Van Agterveld M, Van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, & J Van Embden (1997). Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *Journal of Clinical Microbiology* 35:907-914

Mojica FJM, Diez-Villasenor C, Soria E, & G Juez (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular Microbiology* 36:244-246

Mojica FJM, Diez-Villasenor C, Garcia-Martinez J, & E Soria (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution* 60:174-182

Pourcel C, Savignol G, & G Vergnaud (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA and provide additional tools for evolutionary studies. *Microbiology* 151:653-663

Saunders NFW, Goodchild A, Raftery M, Guilhaus M, Curmi PMG, & R Cavicchioli (2005). Predicted roles for hypothetical proteins in the low-temperature expressed proteome of the antarctic archaeon *Methanococcoides burtonii*. *Journal of Proteome Research* 4:464-472

Mongodin EF, Hance IR, DeBoy RT, Gill SR, Daugherty S, Huber R, Fraser CM, Stetter K, & KE Nelson (2005). Gene transfer and genome plasticity in *Thermotoga maritima*, a model hyperthermophilic species. *Journal of Bacteriology* 187:4935-4944

Peng X, Brugger K, Shen L, She Q, & RA Garrett (2003). Genus-specific protein binding to the large clusters of DNA repeats (Short Regularly Spaced Repeats) present in *Sulfolobus* genomes. *Journal of Bacteriology* 185:2410-2417

[0674] All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and system of the present invention will be apparent to those skilled in the art without departing from the scope and spirit of the present invention. Although the present invention has been described in connection with specific preferred embodiments, it should be understood that the invention as

claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in biochemistry, microbiology and molecular biology or related fields are intended to be within the scope of the following claims.

[0675] The invention also relates to the following aspects as defined in the following numbered paragraphs:

1. 1. Use of one or more *cas* genes or proteins for modulating resistance in a cell against a target nucleic acid or a transcription product thereof.
2. 2. Use according to paragraph 1, wherein the one or more *cas* genes or proteins are used in combination with two or more CRISPR repeats.
3. 3. Use according to paragraph 2, wherein the one or more *cas* genes or proteins and/or the two or more CRISPR repeats are or are derivable (preferably, derived) from the same cell.
4. 4. Use according to paragraph 2 or paragraph 3, wherein the one or more *cas* genes or proteins and the two or more CRISPR repeats naturally co-occur in the same cell.
5. 5. Use according to any of the preceding paragraphs, wherein the one or more *cas* genes or proteins are used in combination with one or more CRISPR spacers.
6. 6. Use according to paragraph 5, wherein the CRISPR spacer(s) is or is derivable (preferably, derived) from an organism that is different to the cell from which the one or more *cas* genes or proteins and/or the two or more CRISPR repeats are or are derivable (preferably, derived).
7. 7. Use according to paragraph 6, wherein the spacer is obtained from a cell which is resistant to a target nucleic acid.
8. 8. Use according to paragraph 5, wherein the CRISPR spacer is a synthetic nucleic acid sequence.
9. 9. Use according to any of paragraphs 5-8, wherein the CRISPR spacer(s) have homology to the target nucleic acid.
10. 10. Use according to paragraph 9, wherein the CRISPR spacer(s) have 100% identity to the target nucleic acid over at least the length of the CRISPR spacer core.
11. 11. Use according to any of the preceding paragraphs, wherein the one or more *cas* genes or proteins are used in combination with at least one or more CRISPR spacers and at least two or more CRISPR repeats.
12. 12. Use according to any of the preceding paragraphs, wherein the target nucleic acid or transcription product thereof is or is derivable (preferably, derived) from bacteriophage DNA.
13. 13. Use according to any of the preceding paragraphs, wherein the target nucleic acid or transcription product thereof is or is derivable (preferably, derived) from plasmid DNA.
14. 14. Use according to any of the preceding paragraphs, wherein the target nucleic acid or transcription product thereof is or is derivable (preferably, derived) from a mobile genetic element.
15. 15. Use according to paragraph 14, wherein the target nucleic acid or transcription product thereof is or is derivable (preferably, derived) from a transposable element or an

insertion sequence.

16. 16. Use according to any of the preceding paragraph, wherein the target nucleic acid or transcription product thereof is or is derivable (preferably, derived) from an antibiotic resistance gene.
17. 17. Use according to any of the preceding paragraphs, wherein the target nucleic acid or transcription product thereof is or is derivable (preferably, derived) from a nucleic acid encoding a virulence factor.
18. 18. Use according to paragraph 17, wherein the virulence factor is selected from the group consisting of a toxin-, an internalin- and a hemolysin-encoding nucleic acid.
19. 19. Use of a recombinant nucleic acid sequence comprising at least one *cas* gene and at least two CRISPR repeats together with at least one CRISPR spacer, wherein at least one CRISPR spacer is heterologous to at least one *cas* gene and/or at least two CRISPR repeats to modulate resistance against a target nucleic acid or transcription product thereof.
20. 20. A nucleic acid sequence consisting essentially of at least one *cas* gene.
21. 21. A nucleic acid sequence consisting essentially of at least one *cas* gene and at least two CRISPR repeats.
22. 22. A nucleic acid sequence consisting essentially of at least one *cas* gene and at least one CRISPR spacer.
23. 23. A nucleic acid sequence consisting essentially of at least one *cas* gene, at least one CRISPR spacer and at least two CRISPR repeats.
24. 24. A recombinant nucleic acid sequence comprising at least one *cas* gene and at least two CRISPR repeats together with at least one CRISPR spacer, wherein the CRISPR spacer is heterologous to the at least one *cas* gene and/or the at least two CRISPR repeats.
25. 25. The nucleic acid sequence according to any of paragraphs 20-24, wherein the one or more *cas* genes and the two or more CRISPR repeats are or are derivable (preferably, derived) from the same cell.
26. 26. The nucleic acid sequence according to any of paragraphs 20-25, wherein the one or more *cas* genes and the two or more CRISPR repeats naturally co-occur in the same cell.
27. 27. The nucleic acid sequence according to any of paragraphs 20-26, wherein the CRISPR spacers are or are derivable (preferably, derived) from an organism that is different to the cell from which the one or more *cas* genes and/or the two or more CRISPR repeats are or are derivable (preferably, derived).
28. 28. A construct comprising one or more of the nucleic acid sequences according to any of paragraphs 20-27.
29. 29. A vector comprising one or more of the nucleic acid sequences according to any of paragraphs 20-27 or one or more of the constructs according to paragraph 24.
30. 30. A cell comprising the nucleic acid sequence according to any of paragraphs 20-27, the construct according to paragraph 28 or the vector according to paragraph 29.
31. 31. The cell according to paragraph 30, wherein the cell is a recipient cell or a host cell.
32. 32. A method for modulating (e.g. conferring or increasing) the resistance of a cell against a target nucleic acid or a transcription product thereof comprising the steps of:

1. (i) identifying a sequence (eg. a conserved sequence) in an organism (preferably, a sequence essential to the function or survival of the organism);
 2. (ii) preparing a CRISPR spacer which is homologous to the identified sequence;
 3. (iii) preparing a nucleic acid (eg. a recombinant nucleic acid) comprising at least one *cas* gene and at least two CRISPR repeats together with the CRISPR spacer; and
 4. (iv) introducing said nucleic acid into a cell thus to render the cell resistant to said target nucleic acid or transcription product thereof.
33. 33. A method for modulating (eg. conferring or increasing) the resistance of a cell against a target nucleic acid or a transcription product thereof comprising the steps of:
1. (i) identifying one or more CRISPR spacers or pseudo CRISPR spacers in an organism resistant to the target nucleic acid or transcription product thereof;
 2. (ii) preparing a recombinant nucleic acid comprising at least one *cas* gene or protein and at least two CRISPR repeats together with said identified one or more spacers; and
 3. (iii) introducing said recombinant nucleic acid into a cell thus to render the cell resistant to said target nucleic acid or transcription product thereof.
34. 34. A method for modulating (eg. conferring or increasing) the resistance of a cell comprising at least one or more *cas* genes or proteins and two or more CRISPR repeats against a target nucleic acid or a transcription product thereof comprising the steps of:
1. (i) identifying one or more CRISPR spacers in an organism resistant to the target nucleic acid or transcription product thereof; and
 2. (ii) modifying the sequence of one or more CRISPR spacer(s) in the cell such that the CRISPR spacer(s) has homology to the CRISPR spacer(s) in the organism.
35. 35. The method according to any of paragraphs 32-34, wherein the one or more *cas* genes or proteins and/or the two or more CRISPR repeats are or are derivable (preferably, derived) from the same cell.
36. 36. The method according to any of paragraphs 32-35, wherein the spacers are or are derivable (preferably, derived) from an organism that is different to the cell comprising the one or more *cas* genes or proteins and/or the two or more CRISPR repeats.
37. 37. The method according to any of paragraphs 32-35, wherein the one or more *cas* genes or proteins and the two or more CRISPR repeats naturally co-occur in the same cell.
38. 38. The method according to any of paragraphs 32-37 wherein said modification comprises inserting one or more CRISPR spacers and/or pseudo CRISPR spacers into the cell.
39. 39. The method according to paragraph 38, wherein the spacer of the cell has 100% homology to the CRISPR spacer or pseudo CRISPR spacer of the organism.
40. 40. The method according to paragraph 38 or paragraph 39, wherein said modification comprises genetically engineering the CRISPR spacer of the cell.
41. 41. The method according to paragraph 40, wherein all or part of the spacer in the cell is modified.
42. 42. The method according to paragraph 40 or paragraph 41, wherein said modification comprises the modification of a recombinant spacer.

43. 43. The method according to any of paragraphs 40-42, wherein said modification occurs through spontaneous mutation or mutagenesis.
44. 44. A method for modulating (eg. reducing or decreasing) the resistance of a cell comprising at least one or more *cas* genes or proteins and two or more CRISPR repeats against a target nucleic acid or a transcription product thereof comprising the steps of:
1. (i) identifying one or more CRISPR spacers in an organism that is substantially resistant to the target nucleic acid or transcription product thereof; and
 2. (ii) modifying the sequence of at least one or more CRISPR spacer(s) in the cell such that the CRISPR spacer(s) has a reduced degree of homology to the spacer(s) in the organism.
45. 45. The method according to paragraph 44, wherein the at least one or more CRISPR spacer(s) in the cell are deleted.
46. 46. The method according to paragraph 44 or paragraph 45, wherein at least one or more CRISPR repeat(s) in the cell are deleted.
47. 47. The method according to any of paragraphs 44-46, wherein one or more *cas* genes are deleted,
48. 48. The method according to any of paragraphs 44-47, wherein CRISPR and/or one or more *cas* genes are deleted.
49. 49. A method for modulating (eg. reducing or decreasing) the resistance of a cell comprising at least one or more *cas* genes or proteins and two or more CRISPR repeats against a target nucleic acid or a transcription product thereof comprising modifying the one or more *cas* genes or proteins and/or two or more CRISPR repeats in the cell.
50. 50. The method according to paragraph 49, wherein the one or more *cas* genes or proteins and/or two or more CRISPR repeats in the cell are deleted.
51. 51. A method for identifying a CRISPR spacer or pseudo CRISPR spacer for use in modulating the resistance of a cell against a target nucleic acid or a transcription product thereof comprising the steps of:
1. (i) preparing a cell comprising at least two CRISPR repeats and at least one *cas* gene or protein;
 2. (ii) identifying at least one CRISPR spacer or pseudo CRISPR spacers in an organism that is substantially resistant to the target nucleic acid or transcription product thereof;
 3. (iii) modifying the sequence of the CRISPR spacer in the cell such that the CRISPR spacer has homology to the spacer of the organism; and
 4. (iv) determining if the cell modulates resistance against the target nucleic acid or transcription product thereof,
- wherein modulation of the resistance of the cell against the target nucleic acid or transcription product thereof is indicative that the CRISPR spacer modulates the resistance of the cell.
52. 52. A method for identifying a *cas* gene for use in modulating the resistance of a cell against a target nucleic acid or transcription product thereof comprising the steps of:
1. (i) preparing a cell comprising at least one CRISPR spacer and at least two CRISPR repeats;
 2. (ii) engineering the cell such that it comprises at least one *cas* gene; and

3. (iii) determining if the cell modulates resistance against the target nucleic acid or transcription product thereof,
wherein modulation of the resistance of the cell against the target nucleic acid or transcription product thereof is indicative that the *cas* gene can be used to modulate the resistance of the cell.
53. 53. A method for identifying a CRISPR repeat for use in modulating the resistance of a cell against a target nucleic acid or transcription product thereof comprising the steps of:
 1. (i) preparing a cell comprising at least one CRISPR spacer and at least one *cas* gene;
 2. (ii) engineering the cell such that it contains the CRISPR repeat; and
 3. (iii) determining if the cell modulates resistance against the target nucleic acid or transcription product thereof,
wherein modulation of the resistance of the cell against the target nucleic acid or transcription product thereof is indicative that the CRISPR repeat can be used to modulate resistance.
54. 54. A method for identifying a functional combination of a *cas* gene and a CRISPR repeat comprising the steps of:
 1. (a) determining the sequences of the *cas* gene and the CRISPR repeat;
 2. (b) identifying one or more clusters of *cas* genes as determined by sequence comparison analysis;
 3. (c) identifying one or more clusters of CRISPR repeats; and
 4. (d) combining those *cas* gene and CRISPR repeat sequences that fall within the same cluster,
wherein the combination of the *cas* gene and CRISPR repeat sequences within the same cluster is indicative that the combination is a functional combination.
55. 55. The method according to paragraph 54, wherein the nucleotide sequences of the *cas* gene and the CRISPR repeat are or are derivable (preferably, derived) from the same or different strains.
56. 56. The method according to paragraph 54, wherein the nucleotide sequences of the *cas* gene and the CRISPR repeat are or are derivable (preferably, derived) from the same or different species.
57. 57. The method according to paragraph 54, wherein the nucleotide sequences of the *cas* gene and the CRISPR repeat are or are derivable (preferably, derived) from the same or different genera.
58. 58. The method according to paragraph 54, wherein the nucleotide sequences of the *cas* gene and the CRISPR repeat are or are derivable (preferably, derived) from the same or different organisms.
59. 59. A method for modulating the lysotype of a bacterial cell comprising one or more *cas* genes or proteins and two or more CRISPR repeats comprising the steps of:
 1. (i) identifying one or more pseudo CRISPR spacers in the genomic sequence of a bacteriophage against which resistance is to be modulated; and
 2. (ii) modifying the sequence of one or more CRISPR spacers of the bacterial cell such that the CRISPR spacer(s) of the bacterial cell has homology to the pseudo CRISPR spacer(s) of the bacteriophage against which resistance is to be

modulated.

60. 60. A method for modulating (eg. conferring or increasing) the resistance of a bacterial cell against a bacteriophage comprising the steps of:
 1. (i) identifying a sequence (eg. a conserved sequence) in a bacteriophage (preferably, a sequence essential to the function or survival of the bacteriophage);
 2. (ii) preparing a CRISPR spacer which is homologous to the identified sequence;
 3. (iii) preparing a nucleic acid comprising at least one *cas* gene and at least two CRISPR repeats together with the CRISPR spacer; and
 4. (iv) introducing said nucleic acid into the bacterial cell thus to render the bacterial cell resistant to said target nucleic acid or transcription product thereof.
61. 61. A method for modulating (eg. conferring or increasing) the resistance of a bacterial cell against a target nucleic acid or transcription product in a bacteriophage thereof comprising the steps of:
 1. (i) identifying one or more pseudo CRISPR spacers in a bacteriophage genome that is capable of providing resistance to the target nucleic acid or transcription product thereof;
 2. (ii) preparing a recombinant nucleic acid comprising at least one *cas* gene and at least two CRISPR repeats together with said identified one or more pseudo CRISPR spacers; and
 3. (iii) introducing said recombinant nucleic acid into said bacterial cell thus to render the bacterial cell resistant to said target nucleic acid or transcription product thereof.
62. 62. A method for modulating the resistance of a bacterial cell comprising one or more *cas* genes or proteins and two or more CRISPR repeats against a target nucleic acid or transcription product thereof in a bacteriophage comprising the steps of:
 1. (i) identifying one or more pseudo CRISPR spacers in a bacteriophage that is capable of providing resistance to a target nucleic acid or transcription product thereof;
 2. (ii) identifying one or more CRISPR spacers in a bacterial cell in which resistance is to be modulated; and
 3. (iii) modifying the sequence of the CRISPR spacer(s) in the bacterial cell in which resistance is to be modulated such that the CRISPR spacer(s) has a higher degree of homology to the pseudo CRISPR spacer(s) of the bacteriophage against which resistance is to be modulated.
63. 63. The method according to paragraph 61 or paragraph 62, wherein the target nucleic acid in the bacteriophage is a highly conserved nucleic acid sequence.
64. 64. The method according to any of paragraphs 61-63, wherein the target nucleic acid in the bacteriophage encodes a host specificity protein.
65. 65. The method according to any of paragraphs 61-64, wherein the target nucleic acid in the bacteriophage encodes a protein that is essential for survival, replication or growth of the bacteriophage.
66. 66. The method according to paragraph 65, wherein the target nucleic acid in the bacteriophage encodes a helicase, a primase, a head or tail structural protein, a protein with a conserved domain (eg. holin, lysin, and others) or a conserved sequences

amongst important phage genes.

67. 67. A method for determining the resistance of a cell against a target nucleic acid or a transcription product thereof comprising identifying one or more functional CRISPR repeat-cas combinations and one or more CRISPR spacers in the cell.
68. 68. The method according to paragraph 67, comprising the additional step of comparing the sequence of the one or more CRISPR spacers in the cell with the sequence of the target nucleic acid .
69. 69. The method according to paragraph 67 or paragraph 68, comprising the additional step of determining the resistance profile of the cell.
70. 70. A cell obtained or obtainable by the method of any of paragraphs 32-69.
71. 71. A CRISPR spacer or pseudo CRISPR spacer obtained or obtainable by the method of paragraph 51.
72. 72. A cas gene obtained or obtainable by the method of paragraph 52.
73. 73. A CRISPR repeat obtained or obtainable by the method of paragraph 53.
74. 74. A functional combination obtained or obtainable by the method of any of paragraphs 54-58.
75. 75. A recombinant CRISPR locus comprising a CRISPR spacer or pseudo CRISPR spacer according to paragraph 71, and/or a cas gene according to paragraph 72, and/or a CRISPR repeat according to paragraph 73 and/or a functional combination according to paragraph 74.
76. 76. Use of a cell according to paragraph 70, a CRISPR spacer or pseudo CRISPR spacer according to paragraph 71, a cas gene according to paragraph 72, a CRISPR repeat according to paragraph 73 or a functional combination according to paragraph 74 for modulating the resistance of a cell against a target nucleic acid or a transcription product thereof.
77. 77. A cell culture comprising a cell according to paragraph 70, a CRISPR spacer or pseudo CRISPR spacer according to paragraph 71, a cas gene according to paragraph 72, a CRISPR repeat according to paragraph 73 or a functional combination according to paragraph 74 for modulating the resistance of a cell against a target nucleic acid or a transcription product thereof.
78. 78. The cell culture according to paragraph 77, wherein said culture is a starter culture or a probiotic culture.
79. 79. A food product or feed comprising the culture according to paragraph 77 or paragraph 78.
80. 80. A process for preparing a food product or feed comprising the use of the culture according to paragraph 77 or paragraph 78.
81. 81. A food product of feed obtained or obtainable by the process of paragraph 80.
82. 82. Use of the culture according to paragraph 77 or paragraph 78 for preparing a food product.
83. 83. A nucleotide sequence comprising or consisting of the sequence set forth in any of SEQ ID Nos. 7-10 and SEQ ID Nos. 359 to 405 or a variant, fragment, homologue or derivative thereof.
84. 84. An amino acid sequence encoded by the nucleotide sequence according to paragraph 83.

85. 85. A construct or vector comprising one or more of the nucleotide sequences according to paragraph 83.
86. 86. A host cell into which has been incorporated one or more of the nucleotide sequences according to paragraph 83 or the construct or vector according to paragraph 85.

SEQUENCE LISTING

[0676]

<110> DANSICO A/S

<120> USE

<130> 14923.0028 PCT

<140> PCT/US06/033167

<141> 2006-08-25

<150> 60/711,396

<151> 2005-08-26

<150> 60/747,683

<151> 2006-05-19

<160> 712

<170> PatentIn version 3.3

<210> 1

<211> 36

<212> DNA

<213> *Streptococcus thermophilus*

<400> 1

gttttgtac tctcaagatt taagtaactg tacaac 36

<210> 2

<211> 36

<212> DNA

<213> *Streptococcus thermophilus*

<400> 2

gttttgtat tctcaagatt taagtaactg tacagt 36

<210> 3
<211> 36
<212> DNA
<213> *Streptococcus thermophilus*

<400> 3
gtttttgtac tctcaagatt taagtaactg tacagt 36

<210> 4
<211> 36
<212> DNA
<213> *Streptococcus thermophilus*

<400> 4
gtttttgtac tctcaagatt taagtaaccg tacaac 36

<210> 5
<211> 36
<212> DNA
<213> *Streptococcus thermophilus*

<400> 5
gtttttgtac tctcaagatt taagtaactg tgcaac 36

<210> 6
<211> 36
<212> DNA
<213> *Streptococcus thermophilus*

<400> 6
gtttttgtac tctcaagatt taagtagctg tacagt 36

<210> 7
<211> 36
<212> DNA
<213> *Streptococcus thermophilus*

<400> 7
gtttttgtac tctcaagata taagtaactg tacaac 36

<210> 8
<211> 36
<212> DNA
<213> *Streptococcus thermophilus*

<400> 8
gtttttgtac tctcaagatc taagtaactg tacaac 36

<210> 9
 <211> 36
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 9
 gttttgtac tctcaagatg taagtaactg tacaac 36

 <210> 10
 <211> 36
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 10
 gtctttgtac tctcaagatt taagtaactg tacaac 36

 <210> 11
 <211> 36
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 11
 aaaaaagtcc cctctcgagg taattaggtt tatatc 36

 <210> 12
 <211> 36
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 12
 gttccgtcc cctctcgagg taattaggtt tatatc 36

 <210> 13
 <211> 36
 <212> DNA
 <213> *Streptococcus agalactiae*

 <400> 13
 gttttagagc tgtgtgttt cgaatggttc caaaac 36

 <210> 14
 <211> 36
 <212> DNA
 <213> *Streptococcus agalactiae*

 <400> 14
 gttttaagc tgtgctgta ttatgctagg gcacca 36

<210> 15
 <211> 36
 <212> DNA
 <213> *Streptococcus agalactiae*

 <400> 15
 gttttagagc tgtgctgttt cgaatgggtc caaaac 36

 <210> 16
 <211> 36
 <212> DNA
 <213> *Streptococcus agalactiae*

 <400> 16
 gttttagagc tgtgctgtta ttatgctagg acatca 36

 <210> 17
 <211> 36
 <212> DNA
 <213> *Streptococcus mutans*

 <400> 17
 gttttagagc catgtaggt actgatttac taaaat 36

 <210> 18
 <211> 36
 <212> DNA
 <213> *Streptococcus pyogenes*

 <400> 18
 gttttagagc tatgctgttt tgaatgggtc caaaac 36

 <210> 19
 <211> 36
 <212> DNA
 <213> *Streptococcus pyogenes*

 <400> 19
 gttttagagc tatgctgttt tgaatgggtc ccattc 36

 <210> 20
 <211> 33
 <212> DNA
 <213> *Streptococcus pyogenes*

 <400> 20
 ctttcaatcc actcaccat gaagggtag acg 33

<210> 21
 <211> 33
 <212> DNA
 <213> Streptococcus pyogenes

 <400> 21
 atttcaatcc actcacccat gaagggtag act 33

 <210> 22
 <211> 33
 <212> DNA
 <213> Streptococcus pyogenes

 <400> 22
 atttcaatcc actcacccat gaagggtag acc 33

 <210> 23
 <211> 30
 <212> DNA
 <213> streptococcus thermophilus

 <400> 23
 agaacgtatt ccaaaacctc ttacgatta 30

 <210> 24
 <211> 30
 <212> DNA
 <213> streptococcus thermophilus

 <400> 24
 ttaactgtta tcaaatgat aagatagtct 30

 <210> 25
 <211> 30
 <212> DNA
 <213> Streptococcus thermophilus

 <400> 25
 cgtagtggt tattcaagta aaataattaa 30

 <210> 26
 <211> 30
 <212> DNA
 <213> Streptococcus thermophilus

 <400> 26
 tccttcacg ggtagcacac taacatacac 30

<210> 27
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 27
gttggaatg caaacaacct ttatgaaccg 30

<210> 28
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 28
tttatttct tgcgataacg ttccacctt 30

<210> 29
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 29
agattataag gaacacaacc aactatatag 30

<210> 30
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 30
acgacatcaa gctgattgtc ttctacataa 30

<210> 31
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 31
tttgaatac tgaatgttt actgaaaatc 30

<210> 32
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 32
acaccactat cttttctcc tgaaaatgaa 30

<210> 33
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 33
gtaattccac gaaattatca accttatgca 30

<210> 34
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 34
ttggaggatt gcccattatt cccaagagt 29

<210> 35
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 35
gagaggcggtt aaatatagaa atgcaagatt 30

<210> 36
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 36
ttttaacgtc atcagtccac cgccttaaat 30

<210> 37
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 37
cacctctttc gatggaaagg tacccttcta 30

<210> 38
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 38
gaccaaagtt tgattataga gctatacacc 30

<210> 39
 <211> 30
 <212> DNA
 <213> Streptococcus thermophilus

 <400> 39
 accatcattc ttaccattac aactgtaatg 30

 <210> 40
 <211> 30
 <212> DNA
 <213> streptococcus thermophilus

 <400> 40
 atacgaattc ggctgcaca attacaattc 30

 <210> 41
 <211> 32
 <212> DNA
 <213> Streptococcus thermophilus

 <400> 41
 tatcaacgca atcattacaa caactcaaa ca 32

 <210> 42
 <211> 30
 <212> DNA
 <213> Streptococcus thermophilus

 <400> 42
 atctacgtgt caatacatat caaaaacag 30

 <210> 43
 <211> 29
 <212> DNA
 <213> streptococcus thermophilus

 <400> 43
 atttttagaa atttctgata taataatga 29

 <210> 44
 <211> 30
 <212> DNA
 <213> Streptococcus thermophilus

 <400> 44
 ttgttgaac aaggacgact tggtaaacta 30

<210> 45
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 45
catattaagc tgactgggcc taatgctttt 30

<210> 46
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 46
ttcatagcat accgtagttg taaaatctat 30

<210> 47
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 47
aacatttagg gaatgaaatt gataagactg 30

<210> 48
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 48
aacatgagaa actgtagaaa acaagcaata 30

<210> 49
<211> 31
<212> DNA
<213> *Streptococcus thermophilus*

<400> 49
tggtaagat ggcagtcata aatggcacat t 31

<210> 50
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 50
aagggttgaa aaatgttgtt atatcaaacg 30

<210> 51
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 51
ttctggtagt ggatttagtc aaacagatgt 30

<210> 52
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 52
tccatagagc gtcttaaaca aagaatagtc 30

<210> 53
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 53
ttatgattga atgacatggt tgtataagta 30

<210> 54
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 54
tttcttagg aataccaggg agttcagctt 30

<210> 55
<211> 30
<212> DNA
<213> streptococcus thermophilus

<400> 55
tggcagagat tacacagcaa cggaaacagc 30

<210> 56
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 56
gggtatcatt gtatctagtg atggacctga 30

<210> 57
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 57
atttgaaaaa tgcacaacag cgtttgatag 30

<210> 58
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 58
gagctaccag ctaccccgta tgtcagagag 30

<210> 59
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 59
cgttcctttt ttcaaggtaa tctttgaaag 30

<210> 60
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 60
aagtccgtaa gcaccagttc caatcgtcat 30

<210> 61
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 61
ttgaatacca atgccagctt cttttaaggc 30

<210> 62
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 62
aacctcatat atggggaaaa ttgtaagta 30

<210> 63
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 63
 taacttcatt agtgtagtgtg taattagcat 30

 <210> 64
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 64
 ttagctaccc aaatatcttc tgttttccaa 30

 <210> 65
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 65
 gagttttcaa tattggcaca ggagacaatt 30

 <210> 66
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 66
 tgatactatt ttagtcagat atgaaatac 30

 <210> 67
 <211> 31
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 67
 tcatcaatgt ttaaagccca acaatacatg a 31

 <210> 68
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 68
 tagatttaat cagtaatgag ttaggcataa 30

<210> 69
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 69
aggaaaatag catgagcgta caacaatcta 30

<210> 70
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 70
tgtctatcac gcttctaag tgcataaaaa 30

<210> 71
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 71
atgtcaccaa tcactaaaga acctacgctg 30

<210> 72
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 72
aacatcttcc tctccgattg caaatagtcg 30

<210> 73
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 73
catatttggg gcccggtcga taaagagta 29

<210> 74
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 74
cattaaatcg ctgaagcag acattgaagc 30

<210> 75
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 75
gacttatctt ggaaggtagt gaaggcactt 30

<210> 76
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 76
tccttgccat ctgcactgta agccaagca 30

<210> 77
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 77
tagtacgcat aatcaattca tcaagcttga 30

<210> 78
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 78
gtagtgaccc aaaattctat gaccttgaaa 30

<210> 79
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 79
agattgtggt gcttacggaa aattccttgt 30

<210> 80
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 80
tggcaagaag tgtaagagat gcaatggata 30

<210> 81
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 81
tttattatca ttattcttct tccaagcgt 30

<210> 82
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 82
ttttatagaa ttggtggtg aacttttca 30

<210> 83
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 83
aatgggtcac agattgcat aataaggag 29

<210> 84
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 84
ccgaggtcac tttagaacc aaaaataag 30

<210> 85
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 85
atgagagaac acagtataga ccctgataca 30

<210> 86
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 86
cagtattaat gaggttggg tggcattcc 30

<210> 87
<211> 31
<212> DNA
<213> *Streptococcus thermophilus*

<400> 87
ccatactctc tatcagttca ttaattctt c 31

<210> 88
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 88
taatatgtcg ctctactgat tccaaaacgg 30

<210> 89
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 89
atgaattaca ttcattgattt tatcgagttt 30

<210> 90
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 90
cgtgccattg ttcggtcgg acgtgggca 29

<210> 91
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 91
ctttctaagt tgaattaaat tcaagtttg 30

<210> 92
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 92
tcgctactat ggtaacgat gaggaactct 30

<210> 93
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 93
agcaacttta aaactaaaag agctacttga 30

<210> 94
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 94
aaaaccctac acagtgtgtg agatgtgtca 30

<210> 95
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 95
aatgggtcac agattgcat aataaggagg 30

<210> 96
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 96
tttttaaaa tccgtcatgc tatactatat 30

<210> 97
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 97
aattcaaact ttctccaata ataccctcca 30

<210> 98
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 98
catgcttca gttaataaga cgtgggacta 30

<210> 99
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 99
tggaaggggt gtctagtga gaaattgtcg 30

<210> 100
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 100
ctcgaagcgc ttcatgccc tattcctttc 30

<210> 101
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 101
atgtctaagg tatccactcg tgaaatcat 29

<210> 102
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 102
atattaatgg aaatttcatt caaacgcagt 30

<210> 103
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 103
tagagagttt atatcctgat ggaatcgatg 30

<210> 104
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 104
tggcgaatta gagagccaat ggcaagcaag 30

<210> 105
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 105
agaagaccaa taaactgag aaaaagcaag 30

<210> 106
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 106
aaatggtcgt ttaattgta atgtcaaagc 30

<210> 107
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 107
caattgattc taaaatgctt ggtacacgta 30

<210> 108
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 108
tcttcgtgtt atcacagctt ctacacgttg 30

<210> 109
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 109
gaaatctcat tgaaaccaac ttcaagacca 30

<210> 110
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 110
tgcttggtag ttgatgcact gcattagtaa 30

<210> 111
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 111
aatgtaccgg aatagcgta cattgcacat 30

<210> 112
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 112
ttcataaatt ctacttttc ctgctattc 30

<210> 113
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 113
tgtcgaaaaa attacctagt cacgacagac 30

<210> 114
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 114
caacaattac ttatgcatta ggaacatctg 30

<210> 115
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 115
aattcgtgaa aaacaataaa aacaaaaaaaa 30

<210> 116
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 116
taacatttct gtccatttct tccttgatgc 30

<210> 117
<211> 28
<212> DNA
<213> *Streptococcus thermophilus*

<400> 117
caaggcaact caaccaacca aattgacc 28

<210> 118
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 118
ctaaaatcgt aaatggtaag ttgcacgatg 30

<210> 119
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 119
aacgtaagga gtttttttat ttctttgtta 30

<210> 120
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 120
gtggaaaatt tcacacccta catatatcaa 30

<210> 121
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 121
cctctgctaa tgacttaaac ggctcgttt 30

<210> 122
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 122
aaaatcaaag ttttgggttt gtctacgttg 30

<210> 123
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 123
atatgtacat acctaaagaa aacacgggca 30

<210> 124
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 124
cgttgtcaaa atatgtgatt actttgtatt 30

<210> 125
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 125
ccatagctgt aatgttggtt gtgactgctt 30

<210> 126
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 126
cgctaagttt ggctttaagt ataacaagct 30

<210> 127
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 127
aaagtacgct tcaaggcacg ttgaagacat 30

<210> 128
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 128
cttttaacg tgtagcgctc ttagctttg 30

<210> 129
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 129
ttggcttcgt gaataat ttt taaaacgcat 30

<210> 130
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 130
tgttgaatca atacgctgaa acacactccc 30

<210> 131
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 131
cgttatcagt tgaaagtttc aactcgtaag 30

<210> 132
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 132
taaactagtt ggcatctatg ctccaggaag 30

<210> 133
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 133
tagaccacca tagccgagtt gtcttttcg 30

<210> 134
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 134
acatcccact ttctgggttt ttagccatg 30

<210> 135
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 135
agtatggcta ttgcctgat actcatccac 30

<210> 136
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 136
cgctcttgac gtggctggcg acatctacgc 30

<210> 137
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 137
gagtacatgg agtttctgct agatacacta 30

<210> 138
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 138
taagtatatga aatataaagt tattgtcta 29

<210> 139
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 139
aacgttatga catttaggag ctccaaatt 30

<210> 140
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 140
aacacagcaa gacaaaagga tgacacttt 29

<210> 141
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 141
caaccataac ttacgcatca ggtacatctg 30

<210> 142
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 142
acacgcgctt acctcgata tcaaattca 29

<210> 143
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 143
tgcccgcaaa ctacgatac acaacagcat 30

<210> 144
<211> 31
<212> DNA
<213> *Streptococcus thermophilus*

<400> 144
ctcaagctct tcatctgtga taggtgttt g 31

<210> 145
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 145
atcactctt gatagtatct caaacgctgg 30

<210> 146
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 146
gaaacagtca gaccagctaa ttcgccaatt 30

<210> 147
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 147
 atatttcgaa agatacaagg acacttacac 30

 <210> 148
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 148
 gcggatgaaa cacaacttca attgtattca 30

 <210> 149
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 149
 taatgctaca tctcaaagga tgatcccaga 30

 <210> 150
 <211> 31
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 150
 acgtctgtct aactggaaag tacctgctaa t 31

 <210> 151
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 151
 ctgttctcta atcgagaggc gcgtgattga 30

 <210> 152
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 152
 aaacctcact agtcacttag tgcggtagg 30

<210> 153
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 153
tattaagttt agtcccaggt ttcttatcgt 30

<210> 154
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 154
aaaccaataa acataccgat tgctgccaat 30

<210> 155
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 155
gcaaacgtta gccaggaaa gcatcatgaa 30

<210> 156
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 156
aagagcaaaa aataactcta gctctcgtcc 30

<210> 157
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 157
aagaaacctc taagtgagc atttaatgat 30

<210> 158
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 158
atatagtttt aaactttctt gaccttctg 29

<210> 159
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 159
acgttgatga atattgtga taaacttta 29

<210> 160
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 160
caagaagtga acaaagtaca cgctggaagt 30

<210> 161
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 161
gacagcaaga tacacgtagt tgatgaattg 30

<210> 162
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 162
taagaaatca acgcagattt ttagccaaca 30

<210> 163
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 163
taaccaata attacagtga agcacaatag 30

<210> 164
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 164
caggcgtaag gtatgctaata tataacgat 29

<210> 165
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 165
gctatcgaac taatagctta gaggaactca 30

<210> 166
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 166
gtggaatatt aagcccgaat tgtgcagca 30

<210> 167
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 167
tattgcaata ttgcgttg ggaaaccttc 30

<210> 168
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 168
cgtctgtcta actggaaagt accggctaata 30

<210> 169
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 169
aaagagatgt acccatccat tctaacaggt 30

<210> 170
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 170
ggggagttga ttcttacat caaaacaatg 30

<210> 171
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 171
catcaaagtt gaaaaggact acaacagccc 30

<210> 172
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 172
cttaaattta gagcgtggga tctgaatat 30

<210> 173
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 173
ataaccgat ggcacatctg aaactggctg 30

<210> 174
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 174
taactcatat gtatcttgac caactatfff 30

<210> 175
<211> 30
<212> DNA
<213> streptococcus thermophilus

<400> 175
aaatagcacc tctaagcgtt aatggtattc 30

<210> 176
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 176
aatatctaca ggtcactaca aagctacgct 30

<210> 177
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 177
gttgggggtgt gtttgtaacg gcgtatgcta 30

<210> 178
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 178
tcaatcaggt gacggtgatg cttatattaa 30

<210> 179
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 179
catacatgat agtttgtaa cacttttgat 30

<210> 180
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 180
tcagcatttg gtttacctga cccacgtctg 30

<210> 181
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 181
caatcaacag gtttgactga ttataacggt 30

<210> 182
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 182
tagctacaca tgaattttat tacaatgggtg 30

<210> 183
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 183
cttacgtttg aaaagaatat caaatcaatg 30

<210> 184
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 184
ttaaaaaagg gcctttctct aaatcaagta 30

<210> 185
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 185
tgtgaacgt atctgtccac tgtgtggcca 30

<210> 186
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 186
ccgttctca aacgttaaattccaaggtgt 30

<210> 187
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 187
gctgcgatta tgacaatgct gtctgtaagg 30

<210> 188
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 188
gaagaattta ttaataaaga tgggtctgct 30

<210> 189
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 189
aggcagaaaa gaagtatttt ggtaagtatg 30

<210> 190
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 190
aaatggttta tcgacaagaa aatgaagct 29

<210> 191
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 191
ccaaatttgc attatacaaa acgctccttc 30

<210> 192
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 192
atcctaactg ctttgctaac tacatcatgg 30

<210> 193
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 193
taacaagata agattagcgt ctcaacat 29

<210> 194
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 194
aaaagcctat gttgcccac tttgtggaag 30

<210> 195
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 195
 tgcactttc tcttctggg ttgtgcaat 30

 <210> 196
 <211> 31
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 196
 catactttc catctgttg ttgttgaaa a 31

 <210> 197
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 197
 tgagagtgc tgatggatt attggcagcc 30

 <210> 198
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 198
 ggggttatt tccatttac cgtctatcta 30

 <210> 199
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 199
 tatcacgcc atttcattt cgccatctgt 30

 <210> 200
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 200
 aacattttaa tataatttct aaatctattg 30

<210> 201
 <211> 30
 <212> DNA
 <213> Streptococcus thermophilus

 <400> 201
 tacaaaattc cttcaaacgc tatttattga 30

 <210> 202
 <211> 29
 <212> DNA
 <213> Streptococcus thermophilus

 <400> 202
 agagtttgaa aattattttt cagtttcta 29

 <210> 203
 <211> 30
 <212> DNA
 <213> Streptococcus thermophilus

 <400> 203
 ttctcatct ttctccgctt ttgctagctt 30

 <210> 204
 <211> 31
 <212> DNA
 <213> Streptococcus thermophilus

 <400> 204
 ttgagcgttc tagtgtgtgg cttgtaatga a 31

 <210> 205
 <211> 30
 <212> DNA
 <213> streptococcus thermophilus

 <400> 205
 tgaaagaaat acaatacaac gataatgacc 30

 <210> 206
 <211> 30
 <212> DNA
 <213> streptococcus thermophilus

 <400> 206
 ctagttttaa gagatagctc tctaagtagg 30

<210> 207
<211> 29
<212> DNA
<213> Streptococcus thermophilus

<400> 207
aaattcgaca taagcactac agttatatt 29

<210> 208
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 208
ctattttcga gagaacgtca gtcattttaa 30

<210> 209
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 209
gtgctaacta tatcagtcgc atcaataaca 30

<210> 210
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 210
ttagcgggtga ttggaataga ataagcgaat 30

<210> 211
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 211
cttctacagc agttaagac acattatcat 30

<210> 212
<211> 29
<212> DNA
<213> Streptococcus thermophilus

<400> 212
cgtatcgaaa acggcgataa tccaacagt 29

<210> 213
 <211> 31
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 213
 caataccttt ttttaattca tcttgataag t 31

 <210> 214
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 214
 ttaagaacaa tatcatcaat acgactttca 30

 <210> 215
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 215
 catctatcaa attcaaattc ggataaacta 30

 <210> 216
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 216
 tgagagtgtc tgatggattt attggaacc 30

 <210> 217
 <211> 29
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 217
 acctcataca tggggaaaac ttgtaagta 29

 <210> 218
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 218
 tatttcacga atttctacac tttcaacct 30

<210> 219
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 219
 ctgaaacctt gtttgaagc gcttgaagt 30

 <210> 220
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 220
 gtcaattgat actgcaatct ctttaacatt 30

 <210> 221
 <211> 31
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 221
 acttcaatat ggtcaacatc ttgatcaccg a 31

 <210> 222
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 222
 taaactcgac aaaagcacta catgaatatt 30

 <210> 223
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 223
 attttttaag gaaaggagga aaataatata 30

 <210> 224
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 224
 cgttcaaaac agcgaaaact taaccctaac 30

<210> 225
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 225
cattaagtcg cttgaggcag acattgaagc 30

<210> 226
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 226
ccaaactcaa attgtctata ataataaccg 30

<210> 227
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 227
tatctctatt tcaggtgggt taaaacattc 30

<210> 228
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 228
aaacgaagat ggaagcgttg atgtttattc 30

<210> 229
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 229
gattgcattt gccagtattt cttttgatta 30

<210> 230
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 230
tgaagacaac ggaaacaatc aacctatta 29

<210> 231
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 231
acttctttt taatgtcatc taagacaata 30

<210> 232
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 232
gccaatgatg ttcaattcgt taatggaatt 30

<210> 233
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 233
tcaacatggg atatttcgtt ggtcaggatg 30

<210> 234
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 234
tatggctctc ttgttgaat aaagatgatt 30

<210> 235
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 235
ataacatagc agtctatttc ttgctgatg 30

<210> 236
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 236
gttaccacgc gccctactgt attagtggag 30

<210> 237
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 237
 tacataccca aggttgtaag tcgttaaatt 30

 <210> 238
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 238
 tgaagtagt caatattcac ttctgataac 30

 <210> 239
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 239
 gatagcaata gctttctga cctaaaagac 30

 <210> 240
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 240
 gaggtctgta attcattcc ctgtaatct 30

 <210> 241
 <211> 29
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 241
 aaaggttct ctaaacacat gcggaatat 29

 <210> 242
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 242
 gtcatagtac caagcacaaa taacgtagt 30

<210> 243
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 243
gtgtatttag taatggtgat ttttaaatt 30

<210> 244
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 244
cattcatttt ttatatatca ataaaacttt 30

<210> 245
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 245
ggggattctt atttcactgt agttacgatg 30

<210> 246
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 246
caaaaattga tgcacaatt aataaagggtg 30

<210> 247
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 247
ctatttctga caatggttga aattgtgttc 30

<210> 248
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 248
cttttttaa attaatttat cgtaagcaa 29

<210> 249
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 249
aacaaactta tgagaacggt tgaacggctt 30

<210> 250
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 250
agcccgctta ttgcttcagt tggtttatat 30

<210> 251
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 251
tgagcaaca agaatgatta actctaatagc 30

<210> 252
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 252
ttgatggat atcattgata aactatacga 30

<210> 253
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 253
taacgaaagc aataccaatc gtgctaaagc 30

<210> 254
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 254
tattcctatg gtcgatattc gaacagtcaa 30

<210> 255
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 255
caggggacaa ggacttgac ccaacagaag 30

<210> 256
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 256
agaaacacct aatggtctct tagaaccga 30

<210> 257
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 257
aagaagttaa agacaacttt gttaaagact 30

<210> 258
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 258
gaaaaagcat ccatgatagt gcttagacct 30

<210> 259
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 259
cggaatggta taaagaatac aaagaaaacg 30

<210> 260
<211> 29
<212> DNA
<213> Streptococcus thermophilus

<400> 260
ccaagtatca cgcaaagaaa tcaacgaga 29

<210> 261
<211> 31
<212> DNA
<213> *Streptococcus thermophilus*

<400> 261
ttgacctgtt ttccttgtt aactagaata g 31

<210> 262
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 262
agagcactag catactgttt agtccgaacg 30

<210> 263
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 263
aggcaaggta ttgatccaa cagaagccaa 30

<210> 264
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 264
catgatttac aaccacgcgc tagaccaag 29

<210> 265
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 265
acctagaagc attgagcgt atattgattg 30

<210> 266
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 266
aattttgccc cttctttgcc ccttgactag 30

<210> 267
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 267
 taatagtta ccaaatcgtc cttgttcaa 30

 <210> 268
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 268
 accattagca atcattgtg cccattgagt 30

 <210> 269
 <211> 31
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 269
 acgtctgtct aactggaaag tacctgttaa t 31

 <210> 270
 <211> 29
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 270
 ttttatact ttgggtaatt acaaaatag 29

 <210> 271
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 271
 aagaaagaaa tattctagat atagatataa 30

 <210> 272
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 272
 caacgaccaa cacaacaact aaagttactg 30

<210> 273
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 273
tgattatggg tgttaaaca ggagcttatg 30

<210> 274
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 274
tgagtggtaa gtacaaatac gcaggactga 30

<210> 275
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 275
ttatttcctc ctttcctaa aaaaattaga 30

<210> 276
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 276
ggatgtatct gttgaaagag gtgtgtatat 30

<210> 277
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 277
aataggtgaa aaatatgcaa gtcacacaaa 30

<210> 278
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 278
aaaatggcat taaaaattaa cataggaata 30

<210> 279
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 279
tatcagctcg taaatgttcg atagactctt 30

<210> 280
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 280
attccattaa cgtatttgac ttcactagct 30

<210> 281
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 281
ctgttaccga tccaagagca gacatcatac 30

<210> 282
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 282
aagaagcggg taaatgcttc aactgaatag 30

<210> 283
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 283
aattgctaaa catctaaaag acttaacggg 30

<210> 284
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 284
gatgaagatt tgactgatga taaagagaaa 30

<210> 285
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 285
gacatcagaa agcagtttat aaatatttta 30

<210> 286
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 286
tttgaattta acaaccttga tttgatatc 30

<210> 287
<211> 31
<212> DNA
<213> Streptococcus thermophilus

<400> 287
tgatacggtc aaagtttttc cactaatagc g 31

<210> 288
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 288
atggttttca tttctgaac ccctaagagg 30

<210> 289
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 289
aagttattga aaaacgcca catgatgagt 30

<210> 290
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 290
atataagtcc tcctattaat atccacaata 30

<210> 291
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 291
ttgcctcaag agatcctgct tgttgccaag 30

<210> 292
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 292
tcccatagtt ttaatgagtc ggttaactta 30

<210> 293
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 293
gtgtactaaa agtgtgctaa gttcataagg 30

<210> 294
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 294
atatagtgat tgtatccagc tgcggcgtag 30

<210> 295
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 295
aaaagcaaata cgcgagtata aaggatata 29

<210> 296
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 296
ttttaattga tctagacacc ctatgaaata 30

<210> 297
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 297
acagaggaga gaaaccatgg ctattttaga 30

<210> 298
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 298
tggcagcagt gaattcgatg ccgagcaat 29

<210> 299
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 299
ccaaggaata ccaggtccta aagggtgccga 30

<210> 300
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 300
ctaaatgaac tacaacaaca gcttgatga 29

<210> 301
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 301
tacctaaca ttttcgatat ttttcaaatt 30

<210> 302
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 302
tttgactgct ttttatctg aattgtaatt 30

<210> 303
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 303
cagtaaccta aagctctatc aagcctattt 30

<210> 304
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 304
cgtcaagctg acagaccttg acaacaatc 30

<210> 305
<211> 29
<212> DNA
<213> Streptococcus thermophilus

<400> 305
aggcataaat aacattgata accctaaca 29

<210> 306
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 306
gccaacgagg tcaaataatgt caacggcatt 30

<210> 307
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 307
gaaataggaa ctcaaaggt aatttcttta 30

<210> 308
<211> 30
<212> DNA
<213> streptococcus thermophilus

<400> 308
atttagagca aggaaagcag tacatcatta 30

<210> 309
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 309
ctgtaatcat ttttaaatca ggattatcaa 30

<210> 310
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 310
ttaaattgat cctagtattt ttgtactata 30

<210> 311
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 311
ccatcagcca actgtatcgg ctactttcta 30

<210> 312
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 312
atgctcttgg cgactatctc atggagcgtg 30

<210> 313
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 313
aggaaaaaac ccaaacaacc caaaatgtta 30

<210> 314
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 314
tctaattctg tcaccacgac tatatcgcca 30

<210> 315
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 315
aatctgtgtg ggaagtaaag attgaagatg 30

<210> 316
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 316
atagtttggt aagtcatacc cattaaattg 30

<210> 317
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 317
tccacatgat taaaagcca cgcaagacct 30

<210> 318
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 318
gaagacaaa attgacaat gagtcctgc 29

<210> 319
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 319
attatattta agttgtaaat gttgcttttc 30

<210> 320
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 320
gcagacattg gctcaacaag tgattatgaa 30

<210> 321
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 321
tgttctcata aatgccttt ccttttatg 30

<210> 322
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 322
cttatcaaac atcaaggatt gtagatgagg 30

<210> 323
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 323
atttcattag tagcttgata aatgtttcta 30

<210> 324
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 324
gaaaatacta tactttaaaa gaaatttaa 30

<210> 325
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 325
tctcctccga cataatcttt tgtctttccg 30

<210> 326
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 326
acaaaagcac tgccacctat agaagcattt 30

<210> 327
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 327
aaaaacttta tgctatccgt gtcagtatat 30

<210> 328
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 328
tttcaatga ttgaaagccc ataactaaca 30

<210> 329
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 329
ctttcatagt tgttacgaaa tgtttggcat 30

<210> 330
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 330
cgatttgcaa tatgatgata ttgatgaatt 30

<210> 331
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 331
ttagatgct agtcctaaga ctgtagagac 30

<210> 332
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 332
gtaatcaagc gtatataagt caggactatc 30

<210> 333
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 333
ataacagaag gagtagggga cgtaggcgcg 30

<210> 334
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 334
ttatttgata ggaatgtcag taattttga 30

<210> 335
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 335
aacatttcag cgcttactta tcaatctaata 30

<210> 336
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 336
gtattagtag gcatacgatt atggaagta 29

<210> 337
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 337
catatatata tatatattta ttttaaata 30

<210> 338
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 338
ttgtcataat aattaaatcc aataggactt 30

<210> 339
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 339
gaaaatttct gttgtgttct taatattagc 30

<210> 340
<211> 30
<212> DNA
<213> *Streptococcus vestibularis*

<400> 340
gtacttcaaa ggttctaact acataacaca 30

<210> 341
<211> 30
<212> DNA
<213> *Streptococcus vestibularis*

<400> 341
taaaaccaga tgggtgttct tctgatacta 30

<210> 342
<211> 30
<212> DNA
<213> *Streptococcus vestibularis*

<400> 342
cattttcttc agtcaattcg ttctcaagcg 30

<210> 343
<211> 30
<212> DNA
<213> *Streptococcus vestibularis*

<400> 343
aaaggacggg ggcaatgaac aaacgacaac 30

<210> 344
<211> 29
<212> DNA
<213> *Streptococcus vestibularis*

<400> 344
taatatcatt gatagcttca tcaaaggct 29

<210> 345
<211> 29
<212> DNA
<213> Streptococcus vestibularis

<400> 345
taaattgttc cttgactccg aactgccct 29

<210> 346
<211> 30
<212> DNA
<213> Streptococcus vestibularis

<400> 346
aaacaatcgt ttatctatcc tcaaaggatg 30

<210> 347
<211> 30
<212> DNA
<213> Streptococcus vestibularis

<400> 347
ataaaaaaac gcctcaaaaa ccgagacaac 30

<210> 348
<211> 30
<212> DNA
<213> Streptococcus vestibularis

<400> 348
tggaaatccc ttatatcgac aaatacgta 30

<210> 349
<211> 30
<212> DNA
<213> Streptococcus vestibularis

<400> 349
ttccagtcg ttgatttta ttgaataccc 30

<210> 350
<211> 30
<212> DNA
<213> Streptococcus vestibularis

<400> 350
ggacatcgaa caagtcaatg ccgtaagctt 30

<210> 351
 <211> 30
 <212> DNA
 <213> Streptococcus vestibularis

 <400> 351
 aatctttaac cggattgtag aaccgttcgg 30

 <210> 352
 <211> 30
 <212> DNA
 <213> Streptococcus vestibularis

 <400> 352
 tgcctttaaa ataactagat ttaccatca 30

 <210> 353
 <211> 29
 <212> DNA
 <213> Streptococcus vestibularis

 <400> 353
 gagcaagcac aagcaagctt tactatcct 29

 <210> 354
 <211> 30
 <212> DNA
 <213> Streptococcus vestibularis

 <400> 354
 cagattgggt tatcgaaca ggtcgcaagt 30

 <210> 355
 <211> 31
 <212> DNA
 <213> Streptococcus vestibularis

 <400> 355
 caaaagctgt tggtaacgg tgcttgggc a 31

 <210> 356
 <211> 30
 <212> DNA
 <213> Streptococcus vestibularis

 <400> 356
 ctgtttttc ctctggggtc tctgcgact 30

<210> 357
<211> 30
<212> DNA
<213> Streptococcus vestibularis

<400> 357
gaaataaact gcccaaacat tttattttc 30

<210> 358
<211> 30
<212> DNA
<213> Streptococcus vestibularis

<400> 358
tgagtaagcg acaagctaga aatcaagtca 30

<210> 359
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 359
atagctaaga tggaagaagc atcaagcacc 30

<210> 360
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 360
cagtatttca aacgctggat acaacaagat 30

<210> 361
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 361
cctactcagt ggacacctgc aattgaagac 30

<210> 362
<211> 29
<212> DNA
<213> Streptococcus thermophilus

<400> 362
cgattggaac gggtgcttat ggcctaac 29

<210> 363
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 363
 gcgaacaatt gaattgtta gaaaatgtcg 30

 <210> 364
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 364
 gaagcattta ttaatataga tgggtctgct 30

 <210> 365
 <211> 28
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 365
 tgcgacgta tctgtccact gtgtgcca 28

 <210> 366
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 366
 ttttatact ttgggtaaat tacaaaatag 30

 <210> 367
 <211> 31
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 367
 tcaagggtgc gccttatgga aaagatgctt g 31

 <210> 368
 <211> 31
 <212> DNA
 <213> *Streptococcus thermophilus*

 <400> 368
 tgtaaaaatt tctagacgtt tagacacttt a 31

<210> 369
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 369
aaatgatgat tgaatgcttg agatagcagt 30

<210> 370
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 370
aataagaagt tcttgacgac caaccgacat 30

<210> 371
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 371
tcgtcaacgt cgatacagaa caacgtgctt 30

<210> 372
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 372
tgattagcaa atttaaaaca ggatatttgg 30

<210> 373
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 373
aaagacaagc ccaagggatt gaactagcaa 30

<210> 374
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 374
cgaacagttg gcgagaaatc cgtctggcgt 30

<210> 375
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

<400> 375
 ctacattatt gatcatgttt ttctcctgt 30

<210> 376
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

<400> 376
 tagaaggctc tggaaataca aagcaattct 30

<210> 377
 <211> 31
 <212> DNA
 <213> *Streptococcus thermophilus*

<400> 377
 tagaaggctc tggtaaatac aaagcaattc t 31

<210> 378
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

<400> 378
 tctgatggct cttggtaggg aactggatat 30

<210> 379
 <211> 30
 <212> DNA
 <213> *Streptococcus thermophilus*

<400> 379
 ttgatggct cttggtaggg aactggatat 30

<210> 380
 <211> 31
 <212> DNA
 <213> *Streptococcus thermophilus*

<400> 380
 tttgatggc tcttgtagg gaactggata t 31

<210> 381
<211> 29
<212> DNA
<213> Streptococcus thermophilus

<400> 381
acagaacaaa atggtagaat atatcatct 29

<210> 382
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 382
ccctggacaa gctatcagca catatccttg 30

<210> 383
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 383
cgctgttgat gtaaccgct ttatatatat 30

<210> 384
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 384
gaatgaatgt attagagcaa gcactgacc 30

<210> 385
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 385
tagacgaaaa ggaaggaaaa tagcatgagc 30

<210> 386
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 386
ataactcgat tgctaactta agcaagcagt 30

<210> 387
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 387
ctgcatgtgt aaccatgact tcttcgtcgt 30

<210> 388
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 388
cttcgctgga aacttcgtag tcatacatatc 30

<210> 389
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 389
aagaccgctg tactggttgg tattcgtacc 30

<210> 390
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 390
caaccaagcg aacacagcag tagcaccgca 30

<210> 391
<211> 29
<212> DNA
<213> *Streptococcus thermophilus*

<400> 391
atgatgatga agtatcgtca tctactaac 29

<210> 392
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 392
cttcacctca aatcttagag atggactaaa 30

<210> 393
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 393
aaaaggtgcg tatgaaactc atcccagcgg 30

<210> 394
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 394
aagggtttaa gtcctcata gagtggaaaa 30

<210> 395
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 395
cctcaaagct taaaattggg ctgaagtaga 30

<210> 396
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 396
gcaatttatt cgcttgatgt actcacgttt 30

<210> 397
<211> 30
<212> DNA
<213> streptococcus thermophilus

<400> 397
tatttattgc aaatggttac catatttta 30

<210> 398
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 398
tatttagca ctacggtatc agcgtatctc 30

<210> 399
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 399
tgctacgtgc tctggacggg cgctatcagc 30

<210> 400
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 400
aaatgaacag acaagaagca acagaaattg 30

<210> 401
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 401
aagttgatcg tatctattta gaatatcgca 30

<210> 402
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 402
attcactttg acagatacta atgctacatc 30

<210> 403
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 403
caagcagtgt aaaggtgggt tatatgttaa 30

<210> 404
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 404
catagtatag ccgtcttctt tgattgattg 30

<210> 405
<211> 30
<212> DNA
<213> *Streptococcus thermophilus*

<400> 405
ccatgggtgc taaagtgat gactaccgct 30

<210> 406
<211> 36
<212> DNA
<213> *Streptococcus thermophilus*

<400> 406
tttctaggaa tgggtaatta tagcgagcta gaaagc 36

<210> 407
<211> 38
<212> DNA
<213> *Streptococcus thermophilus*

<400> 407
agttgggaag gtcttggaat atctatggca aaaaacct 38

<210> 408
<211> 35
<212> DNA
<213> *Streptococcus thermophilus*

<400> 408
tatatggttc aaatgcgatt caaagactat tcaaa 35

<210> 409
<211> 30
<212> DNA
<213> *Streptococcus agalactiae*

<400> 409
taattgcaa tgcttacaat atcttcgtca 30

<210> 410
<211> 30
<212> DNA
<213> *Streptococcus agalactiae*

<400> 410
atgttctgaa ttacctttct cgacactccg 30

<210> 411
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 411
accatcaagg ctcttatctg cagattgta 30

<210> 412
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 412
aaatggttgc caatgacttt ctagagtga 30

<210> 413
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 413
acaaaatctt ttgttgctcc tggacgtatt 30

<210> 414
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 414
atgtaaggta ttgtaaaact tcttctgcg 30

<210> 415
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 415
actgttccta taattaaaat aaaagaggta 30

<210> 416
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 416
tgtccagta aaaagtaatt ttaaagcatt 30

<210> 417
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 417
cgctcgattg atgctatcaa ctatattgaa 30

<210> 418
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 418
ttcttcaaga gaactgttag aacagcttca 30

<210> 419
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 419
aaggctacttt tagcttggtc ttgtggtgtt 30

<210> 420
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 420
acagctactg taaattctgc ttctacggtt 30

<210> 421
<211> 30
<212> DNA
<213> streptococcus agalactiae

<400> 421
tagtgcagtt gtcaaggaga ttgtgagcga 30

<210> 422
<211> 30
<212> DNA
<213> streptococcus agalactiae

<400> 422
tttaaccttt gaaaatgtga aaggctcgta 30

<210> 423
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 423
gcgatgatgg taagtcac tggacagcgt 30

<210> 424
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 424
ttttacacac gatgtcagat ataattgcaa 30

<210> 425
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 425
agtactgcac taggaattgt agagatcaaa 30

<210> 426
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 426
cgtaccatct atcaattac cgcaagctgt 30

<210> 427
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 427
ttaaagatt taaactatca agcgtcaatt 30

<210> 428
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 428
ttctaaatgc tggtagctgc ttgcataaa 30

<210> 429
 <211> 31
 <212> DNA
 <213> Streptococcus agalactiae

 <400> 429
 ttgctgctag acccaaacag ttattttta g 31

 <210> 430
 <211> 30
 <212> DNA
 <213> Streptococcus agalactiae

 <400> 430
 tccttttta gataatgtgc gatcacggac 30

 <210> 431
 <211> 29
 <212> DNA
 <213> Streptococcus agalactiae

 <400> 431
 tttaccaat gcttccatat cgcttatat 29

 <210> 432
 <211> 30
 <212> DNA
 <213> Streptococcus agalactiae

 <400> 432
 tggttatata ttactaatc catcagcatt 30

 <210> 433
 <211> 30
 <212> DNA
 <213> Streptococcus agalactiae

 <400> 433
 aagctaattc tcatctcacc gagatggata 30

 <210> 434
 <211> 30
 <212> DNA
 <213> Streptococcus agalactiae

 <400> 434
 aaaaactctt accacttaca tacatgtatg 30

<210> 435
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 435
gctggagatt ttacaagcag ttgaatttc 30

<210> 436
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 436
atcacaccag tcgttatgat ggatgactat 30

<210> 437
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 437
tgtcaacagt acgtgagacg agtgtgtagg 30

<210> 438
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 438
tgaagttgat ggatatgttg atttagagct 30

<210> 439
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 439
taatcatttt atgagagata ccgcctcaag 30

<210> 440
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 440
tttaaagaga tatctgttc atcttgcgga 30

<210> 441
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 441
aatcacttct gcataaatat cttttacttc 30

<210> 442
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 442
aaacatccgc aacgggataa ataaagctag 30

<210> 443
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 443
agtttcttgt gggttagctt gtccaccgta 30

<210> 444
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 444
gaacatgaaa gattttaaaa aagaacattt 30

<210> 445
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 445
agaggggaaa atatcaatgc cgaatgctga 30

<210> 446
<211> 30
<212> DNA
<213> Streptococcus agalactiae

<400> 446
gatggtacaa aatcatttgt tggactgat 30

<210> 447
<211> 30
<212> DNA
<213> Streptococcus mutans

<400> 447
aaaaggaaac gccattaatt aatatggtga 30

<210> 448
<211> 30
<212> DNA
<213> Streptococcus mutans

<400> 448
gattgaacca gctagcgcag ttagtgctct 30

<210> 449
<211> 30
<212> DNA
<213> Streptococcus mutans

<400> 449
cgctaaaagc tgttgtgtca tcatagttag 30

<210> 450
<211> 30
<212> DNA
<213> Streptococcus mutans

<400> 450
taaataatattt caattagaca atagacaaac 30

<210> 451
<211> 30
<212> DNA
<213> Streptococcus mutans

<400> 451
tgcctatgta ttcggacatg acttgccaca 30

<210> 452
<211> 30
<212> DNA
<213> Streptococcus mutans

<400> 452
atgtgaaaag aaagtaacta ctacatttga 30

<210> 453
<211> 30
<212> DNA
<213> Streptococcus pyogenes

<400> 453
tgcgctgggt gatttcttct tgcgctttt 30

<210> 454
<211> 30
<212> DNA
<213> Streptococcus pyogenes

<400> 454
ttatatgaac ataactcaat ttgtaaaaaa 30

<210> 455
<211> 30
<212> DNA
<213> Streptococcus pyogenes

<400> 455
aggaatatcc gcaataatta attgcgctct 30

<210> 456
<211> 30
<212> DNA
<213> Streptococcus pyogenes

<400> 456
taaatttgtt tagcaggtaa accgtgcttt 30

<210> 457
<211> 30
<212> DNA
<213> Streptococcus pyogenes

<400> 457
ttcagcacac tgagacttgt tgagttccat 30

<210> 458
<211> 33
<212> DNA
<213> Streptococcus pyogenes

<400> 458
ctgtgacatt gcgggatgta atcaaagtaa aaa 33

<210> 459

<211> 32

<212> DNA

<213> *Streptococcus pyogenes*

<400> 459

aaagcaaacc tagcagaagc agaaaatgac tt 32

<210> 460

<211> 34

<212> DNA

<213> *Streptococcus pyogenes*

<400> 460

tgatgtaatt ggtgattttc gtgatatgct tttt 34

<210> 461

<211> 5849

<212> DNA

<213> *Streptococcus thermophilus*

<400> 461

atgagtgact	tagtttttagg	acttgatatc	ggtatagggt	ctgttggtgt	aggtatcctt	60
aacaaagtga	caggagaaat	tatccataaa	aactcacgca	tcttcccagc	agctcaagca	120
gaaaataacc	tagtacgtag	aacgaatcgt	caaggaagac	gcttgacacg	acgtaaaaaa	180
catcgtatag	ttcgttttaa	tcgtctattt	gaggaaagtg	gattaatcac	cgattttacg	240
aagatttcaa	ttaatcttaa	cccatatcaa	ttacgagtta	agggcttgac	cgatgaattg	300
tctaatagag	aactgtttat	cgctcttaaa	aatatgggtga	aacaccgtgg	gattagttac	360
ctcgatgatg	ctagtgatga	cggaaattca	tcagtaggag	actatgcaca	aattgttaag	420
gaaaatagta	aacaattaga	aactaagaca	ccgggacaga	tacagttgga	acgctaccaa	480
acatatggtc	aattacgtgg	tgattttact	gttgagaaag	atggcaaaaa	acatcgcttg	540
attaatgtct	ttccaacatc	agcttatcgt	tcagaagcct	taaggatact	gcaaactcaa	600
caagaattta	atccacagat	tacagatgaa	tttattaatc	gttatctcga	aattttaact	660
ggaaaacgga	aatattatca	tggacccgga	aatgaaaagt	cacggactga	ttatggtcgt	720
tacagaacga	gtggagaaac	tttagacaat	atTTTTggaa	ttctaattgg	gaaatgtaca	780
ttttatccag	aagagtttag	agcagcaaaa	gcttcctaca	cggctcaaga	attcaatttg	840
ctaaatgatt	tgaacaatct	aacagttcct	actgaaacca	aaaagttgag	caaagaacag	900
aagaatcaaa	tcattaatta	tgtcaaaaat	gaaaaggcaa	tggggccagc	gaaacttttt	960
aaatatatcg	ctaagttact	ttcttgat	gttgagata	tcaagggata	ccgtatcgac	1020
aatcaggta	aggctgagat	tcatactttc	gaagcctatc	gaaaaatgaa	aacgcttgaa	1080
accttagata	ttgaacaaat	ggatagagaa	acgcttgata	aattagccta	tgtcttaaca	1140
ttaaactctg	agagggaagg	tattcaagaa	gccttagaac	atgaatttgc	tgatggtagc	1200
tttagccaga	agcaagttag	cgaattgggt	caattccgca	aagcaaatag	ttccattttt	1260
ggaaaaggat	ggcataat	ttctgtcaaa	ctgatgatgg	agttaattcc	agaattgtat	1320

gagacgtcag aagagcaaat gactatcctg acacgacttg gaaaacaaaa acgacttcgt 1380
 cttaaataa aacaaaatat ttcaaataaa acaaaatata tagatgagaa actattaact 1440
 gaagaaatct ataatcctgt tgttgctaag tctgttcgcc aggcataaaa aatcgtaaat 1500

gcggcgatta aagaatacgg agactttgac aatattgtca tcgaaatggc tcgtgaacaa 1560
 aatgaagatg atgaaaagaa agctattcaa aagattcaaa aagccaacaa agatgaaaaa 1620
 gatgcagcaa tgcttaaggc tgctaaccac tataatggaa aggcagaatt accacatagt 1680
 gttttccacg gtcataagca attagcgact aaaatccgcc tttggcatca gcaaggagaa 1740
 cgttgccttt atactggtaa gacaatctca atccatgatt tgataaataa tcctaatacag 1800
 tttgaagtag atcatatttt acctctttct atcacattcg atgatagcct tgcaaataag 1860
 gttttggttt atgcaactgc taaccaagaa aaaggacaac gaacacctta tcaggcttta 1920
 gatagtatgg atgatgcgtg gtctttccgt gaattaaaag cttttgtacg tgagtcaaaa 1980
 acactttcaa acaagaaaaa agaatacctc ctacagaag aagatatttc aaagtttgat 2040
 gttcgaaaga aatttattga acgaaatctt gtagatacaa gatacgcttc aagagttgtc 2100
 ctcaatgccc ttcaagaaca ctttagagct cacaagattg atacaaaagt ttccgtggtt 2160
 cgtggccaat ttacatctca attgagacgc cattggggaa ttgagaagac tcgtgatact 2220
 tatcatcacc atgctgtcga tgcattgatt attgccgcct caagtcagtt gaatttgtgg 2280
 aaaaaacaaa agaataccct tgtaagttat tcagaagaac aactccttga tattgaaaca 2340
 ggtgaactta ttagtgtatg tgagtacaag gaatctgtgt tcaaagcccc ttatcaacat 2400
 tttgttgata cattgaagag taaagaatth gaagacagta tcttattctc atatcaagtg 2460
 gattctaagt ttaatcgtaa aatatacagat gccactattt atgagacaag acaggctaaa 2520
 gtgggaaaaa ataagaagga tgaaacttat gtcttaggga aaatcaaaga tatctatact 2580
 caggatggtt atgatgcctt tatgaagatt tataagaagg ataagtcaaa attcctcatg 2640
 tatcgtcacg accccacaaac ctttgagaaa gttatcgagc caattttaga gaactatcct 2700
 aataagcaaa tgaatgaaaa agggaaagag gtaccatgta atcctttcct aaaatataaa 2760
 gaagaacatg gctatatctg taaatatagt aaaaaaggca atggctcctga aatcaagagt 2820
 cttaaatact atgatagtaa gcttttaggt aatcctattg atattactcc agagaatagt 2880
 aaaaataaag ttgtcttaca gtcattaaaa ccttgagaaa cagatgtcta tttcaataag 2940
 gctactggaa aatacgaat ccttgattta aaatatgctg atctacaatt tgagaaaggg 3000
 acaggaacat ataagatttc ccaggaaaaa tacaatgaca ttaagaaaaa agagggtgta 3060
 gattctgatt cagaattcaa gtttacactt tataaaaatg atttgttact cgtaaagat 3120
 acagaaacaa aagaacaaca gcttttccgt tttctttctc gaactttacc taaacaaaag 3180
 cattatgttg aattaaaacc ttatgataaa cagaaatttg aaggagggtga ggcgttaatt 3240
 aaagtgttgg gtaacgttgc taatggtggt caatgcataa aaggactagc aaaatcaaat 3300
 atttctattt ataaagtaag aacagatgtc ctaggaaatc agcatatcat caaaaatgag 3360
 ggtgataagc ctaagctaga tttttaatat taattgttag aaagtgttgc aattatagtt 3420
 atcatatgct ataataatcg tgtaagggac gccttacaca gttacttaaa tcttgagaa 3480
 gctacaaaga taaggcttca tgccgaaatc aacaccctgt cattttatgg cagggtgttt 3540

tcgttatitta aagaggagaa gaaatgactt ggagagtgtg acatgtcagt caaagtgaga 3600
 agatgcgctt aaagcttgat aacttattag tgcaaaaaat gggacaagag tttacgggtgc 3660
 cactaagtga tatttcgata atcgttgagc aaggtgggga tacagtgtgt acccttcgtc 3720

tattaagtgc cttaagtaaa tataatatgt ccttggtcgt ttgtgataac gaacatttac	3780
caacaggaat ttatcactca caaaatgggc acttttagagc gtacaagcgc ttgaaagaac	3840
agctggattg gtctcagaaa caaaaggaca aggcattggca gattgtaact tattataaaa	3900
tcaataacca agaggatggt cttagcatgt ttgaaaaaag tctggacaac attagattac	3960
tttcagacta taaagagcag atagaacctg gtgatagaac gaatagagag ggacatgctg	4020
ccaagggtcta ctttaatgag ctctttggta aacaatttgt cagagtaact cagcaagaag	4080
ctgatgtcat caatgctggt ttaaactatg gctatgctat catgagggtc cagatggcta	4140
gaatagtggc gggttatggt ttaaatggcc tattaggaat cttccataaa aatgaatata	4200
atcagtttaa tttggttgac gatttgatgg agccatttag acagattgta gatgtttggg	4260
tatatgataa tctacgagat caggaattcc ttaagtatga gtatagggtg ggattgacag	4320
atttactcaa tgcataaact aaatatggca aagagacttg ctcagtgaac gttgctatgg	4380
acaaatatgt caaaggcttt atcaaatata tttcggaaaa agatagtagt aaatttcact	4440
gcccagtggt atcaagttaa gagtggagaa aataagatga ggtatgaagc attgagatta	4500
ttatgttttt ttgatttacc aatggaatcc aaggatgaaa aaagaatata tcgtaatttt	4560
cgtaaagaat taatttcaaa tgggtttgaa atgttacaat tttcgggtcta ctatcgact	4620
tgtcctaata gaagctttgc aaataaattt tataagaagt taaagattag caatcttcct	4680
gctgggaatg tgagactttt ggcagttact gaaaaacaat tttcagagat gacattaatt	4740
ataggtggta aaactaagca agaagaaatc gtcagtgaac ataagttggt ggttatatga	4800
aatattttgt acaacatcct tacaagaac gtattgaatt aaatattggt gcaatcacac	4860
aaattgttgg tcagaataaa gaactcaaat attatatgtg gcaaattttg agctggtatt	4920
ttggcggaaa aaaatactca agtgaggact taagtatttt tgattatgag gaacctacta	4980
tacttgatga gtctggagaa atagtgaagc gaagtagcta tcactatata gacatttcaa	5040
gttttaagga tttactggag cagatggaat acaagaaagg aacacttgct cagggttacc	5100
ttagtaaaat tctcaatcag gttgatattg taggccattt ggagaaaatt aatgaacaag	5160
tagagcttat agaaggagca atgaatcagc atataaactt aaactgtggt cagggtgag	5220
accatttgga gaatcacctc ctaacactag accaattact ttcaaaaaat tttagtccct	5280
tttttgctat cgagaataag aatttatctt ttgaatgggt ttcaataact gataaacttt	5340
ctctctttct agaaatgtta gaccgccttc tgtcacaac aacagagaag tatctcattg	5400
tgctaaaaaa tattgatggc tttatctcag aagaatctta tactattttt tataggcaaa	5460
tctgtcatct ggtcaagaag tatccaaatc taacctttat tttgtttcct agtgaccaag	5520
gctattttaa aattgatgaa gaaaatagta ggttcgtcaa tttttatct gaccaggtgg	5580
agcatttgta tgatgttgag tttatgtatg aaagagtaat gaaatattat ccaagtaatg	5640
attttccgac gagagaagggt tttagatgt ctttagaaac tgtgacacct tatttattga	5700
caaaaatgct gagacaacct agtctctcac ttgttgattc agtaatattg aatatactaa	5760
atcagttggt tcattttagt taccgtataa gatattctca gacacctgat aaggaaactat	5820
tacataaatt tttagaaagt aaggattga	5849

<210> 462

<211> 3387

<212> DNA

<213> *Streptococcus thermophilus*

<400> 462

atgagtgact	tagtttttagg	acttgatatc	ggtatagggtt	ctgttggtgt	aggtatcctt	60
aacaaagtga	caggagaaat	tatccataaa	aactcacgca	tcttcccagc	agctcaagca	120
gaaaataacc	tagtacgtag	aacgaatcgt	caaggaagac	gcttgacacg	acgtaaaaaa	180
catcgtatag	ttcgttttaa	tcgtctattt	gaggaaagt	gattaatcac	cgattttacg	240
aagatttcaa	ttaatcttaa	cccatatcaa	ttacgagtta	agggcttgac	cgatgaattg	300
tctaatagag	aactgtttat	cgctcttaaa	aatatgggtga	aacaccgtgg	gattagttac	360
ctcgatgatg	ctagtgtatga	cggaatttca	tcagtaggag	actatgcaca	aattgttaag	420
gaaaatagta	aacaattaga	aactaagaca	ccgggacaga	tacagttgga	acgctacca	480
acatatggtc	aattacgtgg	tgattttact	gttgagaaag	atggcaaaaa	acatcgcttg	540
attaatgtct	ttccaacatc	agcttatcgt	tcagaagcct	taaggatact	gcaaactcaa	600
caagaattta	atccacagat	tacagatgaa	tttattaatc	gttatctcga	aattttaact	660
ggaaaacgga	aatattatca	tggacccgga	aatgaaaagt	cacggactga	ttatggtcgt	720
tacagaacga	gtggagaaac	tttagacaat	atttttggaa	ttctaattgg	gaaatgtaca	780
ttttatccag	aagagtttag	agcagcaaaa	gcttcttaca	cggtcaaga	attcaatttg	840
ctaaatgatt	tgaacaatct	aacagttcct	actgaaacca	aaaagttgag	caaagaacag	900
aagaatcaaa	tcattaatta	tgtcaaaaat	gaaaaggcaa	tggggccagc	gaaacttttt	960
aaatatatcg	ctaagttact	ttcttgtgat	gttgacagata	tcaagggata	ccgtatcgac	1020
aatcaggta	aggctgagat	tcatactttc	gaagcctatc	gaaaaatgaa	aacgcttgaa	1080
accttagata	ttgaacaaat	ggatagagaa	acgcttgata	aattagccta	tgtcttaaca	1140
ttaaacactg	agaggggaagg	tattcaagaa	gccttagaac	atgaatttgc	tgatggtagc	1200
tttagccaga	agcaagttag	cgaattgggt	caattccgca	aagcaaatag	ttccattttt	1260
ggaaaaggat	ggcataattt	ttctgtcaaa	ctgatgatgg	agttaattcc	agaattgtat	1320
gagacgtcag	aagagcaaat	gactatcctg	acacgacttg	gaaaacaaaa	acgacttcgt	1380
cttcaaataa	aacaaatat	ttcaaataaa	acaaaatata	tagatgagaa	actattaact	1440
gaagaaatct	ataatcctgt	tgttgctaag	tctgttcgcc	aggctataaa	aatcgtaaat	1500
gcggcgatta	aagaatacgg	agactttgac	aatattgtca	tcgaaatggc	tcgtgaaaca	1560
aatgaagatg	atgaaaagaa	agctattcaa	aagattcaaa	aagccaacaa	agatgaaaaa	1620
gatgcagcaa	tgcttaaggc	tgctaacca	tataatggaa	aggctgaatt	accacatagt	1680
gttttccacg	gtcataagca	attagcgact	aaaatccgcc	tttggcatca	gcaaggagaa	1740
cgttgccttt	atactggtaa	gacaatctca	atccatgatt	tgataaataa	tcctaatacag	1800
tttgaagtag	atcatatttt	acctctttct	atcacattcg	atgatagcct	tgcaaataag	1860
gttttggttt	atgcaactgc	taaccaagaa	aaaggacaac	gaacacctta	tcaggcttta	1920
gatagtatgg	atgatgcgtg	gtctttccgt	gaattaaaag	cttttgtacg	tgagtcaaaa	1980
acactttcaa	acaagaaaa	agaatacctc	cttacagaag	aagatatctc	aaagtttgat	2040
gttcgaaaga	aatttattga	acgaaatctt	gtagatacaa	gatacgcttc	aagagttgtc	2100
ctcaatgcc	ttcaagaaca	ctttagagct	cacaagattg	atacaaaagt	ttccgtgggt	2160
cgtggccaat	ttacatctca	attgagacgc	cattggggaa	ttgagaagac	tcgtgatact	2220
tatcatcacc	atgctgtcga	tgcatgtgatt	attgccgcct	caagtcagtt	gaatttgtgg	2280
aaaaaacaaa	agaataacct	tgtaagttat	tcagaagaac	aactccttga	tattgaaaca	2340
atataatata	ttatataata	taataataaa	gaatatatat	tcaaaagccc	ttatataaat	2400

```

ggtagaattt ttatgtatga ttagtataag gaattctgtt tcaagcccc ttatcaaat 2400
tttgttgata cattgaagag taaagaattt gaagacagta tcttattctc atatcaagtg 2460
gattctaagt ttaatcgtaa aatatcagat gccactatct atgcgacaag acaggctaaa 2520
gtgggaaaag ataagaagga tgaacttat gtcttaggga aaatcaaaga tatctatact 2580
caggatgggt atgatgcctt tatgaagatt tataagaagg ataagtcaaa attcctcatg 2640
tatcgtcacg acccaaaaac ctttgagaaa gttatcgagc caatctttaga gaactatcct 2700
aataagcaaa tgaatgaaaa aggaaaagag gtaccatgta atcctttcct aaaatataaa 2760
gaagaacatg gctatattcg taaatatagt aaaaaaggca atggtcctga aatcaagagt 2820
cttaataact atgatagtaa gcttttaggt aatcctattg atattactcc agagaatagt 2880
aaaaataaag ttgtcttaca gtcattaaaa ccttgagaa cagatgtcta tttcaataag 2940
gctactggaa aatacgaat ccttggatta aaatatgctg atctacaatt tgagaaaggg 3000
acaggaacat ataagatttc ccaggaaaaa tacaatgaca ttaagaaaaa agagggtgta 3060
gattctgatt cagaattcaa gtttacactt tataaaaatg atttgttact cgtaaagat 3120
acagaaacaa aagaacaaca gcttttccgt tttctttctc gaactttacc taaacaaaag 3180
cattatgttg aattaaaacc ttatgataaa cagaaatttg aaggagggtga ggcgttaatt 3240
aaagtgttgg gtaacgttgc taatgggtgt caatgcataa aaggactagc aaaatcaaat 3300
atttctatct ataaagtaag aacagatgtc ctaggaaatc agcatatcat caaaaatgag 3360
ggtgataagc ctaagctaga tttttaa 3387

```

<210> 463

<211> 912

<212> DNA

<213> *Streptococcus thermophilus*

<400> 463

```

atgacttgga gagttgtaca tgtcagtc aaagtgaaga tgcgcttaaa gcttgataac 60
ttattagtgc aaaaaatggg acaagagttt acggtgccac taagtgatat ttcgataatc 120
gttgcagaag gtggggatag agttgttacc cttcgtctat taagtgcctt aagtaaatat 180
aatattgcct tggtcgtttg tgataacgaa catttaccaa caggaattta tcaactcaca 240
aatgggcact tttagagcgt caagcgcttg aaagaacagc tggattggtc tcagaaacaa 300
aaggacaagg catggcagat tgtaacttat tataaaatca ataaccaaga ggatgttcta 360
gccatgtttg aaaaaagtct ggacaacatt agattacttt cagactataa agagcagata 420
gaacctggtg atagaacgaa tagagagggg catgctgcca aggtctactt taatgagctc 480
tttggtaaac aatttgtcag agtaactcag caagaagctg atgtcatcaa tgctggttta 540
aactatggct atgctatcat gagggctcag atggctagaa tagtggcggg ttatggttta 600
aatggcctat taggaatctt ccataaaaat gaatacaatc agtttaattt ggttgacgat 660
ttgatggagc catttagaca gattgtagat gtttggttat atgataatct acgagatcag 720
gaattcctta agtatgagta taggttgagg ttgacagatt tactcaatgc taaaatcaaa 780
tatggcaaa agacttgctc agtgacagtt gctatggaca aatatgtcaa aggccttatc 840
aaatatatct cggaaaaaga tagtagtaaa tttcactgcc cagtgggtatc aagtttagag 900
tgagaaaaat aa 912

```

<210> 464

<211> 324

<212> DNA

<213> *Streptococcus thermophilus*

<400> 464

```

atgaggatg aagcattgag attattatgt ttttttgatt taccaatgga atccaaggat    60
gaaaaaagaa tatatcgtaa ttttcgtaaa gaattaattt caaatgggtt tgaaatgtta    120
caattttcgg tctactatcg cacttgcctt aatagaagct ttgcaaataa attttataag    180
aagttaaaga ttagcaatct tcctgctggg aatgtgagac ttttggcagt tactgaaaaa    240
caattttcag agatgacatt aattataggt ggtaaaacta agcaagaaga aatcgtcagt    300
gataataagt tggtggttat atga                                           324

```

<210> 465

<211> 1053

<212> DNA

<213> *Streptococcus thermophilus*

<400> 465

```

atgaaatatt ttgtacaaca tccttacaaa gaacgtattg aattaaatat tggtgcaatc    60
acacaaattg ttggtcagaa taaagaactc aaatattata tttggcaaat ttgagctgg    120
tattttggcg gaaaaaata ctcaagtgag gacttaagta tttttgatta tgaggaaacct    180
actatacttg atgagtctgg agaaatagtg aagcgaagta gctatcacta tatcgacatt    240

tcaagtttta aggatttact ggagcagatg gaatacaaga aaggaacact tgctcagggt    300
taccttagta aaattctcaa tcaggttgat attgtaggcc atttgagaga aattaatgaa    360
caagtagagc ttatagaagg agcaatgaat cagcatataa acttaaactg tggtcagggtg    420
gagtaccatt tggagaatca ccttctaaca ctagaccaat tactttcaaa aaattttagt    480
cccttttttg ctatcgagaa taagaattta tcttttgaat gggtttcaaa tactgataaa    540
ctttctctct ttctagaaat gttagaccgc cttctgtcac aaacaacaga gaagtatctc    600
attgtgctaa aaaatattga tggctttatc tcagaagaat cttatactat tttttatagg    660
caaatctgtc atctgggtcaa gaagtatcca aatctaacct ttattttggt tcctagtgc    720
caaggctatt taaaaattga tgaagaaaat agtaggttcg tcaatatttt atctgaccag    780
gtggagcatt tgtatgatgt tgagtttatg tatgaaagag taatgaaata ttatccaagt    840
aatgattttc cgacgagaga aggttttagg atgtctttag aaactgtgac acctatttta    900
ttgacaaaaa tgctgagaca acctagtctc tcacttggtg attcagtaat attgaatatc    960
ctaaatcagt tgtttcattt tagttaccgt ataagatatt ctgagacacc tgataaggaa   1020
ctattacata aatttttaga aagtaaggat tga                                           1053

```

<210> 466

<211> 4123

<212> DNA

<213> *Streptococcus thermophilus*

<400> 466

```

atgagcgatt tatatagtca aagggtccaat tattacctgt cttatctga acaaagaatt    60

```

atcattaaaa atgataataa agagattgtc aaagaagtgt ccatttcact cgttgataat	120
gtattacttt ttggtaatgc acaactgacc acccaactca tcaaagcctt gtcaaagaac	180
aagggtgaatg ttactatatt ctcaaagtgt ggtcaattta tttctagtat tgaaaccac	240
aggcaggacg aattccaaaa gcaagagttg caagcaaagg cttattttga agaggatttc	300
cgtttagagg ttgcgaggag tattgctacg accaagggtga ggcacccaat tgccttactt	360
agagagtttg atacggatgg tctactagat acctcagatt attctagggt tgaagatagt	420
gtcaatgata ttcagaaagc ttattccatt acagaaatta tgggttacga aggtcgccct	480
gcgaaatcct atttttacta tctgaattta ctctgttcta atgactttca ttttaatggt	540
aggagtagac ggcctgggga ggattgtttt aacagtgcct tcaattttgg ctatagtatc	600
ttatatctct gcttaatggg ctgattaaga aaaacgggct aagcttggga tttggggtaa	660
ttcacaagca tcatcagcat catgcgacct tggccagtga tttaatggaa gaatggagac	720
ctatcatcgt cgataatacg cttatggagt tggtagaaa tggtaaactt cttttaagtc	780
attttgaaaa taaggatcaa gacttcatac tcacccatga aggagagaa atctttgcac	840
gggctttacg ttcaagaata ttagaagtcc atcagtatat tgagttagat aaaaaacgct	900
attcttttct ttatacagca gataggcaaa tcaagagttt gattagggtt tttagagaac	960
ttgaccctag tctctatgag acaagttaca caggagggca ttaatgggac tttactttaa	1020
cctcagcgaa gaagagcgtg agtttgccaa acaaaaaaac catgttttgt ctgattattt	1080
atgatattcg aagtaacaaa cgtagactta aactctcgaa attacttgag ggttatggcg	1140
tgagggtgca aaaatcctgt ttcgaaagtc acctgtcaag aaatgattat cagtctctcc	1200
ttaaggatat cgagggtctt tccaaggctg atgaagaaga cagcataata gtgtatgtgc	1260
caaccaaaga agaggtgact agtttttagc cctaccatag tgctgaaaaa ttagatgaca	1320
ttctcttccc ctaagccttt atagaccttt aatcatatgg tacactatag atagtgtttc	1380
cagaggctct taaggaaatc aaagatagag agacacttca aagattttgt agatatatgg	1440
aagcattagt agcctatttc aagttttatg gaggtaaaga ttaatgacat tcgctaagat	1500
taaattttca gtcaaaatc gtttagagac aggcctccat attggtggaa gcgatgcttt	1560
tgcagccatt ggtgcaatcg attcgctgt tattaagat cctattacca acctaccgat	1620
cattcctggt tcaagtctca aaggaaaaat gagaacgctt cttgccaagg tttataatga	1680
aaaggtagct gagaaaccaa gcgatgacag tgatattctt agccgtttat ttgggaatag	1740
taaagataaa cgattcaaaa tgggacgctt gatttttcgt gatgccttct tgtcaaacgc	1800
tgatgagcta gactctcttg gggtaagaag ttatacagaa gtaaaatttg aaaatacaat	1860
tgaccgtatc actgccgaag ctaatccaag acaaattgaa cgtgctattc gtaccagtac	1920
tttgatttc gagttgattt atgaaattac agatgagaat gaaaatcaag tcgaagaaga	1980
ttccaaagtg attcgagatg gtttaaaact gcttgaactt gattatcttg gtggttctgg	2040
atctcgaggt tacggtgaag ttgcttttga aaacctcaaa gctactaccg tatttggtaa	2100
ttatgatgtt aaaacattaa atgaactttt aactgcggag gtctaatatg acctataaac	2160
tgtatattat gacctttcag aatgctcatt ttggttcggg cactcttgat agctcaaaat	2220
taacattctc agcagaccgt atcttctcag cactagtgtc agaatcccta aaaatgggaa	2280
aactcgatgc atttcttgcg gaagctaacc aagacaagtt cacgctcaca gatgcctttc	2340
catttcaatt tgggtccctt ttgccgaaac ctattgggtt tcccaaacad gaccaaatag	2400
atcaatcagt tgatgtcaaa gaggttcgcc gtcaagcaaa attgtctaag aaactgcaat	2460
*****	2520

```

TCTCTGCTCT AGAAAATGCT GACGATTATA TCAATGGAGA GTCATTGAA AATGAAGAGC 2520
atgcagtcac cgatactgtg acaaaaaatc aaccacataa ggacggcaat ctttatcagg 2580
tagctacaac cagattttca aatgatacgt cgctttacgt catcgcaaac gaatctgatt 2640
tgcttaatga gttgatgtct agtcttcagt attcaggtct tggtggaag cgttcaagtg 2700
gttttggtcg ttttgagtta gatattcaaa atatccact agaattgtca gatagactga 2760
ctaagaatca ttcagataaa gtgatgagtc ttacgacagc acttcctgta gatgctgacc 2820
ttgaagaagc aatggaagat ggacattact tattaactaa atcaagtggg tttgcattta 2880
gtcatgccac caatgagaat tatcgtaagc aggatcttta caaatttgct tctggttcaa 2940
cttttagtaa aacatttgaa ggtcagattg ttgatgtgag accacttgat ttccctcatg 3000

ctgttttaaa ttatgctaaa ccactcttct ttaaattgga ggtataaaaa tgaaaaatga 3060
ctatagaaca tttaaattaa gcctcctgac acttgctcca attcatattg gtaatggaga 3120
gaagtatacc tctagagaat ttatctatga aaataaaaag ttttactttc ctgacatggg 3180
gaaattctat aataaaatgg tggagaagag gcttgctgaa aagtttgaag catttctaata 3240
tcaaactcgt ccaaatgcac gtaataatcg tcttatttcc ttcttaaatg ataaccgaat 3300
tgcagagcgt tcttttggag gttatagtat ctctgaacaa ggtttagaat cggacaaaaa 3360
tcctgattca accggagcta ttaacgaagt taataaattt attcgagatg cttttggaaa 3420
tccctacatt cctggtagct cactaaaagg tgctattcgt accattttta tgaatactac 3480
ccctaagtgg aataatgaaa atgctgtaaa tgactttgga agatttccga aagagaataa 3540
gaaccttata ccttggggac caaaaaaggg aaaagaatac gatgatttgt ttaacgcaat 3600
tcgtgtgagt gatagtaagc cttttgataa taagagtctt atcttagtgc agaaatggga 3660
ttattcagcg aaaacaaata aagctaaacc acttcccttg tatagagaat caatctctcc 3720
attaacaaaa attgaatttg agattacaac aaccactgat gaagctggaa gattgattga 3780
agaattaggt aagagagcac aagcgtttta taaagactat aaggcatttt tcctatctga 3840
atttcctgat gataagattc aagccaatct acaataccca atttatttag gtgcggggag 3900
cgggtgcttg acaaagactc tatttaagca agctgatggg attttacaaa gacgatacag 3960
tcgaatgaaa actaaatgg ttaaaaaagg agttcttaag ctcacaaaag cacctcttaa 4020
aacagttaag attccatctg gtaatcattc attagtcaag aaccacgagt ccttttatga 4080
aatgggaaaa gctaatttca tgattaagga gattgataaa tga 4123

```

<210> 467

<211> 624

<212> DNA

<213> *Streptococcus thermophilus*

<400> 467

```

atgagcgatt tatatagtca aaggctcaat tattacctgt ccttatctga acaaagaatt 60
atcattaaaa atgataataa agagattgtc aaagaagtgt ccatttcact cgttgataat 120
gtattacttt ttggtaatgc acaactgacc acccaactca tcaaagcctt gtcaaagaac 180
aagggtgaatg tttactattt ctcaaatgtt ggtcaattta tttctagtat tgaaaccac 240
aggcaggacg aattccaaaa gcaagagttg caagcaagg cttattttga agaggatttc 300
cgttttagagg ttgcgaggag tattgctacg accaagggtg ggcaccaat tgccttactt 360
agagagtttg atacggatgg tctactagat acctcagatt attctagggt tgaagatagt 420

```

gtcaatgata ttcagaaagc ttattccatt acagaaatta tgggttacga aggtcgcctt 480
 gcgaaatcct atttttacta tctgaattta ctcgttccta atgactttca ttttaatggt 540
 aggagtagac ggcctgggga ggattgtttt aacagtgccc tcaattttgg ctatagtatc 600
 ttatattcctt gcttaatggg ctga 624

<210> 468

<211> 396

<212> DNA

<213> *Streptococcus thermophilus*

<400> 468

ttgcttaatg ggctgattaa gaaaaacggg ctaagcttgg gatttggggg aattcacaag 60
 catcatcagc atcatgcgac cttggccagt gatttaaatg aagaatggag acctatcatc 120
 gtcgataata cgcttatgga gttggtacga aatggtaaac ttcttttaag tcattttgaa 180
 aataaggatc aagacttcat actcaccat gaaggcagag aaatctttgc acgggcttta 240
 cgttcaagaa tattagaagt ccatacagat attgagttag ataaaaaacg ctattctttt 300
 ctttatacag cagataggca aatcaagagt ttgattaggg cttttagaga acttgaccct 360
 agtctctatg agacaagtta cacaggaggg catata 396

<210> 469

<211> 273

<212> DNA

<213> *Streptococcus thermophilus*

<400> 469

atgttttgtc tgattattta tgatattcga agtaacaaac gtagacttaa actctcgaaa 60
 ttacttgagg gttatggcgt gagggcgcaa aaatcctgtt tcgaagtcga cctgtcaaga 120
 aatgattatc agtctctcct taaggatcgc gagggcttct ccaaggctga tgaagaagac 180
 agcataatag tgtatgtgcc aaccaagaa gaggtgacta gttttagccc ctaccatagt 240
 gctgaaaaat tagatgacat tctcttcccc taa 273

<210> 470

<211> 663

<212> DNA

<213> *Streptococcus thermophilus*

<400> 470

atgacattcg ctaagattaa attttcagct caaattcggt tagagacagg cctccatatt 60
 ggtggaagcg atgcttttgc agccattggg gcaatcgatt cgctgttat taaagatcct 120
 attaccaacc taccgatcat tcttggttca agtctcaaag gaaaaatgag aacgcttctt 180
 gccagggttt ataataaaaa ggtagctgag aaaccaagcg atgacagtga tattcttagc 240
 cgttttattg ggaatagtaa agataaacga ttcaaaatgg gacgcttgat ttttcgtgat 300
 gccttcttgt caaacgctga tgagctagac tctcttgggg taagaagtta tacagaagta 360
 aaatttgaaa atacaattga ccgtatcact gccgaagcta atccaagaca aattgaacgt 420
 gctattcgtg ccagtacttt tgatttcgag ttgatttatg aaattacaga tgagaatgaa 480
 atcagatcgc aagaaatgta ccaactgatt ccaactgatt taaactgatt taaactgatt 540

```

aatcaagtcg aagaagattc caaagtcatt cgagatggcc caaaactgct tgaacttgat 340
tatcttggtg gttctggatc tcgaggttac ggtaagggtg cttttgaaaa cctcaaagct 600
actaccgtat ttggttaatta tgatgttaaa acattaaatg aacttttaac tgcggaggtc 660
taa 663

```

<210> 471

<211> 900

<212> DNA

<213> *Streptococcus thermophilus*

<400> 471

```

atgacctata aactgtatat tatgaccttt cagaatgctc attttggttc gggcactctt 60
gatagctcaa aattaacatt ctacgagac cgtatcttct cagcactagt gctagaatcc 120
ctaaaaatgg gaaaactcga tgcatttctt gcggaagcta accaagacaa gttcacgctc 180
acagatgcct ttccatttca atttgggtccc tttttgccga aacctattgg ttatcccaaa 240
catgaccaa tagatcaatc agttgatgtc aaagagggtc gccgtcaagc aaaattgtct 300
aagaaactgc aatttcttgc tctagaaaat gttgacgatt atatcaatgg agagtatttt 360
gaaaatgaag agcatgcagt catcgatact gtgacaaaaa atcaaccaca taaggacggc 420
aatctttatc aggtagctac aaccagattt tcaaatgata cgtcgcttta cgtcatcgca 480
aacgaatctg atttgcttaa tgagttgatg tctagtcttc agtattcagg tcttggtgga 540
aagcgttcaa gtggttttgg tcgttttgag ttagatattc aaaatatccc actagaattg 600
tcagatagac tgactaagaa tcattcagat aaagtgatga gtcttacgac agcacttcct 660
gtagatgctg acctgaaga agcaatggaa gatggacatt acttattaac taaatcaagt 720
ggttttgcat ttagtcatgc caccaatgag aattatcgta agcaggatct ttacaaattt 780
gcttctggtt caacttttag taaaacattt gaaggtcaga ttgttgatgt gagaccactt 840
gatttccttc atgctgtttt aaattatgct aaaccactct tctttaaatt ggaggtataa 900

```

<210> 472

<211> 1074

<212> DNA

<213> *Streptococcus thermophilus*

<400> 472

```

atgaaaaatg actatagaac atttaaatta agcctcctga cacttgctcc aattcatatt 60
ggtaatggag agaagtatac ctctagagaa tttatctatg aaaataaaaa gttttacttt 120
cctgacatgg ggaaattcta taataaaatg gtggagaaga ggcttgctga aaagtttgaa 180
gcatttctaa ttcaaactcg tccaaatgca cgtaataatc gtcttatttc cttcttaaat 240
gataaccgaa ttgcagagcg ttcttttggg ggttatagta tctctgaaac aggttttaga 300
tcggacaaaa atcctgattc aaccggagct attaacgaag ttaataaatt tattcgagat 360
gcttttggaa atccctacat tcctggtagc tcactaaaag gtgctattcg taccatttta 420
atgaatacta cccctaagtg gaataatgaa aatgctgtaa atgactttgg aagatttccg 480
aaagagaata agaaccttat ccttggggga ccaaaaaagg gaaaagaata cgatgatttg 540
tttaacgcaa ttcgtgtgag tgatagtaag ctttttgata ataagagtct tatcttagtg 600
cagaaatggg attattcagc gaaaacaaat aaagctaaac cacttccttt gtatagagaa 660

```

tcaatctctc cattaacaaa aattgaattt gagattacaa caaccactga tgaagctgga	720
agattgattg aagaattagg taagagagca caagcgtttt ataaagacta taaggcattt	780
ttcctatctg aatttcctga tgataagatt caagccaatc tacaataccc aattttattta	840
ggtgcgggga gcggtgcttg gacaaagact ctatttaagc aagctgatgg tattttacaa	900
agacgataca gtcgaatgaa aactaaaatg gttaaaaaag gagttcttaa gctcacaaaa	960
gcacctctta aaacagttaa gattccatct ggtaatcatt cattagtcaa gaaccacgag	1020
tccttttatg aaatgggaaa agctaatttc atgattaagg agattgataa atga	1074

<210> 473

<211> 5832

<212> DNA

<213> *Streptococcus thermophilus*

<400> 473

atgagtgact tagtttttagg acttgatatc ggtatagggt ctgttggtgt aggtatcctt	60
aacaaagtga caggagaaat tatccataaa aactcacgca tcttcccagc agctcaagca	120
gaaaataacc tagtacgtag aacgaatcgt caaggaagac gcttgacacg acgtaaaaaa	180
catcgtatag ttcgttttaa tcgtctatct gaggaaagtg gattaatcac cgattttacg	240
aagatttcaa ttaatcttaa cccatatcaa ttacgagtta agggcttgac cgatgaattg	300
tctaataaag aactgtttat cgctcttaaa aatatgggtg aacaccgtgg gattagttac	360
ctcgatgatg ctagtgatga cggaaattca tcagtaggag actatgcaca aattgttaag	420
gaaaatagta acaattaga aactaagaca ccgggacaga tacagttgga acgctacca	480
acatatggtc aattacgtgg tgattttact gttgagaaag atggcaaaaa acatcgcttg	540
attaatgtct ttccaacatc agcttatcgt tcagaagcct taaggatact gcaaactcaa	600
caagaattta attcacagat tacagatgaa tttattaatc gttatctcga aattttaact	660
ggaaaacgga aatattatca tggacccgga aatgaaaagt cacggactga ttatggctgt	720
tacagaacga atggagaaac tttagacaat atttttggaa ttctaattgg gaaatgtaca	780
ttttatccag acgagtttag agcagcaaaa gcttctaca cggctcaaga attcaatttg	840
ctaaatgatt tgaacaatct aacagttcct actgaaacca aaaagttgag caaagaacag	900
aagaatcaaa tcattaatta tgtcaaaaat gaaaaggtaa tggggccagc gaaacttttt	960
aaatatatcg ctaaattact ttcttgatg gttgcagata tcaagggaca ccgtatcgac	1020
aaatcaggta aggctgagat tcatactttc gaagcctatc gaaaaatgaa aacgcttgaa	1080
accttagata ttgagcaaat ggatagagaa acgcttgata aattagccta tgtcttaaca	1140
ttaaacactg agagggaagg tattcaagaa gctttagaac atgaatttgc tgatggtagc	1200
tttagccaga agcaagtga cgaattggtt caattccgca aagcaaatag ttccattttt	1260
ggaaaaggat ggcataatct ttctgtcaaa ctgatgatgg agttaattcc agaattgtat	1320
gagacgtcag aagagcaaat gactatcctg acacgacttg gaaaacaaaa aacaacttcg	1380
tcttcaataa aaacaaaata tatagatgag aaactattaa ctgaagaaat ctataatcct	1440
gttggtgcta agtctgttcg ccaggctata aaaatcgtaa atgcggcgat taaagaatac	1500
ggagactttg acaatattgt catcgaaatg gctcgtgaaa caaatgaaga tgatgaaaag	1560
aaagctattc aaaagattca aaaagccaac aaagatgaaa aagatgcagc aatgcttaag	1620

gctgctaacc aatataatgg aaaggctgaa ttaccacata gtgtttttcca cggtcataag	1680
caatttagcga ctaaaatccg cctttggcat cagcaaggag aacgttgcct ttatactggt	1740
aagacaatct caatccatga tttgataaat aatcctaatc agtttgaagt agatcatatt	1800
ttacctcttt ctatcacatt cgatgatagc cttgcaaata aggttttgggt ttatgcaact	1860
gctaaccaag aaaaaggaca acgaacacct tatcaggctt tagatagtat ggatgatgcg	1920
tggctctttcc gtgaattaaa agcttttgta cgtgagtcaa aaacactttc aaacaagaaa	1980
aaagaatacc tccttacaga agaagatatt tcaaagtttg atgttcgaaa gaaatttatt	2040
gaacgaaatc ttgtagatac aagatacgct tcaagagttg tcctcaatgc ccttcaagaa	2100
cacttttagag ctcaacaagat tgatacaaaa gtttccgtgg ttcgtggcca atttacatct	2160
caattgagac gccattgggg aattgagaag actcgtgata cttatcatca ccatgctgtc	2220
gatgcattga ttattgccgc ctcaagtcag ttgaatttgt ggaaaaaaca aaagaatacc	2280
cttgtaagtt attcagaaga acaactcctt gatattgaaa cagggtgaact tattagtgat	2340
gatgagtaca aggaatctgt gttcaaagcc ccttatcaac attttgttga tacattgaag	2400
agtaaagaat ttgaagacag tatcttattc tcatatcaag tggattctaa gtttaatcgt	2460
aaaatatcag atgccactat ttatgcgaca agacaggcta aagtgggaaa agataagaag	2520
gatgaaactt atgtcttagg gaaaatcaaa gatatctata ctcaggatgg ttatgatgcc	2580
tttatgaaga ttataaaga ggataagtca aaattcctca tgtatcgta cgaccacaa	2640
acctttgaga aagttatcga gccaatttta gagaactatc ctaataagga aatgaatgaa	2700
aaagggaaag aagtaccatg taatcctttc ctaaaatata aagaagaaca tggctatatt	2760
cgtaaataata gtaaaaaagg caatggtcct gaaatcaaga gtcttaaata ctatgatagt	2820
aagcttttag gtaatcctat tgatattact ccagagaata gtaaaaaata agttgtctta	2880
cagtcattaa aaccttgag aacagatgtc tatttcaata aaaatactgg taaatatgaa	2940
attttaggac tgaatatgac tgatttaca tttgaaaaga agacaggaac atataagatt	3000
tcccaggaaa aatacaatgg cattatgaaa gaagaggggtg tagattctga ttcagaattc	3060
aagtttacac ttataaaaa tgatttggtta ctggttaaag atacagaaac aaaagaacaa	3120
cagcttttcc gtttcttttc tcgaactatg cctaattgtg aatattatgt agagttaaag	3180
ccttattcaa aagataaatt tgagaagaat gagtcactta ttgaaatttt aggttctgca	3240
gataagtcag gacgatgtat aaaagggcta ggaaatcaa atatttctat ttataaggta	3300
agaacagatg tcctaggaaa tcagcatatc atcaaaaatg agggtgataa gcctaagcta	3360
gattttlaa atlaatttgtt aaaaaagtgt tgcaattata gttatcatat gctataataa	3420
tcgtgtaagg gacgccttac acagttactt aaatcttgca gaagctacaa agataaggct	3480
tcatgccgaa atcaacaccc tgtcatttta tggcaggggtg ttttcgttat ttaaagagga	3540
gaagaaatga cttggagagt tgtacatgtc agtcaaagtg agaagatgcg cttaaagctt	3600
gataacttat ragtgcaaaa gatgggacaa gagtttacgg tgccactaag tgatatttcg	3660
ataatcgttg cagaaggtgg ggatacagtt gttacccttc gtctattaag tgccttaagt	3720
aaatataata ttgccctgggt cgtttgatgat aacgaacatt taccaacagg aattttatcac	3780
tcacaaaatg ggcactttag agcgtacaag cgcttgaaag aacagctgga ttggtctcag	3840
aaacaaaagg aaaaggcatg gcagattgta acttattata aaatcaataa ccaagaggat	3900
gtcctagcca tgtttgaaaa aagtctggac aacattagat tactttcaga ctataaagag	3960
cagatagaac ctggtgatag aacgaataga gagggacatg ctgccaaggt ctactttaat	4020
gagctctttg gtaaacaaatt tgtcagagta actcagcaag aagctgatgt catcaatgct	4080

gggtttaaact atggctatgc tatcatgagg gctcagatgg ctagaatagt ggcgggttat 4140
 gggtttaaatg gcctattagg aatcttccat aaaaatgaat acaatcagtt taatttggtt 4200
 gacgatttga tggagccatt tagacagatt gtagatgttt gggatatatga taatctacga 4260
 gatcaggaat tccttaagta tgagtatagg ttgggattga cagatttact caatgctaaa 4320
 atcaaatatg gcaaagagac ttgctcagtg acagttgcta tggacaaaata tgtcaaaggc 4380
 tttatcaaat atatttcgga aaaagatagt agtaaatttc actgccagtt ggtatcaagt 4440
 ttagagtggg gaaaataaga tgaggatga agcattgaga ttattatgtt ttttgattt 4500
 accaatggaa tccaaggatg aaaaaagaat atatcgtaat tttcgtaaag aattaatttc 4560
 aaatgggttt gaaatgttac aattttcggc ctactatcgc acttgctccta atagaagctt 4620
 tgcaataaaa ttttataaga agttaagat gagcaatctt cctgctggga atgtgagact 4680
 tttggcagtt actgaaaaac aattttcaga gatgacatta attataggtg gtaaaactaa 4740
 gcaagaagaa atcgtcagtg ataataagtt ggtgatcata tgaaattttt tgtacaacat 4800
 ccttacaag aacgtattga attaatatt ggtgcaatca cacaaattgt tggtcagaat 4860
 aatgaactca aatattatac ttggcagatt ttgagctggg attttggtgg aaaaaatac 4920
 tcaagtggg acttaagtat ttttgattat gaggagccta ccatacttga tgaggccaga 4980
 gaaatagtga aacgaagtag ctatcactat atcgacattt caagttttaa ggatttactg 5040
 gagcagatgg aatacaagaa aggaacactt gctcaggggt accttcgtaa aattgtcaat 5100
 caagttgata ttgtaggcca ttgggagaaa attaatgaac aagtagagct tattgaagaa 5160
 gctatgaatc ggcataataa cttaactgt ggacaggtag aataccattt ggagaatctc 5220
 cctctaacac tagaccaact actcacaaaa aatttttagcc cttttttgc cattgagaac 5280
 aagaatctat cttttgaatg ggtttctaatt attgataaac tatccctctt tttagaaatg 5340
 ttagaccatc ttctttcaca aacaacagag aagtatctca ttgtgctaaa aaatattgat 5400
 ggctttatct cagaagaatc ttatactatt ttttataggc aaatctgtca tctgggtcaag 5460
 aagtatcaa atctaactt tattttgttt cctagtacc aaggctattt aaaaattgat 5520
 gaagaaaata gtaggttcgt caatatttta tctgaccagg tggaacattt gtatgatgtt 5580

 gagtttatgt atgaaaggtt aatgaaatat tatccaagta atgattttcc gacgagagaa 5640
 ggtttttagga tgtctttaga aactgtgaca cttattttat tgacaaaaat gctgagacaa 5700
 cctagtctct cacttgttga tttagtaata ttgaatatcc taaatcagct gtttcatttt 5760
 agttaccgta taagatgttc tcagacacct gataaggaac tattacagaa attttttagaa 5820
 agtaaggatt ga 5832

<210> 474

<211> 3369

<212> DNA

<213> *Streptococcus thermophilus*

<400> 474

atgagtgact tagttttagg acttgatatc ggtatagggt ctgttgggtg aggtatcctt 60
 aacaaagtga caggagaaat tatccataaa aactcacgca tcttcccagc agctcaagca 120
 gaaaataacc tagtacgtag aacgaatcgt caaggaagac gcttgacacg acgtaaaaaa 180
 catcgtatag ttcgtttaaa tcgtctattt gaggaagtg gattaatcac cgattttacg 240
 aagatttcaa ttaatcttaa cccatatcaa ttacgagtta agggcttgac cgatgaattg 300

tctaataag	aactgtttat	cgctcttaaa	aatatggtga	aacaccgtgg	gattagttac	360
ctcgatgatg	ctagtgatga	cggaaattca	tcagtaggag	actatgcaca	aattgttaag	420
gaaaatagta	aacaattaga	aactaagaca	ccgggacaga	tacagttgga	acgctacca	480
acatatggtc	aattacgtgg	tgattttact	gttgagaaag	atggcaaaaa	acatcgcttg	540
attaatgtct	ttccaacatc	agcttatcgt	tcagaagcct	taaggatact	gcaaactcaa	600
caagaattta	attcacagat	tacagatgaa	tttattaatc	gttatctcga	aattttaact	660
ggaaaacgga	aatattatca	tggacccgga	aatgaaaagt	cacggactga	ttatggctgt	720
tacagaacga	atggagaaac	tttagacaat	atttttggaa	ttctaattgg	gaaatgtaca	780
ttttatccag	acgagtttag	agcagcaaaa	gcttcctaca	cggctcaaga	attcaatttg	840
ctaaatgatt	tgaacaatct	aacagttcct	actgaaacca	aaaagttgag	caaagaacag	900
aagaatcaaa	tcattaatta	tgtcaaaaat	gaaaaggtaa	tggggccagc	gaaacttttt	960
aaatatatcg	ctaaattact	ttcttgtgat	gttgacagata	tcaagggaca	ccgtatcgac	1020
aatcaggta	aggctgagat	tcatactttc	gaagcctatc	gaaaaatgaa	aacgcttgaa	1080
accttagata	ttgagcaaat	ggatagagaa	acgcttgata	aattagccta	tgtcttaaca	1140
ttaaacactg	agaggggaagg	tattcaagaa	gctttagaac	atgaatttgc	tgatggtagc	1200
tttagccaga	agcaagttag	cgaattgggt	caattccgca	aagcaaatag	ttccattttt	1260
ggaaaaggat	ggcataattt	ttctgtcaaa	ctgatgatgg	agttaattcc	agaattgtat	1320
gagacgtcag	aagagcaaat	gactatcctg	acacgacttg	gaaaacaaaa	aacaacttcg	1380
tcttcaata	aaacaaaata	tatagatgag	aaactattaa	ctgaagaaat	ctataatcct	1440
gttgttgcta	agtctgttcg	ccaggctata	aaaatcgtaa	atgcggcgat	taaagaatac	1500
ggagactttg	acaatattgt	catcgaaatg	gctcgtgaaa	caaataaaga	tgatgaaaag	1560
aaagctattc	aaaagattca	aaaagccaac	aaagatgaaa	aagatgcagc	aatgcttaag	1620
gctgctaacc	aatataatgg	aaaggctgaa	ttaccacata	gtgttttcca	cggtcataag	1680
caattagcga	ctaaaatccg	cctttggcat	cagcaaggag	aacgttgcct	ttatactggt	1740
aagacaatct	caatccatga	tttgataaat	aatcctaata	agtttgaagt	agatcatatt	1800
ttacctcttt	ctatcacatt	cgatgatagc	cttgcaata	aggttttggt	ttatgcaact	1860
gctaaccaag	aaaaaggaca	acgaacacct	tatcaggcct	tagatagtat	ggatgatgcg	1920
tggtctttcc	gtgaattaaa	agcttttgta	cgtgagtcaa	aaacactttc	aaacaagaaa	1980
aaagaatacc	tccttacaga	agaagatatt	tcaaagtttg	atgttcgaaa	gaaatttatt	2040
gaacgaaatc	ttgtagatac	aagatacgct	tcaagagttg	tcctcaatgc	ccttcaagaa	2100
cacttttagag	ctcacaagat	tgatacaaaa	gtttccgtgg	ttcgtggcca	atttacatct	2160
caattgagac	gccattgggg	aattgagaag	actcgtgata	cttatcatca	ccatgctgtc	2220
gatgcattga	ttattgccgc	ctcaagtcag	ttgaatttgt	ggaaaaaaca	aaagaatacc	2280
cttgtaagtt	attcagaaga	acaactcctt	gatattgaaa	cagggtgaact	tattagtgat	2340
gatgagtaca	aggaatctgt	gttcaaagcc	ccttatcaac	attttgttga	tacattgaag	2400
agtaaagaat	ttgaagacag	tatcttattc	tcatatcaag	tggattctaa	gtttaatcgt	2460
aaaatatcag	atgccactat	ttatgcgaca	agacaggcta	aagtgggaaa	agataagaag	2520
gatgaaactt	atgtcttagg	gaaaatcaaa	gatatctata	ctcaggatgg	ttatgatgcc	2580
tttatgaaga	tttataagaa	ggataagtca	aaattcctca	tgtatcgtca	cgaccacaaa	2640

```

acctttgaga aagttatcga gccaatTTta gagaactatc ctaataagga aatgaatgaa 2700
aaagggaaaag aagtaccatg taatcctttc ctaaaatata aagaagaaca tggctatatt 2760
cgtaaataata gtaaaaaagg caatggctct gaaatcaaga gtcttaaata ctatgatagt 2820
aagcttttag gtaatcctat tgatattact ccagagaata gtaaaaaata agttgtctta 2880
cagtcattaa aaccttggag aacagatgtc tatttcaata aaaatactgg taaatatgaa 2940
atTTtaggac tgaaatatgc tgatttaca tttgaaaaga agacaggaac atataagatt 3000
tcccaggaaa aatacaatgg cattatgaaa gaagagggtg tagattctga ttcagaattc 3060
aagtttacac tttataaaaa tgatttgTTa ctcgttaaag atacagaaac aaaagaacaa 3120
cagcttttcc gttttctttc tcgaactatg cctaattgtga aatattatgt agagttaaag 3180
ccttattcaa aagataaatt tgagaagaat gagtcaCTta ttgaaatttt aggttctgca 3240
gataagtCag gacgatgtat aaaagggcta ggaaaaatcaa atatttctat ttataaggta 3300
agaacagatg tcctaggaaa tcagcatatc atcaaaaatg agggtgataa gcctaagcta 3360
gatttttaa 3369

```

<210> 475

<211> 912

<212> DNA

<213> *Streptococcus thermophilus*

<400> 475

```

atgacttgga gagttgtaca tgtcagtcaa agtgagaaga tgcgcttaaa gcttgataac 60
ttattagtgc aaaagatggg acaagagttt acgggtccac taagtgatat ttcgataatc 120
gttgCagaag gtggggatac agttgttacc cttcgtctat taagtgcctt aagtaaatat 180
aatattgcct tggtcgTTtT tgataacgaa catttaccaa caggaattta tcactcacia 240
aatgggcact ttagagcgta caagcgcttg aaagaacagc tggattggtc tcagaacaa 300
aagggaaaagg catggcagat tgtaacttat tataaaatca ataaccaaga ggatgtccta 360
gccatgtttg aaaaaagtct ggacaacatt agattacttt cagactataa agagcagata 420
gaacctggtg atagaacgaa tagagagggg catgctgcca aggtctactt taatgagctc 480
tttggtaaac aatttgtcag agtaactcag caagaagctg atgtcatcaa tgctggttta 540
aactatggct atgctatcat gagggctcag atggctagaa tagtggcggg ttatggttta 600
aatggcctat taggaatctt ccataaaaat gaatacaatc agtttaattt ggttgacgat 660
ttgatggagc catttagaca gattgtagat gtttgggtat atgataatct acgagatcag 720
gaattcctta agtatgagta taggttggga ttgacagatt tactcaatgc taaaatcaaa 780
tatggcaaag agacttgctc agtgacagtt gctatggaca aatatgtcaa aggctttatc 840
aaatatattt cggaaaaaga tagtagtaaa tttcactgcc cagtggatc aagtttagag 900
tgagaaaaat aa 912

```

<210> 476

<211> 324

<212> DNA

<213> *Streptococcus thermophilus*

<400> 476

```

atgaggtatg aagcattgag attattatgt ttttttgatt taccaatgga atccaaggat 60

```

gaaaaaagaa tatatcgtaa ttttcgtaaa gaattaattht caaatgggtt tgaaatgtta 120
 caatthttcgg tctactatcg cacttgctct aatagaagct ttgcaaataa atthttataag 180
 aagttaaaga tgagcaatct tcttgctggg aatgtgagac tthttggcagt tactgaaaaa 240
 caatthttcag agatgacatt aattataggt ggtaaaacta agcaagaaga aatcgtcagt 300
 gataataagt tggatgatcat atga 324

<210> 477

<211> 1053

<212> DNA

<213> *Streptococcus thermophilus*

<400> 477

atgaaattht ttgtacaaca tctttacaaa gaacgtattg aattaaatat tggatgcaatc 60
 acacaaattg ttggtcagaa taatgaactc aaatattata cttggcagat tttgagctgg 120
 tathtttggtg gaaaaaata ctcaagttag gacttaagta tthttgatta tgaggagcct 180
 accatacttg atgaggccag agaaatagtg aaacgaagta gctatcacta tatcgacatt 240

 tcaagthtta aggatttact ggagcagatg gaatacaaga aaggaaactc tgctcagggt 300
 taccttcgta aaattgtcaa tcaagttgat attgtaggcc atthtgagaa aattaatgaa 360
 caagtagagc ttattgaaga agctatgaat cggcatataa acttaaaactg tggacaggta 420
 gaataccatt tggagaatct ccttctaaca cttagaccaac tactcacaaa aaatthttagc 480
 ccattthttg ccattgagaa caagaatcta tctthtgaat gggthttctaa tattgataaa 540
 ctatccctct tthtagaaat gtttagaccat cttctttcac aaacaacaga gaagtatctc 600
 attgtgctaa aaaatattga tggctthtct tcagaagaat cttatactat tthttatagg 660
 caaatctgtc atctgggtcaa gaagtatcca aatctaacct ttathttgth tcctagttag 720
 caaggctatt taaaaattga tgaagaaaat agtaggttcg tcaatathtt atctgaccag 780
 gtggaacatt tgtatgatgt tgagthttatg tatgaaaggg taatgaaata ttatccaagt 840
 aatgathttc cgacgagaga aggtthttagg atgtctthtag aaactgtgac acctatthta 900
 ttgacaaaaa tgctgagaca acctagtctc tcaattgttg attcagtaat attgaatctc 960
 ctaaactcagc tgtthtctth tagttaccgt ataagatgth ctgagacacc tgataaggaa 1020
 ctattacaga aathttthaga aagtaaggat tga 1053

<210> 478

<211> 7900

<212> DNA

<213> *Streptococcus thermophilus*

<400> 478

atgagcgatt tatatagtca aagggtccaat tattacctgt ccttatctga acaaagaatt 60
 atcattaaaa atgataataa agagattgtc aaagaagthg ccatttctact cgttgataat 120
 gtattactth ttggtaatgc acaactgacc acctaaactc tcaaagcctt gtcaaagaac 180
 aagggtgaatg thtactatth ctcaaagtht ggtcaatthta thtctagtht tgaaacctac 240
 aggcaggacg aattccaaaa gcaagagtht caagcaaagg cttathttga agaggattth 300
 cgtthtagagg ttgcgaggag tathgtctag acctaaagthg ggcacctaat tgccttactt 360

agagagtttg atacggatgg tctactagat acctcagatt attctagggt tgaagatagt 420
 gtcaatgata ttcagaaagc ttattccatt acagaaatta tgggttacga aggtcgcctt 480
 gcgaaatcct atttttacta tctgaattta ctcgttccta atgactttca ttttaatggt 540
 aggagtagac ggcctgggga ggattgtttt aacagtggcc tcaattttgg ctatagtatc 600
 ttatatctct gcttaatggg ctgattaaga aaaacgggct aagcttggga tttggggtaa 660
 ttcacaagca tcatcagcat catgcgacct tggccagtga tttaatggaa gaatggagac 720
 ctatcatcgt cgataatacg cttatggagt tggtagaaa tggtaaactt cttttaagtc 780
 attttgaaaa taaggatcaa gacttcatac tcacccatga aggcagagaa atctttgcac 840
 gggctttacg ttcaagaata ttagaagtcc atcagtatat tgagttagat aaaaaacgct 900
 attcttttct ttatacagca gataggcaaa tcaagagttt gattagggct tttagagaac 960

 ttgaccctag tctctatgag acaagttaca caggagggca ttaatgggac tttacttta 1020
 cctcagcgaa gaagagcgtg agtttgccaa acaaaaaacc atgttttgct tgattattta 1080
 tgatattcga agtaacaaac gtagacttaa actctcgaaa ttacttgagg gttatggcgt 1140
 gagggtgcaa aaatcctgtt tcgaagtcaa cctgtcaaga aatgattatc agtctctcct 1200
 taaggatatc gagggcttct acaaggctga tgaagaagac agcataatag tgtatgtgac 1260
 aaccaaagaa gaggtgacta gttttagccc ctaccatagt gctgaaaaat tagatgacat 1320
 tctcttcttc taagccttta tagaccttta atcatatggg acactataga tagtgtttcc 1380
 agtaggtcct acatcttggt cctctagcaa ctgcctagag cacaagatat ggggatataa 1440
 acctaatcct ctcgagaggg gacggaaacg ctttctagct cgctataatt acccattcct 1500
 agaaagatat aaacctaat acctcgagag gggacggaaa ctttgaatag tctttgaatc 1560
 gcatttgaac catatagata taaacctaat tacctcgaga ggggacggaa acagggtttt 1620
 tgccatagat tttccaagac cttcccaact gatataaacc taattacctc gagaggggac 1680
 ggaaacgctt tctagctcgc tataattacc cattcctaga aagatataaa cctaattacc 1740
 tcgagagggg acttttttga aaattttgaa aacagtattg ataccgcttc cagaaagtgt 1800
 tagactaaaa gcacattaag ggcgccccaa tgagttgaaa agtactttca gcttttggg 1860
 ttttttcata caaagatgaa ggagtcgaat gaaaaaatta gtatttactt ttaaaaggat 1920
 cgaccatcct gcacaagatt tggctgttaa atttcatggc ttcttgatgg agcagttgga 1980
 tagtgactat gttgattatc tgcacagca gcaaacaat ccctatgcga ccaaggtaat 2040
 ccaagggaaa gaaacacgc agtggttgt acatctgctc acagacgaca tcgaggataa 2100
 ggtttttatg accttattac agattaaaga ggtgtcctta aacgatctgc ctaaactcag 2160
 tgtcgaaaaa gttgagattc aggagtggg ggcagataaa ctgttagaga ttttcaatag 2220
 tgaggaaaaa caaacctatt tttcaattat ttttgagact ccaacagggt ttaaatctca 2280
 aggttcctac gtcacttcc cgtctatgcg tttgattttt caaagtttga tgcaaaagta 2340
 tggaagggtg gttgaaaatc aacctgaaat tgaagaggat accttagatt acctatctga 2400
 acacagcact atcacgaatt atcgcttggg gacgagttat ttcagggtgc acaggcaacg 2460
 aattcctgcc tttagaggaa agttaacctt taaagtacaa ggcgccccaa ctctaaaagc 2520
 ttatgtcaaa atgcttctaa cattcggtga atattcaggt cttggcatga aaacgagtct 2580
 cggtatggga gggataaagc ttgaagaaag aaaagattga tttattttac ggagctcttt 2640
 tgcattgatg cgtaagggtc attcaaaggg cgacaggaga acgaaaaaaa cacgccttgg 2700
 taggcgcgga ttggtttgat gagattgctg ataataagtt ttttccgat caaattagat 2760

atcacatggc taactaccag agtgataaac ttggaaatga ccatcttgct tacataactt	2820
atatcgctga taacattgcc tctgggtgctg acagaagaca gtcaaagag gagagtgcg	2880
aggatacatc agctaagatt tgggatacct atacaaacca ggctgatatt tttaacgttt	2940
ttggggcaca aacggataaa cgctacttta aaccgacggg tctaaacttg aaatctaaac	3000
ctaactttgc gtcggcaaca tatgaacctt tctcaaaagg tgattatgct gcaattgcga	3060
ctcgtatcaa aaatgaattg gcagaatttg agtttaatca agtacaaatt gactctttgt	3120
taaatctgtt cgaagcaacc ctctcttttg tgccttcttc gactaatact aaagaaatcg	3180
ctgatatttc acttgctgat catagtcgtc tgacagcagc ttttgctcta gccatctatg	3240
attacttgga agacaaaggt cgtcataact ataaggagga cttgtttact aaagcatcag	3300
ccttttatga ggaagaagct tttctcctag cttagctttga cttatcaggg attcaagact	3360
ttatctataa tattaatatt gcgacgaatg gtgctgctaa acaattgaag gctagatctt	3420
tatatcttga ctttatgagc gagtatatag cagacagttt acttgataaa ctaggcctca	3480
atcgggctaa tatgctctat gtcggtgggg gacatgctta ctttgctcta gccaatctg	3540
aaaaaacggt agaaacactc gttcaatttg aaaaagattt caatcaattt ttattggcaa	3600
atttccaaac cagattatat gttgcctttg gttggggaag ctttgcggt aaggatatca	3660
tgagcgaact gaactcacct gaaagctata gacaggtcta tcaaaaggct agtcgcatga	3720
tttctgagaa aaaaatctca aggtatgatt atcaaacctt tatgttggtg aacaggggcg	3780
gtaaatcttc tgaaagagag tgcgagattt gtcattccgt tgagaattta gttgcttatc	3840
atgacaaaaa agtgtgtgac atttgtcgag gcttgatca attttctaa gagattgccc	3900
atgaccattt cattatcact gaaaatgaag ggcttcctat tgggtccgaac gcatgtctta	3960
aggggtgtgc atttgaaaag ctgagccaag aagctttttc ccgtgtctat gtcaaaaatg	4020
actataaggc tggtagacgtt aaggcaaccc atgtttttgt tggagattac cagtatgatg	4080
aaatatacaa ttatgctgcc ttatctaaaa acgaaaatgg gttaggtatt aaacgtttag	4140
ctgttgtagc tcttgacgtg gatgatttgg gaggcagcct tatggctggc ttctccaac	4200
aaggaaatgg gcaatatagt actctatcac gctcagccac tttctctcga agcatgagtc	4260
ttttcttcaa ggtttatatt aaccagtttg ctagtataa gaagctctct atcatctatg	4320
ccggtgggga tgatgttttt gctattggct cttggcaaga tattattgcc ttactgttg	4380
aacttcgtga gaacttcatt aaatggacaa atggaaaact aacactatca gctggtatcg	4440
gtctgtttgc tgataagacc cctattagct taatggcaca tcaaacaggg gagctagaag	4500
aaacagctaa aggcaatgag aaagatagta tttcactctt tagttccgac tataccttta	4560
aatttgatcg gtttatcact aatgtttacg acgataagtt agagcagatt cgctatttct	4620
ttaatcacca agatgaacga ggcaagaatt tcatttataa attgattgaa ttgcttcgaa	4680
attatgatcg tatgaatatg gcacgtttag cttattattt aacacgactt gaagaattga	4740
cgcgtgaaac agacagggat aaatttaaaa catttaaaaa ttatttctat tcttggtaga	4800
caaataagga tgataaggat agaaaagaag cagagttagc cttgcttctc tatatctatg	4860
agattagaaa ggattaggat atgacaatct tgactgatga gaattacgtt gatattgcag	4920
aaaaagcaat tctaaaacta gaaagaaata ctaggacag aaagaatcct gatgccttct	4980
ttcttacaac aagtaagctc agaaacttgc tgagcttaac tagtacactt tttgatgaga	5040
gtaagggtcaa agaatatgat gctctccttg atcgtattgc ttatttaaga gtacaatttg	5100
tctaccaagc aggtagagag attgcagtaa aagatctgat agaaaaggct caaattcttg	5160
aggtctttaa ggaatcaaa gataagacga cacttcaaaa attttctaga tatatgaaa	5220

aggcctctaa ggaattcaaa gatagagaga cactctcaag atctctgaga tatctggag 5220
 cattagtagc ctatttcaag ttttatggag gtaaagatta atgacattcg ctaagattaa 5280
 attttcagct caaattcggt tagagacagg cctccatatt ggtggaagcg atgcttttgc 5340
 agccattggt gcaatcgatt cgctgttat taaagatcct attaccaacc taccgatcat 5400
 tcctggttca agtctcaag gaaaaatgag aacgcttctt gccaaaggtt ataataaaaa 5460
 ggtagctgag aaaccaagcg atgacagtga tattcttagc cgtttatttg ggaatagtaa 5520
 agataaacga ttcaaaatgg gacgcttgat ttttcgtgat gccttcttgt caaacgctga 5580
 tgagctagac tctcttgggg taagaagtta tacagaagta aaatttgaaa atacaattga 5640
 ccgtatcact gccgaagcta atccaagaca aattgaacgt gctattcgta ccagtacttt 5700
 tgatttcgag ttgatttatg aaattacaga tgagaatgaa aatcaagtcg aagaagattt 5760
 caaagtgatt cgagatggtt taaaactgct tgaacttgat tatcttgggt gttctggatc 5820
 tcgaggttac ggtaaggttg cttttgaaaa actcaaagct actaccgtat ttggttaatta 5880
 tgatgttaaa acattaaatg aacttttaac tgcggagggtc taatatgacc tataaactgt 5940
 atattatgac ctttcagaat gctcattttg gttcgggcac tcttgatagc tcaaaattaa 6000
 cattctcagc agaccgtatc ttctcagcac tagtgctaga atccctaaaa atgggaaaac 6060
 tcgatgcatt tcttgcggaa gctaaccaag acaagttcac gctcacagat gcctttccat 6120
 ttcaatttgg tccctttttg ccgaaaccga ttggttatcc caaacatgac caaatagatc 6180
 aatcagttga tgtcaaagag gttcgccgtc aagcaaaatt gtctaagaaa ctgcaatttc 6240
 ttgctctaga aaatgttgac gattatctca atggagagtt atttgaaaat gaagagcatg 6300
 cagtcacga tactgtgaca aaaaatcaac cacataagga cggcaatctt tatcaggtag 6360
 ctacaaccag attttcaaat gatacgtcgc ttacgtcat cgcaaacgaa tctgatttgc 6420
 ttaatgagtt gatgtctagt cttcagtatt cagggtcttg tggaagcgt tcaagtgggt 6480
 ttggtcgttt tgagttagat attcaaaata tcccactaga attgtcagat agactgacta 6540
 agaatcattc agataaagtg atgagtctta cgacagcact tcctgtagat gctgaccttg 6600
 aagaagcaat ggaagatgga cattacttat taactaaatc aagtggtttt gcatttagtc 6660
 atgctacca a tgagaattat cgtaagcagg atctttacaa atttgcttct ggttcaactt 6720
 ttagtaaaac atttgaaggt cagattgttg atgtgagacc acttgatttc cctcatgctg 6780
 ttttaaatta tgctaaacca ctcttcttta aattggaggt ataaaaatga aaaatgacta 6840
 tagaacattt aaattaagcc tcctgacact tgctccaatt catattggta atggagagaa 6900
 gtatacctct agagaattta tctatgaaaa taagaagttt tactttcctg acatggggaa 6960
 attctataat aaaatggtgg agaagaggct tgctgaaaag tttgaagcat ttctaattca 7020
 aactcgtcca aatgcacgta ataatcgtct tatttccttc ttaaatgata accgaattgc 7080

 agagcgttct tttggaggtt atagtatctc tgaacagggt ttagaattcgg acaaaaatcc 7140
 tgattcagcc ggagctatta acgaagttaa taaatttatt cgagatgctt ttggaaatcc 7200
 ctacattcct ggtagctcac taaaagggtc tattcgtacc attttaatga atactacccc 7260
 taagtggaat aatgaaaatg ctgtaaatga ctttggaga tttccgaaag agaataagaa 7320
 ccttatccct tggggaccaa aaaagggaaa agaatacgat gatttgttta acgcaattcg 7380
 tgtgagtgat agtaagcctt ttgataataa gagtcttatt ttagtgaga aatgggatta 7440
 ttcagcgaaa acaataaag ctaaacact tcccttgat agagaatcaa tctctccatt 7500
 aacaaaaatt gaatttgaga ttacaacaac cactgatgaa gctggaagat tgattgaaga 7560
 attaggttaag agagcacaag cgttttataa agactataag gcatttttcc tatctgaatt 7620

```

tcctgatgat aagattcaag ccaatctaca ataccaatt tatttaggtg cggggagcgg 7680
tgcttggaaca aagactctat ttaagcaagc tgatgggtatt ttacaaagac gatacagtcg 7740
aatgaaaact aaaatgggta aaaaaggagt tcttaagctc acaaaagcac ctcttaaaac 7800
agttaagatt ccatctggta atcattcatt agtcaagaac cacgagtcct tttatgaaat 7860
gggaaaagct aatttcatga ttaaggagat tgataaatga 7900

```

<210> 479

<211> 624

<212> DNA

<213> *Streptococcus thermophilus*

<400> 479

```

atgagcgatt tatatagtca aagggtccaat tattacctgt ctttatctga acaagaatt 60
atcattaaaa atgataataa agagattgtc aaagaagtgt ccatttcact cggtgataat 120
gtattacttt ttggtaatgc acaactgacc acccaactca tcaaagcctt gtcaaagaac 180
aagggtgaatg ttactatatt ctcaaagtgt ggtcaattta tttctagtat tgaaaccac 240
aggcaggagc aattccaaaa gcaagagtgt caagcaaagg cttatattga agaggatttc 300
cgtttagagg ttgcgaggag tattgctacg accaagggtga ggcaaccaat tgccttactt 360
agagagtttg atacggatgg tctactagat acctcagatt attctagggt tgaagatagt 420
gtcaatgata ttcagaaagc ttattccatt acagaaatta tgggttacga aggtcgctt 480
gcgaaatcct atttttacta tctgaattta ctcgttccta atgactttca ttttaatggt 540
aggagtagac ggcctgggga ggattgtttt aacagtgcct tcaattttgg ctatagtatc 600
ttatattctt gcttaatggg ctga 624

```

<210> 480

<211> 396

<212> DNA

<213> *Streptococcus thermophilus*

<400> 480

```

ttgcttaatg ggctgattaa gaaaaacggg ctaagcttgg gatttggggg aattcacaag 60
catcatcagc atcatcgac cttggccagt gatttaatgg aagaatggag acctatcatc 120

gtcgataata cgcttatgga gttggtacga aatggtaaac tttttttaaag tcattttgaa 180
aataaggatc aagacttcat actcaccat gaaggcagag aaatctttgc acgggcttta 240
cgttcaagaa tattagaagt ccatcagtat attgagttag ataaaaaacg ctattctttt 300
ctttatacag cagataggca aatcaagagt ttgattaggg cttttagaga acttgaccct 360
agtctctatg agacaagtta cacaggaggg cattaa 396

```

<210> 481

<211> 330

<212> DNA

<213> *Streptococcus thermophilus*

<400> 481

```

atggaacttt actttaacct caacgaagaa gacgatgaat ttaccaaac aaaaaccatg 60

```

```

ttttgtctga ttatttatga tattcgaagt aacaaacgta gacttaaact ctcgaaatta 120
cttgaggggt atggcgtgag ggtgcaaaaa tcctgtttcg aagtcaacct gtcaagaaat 180
gattatcagt ctctccttaa ggatatcgag ggcttctaca aggctgatga agaagacagc 240
ataatagtgt atgtgacaac caaagaagag gtgactagtt ttagccccta ccatagtgtc 300
gaaaaattag atgacattct cttcttctaa 330

```

<210> 482

<211> 732

<212> DNA

<213> *Streptococcus thermophilus*

<400> 482

```

atgaaaaaat tagtatttac ttttaaaagg atcgaccatc ctgcacaaga tttggctggt 60
aaatttcatt gcttcttgat ggagcagttg gatagtgact atgttgatta tctgcatcag 120
cagcaaaaaa atccctatgc gaccaaggta atccaaggga aagaaaacac gcagtgggtt 180
gtacatctgc tcacagacga catcgaggat aagggtttta tgaccttatt acagattaaa 240
gagggtgtct taaacgatct gcctaaactc agtgtcgaaa aagttgagat tcaggagttg 300
ggggcagata aactgttaga gattttcaat agtgaggaaa atcaaaccta tttttcaatt 360
atttttgaga ctccaacagg ttttaaatct caaggttcct acgtcatctt cccgtctatg 420
cgtttgattt ttcaaagttt gatgcaaaag tatggaagggt tgggtgaaaa tcaacctgaa 480
attgaagagg ataccttaga ttacctatct gaacacagca ctatcacgaa ttatcgcttg 540
gagacgagtt atttcagggt gcacaggcaa cgaattcctg ccttttagagg aaagttaacc 600
tttaagtagc aaggcgccca aactctaaaa gcttatgtca aaatgcttct aacattcggt 660
gaatattcag gtcttggcat gaaaacgagt ctcggtatgg gagggataaa gcttgaagaa 720
agaaaagatt ga 732

```

<210> 483

<211> 2277

<212> DNA

<213> *Streptococcus thermophilus*

<400> 483

```

ttgaagaaag aaaagattga tttattttac ggagctcttt tgcattgatat cggtaagggtc 60
attcaaaggg cgacaggaga acgaaaaaaa cagccttgg taggcgcgga ttggtttgat 120
gagattgctg ataatacagt tatttccgat caaattagat atcacatggc taactaccag 180
agtgataaac ttggaaatga ccatcttgct tacataactt atatcgctga taacattgcc 240
tctggtgtcg acagaagaca gtcaaatgag gagagtgcg aggatacatc agctaagatt 300
tgggatacct atacaaacca ggctgatatt tttaacgttt ttggggcaca aacggataaa 360
cgctacttta aaccgacggt tctaaacttg aaatctaaac ctaactttgc gtcggcaaca 420
tatgaacctt tctcaaaagg tgattatgcg gcaattgcga ctcgtatcaa aaatgaattg 480
gcagaatttg agtttaatca agtacaaatt gactctttgt taaatctgtt cgaagcaacc 540
ctctcttttg tgcttcttc gactaatact aaagaaatcg ctgatatttc acttgctgat 600
catagtcgtc tgacagcagc ttttgcctta gccatctatg attacttgga agacaaagggt 660

```

cgtcataact ataaggagga cttgtttact aaagcatcag ccttttatga ggaagaagct	720
tttctcctag ctagctttga cttatcaggg attcaagact ttatctataa tattaatatt	780
gcgacgaatg gtgctgctaa acaattgaag gctagatctt tatactctga ctttatgagc	840
gagtatatag cagacagttt acttgataaa ctaggcctca atcgggctaa tatgctctat	900
gtcgggtggg gacatgctta ctttgccta gccaaactg aaaaaacggt agaaacactc	960
gttcaatttg aaaaagattt caatcaattt ttattggcaa atttccaaac cagattatat	1020
gttgcccttg gttggggaag ctttgcggct aaggatatca tgagcgaact gaactcacct	1080
gaaagctata gacaggctta tcaaaaaggct agtcgcatga tttctgagaa aaaaatctca	1140
aggtatgatt atcaaaccct tatgttggtt aacaggggct gtaaatcttc tgaaagagag	1200
tgcgagattt gtcattccgt tgagaattta gttgcttatt atgaccaaaa agtgtgtgac	1260
atttgctgag gcttgatca attttctaaa gagattgcc atgaccattt cattatcact	1320
gaaaatgaag ggcttcctat tgggccgaac gcatgtctta aggggtgtgc atttgaaaag	1380
ctgagccaag aagctttttc ccgtgtctat gtcaaaaatg actataaggc tggtagagtt	1440
aaggcaaccc atgtttttgt tggagattac cagtatgatg aaatatacaa ttatgctgcc	1500
ttatctaaaa acgaaaatgg gttaggattt aaacgttttag ctgttgtagc tcttgacgtg	1560
gatgatttgg gacgagcctt tatggctggc ttctcccaac aaggaaatgg gcaatatagt	1620
actctatcac gctcagccac tttctctcga agcatgagtc ttttcttcaa ggtttatatt	1680
aaccagtttg ctagtataa gaagctctct atcatctatg ccggtgggga tgatgttttt	1740
gctattggct cttggcaaga tattattgcc tttactgttg aacttcgtga gaacttcatt	1800
aatggacaa atggaaaact aacactatca gctggtatcg gtctgtttgc tgataagacc	1860
cctattagct taatggcaca tcaaacaggg gagctagaag aaacagctaa aggcaatgag	1920
aaagatagta tttcactctt tagttccgac tataccttta aatttgatcg gtttatcact	1980
aatgtttacg acgataagtt agagcagatt cgctatttct ttaatcacca agatgaacga	2040
ggcaagaatt tcatttataa attgattgaa ttgcttcgaa attatgatcg tatgaatatg	2100
gcacgttttag cttattatatt aacacgactt gaagaattga cgcgtgaaac agacagggat	2160
aaatttaaaa catttaaaaa tttattctat tcttggtaca caaataagga tgataaggat	2220
agaaaagaag cagagttagc cttgcttctc tatactctatg agattagaaa ggattag	2277
<210> 484	
<211> 381	
<212> DNA	
<213> <i>Steptococcus thermophilus</i>	
<400> 484	
atgacaatct tgactgatga gaattacggt gatattgcag aaaaagcaat tctaaaacta	60
gaaagaaata ctaggaacag aaagaatcct gatgccttct ttcttacaac aagtaagctc	120
agaaacttgc tgagcttaac tagtacactt tttgatgaga gtaagggtcaa agaatatgat	180
gctctccttg atcgtattgc ttatttaaga gtacaatttg tctaccaagc aggtagagag	240
attgcagtaa aagatctgat agaaaaggct caaattcttg aggcctctta ggaaatcaaa	300
gatagagaga cacttcaaag attttgtaga tatatggaag cattagtagc ctatttcaag	360
ttttatggag gtaaaagatta a	381

<210> 485

<211> 663

<212> DNA

<213> *Streptococcus thermophilus*

<400> 485

```

atgacattcg ctaagattaa attttcagct caaattcggt tagagacagg cctccatatt      60
ggtggaagcg atgcttttgc agccattggg gcaatcgatt cgcttggtat taaagatcct    120
attaccaacc taccgatcat tcctgggtca agtctcaaag gaaaaatgag aacgcttctt    180
gccaagggtt ataatgaaaa ggtagctgag aaaccaagcg atgacagtga tattcttagc    240
cgttttattg ggaatagtaa agataaacga ttcaaaatgg gacgcttgat ttttcgtgat    300
gccttcttgt caaacgctga tgagctagac tctcttgggg taagaagtta tacagaagta    360
aaatttgaaa atacaattga ccgtatcact gccgaagcta atccaagaca aattgaacgt    420
gctattcgta ccagtacttt tgatttcgag ttgatttatg aaattacaga tgagaatgaa    480
aatcaagtcg aagaagatgt caaagtgatt cgagatgggt taaaactgct tgaacttgat    540
tatcttggtg gttctggatc tcgaggttac ggtaagggtg cttttgaaaa actcaaagct    600
actaccgtat ttggttaatta tgatgttaaa acattaaatg aacttttaac tgcggagggtc    660
taa

```

<210> 486

<211> 900

<212> DNA

<213> *Streptococcus thermophilus*

<400> 486

```

atgacctata aactgtatat tatgaccttt cagaatgctc attttggttc gggcactcct      60
gatagctcaa aattaacatt ctacagcagac cgtatcttct cagcactagt gctagaatcc    120
ctaaaaatgg gaaaactcga tgcatttctt gcggaagcta accaagacaa gttcacgctc    180
acagatgcct ttccatttca attttggtccc tttttgccga aaccgattgg ttatcccaaa    240
catgaccaa tagatcaatc agttgatgtc aaagagggtc gccgtcaagc aaaattgtct    300
aagaaactgc aatttcttgc tctagaaaat gttgacgatt atctcaatgg agagttatgt    360
gaaaatgaag agcatgcagt catcgatact gtgacaaaaa atcaaccaca taaggacggc    420
aatctttatc aggtagctac aaccagattt tcaaatgata cgtcgcttta cgtcatcgca    480
aacgaatctg atttgcttaa tgagttgatg tctagtcttc agtattcagg tcttggtgga    540
aagcgttcaa gtggttttgg tcgttttgag ttagatatcc aaaatatccc actagaattg    600
tcagatagac tgactaagaa tcattcagat aaagtgatga gtcttacgac agcacttcct    660
gtagatgctg acctgaaga agcaatggaa gatggacatt acttattaac taaatcaagt    720
ggttttgcat ttagtcatgc taccaatgag aattatcgta agcaggatct ttacaaattt    780
gcttctggtt caacttttag taaaacattt gaaggtcaga ttgttgatgt gagaccactt    840
gatttccctc atgctgtttt aaattatgct aaaccactct tctttaaatt ggaggtataa    900

```

<210> 487

<211> 1074

<212> DNA

<213> *Streptococcus thermophilus*

<400> 487

```

atgaaaaatg actatagaac atttaaatta agcctcctga cacttgctcc aattcatatt      60
ggtaatggag agaagtatac ctctagagaa tttatctatg aaaataagaa gttttacttt      120
cctgacatgg ggaatttcta taataaaatg gtggagaaga ggcttgctga aaagtttgaa      180
gcattttctaa ttcaaactcg tccaaatgca cgtaataatc gtcttatttc cttcttaaat      240
gataaccgaa ttgcagagcg ttcttttgga ggttatagta tctctgaaac aggtttagaa      300
tcggacaaaa atcctgattc agccggagct attaacgaag ttaataaatt tattcgagat      360
gcttttgga atccctacat tcctggtagc tcactaaaag gtgctattcg taccatttta      420
atgaatacta cccctaagtg gaataatgaa aatgctgtaa atgactttgg aagatttccg      480
aaagagaata agaaccttat cccttgggga caaaaaagg gaaaagaata cgatgatttg      540
tttaacgcaa ttcgtgtgag tgatagtaag ctttttgata ataagagtct tatcttagtg      600
cagaaatggg attattcagc gaaaacaaat aaagctaaac cacttccctt gtatagagaa      660
tcaatctctc cattaacaaa aattgaattt gagattacaa caaccactga tgaagctgga      720
agattgattg aagaattagg taagagagca caagcgtttt ataaagacta taaggcattt      780
ttcctatctg aatttcctga tgataagatt caagccaatc tacaataccc aattttattta      840
gggtgcggga gcggtgcttg gacaaagact ctatttaagc aagctgatgg tattttacaa      900
agacgataca gtcgaatgaa aactaaaatg gttaaaaaag gagttcttaa gctcacaaaa      960

gcacctctta aaacagttaa gattccatct ggtaatcatt cattagtcaa gaaccacgag      1020
tccttttatg aaatgggaaa agctaatttc atgattaagg agattgataa atga          1074

```

<210> 488

<211> 5974

<212> DNA

<213> *Streptococcus agalactiae*

<400> 488

```

atgaataagc catattcaat aggccttgac atcggtagta attccgtcgg atggagcatt      60
attacagatg attataaagt acctgctaag aagatgagag ttttagggaa cactgataaa      120
gaatatatta agaagaatct cataggtgct ctgctttttg atggcgggaa tactgctgca      180
gatagacgct tgaagcgaac tgctcgctct cgttatacac gtcgtagaaa tcgtattcta      240
tattttacaag aaatttttgc agaggaaatg agtaaagttg atgatagttt ctttcacgca      300
ttagaggatt cttttctagt tgaggaagat aagagaggga gcaagtatcc tatctttgca      360
acattgcagg aagagaaaga ttatcatgaa aaattttcga caatctatca tttagaaaaa      420
gaattagctg acaagaaaga aaaagcagac cttcgtctta tttatattgc tctagctcat      480
atcattaaat ttagagggca tttcctaatt gaggatgata gctttgatgt caggaatata      540
gacatttcaa aacaatatca agatttttta gaaatcttta atacaacttt tgaaaataat      600
gatttggtat ctcaaacgt tgacgtagag gcaatactaa cagataagat tagcaagtct      660
gcgaagaaag atcgtatttt agcgcagtat cctaaccaa aatctactgg catTTTTGCA      720
gaatttttga aattgattgt cggaatcaa gctgacttca agaaatattt caatttggag      780
gataaaacgc cgcttcaatt cgctaaggat agctacgatg aagattttaga aaatcttctt      840

```

ggacagattg gtgatgaatt tgcagactta ttctcagcag cgaaaaagtt atatgatagt 900
 gtccttttgt ctggcattct tacagtaate gacctcagta ccaaggcgcc actttcagct 960
 tctatgattc agcgttatga tgaacataga gaggacttga aacagttaaa acaattcgta 1020
 aaagcttcat tgccggaaaa atatcaagaa atatttgctg attcatcaaa agatggctac 1080
 gctggttata ttgaaggtaa aactaatcaa gaagcttttt ataaatacct gtcaaaattg 1140
 ttgaccaagc aagaagatag cgagaatttt cttgaaaaaa tcaagaatga agatttcttg 1200
 agaaaacaaa ggacctttga taatggctca attccacacc aagtccattt gacagagctg 1260
 aaagctatta tccgccgtca atcagaatac tatcccttct tgaaagagaa tcaagatagg 1320
 attgaaaaaa tccttacctt tagaattcct tattatatcg ggccactagc acgtgagaag 1380
 agtgattttg catggatgac tcgcaaaaca gatgacagta ttcgaccttg gaattttgaa 1440
 gacttggttg ataaagaaaa atctgcggaa gcttttatcc atcgtatgac caacaatgat 1500
 ttttatcttc ctgaagaaaa agttttacca aagcatagtc ttatttatga aaaatttacg 1560
 gtctataatg agttgactaa ggtagatat aaaaatgagc aaggtgagac ttattttttt 1620
 gatagcaata ttaaacaaga aatctttgat ggagtattca aggaacatcg taaggatatcc 1680

aagaagaagt tgctagattt tctggctaaa gaatatgagg agtttaggat agtagatgtt 1740
 attggtctag ataaagaaaa taaagctttc aacgcctcat tgggaactta ccacgatctc 1800
 gaaaaaatac tagacaaaga ttttctagat aatccagata atgagtctat tctggaagat 1860
 atcgtccaaa ctctaactt atttgaagac agagaaatga ttaagaagcg tcttgaaaac 1920
 tataaagatc tttttacaga gtcacaacta aaaaaactct atcgtcgtca ctatactggc 1980
 tggggacgat tgtctgctaa gttaatcaat ggtattcgag ataaagagag tcaaaaaaca 2040
 atcttgact atcttattga tgatggtaga tctaactcga actttatgca gttgataaat 2100
 gatgatggtc tatctttcaa atcaattatc agtaaggcac aggctggtag tcattcagat 2160
 aatctaaaag aagttgtagg tgagcttgca ggtagccctg ctattaaaaa gggaattcta 2220
 caaagtttga aaattgttga tgagcttggt aaagtcattg gatacgaacc tgaacaaatt 2280
 gtggttgaga tggcgctga gaatcaaaca acaaatcaag gtcgtcgtaa ctctcgacaa 2340
 cgctataaac ttcttgatga tggcgtaag aatctagcta gtgacttgaa tggcaatatt 2400
 ttgaagaat atcctacgga taatcaagcg ttgcaaatg aaagactttt cctttactac 2460
 ttacaaaacg gaagagatat gtatacaggg gaagctctag atattgacaa ttttaagtcaa 2520
 tatgatattg accacattat tcctcaagct ttcataaaag atgattctat tgataatcgt 2580
 gttttggtat catctgctaa aaatcgtgga aagtcagatg atgttcctag ctttgaaatt 2640
 gtaaaagatt gtaaaagttt ctggaaaaaa ttacttgatg ctaagttaat gagtcagcgt 2700
 aagtatgata atttgactaa ggcagagcgc ggaggcctaa cttccgatga taaggcaaga 2760
 tttatccaac gtcagttggt tgagacacga caaattacca agcatgttgc ccgtatcttg 2820
 gatgaacgct ttaataatga gcttgatagt aaaggtagaa ggatccgcaa agttaaaatt 2880
 gtaaccttga agtcaaat tggtttcaaat ttccgaaaag aatttgatt ctataaaatt 2940
 cgtgaagtta acaattatca ccatgcacat gatgcctatc ttaatgcagt agttgctaaa 3000
 gctattctaa ccaaatatcc tcagttagag ccagaatttg tctacggcga ctatccaaaa 3060
 tataatagtt acaaaacgcg taaatccgct acagaaaagc tatttttcta ttcaaatatt 3120
 atgaacttct ttaaaactaa ggtaacttta gcggatggaa ccgttggtgt aaaagatgat 3180
 attgaagtta ataatgatac gggtgaaatt gtttgggata aaaagaaaca ctttgcgaca 3240
 gttagaaaag tcttgcata ccttcagaac aatatcgtga agaagacaga gattcagaca 3300

```

ggtggtttct ctaaggaatc aatcttggcg catggttaact cagataagtt gattccaaga 3360
aaaaaagagg atattttattt agatcctaag aaatatggag gttttgatag tccgatagta 3420
gcttactctg ttttagttgt agctgataac aaaaagggtg aagcacaaaa actaaaaaca 3480
gttacggaac ttttaggaat taccatcatg gagagggtcca gatttgagaa aaatccatca 3540
gctttccttg aatcaaaagg ctattttaat attagggtcg ataaactaat tattttgccc 3600
aagtatagtc tgttcgaatt agaaaatggg cgctgcgat tacttgctag tgctggtgaa 3660
ttacaaaaag gtaatgagct agccttacca acacaattta tgaagttctt ataccttgca 3720

agtcgttata atgagtcaaa aggtaaacca gaggagattg agaagaaaca agaatttgta 3780
aatcaacatg tctcttattt tgatgacatc cttcaattaa ttaatgattt ttcaaaacga 3840
gttattctag cagatgctaa tttagagaaa atcaataagc tttaccaaga taataaggaa 3900
aatatatcag tagatgaact tgctaataat attatcaatc tatttacttt taccagtcta 3960
ggagctccag cagcttttaa attttttgat aaaatagttg atagaaaacg ctatacatca 4020
actaaagaag tacttaattc taccctaatt catcaatcta ttactggact ttatgaaaca 4080
cgtattgatt tgggtaagtt aggagaagat tgatatggca ggttggcgaa ccgttgttgt 4140
aaatacacat tctaagctct cttataaaaa taatcatctg atttttaaag attcttatca 4200
gacggaaatg attcatctat cagagattga cattctaadc atggaacaaa cagatatcgt 4260
tttgctgacc atgctgatta aacgtttggt tgatgaaaat attttagtta tattttgtga 4320
cgataaacgc ttgccaacag ctatgttaat gccgtactat gccagacatg attcgagttt 4380
acaattatct aggcagatgt catggattga agatgtcaaa gcagatgttt ggacatcaat 4440
tattgcacaa aaaattttga atcagtcctt ttatctcggg gagtgttctt tctttgaaaa 4500
atcccagttt attatgaatc tctaccatga cttagaacct tttgatcctt ctaatcgtga 4560
ggggcatgct gctaggattt atttcaatac actttttgga aatgatattt caagagagca 4620
ggataatcca ataaatgctg gtttagacta cggatattca ttgcttttga gtatgtttgc 4680
gcgtgaagtt gtttaagtgt gttgcatgac acaatttggc ttgaagcatg ctaatcaatt 4740
taatcagttc aacctagcaa gcgatattat ggaaccattt cgcccaatcg ttgataggat 4800
tatttatgaa aataggcaga gtgattttgt caaaatgaaa agagaactct tttctatgtt 4860
ttcagagaca tacagctaca atggtaaaga aatgtatctc tcaaattatt tcagcgacta 4920
taccaaaaaa gttattaagt cgctaaatag tgatgggaat ggaattccgg agtttaggat 4980
atgagttatc ggtatatgag aatgatattt atgtttgata tgcctactga aacagcagaa 5040
gaacggaagg cgtatcgtaa gtttagaaag tttctcttga gcgaaggctt tatcatgcat 5100
cagttttctg tttatagtaa attattactc aataatacag ctaataatgc tatgatagg 5160
cggtttaaag tgaataatcc taaaaagggt aatatcacac tcttaacagt tacggaaaaa 5220
caatttgcca gaatgggtta cctccatgga gaacgcaaca caagtgttgc caactctgat 5280
agtcgcttgg ttttctagg agattcttat gatcaagatt aattttccaa ttttagatga 5340
accattagtg ttaagtaatg ctacgatatt aacgatagaa gatgtttcag tttattcttc 5400
attggtgaaa catttttatc aatatgacgt agatgaacat ttgaaattat ttgatgataa 5460
gcagaaaagt ctgaaggcaa cagagttaat gctggttaca gatatcttag gatacagatg 5520
caactcagca cctattctaa agttgatata tgggtgactta gaaaatcaat tcaacgaaaa 5580
gccagaagtg aaatcaatgg tagaaaaatt agcagctact attacagaac ttatcgcat 5640
tgagtgtcta gagaatgagc ttgatttaga atacgatgaa attaagattt tagaactcat 5700
taaggcactc gaagtcaaaa ttgaagaca gaggacact atctttgaaa aatottttga 5760

```

aattatacaa gtttaccatt atttaacgaa aaagaatctc ttggtttttg ttaatagcgg 5820
 agcttatctt accaaagatg aagttataaa attatgtgaa tacatcaatt taatgcaaaa 5880
 gtcagtactc tttctagaac ctagaagact ctatgattta ccgcaatatg ttattgataa 5940
 ggattatttc ttgataggcg aaaatatggg ataa 5974

<210> 489
 <211> 4113
 <212> DNA
 <213> *Streptococcus agalactiae*

<400> 489
 atgaataagc catattcaat aggccttgac atcggtagta attccgtcgg atggagcatt 60
 attacagatg attataaagt acctgctaag aagatgagag ttttagggaa cactgataaa 120
 gaatatatta agaagaatct cataggtgct ctgctttttg atggcgggaa tactgctgca 180
 gatagacgct tgaagcgaac tgctcgtcgt cgttatacac gtcgtagaaa tcgtattcta 240
 tatttacaag aaatttttgc agaggaaatg agtaaagttg atgatagttt ctttcatcga 300
 ttagaggatt cttttctagt tgaggaagat aagagagggg gcaagtatcc tatctttgca 360
 acattgcagg aagagaaaga ttatcatgaa aaattttcga caatctatca ttgagaaaa 420
 gaattagctg acaagaaaga aaaagcagac cttcgtctta tttatattgc tctagctcat 480
 atcattaaat ttagagggca tttcctaatt gaggatgata gctttgatgt caggaataca 540
 gacatttcaa aacaatatca agatttttta gaaatcttta atacaacttt tgaaaataat 600
 gatttgttat ctcaaaacgt tgacgtagag gcaatactaa cagataagat tagcaagtct 660
 gcgaagaaag atcgtatttt agcgcagtat cctaaccaaa aatctactgg catttttgca 720
 gaatttttga aattgattgt cggaatcaa gctgacttca agaaatattt caatttggag 780
 gataaaacgc cgcttcaatt cgctaaggat agctacgatg aagatttaga aaatcttctt 840
 ggacagattg gtgatgaatt tgcagactta ttctcagcag cgaaaaagtt atatgatagt 900
 gtccttttgt ctggcattct tacagtaatc gacctcagta ccaaggcgcc actttcagct 960
 tctatgatcc agcgttatga tgaacataga gaggacttga aacagttaaa acaattcgta 1020
 aaagcttcat tgccggaaaa atatcaagaa atatttgctg attcatcaaa agatggctac 1080
 gctgggtata ttgaaggtaa aactaatcaa gaagcttttt ataaatacct gtcaaaattg 1140
 ttgaccaagc aagaagatag cgagaatttt cttgaaaaaa tcaagaatga agatttcttg 1200
 agaaaacaaa ggacctttga taatggctca attccacacc aagtccattt gacagagctg 1260
 aaagctatta tccgccgtca atcagaatac tatcccttct tgaaagagaa tcaagatagg 1320
 attgaaaaaa tccttacctt tagaattcct tattatatcg ggccactagc acgtgagaag 1380
 agtgattttg catggatgac tcgcaaaaca gatgacagta ttcgaccttg gaattttgaa 1440
 gacttggttg ataaagaaaa atctgcggaa gcttttatcc atcgtatgac caacaatgat 1500
 ttttatcttc ctgaagaaaa agtttiacca aagcatagtc ttatttatga aaaatttacg 1560
 gtctataatg agttgactaa ggttagatat aaaaatgagc aaggtgagac ttattttttt 1620

gatagcaata ttaaacaaga aatctttgat ggagtattca aggaacatcg taaggtatcc 1680
 aagaagaagt tgctagattt tctggctaaa gaatatgagg agtttaggat agtagatgtt 1740
 attggtctag ataaagaaaa taaagctttc aacgcctcat tgggaactta ccacgatctc 1800
 gaaaaaatac tagacaaaga ttttctagat aatccagata atgagtctat tctggaagat 1860
 atcgtccaaa ctctaacatt atttgaagac agagaaatga ttaagaagcg tcttgaaaac 1920
 tataaagatc tttttacaga gtcacaacta aaaaaactct atcgtcgtca ctatactggc 1980

tggggacgat	tgtctgctaa	gttaatcaat	ggtattcgag	ataaagagag	tcaaaaaaca	2040
atcttggact	atctttattga	tgatggtaga	tctaactgca	actttatgca	gttgataaat	2100
gatgatggtc	tatcttttcaa	atcaattatc	agtaaggcac	aggctggtag	tcattcagat	2160
aatctaaaag	aagttgtagg	tgagcttgca	ggtagccctg	ctattaaaaa	gggaattcta	2220
caaagtttga	aaattgttga	tgagcttggt	aaagtcattg	gatacgaacc	tgaacaaatt	2280
gtggttgaga	tggcgcgtga	gaatcaaaca	acaaatcaag	gtcgtcgtaa	ctctcgacaa	2340
cgctataaac	ttcttgatga	tggcgttaag	aatctagcta	gtgacttgaa	tggcaatatt	2400
ttgaaagaat	atcctacgga	taatcaagcg	ttgcaaaatg	aaagactttt	cctttactac	2460
ttacaaaacg	gaagagatat	gtatacaggg	gaagctctag	atattgacaa	tttaagtcaa	2520
tatgatattg	accacattat	tcctcaagct	ttcataaaag	atgattctat	tgataatcgt	2580
gttttggtat	catctgctaa	aaatcggtga	aagtcagatg	atgttcctag	ccttgaaatt	2640
gtaaaagatt	gtaaagtttt	ctggaaaaaa	ttacttgatg	ctaagttaat	gagtcagcgt	2700
aagtatgata	atttgactaa	ggcagagcgc	ggaggcctaa	cttccgatga	taaggcaaga	2760
tttatccaac	gtcagttggt	tgagacacga	caaattacca	agcatgttgc	ccgtatcttg	2820
gatgaacgct	tttaataatga	gcttgatagt	aaaggtagaa	ggatccgcaa	agttaaaatt	2880
gtaaccttga	agtcaaattt	ggtttcaaatt	ttccgaaaag	aatttggtatt	ctataaaatt	2940
cgtgaagtta	acaattatca	ccatgcacat	gatgcctatc	ttaatgcagt	agttgctaaa	3000
gctattctaa	ccaaatatcc	tcagtttagag	ccagaatttg	tctacggcga	ctatccaaaa	3060
tataatagtt	acaaaacgcg	taaatccgct	acagaaaagc	tatttttcta	ttcaaattatt	3120
atgaacttct	ttaaaactaa	ggtaacttta	gcggatggaa	ccgttggtgt	aaaagatgat	3180
attgaagtta	ataatgatac	gggtgaaatt	gtttgggata	aaaagaaaca	ctttgcgaca	3240
gttagaaaag	tcttgtcata	ccctcagaac	aatatcgtga	agaagacaga	gattcagaca	3300
ggtggtttct	ctaaggaatc	aatcttggcg	catggtaact	cagataagtt	gattccaaga	3360
aaaacgaagg	atattttatt	agatcctaag	aaatatggag	gttttgatag	tccgatagta	3420
gcttactctg	tttttagttgt	agctgatatc	aaaaagggtg	aagcacaaaa	actaaaaaca	3480
gttacggaac	tttttaggaat	taccatcatg	gagaggtcca	gatttgagaa	aatccatca	3540
gctttccttg	aatcaaaaagg	ctattttaa	attagggctg	ataaactaat	tattttgccc	3600
aagtatagtc	tgttcgaatt	agaaaaatggg	cgtcgtcgat	tacttgctag	tgctggtgaa	3660
ttacaaaaag	gtaatgagct	agccttacca	acacaattta	tgaagttcct	ataccttgca	3720
agtcgttata	atgagtcaaa	aggtaaacca	gaggagattg	agaagaaaca	agaatttgta	3780
aatcaacatg	tctcttattt	tgatgacatc	cttcaattaa	ttaatgattt	ttcaaaacga	3840
gttattctag	cagatgctaa	tttagagaaa	atcaataagc	tttaccaaga	taataaggaa	3900
aatatatacag	tagatgaact	tgctaataat	attatcaatc	tatttacttt	taccagtcta	3960
ggagctccag	cagcttttaa	attttttgat	aaaatagttg	atagaaaacg	ctatacatca	4020
actaaagaag	tacttaattc	taccctaatt	catcaatcta	ttactggact	ttatgaaaca	4080
cgtattgatt	tgggtaagtt	aggagaagat	tga			4113

<210> 490

<211> 870

<212> DNA

<213> *Streptococcus agalactiae*

<400> 490

```

atggcaggtt ggcgaaccgt tgttgtaaat acacattcta agctctctta taaaaataat    60
catctgattt ttaaagattc ttatcagacg gaaatgattc atctatcaga gattgacatt    120
ctaatacatg aaacaacaga tatcgttttg tcgaccatgc tgattaaacg tttgggtgat    180
gaaaatatat tagttatatt ttgtgacgat aaacgcttgc caacagctat gttaatgccg    240
tactatgcca gacatgattc gagtttaca ttatctaggc agatgtcatg gattgaagat    300
gtcaaagcag atgtttggac atcaattatt gcacaaaaaa ttttgaatca gtctttttat    360
ctcggtgagt gttctttctt tgaaaaatcc cagtctatta tgaatctcta ccatgactta    420
gaaccttttg atccttctaa tcgtgagggg catgctgcta ggattttatt caatacactt    480
tttggaatg atttttcaag agagcaggat aatccaataa atgctgggtt agactacgga    540
tattcattgc ttttgagtat gtttgcgcgt gaagttgta agtgtgggtg catgacacaa    600
tttggttga agcatgctaa tcaatttaac cagtccaacc tagcaagcga tattatggaa    660
ccatttcgcc caatcgttga taggattatt tatgaaaata ggcagagtga ttttgtcaaa    720
atgaaaagag aactcttttc tatgttttca gagacataca gctacaatgg taaagaaatg    780
tatctctcaa atattgtcag cgactatacc aaaaaagtta ttaagtcgct aaatagtgat    840
gggaatggaa ttccggaggt taggatatga    870

```

<210> 491

<211> 327

<212> DNA

<213> *Streptococcus agalactiae*

<400> 491

```

atgcgaatga ttttaatggt tgatatgcct actgaaacag cagaagaacg gaaggcgat    60
cgtaagttaa gaaagtttct cttgagcgaa ggctttatca tgcatacagt ttctgtttat    120
agtaaattat tactcaataa tacagctaata atgctatga taggtcggct taaagtgaat    180
aatcctaaaa agggtaatat cacactctta acagttacgg aaaaacaatt tgcgagaatg    240

gtttacctcc atggagaacg caacacaagt gttgccaact ctgatagtcg cttgggtttc    300
ctaggagatt cttatgatca agattaa    327

```

<210> 492

<211> 666

<212> DNA

<213> *Streptococcus agalactiae*

<400> 492

```

atgatcaaga ttaattttcc aatttttagat gaaccattag tgttaagtaa tgctacgatt    60
ttaacgatag aagatgtttc agttttattct tcattgggtga aacattttta tcaatatgac    120
gtagatgaac atttgaaatt atttgatgat aagcagaaaa gtctgaaggc aacagagtta    180
atgctgggtta cagatatctt aggatacgat gtcaactcag cacctattct aaagttgata    240
catggtgact tagaaaatca attcaacgaa aagccagaag tgaaatcaat ggtagaaaaa    300
ttagcagcta ctattacaga acttatcgca tttgagtgtc tagagaatga gcttgattta    360
gaatacgatg aaattaagat tttagaactc attaaggcac tgggagtcaa aattgagaca    420

```

cagagcgaca ctatctttga aaaatgtttt gaaattatac aagtttacca ttatttaacg 480
 aaaaagaatc tcttggtttt tgtaaatagc ggagcttatc ttaccaaaga tgaagttata 540
 aaattatgtg aatacatcaa tttaatgcaa aagtcagtag tctttctaga acctagaaga 600
 ctctatgatt taccgcaata tgttattgat aaggattatt tcttgatagg cgaaaatatg 660
 gtataa 666

<210> 493

<211> 5995

<212> DNA

<213> *Streptococcus agalactiae*

<400> 493

atgaataagc catattcaat aggccttgac atcggtagta attccgtcgg atggagcatt 60
 attacagatg attataaagt acctgctaag aagatgagag ttttagggaa cactgataaa 120
 gaatatatta agaagaatct cataggtgct ctgctttttg atggcgggaa tactgctgca 180
 gatagacgct tgaagcgaac tgctcgtcgt cggtatacac gtcgtagaaa tcgtattcta 240
 tatttacaag aaatttttgc agaggaaatg agtaaagttg atgatagttt ctttcacgca 300
 ttagaggatt cttttctagt tgaggaagat aagagaggta gcaagtatcc tatctttgca 360
 acaatgcagg aggagaaata ttatcatgaa aaatttccga caatctatca tttagaaaaa 420
 gaattggctg acaagaaaga aaaagcagac cttcgtcttg tttatctggc tctagctcat 480
 atcattaaat tcagagggca tttcctaatt gaggatgata gatttgatgt gaggaatacc 540
 gatattcaaa aacaatatca agccttttta gaaatttttg atactacctt tgaaaataat 600
 catttgttat ctcaaaatgt agatgtagaa gcaattctaa cagataagat tagcaagtct 660
 gcgaagaagg atcgcatctt agcgcagtag cctaaccaaa aatctactgg ttttttgca 720
 gaatttttga aattgattgt cggaaatcaa gctgacttca agaaacattt caatttggag 780

gataaaacac cgcttcaatt cgctaaggat agctacgatg aagattttaga aaatcttctt 840
 ggacagattg gtgatgaatt tgcagactta ttctcagtag cgaaaaagct atatgatagt 900
 gttcttttat ctggcattct tacagtaact gatctcagta ccaaggcgcc actttctgcc 960
 tctatgattc agcgttatga tgaacatcat gaggacttaa agcatctaaa acaattcgta 1020
 aaagcttcat tacctgaaaa ttatcgggaa gtatttgctg attcatcaaa agatggctac 1080
 gctggctata ttgaaggcaa aactaatcaa gaagcttttt ataaatatct gttaaaattg 1140
 ttgaccaaac aagaaggtag cgagtatttt cttgagaaaa ttaagaatga agattttttg 1200
 agaaaacaga gaacctttga taatggctca atcccgcac aagtcattt gacagaattg 1260
 agggctatta ttcgacgtca atcagaatac tatccattct tgaaagagaa tcaagatagg 1320
 attgaaaaaa tccttacctt tagaattcct tattatgtcg ggccactagc acgtgagaag 1380
 agtgattttg catggatgac tcgcaaaaca gatgacagta ttcgaccttg gaattttgaa 1440
 gacttggttg ataaagaaaa atctgcggaa gcttttatcc atcgcatgac caacaatgac 1500
 ctctatcttc cagaagaaaa agttttacca aagcatagtc ttatttatga aaaatttact 1560
 gtttacaatg aattaacgaa gggttagattt ttggcagaag gcttttaaga ttttcaattt 1620
 ttaaatagga agcaaaaaga aactatcttt aacagcttgt ttaaggaaaa acgtaaagta 1680
 actgaaaagg atattattag ttttttgaat aaagttgatg gatatgaagg aattgcaatc 1740
 aaaggaattg aaaaacattt taacactaac ctttcaacct atcatgattt taaaaaata 1800

```

cttggcaagg atttccttga taatacagat aacgagctta ttttgaaga tatcgtccaa 1860
actctaacct tatttgaaga tagagaaatg attaagaagt gtcttgacat ctataaagat 1920
ttttttacag agtcacagct taaaaagctc tatcgccgtc actatactgg ctggggacga 1980
ttgtctgcta agctaataaa tggcatccga aataaagaga atcaaaaaac aatcttggac 2040
tatcttattg atgatggaag tgcaaaccca aacttcacgc agttgataaa tgatgatgat 2100
ctatcattta aaccaattat tgacaaggca cgaactggta gtcattcggg taatctgaaa 2160
gaagttgtag gtgaacttgc tggtagccct gctattaaaa aagggattct acaaagtittg 2220
aaaatagttg atgagctggg taaagtcacg ggctatgaac ctgaacaaat cgtgggttgaa 2280
atggcacgtg agaaccacaa gacagcaaaa ggattaagtc gttcacgaca acgcttgaca 2340
accttgagag aatctcttgc taatttgaag agtaatatct tggaagagaa aaagcctaag 2400
tatgtgaaaag atcaagttga aaatcatcat ttatctgatg accgtctttt cctttactac 2460
ttacaaaacg gaagagatat gtatacaaaa aaggctctgg atattgataa tttaagtcaa 2520
tatgatattg accacattat tcctcaagct ttcataaaaag atgattctat tgataatcgt 2580
gttttggtat catctgctaa aaatcgtgga aaatcagatg atgttcctag cattgaaatt 2640
gtaaaagctc gcaaaatggt ctggaaaaat ttactggatg ctaagttaat gagtcagcgt 2700
aagtatgata atttgactaa ggcagagcgc ggaggcctaa cttccgatga taaggcaaga 2760
tttatccaac gtcagttggg tgagactcga caaattacca agcatgtagc tcgtatcttg 2820

gatgaacgct tcaataatga agttgataat ggtaaaaaga tttgcaaggg taaaattgta 2880
accttgaagt caaatttggg ttcaaatttc cgaagaagaat ttggattcta taaaattcgt 2940
gaagttaatg attatcacca tgcacacgat gcttatctta atgcagtagt tgccaaagct 3000
attctaacca aatatccaca gttagagcca gagtttgtct acggaatgta tagacagaaa 3060
aaactttcga aaatcgttca tgaggataag gaagaaaaat atagtgaagc aaccaggaaa 3120
atgtttttct actccaactt gatgaatatg ttcaaaagag ttgtgagggt agcagatggt 3180
tctattgttg taagaccagt aatagaaact ggtagatata tgagaaaaac tgcattgggt 3240
aaaaagaaac actttgcgac agttagaaaa gtcttgtcat accctcagaa caatatcgtg 3300
aagaagacag agattcagac aggtggtttc tctaagggaat caatcttggc gcatggtaac 3360
tcagataagt tgattccaag aaaaacgaag gatatttatt tagatcctaa gaaatatgga 3420
ggttttgata gtccgatagt agcttactct gttttagttg tagctgatat caaaaaaggt 3480
aaagcacaaa aactaaaaac agttacggaa cttttaggaa ttaccatcat ggagaggtcc 3540
agatttgaga aaaatccatc agctttcctt gaatcaaaag gttattttaa tattagggac 3600
gataaattaa tgattttacc gaagtatatg ctgttcgaat tagaaaatgg gcgtcgtcga 3660
ttacttgcta gtgctggtga attacaaaaa ggtaacgagc tagccttacc aacacaattt 3720
atgaagttct tataccttgc aagtcgttat aatgagtcaa aaggtaaacc agaggagatt 3780
gagaagaac aagaatttgt aaatcaacat gtctcttatt ttgatgacat ctttcaatta 3840
attaatgatt tttcaaaacg agttattcta gcagatgcta atttagagaa aatcaataag 3900
ctttaccagg ataataagga aaatatacca gtagatgaac ttgctaataa tattatcaat 3960
ctatttactt ttaccagtct aggagctcca gcagctttta aattttttga taaaatagtt 4020
gatagaaaac gctatacatc aactaaagaa gtacttaatt ctactctaatt ccatcaatct 4080
attactggac tttatgaaac acgtattgat ttgggtaaat taggagaaga ttgatatggc 4140
aggttggcga actgttgttg taaatacaca ttctaagctc tcttataaaa ataataatct 4200

```

gattttttaa gattcttatt agacggaaat gattcatctt tcagagattg atattctaatt 4260
 catggaaacg acagatatgt ttttgtcgac tatgctgatt aaacgtttgg ttgatgaaaa 4320
 tatttttagtc atattttgtg atgataaacg cttgccaaaca gctatgttaa tgccgtacta 4380
 tgctagacat gattcgagtt tacaattatc taggcagatg tcatggattg aggatgtcaa 4440
 agcggatgtt tggacatcaa ttattgcaca aaaaattttg aatcagtcct tttatctcgg 4500
 tgagtgttct ttttttga aaatcccagtc tattatgaat ctctatcatg atttagaatc 4560
 ttttgaccct tccaatcgtg aagggtcatgc agctaggatt tatttcaata cactttttgg 4620
 aaatgatttt tcaagagagc aggataatcc aataaatgct ggttttagact atggatattc 4680
 tctgattttg agtatgtttg cgcgtgaagt tgtaagtgt ggttgcata cacaatttgg 4740
 cttaaagcat gctaataat ttaatcagtt caacctagca agcgatatta tggaaccatt 4800
 tcgccaatc gttgatagga ttatttatga aaataggcag agtgattttg tcaaaatgaa 4860

 aagagaactc ttttctatgt tttcagagac atacagctac aacggtaaag aaatgtatct 4920
 ttcaaatatt gtcagcgatt acacaaaaa agttattaag tcgctaaata gtgatgggaa 4980
 tggaattccg gagtttagga tatgagttat cggatatga gaatgatttt aatgtttgat 5040
 atgcctactg aaacagtaga agaacgtaag gcgtatcgta agtttagaaa gtttctgttg 5100
 agcgaagggt ttattatgca tcagttctct gtttatagta aattattgct caataatata 5160
 gctaataatg ccgatagtag tcggcttaaa gtgaataatc ctaagaaagg gagtataact 5220
 cttttgacag ttaccgagaa gcagtttgca aggatgggtt atctacatgg tgaacataat 5280
 atgagtgttg ccaactctga tagtcgcttg gttttcctag gagattctta tgatcaagat 5340
 taattttcca attttagatg aaccattagt gttaagtaat gctacgattt taacgataga 5400
 agatgtttca gtttattctt cattgggtgaa acatttttat caatatgacg tagatgaaca 5460
 tttgaaatta tttgatgata agcagaaaag tctgaaggca acggagttaa tgtttagttac 5520
 agatatctta ggatacagat tcaactcagc acctattcta aagttgatac atgggtgactt 5580
 agaaaatcaa ttcaacgaaa agccagaagt gaaatcaatg gtagaaaaat tagcagctac 5640
 tattacagaa cttatcgcat ttgagtgtct agagaatgag cttgatttag aatacagatga 5700
 aattacgatt ttagaactca ttaaggcact gggagtcaaa attgagacac agagcgacac 5760
 tatctttgaa aaatgttttg aaattataca agtttaccat tatttaacga aaaagaatct 5820
 cttagttttt gttaatatagc gagcttatct taccaaagat gaagttataa aattatgtga 5880
 atacatcaat ttaatgcaaa agtcagtact ctttctagaa cctagaagac tctatgattt 5940
 accgcaatat gttattgata aggattattt cttgataggc gaaaatatgg tataa 5995

<210> 494

<211> 4134

<212> DNA

<213> *Streptococcus agalactiae*

<400> 494

atgaataagc catattcaat aggccttgac atcggtagta attccgtcgg atggagcatt 60
 attacagatg attataaagt acctgctaag aagatgagag ttttagggaa cactgataaa 120
 gaatatatta agaagaatct cataggtgct ctgctttttg atggcgggaa tactgctgca 180
 gatagacgct tgaagcgaac tgctcgtcgt cgttatacac gtcgtagaaa tcgtattcta 240
 tatttacaag aaatttttgc agaggaaatg agtaaagttg atgatagttt ctttcatcga 300

ttagaggatt cttttctagt tgaggaagat aagagaggta gcaagtatcc tatctttgca 360
 acaatgcagg aggagaaata ttatcatgaa aaatttccga caatctatca tttgagaaaa 420
 gaattggctg acaagaaaga aaaagcagac cttcgtcttg tttatctggc tctagctcat 480
 atcattaaat tcagagggca tttcctaatt gaggatgata gatttgatgt gaggaatacc 540
 gatattcaaa aacaatatca agccttttta gaaatttttg atactacctt tgaaaataat 600
 catttgttat ctcaaaatgt agatgtagaa gcaattctaa cagataagat tagcaagtct 660

 gcgaagaagg atcgcatctt agcgcagtat cctaaccaaa aatcta[~]ctgg tatttttgca 720
 gaatttttga aattgattgt cggaatatca gctgacttca agaaacattt caatttgag 780
 gataaaacac cgcttcaatt cgctaaggat agctacgatg aagattttaga aaatcttctt 840
 ggacagattg gtgatgaatt tgcagactta ttctcagtag cgaaaaagct atatgatagt 900
 gttcttttat ctggcattct tacagtaact gatctcagta ccaaggcgcc actttctgcc 960
 tctatgattc agcgttatga tgaacatcat gaggacttaa agcatctaaa acaattcgta 1020
 aaagcttcat tacctgaaaa ttatcgggaa gtatttgctg attcatcaaa agatggctac 1080
 gctggctata ttgaaggcaa aactaatcaa gaagcttttt ataaatatct gttaaaattg 1140
 ttgaccaaac aagaaggtag cgagtatttt ctgagaaaa ttaagaatga agattttttg 1200
 agaaaacaga gaacctttga taatggctca atcccgcatc aagtccattt gacagaattg 1260
 agggctatta ttcgacgtca atcagaatac tatccattct tgaaagagaa tcaagatagg 1320
 attgaaaaaa tccttacctt tagaattcct tattatgtcg ggccactagc acgtgagaag 1380
 agtgattttg catggatgac tcgcaaaaca gatgacagta ttcgaccttg gaattttgaa 1440
 gacttggttg ataaagaaaa atctgcggaa gcttttatcc atcgcatgac caacaatgac 1500
 ctctatcttc cagaagaaaa agttttacca aagcatagtc ttattttatga aaaatttact 1560
 gtttacaatg aattaacgaa ggtagattt ttggcagaag gctttaaaga ttttcaattt 1620
 ttaaatagga agcaaaaaga aactatcttt aacagcttgt ttaaggaaaa acgtaaagta 1680
 actgaaaagg atattattag ttttttgaat aaagttagtg gatatgaagg aattgcaatc 1740
 aaaggaattg agaaacagtt taacgctagc ctttcaacct atcatgatct taaaaaata 1800
 cttggcaagg atttccttga taatacagat aacgagctta ttttgaaga tatcgtccaa 1860
 actctaacct tatttgaaga tagagaaatg attaagaagt gtcttgacat ctataaagat 1920
 ttttttacag agtcacagct taaaaagctc tatcgccgtc actatactgg ctggggacga 1980
 ttgtctgcta agctaataaa tggcatccga aataaagaga atcaaaaaac aatcttggac 2040
 tatcttattg atgatggaag tgcaaaccca aacttcatgc agttgataaa tgatgatgat 2100
 ctatcattta aaccaattat tgacaaggca cgaactggta gtcattcgga taatctgaaa 2160
 gaagttgtag gtgaacttgc tggtagccct gctattaaaa aagggttctt acaaagtttg 2220
 aaaatagttg atgagctggt taaagtcagtg ggctatgaac ctgaacaaat cgtgggtgaa 2280
 atggcacgtg agaaccaaac gacagcaaaa ggattaagtc gttcacgaca acgcttgaca 2340
 accttgagag aatctcttgc taatttgaag agtaatat[~]tt tggaaagagaa aaagcctaag 2400
 tatgtgaaag atcaagttga aaatcatcat ttatctgatg accgtctttt cctttactac 2460
 ttacaaaacg gaagagatat gtatacaaaa aaggctctgg atattgataa tttaagtcaa 2520
 tatgatattg accacattat tcctcaagct ttcataaaag atgattctat tgataatcgt 2580
 gttttggtat catctgctaa aaatcgtgga aaatcagatg atgttcctag cattgaaatt 2640
 gtaaaagctc gcaaaatgtt ctggaaaaat ttactggatg ctaagttaat gagtcagcgt 2700

```

aagtatgata atttgactaa ggcagagcgc ggaggcctaa cttccgatga taaggcaaga 2760
tttatccaac gtcagttggg tgagactcga caaattacca agcatgtagc tcgtatcttg 2820
gatgaacgct tcaataatga agttgataat ggtaaaaaga ttgcaagggt taaaattgta 2880
accttgaagt caaatttggg ttcaaatttc cgaaaagaat ttggattcta taaaattcgt 2940
gaagttaatg attatcacca tgcacacgat gcttatctta atgcagtagt tgccaaagct 3000
attctaacca aatatccaca gttagagcca gagtttgtct acggaatgta tagacagaaa 3060
aaactttcga aaatcgttca tgaggataag gaagaaaaat atagtgaagc aaccaggaaa 3120
atgtttttct actccaactt gatgaatatg ttcaaagag ttgtgaggtt agcagatggt 3180
tctattgttg taagaccagt aatagaaact ggtagatata tgagaaaaac tgcattggat 3240
aaaaagaac actttgcgac agttagaaaa gtcttgtcat accctcagaa caatatcgtg 3300
aagaagacag agattcagac aggtggtttc tctaaggaat caatcttggc gcatggtaac 3360
tcagataagt tgattccaag aaaaacgaag gatatttatt tagatcctaa gaaatatgga 3420
ggttttgata gtccgatagt agcttactct gttttagttg tagctgatat caaaaagggt 3480
aaagcacaaa aactaaaaac agttacggaa cttttaggaa ttaccatcat ggagagggtcc 3540
agatttgaga aaaatccatc agctttcctt gaatcaaaag gttattttaa tattagggac 3600
gataaattaa tgattttacc gaagtatagt ctgttcgaat tagaaaatgg gcgtcgtcga 3660
ttacttgcta gtgctggtga attacaaaaa ggtaacgagc tagccttacc aacacaattt 3720
atgaagtctt tataccttgc aagtcgttat aatgagtcaa aaggtaaacc agaggagatt 3780
gagaagaac aagaatttgt aaatcaacat gtctcttatt ttgatgacat cttcaatta 3840
attaatgatt tttcaaaacg agttattcta gcagatgcta atttagagaa aatcaataag 3900
ctttaccagg ataataagga aaatatacca gtagatgaac ttgctaataa tattatcaat 3960
ctatttactt ttaccagtct aggagctcca gcagctttta aattttttga taaaatagtt 4020
gatagaaaac gctatacatc aactaaagaa gtacttaatt ctactcta atccatcaatct 4080
attactggac tttatgaaac acgtattgat ttgggtaaat taggagaaga ttga 4134

```

<210> 495

<211> 870

<212> DNA

<213> *Streptococcus agalactiae*

<400> 495

```

atggcagggt ggcgaactgt tgttgtaaat acacattcta agctctctta taaaaataat 60
catctgattt ttaaagattc ttatcagacg gaaatgattc atctttcaga gattgatatt 120
ctaatacatg aaacgacaga tattgttttg tcgactatgc tgattaaacg tttggttgat 180
gaaaatattt tagtcatatt ttgtgatgat aaacgcttgc caacagctat gttaatgccg 240
tactatgcta gacatgattc gagtttacaa ttatctaggc agatgtcatg gattgaggat 300
gtcaaagcgg atgtttggac atcaattatt gcacaaaaaa ttttgaatca gtccttttat 360
ctcggtgagt gttctttctt tgaaaaatcc cagtctatta tgaatctcta tcatgattta 420

gaatcttttg acccttccaa tcgtgaagggt catgcagcta ggattttatt caatacactt 480
tttggaatg atttttcaag agagcaggat aatccaataa atgctggttt agactatgga 540
tattctctga ttttgagtat gtttgcgcgt gaagttgtta agtgtggttg catgacacaa 600
tttggttaa agcatgctaa tcaatttaat cagttcaacc tagcaagcga tattatggaa 660

```

ccatttcgcc caatcgttga taggattatt tatgaaaata ggcagagtga ttttgtcaaa	720
atgaaaagag aactcttttc tatgttttca gagacataca gctacaacgg taaagaaatg	780
tatctttcaa atattgtcag cgattacacc aaaaaagtta ttaagtcgct aaatagtgat	840
gggaatggaa ttccggagtt taggatatga	870

<210> 496

<211> 342

<212> DNA

<213> *Streptococcus agalactiae*

<400> 496

atgagttatc ggtatatgag aatgatttta atgtttgata tgcctactga aacagtagaa	60
gaacgtaagg cgtatcgtaa gtttagaaag tttctgttga gcgaagggtt tattatgcat	120
cagtttctctg tttatagtaa attattgctc aataatacag ctaataatgc catgataggt	180
cggcttaaag tgaataatcc taagaaaggg agtataactc ttttgacagt taccgagaag	240
cagtttgcaa ggatgggttta tctacatggt gaacataata tgagtgttgc caactctgat	300
agtcgcttgg ttttcctagg agattcctat gatcaagatt aa	342

<210> 497

<211> 666

<212> DNA

<213> *Streptococcus agalactiae*

<400> 497

atgatcaaga ttaattttcc aatttttagat gaaccattag tgtaagtaa tgctacgatt	60
ttaacgatag aagatgtttc agtttattct tcattggtga aacattttta tcaatatgac	120
gtagatgaac atttgaaatt atttgatgat aagcagaaaa gtctgaaggc aacggagtta	180
atgttagtta cagatatctt aggatacgat gtcaactcag cacctattct aaagttgata	240
catggtgact tagaaaatca attcaacgaa aagccagaag tgaaatcaat ggtagaaaaa	300
ttagcagcta ctattacaga acttategca tttgagtgtc tagagaatga gcttgattta	360
gaatacgatg aaattacgat tttagaactc attaaggcac tgggagtcaa aattgagaca	420
cagagcgaca ctatctttga aaaatgtttt gaaattatac aagtttacca ttatttaacg	480
aaaaagaatc tcttagtttt tgtaaatagc ggagcttatt ttaccaaaga tgaagttata	540
aaattatgtg aatacatcaa tttaatgcaa aagtcagtac tctttctaga acctagaaga	600
ctctatgatt taccgcaata tgttattgat aaggattatt tcttgatagg cgaaaatatg	660
gtataa	666

<210> 498

<211> 6580

<212> DNA

<213> *Streptococcus mutans*

<400> 498

atgaaaaaac cttactctat tggacttgat attggaacca attctgttgg ttgggctgtt	60
gtgacagatg actacaaagt tcctgctaag aagatgaagg ttctgggaaa tacagataaa	120

agtcatatcg agaaaaatth gcttggcgct ttattatttg atagcgggaa tactgcagaa 180
 gacagacggt taaagagaac tgctcgccgt cgttacacac gtcgcagaaa tcgtatttta 240
 tatttgcaag agattttttc agaagaaatg ggcaaggtag atgatatgtt ctttcatcgt 300
 ttagaggatt cttttcttgt tactgaggat aaacgaggag agcgccatcc catttttggg 360
 aatcttgaag aagaagttaa gtatcatgaa aattttccaa ccatttatca tttgcggcaa 420
 tatcttgcgg ataattccaga aaaagttgat ttgcgttttag ttattttggc tttggcacat 480
 ataattaagt ttagaggatca ttttttaatt gaaggaaagt ttgatacacg caataatgat 540
 gtacaaagac tgtttcaaga attttttagca gtctatgata atacttttga gaatagtctg 600
 cttcaggagc aaaatgttca agttgaagaa attctgactg ataaaatcag taaatctgct 660
 aagaaagata gagttttgaa actttttcct aatgaaaagt ctaatggccg ctttgcagaa 720
 tttctaaaac taattgttgg taatcaagct gattttaaaa agcattttga attagaagag 780
 aaagcaccat tgcaattttc taaagatact tatgaagaag agttagaagt actattagct 840
 caaattggag ataattacgc agagctcttt ttatcagcaa agaaactgta tgatagtatc 900
 cttttatcag ggattttaac agttactgat gttggtacca aagcgctttt atctgcttcg 960
 atgattcagc gatataatga acatcagatg gatttagctc agcttaaaca attcattcgt 1020
 cagaaattat cagataaata taacgaagtt ttttctgatg tttcaaaaga cggctatgcg 1080
 gggtatattg atgggaaaac aaatcaagaa gctttttata aataccttaa aggtctatta 1140
 aataagattg agggaagtgg ctatttcctt gataaaattg agcgtgaaga ttttctaaga 1200
 aagcaacgta cctttgacaa tggctctatt ccacatcaga ttcattctca agaaatgcgt 1260
 gctatcattc gtagacaggc tgaattttat ccgttttttag cagacaatca agataggatt 1320
 gagaaattat tgactttccg ttttccctac tatgttggtc cattagcgcg cggaaaaagt 1380
 gattttgctt ggttaagtgc gaaatcggtc gataaaatta caccatggaa ttttgatgaa 1440
 atcgttgata aagaatcctc tgcagaagct tttatcaatc gtatgacaaa ttatgatttg 1500
 tacttgccaa atcaaaaagt tcttcctaaa catagtttat tatacgaaaa atttactgtt 1560
 tacaatgaat taacaaaggt taaatataaa acagagcaag gaaaaacagc attttttgat 1620
 gccaatatga agcaagaaat ctttgatggc gtatttaagg tttatcgaaa agtaactaaa 1680
 gataaattaa tggatttcct tgaaaaagaa tttgatgaat ttcgtattgt tgatttaaca 1740
 ggtctggata aagaaaaata agtatttaac gcttcttatg gaacttatca tgatttgtgt 1800
 aaaatttttag ataaagattt tctcgataat tcaaagaatg aaaagatttt agaagatatt 1860

 gtgttgacct taacgttatt tgaagataga gaaatgatta gaaaacgtct agaaaattac 1920
 agtgatttat tgaccaaaaga acaagtgaag aagctggaaa gacgtcatta tactggttgg 1980
 ggaagattat cagctgagtt aattcatggt attcgcaata aagaaagcag aaaaacaatt 2040
 cttgattatc tcattgatga tggcaatagc aatcggaact ttatgcaact gattaacgat 2100
 gatgctcttt ctttcaaaga agagattgct aaggcacaag ttattggaga aacagacaat 2160
 ctaaatcaag ttgttagtga tattgctggc agccctgcta ttaaaaaagg aattttacaa 2220
 agcttgaaga ttgttgatga gcttgtcaaa attatgggac atcaacctga aaatatcgtc 2280
 gtggagatgg cgcgtgaaaa ccagtttacc aatcagggac gacgaaattc acagcaacgt 2340
 ttgaaagggt tgacagattc tattaaagaa tttggaagtc aaattcttaa agaacatccg 2400
 gttgagaatt cacagttaca aaatgataga ttgtttctat attatttaca aaacggcaga 2460
 gatatgtata ctggagaaga attggatatt gattatctaa gccagtatga tatagaccat 2520
 2580

attatcccg	aagcttttat	aaaggataat	tctattgata	atagagrat	gactagctca	2580
aaggaaaatc	gtggaaaatc	ggatgatgta	ccaagtaaag	atgttggtcg	taaaatgaaa	2640
tcctattgga	gtaagctact	ttcggcaaag	cttattacac	aacgtaaatt	tgataatttg	2700
acaaaagctg	aacgaggtgg	attgaccgac	gatgataaag	ctggattcat	caagcgtcaa	2760
ttagtagaaa	cacgacaaat	taccaaacat	gtagcacgta	ttctggacga	acgattttaa	2820
acagaaacag	atgaaaacaa	caagaaaatt	cgtaagtaa	aaattgtgac	cttgaaatca	2880
aatcttggtt	ccaatttccg	taaagagttt	gaactctaca	aagtgcgtga	aattaatgac	2940
tatcatcatg	cacatgatgc	ctatctcaat	gctgtaattg	gaaaggcttt	actaggtgtt	3000
taccacaaat	tggaacctga	atttggttat	ggtgattatc	ctcattttca	tggaacataa	3060
gaaaataaag	caactgctaa	gaaatttttc	tattcaaata	ttatgaactt	ctttaaaaaa	3120
gatgatgtcc	gtactgataa	aaatggtgaa	attatctgga	aaaaagatga	gcataattct	3180
aataattaaa	aagtgccttc	ttatccacaa	gtaaatattg	ttaagaaagt	agaggagcaa	3240
acgggaggat	tttctaaaga	atctatcttg	ccgaaaggta	attctgacaa	gcttattcct	3300
cgaaaaacga	agaaatttta	ttgggatacc	aagaaatatg	gaggatttga	tagcccgatt	3360
gttgcttatt	ctattttagt	tattgctgat	attgaaaaag	gtaaatctaa	aaaattgaaa	3420
acagtcaaag	ccttagttgg	tgctactatt	atggaaaaga	tgacttttga	aagggatcca	3480
gttgcttttc	ttgagcgaaa	aggctatcga	aatgttcaag	aagaaaatat	tataaagtta	3540
ccaaaatata	gtttatttaa	actagaaaac	ggacgaaaaa	ggctattggc	aagtgcctagg	3600
gaacttcaaa	agggaaatga	aatcgttttg	caaatacatt	taggaacctt	gctttatcac	3660
gctaaaaata	ttcataaagt	tgatgaacca	aagcatttgg	actatgttga	taaacataaa	3720
gatgaattta	aggagtgtct	agatgttgty	tcaaactttt	ctaaaaata	tacttttagca	3780
gaaggaaatt	tagaaaaaat	caaagaatta	tatgcacaaa	ataatggtga	agatcttaaa	3840
gaattagcaa	gttcatttat	caacttatta	acatttactg	ctataggagc	accggctact	3900
tttaaatctt	ttgataaaaa	tattgatcga	aaacgatata	cttcaaactac	tgaaattctc	3960
aacgctaccc	tcatccacca	atccatcacc	ggcttttatg	aaacgcggat	tgatctcaat	4020
aagttaggag	gagactaatg	ggctggcgga	cagtggttgt	taatacgcac	tccaagttgt	4080
cttataagaa	caaccacttg	atttttaaag	atgcttatca	gacagagatg	attcatctgt	4140
ctgagattga	catcttatta	cttgagacaa	cagatattgt	tttgtaact	atgctaata	4200
aacgcttggt	tgatgagaat	attttgggtca	ttttttgtga	tgacaaacgt	ctgccaacag	4260
ccatgctcat	gccttactat	gcgcgtcacg	attccagctt	gcagctgagt	catcagattt	4320
cttgacaga	agaagtgaag	tgcatgtctt	ggacaacaat	catcgtcaa	aagattttga	4380
atcagtcacg	ttatttgga	gaatgttttt	attttgaaaa	atctcagtc	attatggatt	4440
tatatcatga	cttagagcct	tttgacccta	gtaatcgaga	aggacattct	gcgcggattt	4500
atttcaatac	cttatttgga	aatgtttttt	ccagagaaca	agataatgat	attaatgcag	4560
gtcttgacta	tggttatacg	ctgctgttaa	gtatgtttgc	gcgtgaagtg	gttgatctcg	4620
gctgtatgac	acaatttggt	ctcaagcatg	ccaaccaatt	caatcagttt	aactttgcc	4680
gtgatattat	ggagcctttt	cgtccaattg	ttgaccgtat	tgtttatgaa	aatcgaaata	4740
actcttttat	taaaataaaa	cgtgagctat	tcagcatgtt	ttcagacacc	tatctttata	4800
ataataagga	gatgtatttg	acaaatattg	tcagcgatta	taccaaaaag	gtaatcaagg	4860
cgctgaataa	tgatgggaaa	ggagttcctg	agtttaggat	atgagttacc	gatatatgcg	4920
aatgatttta	atgtttgata	tgccaacaga	tactgctgag	gaacgcaaag	cttatcgtaa	4980

atttcggaaa tttttactga gcgaagggtt catcatgcat cagttttcag tatacagcaa 5040
 gctgcttttg aataactctg ccaatacagc catgattgcc cgcttgaagg agaataatcc 5100
 aaagaagggc aatatcacct tgttgaccgt gactgaaaag cagtttgccc gtatgattta 5160
 cctgaatggt gagcgtgata ctagcattgc taattcggat tcacgactgg tctttctagg 5220
 ggaggctttt cctgatgaaa cttaattttc ctatattgga tgaaccaata actcttgaaa 5280
 aatctacgat tttggtatta gaagatgtgc aagtttttgc tcaaatggtg agaaatcttt 5340
 atcaatatga tgaagatagt gaacttaaat tttttaatag aaaatttaag agtctgaaac 5400
 catctgagtt aatgcttgtg acagatatat taggttatga tgtcaatgcc ccgtccttgc 5460
 tgaagttggt tcacgctgat ttagaaaatc agtttaatga aaaaccagag gttaagtcta 5520
 tggttgaaaa actggcaaat accattacgg aattaattgc ttatgaatgt ttagaaaatg 5580
 aattggactt agaatatgat gagattacta ttttagagtt aatcaaagct ttaggcgtca 5640
 aaattgaaac acaaagtgat accatttttg aaaaaatgtt tgaagtcctt caagtttata 5700
 agtatctaaa taaaaagaag cttctcgttt ttatcaatac tttatcctat tttaaaagag 5760
 aagaaatcgc gcaaatctta gaatatattc acttatccga tatggttgtt ttatttcttg 5820
 aaccccgtaa aattgatggt ttigtcaat atattttaga tgaagattat ttcttgataa 5880
 cagaagcaa caactaaata cgaataataa gatagtttct aaatcagggg ctgtctttta 5940

 ttatggattg acaaatgcgt ataatgcgta taaaataaaa agagaaatgt tatttgccat 6000
 taacagggaa agaattagct aaattagcga taaacaatgg atgggaagaa gttcgggtga 6060
 gaggaagtca tcatcatctt aagaaagatg gagtatctta tattgtgacg attcctattc 6120
 atggaaataa agtgcttaaa attggtcttg aaaagaaact cttaagggat ttaaatttat 6180
 tatgatagag gaggaagtcg tcatgttaaa atcatatcct gtaatttttc ataaggaaga 6240
 ggaaggggat tgggttgaat ttcctgaatt tggcgggtgg acgcaagggg aagatttgga 6300
 agaagccatg aagaacgctc gtcagatgtt agaaagtgtg ttggcatctt atcttgatga 6360
 agggtttggt ctaccattt caagcgatat tcagaaaata tctgttgaag atggttttgc 6420
 gaccatgatt caagctgac ctagtcctta tctcaaaaat aacaaagcta ttcggaaaaa 6480
 tgttaccgtg cctgagtggt tgatacgatt agcagaccgt gaccgagtaa attattctga 6540
 agtattaaca aaggcttttg aaaagaaact acaattataa 6580

<210> 499

<211> 4038

<212> DNA

<213> *Streptococcus mutans*

<400> 499

atgaaaaaac ctactctat tggacttgat attggaacca attctgttgg ttgggctggt 60
 gtgacagatg actacaaagt tctgctaag aagatgaagg ttctgggaaa tacagataaa 120
 agtcatatcg agaaaaattt gcttggcgct ttattatttg atagcgggaa tactgcagaa 180
 gacagacggt taaagagaac tgctcgccgt cgttacacac gtcgcagaaa tcgtatttta 240
 tatttgcaag agattttttc agaagaaatg ggcaaggtag atgatagttt cttcatcgt 300
 ttagaggatt cttttcttgt tactgaggat aaacgaggag agcgccatcc catttttggg 360
 aatcttgaaag aagaagttaa gtatcatgaa aattttccaa ccatttatca tttgcggcaa 420
 tatcttgagg ataatccaga aaaagttgat ttgcgtttag tttatttggc tttggcacat 480

ataattaagt ttagagggtca ttttttaatt gaaggaaagt ttgatacacg caataatgat 540
 gtacaaagac tgtttcaaga attttttagca gtctatgata atacttttga gaatagtctg 600
 cttcaggagc aaaatgttca agttgaagaa attctgactg ataaaatcag taaatctgct 660
 aagaaagata gagttttgaa actttttcct aatgaaaagt ctaatggccg ctttgcagaa 720
 tttctaaaac taattgtttg taatcaagct gattttaaaa agcattttga attagaagag 780
 aaagcaccat tgcaattttc taaagatact tatgaagaag agttagaagt actattagct 840
 caaattggag ataattacgc agagctcttt ttatcagcaa agaaactgta tgatagtatc 900
 cttttatcag ggatttttaac agttactgat gttggtacca aagcgctttt atctgcttcg 960
 atgattcagc gatataatga acatcagatg gatttagctc agcttaaaca attcattcgt 1020
 cagaaattat cagataaata taacgaagtt ttttctgatg tttcaaaaga cggctatgcg 1080
 ggttatattg atgggaaaac aaatcaagaa gctttttata aataccttaa aggtctatta 1140
 aataagattg agggaagtgg ctatttcctt gataaaattg agcgtgaaga ttttctaaga 1200

aagcaacgta cttttgacaa tggctctatt ccacatcaga ttcattcttca agaaatgcgt 1260
 gctatcattc gtagacaggc tgaattttat ccgtttttag cagacaatca agataggatt 1320
 gagaaattat tgactttccg tattccctac tatgttggc cattagcgcg cggaaaaagt 1380
 gattttgctt gggttaagtcg gaaatcggct gataaaatta caccatggaa ttttgatgaa 1440
 atcgttgata aagaatcctc tgcagaagct tttatcaatc gtatgacaaa ttatgatttg 1500
 tacttgccaa atcaaaaagt tcttcctaaa catagtttat tatacgaaaa atttactgtt 1560
 tacaatgaat taacaaagggt taaatataaa acagagcaag gaaaaacagc attttttgat 1620
 gccaatatga agcaagaaat ctttgatggc gtattttaagg tttatcgaaa agtaactaaa 1680
 gataaattaa tggatttcct tgaaaaagaa tttgatgaat ttcgtattgt tgatttaaca 1740
 ggtctggata aagaaaataa agtatitaaac gcttcttatg gaacttatca tgatttgtgt 1800
 aaaatttttag ataaagattt tctcgataat tcaaagaatg aaaagatttt agaagatatt 1860
 gtgttgacct taacgttatt tgaagataga gaaatgatta gaaaacgtct agaaaattac 1920
 agtgatttat tgaccaaaga acaagtgaag aagctggaaa gacgtcatta tactggttg 1980
 ggaagattat cagctgagtt aattcatggt attcgcaata aagaaagcag aaaaacaatt 2040
 cttgattatc tcattgatga tggcaatagc aatcggaact ttatgcaact gattaacgat 2100
 gatgctcttt ctttcaaaga agagattgct aaggcacaag ttattggaga aacagacaat 2160
 ctaaatacag ttgttagtga tattgctggc agccctgcta ttaaaaaagg aatttttaca 2220
 agcttgaaga ttgttgatga gcttgctaaa attatgggac atcaacctga aaatatcgtc 2280
 gtggagatgg cgcgtgaaaa ccagtttacc aatcagggac gacgaaattc acagcaacgt 2340
 ttgaaagggt tgacagattc tatiaaagaa tttggaagtc aaattcttaa agaacatccg 2400
 gttgagaatt cacagttaca aaatgataga ttgtttctat attatttaca aaacggcaga 2460
 gatatgtata ctggagaaga attggatatt gattatctaa gccagtatga tatagaccat 2520
 attatcccgc aagcttttat aaaggataat tctattgata atagagtatt gactagctca 2580
 aaggaaaatc ttggaaaatc ggatgatgta ccaagtaaag atgttggttcg taaaatgaaa 2640
 tcctattgga gtaagctact ttcggcaaag cttattacac aacgtaaatt tgataatttg 2700
 acaaaaagctg aacgaggtgg attgaccgac gatgataaag ctggattcat caagcgtcaa 2760
 ttagtagaaa cagacaaat taccaaacat gtagcacgta ttctggacga acgatttaat 2820
 acagaaacag atgaaaacaa caagaaaatt cgtcaagtaa aaattgtgac cttgaaatca 2880

aatcttgttt ccaatttccg taaagagttt gaactctaca aagtgcgtga aattaatgac	2940
tatcatcatg cacatgatgc ctatctcaat gctgtaattg gaaaggcttt actagggtgtt	3000
taccacaat tggaaacctga atttgtttat ggtgattatc ctcattttca tggacataaa	3060
gaaaataaag caactgctaa gaaatttttc tattcaaata ttatgaactt ctttaaaaaa	3120
gatgatgtcc gtactgataa aaatggtgaa attatctgga aaaaagatga gcatatttct	3180
aatattaaaa aagtgccttc ttatccacaa gttaatattg ttaagaaagt agaggagcaa	3240
acgggaggat tttctaaaga atctatcttg ccgaaaggta attctgacaa gcttattcct	3300
cgaaaaacga agaaatttta ttgggatacc aagaaatatg gaggatttga tagcccgatt	3360
gttgcttatt ctattttagt tattgctgat attgaaaaag gttaatctaa aaaattgaaa	3420
acagtcaaag ccttagttgg tgtcactatt atggaaaaga tgacttttga aagggatcca	3480
gttgcttttc ttgagcgaaa aggctatcga aatgttcaag aagaaaatat tataaagtta	3540
ccaaatata gtttatttaa actagaaaac ggacgaaaaa ggctattggc aagtgcctagg	3600
gaacttcaaa agggaaatga aatcgttttg ccaaatacatt taggaacctt gctttatcac	3660
gctaaaaata ttcataaagt tgatgaacca aagcatttgg actatgttga taaacataaa	3720
gatgaattta aggagtgtct agatgttgtg tcaaactttt ctaaaaata tactttagca	3780
gaaggaaatt tagaaaaat caaagaatta tatgcacaaa ataatggtga agatcttaaa	3840
gaattagcaa gttcatttat caacttatta acatttactg ctataggagc accggctact	3900
tttaaatctt ttgataaaaa tattgatcga aaacgatata cttcaactac tgaaattctc	3960
aacgctaccc tcattccacca atccatcacc ggtctttatg aaacgcggat tgatctcaat	4020
aagttaggag gagactaa	4038

<210> 500

<211> 867

<212> DNA

<213> Streptococcus mutans

<400> 500

atgggctggc ggacagtggg tgtaataacg cattccaagt tgtcttataa gaacaaccac	60
ttgattttta aagatgctta tcagacagag atgattcatc tgtctgagat tgacatctta	120
ttacttgaga caacagatat tgttttgtca actatgctaa tcaaacgctt ggttgatgag	180
aatatttttg tcattttttg tgatgacaaa cgtctgccaa cagccatgct catgccttac	240
tatgcgcgtc acgattccag cttgcagctg agtcatcaga tttcttggac agaagaagtg	300
aatgcgatg tctggacaac aatcatcgct caaaagattt tgaatcagtc atgttatttg	360
ggagaatgtt tttattttga aaaatctcag tcaattatgg atttatatca tgacttagag	420
ccttttgacc ctagtaatcg agaaggacat tctgcgcgga tttatttcaa taccttattt	480
ggaaatgtt tttccagaga acaagataat gatattaatg caggcttga ctatggttat	540
acgctgctgt taagtatgtt tgcgcgtgaa gtggtgtat ctggctgtat gacacaattt	600
ggtctcaagc atgccaacca attcaatcag tttaactttg ccagtgatat tatggagcct	660
tttcgtccaa ttgttgaccg tattgtttat gaaaatcgaa ataactcttt tattaaaata	720
aaacgtgagc tattcagcat gttttcagac acctatcttt ataataataa ggagatgtat	780
ttgacaaata ttgtcagcga ttataccaaa aaggtaatca aggcgctgaa taatgatggg	840
aaaggagttc ctgagtttag gatatga	867

<210> 501

<211> 330

<212> DNA

<213> *Streptococcus mutans*

<400> 501

```

atgcgaatga ttttaatggt tgatatgccac acagatactg ctgaggaacg caaagcttat      60
cgtaaatttc ggaaattttt actgagcgaa ggtttcatca tgcattcagtt ttcagtatac      120
agcaagctgc ttttgaataa ctctgccaat acagccatga ttgcccgctt gaaggagaat      180
aatccaaaga agggcaatat caccttggtg accgtgactg aaaagcagtt tgcccgtatg      240
atttacctga atggtgagcg tgatactagc attgctaatt cggattcacg actggtcttt      300
ctaggggagg cttttcctga tgaaacttaa                                     330

```

<210> 502

<211> 573

<212> DNA

<213> *Streptococcus mutans*

<400> 502

```

atggtgagaa atctttatca atatgatgaa gatagtgaac tttaaattttt taatagaaaa      60
tttaagagtc tgaaaccatc tgagttaatg cttgtgacag atattttagg ttatgatgtc      120
aatgccccgt ccttgctgaa gttggttcac gctgatttag aaaatcagtt taatgaaaaa      180
ccagaggtta agtctatggt tgaaaaactg gcaaatacca ttacggaatt aattgcttat      240
gaatgttttag aaaatgaatt ggacttagaa tatgatgaga ttactatttt agagttaatc      300
aaagcttttag gcgtcaaaat tgaaacacaa agtgatacca tttttgaaaa aatgtttgaa      360
gtccttcaag tttataagta tctaaataaa aagaagcttc tcgtttttat caatacttta      420
tcctatttta aaagagaaga aatcgcgcaa attctagaat atattcactt atccgatatg      480
gttggtttat ttcttgaacc ccgtaaaatt gatgggtttg ctcaatatat tttagatgaa      540
gattatttct tgataacaga aagcaacaac taa                                     573

```

<210> 503

<211> 378

<212> DNA

<213> *Streptococcus mutans*

<400> 503

```

atgttaaaaa catatcctgt aatttttcat aaggaagagg aagggtattg ggttgaattt      60
cctgaatttg gcggtggtac gcaaggggaa gatttggaag aagccatgaa gaacgctcgt      120
cagatgttag aaagtgtggt ggcattctat cttgatgaag ggttggttct acccatttca      180
agcgatattc agaaaatata tgttgaagat ggttttgcca ccatgattca agctgacct      240
agtccttata tcaaaaataa caaagctatt cggaaaaatg ttaccgtgcc tgagtgggtg      300
atacgattag cagaccgtga ccgagtaaatt tattctgaag tattaacaaa ggcttttgaa      360
aagaaactac aattataa                                     378

```

<210> 504

<211> 5966

<212> DNA

<213> *steptococcus pyogenes*

<400> 504

atggataaga aatactcaat aggccttagat atcggcacaa atagcgtcgg atgggcggtg	60
atcactgatg aatataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc	120
cacagtatca aaaaaaatct tataggggct cttttatttg acagtggaga gacagcggaa	180
gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt	240
tatctacagg agattttttc aaatgagatg gcgaaagtag atgatagttt ctttcacgca	300
cttgaagagt cttttttggt ggaagaagac aagaagcatg aacgtcatcc ttttttggga	360
aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa	420
aaattggtag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgc	480
atgattaagt ttcgtggtca ttttttgatt gagggagatt taaatcctga taatagtgt	540
gtggacaaac tatttatcca gttggtacaa acctacaatc aattatttga agaaaaccct	600
attaacgcaa gtggagtaga tgctaaagcg attccttctg cacgattgag taaatcaaga	660
cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggctt atttgggaat	720
ctcattgctt tgtcattggg tttgaccctt aattttaaat caaattttga tttggcagaa	780
gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg	840
caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttatc agatgctatt	900
ttactttcag atatcctaag agtaaatact gaaataacta aggcctcccct atcagcttca	960
atgattaaac gctacgatga acatcatcaa gacttgactc ttttaaagc tttagtctga	1020
caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca	1080
ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta	1140
gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc	1200
aagcaacgga cctttgacaa cggctctatt ccccatcaaa ttcacttggg tgagctgcat	1260
gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt	1320
gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt	1380
cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa	1440
gttgctgata aagggtgctt agctcaatca tttattgaac gcatgacaaa ctttgataaa	1500
aatcttccaa atgaaaaagt actaccaaaa catagtttgc tttatgagta ttttacggtt	1560
tataacgaat tgacaaaggt caaatatggt actgaaggaa tgcgaaaacc agcatttctt	1620
tcagggtgaa agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc	1680
gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt	1740
tcaggagttg aagatagatt taatgcttca ttaggtacct accatgattt gctaaaaatt	1800
attaaagata aagatttttt ggataatgaa gaaaatgaag atatcttaga ggatattgtt	1860
ttaacattga ccttatttga agatagggag atgattgagg aaagacttaa aacatatgct	1920
cacctctttg atgataaggt gatgaaacag cttaaacgtc gccgtttatac tggttgggga	1980
cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta	2040
gattttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat	2100
agtttgacat ttaaagaaga cattcaaaaa gcacaagtgt ctggacaagg cgatagtta	2160

catgaacata ttgcaaat	ttt agctggtagc	cctgctatta	aaaaaggtat	tttacagact	2220
gtaaaagttg ttgatgaatt	gggtcaaagta	atggggcggc	ataagccaga	aaatatcggt	2280
attgaaatgg cacgtgaaaa	tcagacaact	caaaagggcc	agaaaaattc	gcgagagcgt	2340
atgaaacgaa tcgaagaagg	tatcaaagaa	ttaggaagtc	agattcttaa	agagcatcct	2400
gttgaaaata ctcaattgca	aaatgaaaag	ctctatctct	attatctcca	aaatggaaga	2460
gacatgtatg tggaccaaga	attagatatt	aatcgtttaa	gtgattatga	tgtcgatcac	2520
attgttccac aaagtttcct	taaagacgat	tcaatagaca	ataaggctct	aacgcgttct	2580
gataaaaatc gtggtaaatc	ggataacggt	ccaagtgaag	aagtagtcaa	aaagatgaaa	2640
aactattgga gacaacttct	aaacgccaa	ggttaatcactc	aacgtaagtt	tgataattta	2700
acgaaagctg aacgtggagg	tttgagtga	cttgataaag	ctggttttat	caaacgcaa	2760
ttggttgaaa ctgcgcaaat	cactaagcat	gtggcacaaa	ttttggatag	tcgcatgaat	2820
actaaatacg atgaaaatga	taaacttatt	cgagagggtta	aagtgtattac	cttaaaatct	2880
aaattagttt ctgacttccg	aaaagatttc	caattctata	aagtacgtga	gattaacaat	2940
taccatcatg cccatgatgc	gtatctaaat	gccgtcggtg	gaactgcttt	gattaagaaa	3000
tatccaaaac ttgaatcgga	gtttgtctat	ggtgattata	aagtttatga	tgttcgtaaa	3060
atgattgcta agtctgagca	agaaataggc	aaagcaaccg	caaaatattt	cttttactct	3120
aatatcatga acttcttcaa	aacagaaatt	acacttgcaa	atggagagat	tcgcaaacgc	3180
cctctaactg aaactaatgg	ggaaactgga	gaaattgtct	gggataaagg	gcgagatttt	3240
gccacagtgc gcaaagtatt	gtccatgccc	caagtcaata	ttgtcaagaa	aacagaagta	3300
cagacaggcg gatttctcaa	ggagtcaatt	ttaccaaaaa	gaaattcgga	caagcttatt	3360
gctcgtaaaa aagactggga	tccaaaaaaa	tatgggtggt	ttgatagtcc	aacggtagct	3420
tattcagtcc tagtggttgc	taagggtgga	aaagggaaat	cgaagaagtt	aaaatccggt	3480
aaagagttag tagggatcac	aattatggaa	agaagttcct	ttgaaaaaaa	tccgattgac	3540
tttttagaag ctaaaggata	taaggaagtt	aaaaaagact	taatcattaa	actacctaaa	3600
tatagtcttt ttgagttaga	aaacggtcgt	aaacggatgc	tggctagtgc	cggagaatta	3660
caaaaaggaa atgagctggc	tctgccaa	gc aaatatgtga	attttttata	tttagctagt	3720
cattatgaaa agttgaaggg	tagtccagaa	gataacgaac	aaaaacaatt	gtttgtggag	3780
cagcataagc attatttaga	tgagattatt	gagcaaatca	gtgaattttc	taagcgtggt	3840
atttttagcag atgccaattt	agataaagtt	cttagtgcat	ataacaaaca	tagagacaaa	3900
ccaatacgtg aacaagcaga	aaatattatt	catttattta	cgttgacgaa	tcttgagact	3960
cccgtgctt ttaaataattt	tgatacaaca	attgatcgta	aacgatatac	gtctacaaaa	4020
gaagtttttag atgccactct	tatccatcaa	tccatcactg	gtctttatga	aacacgcatt	4080
gatttgagtc agctaggagg	tgactgatgg	ctggttggcg	tactgttggtg	gtaaataccc	4140
actcgaaatt atcctataag	aataatcatc	tgatttttaa	ggatgcctat	aaaacggagc	4200
tgatccattt atcagaaatt	gatattttgt	tattagaaac	gaccgatatt	gtcttgacca	4260
ctatgctggt aaaacggcta	gtggatgaga	atgtccttgt	catattctgt	gatgataaac	4320
gattaccaac agctatgctg	atgccttttt	atggtcgtca	tgattcgagt	ttacagcttg	4380
ggaaacaaat gtcctggtca	gaaacagtca	aatcgacagg	ttggacgacg	attattgctc	4440
aaaagatttt gaatcaatct	tgctatctag	gagcatgctc	ctattttgaa	aaatcccaat	4500
ctattatgga ttatatcat	ggtttgga	atgttgatcc	gagtaatcga	gaagggcag	4560
cagcgagaat ttattttaat	acactttttg	ggaacgattt	ctcaagagat	ttggagcatc	4620

```

caatcaatgc aggtctggat tatggttata ctttattatt gagtatgttt gcgcgtgaag 4680
tggttgtgtc tggatgtatg actcagtttg ggcttaaaca cgctaatacag tttaatacagt 4740
tcaattttgc tagcgatatt atggaacat ttaggccttt agtggataag attgtttatg 4800
aaaatcgaaa tcagcctttt cccaaaataa agagagagtt atttactttg ttttcagata 4860
cattttcata taatggtaaa gagatgtatc tcacgaatat tattagcgat tataactaaa 4920
aagttgtcaa agctctgaat aatgaaggga aaggagttcc tgaatttagg atatgagtta 4980
tagatatatg agaatgatac ttatgtttga tatgccgacg gacaccgctg aggaacgaaa 5040
agcctatcga aaatttcgga aatttttact tagtgaaggg tttatcatgc atcaattttc 5100
tatttatagt aagttgctgt tgaataatac agctaacaat gccatgattg gtcggctgag 5160
ggagcataat cctaataaag gaaatattac attactaacg gtcacggaaa aacagtttgc 5220
acgaatgatt tattttacatg gtgaaagaaa taattgtatt gcaaactccg atgaaagact 5280
tgtatttctt ggggaggcct ttgatgaatc ttaatttttc cttactagat gaaccgattc 5340
cattaagagg cggtaacaatt cttgtgctcg aagatgtctg tgtattttca aaaatagtgc 5400
aatattgtta ccaatatgag gaagattctg aacttaaat ttttgatcac aagatgaaaa 5460
caatcaaaga atcagaaatc atgcttgtaa cagatatttt aggatttgat gttaactcct 5520
caaccatttt aaaattgatt catgcagatt tagaatctca atttaatgag aaacccgaag 5580
tgaaatcgat gattgacaaa ttggttgcta cgattacaga actgattgtc tttgaatgct 5640
tagaaaatga attagattta gagtatgatg aaatcacaat cctggaattg attaagtcct 5700
taggagtaaa agtagaaacg caaagtgata ctatttttga aaaatgtcta gagatacttc 5760
aaattttcaa atatctcact aagaaaaagt tgcttatttt tgtcaatagc ggagcttttc 5820
taacaaagga tgaagtggct agtttacaag agtatatatc attgacaaat ttaacagttc 5880
tctttttaga accacgtgaa ctatatgatt ttccgcagta tatttttagat gaagattatt 5940
tcttaataac taaaaatatg gtataa 5966

```

<210> 505

<211> 4107

<212> DNA

<213> *Streptococcus pyogenes*

<400> 505

```

atggataaga aatactcaat aggccttagat atcggcacaa atagcgtcgg atgggcggtg 60
atcactgatg aatataaggt tccgtctaaa agtttcaagg ttctgggaaa tacagaccgc 120
cacagtatca aaaaaaatct tataggggct cttttatttg acagtggaga gacagcggaa 180
gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240
tatctacagg agattttttc aaatgagatg gcgaaagtag atgatagttt ctttcacga 300
cttgaagagt cttttttggt ggaagaagac aagaagcatg aacgtcatcc tatttttggg 360
aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420
aaattggtag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgc 480
atgattaagt ttctgtgtca ttttttgatt gagggagatt taaatcctga taatagtgat 540
gtggacaaac tatttatcca gtgggtacaa acctacaatc aattatttga agaaaaccct 600
attaacgcaa gtggagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga 660
cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggcct atttgggaa 720
ctcattgatt tctcattgag tttagccctt aatttttaatt ccaattttta tttagcagaa 780

```

ctcattgctt	tgctattggy	cttgaccccc	aattcttaaat	caaatcttga	cttgctagaa	700
gatgctaaat	tacagctttc	aaaagatact	tacgatgatg	atttagataa	tttattggcg	840
caaattggag	atcaatatgc	tgatttggtt	ttggcagcta	agaatttatc	agatgctatt	900
ttactttcag	atatcctaag	agtaaatact	gaaataacta	aggctccccct	atcagcttca	960
atgattaaac	gctacgatga	acatcatcaa	gacttgactc	ttttaaaagc	tttagttcga	1020
caacaacttc	cagaaaagta	taaagaaatc	ttttttgatc	aatcaaaaaa	cggatatgca	1080
ggtttatattg	atgggggagc	tagccaagaa	gaattttata	aattttatcaa	accaatttta	1140
gaaaaaatgg	atggtactga	ggaattattg	gtgaaactaa	atcgtgaaga	tttgctgcgc	1200
aagcaacgga	cctttgacaa	cggctctatt	ccccatcaaa	ttcacttggg	tgagctgcat	1260
gctattttga	gaagacaaga	agacttttat	ccatttttaa	aagacaatcg	tgagaagatt	1320
gaaaaaatct	tgacttttcg	aattccttat	tatgttggtc	cattggcgcg	tggcaatagt	1380
cgttttgcat	ggatgactcg	gaagtctgaa	gaaacaatta	ccccatggaa	ttttgaagaa	1440
gttgctgata	aagggtgcttc	agctcaatca	tttattgaac	gcatgacaaa	ctttgataaa	1500
aatcttccaa	atgaaaaagt	actaccaaaa	catagtttgc	tttatgagta	ttttacgggt	1560
tataacgaat	tgacaaaggt	caaatatggt	actgaaggaa	tgcgaaaacc	agcatttctt	1620
tcagggtgac	agaagaaagc	cattgttgat	ttactcttca	aaacaaatcg	aaaagtaacc	1680
gttaagcaat	taaaaagaaga	ttatttcaaa	aaaatagaat	gttttgatag	tgttgaaatt	1740
tcaggagttg	aagatagatt	taatgcttca	ttaggtacct	accatgattt	gctaaaaatt	1800
attaagata	aagatttttt	ggataatgaa	gaaaatgaag	atatcttaga	ggatattggt	1860
ttaacattga	ccttatttga	agatagggag	atgattgagg	aaagacttaa	aacatatgct	1920
cacctctttg	atgataaggt	gatgaaacag	cttaaacgtc	gccgttatac	tggttgggga	1980
cgtttgtctc	gaaaattgat	taatggtatt	agggataagc	aatctggcaa	aacaatatta	2040
gattttttga	aatcagatgg	ttttgccaat	cgcaatttta	tgcagctgat	ccatgatgat	2100
agtttgacat	ttaaagaaga	cattcaaaaa	gcacaagtgt	ctggacaagg	cgatagttta	2160
catgaacata	ttgcaaattt	agctggtagc	cctgctatta	aaaaagggtat	tttacagact	2220
gtaaaagtig	ttgatgaatt	ggtcaaagta	atggggcggc	ataagccaga	aaatatcggt	2280
attgaaatgg	cacgtgaaaa	tcagacaact	caaaagggcc	agaaaaattc	gcgagagcgt	2340
atgaaacgaa	tcgaagaagg	tatcaaagaa	ttaggaagtc	agattcttaa	agagcatcct	2400
gttgaaaata	ctcaattgca	aaatgaaaag	ctctatctct	attatctcca	aaatggaaga	2460
gacatgtatg	tggaccaaga	attagatatt	aatcgtttaa	gtgattatga	tgtcgatcac	2520
attgttccac	aaagtttcct	taaagacgat	tcaatagaca	ataaggctct	aacgcgttct	2580
gataaaaatc	gtggtaaatc	ggataacggt	ccaagtgaag	aagtagtcaa	aaagatgaaa	2640
aactattgga	gacaacttct	aaacgccaag	ttaatcactc	aacgtaagtt	tgataattta	2700
acgaaagctg	aacgtggagg	tttgagtga	cttgataaag	ctggttttat	caaacgcaa	2760
ttggttgaaa	ctcgccaaat	cactaagcat	gtggcacaaa	ttttggatag	tcgcatgaat	2820
actaaatacg	atgaaatga	taaacttatt	cgagagggtta	aagtgtttac	cttaaaatct	2880
aaattagttt	ctgacttccg	aaaagatttc	caattctata	aagtacgtga	gattaacaat	2940
taccatcatg	cccatgatgc	gtatctaaat	gccgtcgttg	gaactgcttt	gattaagaaa	3000
tatccaaaac	ttgaatcgga	gtttgtctat	ggtgattata	aagtttatga	tgttcgtaaa	3060
atgattgcta	agtctgagca	agaaataggc	aaagcaaccg	caaaaatatt	cttttactct	3120
aatatcatga	acttcttcaa	aacagaaatt	acacttgcaa	atggagagat	tcgcaaacgc	3180

cctctaatacg aaactaatgg ggaaactgga gaaattgtct gggataaagg gcgagatttt	3240
gccacagtgc gcaaagtatt gtccatgccc caagtcaata ttgtcaagaa aacagaagta	3300
cagacaggcg gattctccaa ggagtcaatt ttacaaaaaa gaaattcgga caagcttatt	3360
gctcgtaaaa aagactggga tccaaaaaaa tatggtgggt ttgatagtc aacggtagct	3420
tattcagtcc tagtggttgc taagggtgga aaagggaaat cgaagaagtt aaaatccgtt	3480
aaagagttac tagggatcac aattatggaa agaagttcct ttgaaaaaa tccgattgac	3540
tttttagaag ctaaaggata taagggaagt aaaaaagact taatcattaa actacctaaa	3600
tatagtcctt ttgagttaga aaacggctgt aaacggatgc tggctagtgc cggagaatta	3660
caaaaaggaa atgagctggc tctgccaagc aaatatgtga attttttata tttagctagt	3720
cattatgaaa agttgaaggg tagtccagaa gataacgaac aaaaacaatt gtttgtggag	3780
cagcataagc attattttaga tgagattatt gagcaaatca gtgaattttc taagcgtgtt	3840
atttttagcag atgccaattt agataaagtt cttagtgc atatacaaa tagagacaaa	3900
ccaatacgtg aacaagcaga aaatattatt catttattta cgttgacgaa tcttgagct	3960
cccgtgctt ttaaataattt tgatacaaca attgatcgta aacgatatac gtctacaaaa	4020
gaagtttttag atgccactct tatccatcaa tccatcactg gtctttatga aacacgcatt	4080
gatttgagtc agctaggagg tgactga	4107

<210> 506

<211> 870

<212> DNA

<213> *Streptococcus pyogenes*

<400> 506

atggctggtt ggcgtactgt tgtggtaaat acccactcga aattatccta taagaataat	60
catctgattt ttaaggatgc ctataaaacg gagctgatcc atttatcaga aattgatatt	120
ttgttattag aaacgaccga tattgtcttg tccactatgc tggtaaaacg gctagtggat	180
gagaatgtcc ttgtcatatt ctgtgatgat aaacgattac caacagctat gctgatgcct	240
ttttatggtc gtcattgattc gagtttacag cttgggaaac aaatgtcctg gtcagaaaca	300
gtcaaatcgc aggtttggac gacgattatt gctcaaaaga ttttgaatca atcttgctat	360
ctaggagcat gtcctatatt tgaaaaatcc caatctatta tggatttata tcatggtttg	420
gaaaattttg atccgagtaa tcgagaaggg catgcagcga gaatttattt taatacactt	480
tttgggaacg atttctcaag agatttggag catccaatca atgcaggctt ggattatggt	540
tatactttat tattgagtat gtttgcgcgt gaagtgggtt tgtctggatg tatgactcag	600
tttgggctta aacacgctaa tcagtttaat cagttcaatt ttgctagcga tattatggaa	660
ccatttaggc ctttagtgga taagattgtt tatgaaaatc gaaatcagcc ttttcccaaa	720
ataaagagag agttattttac tttgttttca gatacatatt catataatgg taaagagatg	780
tatctcacga atattattag cgattatact aaaaaagttg tcaaagctct gaataatgaa	840
gggaaaggag ttcctgaatt taggatatga	870

<210> 507

<211> 342

<212> DNA

<213> *Streptococcus pyogenes*

<400> 507

```
atgagttata gatatatgag aatgatactt atgtttgata tgccgacgga caccgctgag    60
gaacgaaaaag cctatcgaaa atttcggaaa tttttactta gtgaaggggt tatcatgcat    120
caattttcta tttatagtaa gttgctgttg aataatacag ctaacaatgc catgattggt    180
cggctgaggg agcataatcc taataaagga aatattacat tactaacggt cacggaaaaa    240
cagtttgcac gaatgattta tttacatggt gaaagaaata attgtattgc aaactccgat    300
gaaagacttg tatttcttgg ggaggctttt gatgaatctt aa                        342
```

<210> 508

<211> 663

<212> DNA

<213> *Streptococcus pyogenes*

<400> 508

```
atgaatctta atttttcctt actagatgaa ccgattccat taagaggcgg tacaattctt    60
gtgctcgaag atgtctgtgt attttcaaaa atagtgcatt attgttacca atatgaggaa    120
gattctgaac ttaaattttt tgatcacaag atgaaaacaa tcaaagaatc agaaatcatg    180
cttgtaacag atatttttagg atttgatgtt aactcctcaa ccatttttaa attgattcat    240
gcagatttag aatctcaatt taatgagaaa cccgaagtga aatcgatgat tgacaaattg    300
gttgctacga ttacagaact gattgtcttt gaatgcttag aaaatgaatt agatttagag    360
tatgatgaaa tcacaatcct ggaattgatt aagtccttag gagtaaaagt agaaacgcaa    420
agtgatacta tttttgaaaa atgtctagag atacttcaaa ttttcaaata tctcactaag    480
aaaaagttgc ttatttttgt caatagcggg gcttttctaa caaaggatga agtggctagt    540
ttacaagagt atatatcatt gacaaattta acagttctct ttttagaacc acgtgaacta    600
tatgattttc cgcagtatat tttagatgaa gattatttct taataactaa aaatatggta    660
taa                                                                663
```

<210> 509

<211> 8020

<212> DNA

<213> *Streptococcus pyogenes*

<400> 509

```
atgagaatga ttttagcaca ctatgactgt aaaaagata aaaagcaatc tttagatgag    60
catttatggc atgtggcctg ttctagtcga caggaagcat ctataattgg tcaaggagat    120
gtgctttttt taattggtct ttaccacgac ctgggcaaag ctgacgaac ctttcaagat    180
aaattattaa ataatccaaa tcggcatgtt gatcactctt atgcaggggc aaaatactta    240
tgttctatta ttgggcctca tctaaaaaac cgagggggtg ataaaaatga gagaatgaca    300
ttcaacgaaa tgggtgggta tgtcatctct gctcatcatg ggatgtatga tttatgctac    360
tattttgacg atgctgaata ttatggcttt aataagttta aaaatcgtat caatagagac    420
ttagatgggt atcactatca tgaagatatt aaagggtagc ctctaaaatt agaaaaaaaa    480
ttatgtgatt atggctacaa agatttaagg gagcttattg ataaagcttt tgataattac    540
caacaagcca tttcttctt aaactaocaa aataaagata actaodatta ttatcaotct    600
```

```

tgtatggtga gactttactt gtcactctta aaaaacgctg atattttggga cacagtaaat 660
gcctatggcc ttaagataag tcctatggat aaaacagagc gatcctttct aaaacactcc 720
tatttagcgg ccattgaaca aaaatatgct agctttggac agccaaacaa tcagttgaac 780
actattcggg cagaaatcgc tgagcgtggt aaagaaagag gtaaacgaga ttccaagggg 840
atttatcgct tagatttacc gacaggagct ggcaagacta atcttagtat gcgttatgcg 900
tttcaccaat tagttcatca cgacaaatca aggttttttt acataactcc ctttctttcg 960
gttcttgagc aaaatgcttc cgaaattaga aaagttacag gtgaccttgg cgttctagaa 1020

caccattcca atgtggtgaa acaggcta atgataagga cagtttattg 1080
tcagcttatc ttagtgatag ctgggacagt caagtagtct tgacttctat ggttcaattt 1140
ttccaaacac ttttcaaac aaaatcagct aatctgagac gtttttcaag ttgattaat 1200
agtgttgga ttttagatga agttcaatcc ctgcctattg aagtcaccac ttgtttaat 1260
ttaacgatga attttttaaa taaagtattg gatacaacca tcgttctttg cacagcgaca 1320
caacctgctt atgattcttc agagattgac catcgtatct gttatggagg gaacctggga 1380
gaattagctg aaatagttga gttaacgatt gaagaaaaac agattttttc aaggacagag 1440
cttagaaaat ttgatgatag tgatcagaaa gttcacttga ctgatgttat taaccttatt 1500
ctaggtgagg aaaactcagt tcttgctatt ttaatacga aaaaaacggt tcataactgc 1560
tatactatgc taaaagacat gactgataga ccggtctatc agctttcgac aaatatgtgt 1620
gcgcagcata gacttgactt gattgctaag atcaaaacgg agttacaaaa taatatccct 1680
attatttgta ttagcacgca attaatgaa gcagggttag atgttgattt tcatcgcgtc 1740
attcgttctt actcagggat tgattctatt gttcaggctg ctggacggtg taaccgagaa 1800
ggcaaacgag ataaagggca agtcactctt gtcaatctga ccaatgaaga ggaaaatatt 1860
tctaggctga cagaaataaa aactaaaaaa gaagccacag aatctattct tcataagatt 1920
gggtctccaa ttgatattct aactttaaac cgtgactttt ttgagtatta ttatgccaat 1980
aatcaggggc tgatggatta tcttttgaa gacaacctat caatctacga ctatttaagc 2040
cttaatatat atcagacggc aaataaaaag ttcaaaggta agttaaaaca agcttttaaa 2100
acagcaggag ccaaaatgaa cctcatcaat aatgatatga taggaattct cgtaccttat 2160
ggcgaagctg agaaaaaatt ggcttattta gaagaattag gtgtgtcaca ttttttatca 2220
gcaaaagatt atcaaacgat aaaatcatta ctaaaagagt tacaaccttt tacgggttaat 2280
gtccgcgaga acgatcctct ttttgagaca acaaaatctt atctaaatgg tcagattctg 2340
gttttgacgt cggagtatta tgacacggaa agaggagtta aatacgattc agctagcttt 2400
tacttctaac tcaaaacgaa agaagattaa caaaagggtg ttagaggacc ttgttaacct 2460
gccaatcatc attagtaatt attatcaatt tagactattt aataaaatta gattacaaaa 2520
aaacagaagg aggaaagtag cttgtacaga tctagagact tctacgtgag agtaagtggg 2580
cagcgagctc tttttacaaa tccagccaca aaagggggat cggaacgctc atcctattcg 2640
gttccgacta gacaggcact gaatggtatc gttgatgcca tctattataa gccgaccttt 2700
actaatatcg tcacagaggt taaggttatt aaccagattc aaaccgaatt acagggtgtc 2760
agggctctgt tacatgatta tagtgcagat ttaagttatg tatcctattt gagtgatgtt 2820
gtttatctga tcaagtttca ttttgtttgg aatgaagata gaaaagattt gaactcagat 2880
agacttccag ctaaacatga agccattatg gagcgttcta ttcgtaaagg gggacgtcga 2940
gatgtgtttt tgggtacaag agaatgttta gggctttag atgatatcag ccaagaagag 3000

```

tatgagacta ctgtgtcgtt ttataatggt gtcaatatcg acttggaat catgttccat	3060
tcctttgcct atccgaagga caaaaagaca ccattaaaat catactttac aaagactgtg	3120
atgaaaaatg gagtcattac gtttaaagca cagtctgaat gcgatattgt taacacgctt	3180
tctagtattg cttttaagc accagaggag ataaaatcgg ttaacgatga atgcatggag	3240
tatgatgcca tggagaaagg agaaaactga tggatTTTT tacttctctc ttgaagactt	3300
atgaaaaagc agagctagca gacttggttg atcatcaaaa aagaaataat gagccggtt	3360
tactgccgat ttatcatag agtttaaagt caaatggtta aaatatcatt tcagtgaac	3420
ttgacaaaga tggccagttt cacaaggcag aatttatggc agataagcaa atgattattt	3480
ttcctgtaac ggctgattct gttgctaggt caggtagtca tctgcaccg catcccctag	3540
tcgataaatt tgcttattat agtgctgaaa tggggcagat tcagtatgat tcttttcata	3600
agcaactgaa taactggatt gattattgtg aggagggtga tgtcaagaaa tttttaacct	3660
ttgttcagca gttcattttt aagccagaat ttctaactt gattcttgat tctttaattg	3720
gtcctgatta tcaacataat caattaaaag tcacattttg tgatgccact ggaaaagaaa	3780
aattaattga tttatcagct tgcttttttag aattttcaat tgatcagttc cagggttcta	3840
aaaatgaatc ggtttcgaca tttaaagcct tacaccaatc ctatatttct tttgttgaag	3900
ccaatcgtga aaatcctcgtt atttgtaata ttagtggaag agaggaaacag cttaccgata	3960
agcatagagg tttgatgggg aatgctaaaa tcatctctgt tagtaataaa agagaagctt	4020
ataaaggacg ttttagagaa cgcgaagacg ttttttagtgt tggctatgaa acttccgaaa	4080
agattcattt aatgtcgaag taccttttag aaaataaaaa taccagtact tggttagggt	4140
cttctcaata tttaatcaac tggttcagcg atgatttaac aaatgatagt cggttggata	4200
ttgtatcacc aatctttgat gatggacttg aagaagatga tgatgacgat acgcctcctg	4260
ttataacatt agcaactgaa gacaataaaa gaattggtta atcattcatc aagggacaaa	4320
aattatttgc taatgatgcc acttactacg ttgctatttt gaataaaaacc agcaatgggc	4380
ggattgcttt aaaatatttt cgtcagcttc aagcgtccca attactcacc aatcttaaca	4440
agtggcagga aacatacagt tgggagtcgc gatctaagtt tgggaaaagt cgcttaagaa	4500
cccctacttt tcatgacatc cttaatgtgt cctacggggt tgatagggat cgcttccttg	4560
aattagataa tgataacttc aaaagtgatc aaattcaaaa gttagtggca agtttgattg	4620
atggtaaacc gatgccacag tccattgtca aaaagttagg taacaatggt aaagaacgac	4680
atcgttaccg taagcactgg tatcaagttg agcagggtctg cttagcaatt ttacacaaac	4740
aaaatgggga ggaattttca ccgatgctag atcataccaa tcaaaatcgt tcctatcttt	4800
ttggacgatt attagcaatt tttgaattaa tcgagacctt gcgttatggc ttggatggaa	4860
acaataacga ccgtattacc aatgctgaac gttattggac agcctatact ggacaacaa	4920
caaaattgat gatgttattg gaaaataaaa ttaagcctta cgaagaacca ttgaaattaa	4980
atcgtcgtgg cagttggatg aaattagaaa aagaaaaaga agagatttta gaactgttaa	5040
atcctctgtt agaaacagaa acaatggaaa aacccttaga ttaccgcttt atttttgggt	5100
attatgctga gaaaaactat tactatacaa acaaaaacac ggaagtaaca gaaagtgagg	5160
agtaaaaaga tgttggaaaca caaaattgat tttatggtaa ctcttgaagt gaaagaagca	5220
aatgcaaatg gtgatccctt aaatggaaac atgcctcgta cagatgccaa aggatatggt	5280
gtgatgagtg atgtctccat taaacgtaag attcgtaatc gtttgcaaga tatggggaag	5340
tctatttttg tgcaagctaa tgagcgtatt gaagatgatt ttcgttccact ggaaaaacgc	5400
ttcttttttg tgcagctaa tgagcgtatt gaagatgatt ttcgttccact ggaaaaacgc	5460

ttttcgcaac	attttacagc	taagacacct	gacaaagaaa	ttgaagaaaa	agcaaatgca	5460
ttatggtttg	atgttcgtgc	ttttggacaa	gtttttactt	atctgaaaaa	atcaattggg	5520
gtgcgtggac	cagtttccat	cagtatggct	aagtccttgg	agccaattgt	catttccagc	5580
cttcaaatta	cgcgtagtac	caatggtatg	gaagctaaga	ataatagtgg	ccgctcttct	5640
gatacgatgg	ggacaaaaca	ttttgtagat	tatggtgtgt	atgtacttaa	aggttctatc	5700
aatgcttatt	ttgctgaaaa	gactgggttt	tctcaggaag	atgctgaggc	tattaaagaa	5760
gttttggtta	gcttgtttga	aaatgatgcg	tcgtctgcac	gtccggaagg	ctctatgcga	5820
gtttgtgaag	tcttttggtt	tacgcattca	agcaaattgg	gaaatgtttc	aagtgcgcgt	5880
gtctttgact	tgttagagta	tcacaaatca	atagaagaaa	aaagcactta	tgacgcttat	5940
cagattcatc	taaatcaaga	aaaattggct	aaatatgaag	cgaaagggtt	aacgcttgaa	6000
atcctagaag	gactctagta	tggcttatgc	cgaagatgat	tatttaatgc	tgtcaggtat	6060
tcagcatttc	caattttgta	aacgtcaatg	ggcgttgatc	catattgagc	aacaatggct	6120
tgataatgaa	gcgacagcgc	atggacaggt	tttacatact	aaagcagata	acccttacat	6180
taaagaaaaa	cgaaaagagc	ttttggcttc	acgtgctatg	cccatttctt	ctgcagaact	6240
tggactttca	ggaattatgg	atgttgtgga	attttataaa	gatgatcaag	gtgtgtcttt	6300
gaggggaaaa	cgtgggaaat	ggttaccaaa	agtttgtgga	tacaagcgcg	gaaaacctaa	6360
aaaagatacc	agagatatgt	tccagttggt	ggctcagacc	atgtgtttag	aagaaacgct	6420
agactgcgac	attaacgaag	gttgtcttta	ttaccatagt	gtcaatcaaa	gagtgattgt	6480
tcctatgaca	tcagctttgc	gtcaagaagt	gaaggaatta	gccgcagaga	tgcattgaggt	6540
ttatcagagt	caaatgctac	ctaaagcagc	ttattttaaa	aactgtcagc	tttgttcttt	6600
agtcgatatt	tgtaagccca	ggttgagtaa	aaaaacaagg	agtgtgtcgc	gttacatcaa	6660
tgaggctatg	accagtggag	agatggacct	atgaagaagt	tgctaaatac	cttgattttg	6720
acgcaagaag	atttttatgt	cactaaagag	ggcgataaca	ttgttatcaa	gcaagaagggt	6780
aaggttctca	aacggtttcc	gtttcggatt	attgacggta	ttgtctgttt	ttcttatttg	6840
ggtgtgtcgt	ctgctttggt	gaagttatgt	acggagaatc	agattaattt	atcgtttcat	6900
acaccacaag	ggcgtttttg	tggctgctat	attggttcaa	ccaatgggaa	tgtgttggtg	6960
cgtagagaac	attatcgttt	atctgatcgt	gaggaatctt	tggaaatcgc	aaagcggttt	7020
attttggtta	aaatttccaa	ctcaaggaaa	tacttgctac	gctttaaacg	agatcatcgt	7080
caacagattg	ataccaagct	ttttgaggct	gttaatgacg	aattgatatg	ggcttttagag	7140
atggttcagg	cagcagataa	taaagactct	ttaagaggga	ttgaaggcca	agctgctaatt	7200
cagtattttc	gcataattta	tgacctgggt	ttgacggaca	aaaaaacggt	ttacttccaa	7260
ggtcggagta	aacgaccacc	cttagattgt	gttaatgccc	tcttgtcttt	tggttacagt	7320
ttactgacct	ttgaatgtca	atctgccttg	gaagctgtcg	gattagacag	ttacgttggt	7380
ttctttcaca	cggatcgctc	tgggctgctc	agtttagcgc	ttgatttagt	tgaagagttc	7440
cgtcatata	ttgtagatcg	ttttgtcttt	tcattaatta	ataaaggaca	acttcagaaa	7500
aaacactttg	agggttaaaga	aaatggtagt	attttattga	cggaaaatgg	cagagctatt	7560
tttattgatt	tgtggcagaa	gcgtaagcat	actgaggtag	aacatccttt	tacaaaagag	7620
aaagtaaaac	ttatgttatt	accctatgta	caagcgcagc	ttttagctaa	ggctatacga	7680
ggagatttag	aaagctatcc	accttttatg	gtttaggaga	tgttatatga	tggtttttagt	7740
cacttatgat	gtaaaatcgg	aaacacctgc	tggtagaaaa	agattgcgtc	atgttgccaa	7800
actctgtgtg	gactatgggc	aacgtgttca	aaattctgtt	tttgaatgtt	ctgtgacacc	7860

cgcagaattt	gtggatataa	agcaccgcct	aacacaaatc	attgatgaga	aaactgatag	7920
tattcgcttt	tatttattgg	ggaaaaattg	gcagaggcgt	gtggaaacac	ttggtcgctc	7980
agacagctat	gaccagata	aagggtgtct	attattgtaa			8020

$\langle 210 \rangle$ 510

<211> 2409

<212> DNA

<213> Streptococcus pyogenes

$\langle 400 \rangle$ 510

atgagaatga	ttttagcaca	ctatgactgt	aaaaaagata	aaaagcaatc	tttagatgag	60
catttatggc	atgtggcctg	ttctagtctg	caggaagcat	ctataattgg	tcaaggagat	120
gtgctttttt	taattggctt	ttaccacgac	ctggggcaaag	ctgatcgaac	ctttcaagat	180
aaattattaa	ataatccaaa	tgggcatggt	gatcactctt	atgcaggggc	aaaatactta	240
tgttctatta	ttgggcctca	tctaaaaaac	cgagggggtg	ataaaaatga	gagaatgaca	300
ttcaacgaaa	tgggtggggta	tgtcatctct	gctcatcatg	ggatgtatga	tttatgctac	360
tattttgacg	atgctgaata	ttatggcttt	aataagttta	aaaatcgtat	caatagagac	420
ttagatgggt	atcactatca	tgaagatatt	aaagggtacg	ctctaaaatt	agaaaaaaaa	480
ttatgtgatt	atggctacaa	agatttaagg	gagcttattg	ataaagcttt	tgataattac	540
caacaagcca	tgtcttcctt	aaactggcaa	gataagagtg	agtgggatta	ttatcagtct	600
tgtatgggtg	gactttactt	gtcactctta	aaaaacgctg	atatttttga	cacagtaaat	660
gcctatggcc	ttaagataag	tcctatggat	aaaacagagc	gaccccttct	aaaacactcc	720
tattttagcg	ccattgaaca	aaaatatgct	agctttggac	agccaaacaa	tcagttgaac	780
actattcgga	cagaaatcgc	tgagcgtggt	aaagaaagag	gtaaacgaga	ttccaagggg	840
atttatcgct	tagatttacc	gacaggagct	ggcaagacta	atcttagtat	gcgttatgcg	900
tttcaccaat	tagttcatca	cgacaaatca	aggttttttt	acataaactcc	ctttcttttcg	960
gttcttgagc	aaaatgcttc	cgaaattaga	aaagttacag	gtgaccttgg	cgttctagaa	1020
caccattcca	atgtggtgaa	acaggctaata	gaagatgatg	atgataagga	cagtttattg	1080
tcagcttatc	ttagtgatag	ctgggacagt	caagtagtct	tgacttctat	ggttcaattt	1140
ttccaaacac	ttttcaaaac	aaaatcagct	aatctgagac	gtttttcaag	tttgattaat	1200
agtgttgtga	ttctagatga	agttcaatcc	ctgcctattg	aagtcaccac	tttgtttaat	1260
ttaacgatga	atTTTTTaaa	taaagtatat	gatacaacca	tcgttctttg	cacagcgaca	1320
caacctgctt	atgattcttc	agagattgac	catcgtatct	gttatggagg	gaacttggga	1380
gaattagctg	aaatagttga	gttaacgatt	gaagaaaaac	agattttttc	aaggacagag	1440
cttagaaaaat	ttgatgatag	tgatcagaaa	gttcacttga	ctgatgttat	taaccttatt	1500
ctagggtgagg	aaaactcagt	tcttgctatt	tttaatacga	aaaaaacggt	tcataactgc	1560
tatactatgc	taaaagacat	gactgataga	ccggtctatc	agcttttcgac	aaatatgtgt	1620
gcgcagcata	gacttgactt	gattgctaag	atcaaaacgg	agttacaaaa	taatattcct	1680
attatttgta	ttagcacgca	attaattgaa	gcagggttag	atgttgattt	tcatcgcgtc	1740
attcgttcct	actcagggat	tgattctatt	gttcaggctg	ctggacggtg	taaccgagaa	1800
ggcaaacgag	ataaagggca	agtcactctt	gtcaatctga	ccaatgaaga	ggaaaaatatt	1860

tctaggctga cagaaataaa aactaaaaaa gaagccacag aatctattct tcataagatt	1920
gggtctccaa ttgatatctc aactttaaac cgtgactttt ttgagtatta ttatgccaat	1980
aatcagggac tgatggatta tcctttggaa gacaacctat caatctacga ctatttaagc	2040
cttaatatTT atcagacggc aaataaaaaag ttcaaaggta agttaaaaca agcttttaaa	2100
acagcaggag ccaaaatgaa cctcatcaat aatgatatga taggaattct cgtaccttat	2160
ggcgaagctg agaaaaaatt ggcttattta gaagaattag gtgtgtcaca ttttttatca	2220
gcaaaagatt atcaaacgat aaaatcatta ctaaaagagt tacaaccttt tacggttaat	2280
gtccgcgaga acgatcctct ctttgagaca acaaaatctt atctaaatgg tcagattctg	2340
gttttgacgt cggagtatta tgacacggaa agaggagtta aatacgattc agctagcttt	2400
tacttctaa	2409

<210> 511

<211> 729

<212> DNA

<213> *Streptococcus pyogenes*

<400> 511

ttgtacagat ctagagactt ctacgtgaga gtaagtggc agcgagctct ttttacaat	60
ccagccacaa aagggggatc ggaacgctca tcctattcgg ttccgactag acaggcactg	120
aatggtatcg ttgatgccat ctattataag ccgaccttta ctaatatcgt cacagagggt	180
aagggttatta accagattca aaccgaatta cagggtgtca gggctctggt acatgattat	240
agtgcagatt taagttatgt atcctatttg agtcatgttg tttatctgat caagtttcat	300
tttgtttga atgaagatag aaaagatttg aactcagata gacttccagc taaacatgaa	360
gccattatgg agcgttctat tcgtaaaggg ggacgtcgag atgtgttttt gggtaacaaga	420
gaatgttttag ggctttaga tgatattcagc caagaagagt atgagactac tgtgtcgtat	480
tataatggtg tcaatatcga cttgggaatc atgttccatt cctttgccta tccgaaggac	540
aaaaagacac cattaaaatc atactttaca aagactgtga tgaaaaatgg agtcattacg	600
tttaagcac agtctgaatg cgatattgtt aacacgcttt ctagtattgc ttttaagca	660
ccagaggaga taaaatcggg taacgatgaa tgcattggagt atgatgccat ggagaaagga	720
gaaaactga	729

<210> 512

<211> 1896

<212> DNA

<213> *Streptococcus pyogenes*

<400> 512

atggattttt ttacttctct cttgaagact tatgaaaaag cagagctagc agacttggtt	60
gatcatcaaa aaagaataa tgagccggtt ttactgccga tttatcatac gagtttaag	120
tcaaattggtg aaaatatcat ttcagtgaag cttgacaaag atggccagtt tcacaaggca	180
gaatttatgg cagataagca aatgattatt tttcctgtaa cggctgattc tgttgctagg	240
tcaggtagtc atcctgcacc gcatccccta gtcgataaat ttgcttatta tagtgctgaa	300
atggggcaga ttcagtatga ttcttttcat aagcaactga ataactggat tgattattgt	360
naaaagaaatc atatcaaaaa attttttaac tttattcagc aatttatttt naaaacaaa	420

```

gagggagggtg agtgcaggaa accctcaacc cctgcctcgc agctccctccc gaagctcaggaa
tttctaacat tgattcttga ttctttaatt ggtcctgatt atcaacataa tcaattaaaa 480
gtcacatttt gtgatgccac tggaaaagaa aaattaattg atttatcagc ttgcttttta 540
gaattttcaa ttgatcagtt ccagggtctt aaaaatgaat cggtttcgac atttaaagcc 600
ttacaccaat cctatatttc ttttgttgaa gccaatcgtg aaaaatctcg tatttgtaat 660
attagtggac gagaggaaca gcttaccgat aagcatagag gtttgatggg gaatgctaaa 720
atcatctctg ttagtaataa aagagaagct tataaaggac gttttagaga acgcgaagac 780
gttttttagtg ttggctatga aacttccgaa aagattcatt taatgctcaa gtacctttta 840
gaaaataaaa ataccagtac ttggttaggg tcttctcaat atttaataa ctggttcagc 900
gatgatttaa caaatgatag tcggttggat attgtatcac caatcttga tgatggactt 960
gaagaagatg atgatgacga tacgcctcct gttataacat tagcaactga agacaataaa 1020
agaattggta aatcattcat caagggacaa aaattatttg ctaatgatgc cacttactac 1080
gttgctattt tgaataaaac cagcaatggg cggattgctt taaaatattt tcgtcagctt 1140
caagcgtccc aattactcac caatcttaac aagtggcagg aaacatacag ttgggagtcg 1200
cgatctaagt ttgggaaaag tcgcttaaga acccctactt ttcatgacat ccttaatgtg 1260
tcctacgggg ttgataggga tcgcttcctt gaattagata atgataactt caaaagtgat 1320

caaattcaaa agttagtggc aagtttgatt gatggtaaac cgatgcāaca gtccattgtc 1380
aaaaagttag gtaacaatgt taaagaacga catcgttacc gtaagcactg gtatcaagtt 1440
gagcaggtct gcttagcaat ttacacaaa caaaatgggg aggaattttc accgatgcta 1500
gatcatacca atcaaaatcg ttcctatctt tttggacgat tattagcaat ttttgaatta 1560
atcgagacct tgcgttatgg cttggatgga aacaataacg accgtattac caatgctgaa 1620
cgttattgga cagcctatac tggacaacca acaaaattga tgatgttatt ggaaaataaa 1680
attaagcctt acgaagaacc attgaaatta aatcgctcgt gcagttggat gaaattagaa 1740
aaagaaaaag aagagatttt agaactgtta aatcctctgt tagaaacaga aacaatggaa 1800
aaacccttag attaccgctt tatttttggg tattatgctg agaaaaacta ttactatata 1860
aaacaaaaca cggaagtaac agaaagtga gtagtaa 1896

```

<210> 513

<211> 849

<212> DNA

<213> *Streptococcus pyogenes*

<400> 513

```

atgttggaac acaaaattga ttttatggta actcttgaag tgaaagaagc aaatgcaaat 60
ggtgatccct taaatggaaa catgcctcgt acagatgcca aaggatatgg tgtgatgagt 120
gatgtctcca ttaaacgtaa gattcgtaat cgtttgcaag atatggggaa gtctattttt 180
gtgcaagcta atgagcgtat tgaagatgat tttcgttcac tggaaaaacg cttttcga 240
cattttacag ctaagacacc tgacaaagaa attgaagaaa aagcaaatgc attatggttt 300
gatgttcgtg cttttggaca agtttttact tatctgaaaa aatcaattgg ggtgcgtgga 360
ccagtttcca tcagtatggc taagtccttg gagccaattg tcatttccag ctttcaaatt 420
acgcgtagta ccaatgggat ggaagctaag aataatagtg gccgctcttc tgatacgtg 480
gggacaaaac atttttaga ttatgggtgt tatgtactta aaggttctat caatgcttat 540

```

```

tttgctgaaa agactggttt ttctcaggaa gatgctgagg ctattaaaga agttttggtt 600
agcttgtttg aaaatgatgc gtcgtctgca cgtccggaag gctctatgcg agtttgtaa 660
gtcttttggt ttacgcattc aagcaaattg ggaaatgttt caagtgcgcg tgtctttgac 720
ttgtagagt atcatcaatc aatagaagaa aaaagcactt atgacgctta tcagattcat 780
ctaaatcaag aaaaattggc taaatatgaa gcgaaagggt taacgcttga aatcctagaa 840
ggactctag 849

```

<210> 514

<211> 675

<212> DNA

<213> *Streptococcus pyogenes*

<400> 514

```

atggtctatg ccgaagatga ttatttaatg ctgtcaggta ttcagcattt ccaattttgt 60
aaacgtcaat gggcggtgat ccatattgag caacaatggc ttgataatga agcgacagcg 120

catggacagg ttttacatac taaagcagat aacccttaca ttaaagaaaa acgaaaagag 180
cttttggtct cacgtgctat gccatttct tctgcagaac ttggactttc aggaattatg 240
gatgttggtg aattttataa agatgatcaa ggtgtgtctt tgaggggaaa acgtgggaaa 300
tggttaccaa aagttgtgga atacaagcgc ggaaaacctt aaaaagatac cagagatatt 360
gtccagttgg tggctcagac catgtgttta gaagaaacgc tagactgcga cattaacgaa 420
ggttgtcttt attaccatag tgtcaatcaa agagtgattg ttcctatgac atcagctttg 480
cgtcaagaag tgaaggaatt agccgcagag atgcatgagg tttatcagag tcaaatgcta 540
cctaaagcag cttattttta aaactgtcag ctttgttctt tagtcgatat ttgtaagccc 600
aggttgagta aaaaaacaag gagtgtgtcg cgttacatca atgaggctat gaccagtgag 660
gagatggacc tatga 675

```

<210> 515

<211> 1026

<212> DNA

<213> *Streptococcus pyogenes*

<400> 515

```

atgaagaagt tgctaaatac cttgtatttg acgcaagaag atttttatgt cactaaagag 60
ggcgataaca ttgttatcaa gcaagaagggt aaggttctca aacggtttcc gtttcggatt 120
attgacggta ttgtctgttt ttcttatttg ggtgtgtcgt ctgctttggt gaagtattgt 180
acggagaatc agattaatth atcgtttcat acaccacaag ggcgtttttg tggtcgctat 240
attggttcaa ccaatgggaa tgtgttgttg cgtagagaac attatcgtht atctgatcgt 300
gaggaatctt tggaatacgc aaagcggttt attttggtta aaatttccaa ctcaaggaaa 360
tacttgctac gctttaaacg agatcatcgt caacagattg ataccaagct ttttgaggct 420
gttaatgacg aattgatatg ggcttttagag atggttcagg cagcagataa taaagactct 480
ttaagaggga ttgaaggcca agctgctaath cagtattttc gcatatttaa tgacctggtg 540
ttgacggaca aaaaaacgtt ttacttccaa ggtcggagta aacgaccacc cttagattgt 600
gttaatgccc tcttgtcttt tggttacagt ttactgacct ttgaatgtca atctgccttg 660

```

```

gaagctgtcg gattagacag ttacgttggg ttctttcaca cggatcgtcc tgggcgtgct 720
agtttagcgc ttgatttagt tgaagagttc cgctcatata ttgtagatcg ttttgtcttt 780
tcattaatta ataaaggaca acttcagaaa aaacactttg aggttaaaga aaatggtagt 840
attttattga cggaaaatgg cagagctatt tttattgatt tgtggcagaa gcgtaagcat 900
actgaggtag aacatccttt tacaaaagag aaagtaaaac ttatgttatt accctatgta 960
caagcgcagc ttttagctaa ggctatacga ggagatttag aaagctatcc accttttatg 1020
gtttag 1026

```

<210> 516

<211> 294

<212> DNA

<213> *Streptococcus pyogenes*

<400> 516

```

atgatggttt tagtcactta tgatgtaaat acggaacac ctgctggtag aaaaagattg 60
cgtcatgttg ccaaactctg tgtggactat gggcaacgtg ttcaaaattc tgtttttgaa 120
tgttctgtga cccccgcaga atttgtggat ataaagcacc gcttaacaca aatcattgat 180
gagaaaactg atagtattcg cttttattta ttggggaaaa attggcagag gcggtgtggaa 240
acacttggtc gctcagacag ctatgacca gataaagggt tcttattatt gtaa 294

```

<210> 517

<211> 5966

<212> DNA

<213> *Streptococcus pyogenes*

<400> 517

```

atggataaga aatactcaat aggccttagat atcggcaca atagcgtcgg atgggcgggtg 60
atcactgatg attataaggt tccgtctaaa aagctcaagg gtctgggaaa tacagaccgc 120
cacggtatca aaaaaatct tataggggct cttttatttg acagtggaga gacagcggaa 180
gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240
tatctacagg agattttttc aaatgagatg gcgaaagtag atgatagttt ctttcatcga 300
cttgaagagt cttttttggt ggaagaagac aagaagcatg aacgtcatcc tatttttgga 360
aatatagtag atgaagtggc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420
aaattggcag attctactga taaagtggat ttgcgcttaa tctatttggc cttagcgcgt 480
atgattaagt ttcgtgggtc ttttttgatt gagggagatt taaatcctga taatagtgat 540
gtggacaaac tatttatcca gttggtacaa acctacaatc aattatttga agaaaacctt 600
attaacgcaa gtagagtaga tgctaaagcg attccttctg cacgattgag taaatcaaga 660
cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggatt gtttggaat 720
ctcattgctt tgtcattggg attgaccctt aattttaaat caaattttga tttggcagaa 780
gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg 840
caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttatc agatgctact 900
ttactttcag atatcctaag agtaaagtag gaaataacta aggcctcccct atcagcttca 960
atgattaagc gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga 1020
caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca 1080

```

ggttatattg atgggggagc tagccaagaa gaattttata aattttatcaa accaatttta 1140
 gaaaaaatgg atggtactga ggaattattg gcgaaactaa atcgtgaaga tttgctgcgc 1200
 aagcaacgga cctttgacaa cggctctatt ccctatcaaa ttcacttggg tgagctgcac 1260
 gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt 1320
 gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt 1380
 cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa 1440
 gttgtcgata aagggtgcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa 1500

aatcttccaa atgaaaaagt actacaaaa catagtttgc tttatgagta ttttacggtt 1560
 tataacgaat tgacaaaagt caaatatgtt actgagggaa tgcgaaaacc agcatttctt 1620
 tcagggtgac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc 1680
 gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt 1740
 tcaggagttg aagatagatt taatgcttca ttaggtacct accatgattt gctaaaaatt 1800
 attaaagata aagatttttt ggataatgaa gaaaacgaag atatctttaga ggatattgtt 1860
 ttaacattga ctttatttga agatagggag atgattgagg aaagacttaa aacatatgct 1920
 cacctctttg atgataaggt gatgaaacag cttaaacgtc gccgttatac tggttgggga 1980
 cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta 2040
 gattttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat 2100
 agtttgacat ttaagaaga cattcaaaaa gcacaagtg ctggacaagg cgatagttaa 2160
 catgaacata ttgcaaat tt agctggtagc cctgctatta aaaaagggtat tttacagact 2220
 gtaaaagtgt ttgatgaatt ggtcaaagta atggggcggc ataagccaga aaatatcggt 2280
 attgaaatgg cacgtgaaaa tcagacaact caaaagggcc agaaaaattc gcgtgagcgt 2340
 atgaaacgta ttgaagaagg aataaaagaa ctaggaagtg atattctaaa ggagtatcct 2400
 gttgaaaaca ctcaattaca aaatgaaaag ctctatctct attatctcca aaatggaaga 2460
 gacatgtatg tggaccaaga attagatatt aatcgtttta gtgattatga tgtcgatcac 2520
 attgttcac aaagtctcct taaagacgat tcaatagaca ataaggctct aacgcgttct 2580
 gataaaaatc gtggtaaatc ggataacggt ccaagtgaag aagtagtcaa aaagatgaaa 2640
 aactattgga gacaacttct aaacgccaag ttaatcactc aacgtaagtt tgataattta 2700
 acaaaagctg aacgtggagg tttgagtga cttgataaag ttggttttat caaacgcca 2760
 ttggttgaaa ctgcgcaaat cactaagcat gtggcacaaa ttttgatag tcgcatgaat 2820
 actaaatag atgaaaatga taaacttatt cgagagggtta gtagtattac cttaaaatct 2880
 aaattagttt ctgacttccg aaaagatttc caattctata aagtacgtga gattaacaat 2940
 taccatcatg cccatgatgc gtatcttaat gccgtcgttg gaactgcttt gattaagaaa 3000
 tatccaaaac ttgaatcggg tttgtctat ggtgattata aagtttatga tgttcgtaaa 3060
 atgattgcta agtctgagca ggaaataggc aaagcaaccg caaaatattt cttttactct 3120
 aatatcatga acttcttcaa aacagaaatt acacttgcaa atggagagat tcgcaaacgc 3180
 cctctaactg aaactaatgg ggaaactgga gaaattgtct gggataaagg gcgagatttt 3240
 gccacagtgc gcaaaagtatt gtccatgccc caagtcaata ttgtcaagaa aacagaagta 3300
 cagacaggcg gattctccaa ggagtaatt ttacaaaaa gaaattcggg caagcttatt 3360
 gctcgtaaaa aagactggga tccaaaaaaa tatggtggtt ttgatagtcc aacggtagct 3420
 tattcagtc tagtggttgc taagggtgaa aaagggaaat cgaagaagtt aaaatccgtt 3480
 aaaaattac taadattcac aataatgaa aaaaactctt ttaaaaaa tccaattoac 3540

```

tttttagaag ctaaaggata taaggaagtt agaaaagact taatcattaa actacctaaa 3600
tatagtcttt ttgagttaga aaacggtcgt aaacggatgc tggctagtgc cggagaattg 3660
caaaaaggaa atgagctagc tctgccagc aaatatgtga attttttata tttagctagt 3720
cattatgaaa agttgaaggg tagtccagaa gataacgaac aaaaacaatt gtttgtggag 3780
cagcataagc attattttaga tgagattatt gagcaaatca gtgaattttc taagcgtggt 3840
attttagcag atgccaatat agataaagtt cttagtgcac ataacaacaa tagagacaaa 3900
ccaatacgtg aacaagcaga aaatatattt catttattta cgttgacgaa tcttgagact 3960
cccgtgctt ttaaataattt tgatacaaca attgatcgta aacgatatac gtctacaaaa 4020
gaagttttag atgccactct tatccatcaa tccatcactg gtctttatga aacacgcatt 4080
gatttgagtc agctaggagg tgactgatgg ctggttgccg tactgttggt gtaaatatcc 4140
actcgaaatt atcctataag aataatcatc tgatttttaa ggatgcctat aaaacggagc 4200
tgatccattt atcagaaatt gatattttgt tattagaaac gaccgatatt gtcttgacca 4260
ctatgctggt aaaacggcta gtggatgaga atgtccttgt catattctgt gatgataaac 4320
gattaccaac agctatgctg atgccttttt atggtcgtca tgattcgagt ttacagcttg 4380
ggaaacaaat gtcctggtca gaaacagtca aatcgcagggt ttggacgacg attattgctc 4440
aaaagatttt gaatcaatct tgctatctag gagcatgctc ctattttgaa aaatcccaat 4500
ctattatgga ttatatcatc ggtttggaat attttgatcc gagtaatcga gaagggcattg 4560
cagcgagaat ttattttaat acactttttg ggaacgattt ctcaagagat ttggagcatc 4620
caatcaatgc aggtctggat tatggttata ctttattatt gagtatgttt gcgcgtgaag 4680
tggttggtgc tggatgtatg actcaatttg gactcaaaca cgccaatcag tttaatcagt 4740
tcaattttgc tagcgatatt atggaaccat ttaggccttt ggtggataag attgtttatg 4800
aaaatcgaaa tcagcctttt cccaaaataa agagagagtt atttactttg ttttcagata 4860
cattttcata taatggtaaa gagatgtatc tcacgaatat tattagcgat tataactaaa 4920
aagttgtcaa agctctgaat aatgaagga aaggagttcc tgaatttagg atatgagtta 4980
tagatatatg agaatgatac ttatgtttga tatgccgacg gacactgctg aggaacgaaa 5040
agcttatcga aaatttcgga aatttttact tagtgaaggg tttatcatgc atcaattttc 5100
tatttatagt aagttactgt tgaataatac agctaacaac gccatgattg gtcggctgag 5160
ggagcataat ctcataaag gaaatattac attactaaca gtcacagaaa aacagtttgc 5220
acgaatgatt tatttacatg gtgaaagaaa taattgtatt gcaaactccg atgagagact 5280
tgtatttctt ggggaggctt ttgatgaatc ttaattttcc cttattagat gaaccgattc 5340
cattaagagg cggtaacaatt cttgtgctcg aagatgtctg tgtattttca aaaatagtgc 5400
aatattgtta caaatatgag gaagattctg aacttaaatt ttttgatcac aagatgaaaa 5460
ccatcaaaga atcagaaatc atgcttgtaa cagatatatt aggatttgat gtttaactct 5520
caaccatttt aaaattgatt catgcagatt tagaatctca atttaatgag aaacccgaag 5580

tgaaatcgat gattgacaaa ttggttgcta cgattacaga actgattgtc tttgaatgct 5640
tagaaaaatga attagattta gagtatgatg aaatcacaat cctggaattg attaatcctt 5700
taggagtaaa agtagaaacg caaagtgata ctatttttga aaaatgtcta gagatacttc 5760
aaattttcaa atatctcact aagaaaaagt tgcttatttt tgtcaatagc ggagcttttc 5820
taacaaagga tgaagtggct agtttacaag agtatatatc attgacaaat ttaacagttc 5880
tatttttaga agcagatgaa ctatatgatt ttccgagata tatttttagat gaagattatt 5940

```

tttttttaga accatgtgaa ctatctgact tttctgagta ttttttagat gaagattatt 5570
 tcttaataac taaaaatatg gtataa 5966

<210> 518

<211> 4107

<212> DNA

<213> *Streptococcus pyogenes*

<400> 518

atggataaga aatactcaat aggccttagat atcggcacaa atagcgtcgg atgggcggtg 60
 atcactgatg attataagggt tccgtctaaa aagctcaagg gtctgggaaa tacagaccgc 120
 cacggtatca aaaaaaatct tataggggct cttttatttg acagtggaga gacagcggaa 180
 gcgactcgtc tcaaaccggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240
 tatctacagg agattttttc aaatgagatg gcgaaagtag atgatagttt ctttcacatga 300
 cttgaagagt cttttttggt ggaagaagac aagaagcatg aacgtcatcc tttttttgga 360
 aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420
 aaattggcag attctactga taaagtggat ttgcgcttaa tctatttggc cttagcgcgt 480
 atgattaagt ttcgtggtca ttttttgatt gagggagatt taaatcctga taatagtgat 540
 gtggacaaac tatttatcca gttggtacaa acctacaatc aattatttga agaaaaccct 600
 attaacgcaa gtagagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga 660
 cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggatt gtttgggaat 720
 ctcatgtgtt tgtcattggg attgaccctt aattttaaat caaattttga tttggcagaa 780
 gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg 840
 caaattggag atcaatatgc tgatttggtt ttggcagcta agaatttatc agatgctact 900
 ttactttcag atatcctaag agtaaatagt gaaataacta aggcctccct atcagcttca 960
 atgattaagc gctacgatga acatcatcaa gacttgactc ttttaaagc tttagtctga 1020
 caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca 1080
 gggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta 1140
 gaaaaaatgg atggtactga ggaattattg gcgaaactaa atcgtgaaga tttgctgcgc 1200
 aagcaacgga cctttgacaa cggctctatt ccctatcaaa ttcacttggg tgagctgcat 1260
 gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt 1320
 gaaaaaatct tgacttttcg aattccttat tatgttgggc cattggcgcg tggcaatagt 1380

- 1380

cgttttgcgt ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa 1440
 gttgtcgata aagggtgctt agctcaatca tttattgaac gcatgacaaa ctttgataaa 1500
 aatcttccaa atgaaaaagt actacaaaaa catagtgttc tttatgagta ttttacggtt 1560
 tataacgaat tgacaaaagt caaatatgtt actgagggaa tgcgaaaacc agcatttctt 1620
 tcagggtgaa agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc 1680
 gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt 1740
 tcaggagtgt aagatagatt taatgcttca ttaggtacct accatgattt gctaaaaatt 1800
 attaaagata aagatttttt ggataatgaa gaaaacgaag atatcttaga ggatattgtt 1860
 ttaacattga ctttatttga agatagggag atgattgagg aaagacttaa aacatatgct 1920
 cacctctttt atdataaagt atataaacaa cttaaacttc accattatac taattaaaaa 1980

```

cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta 2040
gattttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat 2100
agtttgacat ttaaagaaga cattcaaaaa gcacaagtggt ctggacaagg cgatagttta 2160
catgaacata ttgcaaatTT agctggtagc cctgctatta aaaaggtat tttacagact 2220
gtaaaagtTg ttgatgaatt ggtCaaagta atggggcggc ataagccaga aaatatcgTt 2280
attgaaatgg cacgtgaaaa tcagacaact caaaagggcc agaaaaattc gcgtgagcgt 2340
atgaaacgta ttgaagaagg aataaaagaa ctaggaagtg atattctaaa ggagtatcct 2400
gttgaaaaca ctcaattaca aaatgaaaag ctctatctct attatctcca aaatggaaga 2460
gacatgtatg tggaccaaga attagatatt aatcgtttaa gtgattatga tgtcgatcac 2520
attgttccac aaagtttctt taaagacgat tcaatagaca ataaggTctt aacgcgttct 2580
gataaaaatc gtggtaaatc ggataacgTt ccaagtgaag aagtagtcaa aaagatgaaa 2640
aactattgga gacaacttct aaacgccaag ttaatcactc aacgtaagtt tgataattta 2700
acaaaagctg aacgtggagg tttgagtga cttgataaag ttggttttat caaacgcca 2760
ttggttgaaa ctgcgcaaat cactaagcat gtggcacaaa ttttggatag tcgcatgaat 2820
actaaatacg atgaaaatga taaacttatt cgagaggTta gagtgattac cttaaaatct 2880
aaattagtTt ctgacttccg aaaagatttc caattctata aagtacgtga gattaacaat 2940
taccatcatg cccatgatgc gtatcttaat gccgtcgttg gaactgcttt gattaagaaa 3000
tatccaaaac ttgaatcgga gtttgtctat ggtgattata aagtttatga tgttcgtaaa 3060
atgattgcta agtctgagca ggaaataggc aaagcaaccg caaaatattt cttttactct 3120
aatatcatga acttcttcaa aacagaaatt acacttgcaa atggagagat tcgcaaacgc 3180
cctctaactg aaactaatgg ggaaactgga gaaattgtct gggataaagg gcgagatttt 3240
gccacagtgc gcaaagtatt gtccatgccc caagtcaata ttgtcaagaa aacagaagta 3300
cagacaggcg gattctccaa ggagtcaatt ttaccaaaaa gaaattcgga caagcttatt 3360
gctcgtaaaa aagactggga tccaaaaaaa tatggtggTt ttgatagtcc aacggtagct 3420

tattcagtcc tagtggtTgc taaggTggaa aaagggaaat cgaagaagtt aaaatccgTt 3480
aaagagttac tagggatcac aataatggaa agaagctctt ttgaaaaaga tccgattgac 3540
tttttagaag ctaaaggata taagggaagtt agaaaagact taatcattaa actacctaaa 3600
tatagtcttt ttgagttaga aaacggTcgt aaacggatgc tggctagtgc cggagaattg 3660
caaaaaggaa atgagctagc tctgccaagc aaatatgtga attttttata tttagctagt 3720
cattatgaaa agttgaaggg tagtccagaa gataacgaac aaaaacaatt gtttgtggag 3780
cagcataagc attatttTga tgagattatt gagcaaatca gtgaattttc taagcgtgtt 3840
atttttagcag atgccaattt agataaagtt cttagtgcTt ataacaaca tagagacaaa 3900
ccaatacgtg aacaagcaga aaatattatt catttattta cgttgacgaa tcttggagct 3960
cccgctgctt ttaaattttt tgatacaaca attgatcgta aacgatatac gtctacaaaa 4020
gaagttttag atgccactct tatccatcaa tccatcactg gtctttatga aacacgcatt 4080
gatttgagtc agctaggagg tgactga 4107

```

<210> 519

<211> 870

<212> DNA

<213> *Streptococcus pyogenes*

<400> 519

atggctggtt ggcgtactgt tgtggtaa	60
at acccactcga aattatccta taagaataat	
catctgattt ttaaggatgc ctataaaacg	120
gagctgatcc atttatcaga aattgatatt	
ttgttattag aaacgaccga tattgtcttg	180
tccactatgc tggtaaaacg gctagtggat	
gagaatgtcc ttgtcatatt ctgtgatgat	240
aaacgattac caacagctat gctgatgcct	
ttttatggtc gtcgatgattc gagtttacag	300
cttgggaaac aaatgtcctg gtcagaaaca	
gtcaaatcgc aggtttggac gacgattatt	360
gctcaaaaga ttttgaatca atcttgctat	
ctaggagcat gctcctatct tgaaaaatcc	420
caatctatta tggattttata tcatggtttg	
gaaaattttg atccgagtaa tcgagaaggg	480
catgcagcga gaattttatt taatacactt	
tttgggaacg atttctcaag agatttggag	540
catccaatca atgcagggtct ggattatggt	
tatactttat tattgagtat gtttgcgcgt	600
gaagtgggtg tgtctggatg tatgactcaa	
tttggactca aacacgccaa tcagtttaat	660
cagttcaatt ttgctagcga tattatggaa	
ccatttaggc ctttgggtgga taagattggt	720
tatgaaaatc gaaatcagcc ttttcccaa	
ataaagagag agttattttac tttgttttca	780
gatacatctt catataatgg taaagagatg	
tatctcacga atattattag cgattatact	840
aaaaaagttg tcaaagctct gaataatgaa	
gggaaaggag ttcctgaatt taggatatga	870

<210> 520

<211> 327

<212> DNA

<213> *Streptococcus pyogenes*

<400> 520

atgagaatga tacttatggt tgatatgccg	60
acggacactg ctgaggaaacg aaaagcttat	
cgaaaatttc ggaatttttt acttagtgaa	120
gggtttatca tgcattcaatt ttctatttat	
agtaagtac tgttgaataa tacagctaac	180
aacgccatga ttggtcggct gagggagcat	
aatcctcata aaggaaatat tacattacta	240
acagtcacag aaaaacagtt tgcacgaatg	
atttatttac atggtgaaag aaataattgt	300
attgcaact ccgatgagag acttgatatt	
cttggggagg cttttgatga atcttaa	327

<210> 521

<211> 663

<212> DNA

<213> *Streptococcus pyogenes*

<400> 521

atgaatctta attttccctt attagatgaa	60
ccgattccat taagaggcgg tacaattctt	
gtgctcgaag atgtctgtgt attttcaaaa	120
atagtgcatt attgttaca atatgaggaa	
gattctgaac ttaatttttt tgatcacaag	180
atgaaaacca tcaaagaatc agaaatcatg	
cttgtaacag atatttttagg atttgatggt	240
aactcctcaa ccatttttaa attgattcat	
gcagatttag aatctcaatt taatgagaaa	300
cccgaagtga aatcgatgat tgacaaattg	
gttgctacga ttacagaact gattgtcttt	360
gaatgcttag aaaatgaatt agatttagag	
tatgatgaaa tcacaatcct ggaattgatt	420
aagtccttag gagtaaaagt agaaacgcaa	

```

agtgatacta tttttgaaaa atgtctagag atacttcaaa ttttcaaata tctcactaag    480
aaaaagttgc ttatTTTTgt caatagcggg gcttttctaa caaaggatga agtggctagt    540
ttacaagagt atatatcatt gacaaattta acagttctct ttttagaacc acgtgaacta    600
tatgattttc cgcagtatat tttagatgaa gattatttct taataactaa aaatatggta    660
taa                                                                    663

```

<210> 522

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 522

```

caacacattc aacagattaa tgaagaatac    30

```

<210> 523

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 523

```

tccactcacg tacaaatagt gagtgactc    30

```

<210> 524

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 524

```

gcccttctaa ttggattacc ttccgaggtg    30

```

<210> 525

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 525

```

ctcagtcgtt actggtgaac cagtttcaat    30

```

<210> 526

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 526

```

attgtctatt acgacaacat ggaagatgat    30

```

<210> 527

<211> 31

<212> DNA

<213> Streptococcus thermophilus

<400> 527

gagtttcttt gtcagactct aacacagccg c 31

<210> 528

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 528

ttactagagc gtgtcgtaa ccactttaa 30

<210> 529

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 529

ttcgtaaag tcacctgtg ctagcgttc 30

<210> 530

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 530

ataacggtag caaatataaa cctgttactg 30

<210> 531

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 531

gaagtagcca tacaagaaga tggatcagca 30

<210> 532

<211> 29

<212> DNA

<213> Streptococcus thermophilus

<400> 532

atgtcactga gtgtctaagc attgcgtac 29

<210> 533

<211> 30

<212> DNA
<213> Streptococcus thermophilus

<400> 533
tgaataagca gttcttgacg accaaccgac 30

<210> 534
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 534
tcaacaattg caacatctta taaccactt 30

<210> 535
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 535
ttacgtttga aaagaatatc aaatcaatga 30

<210> 536
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 536
gctctacgac ttcttcacg agttcctgcc 30

<210> 537
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 537
aacacagcaa gacaagagga tgatgctatg 30

<210> 538
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 538
aagtagttga tgacctctac aatggttat 30

<210> 539
<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 539

aataatttat ggtatagctt aatatcattg 30

<210> 540

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 540

aatcaatacg acaagagtta aaatggtctt 30

<210> 541

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 541

aatcgttcaa attctgtttt aggtacattt 30

<210> 542

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 542

aatgacgagg agctattggc acaacttaca 30

<210> 543

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 543

aattaagggc atagaaaggg agacaacatg 30

<210> 544

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 544

acaattcttc atccggtaac tgctcaagtg 30

<210> 545

<211> 30

<212> DNA
<213> Streptococcus thermophilus

<400> 545
acacttggca ggcttattac tcaacagcga 30

<210> 546
<211> 30
<212> DNA
<213> streptococcus thermophilus

<400> 546
ataaactatg aaattttata attttaaga 30

<210> 547
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 547
ataactgaag gataggagct tgtaaagtct 30

<210> 548
<211> 29
<212> DNA
<213> Streptococcus thermophilus

<400> 548
ataatgccgt tgaattacac ggcaagtca 29

<210> 549
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 549
caaccaacgg taacagctac tttttacagt 30

<210> 550
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 550
catagagtgg aaaactagaa acagattcaa 30

<210> 551
<211> 29

<212> DNA
<213> Streptococcus thermophilus

<400> 551
cgacacaaga acgtatgcaa gaggtaag 29

<210> 552
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 552
cgatatttaa aatcatttc ataactcat 30

<210> 553
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 553
cgattgaca atctgctgac cactgtatc 30

<210> 554
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 554
ctgttccttg ttcttttgtt gtatctttc 30

<210> 555
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 555
gagcgagctc gaaataatct taattacaag 30

<210> 556
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 556
gcagtatcag caagcaagct gtagttact 30

<210> 557
<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 557

gctggcgagg aaacgaaca ggcctcaaca 30

<210> 558

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 558

gcttagctgt ccaatccacg aacgtggatg 30

<210> 559

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 559

ggcgtcccaa tctgattaa tacttactcg 30

<210> 560

<211> 30

<212> DNA

<213> streptococcus thermophilus

<400> 560

gttcgctagc gtcattgtgt aacgtattta 30

<210> 561

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 561

tctatatcga ggtcaactaa caattatgct 30

<210> 562

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 562

tgcattgagc acgttcgagt ttaccgtttc 30

<210> 563

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 563

tgtttgacag caaatcaaga ttcgaattgt 30

<210> 564

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 564

ttcattcttc cgtttttgtt tgcgaatcct 30

<210> 565

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 565

tgacttagcg aatttaatcg ctaagatc 30

<210> 566

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 566

tttatacttt atctttttaa agaattgtatt 30

<210> 567

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 567

cctaaaatca tttcaacga gttgcgatac 30

<210> 568

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 568

aataaattgc tatgatacag cgtaccgata 30

<210> 569

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 569

tgctctctat gcgattggac gtctgtctaa 30

<210> 570

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 570

aagaaagata agaaaaaagt aacactactt 30

<210> 571

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 571

tctctttcca tcggtactgg tataatctcat 30

<210> 572

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 572

attggtagcc aagtaaataat caccattgat 30

<210> 573

<211> 30

<212> DNA

<213> streptococcus thermophilus

<400> 573

ttctcaaata tcaccgactg caaaattaca 30

<210> 574

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 574

gcttcctaag tgcataaaaa tcgcaaacgg 30

<210> 575

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 575

tatacctgtc tatgtaaggg aatttaactc 30

<210> 576

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 576

ggtaggtg ctgttggttaa gttgtttaat 30

<210> 577

<211> 30

<212> DNA

<213> streptococcus thermophilus

<400> 577

gtgaaacagg ttatcaaaaa acgtatattg 30

<210> 578

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 578

ttattcttgg aattattaca gaccctacta 30

<210> 579

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 579

gctttcatta tatcacttac tcataaatct 30

<210> 580

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 580

taatcacccc ttttctagc tcttgattga 30

<210> 581

<211> 30

<212> DNA
<213> Streptococcus thermophilus

<400> 581
caagcagtgt aaaggtggtt taaatgttaa 30

<210> 582
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 582
aaccgcgtg gttatgggct tgaggagtgt 30

<210> 583
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 583
atattaatag cgattctatg ctacaacgtg 30

<210> 584
<211> 30
<212> DNA
<213> streptococcus thermophilus

<400> 584
tcatttcta agtaaatacc actgtcaggg 30

<210> 585
<211> 29
<212> DNA
<213> Streptococcus thermophilus

<400> 585
tttcgcaaa gtaagcgaag ctctacgtg 29

<210> 586
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 586
ttctgtagcc actccgtgga tgccttcagc 30

<210> 587
<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 587

ttcttttagtt cggacaccct caacacctat 30

<210> 588

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 588

gctttgattg gacggaaaat ggtatccctg 30

<210> 589

<211> 31

<212> DNA

<213> *Streptococcus thermophilus*

<400> 589

ttcctcatct ttctccgctt ttgctagact t 31

<210> 590

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 590

ttagaccaga tggacagata ttcttcatcg 30

<210> 591

<211> 31

<212> DNA

<213> *Streptococcus thermophilus*

<400> 591

tcatcagagt caacaatcac gggaaagacc t 31

<210> 592

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 592

acactcatcc ttatcctgta gtcaaaaca 30

<210> 593

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 593

cagcactagc cgcaagccct tgtatattaa 30

<210> 594

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 594

tagaaatcaa ggaacttgga tgaaaagtaa 30

<210> 595

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 595

atatgaaagg gaaatgatat gaagaatgaa 30

<210> 596

<211> 31

<212> DNA

<213> Streptococcus thermophilus

<400> 596

ttttgggata caacacgcag tcgttgactt g 31

<210> 597

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 597

gtttgagatg ccaatgtttt tcaatccttg 30

<210> 598

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 598

gtatcaaaag acgcattcat gaagcgagct 30

<210> 599

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 599

aaaaacaatt gaaattcata atcagcgctt 30

<210> 600

<211> 29

<212> DNA

<213> *Streptococcus thermophilus*

<400> 600

gcttttaacg tttaagaga atacctct 29

<210> 601

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 601

gtgacgctgc aatgactgc catagtaatt 30

<210> 602

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 602

atactggtat atagtaattc atacttcac 30

<210> 603

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 603

ttggtttcat atttactcct ttgtgtttg 30

<210> 604

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 604

ctgatttggt ctgttcttt tgcctttt 30

<210> 605

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 605

gcagcagttg agaacttag cgtccagtg 30

<210> 606

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 606

tgctactatg aaggacgctg ttgatactt 30

<210> 607

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 607

tcttctttaa tctttttaa cgtcaacgtt 30

<210> 608

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 608

gatatcatta atagtagc atttctatca 30

<210> 609

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 609

attcattaat atctgcaagg atgtcttgtt 30

<210> 610

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 610

gagaaagtag ccattcggc ccattcgggg 30

<210> 611

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 611

tacttgagtt agctctggaa gtcattatc 30

<210> 612

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 612

ctgcattgt aaccatgact tctcgtcgt 30

<210> 613

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 613

aatttgcat cgacatctac caacgccag 30

<210> 614

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 614

ataaaattat gccacgtttt ggcactagat 30

<210> 615

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 615

atgtctctga ggctgtagta atttactgt 30

<210> 616

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 616

ctttaaagag ttgattaagt gcgttactgt 30

<210> 617

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 617

aaatgggtta tgctgtcaa tatgcgtccc 30

<210> 618

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 618

aaactgaaaa caacacagac aattcaaca 30

<210> 619

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 619

gccccaaatg ctagacgttt gaatgacggc 30

<210> 620

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 620

atgaagaacg tgattcacct acggtatgct 30

<210> 621

<211> 31

<212> DNA

<213> Streptococcus thermophilus

<400> 621

gcttttgcag aattgtctcc agtgccgatt t 31

<210> 622

<211> 30

<212> DNA

<213> streptococcus thermophilus

<400> 622

tgtactctat tgattgcttc atctttatta 30

<210> 623

<211> 31

<212> DNA

<213> Streptococcus thermophilus

<400> 623

cttcaagat actcatcaac cattgatgc a 31

<210> 624

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 624

ctatgtcttt actgttcttc caaaaccacc 30

<210> 625

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 625

tgctacgtgc tctgtacggg cgctatcagc 30

<210> 626

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 626

cgtaggcagcg tggtaggggtt taatagcccg 30

<210> 627

<211> 29

<212> DNA

<213> Streptococcus thermophilus

<400> 627

aagcccaagt cagagcatcc gtccaagcc 29

<210> 628

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 628

attgggttgc ggtaagaact aaacatacca 30

<210> 629

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 629

cacaaaataa ttcggtagtt ttactaact 30

<210> 630

<211> 29

<212> DNA

<213> *Streptococcus thermophilus*

<400> 630

tttgaccgtt tatttagacg tgctaaagt 29

<210> 631

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 631

cttcacctca aatcttagag ctggactaaa 30

<210> 632

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 632

atgtctgaaa aataaccgac catcattact 30

<210> 633

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 633

gaagctcatc atgttaaggc taaaacctat 30

<210> 634

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 634

tagtctaaat agatttcttg caccattgta 30

<210> 635

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 635

attcgtgaaa aaatatcgtg aaataggcaa 30

<210> 636

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 636

tctaggctca tctaaagata aatcagtagc 30

<210> 637

<211> 31

<212> DNA

<213> *Streptococcus thermophilus*

<400> 637

taaaaacatg gggcggcggg aatagtgtaa g 31

<210> 638

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 638

acaaccagca aagagagcgc cgacaacatt 30

<210> 639

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 639

tataacacag gtttagagga tgttatactt 30

<210> 640

<211> 31

<212> DNA

<213> *Streptococcus thermophilus*

<400> 640

ctagaagctc aagcggtaaa agttgatggc g 31

<210> 641

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 641

cttgagggc aagccctcgc cgttcattt 30

<210> 642

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 642

aactaccaag caaatcagca atcaataagt 30

<210> 643

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 643

ctataagtga caatcagcgt agggaatacg 30

<210> 644

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 644

atcagtgcgg tatattacc ctagacgcta 30

<210> 645

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 645

aacagttact attaatacag attccaacgg 30

<210> 646

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 646

aattagggcg tcttccttta ttccgtggtt 30

<210> 647

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 647

atagcttcat tgcgctttt aattgacct 30

<210> 648

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 648

aacaacaaag caatacaac agtaacaacc 30

<210> 649

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 649

ctaaactacg ttgaaggct tcaactccgt 30

<210> 650

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 650

gagggtgaat agtgagtgc ccatgtttgt 30

<210> 651

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 651

agtagagaga ccagcacact actgtactac 30

<210> 652

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 652

cttcgcacga aagtttatta gacaactcgc 30

<210> 653

<211> 30

<212> DNA
<213> Streptococcus thermophilus

<400> 653
tgatagagct agaattgtct ttttaccga 30

<210> 654
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 654
agatactctt gctcgctct gaacaaccag 30

<210> 655
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 655
ggtgaaaaag gttcactgta cgagtactta 30

<210> 656
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 656
tcaatgagtg gtatccaaga cgaaaactta 30

<210> 657
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 657
ccttgctgtg gctctcata cgcccatata 30

<210> 658
<211> 30
<212> DNA
<213> Streptococcus thermophilus

<400> 658
tgtttgggaa accgcagtag ccatgattaa 30

<210> 659
<211> 32

<212> DNA

<213> *Streptococcus thermophilus*

<400> 659

acagagtaca atattgtcct cattggagac ac 32

<210> 660

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 660

ctcatattcg ttagttgctt ttgtcataaa 30

<210> 661

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 661

agaactttat caagataaaa ctactttaa 30

<210> 662

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 662

atagtattaa tttcattgaa aaataattgt 30

<210> 663

<211> 36

<212> DNA

<213> *Streptococcus thermophilus*

<400> 663

gctttctagc tcgtataat taccattcc tagaaa 36

<210> 664

<211> 40

<212> DNA

<213> *Streptococcus thermophilus*

<400> 664

tcaaaatag ttattacctt gtattcata attcaattaa 40

<210> 665

<211> 39

<212> DNA

<213> *Streptococcus thermophilus*

<400> 665

ccacttgctg tgtacatcct accagttccg cctatgatg 39

<210> 666

<211> 63

<212> DNA

<213> *Streptococcus thermophilus*

<400> 666

caaggacagt tattgatttt ataatcacta tgtgggtata aaaacgtcaa aatttcattt 60

gag 63

<210> 667

<211> 41

<212> DNA

<213> *Streptococcus thermophilus*

<400> 667

ttgattcaac ataaaaagcc agttcaattg aacttggtt t 41

<210> 668

<211> 2317

<212> DNA

<213> *Streptococcus thermophilus*

<400> 668

caaggacagt tattgatttt ataatcacta tgtgggtata aaaacgtcaa aatttcattt 60

gagggtttttg tactctcaag atttaagtaa ctgtacaact caacaattgc aacatcttat 120

aaccacttg tttttgtact ctcaagattt aagtaactgt acaactgttt gacagcaaat 180

caagattcga attgtgtttt tgtactctca agatttaagt aactgtacaa caatgacgag 240

gagctattgg cacaacttac agtttttcta ctctcaagat ttaagtaact gtacaaccga 300

tttgacaatc tgctgaccac tgttatcggt tttgtactct caagatttaa gtaactgtac 360

aacacacttg gcaggcttat tactcaacag cgagtttttg tactctcaag atttaagtaa 420

ctgtacaacc tgttccttgt tcttttgttg tatcttttcg tttttgtact ctcaagattt 480

aagtaactgt acaacttcat tcttccgttt ttgtttgcga atcctgtttt tgtactctca 540

agatttaagt aactgtacaa cgctggcgag gaaacgaaca aggcctcaac agtttttcta 600

ctctcaagat ttaagtaact gtacaaccat agagtggaaa actagaaaca gattcaagtt 660

tttgtactct caagatttaa gtaactgtac aacataatgc cgttgaatta cacggcaagg 720

tcagtttttg tactctcaag atttaagtaa ctgtacaacg agcgagctcg aaataatctt 780

aattacaagg tttttgtact ctcaagattt aagtaactgt acaacgttcg ctacgctcat 840

gtggtaacgt atttagtttt tgtactctca agatttaagt aactgtacaa cggcgctcca 900

atcctgatta atacttactc gggttttcta ctctcaagat ttaagtaact gtacaacaac 960

acacgaagac aagaggaatga tcttatggtt tttgtactct caagatttaa gtaactctac 1020

```

aaccgacaca agaacgtatg caagagttca aggtttttgt actctcaaga ttttaagtaac 1080
tgtacaacac aattcttcat ccggttaactg ctcaagtggg ttttgtaactc tcaagattta 1140
agtaactgta caacaattaa gggcatagaa agggagacaa catgggtttt gtactctcaa 1200
gatttaagta actgtacaac cgatatataa aatcattttc ataacttcat gtttttgtac 1260
tctcaagatt taagtaactg tacaacgcag tatcagcaag caagctgtta gttactgttt 1320
ttgtactctc aagatttaag taactgtaca acataaacta tgaaatttta taatttttaa 1380
gagtttttgt actctcaaga ttttaagtaac tgtacaacaa taatttatgg tatagcttaa 1440
tatcattggg ttttgtaactc tcaagattta agtaactgta caactgcac gagcacgttc 1500
gagtttaccg tttcgttttt gtactctcaa gatttaagta actgtacaac tctatatcga 1560
ggtaactaa caattatgct gtttttgtac tctcaagatt taagtaactg tacaacaatc 1620
gttcaaattc tgttttaggt acatttgttt ttgtactctc aagatttaag taactgtaca 1680
acaatcaata cgacaagagt taaaatgggc ttgtttttgt actctcaaga ttttaagtaac 1740
tgtacaacgc ttagctgtcc aatccacgaa cgtggatggg ttttgtaactc tcaagattta 1800
agtaactgta caaccaacca acggtaacag ctacttttta cagtgttttt gtactctcaa 1860

gatttaagta actgtacaac ataactgaag gataggagct tgtaaagtct gtttttgtac 1920
tctcaagatt taagtaactg tacaactaat gctacatctc aaaggatgat ccagagttt 1980
ttgtactctc aagatttaag taactgtaca acaagtagtt gatgacctct acaatggttt 2040
atgtttttgt actctcaaga ttttaagtaac tgtacaacac ctagaagcat ttgagcgtat 2100
attgattggg ttttgtaactc tcaagattta agtaactgta caacaatttt gccccttctt 2160
tgccccttga ctaggttttt gtactctcaa gatttaagta actgtacaac accattagca 2220
atcatttgtg ccatttgagt gtttttgtac tctcaagatt taagtaactg tacagtttga 2280
ttcaacataa aaagccagtt caattgaact tggctttt 2317

```

<210> 669

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 669

tcaacaattg caacatctta taaccactt 30

<210> 670

<211> 2317

<212> DNA

<213> *Streptococcus thermophilus*

<400> 670

```

caaggacagt tattgatttt ataactacta tgtgggtata aaaacgtcaa aatttcattt 60
gaggtttttg tactctcaag atttaagtaa ctgtacaact tacgtttgaa aagaatatca 120
aatcaatgag tttttgtact ctcaagattt aagtaactgt acaactgttt gacagcaaat 180
caagattcga attgtgtttt tgtactctca agatttaagt aactgtacaa caatgacgag 240
gagctattgg cacaacttac agtttttcta ctctcaagat ttaagtaact gtacaaccga 300

```

```

tttgacaatc tgcgtaccac tgttatcgtt ttgtatctct caagatttaa gtaactctac 300
aacacacttg gcaggcttat tactcaacag cgagtttttg tactctcaag atttaagtaa 420
ctgtacaacc tgttccttgt tcttttggtg tatcttttcg tttttgtact ctcaagattt 480
aagtaactgt acaacttcat tcttccgttt ttgtttgcga atcctgtttt tgtactctca 540
agatttaagt aactgtacaa cgctggcgag gaaacgaaca aggcctcaac agtttttgta 600
ctctcaagat ttaagtaact gtacaacat agagtggaaa actagaaaca gattcaagtt 660
tttgactctt caagatttaa gtaactgtac aacataatgc cgttgaatta cacggcaagg 720
tcagtttttg tactctcaag atttaagtaa ctgtacaacg agcgagctcg aaataatctt 780
aattacaagg tttttgtact ctcaagattt aagtaactgt acaacgttcg ctacgctcat 840
gtggtaacgt atttagtttt tgtactctca agatttaagt aactgtacaa cggcgtccca 900
atcctgatta atacttactc gggtttttgta ctctcaagat ttaagtaact gtacaacaac 960
acagcaagac aagaggatga tgctatggtt ttgtactctt caagatttaa gtaactgtac 1020
aaccgacaca agaacgtatg caagagttca aggtttttgt actctcaaga ttaagtaac 1080

tgtacaacac aattcttcat ccggttaactg ctcaagtggg ttttgtactc tcaagattta 1140
agtaactgta caacaattaa gggcatagaa agggagacaa catggttttt gtactctcaa 1200
gatttaagta actgtacaac cgatatataa aatcattttc ataacttcat gttttgtac 1260
tctcaagatt taagtaactg tacaacgcag tatcagcaag caagctgtta gttactgttt 1320
ttgtactctc aagatttaag taactgtaca acataaacta tgaaatttta taatttttaa 1380
gagttttgtt actctcaaga ttttaagtaac tgtacaacaa taatttatgg tatagcttaa 1440
tatcattggt ttttgtactc tcaagattta agtaactgta caactgcac gagcacgttc 1500
gagtttaccg ttctgttttt gtactctcaa gatttaagta actgtacaac tctatatcga 1560
ggtaactaa caattatgct gttttgtac tctcaagatt taagtaactg tacaacaatc 1620
gttcaaattc tgttttaggt acatttggtt ttgtactctc aagatttaag taactgtaca 1680
acaatcaata cgacaagagt taaaatggtc ttgtttttgt actctcaaga ttttaagtaac 1740
tgtacaacgc ttagctgtcc aatccacgaa cgtggatggg ttttgtactc tcaagattta 1800
agtaactgta caaccaacca acggtaacag ctacttttta cagtgttttt gtactctcaa 1860
gatttaagta actgtacaac ataactgaag gataggagct tgtaaagtct gttttgtac 1920
tctcaagatt taagtaactg tacaactaat gctacatctc aaaggatgat ccagagttt 1980
ttgtactctc aagatttaag taactgtaca acaagtagtt gatgacctct acaatggttt 2040
atgtttttgt actctcaaga ttttaagtaac tgtacaacac ctagaagcat ttgagcgtat 2100
attgattggt ttttgtactc tcaagattta agtaactgta caacaatttt gcccttctt 2160
tgcccttga ctaggttttt gtactctcaa gatttaagta actgtacaac accattagca 2220
atcatttggt ccattgagt gttttgtac tctcaagatt taagtaactg tacagtttga 2280
ttcaacataa aaagccagtt caattgaact tggcttt 2317

```

<210> 671

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 671

ttacgtttga aaagaatatc aatcaatga 30

<210> 672

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 672

caaatggata gagaaacgc 19

<210> 673

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 673

ctgataaggt gttcgtgtgc c 21

<210> 674

<211> 24

<212> DNA

<213> Artificial sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 674

ggagcagatg gaatacaaga aagg 24

<210> 675

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 675

gagagactag gttgtctcag ca 22

<210> 676

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 676

acaaacaaca gagaagtatc tcattg 26

<210> 677

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 677

aacgagtaca ctactatctt gtacg 25

<210> 678

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 678

cttccttca tctcgcttt gggt 24

<210> 679

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 679

ttctggtagt gggttagtc aaacagatgt 30

<210> 680

<211> 30

<212> DNA

<213> *Streptococcus thermophilus*

<400> 680

ttctggtagt gggttagtc aaacagatgt 30

<210> 681

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 681

ttctgtagt ggatttagt aaacagatgt 30

<210> 682

<211> 30

<212> DNA

<213> Streptococcus thermophilus

<400> 682

ttctgtagt ggttttagt aaacagatgt 30

<210> 683

<211> 30

<212> DNA

<213> streptococcus thermophilus

<400> 683

ttctgtagt ggttttagt aaacagatgt 30

<210> 684

<211> 23

<212> DNA

<213> Streptococcus thermophilus

<400> 684

tctgtagtg gatttagtca aac 23

<210> 685

<211> 23

<212> DNA

<213> Streptococcus thermophilus

<400> 685

tctgtagtg gatttagtca aac 23

<210> 686

<211> 26

<212> DNA

<213> Streptococcus thermophilus

<400> 686

ggtagtggat ttagtcaaac agatgt 26

<210> 687

<211> 26

<212> DNA

<213> *Streptococcus thermophilus*

<400> 687

ggtagtggct ttagtcaaac agatgt 26

<210> 688

<211> 28

<212> DNA

<213> *Streptococcus thermophilus*

<400> 688

ttctgtagt ggatttagtc aaacagat 28

<210> 689

<211> 28

<212> DNA

<213> *Streptococcus thermophilus*

<400> 689

ttctggcagt ggttttagtc aaacagat 28

<210> 690

<211> 36

<212> RNA

<213> *Streptococcus thermophilus*

<400> 690

guuuuagagc uguguuguuu cgaaugguuc caaaac 36

<210> 691

<211> 222

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic nucleotide construct

<400> 691

caggaaacag ctatgaccat gattacgcc a gcttggtac cgagctcga tccactagta	60
acggccgccga gtgtgctgga attcgccctt aagggcgaat tctgcagata tccatcacac	120
tggcggccgc tcgagcatgc atctagaggg cccaattcgc cctatagtga gtcgtattac	180
aattcactgg ccgtcgtttt acaacgtcgt gactgggaaa ac	222

<210> 692

<211> 84

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 692

tccactcacg tacaatagt gagggtactc gttttgtat tctcaagatt taagtaactg 60

tacagtttga ttcaacataa aaag 84

<210> 693

<211> 8

<212> PRT

<213> *Streptococcus thermophilus*

<400> 693

Asp Cys Phe Cys Thr Leu Lys Ile
1 5

<210> 694

<211> 21

<212> PRT

<213> *Streptococcus thermophilus*

<400> 694

Val Thr Val Gln Leu Leu Thr Ser Ser Val His Arg Leu Lys Cys Phe
1 5 10 15Cys Thr Leu Lys Ile
20

<210> 695

<211> 16

<212> PRT

<213> *Streptococcus thermophilus*

<400> 695

Val Thr Val Gln Pro Pro Leu Ser Met Glu Arg Tyr Pro Ser Ser Phe
1 5 10 15

<210> 696

<211> 10

<212> PRT

<213> *Streptococcus thermophilus*

<400> 696

Lys Ile Val Phe Val Leu Ser Arg Phe Lys
1 5 10

<210> 697

<211> 4

<212> PRT

<213> *Streptococcus thermophilus*

<400> 697

Leu Tyr Asn Phe
1

<210> 698

<211> 16

<212> PRT

<213> steptococcus thermophilus

<400> 698

Arg His Gln Ser Thr Ala Leu Asn Val Phe Val Leu Ser Arg Phe Lys
1 5 10 15

<210> 699

<211> 15

<212> PRT

<213> Steptococcus thermophilus

<400> 699

Leu Tyr Asn His Leu Phe Arg Trp Lys Gly Ile Leu Leu Val Phe
1 5 10 15

<210> 700

<211> 22

<212> PRT

<213> Steptococcus thermophilus

<400> 700

Arg Leu Phe Leu Tyr Ser Gln Asp Leu Ser Asn Cys Thr Thr Phe Asn
1 5 10 15Val Ile Ser Pro Pro Pro
20

<210> 701

<211> 22

<212> PRT

<213> Steptococcus thermophilus

<400> 701

Met Phe Leu Tyr Ser Gln Asp Leu Ser Asn Cys Thr Thr Thr Ser Phe
1 5 10 15Asp Gly Lys Val Ser Phe
20

<210> 702

<211> 143

<212> DNA

<213> Steptococcus thermophilus

<400> 702

agattgtttt tgtactctca agatttaagt aactgtacaa cttttaacgt catcagtcca 60
agacttttaa ttttttttata ctatcagact ttacgttact ttacgttttt ttatttttact 120

ccgccccaaa tgcctcttga cctcccaagat ttaagtaact gtaacacccat cctctctgac 120
 ggaaagggtat ccttctagtt ttt 143

<210> 703

<211> 143

<212> DNA

<213> Streptococcus thermophilus

<400> 703

tctaacaaaa acatgagagt tctaaattca ttgacatggt gaaaattgca gtagtcaggt 60

ggcgggaattt acaaaaacat gagagttcta aattcattga catgttggtg gagaaagcta 120

cctttccata ggaagatcaa aaa 143

<210> 704

<211> 6

<212> PRT

<213> Streptococcus thermophilus

<400> 704

Leu Asn Asn Lys Tyr Glu
 1 5

<210> 705

<211> 15

<212> PRT

<213> Streptococcus thermophilus

<400> 705

Ser Lys Leu Leu Gln Val Val Lys Leu Thr Met Leu Gly Gly Gly
 1 5 10 15

<210> 706

<211> 5

<212> PRT

<213> Streptococcus thermophilus

<400> 706

Ile Asn Lys Tyr Glu
 1 5

<210> 707

<211> 16

<212> PRT

<213> Streptococcus thermophilus

<400> 707

Ser Lys Leu Leu Gln Val Val Val Glu Lys Ser Pro Phe Thr Asp Lys
 1 5 10 15

<210> 708

<211> 8

<212> PRT

<213> Streptococcus thermophilus

<400> 708

Ser	Gln	Lys	Gln	Val	Arg	Leu	Ile
1			5				

<210> 709

<211> 21

<212> PRT

<213> Streptococcus thermophilus

<400> 709

Thr	Val	Thr	Cys	Ser	Lys	Val	Asp	Asp	Thr	Trp	Arg	Arg	Leu	His	Lys
1				5					10					15	

Gln	Val	Arg	Leu	Ile
			20	

<210> 710

<211> 17

<212> PRT

<213> Streptococcus thermophilus

<400> 710

Thr	Val	Thr	Cys	Gly	Gly	Arg	Glu	Ile	Ser	Leu	Tyr	Gly	Glu	Leu	Lys
1				5					10					15	

Gln

<210> 711

<211> 14

<212> PRT

<213> Streptococcus thermophilus

<400> 711

Ile	Thr	Lys	Thr	Ser	Glu	Leu	Asn	Leu	Tyr	Ser	Tyr	Leu	Lys
1				5					10				

<210> 712

<211> 29

<212> PRT

<213> Streptococcus thermophilus

<400> 712

Asp	Val	Ala	Lys	Phe	Thr	Lys	Thr	Ser	Glu	Leu	Asn	Leu	Tyr	Ser	Tyr
1				5					10					15	

Leu	Trp	Arg	Lys	Arg	His	Phe	Pro	Ile	Arg	Arg	Thr	Lys
			20					25				

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

- [WO0114520A](#) [\[0006\]](#)
- [US4683195A](#) [\[0109\]](#)
- [US4683202A](#) [\[0109\]](#)
- [US2225783A](#) [\[0481\]](#)
- [US3024116A](#) [\[0482\]](#)
- [US3403032A](#) [\[0482\]](#)
- [US3932674A](#) [\[0482\]](#)
- [US3897307A](#) [\[0482\]](#)
- [US4621058A](#) [\[0541\]](#)
- [US4423079A](#) [\[0542\]](#)
- [US4140800A](#) [\[0542\]](#)
- [US06033167W](#) [\[0676\]](#)
- [US60711396B](#) [\[0676\]](#)
- [US60747683B](#) [\[0676\]](#)

Non-patent literature cited in the description

- **PEDERSEN et al.** 7th symposium on lactic acid bacteria: genetics, metabolism and applications, 2002, [\[0005\]](#)
- **KOSUGE et al.** Appl. Environ. Microbiol., 1998, vol. 64, 114328-4332 [\[0007\]](#)
- **KOSUGE et al.** FEMS Microbiology Letters, 1994, vol. 123, 1/255-62 [\[0007\]](#)
- **RUSSELL, M. W.T. R. KLAENHAMMER** Efficient system for directed integration into the Lactobacillus acidophilus and Lactobacillus gasseri chromosomes via homologous recombination Applied and Environmental Microbiology, 2001, vol. 67, 4361-4364 [\[0100\]](#)
- **ISHINO et al.** J. Bacteriol., 1987, vol. 169, 5429-5433 [\[0101\]](#)
- **NAKATA et al.** J. Bacteriol., 1989, vol. 171, 3553-3556 [\[0101\]](#)
- **GROENEN et al.** Mol. Microbiol., 1993, vol. 10, 1057-1065 [\[0101\]](#)

- **HOE et al.**Emerg. Infect. Dis., 1999, vol. 5, 254-263 [\[0101\]](#)
- **MASEPOHL et al.**Biochim. Biophys. Acta, 1996, vol. 1307, 26-30 [\[0101\]](#)
- **MOJICA et al.**Mol. Microbiol., 1995, vol. 17, 85-93 [\[0101\]](#)
- **JANSSEN et al.**OMICS J. Integ. Biol., 2002, vol. 6, 23-33 [\[0101\]](#)
- **MOJICA et al.**Mol. Microbiol., 2000, vol. 36, 244-246 [\[0101\]](#) [\[0101\]](#)
- **VAN EMBDEN et al.**J. Bacteriol., 2000, vol. 182, 2393-2401 [\[0101\]](#)
- **DIEFFENBACHDVEKSLER**PCR Primer, a Laboratory ManualCold Spring Harbor Press19950000 [\[0108\]](#)
- **SAMBROOK et al.**Molecular Cloning: A Laboratory ManualCold Spring Harbor Laboratory Press19890000 [\[0111\]](#) [\[0515\]](#) [\[0517\]](#)
- PCR Protocols: A Guide to Methods and ApplicationsAcademic Press19900000 [\[0111\]](#)
- PCR StrategiesAcademic Press19950000 [\[0111\]](#)
- PCR Methods ManualAcademic Press19990000 [\[0111\]](#)
- **HAFT et al.**Computational Biology, 2005, vol. 1, 6e60- [\[0123\]](#)
- **MORINAGA et al.**Biotechnology, 1984, vol. 2, 646-649 [\[0419\]](#)
- **NELSONLONG**Analytical Biochemistry, 1989, vol. 180, 147-151 [\[0419\]](#)
- **SARKARSOMMER**The megaprimer method of site directed mutagenesisBiotechniques, 1990, vol. 8, 404-407 [\[0419\]](#)
- **J. SAMBROOKE. F. FRITSCHT. MANIATIS**Molecular Cloning: A Laboratory ManualCold Spring Harbor Laboratory Press19890000 [\[0419\]](#) [\[0590\]](#)
- **AUSUBEL, F. M. et al.**Current Protocols in Molecular BiologyJohn Wiley & Sons19950000 [\[0419\]](#) [\[0590\]](#)
- **B. ROEJ. CRABTREEA. KAHN**DNA Isolation and Sequencing: Essential TechniquesJohn Wiley & Sons19960000 [\[0419\]](#) [\[0590\]](#)
- Oligonucleotide Synthesis: A Practical ApproachIrl Press19840000 [\[0419\]](#) [\[0590\]](#)
- **D. M. J. LILLEYJ. E. DAHLBERG**Methods of Enzymology: DNA Structure Part A: Synthesis and Physical Analysis of DNA Methods in EnzymologyAcademic Press19920000 [\[0419\]](#)
- **D. KNORR**Food Biotechnology19870000538-39 [\[0481\]](#) [\[0486\]](#)
- **C. PEDERSON**Microbiology of Fermented Foods19790000210-34 [\[0481\]](#)
- **D. KNORR**Food Biotechnology19870000540- [\[0482\]](#)
- **C. PEDERSON**Microbiology of Fermented Foods19790000153-209 [\[0482\]](#)
- **D. KNORR**Food Biotechnology19870000530- [\[0485\]](#) [\[0486\]](#)
- **C. PEDERSON**Microbiology of Fermented Foods19790000135-51 [\[0485\]](#) [\[0486\]](#)
- **AUSUBEL et al.**Current Protocols in Molecular BiologyJohn Wiley & Sons, Inc.19950000 [\[0517\]](#)
- **RAJAGOPAL et al.**J.Dairy Sci., 1990, vol. 73, 894-899 [\[0540\]](#)
- **DEVEREUX et al.**Nucleic Acids ResearchUniversity of Wisconsin, U.S.A.19840000vol. 12, 387- [\[0579\]](#)
- **AUSUBEL et al.**NUCLEIC ACIDS RESEARCH19990000 [\[0579\]](#)
- **ATSCHUL et al.**J. Mol. Biol., 1990, 403-410 [\[0579\]](#)
- **AUSUBEL et al.**J. MOL. BIOL., 1999, 7-58-7-60 [\[0579\]](#)
- FEMS Microbiol Lett, 1999, vol. 174, 2247-50 [\[0579\]](#)
- FEMS Microbiol Lett, 1999, vol. 177, 1187-8 [\[0579\]](#)

- **SIMON RJ et al.**PNAS, 1992, vol. 89, 209367-9371 [\[0588\]](#)
- **HORWELL DC**Trends Biotechnol., 1995, vol. 13, 4132-134 [\[0588\]](#)
- **Methods of Enzymology: DNA Structure Part A: Synthesis and Physical Analysis of DNAD. M. J. LILLEYJ. E. DAHLBERG**Methods in EnzymologyAcademic Press19920000 [\[0590\]](#)
- **RUSSELLKLAENHAMMER**Applied and Environmental Microbiology, 2001, vol. 67, 43691-4364 [\[0636\]](#) [\[0640\]](#)
- **LEVESQUE et al.**Applied and Environmental Microbiology, 2005, vol. 71, 4057-4068 [\[0636\]](#) [\[0637\]](#) [\[0638\]](#)
- **DUPLESSIS et al.**Virology, 2005, vol. 340, 192-208 [\[0637\]](#) [\[0638\]](#)
- **BARRANGOU et al.**Applied and Environmental Microbiology, 2002, vol. 68, 2877-2884 [\[0639\]](#)
- **BOLOTIN AQUINQUIS BSOROKIN AEHRlich SD**Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal originMicrobiology, 2005, vol. 151, 82551-61 [\[0673\]](#)
- **GROENEN PMBUNSCHOTEN AEVAN SOOLINGEN DJD VAN EMBDEN**Nature of DNA polymorphism in the direct repeat cluster of Mycobacterium tuberculosis; application for strain differentiation by a novel typing methodMolecular Microbiology, 1993, vol. 10, 1057-1065 [\[0673\]](#)
- **HOE NNAKASHIMA KGRIGSBY DPAN XDOU SJNAIDICH SGARCIA MKAHN EBERGMIRE-SEAT DJM MUSSER**Rapid molecular genetic subtyping of serotype M1 group A Streptococcus strainsEmerging Infectious Diseases, 1999, vol. 5, 254-263 [\[0673\]](#)
- **JANSEN RVAN EMBDEN JDAGAASTRA WLM SCHOULS**Identification of a novel family of sequence repeats among prokaryotesOMICS, 2002, vol. 6, 23-33 [\[0673\]](#)
- **JANSEN RVAN EMBDEN JDAGAASTRA WLM SCHOULS**Identification of genes that are associated with DNA repeats in prokaryotesMolecular Microbiology, 2002, vol. 43, 1565-1575 [\[0673\]](#)
- **KAMERBEEK JSCHOULS LKOLK AVAN AGTERVELD MVAN SOOLINGEN DKUIJPER SBUNSCHOTEN AMOLHUIZEN HSHAW RGOYAL M**Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiologyJournal of Clinical Microbiology, 1997, vol. 35, 907-914 [\[0673\]](#)
- **MOJICA FJMDIEZ-VILLASENOR CSORIA EG JUEZ**Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondriaMolecular Microbiology, 2000, vol. 36, 244-246 [\[0673\]](#)
- **MOJICA FJMDIEZ-VILLASENOR CGARCIA-MARTINEZ JE SORIA**Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elementsJournal of Molecular Evolution, 2005, vol. 60, 174-182 [\[0673\]](#)
- **POURCEL CSAVIGNOL GG VERGNAUD**CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA and provide additional tools for evolutionary studiesMicrobiology, 2005, vol. 151, 653-663 [\[0673\]](#)
- **SAUNDERS NFWGOODCHILD ARAFTERY MGUILHAUS MCURMI PMGR CAVICCHIOLI**Predicted roles for hypothetical proteins in the low-temperature expressed proteome of the antarctic archaeon Methanococcoides burtoniiJournal of Proteome

Research, 2005, vol. 4, 464-472 [0673]

- **MONGODIN EFHANCE IRDEBOY RTGILL SRDAUGHERTY SHUBER RFRASER CMSTETTER KKE NELSON**Gene transfer and genome plasticity in *Thermotoga maritima*, a model hyperthermophilic species *Journal of Bacteriology*, 2005, vol. 187, 4935-4944 [0673]
- **PENG XBRUGGER KSHEN LSHE QRA GARRETT**Genus-specific protein binding to the large clusters of DNA repeats (Short Regularly Spaced Repeats) present in *Sulfolobus* genomes *Journal of Bacteriology*, 2003, vol. 185, 2410-2417 [0673]

Patentkrav

1. Fremgangsmåde til bibringelse af resistens til en bakteriecelle, som omfatter et eller flere cas-gener eller -proteiner og to eller flere CRISPR-gentagelsesenheder, over for en målnukleinsyre eller et transskriptionsprodukt deraf; hvor nævnte ene eller flere cas-gener eller -proteiner og to eller flere CRISPR-gentagelsesenheder udgør en funktionel kombination;
- 10 hvor nævnte mindst ene eller flere cas-gener eller -proteiner og nævnte to eller flere CRISPR-gentagelsesenheder endvidere er fra samme CRISPR-locus i et genom eller plasmid; hvilken fremgangsmåde omfatter følgende trin:
- (i) identificering af en eller flere CRISPR-spacere i en organisme, som er resistent over for målnukleinsyren eller transskriptionsproduktet deraf;
- 15 (ii) insertion af sekvensen for CRISPR-spaceren eller -spacerne i bakteriecellen; således at cellen bliver i alt væsentligt resistent over for målnukleinsyren eller transskriptionsproduktet deraf; og
- 20 (iii) bestemmelse af, om cellen modulerer resistens over for målnukleinsyren eller transskriptionsproduktet deraf, hvor modulering af cellens resistens over for målnukleinsyren eller transskriptionsproduktet deraf indikerer, at CRISPR-spaceren
- 25 modulerer cellens resistens.
2. Fremgangsmåde til bibringelse af resistens til en bakteriecelle over for en målnukleinsyre eller et transskriptionsprodukt deraf, hvilken fremgangsmåde omfatter følgende trin:
- 30 (i) identificering af en eller flere CRISPR-spacere i en organisme, som er resistent over for målnukleinsyren eller transskriptionsproduktet deraf;
- (ii) fremstilling af en rekombinant nukleinsyre omfattende mindst ét cas-gen og mindst to CRISPR-gentagelsesenheder sammen med den eller de identificerede CRISPR-spacere;
- 35 hvor det mindst ene cas-gen og de mindst to eller flere CRISPR-gentagelsesenheder udgør en funktionel kombination;

hvor det mindst ene cas-gen og de mindst to eller flere CRISPR-gentagelsesenheder endvidere er fra samme CRISPR-locus i et genom eller plasmid;

- 5 (iii) indføring af den rekombinante nukleinsyre i cellen, således at cellen bliver i alt væsentligt resistent over for målnukleinsyren eller transskriptionsproduktet deraf; og
- (iv) bestemmelse af, om cellen modulerer resistens over for målnukleinsyren eller transskriptionsproduktet deraf, hvor modulering af cellens resistens over for målnukleinsyren eller
- 10 transskriptionsproduktet deraf indikerer, at CRISPR-spaceren modulerer cellens resistens.

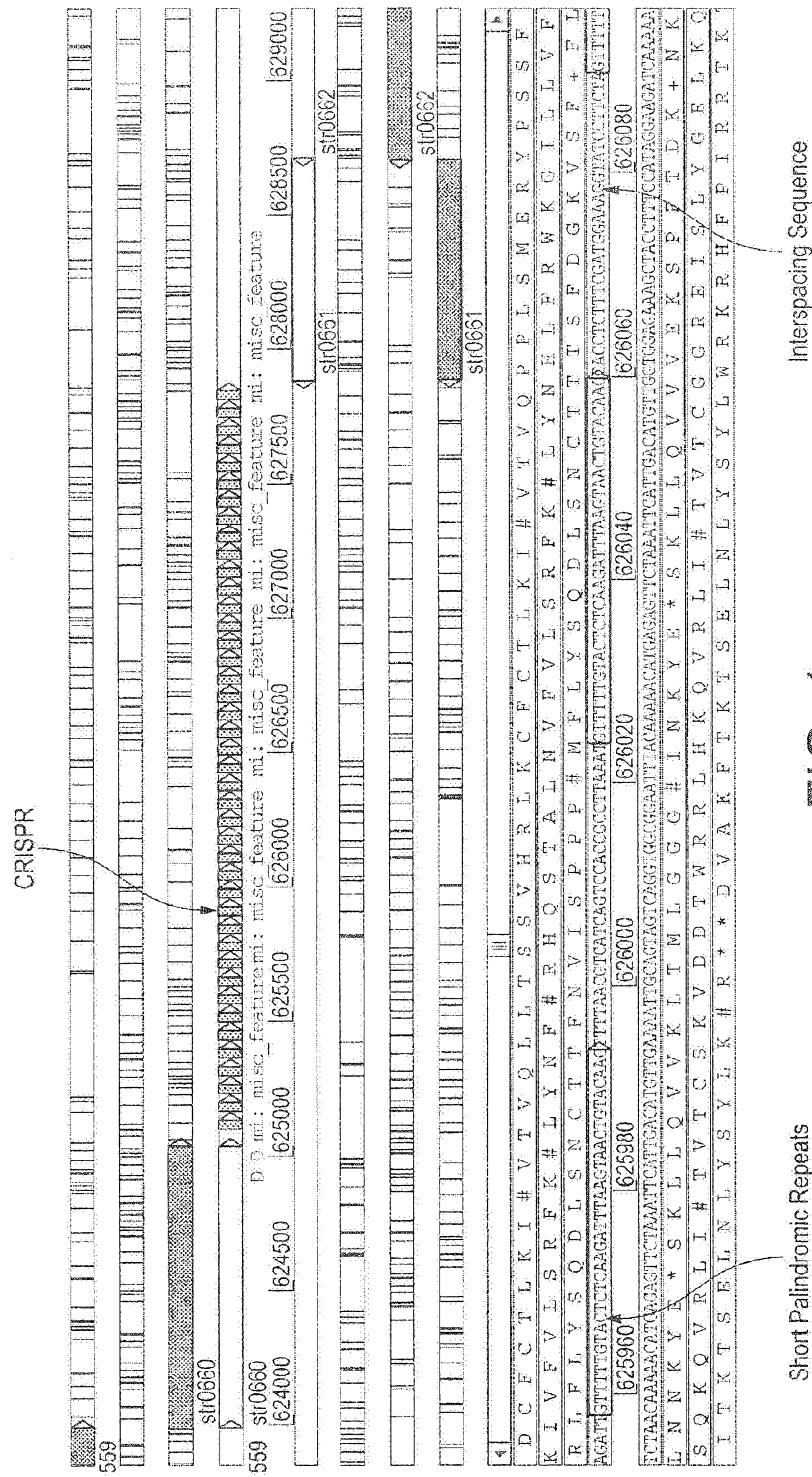
3. Fremgangsmåde ifølge et hvilket som helst af de foregående krav, hvor bakteriecellen er en mælkesyrebakterie
- 15 valgt fra gruppen bestående af en Bifidobacterium-art, en Brevibacterium-art, en Propionibacterium-art, en Lactococcus-art, en Streptococcus-art, en Lactobacillus-art, en Enterococcus-art, en Pediococcus-art, en Leuconostoc-art og en Oenococcus-art.

20

4. Fremgangsmåde ifølge et hvilket som helst af de foregående krav, hvor bakteriecellen er Streptococcus thermophilus.

- 25 5. Fremgangsmåde ifølge et hvilket som helst af de foregående krav, hvor målnukleinsyren eller transskriptionsproduktet deraf er eller er afledt af bakteriofag-DNA.

DRAWINGS



7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530

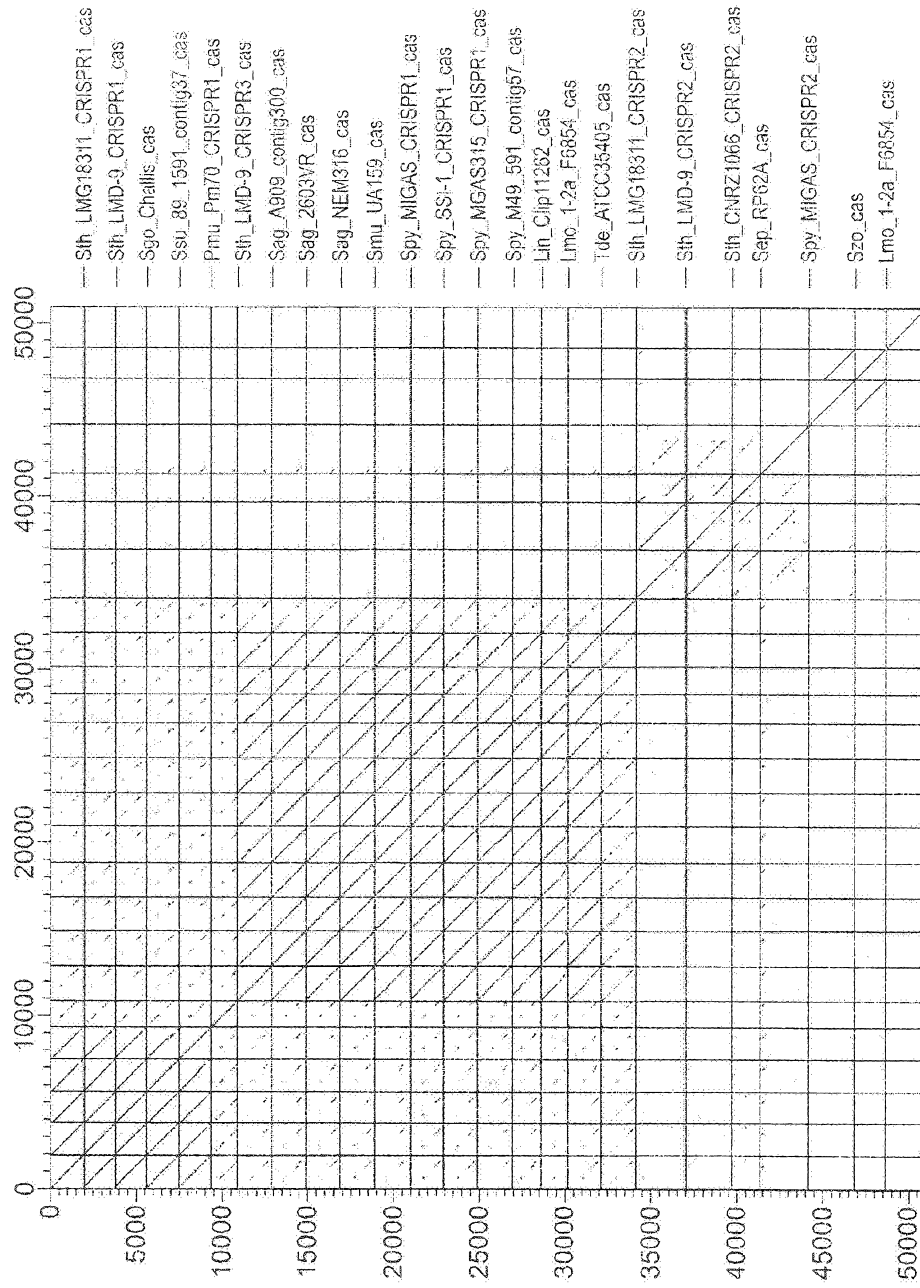


FIG. 2A

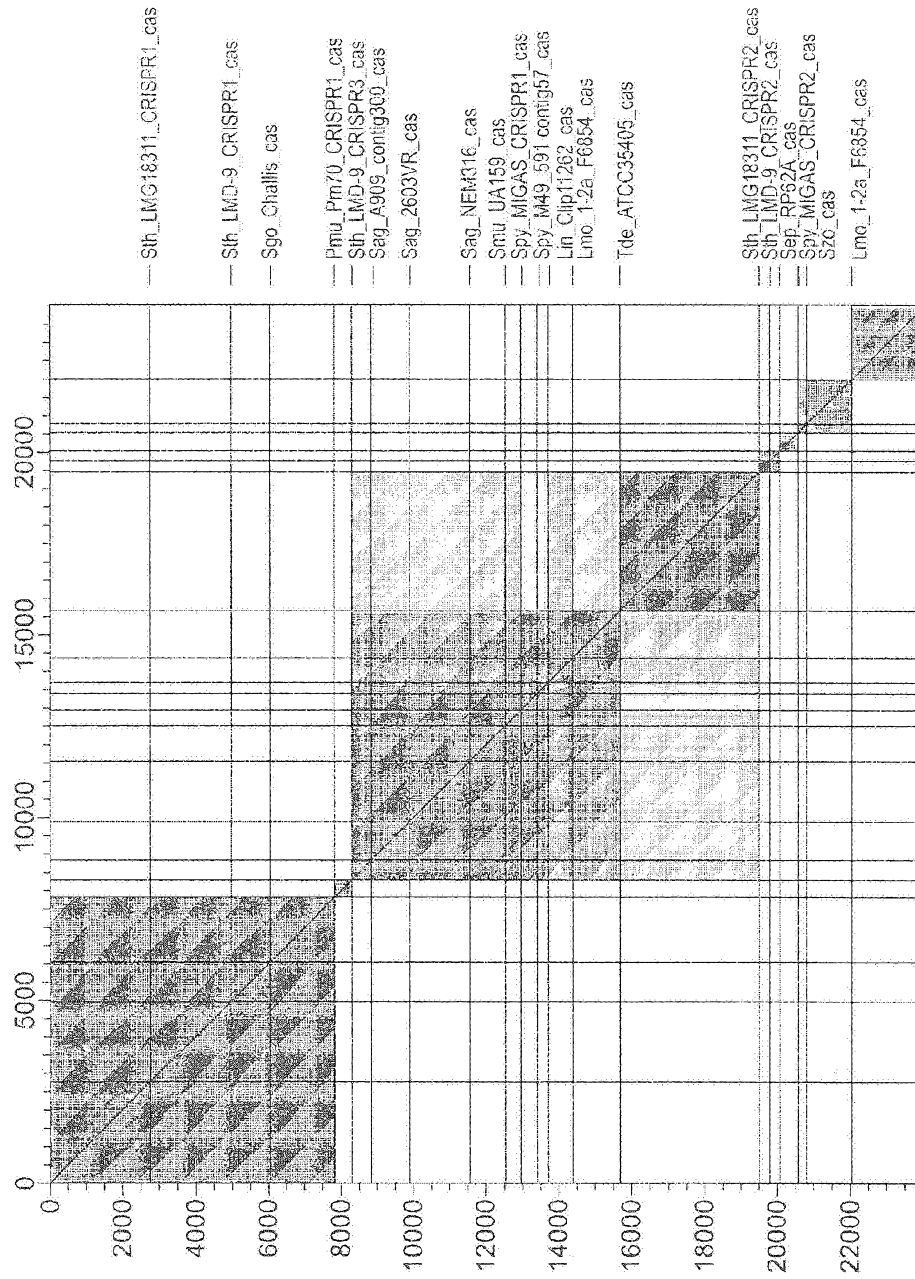


FIG. 2B

interspersing sequence number	Phages															Number of matches	
	Sfi 19	Sfi 21	Sfi 11	DT1	MD2	Sfi 18	Sfi 21	St3	O1205	Di2	2972	Di4	MD1	Q5	Kappa3		7201
1																	3
3																	3
5 and 31																	3
6																	2
8																	1
11																	2
13																	6
17																	2
20																	4
22																	7
23																	2
25																	2
27																	1
28																	2
29																	8
32																	2
34																	2
36																	1
37																	1
41																	2

FIG. 3A

☐ >gi|5523999|gb|AF115102.1|AF115102  S. thermophilus phage Sfi19
 Query: 842 ttctggtagtggatttagtcaaacagatgt 871
 |||||
 Sbjct: 18912 ttctggtagtggatttagtcaaacagatgt 18941

☐ >gi|5524032|gb|AF115103.1|AF115103  S. thermophilus phage Sfi21
 Query: 842 ttctggtagtggatttagtcaaacagatgt 871
 |||||
 Sbjct: 17084 ttctggtagtggatttagtcaaacagatgt 17113

☐ >gi|7669462|gb|AF158600.2|AF158600  S. thermophilus phage Sfi11
 Query: 842 ttctggtagtggatttagtcaaacagatgt 871
 |||||
 Sbjct: 21396 ttctggtagtggatttagtcaaacagatgt 21425

☐ >gi|61383211|gb|AF348736.2|  S. thermophilus phage MD2
 Query: 842 ttctggtagtggatttagtcaaacagatgt 871
 |||||
 Sbjct: 9043 ttctggtagtggatttagtcaaacagatgt 9072

☐ >gi|2444080|gb|U88974.1|  S. thermophilus temperate phage O1205
 Query: 842 ttctggtagtggatttagtcaaacagatgt 871
 |||||
 Sbjct: 34602 ttctggtagtggatttagtcaaacagatgt 34631

☐ >gi|15077558|gb|AF348739.1| S. thermophilus phage DT2
 Query: 843 tctggtagtggatttagtcaaac 865
 |||||
 Sbjct: 2575 tctggtagtggatttagtcaaac 2597

☐ >gi|56718416|gb|AY699705.1|  S. thermophilus phage 2972
 Query: 846 ggtagtggatttagtcaaacagatgt 871
 |||||
 Sbjct: 20033 ggtagtggatttagtcaaacagatgt 20058

☐ >gi|7248462|gb|AF145054.1|AF145054  S. thermophilus phage 7201
 Query: 842 ttctggtagtggatttagtcaaacagat 869
 |||||
 Sbjct: 29631 ttctggcagtggatttagtcaaacagat 29658

FIG. 3B

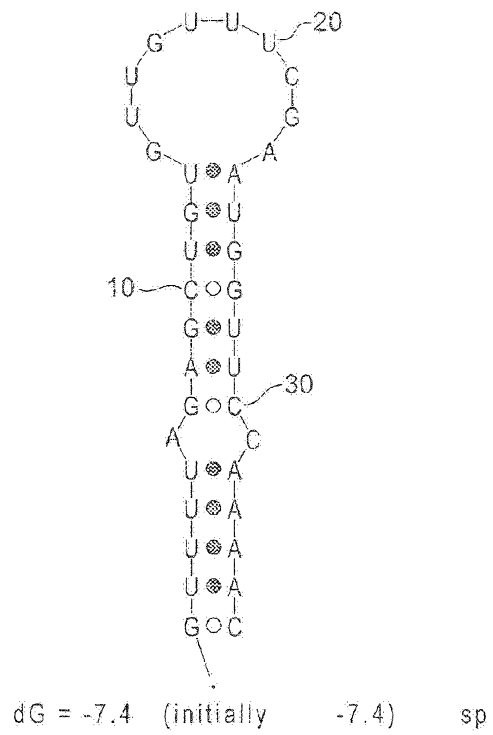


FIG. 4

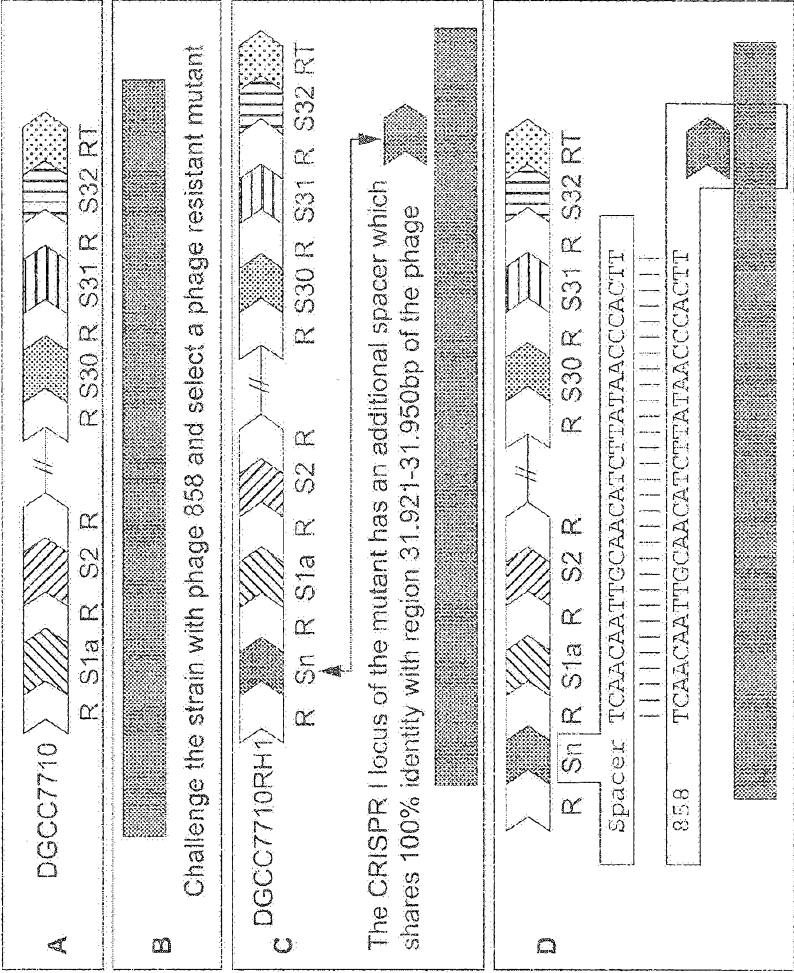


FIG. 5

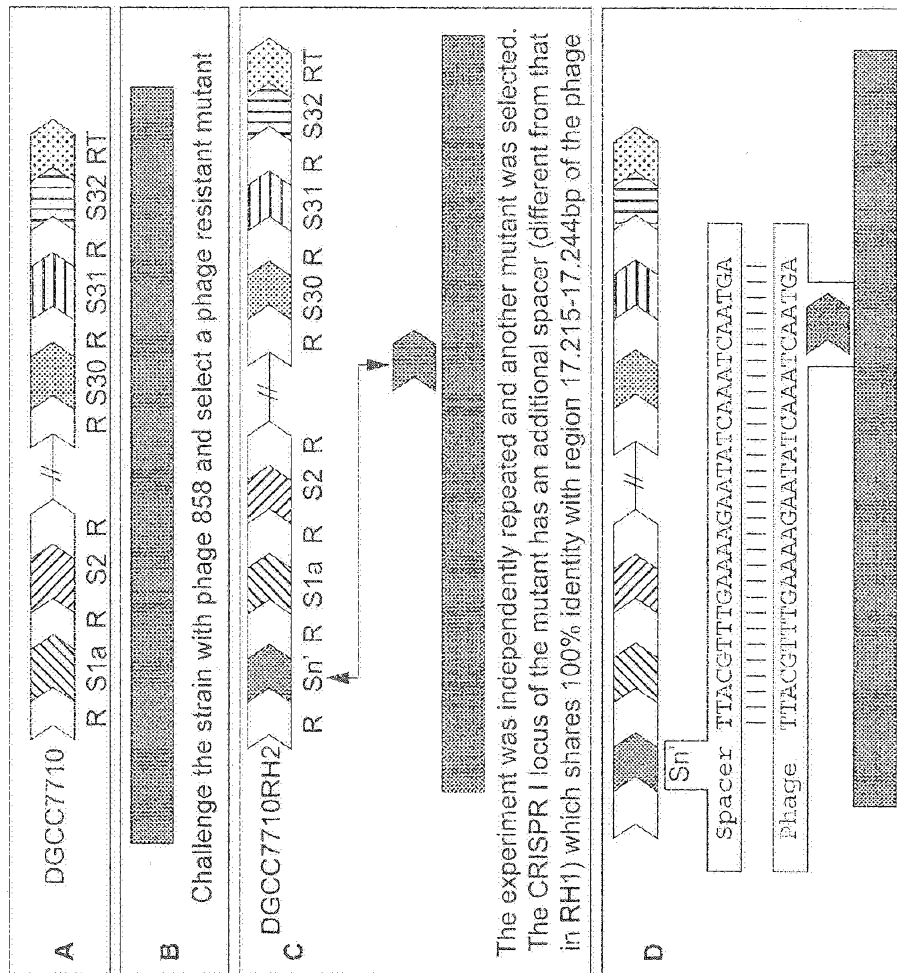


FIG. 6

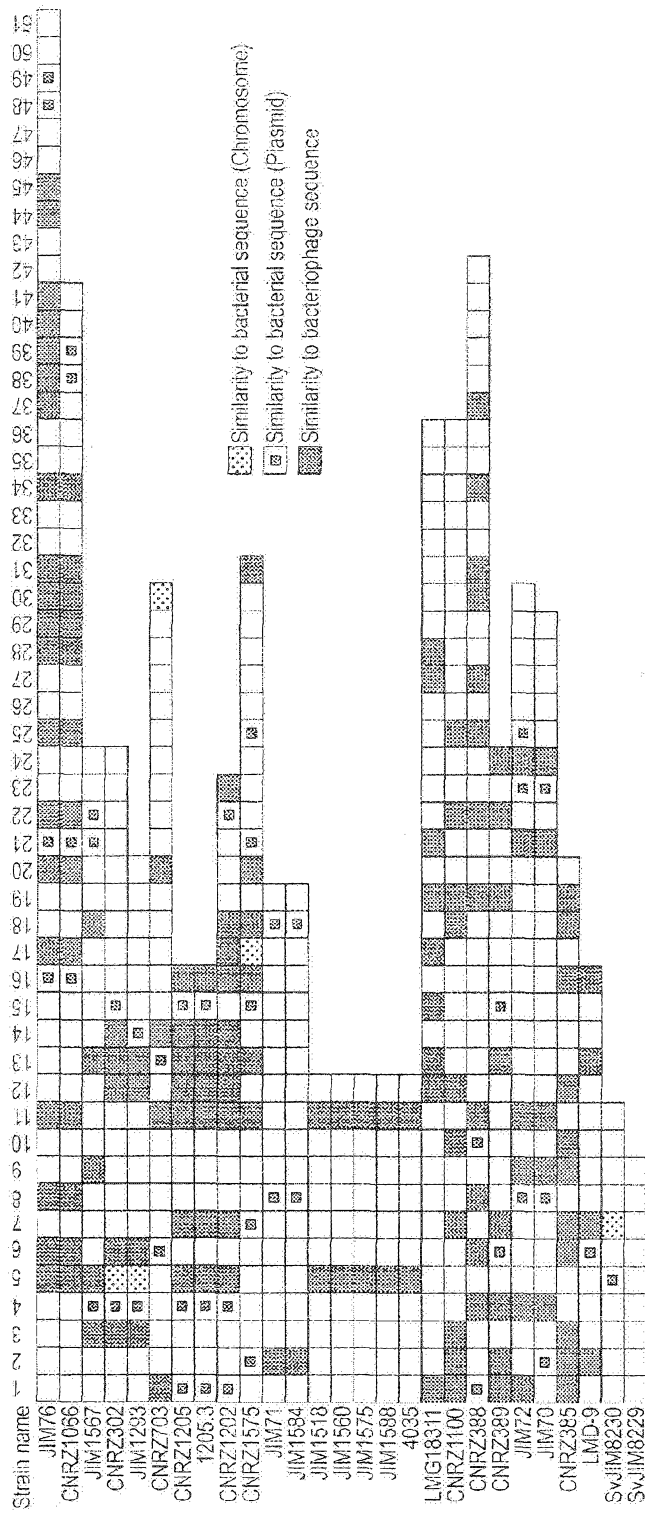


FIG. 8

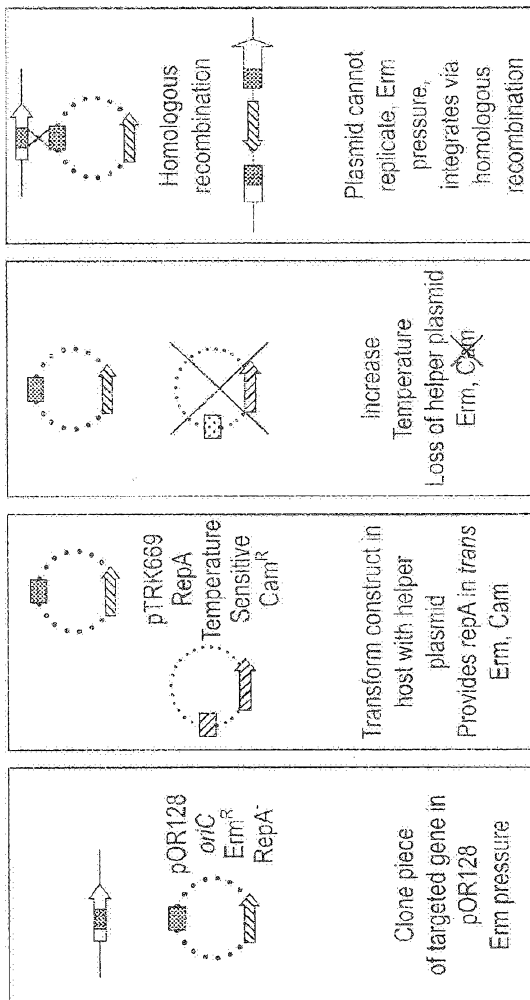


FIG. 9

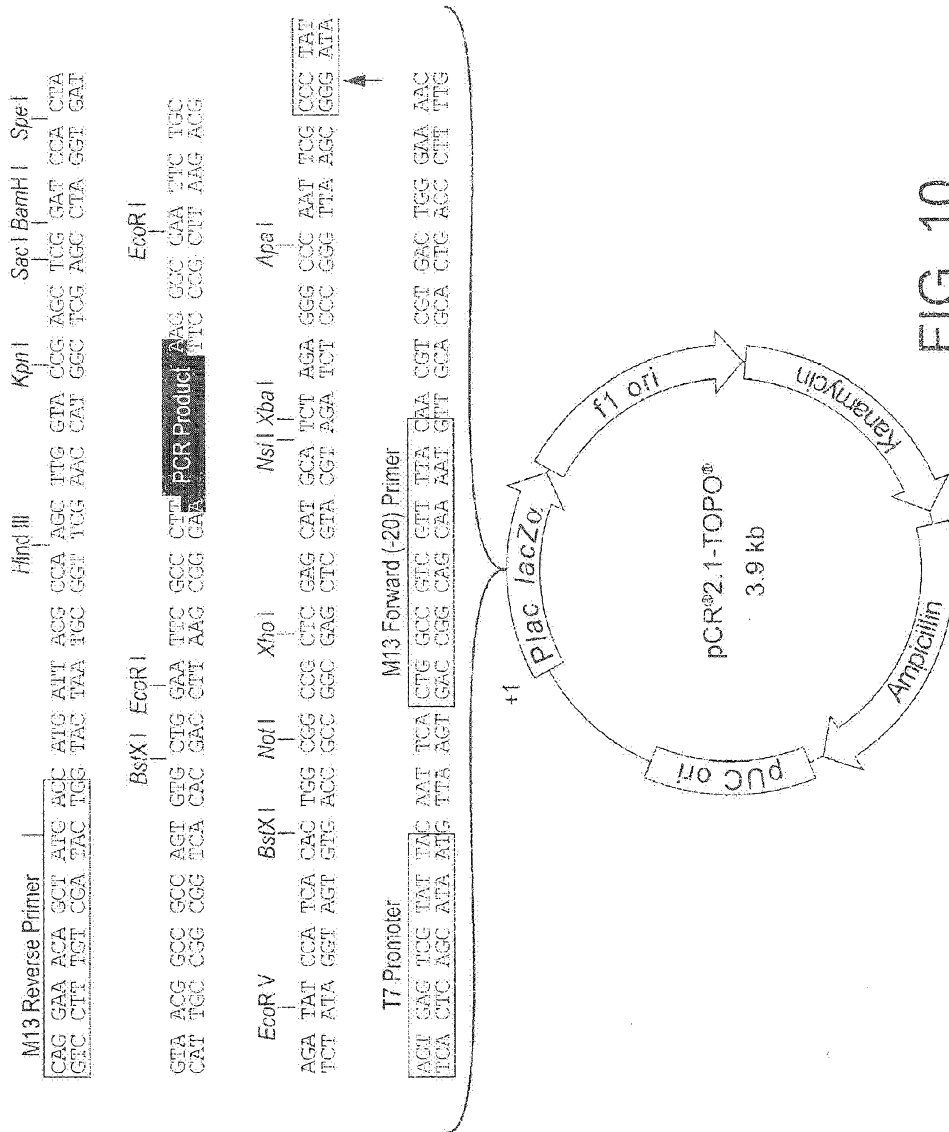


FIG. 10

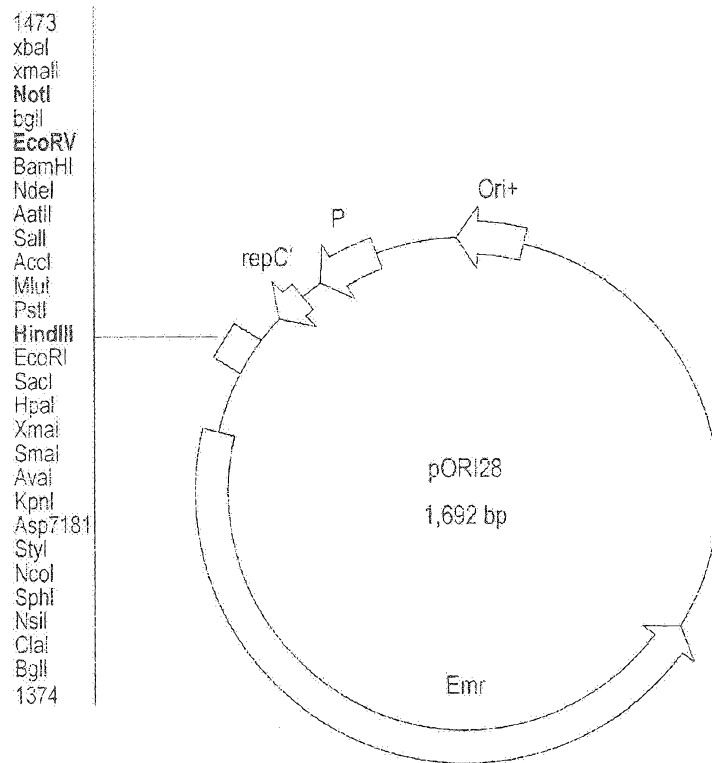


FIG. 11

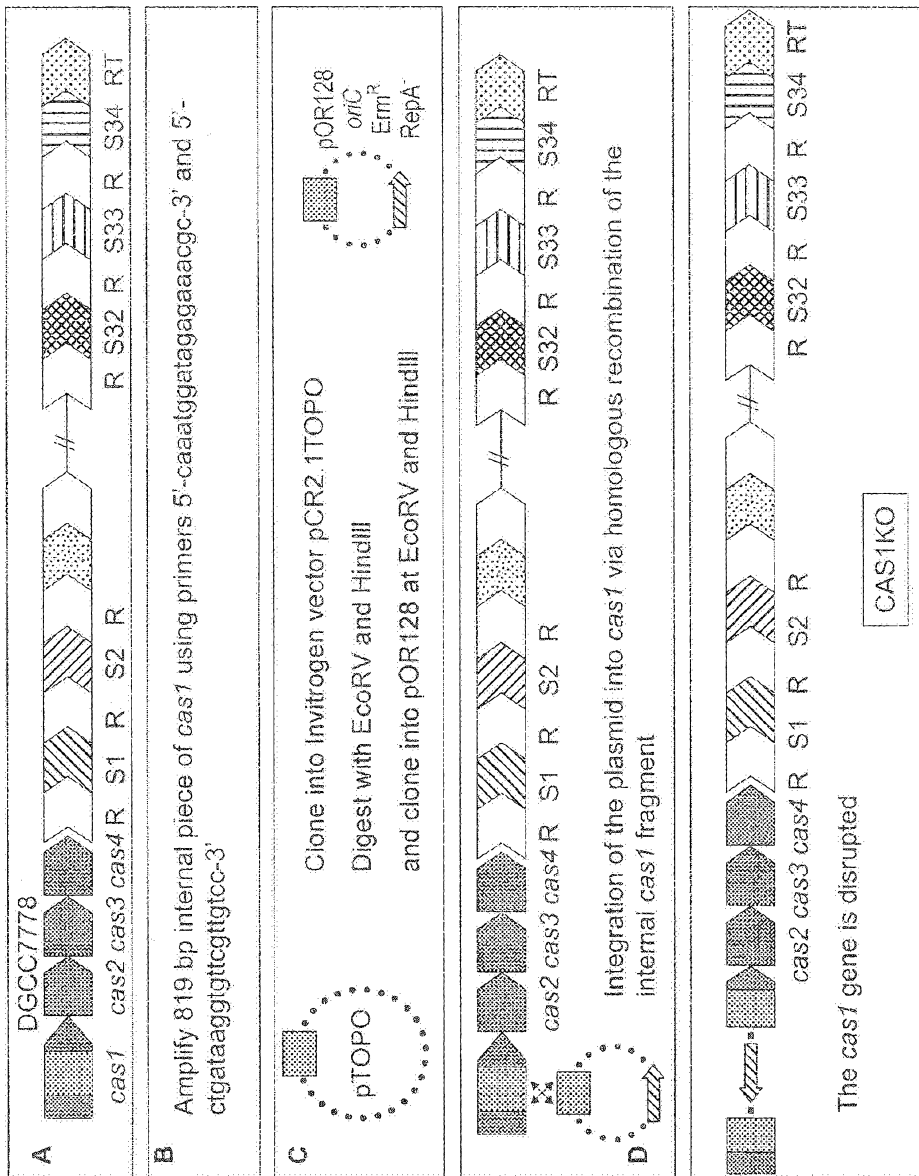


FIG. 12

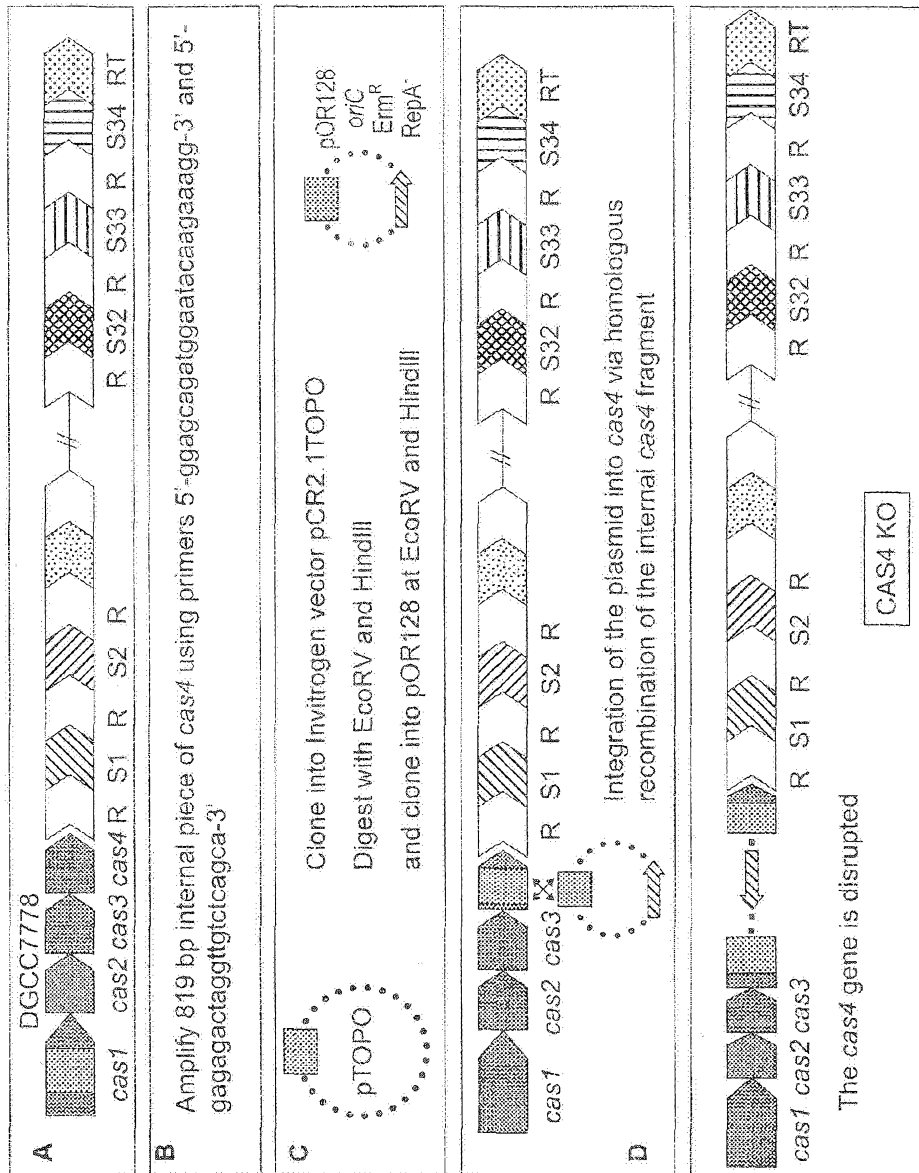


FIG. 13

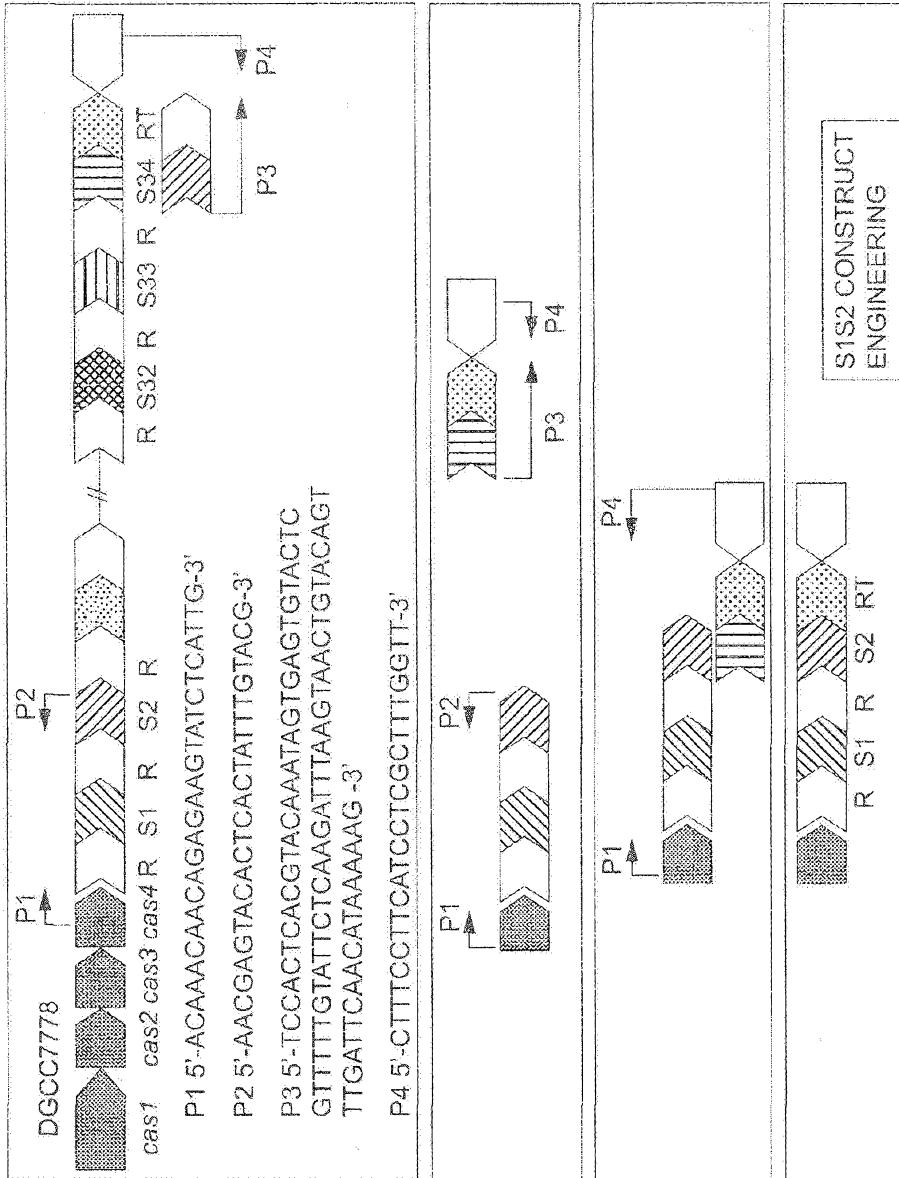


FIG. 14

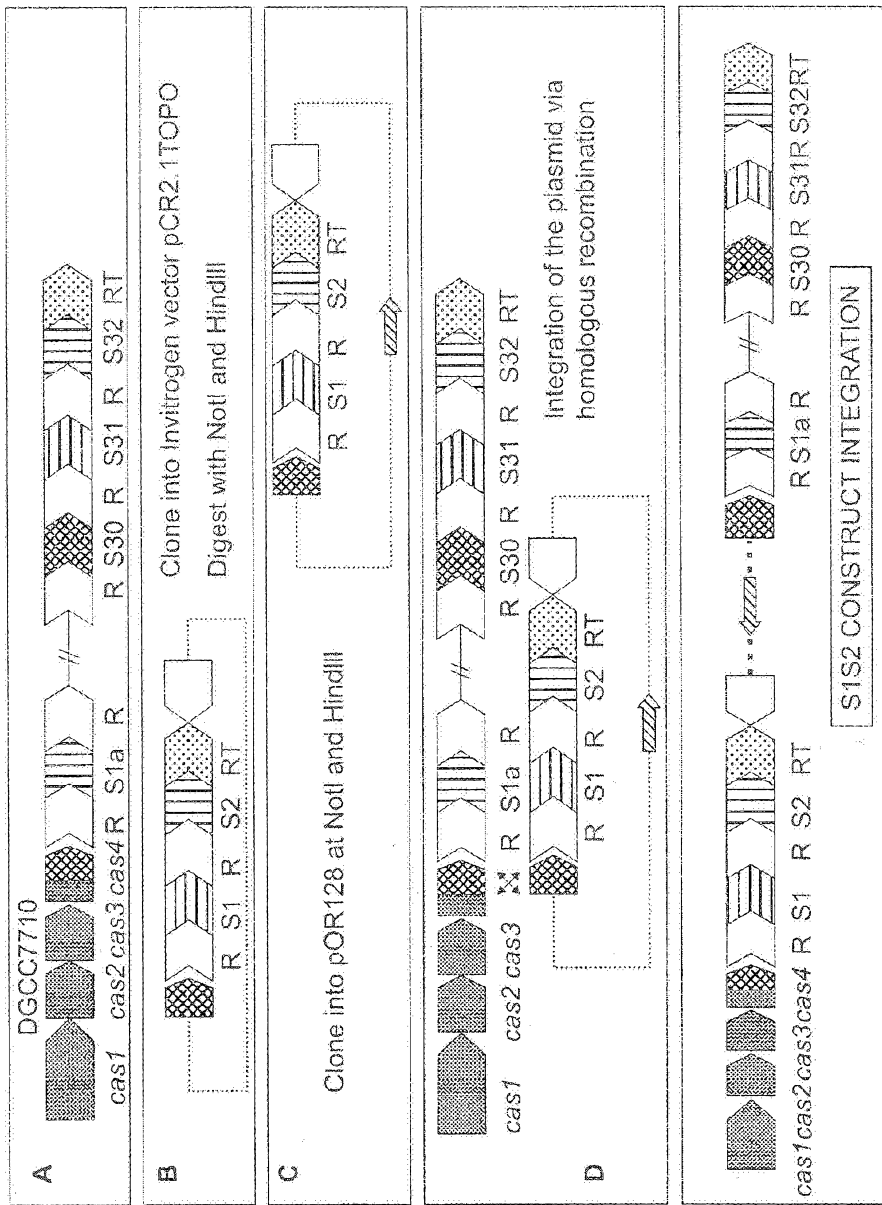


FIG. 16

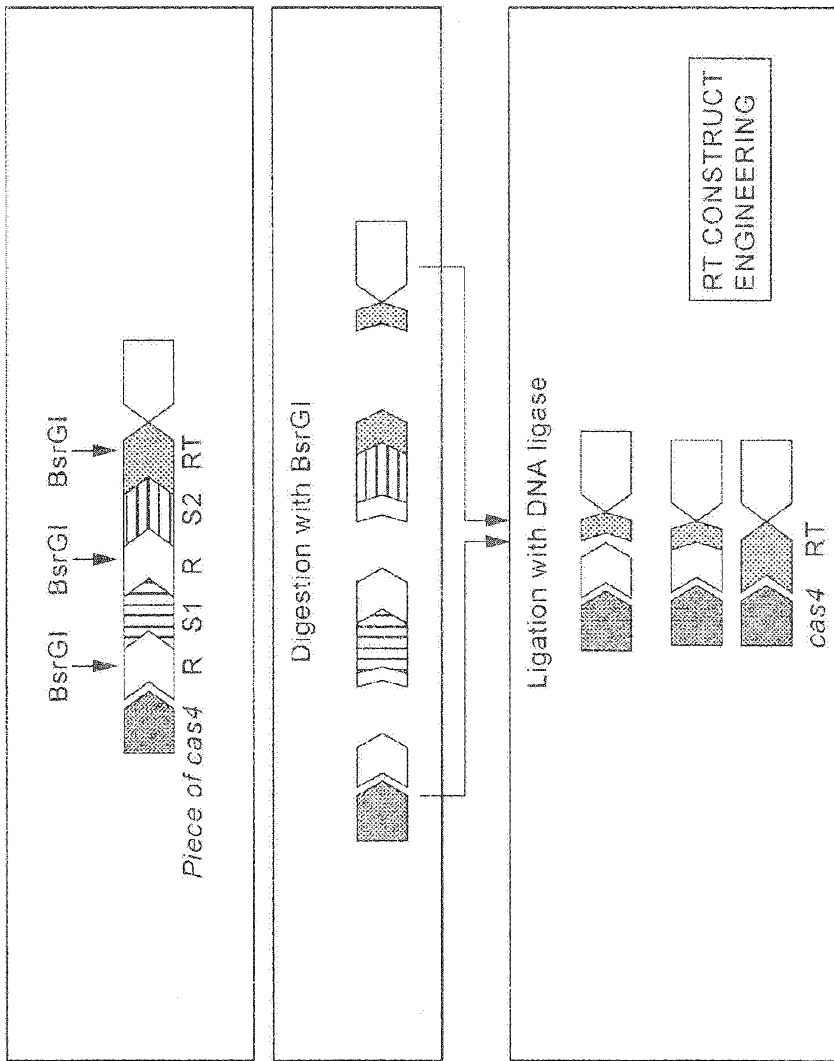


FIG. 17

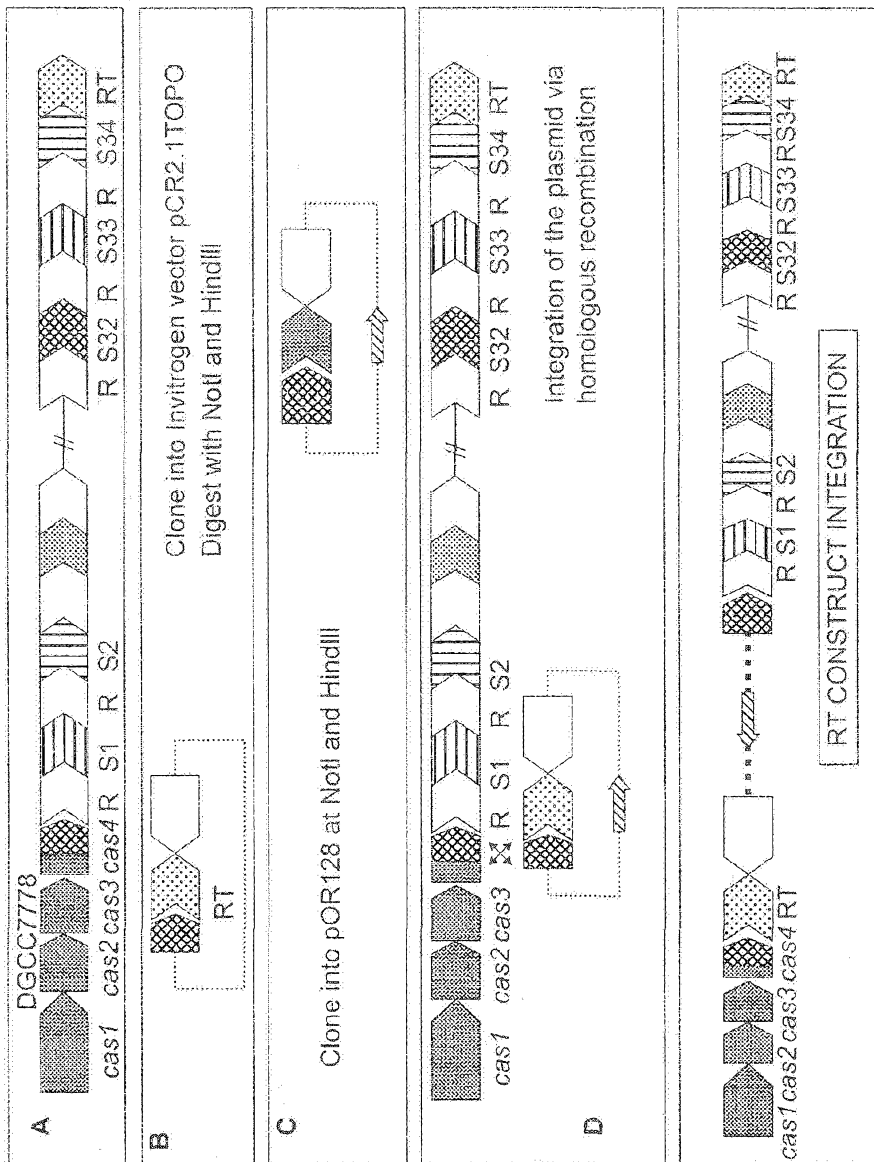


FIG. 18

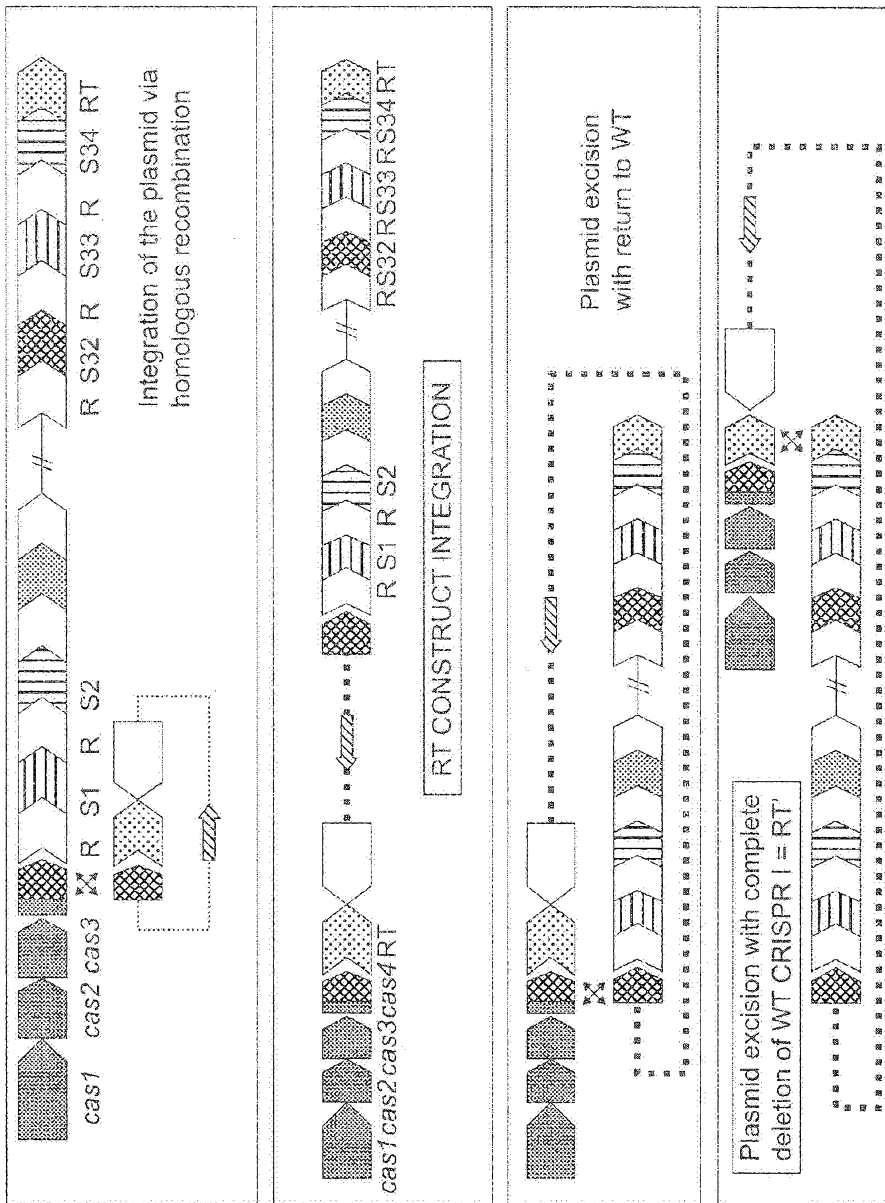


FIG. 19

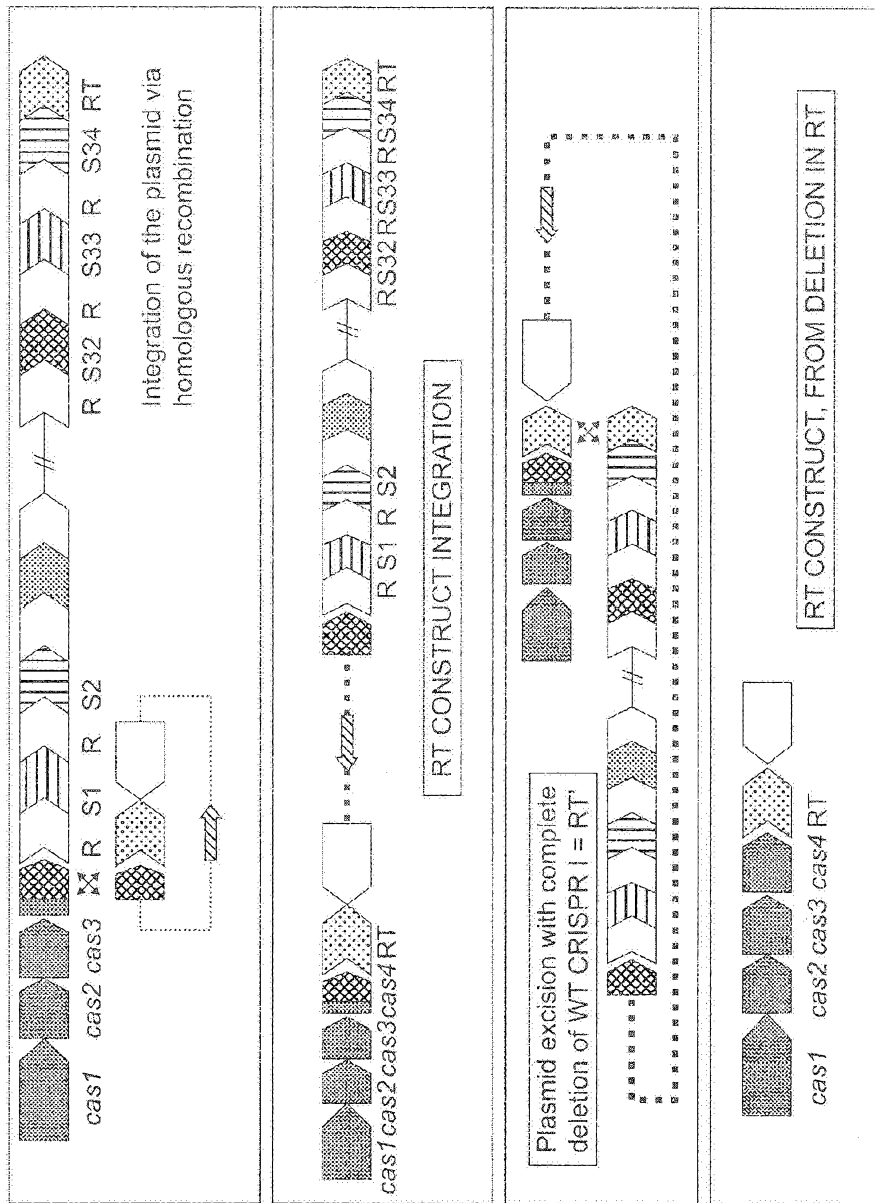


FIG. 20