



(22) Date de dépôt/Filing Date: 2006/08/25

(41) Mise à la disp. pub./Open to Public Insp.: 2007/03/01

(45) Date de délivrance/Issue Date: 2022/08/02

(62) Demande originale/Original Application: 2 619 833

(30) Priorités/Priorities: 2005/08/26 (US60/711,396);  
2006/05/19 (US60/747,683)

(51) Cl.Int./Int.Cl. *C12N 15/63* (2006.01),  
*A23C 9/123* (2006.01), *A23K 10/10* (2016.01),  
*A23L 33/135* (2016.01), *C12N 15/00* (2006.01),  
*C12N 15/113* (2010.01), *C12N 15/31* (2006.01),  
*C12N 15/55* (2006.01), *C12N 15/74* (2006.01),  
*C12N 9/14* (2006.01), *C12N 9/24* (2006.01)

(72) Inventeurs/Inventors:  
BARRANGOU, RODOLPHE, US;  
FREMAUX, CHRISTOPHE, FR;  
BOYVAL, PATRICK, FR;  
ROMERO, DENNIS, US;  
HORVATH, PHILIPPE, FR

(73) Propriétaire/Owner:

(54) Titre : UTILISATION DE GENES ASSOCIES A CRISPR-CAS

(54) Title: USE OF CRISPR ASSOCIATED GENES (CAS)

(57) Abrégé/Abstract:

The present invention relates to the use of one or more cas genes for modulating resistance in a cell against a target nucleic acid or a transcription product thereof.



(73) **Propriétaires(suite)/Owners(continued):**DUPONT NUTRITION BIOSCIENCES APS, DK

(74) **Agent:** BERESKIN & PARR LLP/S.E.N.C.R.L.,S.R.L.

**ABSTRACT**

The present invention relates to the use of one or more cas genes for modulating resistance in a cell against a target nucleic acid or a transcription product thereof.

## USE OF CRISPR ASSOCIATED GENES (CAS)

### FIELD OF INVENTION

The present invention relates to *inter alia* modulating the resistance of a cell  
5 against a target nucleic acid or a transcription product thereof. In particular, the  
present invention relates, in one aspect, to the use of one or more *cas* genes or proteins  
for modulating the resistance of a cell against a target nucleic acid or a transcription  
product thereof.

### 10 BACKGROUND TO THE INVENTION

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  
15 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.

20 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).

(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  
25 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.

(ii) The rotation of selected bacterial strains which are sensitive to different  
30 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.

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

      Pedersen *et al* (7<sup>th</sup> 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  
10        which requires thymidine for DNA replication.

      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.

15        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.

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

20

#### SUMMARY OF THE INVENTION

      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.

25        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  
30        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.

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).  
5 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  
10 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.

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  
15 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.

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  
20 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.

25 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.

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  
30 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.

5           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.

## 10    SUMMARY ASPECTS OF THE PRESENT INVENTION

          In one aspect there is provided 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.

          In a second aspect there is provided the use of a recombinant nucleic acid  
15   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.

          In a third aspect there is provided a nucleic acid sequence consisting essentially  
20   of at least one *cas* gene.

          In a fourth aspect there is provided a nucleic acid sequence consisting essentially of at least one *cas* gene and at least two CRISPR repeats.

          In a fifth aspect there is provided a nucleic acid sequence consisting essentially of at least one *cas* gene and at least one CRISPR spacer.

25           In a sixth aspect there is provided a nucleic acid sequence consisting essentially of at least one *cas* gene, at least one CRISPR spacer and at least two CRISPR repeats.

          In a seventh aspect there is provided 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*  
30   gene and/or the at least two CRISPR repeats.

          In an eighth aspect there is provided a construct comprising one or more of the nucleic acid sequences described herein

In a ninth aspect there is provided a vector comprising one or more of the nucleic acid sequences or one or more of the constructs described herein.

In a tenth aspect there is provided a cell comprising the nucleic acid sequence or the construct or the vector described herein.

5 In an eleventh aspect there is provided 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  
10 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.

In a twelfth aspect there is provided a method for modulating (eg. conferring or  
15 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  
20 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.

In a thirteenth aspect there is provided a method for modulating (eg. conferring or increasing) the resistance of a cell comprising at least one or more *cas* genes or  
25 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  
30 organism.

In a fourteenth aspect there is provided 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  
5 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.

In a fifteenth aspect there is provided 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  
10 transcription product thereof comprising modifying the one or more *cas* genes or proteins and/or two or more CRISPR repeats in the cell.

In a sixteenth aspect there is provided 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)  
15 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)  
20 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.

In a seventeenth aspect there is provided a method for identifying a *cas* gene  
25 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  
30 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.

In an eighteenth aspect there is provided 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.

In a nineteenth aspect there is provided 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.

In a twentieth aspect there is provided 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.

In a twenty-first aspect there is provided 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.

In a twenty-second aspect there is provided 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.

In a twenty-third aspect there is provided 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.

In a twenty-fourth aspect there is provided 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.

In a twenty-fifth aspect there is provided a cell obtained or obtainable by the method(s) described herein.

In a twenty-sixth aspect there is provided a CRISPR spacer or pseudo CRISPR spacer obtained or obtainable by the method(s) described herein.

In a twenty-seventh aspect there is provided a *cas* gene obtained or obtainable by the method(s) described herein.

In a twenty-eighth aspect there is provided a CRISPR repeat obtained or obtainable by the method(s) described herein.

In a twenty-ninth aspect there is provided a functional combination obtained or obtainable by the method(s) described herein.

In a thirtieth aspect there is provided 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.

In a thirty-first aspect there is provided 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.

In a thirty-second aspect there is provided 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.

In a thirty-third aspect there is provided a food product or feed comprising the culture described herein.

In a thirty-fourth aspect there is provided a process for preparing a food product or feed comprising the use of the culture described herein.

In a thirty-fifth aspect there is provided a food product or feed obtained or obtainable by the process described herein.

In a thirty-sixth aspect there is provided the use of the culture described herein for preparing a food product.

In a thirty-seventh aspect there is provided 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.

In a thirty-eighth aspect there is provided an amino acid sequence encoded by the nucleotide sequence described herein.

In a thirty-ninth aspect there is provided a construct or vector comprising one or more of the nucleotide sequences described herein.

In a fortieth aspect there is provided a host cell into which has been incorporated one or more of the nucleotide sequences described herein or the construct or vector described herein.

### PREFERRED EMBODIMENTS

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

5 In some embodiments, 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.

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.

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

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).

15 In some embodiments, the spacer is obtained from a cell which is resistant to a target nucleic acid.

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

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

20 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.

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.

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

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

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

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.

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

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.

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

In some embodiments, the one or more *cas* genes and the two or more CRISPR repeats are or are derivable (preferably, derived) from the same cell.

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

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).

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

In some embodiments, 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.

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.

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.

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

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

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

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

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

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

5 In some embodiments, the at least one or more CRISPR spacer(s) in the cell are deleted.

In some embodiments, at least one or more CRISPR repeat(s) in the cell are deleted.

In some embodiments, one or more *cas* genes are deleted,

10 In some embodiments, CRISPR and/or one or more *cas* genes are deleted.

In some embodiments, the one or more *cas* genes or proteins and/or two or more CRISPR repeats in the cell are deleted.

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

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

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

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

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

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

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

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.

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  
5 comparing the sequence of the one or more CRISPR spacers in the cell with the sequence of the target nucleic acid .

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  
10 determining the resistance profile of the cell.

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

## FIGURES

### *Figure 1*

15 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  
20 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*  
25 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*

30 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. The new spacer that is integrated into the CRISPR1 locus of *Streptococcus thermophilus* strain DGCC7710-RH1 genome between positions 31921 and 31950 bp on the plus strand confers to this strain resistance to phage D858 is shown in step (d), with the spacer given as SEQ ID NO: 534 and the D858 sequence given as SEQ ID NO:669.

#### 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. The new spacer that is integrated into the CRISPR1 locus of *Streptococcus thermophilus* strain DGCC7710-RH2 genome between positions 17215 and 17244 bp on the plus strand confers to *Streptococcus thermophilus* strain DGCC7710-RH2 a resistance to phage D858 is shown in step (d), with the spacer given as SEQ ID NO:535 and the D858 sequence given as SEQ ID NO:671.

#### 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 (indicated by boxes with an “x”), plasmid DNA sequences (black squares) and phage genomic sequences (shaded in black).

5 *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*

5        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*

10       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*

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

*Figure 14*

20       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  
25       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 S1S2 construct.

*Figure 15*

30       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*

5 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 *BglII* 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*

10 A graphical representation of the integration of the RT construct.

*Figure 19*

A graphical representation of the RT' construct.

*Figure 20*

15 A graphical representation of the RT' construct.

## DETAILED DESCRIPTION OF THE INVENTION

### CRISPR LOCUS

CRISPR loci are a distinct class of interspersed short sequence repeats (SSRs)  
 20 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.*  
 25 *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  
 30 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).

The common structural characteristics of CRISPR loci are described in Jansen *et al.* (2002a; "Identification of a novel family of sequence repeats among prokaryotes." OMICS 6:23-33) 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.

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; "Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria." *Molecular Microbiology* 36:244-246).

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*.

Various methods for identifying CRISPR loci are already known in the art. By way of example, Jansen *et al.* (2002; "Identification of genes that are associated with DNA repeats in prokaryotes." *Molecular Microbiology* 43:1565-1575) 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  $p1 = a \dots b \ c \dots d \ p1 \ c \dots d \ p1 \ c \dots d \ p1$ , 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.

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

5        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 and GenBank.

10        The amplification of CRISPR loci has been described in, for example, Mojica *et al.* (2005; "Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria." *Molecular Microbiology* 36:244-246) and Pourcel *et al.* (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). 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  
15        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).

20        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  
25        segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified."

30        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  
 5 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.

Methods for designing PCR primers and PCR cloning are generally known in  
 10 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*  
 15 *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.

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

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

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

25 Nucleotide sequences described herein may be obtained from databases – such as GenBank or the JGI website.

#### CRISPR ORIENTATION

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

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.

5       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 the terminal repeat, and ending just before the stop codon of gene str0661 (located on the opposite DNA strand). The CRISPR trailer is located at the 3' end of the CRISPR locus. The CRISPR trailer is located immediately  
10 downstream of the terminal repeat.

By way of example, the CRISPR leader and CRISPR trailer sequences in the CRISPR1 locus of *Streptococcus thermophilus* strain CNRZ1066 are:

*CRISPR leader*

15       5' -CAAGGACAGTTATTGATTTTATAATCACTATGTGGGTATAAAAACGTCAAATTTTCATTGAG-3'  
(SEQ ID NO: 666)

*CRISPR trailer*

5' -TTGATTCAACATAAAAAGCCAGTTCAATTGAACTTGGCTTT-3' (SEQ ID NO: 667)

20       The CRISPR leader corresponds to positions 625038 to 625100, and the CRISPR trailer corresponds to positions 627845 to 627885 in the full genome (CP000024) of *Streptococcus thermophilus*.

For the avoidance of doubt, "upstream" means in the 5' direction and "downstream" means in the 3' direction.

25

CAS

As used herein, the term "cas gene" has the conventional meaning as used in the art and refers to one or more cas genes that are generally coupled, associated or close to or in the vicinity of flanking CRISPR loci.

30       A comprehensive review of the Cas protein family is presented in Haft *et al.* (2005) *Computational Biology* 1, 6 e60. As described therein, 41 CRISPR-associated (cas) gene families are described, in addition to the four previously known gene

families. It shows that CRISPR systems belong to different classes, with different repeat patterns, sets of genes, and species ranges.

The number of *cas* genes at a given CRISPR locus can vary between species.

5 In one aspect, the present invention relates to the use of one or more *cas* genes or proteins for modulating resistance in a cell (eg. a recipient cell) against a target nucleic acid or a transcription product thereof.

In a further aspect, the present invention relates to the use of one or more *cas* genes or proteins and one more CRISPR spacers for modulating resistance in a cell (eg. a recipient cell) against a target nucleic acid or a transcription product thereof.

10 In some embodiments, one or more of the *cas* genes and/or proteins may naturally 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.

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.

15 In one aspect, the present invention relates to 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.

20 In one aspect, the present invention relates to 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.

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.

Bolotin *et al.* (2005; “Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin.” *Microbiology* 151( 8):2551-61) have recently reported another cluster of three genes associated with CRISPR structures in many bacterial species, named here as *cas1B*, *cas5* and *cas6*.

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

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

The *cas* genes may be *cas1*, *cas2*, *cas3*, *cas4*, *cas1B*, *cas5* and/or *cas6* or a  
10   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*;  
15   *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*  
20   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  
25   combinations thereof.

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  
30   *cas5* or *cas1B* and *cas6* or combinations thereof.

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.

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.

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.

The *cas* gene 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.

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.

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.

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.

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.

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

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

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

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 least 99% of the wild-type sequence.

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*.

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

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

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  
 5 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.

10 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.

15 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 ;  
 20 Cas1, Cas2, Cas3, Cas4, Cas1B, Cas5 and Cas6 or combinations thereof.

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  
 25 combinations thereof.

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.

30 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.

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  
5 Cas proteins, 40 Cas proteins or even 50 or more Cas proteins.

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  
10 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 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  
15 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  
20 proteins may be duplicated a plurality of times.

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  
25 and 7 proteins genes – such as 4, 5, 6, or 7 Cas proteins.

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

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  
30 combinations thereof.

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

In a further aspect, there is provided 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.

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 one or more, preferably, 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.

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

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.

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

## 5 CRISPR REPEAT

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.

As used herein, the term "CRISPR" is synonymous with the term "CRISPR  
10 repeat".

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.

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.

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.

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.

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

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

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

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.

In a further aspect of the present invention, there is also provided 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.

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

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

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

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.

Accordingly, particular combinations of one or more *cas* genes or proteins and one or more, preferably, two or more CRISPR repeats are used in order for a CRISPR  
5 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.

In the context of the CRISPR repeat-*cas* gene or protein combination described  
10 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.

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

Suitably, 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 the same cell (eg. the same recipient cell).

20 In one embodiment, the term "derivable" is synonymous with the term "obtainable".

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

Suitably, the one or more *cas* genes or proteins and/or the one or more,  
25 preferably, 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.

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

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

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

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

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

Accordingly, in a further aspect, there is provided a combination or nucleic acid consisting essentially of at least two CRISPR repeats and at least one *cas* gene or protein.

In one embodiment, the term “consists essentially of” refers to a combination  
15        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.

In one embodiment, the term “consists essentially of” refers to a combination  
20        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.

In a further embodiment, the term “consists essentially of” refers to a combination of at least two CRISPR repeats and at least one *cas* gene or protein only  
25        and excluding at least one further component of a CRISPR locus – preferably excluding at least one further component of a naturally occurring CRISPR locus.

In a further embodiment, the term “consists essentially of” refers 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  
30        (eg. substantially absent).

Suitably, there is provided 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.

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

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 one or more, preferably, two or more CRISPR repeats.

In one embodiment, the CRISPR spacer(s) are or are derivable (preferably,  
10 derived) from an organism (eg. a donor organism) that is different to the cell (eg. 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).

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

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  
20 repeats and 20 *cas* genes or proteins or any other combinations thereof.

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

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  
25 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,  
30 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),.

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

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

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); *cas1*, *cas2*, *cas3* and *cas5*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas1*, *cas2*, *cas3* 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* and *cas1B*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats); *cas1*, *cas2*, *cas3*, *cas4* 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*-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); *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*,  
 5 *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) ; *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*,  
 15 *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  
 20 *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.

25 The *cas* genes may be *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  
 30 (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) or *cas1B* and *cas6*-repeat (wherein the repeat is at least two repeats, preferably, with at least one spacer in between the repeats) or combinations thereof.

The *cas* genes may be a *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); *cas1*, *cas2*, *cas3* and *cas5*-repeat (wherein the repeat is

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

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.

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) or combinations thereof.

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.

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.

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.

In a still further embodiment of the present invention, a first and/or a second  
5 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)  
10 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.

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  
15 further detail herein.

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  
20 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.

In a further aspect, there is provided 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  
25 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.

In a further aspect, there is provided a method for identifying a functional  
30 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.

Suitably, the sequences of the *cas* gene or protein and the CRISPR repeat are or are derivable (preferably, derived) from the same or different strains.

Suitably, the sequences of the *cas* gene or protein and the CRISPR repeat are or are derivable (preferably, derived) from the same or different species.

Suitably, the sequences of the *cas* gene or protein and the CRISPR repeat are or are derivable (preferably, derived) from the same or different genera.

Suitably, the sequences of the *cas* gene or protein and the CRISPR repeat are or are derivable (preferably, derived) from the same or different organisms.

Suitably, the analysis is performed using dotplot analysis.

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

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

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

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

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

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

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

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

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

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

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

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

10 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.

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  
15 whether representing the sense or antisense strand or combinations thereof.

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.

20 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.

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.

25 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  
30 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.

In one embodiment, 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.

Clusters are described in further detail below.

#### 15 CRISPR SPACER

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.

20 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.* 25 2005; “Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria.” *Molecular Microbiology* 36:244-246).

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

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.

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

The number of CRISPR spacers at a given CRISPR loci or locus can vary between species.

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  
10 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.

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  
15 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  
20 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.

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

As described herein, the inventors have surprisingly discovered that the use of  
25 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.

As used herein, the term “specificity of immunity” means that immunity can be  
30 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.

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.

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

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.

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).

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.

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.

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  
5 and/or the one or more, preferably, two or more CRISPR repeats.

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  
10 invention is heterologous to the CRISPR repeat and/or *cas* genes of the CRISPR locus.

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  
15 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.

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  
20 proteins 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 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.

25 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.

30 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.

In a further aspect, there is also provided 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.

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.

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

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.

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

5

#### CRISPR SPACER CORE

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  
10 between about 20 bp and about 58 bp.

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.*  
15 *thermophilus* CRISPR Type 1), the CRISPR spacer core is defined as a continuous stretch of 28 bp.

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  
20 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.

25 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

30 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.

One or more of pseudo CRISPR spacers or CRISPR spacer(s) which is/are  
5 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  
10 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.

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  
15 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.

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)  
20 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 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  
25 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.

30 NUCLEIC ACID SEQUENCE

In a further aspect, there is provided a nucleic acid sequence (eg. a recombinant or an isolated nucleic acid sequence) consisting essentially of at least one *cas* gene or protein.

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.

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.

Accordingly, in one embodiment, there is provided a nucleic acid sequence consisting essentially of at least one *cas* gene and at least two CRISPR repeats.

In a further embodiment, there is provided a nucleic acid sequence consisting essentially of at least one *cas* gene and at least one CRISPR spacer.

In a further embodiment, there is provided a nucleic acid sequence consisting essentially of at least one *cas* gene, at least one CRISPR spacer and at least two CRISPR repeats.

In a further embodiment, there is provided 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).

In a further embodiment, there is provided 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).

In a further embodiment, there is provided 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.

In a further embodiment, there is provided 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.

5 In a further embodiment, there is provided 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.

10 In a further embodiment, there is provided 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.

15 In a further embodiment, there is provided 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.

20 In a further embodiment, there is provided 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.

25 In further embodiment, there is provided 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.

30 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.

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  
 5 moieties that do not deleteriously affect the basic characteristics of the composition.

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  
 10 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  
 15 those described herein.

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

20 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.

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,  
 25 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  
 30 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.

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).

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

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

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

The nucleotide sequence may be the same as a naturally occurring form, or 15 may be derivable (preferably, derived) therefrom.

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

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

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

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

In one embodiment, the target nucleic sequence is or is derivable (preferably, 25 derived) from a mobile genetic element.

In one embodiment, the target nucleic sequence is or is derivable (preferably, derived) from a transposable element or an insertion sequence.

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

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

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

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

5

#### MODULATING RESISTANCE

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.

10 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.

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  
15 that are tolerant of the target nucleic acid or a transcription product thereof.

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.

20 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  
25 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.

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

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.

In one aspect, there is provided a method for modulating (e.g. conferring or  
5 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  
10 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.

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  
15 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.

In one embodiment, the conserved sequence is a sequence that is essential for  
20 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.

In a further aspect, there is provided a method for modulating (eg. conferring or  
25 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)  
30 transforming a cell with said recombinant nucleic acid thus to render the recipient cell resistant to said target nucleic acid or transcription product thereof.

In a further aspect, there is provided 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  
5 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.

In one embodiment, one or more CRISPR spacers in a cell – such as a recipient  
10 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.

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

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.  
20 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.

The genetic engineering may even include, but is not limited to, adding (eg.  
25 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.

In another embodiment, “modulating resistance” means suppressing, reducing  
30 or decreasing the resistance of a cell to a target nucleic acid.

Thus, in a further aspect, there is provided 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.

According to this embodiment, the method comprises the steps of: (i) identifying one or more CRISPR spacers in an organism that is substantially resistant  
5 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.

In another embodiment, there is provided a method for modulating (eg. decreasing) the resistance of a cell comprising one or more *cas* genes or proteins and  
10 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;  
15 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.

20 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.  
25 insertion), deletion (eg. removal) or modification of one or more CRISPR spacers or portions or fragments thereof in the substantially resistant cell.

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.

30 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.

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.

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.

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.

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.

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.

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.

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

## 10 BACTERIOPHAGE

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

The bacteriophage is virulent to the cell.

The bacteriophage may be a virulent or a temperate bacteriophage.

15 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.

The bacteriophage may be a lytic bacteriophage or a lysogenic bacteriophage.

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.

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.

25 The term “bacteriophage” is synonymous with the term “phage”.

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.

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; “Identification of genes that are associated with DNA repeats in prokaryotes.” *Molecular Microbiology* 43:1565-1575; Mojica *et al.* 2005; “Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria.” *Molecular Microbiology* 36:244-246; Haft *et al.*, 2005; “A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes.” *Computational Biology* 1, 6 e60) including, but not limited to *Aeropyrum*, *Pyrobaculum*, *Sulfolobus*,  
 10 *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*,  
 15 *Campylobacter*, *Wolinella*, *Acinetobacter*, *Erwinia*, *Escherichia*, *Legionella*, *Methylococcus*, *Pasteurella*, *Photobacterium*, *Salmonella*, *Xanthomonas*, *Yersinia*, *Treponema* and *Thermotoga*.

By way of example, the bacteriophage include, but are not limited to, those bacteriophage capable of infecting bacteria belonging to the following genera: *Escherichia*,  
 20 *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*.

By way of further example, the bacteriophage include, but are not limited to, those  
 25 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.

By way of further example, the bacteriophage include, but are not limited to, those  
 30 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*  
 35 *casei*, *Bifidobacterium infantis*, *Lactobacillus paracasei*,

*Lactobacillus salivarius*, *Lactobacillus plantarum*, *Lactobacillus reuteri*, *Lactobacillus gasseri*, *Lactobacillus johnsonii* or *Bifidobacterium longum*.

By way of further example, the bacteriophages include, but are not limited to, those bacteriophage capable of infecting any fermentative bacteria susceptible to  
 5 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  
 10 (*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  
 15 (*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*  
 20 *saccharoperbutylacetonicum*).

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

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  
 25 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.

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

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  
5 as well.

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  
10 molecules containing sequences homologous to the spacers (Mojica *et al.* 2005).

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  
15 (eg. provide) resistance against a particular bacteriophage, thus substantially preventing phage attack.

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  
20 structural proteins, proteins with a conserved domain (eg. holing, lysine, and others) or conserved sequences amongst important phage genes.

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  
25 to an internal sequence of a phage gene, and even more "efficient" when this gene encodes "essential" proteins (eg. the antireceptor).

Accordingly, in a further aspect, there is provided 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  
30 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.

In a further aspect, there is provided a method for conferring resistance to a cell  
5 (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)  
10 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.

In a further aspect, there is provided a method for modulating the lysotype of a bacterial cell comprising the steps of: (a) providing one or more pseudo CRISPR  
15 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  
20 complementary or homologous to the one or more pseudo CRISPR spacer(s).

In a further aspect, there is provided 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  
25 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.

In a further aspect, there is provided a method for conferring resistance to a cell  
30 (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 homology to the pseudo CRISPR spacer of the bacteriophage comprising the target nucleic acid.

In a further aspect, there is provided a method for conferring resistance to a cell  
5 (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  
10 bacteriophage comprising the target nucleic acid.

In a further aspect, there is provided 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  
15 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.

In a further aspect, there is provided a method for modulating the lysotype of a bacterial cell comprising the steps of: (i) identifying a pseudo CRISPR spacer in a  
20 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.

25 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.

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  
30 herein.

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

Suitably, the target nucleic acid or transcription product thereof in the bacteriophage is a gene coding for a host specificity protein.

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

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).

Advantageously, bacterial cells may be prepared according to the present  
10 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.

In one embodiment, the term "low susceptibility to bacteriophage  
15 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  
20 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-  
25 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  
30 preferably after about 6 hours, more preferably after about 5 hours and more preferably after about 4 hours.

In a further aspect, there is provided 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.

10 In a further aspect, there is provided 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.

In still a further aspect, there is provided 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.

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.

Bacterial cells may therefore be prepared according to the present invention 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.

In one embodiment, the term "high susceptibility to bacteriophage multiplication" refers to the level of bacteriophage multiplication in a bacterium being  
5 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  
10 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  
15 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.

20 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

In a further embodiment, the target nucleic sequence or a transcription product  
25 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.

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  
30 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  
5 proteins.

Thus, according to this aspect, 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  
10 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.

In still a further aspect, there is provided a further method for conferring resistance to a cell against plasmid transfer comprising the steps of: (a) identifying a  
15 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  
20 render the cell resistant.

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  
25 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),  
30 insertion sequences (IS26) or transposase genes.

Thus, according to this aspect, 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  
 5 mobile genetic element(s) of the cell.

In still a further aspect, there is provided 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  
 10 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.

15 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<sub>tem</sub>*, *bla<sub>rob</sub>*, *bla<sub>shv</sub>*, *aadB*, *aacC1*, *aacC2*, *aacC3*, *aacA4*, *mecA*, *vanA*, *vanH*, *vanX*,  
 20 *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,  
 25 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*.

30 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.

5        Thus, according to this aspect, 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  
10       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.

      In still a further aspect, there is provided a method for modulating the acquisition of antibiotic resistance markers in a cell comprising the steps of: (a)  
15       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  
20       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.

      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  
25       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.

30       Thus, according to this aspect, 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.

5 In still a further aspect, there is provided 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  
10 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.

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

#### MODIFICATION

Nucleic acid sequences may be modified by genetically engineering nucleic acid sequences.

20 All or part of a nucleic acid sequence may be modified.

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

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

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

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

30 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  
5 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  
10 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*  
15 *Enzymology: DNA Structure Part A: Synthesis and Physical Analysis of DNA Methods* in *Enzymology*, Academic Press.

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.

20 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.

Suitably, one or more CRISPR spacers or pseudo CRISPR spacers may be  
25 inserted into at least one CRISPR locus.

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.

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

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

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

5 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.

10 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  
15 CRISPR spacers of a cell.

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.

20 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  
25 remains intact in the cell.

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.

30 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.

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  
5 "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.

In another embodiment, the CRISPR spacer(s) or pseudo CRISPR spacer(s) are  
10 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.

In another embodiment, the intervening nucleotide may be called a leader  
15 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.

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

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  
25 repeats of the cell.

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.

In another embodiment, one or more CRISPR spacers or pseudo CRISPR  
30 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.

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.

5 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.

10 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.

15 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.

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.

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

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

25 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.

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

30 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.

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  
5 *cas* gene in a substantially sensitive cell.

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  
10 then transformed into the cell, using methods such as those described herein.

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  
15 or a vector and the like which is then transformed into the cell, using methods such as those described herein.

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.  
20 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.

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  
25 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.

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.

30 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

5 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.

Rather, for the CRISPR repeat-*cas* combination(s) to be functional they should  
10 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.

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  
15 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.

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  
20 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.

25 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.

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

30

### DETERMINING RESISTANCE

In a further aspect, there is provided 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.

5 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.*

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  
10 or sequenced so as to predict/determine the likely resistance profile of a particular cell.

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.

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

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  
20 more bacteriophage which can be used as a lysotype predictor for microbial selection.

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.

25

### RECIPIENT CELL

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

30 In one embodiment, the recipient cell refers to any cell comprising the recombinant nucleic acid according to the present invention.

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.

5 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.

The recipient cell may comprise one or more genetically engineered CRISPR  
10 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.

The recipient cell may comprise one or more recombinant CRISPR repeats  
15 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.

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.

20 By "naturally occurring" we mean occurring naturally in nature.

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  
25 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.

Suitably, the recipient cell is a prokaryotic cell.

Suitably, the recipient cell is a bacterial cell. Suitable bacterial cells are  
30 described herein.

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.

Suitable species include, but are not limited to *Lactococcus lactis*, including  
 5 *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*.

10 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*  
 15 *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).

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*,  
 20 *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).

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

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.

30 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)).

- 5           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)).
- 10           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)).

- The cell in which resistance is to be modulated may be a bacterium that naturally
- 15           comprises one or more CRISPR loci. CRISPR loci have been identified in more than 40 prokaryotes (Jansen *et al.* 2002b; “Identification of genes that are associated with DNA repeats in prokaryotes.” *Molecular Microbiology* 43:1565-1575; Mojica *et al.* 2005; “Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria.” *Molecular Microbiology* 36:244-246; Haft *et al.* 2005; “A Guild of 45
- 20           CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes.” *Computational Biology* 1, 6 e60) including, but not limited to *Aeropyrum*, *Pyrobaculum*, *Sulfolobus*, *Archaeoglobus*, *Halocarcula*, *Methanobacterium*, *Methanococcus*, *Methanosarcina*, *Methanopyrus*, *Pyrococcus*, *Picrophilus*, *Thermoplasma*, *Corynebacterium*, *Mycobacterium*, *Streptomyces*, *Aquifex*, *Porphyromonas*, *Chlorobium*,
- 25           *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*, *Xanthamonas*, *Yersinia*, *Treponema* and *Thermotoga*.

- 30           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

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  
5 CRISPR spacers are derivable (preferably, derived). These can be the same or different.

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  
10 (preferably, derived). These can be the same or different.

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

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  
15 transcription product thereof.

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.

20 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.

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  
25 immunity against the target nucleic acid or transcription product thereof.

The organism may be a bacterial cell or a bacteriophage.

Suitably, the organism is a bacteriophage.

### HOST CELLS

30 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.

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  
5 prokaryotic (for example bacterial) cells.

Aspects of the present invention also relate to host cells comprising the combination, construct or the vector of the present invention. 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  
10 example, be prokaryotic (for example bacterial) cells.

### CONSTRUCT

In a further aspect, there is provided a construct comprising one or more of the nucleic acid sequences described herein.

15 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  
20 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.

The term "regulatory sequences" includes promoters and enhancers and other  
25 expression regulation signals.

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

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  
30 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

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

Thus, in a further aspect there is provided a vector comprising one or more of  
5 the constructs or sequences described herein.

The term "vector" includes expression vectors and transformation vectors and shuttle vectors.

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  
10 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".

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

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

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  
20 organism.

The vectors may be used *in vitro*, for example for the production of RNA or used to transfect or transform a host cell.

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

### TRANSFECTION

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-  
30 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.

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

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.

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

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

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.

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.

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'.

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

5 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  
10 *Lactobacillus* species including the *Lactobacillus acidophilus*, *Enterococcus* species, *Pediococcus* species, a *Leuconostoc* species and *Oenococcus* species.

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  
15 include the widely used *Lactococcus lactis*, including *Lactococcus lactis* subsp. *lactis* and *Lactococcus lactis* subsp. *cremoris*.

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  
20 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*,  
25 *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.

30 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.

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  
 5 obtain a desired flavour.

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.

The selection of organisms for the starter culture of the invention will depend  
 10 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.

15 The starter culture may even be a dried starter culture.

The starter culture may be a concentrated starter culture.

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

The starter culture may be a frozen starter culture.

20 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.

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

25

#### LACTIC ACID BACTERIA

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

As used herein the term "lactic acid bacteria" refers to Gram positive,  
 30 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.

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

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).

#### 15 PREPARING STARTER CULTURES

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

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.

25 Dried starter cultures for use in the present invention 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, spray-dried, freeze-dried or lyophilised.

The dried starter cultures for use in the present invention 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.

The starter cultures for use in the present invention 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 invention 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.

## 10 PRODUCT

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

Any product, which is prepared from, or comprises, a culture is contemplated in accordance with the present invention. 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.

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

20 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.

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.

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.

30 The cells described herein may be – or may be added to – a food ingredient, a food supplement, or a functional food.

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.

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.

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.

There is also provided a method of preparing a food, the method comprising admixing the cells according to the present invention with a food ingredient (such as a starting material for a food). The method for preparing a food is also another aspect of the present invention.

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.

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.

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

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.

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

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  
5 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%,  
10 more preferably at least 97%, more preferably at least 98%, most preferably at least 99% of the wild-type sequence.

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  
15 most preferably 99% activity of the wild-type polypeptide or nucleotide sequence.

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%,  
20 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.

25 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  
30 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.

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  
5 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.

10 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  
15 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.

Preferably, a functional CRISPR repeat-*cas* combination comprises at least  
20 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  
25 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.

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%,  
30 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.

5 The fragment may be a functional fragment.

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.

10 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".

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.

15 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.

20 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.

% 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.

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  
5 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  
10 by inserting "gaps" in the sequence alignment to try to maximise local homology.

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  
15 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  
20 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.

Calculation of maximum % homology therefore firstly requires the production of an optimal alignment, taking into consideration gap penalties. A suitable computer  
25 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  
30 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).

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.

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.

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

|                  |   |
|------------------|---|
| <b>FOR BLAST</b> |   |
| GAP OPEN         | 0 |
| GAP EXTENSION    | 0 |

20

|                    |            |                |          |
|--------------------|------------|----------------|----------|
| <b>FOR CLUSTAL</b> | <b>DNA</b> | <b>PROTEIN</b> |          |
| WORD SIZE          | 2          | 1              | K triple |
| GAP PENALTY        | 10         | 10             |          |
| GAP EXTENSION      | 0.1        | 0.1            |          |

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.

5 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.

15 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 |

20 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 ornithine (hereinafter referred to as O), pyriylalanine, thienylalanine, naphthylalanine and phenylglycine.

5 Replacements may also be made by unnatural amino acids include; alpha\* and alpha-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\*,  $\beta$ -alanine\*, L- $\alpha$ -amino butyric acid\*, L- $\gamma$ -amino butyric acid\*, L- $\alpha$ -amino isobutyric acid\*, L- $\epsilon$ -amino caproic acid<sup>#</sup>, 7-amino heptanoic acid\*, L-methionine sulfone<sup>#\*</sup>, L-norleucine\*, L-norvaline\*, 10 p-nitro-L-phenylalanine\*, L-hydroxyproline<sup>#</sup>, L-thioprolin\*, methyl derivatives of phenylalanine (Phe) - such as 4-methyl-Phe\*, pentamethyl-Phe\*, L-Phe (4-amino)<sup>#</sup>, L-Tyr (methyl)\*, L-Phe (4-isopropyl)\*, L-Tic (1,2,3,4-tetrahydroisoquinoline-3-carboxyl acid)\*, L-diaminopropionic acid<sup>#</sup> and L-Phe (4-benzyl)\*. The notation \* has been 15 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.

Variant amino acid sequences may include suitable spacer groups that may be 20 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  $\beta$ -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 25 amino acid residues wherein the  $\alpha$ -carbon substituent group is on the residue's nitrogen atom rather than the  $\alpha$ -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.

30 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

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.

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*

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

Strain – *Streptococcus thermophilus* ST0089

Phage – 2972

*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.

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.

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

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).

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.

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.

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.

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.

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

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

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*

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*

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*

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*

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*

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

5

*Example 7*

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.

10

*Example 8*

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.

15

*Example 9*

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.

20

*Example 10*

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.

25

*Example 11*

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

30

(A) *Streptococcus thermophilus* strain DGCC7710RH1

*Streptococcus thermophilus*

*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  
 5 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).

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

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

Virulent bacteriophage

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

Phage resistant mutant

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

WO 2007/025097

PCT/US2006/033167

103

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  
 5 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.

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

10 >CRISPR1\_DGCC7710-RH1 (SEQ ID NO: 668)

```

caaggacagttattgattttataatcactatgtgggtataaaaaacgtcaaaatttcatttgag
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtcaacaattgcaacatcttataaccactt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgtttgacagcaaatcaagattcgaattgt
15 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgacgaggagctattggcacaacttaca
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatttgacaatctgctgaccactgttattc
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacacttggcaggcttattactcaacagcga
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACctgttcttctgttcttttgttgtatcttttc
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtctattcttctcgtttttgttgggaatcct
20 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgctggcgaggaaaacgaacaaggcctcaaca
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcatagagtggaaaactagaaacagattcaa
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataatgccgttgaattacacggcaaggta
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgagcgagctcgaaataatcttaattacaag
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcttcgctagcgtcatgtggttaacytattta
25 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACggcgtcccaatctctgattaatacttactcg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaacacagcaagacaagaggatgatgctatg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgacacaagaacgtatgcaagagttcaag
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacaattcttctcctcggttaactgctcaagt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgaaggcctagaaagggagacaacatg
30 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatattttaaaatcattttcataacttcat
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcagtatcagcaagcaagctgttagttact
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataaactatgaaattttataatttttaaga
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaataatttatggtatagcttaatatcattg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgcacgagcagcttcgagttaccggttc
35 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtctatatcgagggtcaactaacaattatgct
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcggtcaaattctgttttaggtacattt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcaatcagacaagagttaaaatgggtctt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcttagctgtccaatccacgaacgtggatg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcaaccaacggtaacagctactttttacagt
40 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataactgaaggataggagcttgtaaagtct
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtaatgctacatctcaaaggatgatccaga
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaagtagttgatgacctctacaatgggttat
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacctagaagcatttgagcgtatattgattg
45 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattttgccccttctttgccccttgactag
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAGT
ttgattcaacataaaaagccagttcaattgaacttggtttt

```

Legend

50

WO 2007/025097

PCT/US2006/033167

104

Leader sequence:

5' caaggacagttattgattttataatcactatgtgggtataaaaacgtcaaaatttcatttgag 3'

(SEQ ID NO: 666)

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

CRISPR Repeats

Terminal repeat: 5' gttttgtactctcaagatttaagtaactgtacagt 3' (SEQ ID NO: 2)

10

Trailer sequence: 5' ttgattcaacataaaaagccagttcaattgaacttggtttt 3' (SEQ ID NO: 667)

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

15 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    | (SEQ ID NO: 669) |
|        |       |                               |       |                  |
| D858   | 31921 | tcaacaattgcaacatcttataaccactt | 31950 |                  |

20

The new spacer that is integrated into the CRISPR1 locus of *Streptococcus*  
 25 *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

30 *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.

CRISPR1 of *Streptococcus thermophilus* strain DGCC7710-RH2 contains 34 repeats (including the terminal repeat), and thus 33 spacers. When compared to the  
 35 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

WO 2007/025097

PCT/US2006/033167

105

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.

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

5 >CRISPR1\_DGCC7710-RH2 (SEQ ID NO: 670)  
 caaggacagttattgattttataatcactatgtgggtataaaaaacgtcaaaatttcatttgag  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACttacgtttgaaaagaatatcaaatcaatga  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgtttgacagcaaatcaagattcgaattgt  
 10 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgaacgaggagctattggcacaacttaca  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatttgacaatctgctgacctgtttatc  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacacttggcaggcttattactcaacagcga  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACctgttccttgttcttttgtgtatcttttc  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACttcattcttctcgtttttgtttgcgaatcct  
 15 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgctggcgaggaaacgaacaaggcctcaaca  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcatagagtggaaaactagaaacagattcaa  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataatgcogttgaattacacggcaagggtca  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgagcgagctcgaaataatcttaattacaag  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcttcgctagcgtcatgttgtaacgtattta  
 20 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcggttcctcaatcctgattaatacttactcg  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaacacagcaagacaagaggatgatgctatg  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgacacagaacgtatgcaagagttcaag  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacaattcttcatcggtaactgctcaagtg  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatlaagggcatagaaaggagacaacatg  
 25 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatatttaaaatcattttcataacttcat  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcagtatcagcaagcaagctgttagttact  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataaactatgaaattttataatttttaaga...  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatatttatggtatagcttaatatcattg  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgcacgagcacgttcgagtttaccgtttc  
 30 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtctatatcgaggtcaactaacaattatgct  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcggttcaaatctgttttaggtacattt  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcaatacagacaagagttaaaatggtctt  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcttagctgtccaatccacgaacgtggatg  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcaaccaacggtaacagctactttttacagt  
 35 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataactgaaggataggagcttgtaaagtct  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtaatgctacatctcaaaggatgatccaga  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaagtagttgatgacctctacaatggtttat  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacctagaagcatttgagcgtatattgattg  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatittgccccctctttgcccccttgactag  
 40 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaccattagcaatcatttgtgcccattgagt  
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAGT  
 ttgattcaacataaaaagccagttcaattgaacttggtctt

#### Legend

45 Leader sequence:  
 5' caaggacagttattgattttataatcactatgtgggtataaaaaacgtcaaaatttcatttgag 3'  
 (SEQ ID NO: 666)  
 Integrated sequence comprising a CRISPR Repeat in upper case and a CRISPR  
 50 spacer (ie. tagging sequence) in lower case.  
 CRISPR Repeats  
 Terminal repeat; 5' gtttttgtactctcaagatttaagtaactgtacagt 3' (SEQ ID NO: 3)

WO 2007/025097

PCT/US2006/033167

106

Trailer sequence: 5' ttgattcaacataaaaagccagttcaattgaacttggcttt3' (SEQ ID NO: 667)

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

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):

10 spacer 1 ttacgtttgaaaagaatatcaaataatga 30 (SEQ ID NO: 671)  
|||||  
D858 17215 ttacgtttgaaaagaatatcaaataatga 17244

The new spacer that is integrated into the CRISPR1 locus of *Streptococcus*  
15 *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

20 Construct integration and knockout

#### Materials and Methods

##### Strains and plasmids

25 *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

30 *Streptococcus thermophilus* DGCC7778RT

*Streptococcus thermophilus* DGCC7778RT'

*Streptococcus thermophilus* DGCC7710R2 CRISPR mutant resistant to 2972

*Streptococcus thermophilus* DGCC7710R2S1S2

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

WO 2007/025097

PCT/US2006/033167

107

*Escherichia coli* pCR2.1 TOPO 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

5 pTOPOcas4ko contains an integral fragment of cas4

pTOPOS1S2 contains the S1S2 spacer construct

pTOPO RT contains the RT terminal repeat construct

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

pOR1cas1ko contains an integral fragment of cas1

pOR1cas4ko contains an integral fragment of cas4

pORIS1S2 contains the S1S2 spacer construct

purist contains the RT terminal repeat construct

15

#### Primers

##### Cas1

5'-caaatggatagagaacgc-3' and 5'-ctgataaggtgttcgtgtcc-3' (SEQ ID NOS: 672-673, respectively)

##### Cas4

20 5'-ggagcagatggaatācaagaaagg-3' and 5'-gagagactaggtgtctcagca-3' (SEQ ID NOS: 674-675, respectively)

##### S1S2 and RT

P1 5'-acaacaacagagaagtatctcattg-3' (SEQ ID NOS: 676)

P2 5'-aacgagtacactcactattgtacg-3' (SEQ ID NOS: 677)

P3

5'-

25 tccactcacgtacaaatagtgagtgtactcgtttgtattctcaagatttaagtaactgtacagttgattcaacataaaaag-3' (SEQ ID NOS: 692)

P4 5'-cttccttcacctcgtttggtt-3' (SEQ ID NOS: 678)

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).

30

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).

*Streptococcus thermophilus* strains were grown at 37C or 42C in M17 (Difco Laboratories) supplemented with 0.5 % lactose or sucrose. For phage infection, 10mM CaCl<sub>2</sub> 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).

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).

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

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

The parent DGCC7778 is resistant to phage 858.

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

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) *CasI* knockout

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

(3) *Cas4* knockout

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

5

(4) S1S2 construct integration

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

10 SUMMARY

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.

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,

30

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.

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.

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.

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.

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).

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  
5 and 5). Thus any CRISPR locus could be highly structured into a series of regularly spaced DNA hairpins.

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  
10 by, for example, addition, by deletion, or by modification of the spacer sequences or even a complete CRISPR locus.

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

(i) Phage resistance. Particular CRISPR spacers derived from bacteriophage DNA may be added within a bacterial host CRISPR locus as to provide resistance against  
15 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)  
20 or conserved sequences amongst important phage genes.

(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  
25 immunity against plasmid DNA, such as sequences within the plasmid's origin of replication.

(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  
30 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).

(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.

(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.

(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.

(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.

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).

In a further aspect, there is provided the use of a combination of a CRISPR locus and one or more *cas* genes to provide resistance against a defined nucleic acid.

Suitably, the nucleic acid is DNA.

Suitably, the nucleic acid is RNA.

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

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

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

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

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

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

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

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

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

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.

20        In a further aspect, there is provided the use of CRISPR for identification and typing.

In a further aspect, there is provided 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  
25        product thereof.

**TABLE 1**

| Strains         | BIM on <sup>1</sup> | Phage 2972                     |                                    | Phage 858                      |                                      |
|-----------------|---------------------|--------------------------------|------------------------------------|--------------------------------|--------------------------------------|
|                 |                     | Phage sensitivity <sup>2</sup> | Spacer-phage homology <sup>3</sup> | Phage sensitivity <sup>2</sup> | Spacer-phage homology <sup>3</sup>   |
| 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 cas             |
| DGCC7778RT'     | 858                 | S                              | > 10 SNPs                          | S                              | No spacers left                      |
| DGCC7778cas1    | 858                 | S                              | > 10 SNPs                          | S                              | 100% (2 spacers) but cas1 KO         |
| DGCC7778cas4    | 858                 | S                              | > 10 SNPs                          | R                              | 100% (2 spacers) but cas4 KO         |
| DGCC7710-R2     | 2972                | R                              | 100% (1 spacer)                    | S                              | 5 SNPs                               |
| DGCC7710-R2S1S2 | 2972                | S                              | 100% but not next to cas           | 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

- 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.
- 5 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.
- 10 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  
15 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
- 20 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
- 25 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
- 30 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

5

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

- 10 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

- 15 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

The scope of the claims should not be limited by the preferred embodiment and examples, but should be given the broadest interpretation consistent with the description as a whole.

20

**We claim:**

1. A DNA construct comprising a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (*cas*) gene from a CRISPR locus, one or more CRISPR repeats, and a heterologous regulatory sequence operably linked to at least one nucleotide sequence component of the construct, wherein:

the *cas* gene comprises one of the nucleotide sequences of the SEQ ID NO:s selected from the group consisting of SEQ ID NO:s 505-508, 510-516, 518-520 and 521 or a *cas* gene encoding a polypeptide having at least 100% identity to a polypeptide encoded by one of the nucleotide sequences selected from the group consisting of SEQ ID NO:s 505-508, 510-516, 518-520, and 521;

the one or more CRISPR repeats are from the same CRISPR locus as the *cas* gene and originate from a genome or a plasmid; and

the heterologous regulatory sequence is heterologous to the source of the *cas* gene.

2. A DNA construct comprising a *cas* gene comprising one of the nucleotide sequences of the SEQ ID NO:s selected from the group consisting of SEQ ID NO:s 505-507, 510, 512, 514-516, 518-520, and 521 wherein said *cas* gene is operably linked to a heterologous regulatory sequence which is heterologous to the source of the *cas* gene.

3. The construct according to claim 2, wherein said *cas* gene encodes a Cas protein which keeps the ability to form a functional combination with a CRISPR repeat.

4. A vector comprising the construct according to any one of claims 1 to 3.

5. A host cell comprising the construct of any one of claims 1 to 3 or the vector of claim 4.

6. A cell culture comprising the host cell of claim 5.

7. The cell culture according to claim 6, wherein said culture is a starter culture or a probiotic culture.

8. A food product or feed comprising the cell culture of claim 6 or claim 7.

9. A food or feed produced from the host cell of claim 5.

FIGURE 1

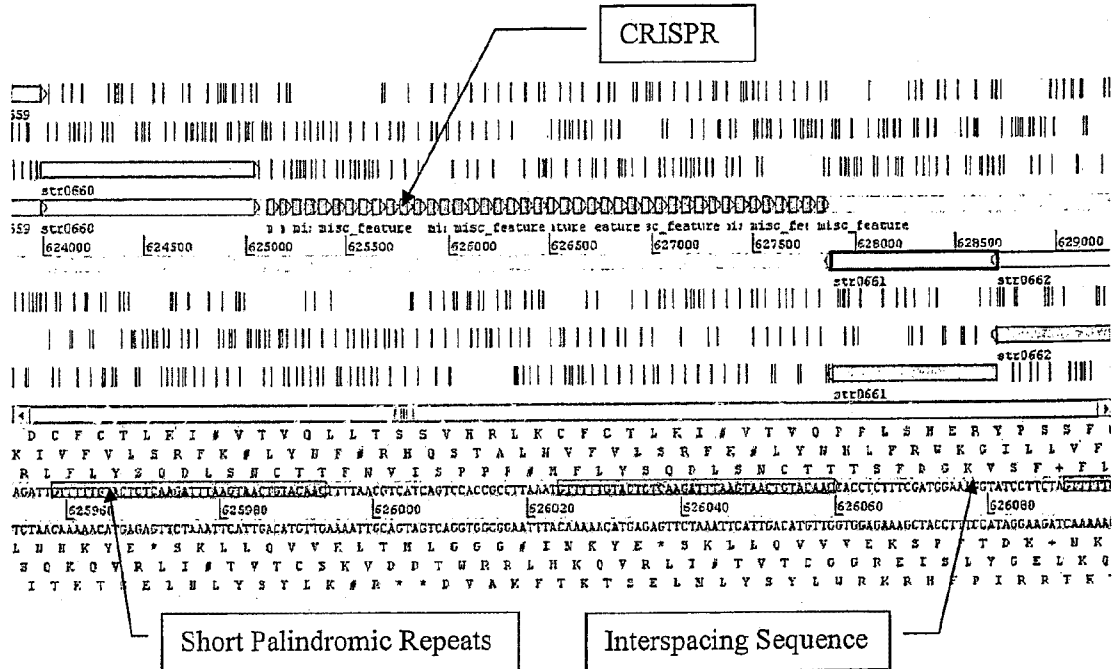


FIGURE 2

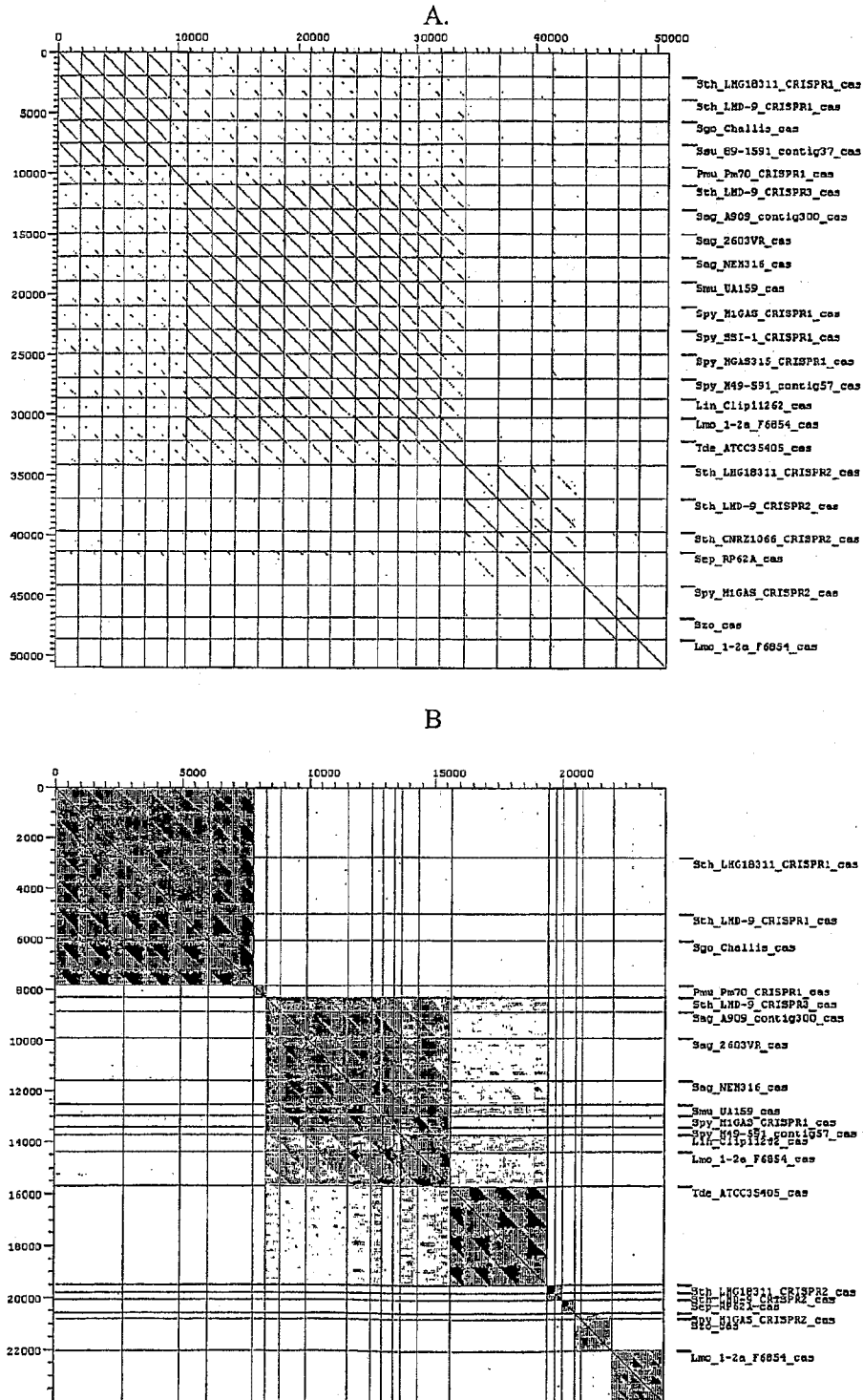


FIGURE 3

A

| Interspersing<br>sequence number | Phages |       |       |     |     |       |       |     |       |     |      |     |     |    |        |      | Number of<br>matches |
|----------------------------------|--------|-------|-------|-----|-----|-------|-------|-----|-------|-----|------|-----|-----|----|--------|------|----------------------|
|                                  | Sfi19  | Sfi21 | Sfi11 | DT1 | MD2 | Sfi18 | Sfi21 | SI3 | O1205 | DI2 | 2972 | DI4 | MD1 | Q5 | KappaJ | 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                    |

B

```

>gi|5523999|gb|AF115102.1|AF115102 S. thermophilus phage Sfi19
Query: 842 ttctggtagtggttttagtcaaacagatgt 871
          |||
Sbjct: 18912 ttctggtagtggttttagtcaaacagatgt 18941

>gi|5524032|gb|AF115103.1|AF115103 S. thermophilus phage Sfi21
Query: 842 ttctggtagtggttttagtcaaacagatgt 871
          |||
Sbjct: 17084 ttctggtagtggttttagtcaaacagatgt 17113

>gi|7669462|gb|AF158600.2|AF158600 S. thermophilus phage Sfi11
Query: 842 ttctggtagtggttttagtcaaacagatgt 871
          |||
Sbjct: 21396 ttctggtagtggttttagtcaaacagatgt 21425

>gi|61383211|gb|AF348736.2| S. thermophilus phage MD2
Query: 842 ttctggtagtggttttagtcaaacagatgt 871
          |||
Sbjct: 9043 ttctggtagtggttttagtcaaacagatgt 9072

>gi|2444080|gb|U88974.1| S. thermophilus temperate phage O1205
Query: 842 ttctggtagtggttttagtcaaacagatgt 871
          |||
Sbjct: 34602 ttctggtagtggttttagtcaaacagatgt 34631

>gi|15077558|gb|AF348739.1| S. thermophilus phage DT2
Query: 843 tctggtagtggttttagtcaaac 865
          |||
Sbjct: 2575 tctggtagtggttttagtcaaac 2597

>gi|56718416|gb|AY699705.1| S. thermophilus phage 2972
Query: 846 ggtagtggttttagtcaaacagatgt 871
          |||
Sbjct: 20033 ggtagtggttttagtcaaacagatgt 20058

>gi|7248462|gb|AF145054.1|AF145054 S. thermophilus phage 7201
Query: 842 ttctggtagtggttttagtcaaacagat 869
          |||
Sbjct: 29631 ttctggcagtggttttagtcaaacagat 29658

```

$$dG = -7.4 \text{ [initially } -7.4] \text{ sp}$$

FIGURE 5

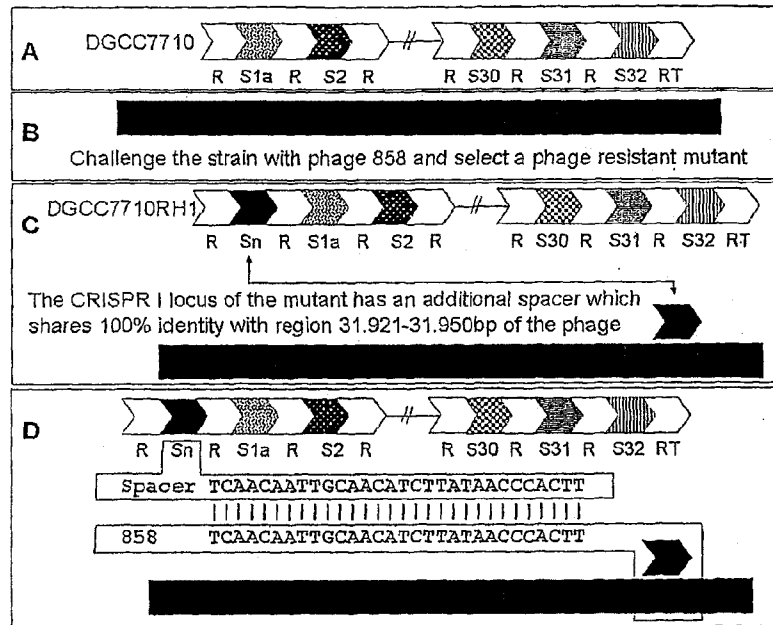


FIGURE 6

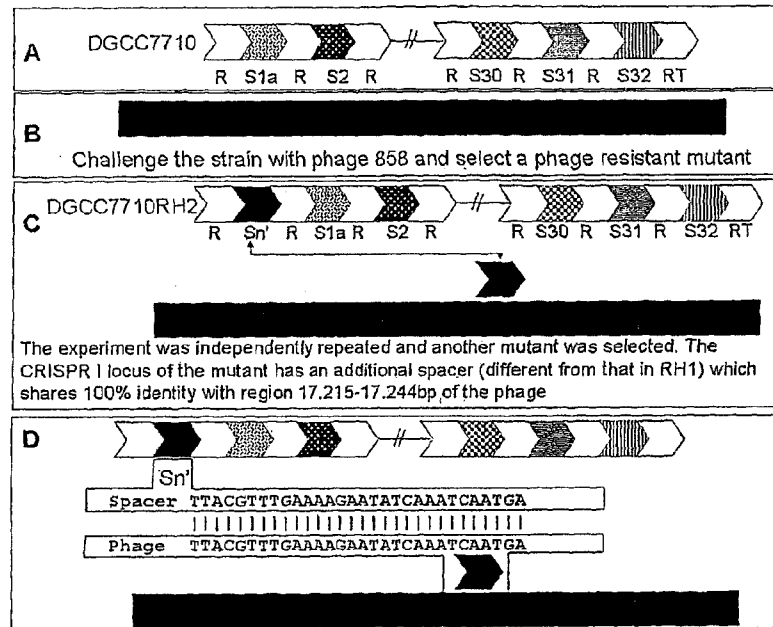


FIGURE 7

[illegible]

FIGURE 8

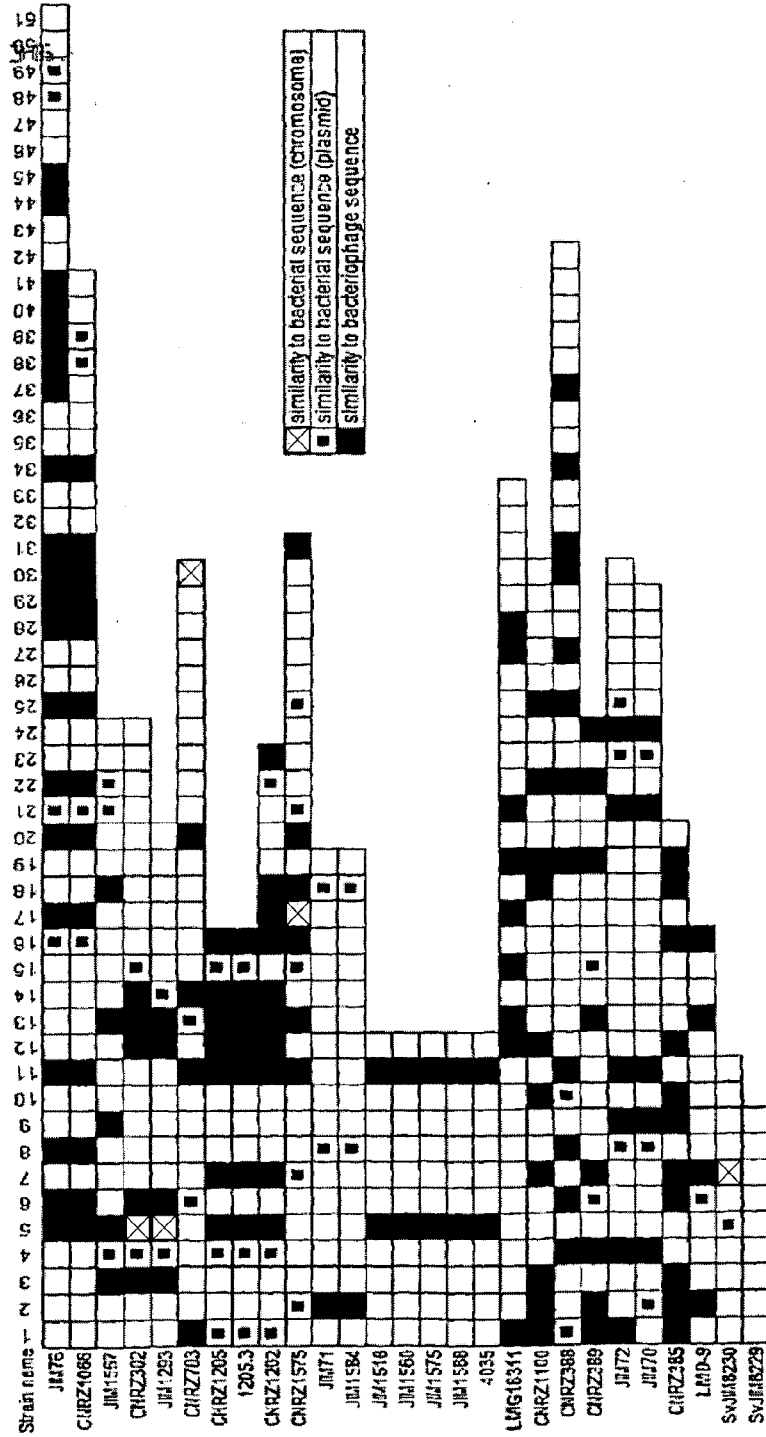


FIGURE 9

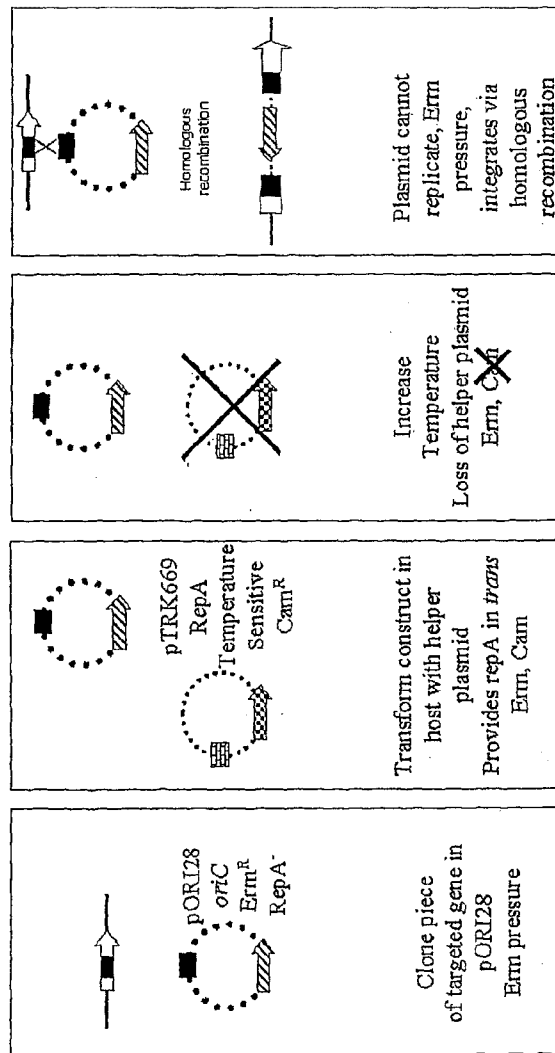
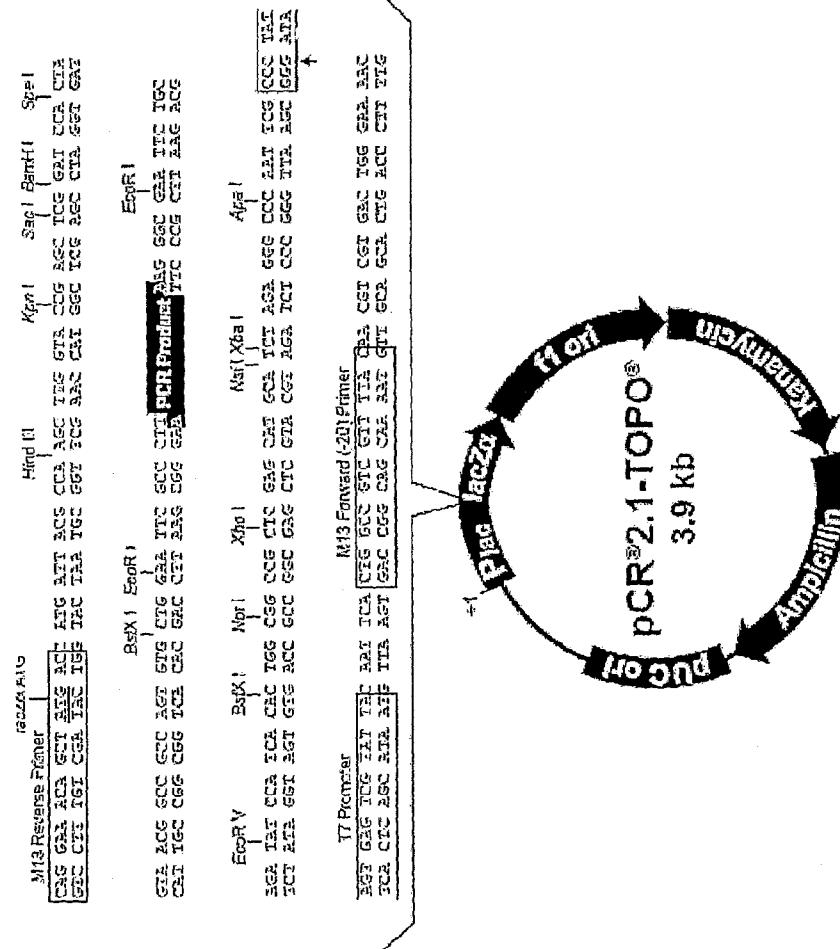


FIGURE 10



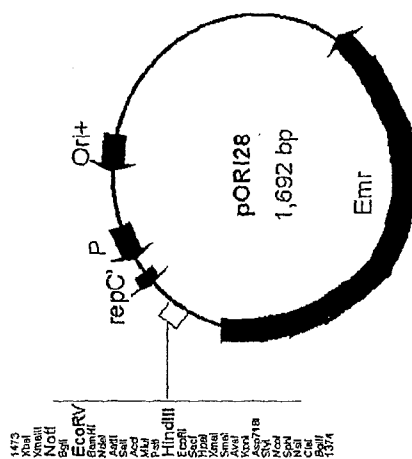


FIGURE 12

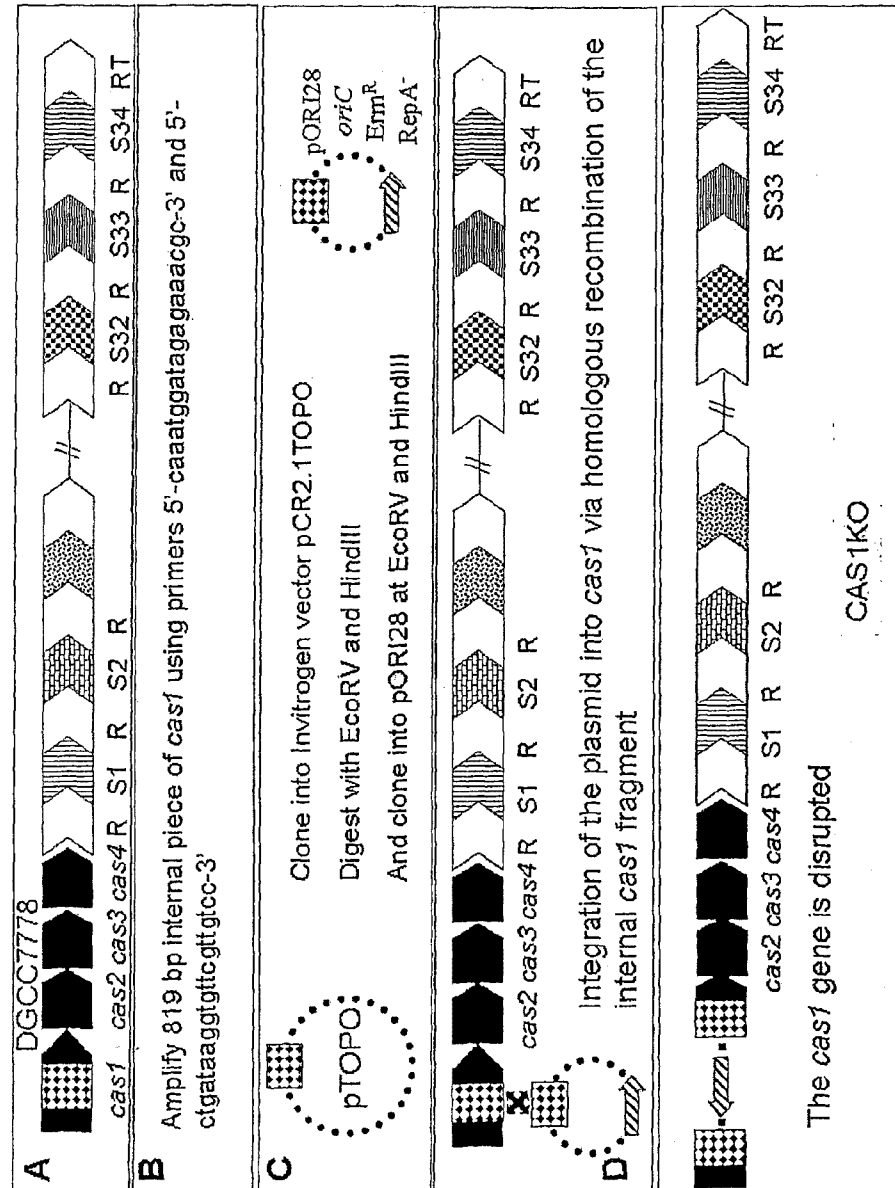


FIGURE 13

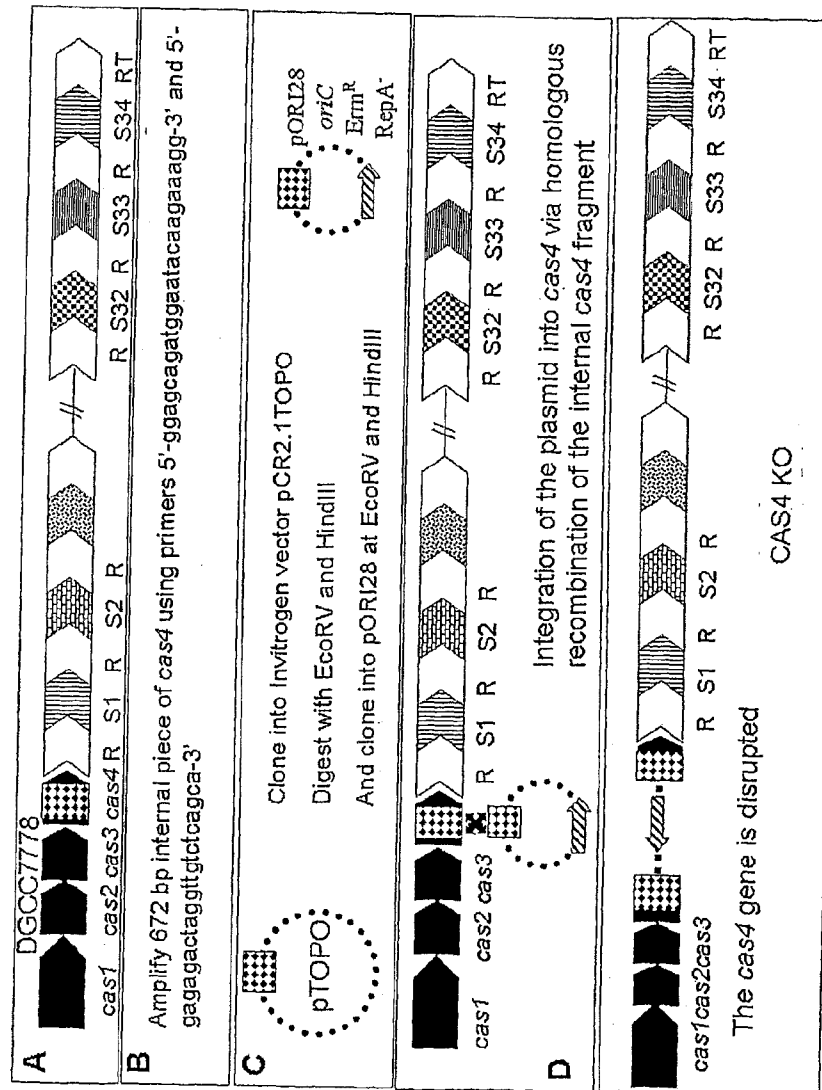


FIGURE 14

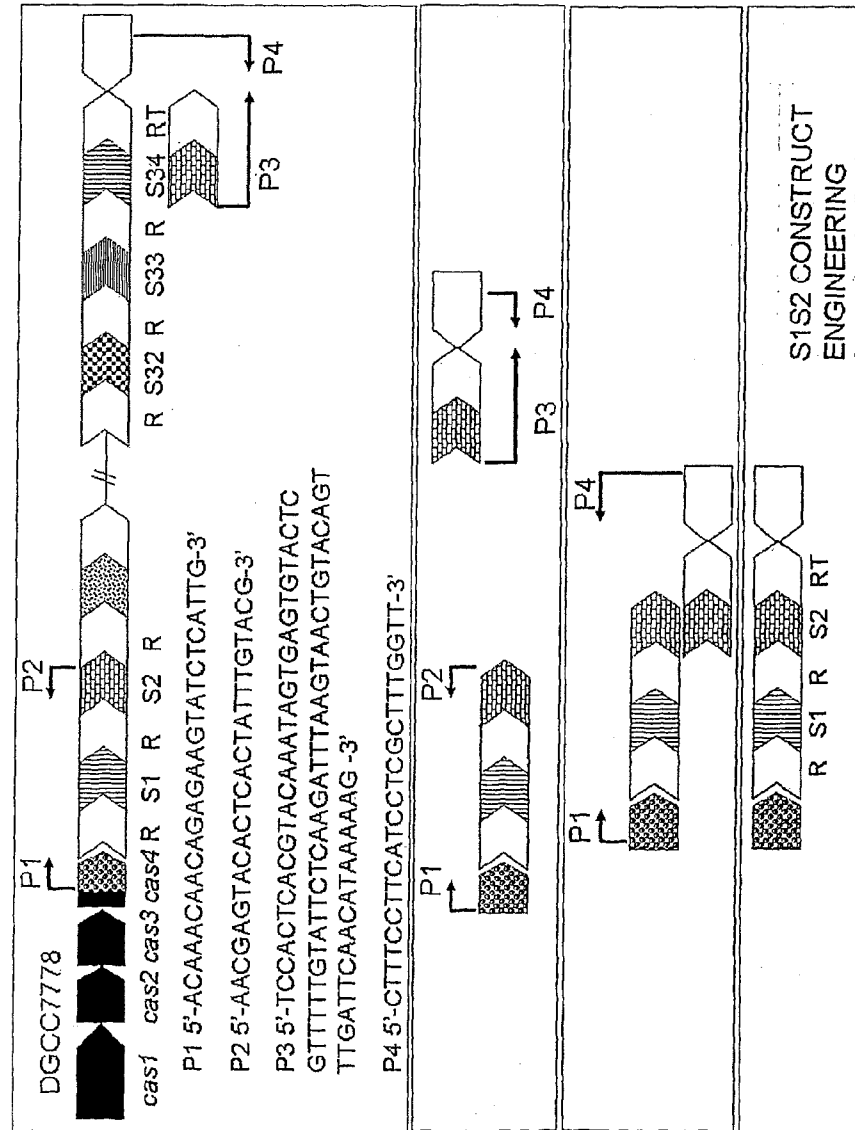
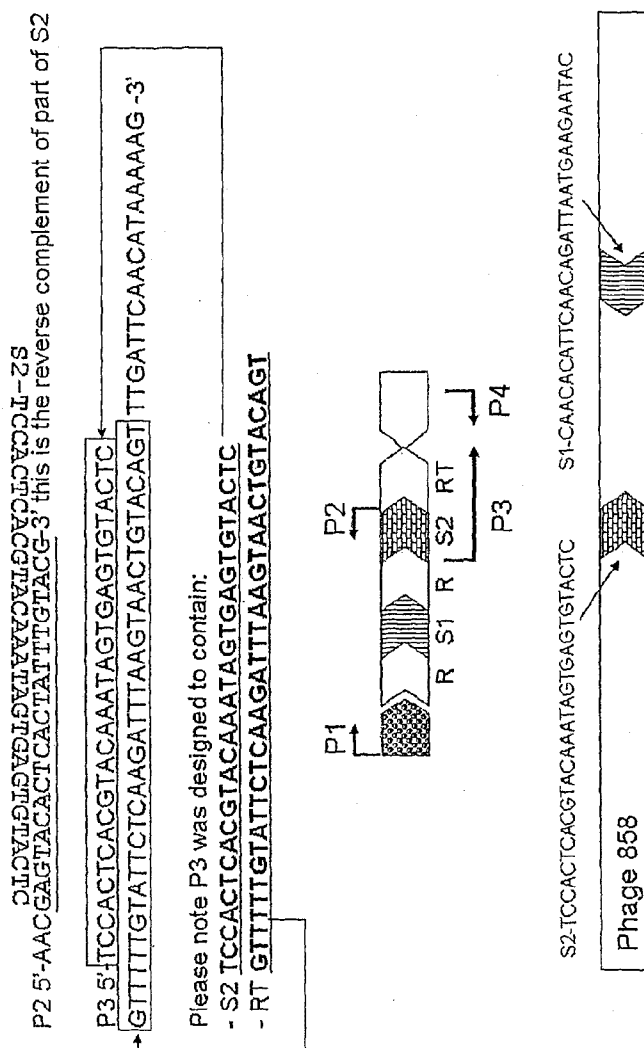


FIGURE 15



S1 and S2 both originate from phage 858. S1 is in ORF 38 at 31,381bp - S2 is in ORF 27 at 25,440bp

FIGURE 16

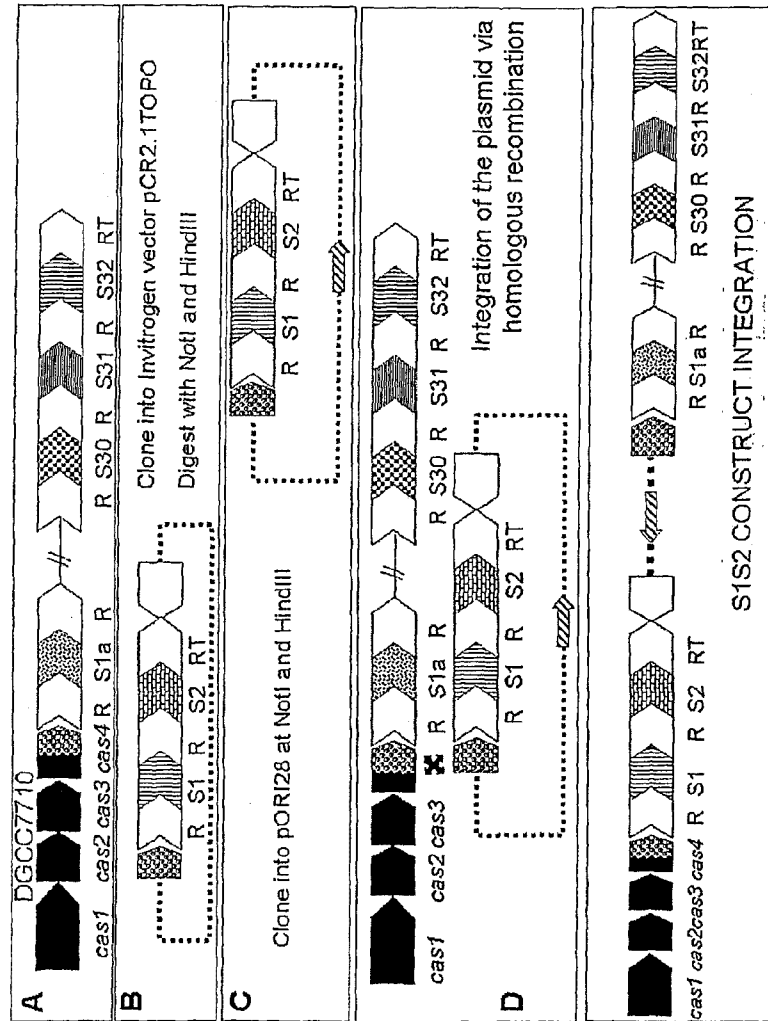


FIGURE 17

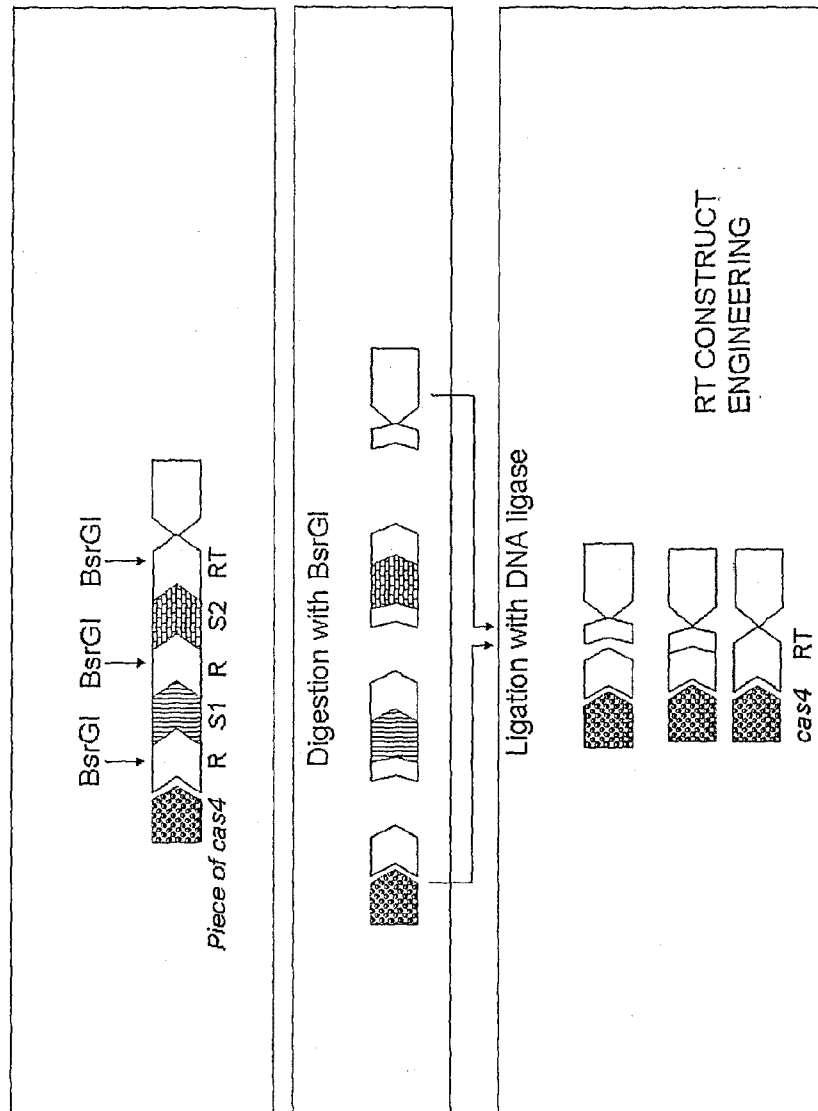


FIGURE 18

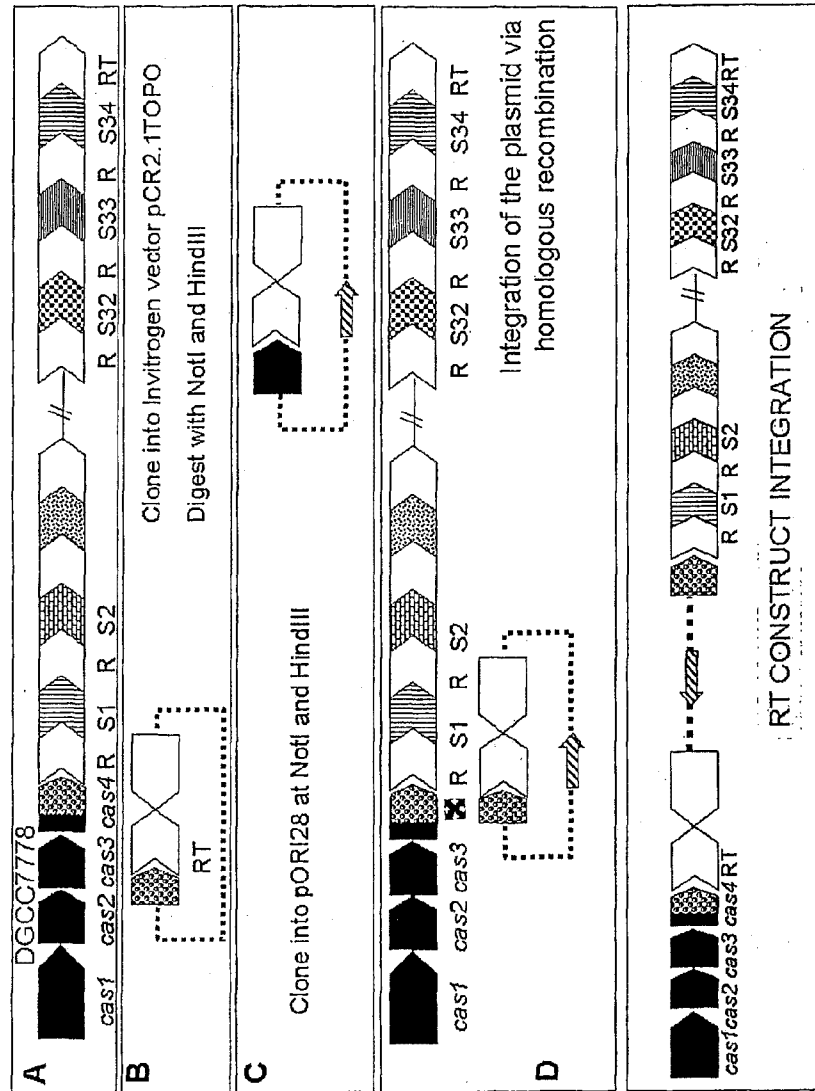


FIGURE 19

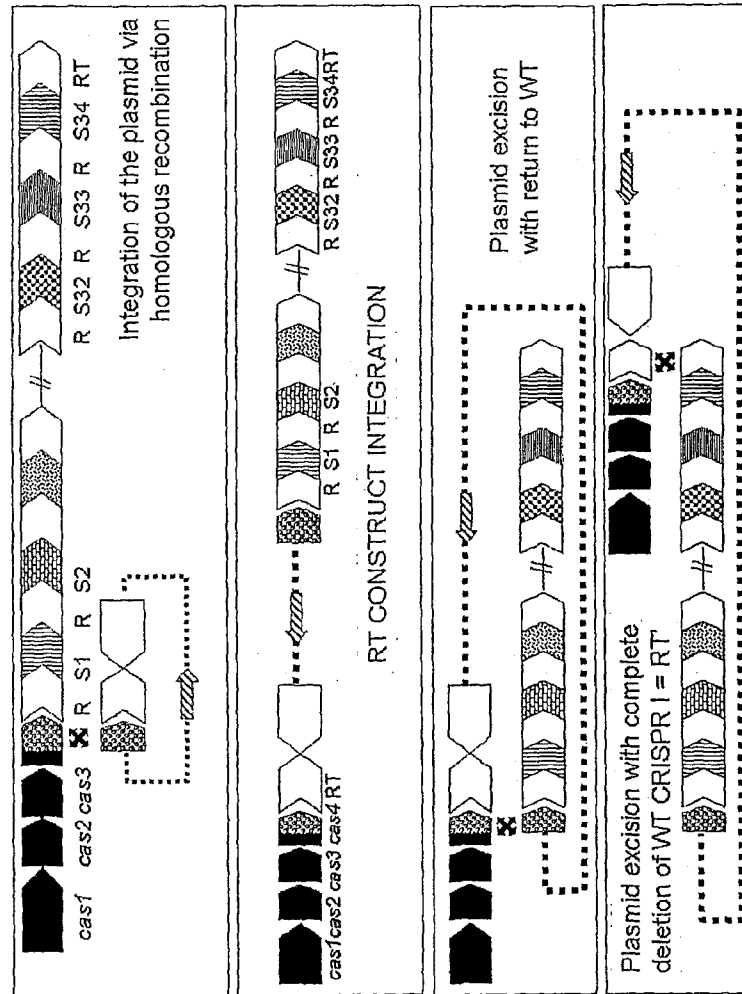


FIGURE 20

