Trends Biotechnol. 34, 575–587 (2016)). These systems are among the simplest CRISPR-Cas systems known as they introduce targeted double stranded DNA breaks (DSBs) based on a ribonucleoprotein (RNP) complex formed by a single Cas endonuclease and an RNA guide.

To date, *Streptococcus pyogenes* Cas9 (SpCas9) is the best characterized and most widely employed Cas9 for genome engineering. Although a few other type-II systems have been characterized, none of them is derived from a thermophilic organism (Nakade, *et al*. Bioengineered 1–9 (2017). doi:10.1080/21655979.2017.1282018). Characterization of such CRISPR-Cas systems would be interesting to gain fundamental insights as well as to develop novel applications.

Although basic genetic tools are available for a number of thermophiles (Taylor *et al*. Microb. Biotechnol. 4, 438–448 (2011); Olson, et al. Curr. Opin. Biotechnol. 33, 130–141 (2015); Zeldes, *et al*. Front. Microbiol. 6, 1209 (2015)), the efficiency of these tools is still too low to enable full exploration and exploitation of this interesting group of organisms. Based on our finding that SpCas9 is not active *in vivo* ≥42°C, we have previously developed an SpCas9-based engineering tool for facultative thermophiles, combining homologous recombination at elevated temperatures and SpCas9-based counter-selection at moderate temperatures (Mougiakos *et al. ACS Synth. Biol.* 6, 849–861 (2017)). However, a Cas9-based editing and silencing tool for obligate thermophiles is not yet available as SpCas9 is not active at or above 42°C (Mougiakos *et al. ACS Synth. Biol.* 6, 849–861 (2017)) and to date no thermophilic Cas9 has been characterized.

SUMMARY

The inventors have discovered and characterised ThermoCas9: an RNA-guided DNA-endonuclease from the CRISPR-Cas type-IIC system of the thermophilic bacterium *Geobacillus thermodenitrificans* T12. The inventors have surprisingly shown its *in vitro* activity over a broad temperature range, demonstrated the importance of the sgRNA-structure for thermostability and applied ThermoCas9 for *in vivo* genome editing across a wide temperature range.

Accordingly, in one aspect, the present disclosure provides a method of modifying the genetic material of a eukaryotic cell, comprising (i) integrating a polynucleotide encoding a ThermoCas9 under the control of a first promoter into the genome of the cell, wherein the expressed ThermoCas9 comprises an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 77% identity

therewith, or an active fragment thereof; (ii) transforming the cell with an expression vector comprising a polynucleotide sequence encoding a guide RNA and under the control of a second promoter, wherein the guide RNA has a nucleic acid sequence which recognizes a nucleic acid sequence comprised at a desired target locus in the genome of the cell, and (iii) transforming the cell with a repair oligonucleotide.

The ThermoCas9 nuclease of the specified sequence and variant and fragments noted above is the ThermoCas9 of the examples of the disclosure described hereinafter. The active fragments of the ThermoCas9 have guide RNA directed endonuclease activity with respect to target sequences in the genome of cells or organisms which are desired to be modified.

The genome of the cell may be modified first to express the ThemoCas9, whether constitutively or inducible, after which the cell is transformed with the expression vector for the gRNA; optionally at the same time or separately of transformation with the repair-oligo. In one possibility, the cell may be transformed simultaneously with all three elements, ThermoCas9 integration vector, gRNA expression vector and repair oligonucleotide.

In alternative example, the disclosure provides a method of modifying the genetic material of a eukaryotic cell, comprising (i) integrating a polynucleotide encoding a ThermoCas9 under the control of a first promoter into the genome of the cell, wherein the expressed ThermoCas9 comprises an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 77% identity therewith, or an active fragment thereof; and (ii) transforming the cell with an expression vector comprising a polynucleotide sequence encoding a guide RNA under the control of the first or of a separate second promoter, wherein the guide RNA has a nucleic acid sequence which recognizes a nucleic acid sequence comprised at a desired target locus in the genome of the cell, and a repair oligonucleotide also under the control of the first or the second promoter, or a separate third promoter.

The repair oligonucleotide is preferably a double-stranded DNA repair oligo; optionally comprising a polynucleotide sequence for insertion into the genome of the cell by way of homologous recombination following guide RNA directed ThermoCas9 endonuclease cutting. Therefore, the repair oligo which does not contain and insert may be used to achieve a deletion of a desired locus or gene segment in the genome of the cell.

In some examples, the first promoter is a constitutive promoter, e.g. *TEF1* promoter. In other examples, the first promoter may be a physical or chemically inducible promoter of the kinds described hereinafter in connection with other examples of the disclosure.

Various combinations of promoter are possible in accordance with the disclosure such that the the second promoter may be a constitutive or inducible promoter. Any third promoter may be a constitutive or inducible promoter.

The cell may be transformed with the expression plasmid(s) and/or the repair-oligo by heat-shock or for example by electroporation.

Certain eukaryotic cells are able to grow at temperatures up to about 62 °C – 63 °C. Various fungi or algae or blue green algae, for example. Therefore the methods of the disclosure may be employed wherein the cell is transformed and/or grown following transformation at a temperature in the range 26 °C - 63 °C; preferably 31°C – 61 °C; more preferably 35 °C – 60 °C; even more preferably 34 °C – 41 °C, e.g. 37 °C. Other temperatures may be used falling within the ranges selected from any of the following ranges, as shown by the upper and lower limit combinations marked “x” in the table below:

Upper Limit (°C) / Lower limit (°C)	62	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33
27	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
28	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
29	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
30	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x			
31	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				
32	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x					
33	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x						
34	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x							
35	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x								
36	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x									
37	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x										
38	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x											
39	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x												
40	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x													
41	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x														
42	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x															
43	x	x	x	x	x	x	x	x	x	x	x	x	x	x																
44	x	x	x	x	x	x	x	x	x	x	x	x	x																	
45	x	x	x	x	x	x	x	x	x	x	x	x																		
46	x	x	x	x	x	x	x	x	x	x	x																			
47	x	x	x	x	x	x	x	x	x	x																				
48	x	x	x	x	x	x	x	x	x																					
49	x	x	x	x	x	x	x	x																						
50	x	x	x	x	x	x	x																							
51	x	x	x	x	x	x																								
52	x	x	x	x	x																									
53	x	x	x	x																										
54	x	x	x																											
55	x	x																												
56	x																													

In some methods of the disclosure the eukaryotic cell is a fungus, particularly a yeast, e.g. *Saccaharomyces* sp., e.g. *S. cerevisiae.*

In another aspect, the present disclosure refers to a polynucleotide expression vector for modifying the genetic material at a target locus of a host organism which comprises the expression vector, the vector comprising:

a) a polynucleotide sequence encoding a ThermoCas9 nuclease, wherein the ThermoCas9 nuclease comprises an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 77% identity therewith, or an active fragment thereof;

b) a polynucleotide sequence encoding a guide RNA, wherein the guide RNA has a nucleic acid sequence which recognizes a nucleic acid sequence comprised in the target locus;

c) a first promoter oriented with respect to polynucleotide sequences of (a) and (b) so as to drive expression thereof in the organism.

The present disclosure also provides a polynucleotide expression vector for modifying the genetic material at a target locus of a prokaryotic host organism which comprises the expression vector, the vector comprising:

a. a polynucleotide sequence encoding a Cas9 nuclease, wherein the Cas9 nuclease comprises an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 77% identity therewith, or an active fragment thereof;
b. a polynucleotide sequence encoding a guide RNA, wherein the guide RNA has a nucleic acid sequence which recognizes a nucleic acid sequence comprised in the target locus;
c. a first promoter oriented with respect to polynucleotide sequences of (a) and (b) so as to drive expression thereof in the organism.

In preferred examples, the prokaryotic organisms are thermophilic bacteria, as further defined hereinafter.

The expression vector of the disclosure may be used alone or in conjunction with a second expression vector.

Preferably, the sequence of (a) is 3' of the promoter and the sequence of (b) is 3' of the sequence of (b).

Additionally, the expression vector may further comprise a polynucleotide sequence which encodes an homologous recombination (HR) fragment under the control of the first promoter or a separate second promoter.

Where a single expression vector is used in methods of the disclosure described herein, and there are first and second promoters, the first promoter may be an inducible promoter. The second promoter may be a constitutive promoter or an inducible promoter. When the second promoter is an inducible promoter it may be the same or different as the first inducible promoter.

The inducible promoter(s) may be physically or chemically inducible. Some preferred chemically inducible promoters include a β-glucosidase promoter inducible with cellobiose or a Pm promoter inducible with 3-methylbenzoate. When the second promoter contolling the HR fragment is a constitutive promoter then this may be the P3 promoter.

The arms of the HR fragment may comprise a nucleic acid sequence which allows recombination upstream and downstream, respectively, of a locus of interest in the host organism. In operation, a double recombination event involving each arm (respectively) of the HR fragment, either side of a locus of interest, serves to replace the locus with the HR

fragment. Therefore, expression of the guide RNA, the ThermoCas9 and the HR fragment by the vector results in the deletion of a locus of interest, e.g. gene of interest, from the genome of the prokaryotic organism.

The locus of interest herein comprises the target sequence recognized by the guide RNA.

The HR fragment may also comprise an insertion element between the upstream and downstream arms thereof. In this mode of operation, a vector of the disclosure when expressed in a transformed host cell results in a substitution at the locus of interest. When the insertion element is a gene of interest then expression of the vector in a transformed host results in a gene substitution, e.g. from a native gene in the organism to a new gene of interest, possibly an heterologous gene from another species or organism. In some instances, the new gene of interest may also be provided with a promoter, optionally an inducible promoter so that expression of the new gene can be switched on in the transformed organism as desired.

The locus of interest as described herein may comprise the PAM sequence 5'-NNNNCNN-3' located 3' of the target sequence; optionally at least 2, 3, 4, 5, 6 or more nucleotides from the target sequence.

In preferred vectors, the nucleotide sequence encoding the guide RNA preferably encodes a single guide RNA (sgRNA).

In another example, the disclosure provides methods employing any of the expression vectors of the disclosure described herein. Therefore, the disclosure provides a method of modifying the genetic material of a prokaryotic organism comprising transforming the organism with a first expression vector which comprises:

(a) a polynucleotide sequence encoding a Cas9 nuclease, wherein the Cas9 nuclease comprises an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 77% identity therewith, or an active fragment thereof;

(b) a polynucleotide sequence encoding a guide RNA, wherein the guide RNA has a nucleic acid sequence which recognizes a nucleic acid sequence comprised in the target locus;

(c) a first promoter oriented with respect to polynucleotide sequences of (a) and (b) so as to drive expression thereof in the organism; and,

a second expression vector comprising a polynucleotide sequence which encodes an homologous recombination (HR) fragment under the control of a promoter.

In another aspect, the present disclosure refers to an *in vitro* method of modifying the genetic material of an organism comprising transforming the organism with a first expression vector being a vector as disclosed herein, and a second expression vector comprising a polynucleotide sequence which encodes an homologous recombination (HR) fragment under the control of a promoter.

In yet another aspect, the present disclosure refers to an *in vitro* method of modifying the genetic material of a prokaryotic organism comprising transforming the organism with an expression vector as disclosed herein.

In yet another aspect, the present disclosure refers to a prokaryotic cell transformed with an expression vector as disclosed herein.

In another example, the disclosure provides a method of modifying the genetic material of a prokaryotic organism comprising transforming the organism with a single expression vector comprising (a), (b), and (c) above and also a polynucleotide sequence which encodes an homologous recombination (HR) fragment under the control of the said first promoter or a separate second promoter.

The organism may be transformed e.g. using electroporation.

In methods of the disclosure, the transformed organism is cultured for a period at a first temperature, and then cultured at a second temperature before or during induction of the promoter of the Cas9 sequence. Advantageously to improve the recombination efficiency, an incubation step may be carried out at the elevated temperature, at which the plasmid cannot replicate.

The first temperature may be 60 °C or less and the second temperature is a higher temperature which may be at least 55 °C. Alternativaely, the first temperature may be 55 °C or less and the second temperature may be greater than 55 °C. The different temperatures to be used may readily be selected by a person of average skill in the art, depending on the particular prokaryotic organism being modified, particularly a thermophilic bacterium.

In some preferred methods, the organism being modified is selected from *Geobacillus thermoglucosidans, Bacillus coagulans* or *Pseudomonas putida*.

Also provided by the disclosure is a prokaryotic cell, e.g. a bacterial cell, transformed with an expression vector of the disclosure as hereinbefore defined.

In accordance with the disclosure is a method of genetically modifying a thermophilic prokaryotic organism comprising transforming with a single expression plasmid as defined above at a first temperature and then subjecting the organism to an elevated temperature during which there is homologous recombination selected over plasmid multiplication. Such elevated temperatures may be in the range range 55°C to 100°C, 60°C to 100°C, 65°C to 100°C, 70°C to 100°C, 75°C to 100°C, 80°C to 100°C, 85°C to 100°C, 90°C to 100°C, 95°C to 100°C.

Further, the present disclosure provides an isolated clustered regularly interspaced short palindromic repeat (CRISPR)-associated (Cas) protein or polypeptide comprising;

a. the amino acid motif EKDGKYYC [SEQ ID NO: 2]; and/or
b. the amino acid motif $X_1X_2CTX_3X_4$ [SEQ ID NO: 3] wherein X_1 is independently selected from Isoleucine, Methionine or Proline, X_2 is independently selected from Valine, Serine, Asparagine or Isoleucine, X_3 is independently selected from Glutamate or Lysine and X_4 is one of Alanine, Glutamate or Arginine; and/or
c. the amino acid motif X_5LKX_6IE [SEQ ID NO: 4] wherein X_5 is independently selected from Methionine or Phenylalanine and X_6 is independently selected from Histidine or Asparagine; and/or
d. the amino acid motif X_7VYSX_8K [SEQ ID NO: 5] wherein X_7 is Glutamate or Isoleucine and X_8 is one of Tryptophan, Serine or Lysine; and/or
e. the amino acid motif $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6] wherein X_9 is Alanine or Glutamate, X_{10} is Glutamine or Lysine, X_{11} is Arginine or Alanine, X_{12} is Asparagine or Alanine and X_{13} is Lysine or Serine.

For the avoidance of doubt proteins, polypeptides or nucleic acids coding for Cas proteins of the present disclosure may also be referred to as “GtCas9” or “ThermoCas9”. “GtCas9” and “ThermoCas9” are used interchangeably throughout the specification and have the same meaning.

A polypeptide in the context of this disclosure may be viewed as a fragment of the full length Cas protein. Such fragments may be inactive and used in ways and for purposes not associated directly with binding, editing and/or cutting of genetic material, for example for standards in assays or raising antibodies or the like.

In preferred examples however, the Cas protein or polypeptide is functional and capable of cleavage, binding, marking or modifying at a temperature in the range 20°C and 100°C, inclusive, when associated with at least one targeting RNA molecule, and a polynucleotide comprising a target nucleic acid sequence recognised by the targeting RNA molecule. Preferably the Cas protein or polypeptide is functional and capable of said cleavage, binding, marking or modifying at a temperature in the range 50°C and 70°C, for example 55°C or 60°C.

In particular examples, the disclosure may provide a Cas protein or polypeptide comprising the amino acid motif EKDGKYYC [SEQ ID NO: 2]. In other examples, the Cas proteins or polypeptides may further comprise the amino acid motif $X_1X_2CTX_3X_4$ [SEQ ID NO: 3] wherein X_1 is independently selected from Isoleucine, Methionine or Proline, X_2 is independently

selected from Valine, Serine, Asparagine or Isoleucine, X_3 is independently selected from Glutamate or Lysine and X_4 is one of Alanine, Glutamate or Arginine.

In other examples the Cas proteins or polypeptides defined herein may additionally further comprise the amino acid motif X_5LKX_6IE [SEQ ID NO: 4] wherein X_5 is independently selected from Methionine or Phenylalanine and X_6 is independently selected from Histidine or Asparagine.

In other examples, the Cas proteins or polypeptides defined herein may additionally further comprise the amino acid motif X_7VYSX_8K [SEQ ID NO: 5] wherein X_7 is Glutamate or Isoleucine and X_8 is one of Tryptophan, Serine or Lysine.

In other examples, the Cas proteins or polypeptides defined herein may additionally further comprise the amino acid motif $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6] wherein X_9 is Alanine or Glutamate, X_{10} is Glutamine or Lysine, X_{11} is Arginine or Alanine, X_{12} is Asparagine or Alanine and X_{13} is Lysine or Serine.

In accordance with the present disclosure, it may be appreciated that a Cas protein or polypeptide of the disclosure may comprise any of the motifs of SEQ ID NOs 2 to 6, either alone or in combination. The following summarises each of the combinations of motifs which may characterize Cas proteins or polypeptides of the disclosure:

EKDGKYYC [SEQ ID NO: 2].

EKDGKYYC [SEQ ID NO: 2]; and $X_1X_2CTX_3X_4$ [SEQ ID NO: 3] wherein X_1 is independently selected from Isoleucine, Methionine or Proline, X_2 is independently selected from Valine, Serine, Asparagine or Isoleucine, X_3 is independently selected from Glutamate or Lysine and X_4 is one of Alanine, Glutamate or Arginine.

EKDGKYYC [SEQ ID NO: 2]; and $X_1X_2CTX_3X_4$ [SEQ ID NO: 3] wherein X_1 is independently selected from Isoleucine, Methionine or Proline, X_2 is independently selected from Valine, Serine, Asparagine or Isoleucine, X_3 is independently selected from Glutamate or Lysine and X_4 is one of Alanine, Glutamate or Arginine; and X_5LKX_6IE [SEQ ID NO: 4] wherein X_5 is independently selected from Methionine or Phenylalanine and X_6 is independently selected from Histidine or Asparagine.

X_{10} is Glutamine or Lysine, X_{11} is Arginine or Alanine, X_{12} is Asparagine or Alanine and X_{13} is Lysine or Serine.

X_5LKX_6IE [SEQ ID NO: 4] wherein X_5 is independently selected from Methionine or Phenylalanine and X_6 is independently selected from Histidine or Asparagine; and X_7VYSX_8K [SEQ ID NO: 5] wherein X_7 is Glutamate or Isoleucine and X_8 is one of Tryptophan, Serine or Lysine; and $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6] wherein X_9 is Alanine or Glutamate, X_{10} is Glutamine or Lysine, X_{11} is Arginine or Alanine, X_{12} is Asparagine or Alanine and X_{13} is Lysine or Serine.

X_5LKX_6IE [SEQ ID NO: 4] wherein X_5 is independently selected from Methionine or Phenylalanine and X_6 is independently selected from Histidine or Asparagine; and X_7VYSX_8K [SEQ ID NO: 5] wherein X_7 is Glutamate or Isoleucine and X_8 is one of Tryptophan, Serine or Lysine.

X_5LKX_6IE [SEQ ID NO: 4] wherein X_5 is independently selected from Methionine or Phenylalanine and X_6 is independently selected from Histidine or Asparagine; and $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6] wherein X_9 is Alanine or Glutamate, X_{10} is Glutamine or Lysine, X_{11} is Arginine or Alanine, X_{12} is Asparagine or Alanine and X_{13} is Lysine or Serine.

X_7VYSX_8K [SEQ ID NO: 5] wherein X_7 is Glutamate or Isoleucine and X_8 is one of Tryptophan, Serine or Lysine; and $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6] wherein X_9 is Alanine or Glutamate, X_{10} is Glutamine or Lysine, X_{11} is Arginine or Alanine, X_{12} is Asparagine or Alanine and X_{13} is Lysine or Serine.

In another example, the present disclosure provides an isolated Cas protein or polypeptide fragment thereof having an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 77% identity therewith, wherein the Cas protein or polypeptide fragment thereof comprises any of the following motifs or amino acids, either alone or in combination:

An RuvC-I domain comprising IGLDIGITSIG [SEQ ID NO: 23], preferably IGLDIGITSIGWAVINLD [SEQ ID NO: 24].

A Bridge domain comprising RSARR [SEQ ID NO: 25], preferably PRRLARSARRRLRRRKHRLERIRRL [SEQ ID NO: 26]; and/or

An α-helical/recognition lobe domain comprising WQLR [SEQ ID NO: 27]; and/or

An α-helical/recognition lobe domain comprising HLAKRRG [SEQ ID NO: 28], preferably LARILLHLAKRRG [SEQ ID NO: 29]; and/or
An α-helical/recognition lobe domain comprising IFAKQ [SEQ ID NO: 30], preferably EIKLIFAKQ [SEQ ID NO: 31]; and/or
An α-helical/recognition lobe domain comprising IWASQR [SEQ ID NO: 32]; and/or
KVGFCTFEPKEKRAPK [SEQ ID NO: 33]; and/or
FTVWEHINKLRL [SEQ ID NO: 34]; and/or

An RuvC-II domain comprising the motif IANPVVMRALTQ [SEQ ID NO: 35], preferably IANPVVMRALTQARKVVNAIIKKYG [SEQ ID NO: 36]; and/or
An RuvC-II domain comprising the motif ELAR [SEQ ID NO: 37], preferably IHIELARE [SEQ ID NO: 38]; and/or

An HNH domain comprising the motif QNGKCAY [SEQ ID NO: 39], preferably IVKFKLWSEQNGKCAY [SEQ ID NO: 40]; and/or
An HNH domain comprising the motif VDHVIP [SEQ ID NO: 41], preferably VDHVIPYSRSLDDSYTNKVL [SEQ ID NO: 42]; and/or

An RuvC-III domain comprising the motif DTRYISRFLAN [SEQ ID NO: 43]; and/or
An RuvC-III domain comprising the motif VYTVNGRITAHLRSRW [SEQ ID NO: 44]; and/or
An RuvC-III domain comprising the motif HHAVDA [SEQ ID NO: 45], preferably HHAVDAAIVA [SEQ ID NO: 46]; and/or

Preferably, the present disclosure provides an isolated Cas protein or polypeptide fragment thereof having an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 77% identity therewith, wherein the Cas protein or polypeptide fragment thereof comprises each of the amino acid motifs [SEQ ID NO: 23] to [SEQ ID NO: 46] in combination.

In another example, the present disclosure provides an isolated Cas protein or polypeptide fragment thereof having an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 77% identity therewith. Preferably the Cas protein or polypeptide is capable of binding, cleavage, marking or modifying at a temperature in the range 20°C and 100°C inclusive. Preferably the Cas protein or polypeptide is capable of said cleavage, binding, marking or modifying at a temperature in the range between 20°C and 70°C, for example 25°C, 55°C, 60°C or 65°C. Preferably the Cas protein or polypeptide is capable of said cleavage, binding, marking or modifying at a temperature in the range between 50°C and 70°C, for example 55°C or 60°C. Preferably the Cas protein or polypeptide is capable of said cleavage, binding,

marking or modifying at a temperature in the range between 30°C and 80°C, at a temperature between 37°C and 78°C, preferably at a temperature above 55°C; more preferably at a temperature between 55°C and 80°C; even more preferably at a temperature between 55°C and 65°C or 60°C and 65°C.

The present disclosure also provides uses of a targeting RNA molecule and a Cas protein or polypeptide provided herein, for binding, cleaving, marking or modifying a target polynucleotide comprising a target nucleic acid sequence. The targeting RNA molecule recognizes the target nucleic acid sequence on a target nucleic acid strand of the polynucleotide.

The target polynucleotide that comprises the target nucleic acid sequence may be double stranded and so comprise a target nucleic acid strand, comprising said target nucleic acid sequence, and a non-target nucleic acid strand, comprising a protospacer nucleic acid sequence. The protospacer nucleic acid sequence is substantially complementary to the target nucleic acid sequence and pairs with it in the double stranded target polynucleotide. The non-target nucleic acid strand may further comprise a protospacer adjacent motif (PAM) sequence directly adjacent the 3' end of the protospacer sequence. The PAM sequence may be at least 6, 7, or 8 nucleic acids in length. Preferably, the PAM sequence has a cytosine in the fifth position. Preferably the PAM sequence comprises the sequence 5'-NNNNC-3', so that from the 5'-end the PAM sequence begins 5'-NNNNC-3'. Additionally or alternatively, the PAM sequence may have an adenine in the eighth position, so that the PAM sequence comprises the sequence 5'-NNNNNNNA-3', and from the 5'-end the PAM sequence begins 5'-NNNNNNNA-3'. Additionally or alternatively, the PAM sequence may have a cytosine in one or more of the first, second, third, fourth, and sixth positions, such that from the 5'-end the PAM sequence begins 5'-CNNNN-3', 5'-NCNNN-3', 5'-NNCNN-3', 5'-NNNCN-3', and/or 5'-NNNNNC-3'. Optionally the PAM sequence comprises, so that from the 5'-end the PAM sequence begins, 5'-CCCCCCNA-3' [SEQ ID NO: 10], and further preferably the PAM sequence comprises, so that from the 5'-end the PAM sequence begins, 5'-CCCCCCAA-3' [SEQ ID NO: 11]. Other preferred PAM sequences include 5'-ATCCCCAA-3' [SEQ ID NO: 21] and 5'-ACGGCCAA-3' [SEQ ID NO: 22].

Preferably, the Cas protein or polypeptide is capable of the binding, cleaving, marking or modifying at a temperature in the range 40°C to 80°C inclusive, preferably in the range 45°C to 80°C inclusive, and further preferably in the range 50°C to 80°C inclusive. For example, the binding, cleaving, marking or modifying occurs at a temperature of 45°C, 46°C, 47°C, 48°C, 49°C, 50°C, 51°C, 52°C, 53°C, 54°C, 55°C, 56°C, 57°C, 58°C, 59°C, 60°C, 61°C, 62°C,

63°C, 64°C, 65°C, 66°C, 67°C, 68°C, 69°C, 70°C, 71°C, 72°C, 73°C, 74°C, 75°C, 76°C, 77°C, 78°C, 79°C or 80°C. More preferably the Cas protein or polypeptide is capable of the binding, cleaving, marking or modifying at a temperature in the range 55 to 65°C. In preferred examples, a Cas protein or polypeptide fragment of the disclosure may comprises an amino acid sequence of at least 75% identity; preferably at least 85%; more preferably at least 90%; even more preferably at least 95% identity to SEQ ID NO: 1.

The Cas protein or polypeptide may be used in combination with a targeting RNA molecule that recognizes a target nucleic acid sequence on the target nucleic acid strand, where the non-target nucleic acid sequence has a PAM sequence directly adjacent the 3' end of the protospacer sequence on the non-target strand, as disclosed herein. Thus, the PAM sequence may comprise the sequence 5'-NNNNC-3', and the Cas protein may bind, cleave, mark or modify the target strand at a temperature in the range 20°C and 100°C inclusive, preferably in the range 30°C and 90°C inclusive, in the range 37°C and 78°C inclusive, in the range 40°C and 80°C inclusive, in the range 50°C and 70°C inclusive, or in the range 55°C and 65°C, inclusive. Preferably from the 5'-end the PAM sequence begins 5'-NNNNC-3' and the Cas protein may bind, cleave, mark or modify the target strand at a temperature in the range 20°C and 100°C inclusive, preferably in the range 30°C and 90°C inclusive, in the range 37°C and 78°C inclusive, in the range 40°C and 80°C inclusive, in the range 50°C and 70°C inclusive, or in the range 55°C and 65°C, inclusive. Preferably from the 5'-end the PAM sequence begins 5'-NNNNNNNA-3' and the Cas protein may bind, cleave, mark or modify the target strand at a temperature in the range 20°C and 100°C inclusive, preferably in the range 30°C and 90°C inclusive, in the range 37°C and 78°C inclusive, in the range 40°C and 80°C inclusive, the range 50°C and 70°C inclusive, or in the range 55°C and 65°C, inclusive. Further preferably the 5'-end of the PAM sequence begins 5'-NNNNCNNA-3' [SEQ ID NO: 47] and the Cas protein may bind, cleave, mark or modify the target strand at a temperature in the range 20°C and 100°C inclusive, preferably in the range 30°C and 90°C inclusive, in the range 37°C and 78°C inclusive, in the range 40°C and 80°C inclusive, in the range 50°C and 70°C inclusive, or in the range 55°C and 65°C, inclusive.

More particularly, a Cas protein or polypeptide of the disclosure may comprise an amino acid sequence with a percentage identity with SEQ ID NO:1 as follows: at least 60%, at least 61%, at least 62%, at least 63%, at least 64%, at least 65%, at least 66%, at least 67%, at least 68%, at least 69%, at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at

least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5% or at least 99.8%. The percentage identity may be at least 89%. The percentage identity may be at least 90%. Preferably the percentage identity will be at least 95%, for example 98%.

The percentage amino acid sequence identity with SEQ ID NO: 1 is determinable as a function of the number of identical positions shared by the sequences in a selected comparison window, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

A Cas protein or polypeptide fragment of the disclosure may be characterised in terms of both the reference sequence SEQ ID NO: 1 and any aforementioned percentage variant thereof as defined by percentage sequence identity, alone or in combination with any of the aforementioned amino acid motifs (i.e. SEQ ID NOS 2 and/or 3 and/or 4 and/or 5 and/or 6) as essential features.

The disclosure provides a use of a targeting RNA molecule as provided herein and a Cas protein or polypeptide of the disclosure for binding, cleaving, marking or modifying a target nucleic acid strand comprising a target nucleic acid sequence. Preferably said binding, cleaving, marking or modifying occurs at a temperature disclosed herein, for example at a temperature of between 20 and 100°C. The disclosure also provides a method of binding, cleaving, marking or modifying a target nucleic acid sequence in a target nucleic acid strand comprising designing a targeting RNA molecule as provided herein and forming a ribonucleoprotein complex comprising the targeting RNA molecule and a Cas protein or polypeptide of the disclosure. Preferably the ribonucleoprotein complex binding, cleaving, marking or modifying the target nucleic acid sequence at a temperature disclosed herein, for example at a temperature of between 37 and 100°C.

The uses and methods of the disclosure may be carried out, and the nucleoproteins of the disclosure formed and used, *in vivo*, for example in bacterial cells. The uses and methods of the disclosure may be carried out, and the nucleoproteins of the disclosure formed and used, *in vivo*, except in human cells. Alternatively the uses and methods of the disclosure may be carried out, and the nucleoproteins of the disclosure formed and used, *in vitro*. The Cas protein of the disclosure may be provided in isolated form, for example when used *in vitro* or when added to cells by transfection, the Cas protein may be heterologously expressed, for example following transient or stable transformation of the cell by nucleic acid encoding the Cas protein, the targeting RNA molecule may be transcribed from an expression vector following transient or stable transformation of the cell by nucleic acid encoding the RNA molecule, and/or the

RNA molecule may be provided in isolated form, for example when used *in vitro* or when added to cells by transfection. In preferred examples, the Cas protein or polypeptide is expressed from the genome of a host cell, following stable intergration of a nucleic acid encoding the Cas protein or polypeptide in the genome of the host cell. Thus the Cas protein and/or RNA molecule may be added to the *in vivo* or *in vitro* environment using any artificial or contrived method for adding a protein or nucleic acid molecule to a cell in which it is not otherwise present.

The polynucleotide comprising the target nucleic acid sequence may be cleaved by the Cas protein, and optionally the cleavage may be DNA cleavage. The target nucleic acid strand comprising the target sequence may be double stranded DNA and the method or use may result in a double stranded break in the polynucleotide comprising the target nucleic acid sequence. The polynucleotide comprising the target nucleic acid sequence may be double stranded DNA, the Cas protein may lack the ability to cut the double stranded DNA and the use or method may result in gene silencing of the polynucleotide.

The Cas protein or polypeptide may be provided for the methods, uses and nucleoproteins of the disclosure at a concentration of 250 nM or less, for example at a concentration of 200 nM or less, 150 nM or less, 100 nM or less, 50 nM or less, 25nM or less, 10 nM or less, 5nM or less, 1nM or less or 0.5 nM or less. Alternatively, the Cas protein or polypeptide may be provided at a concentration of at least 0.5 nM, at least 1nM, at least 5nM, at least 10nM, at least 25nM, at least 50 nM, at least 100 nM, at least 150 nM, at least 200 nM, or at least 250 nM. The PAM sequence of the disclosure may have an adenine in the eighth position, so that the PAM sequence comprises the sequence 5'-NNNNNNNA-3', and the concentration of Cas protein or polypeptide may be 100 nM or less, 50 nM or less, 25nM or less, 10 nM or less, 5nM or less, 1nM or less or 0.5 nM or less. The PAM sequence may comprise the sequence 5'-NNNNCNNA-3' [SEQ ID NO: 47], and the concentration of Cas protein or polypeptide may be 100 nM or less, 50 nM or less, 25nM or less, 10 nM or less, 5nM or less, 1nM or less or 0.5 nM or less. The PAM sequence may comprise the sequence 5'-CCCCCCNA-3' [SEQ ID NO: 10], and the concentration of Cas protein or polypeptide may be 100 nM or less, 50 nM or less, 25nM or less, 10 nM or less, 5nM or less, 1nM or less or 0.5 nM or less.

Also, the disclosure provides nucleic acids encoding any of the aforementioned proteins or polypeptides of the disclosure. The nucleic acids may be isolated or in the form of expression constructs.

In all aforementioned examples of the present disclosure, amino acid residues may be substituted conservatively or non-conservatively. Conservative amino acid substitutions refer to those where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not alter the functional properties of the resulting polypeptide.

Similarly it will be appreciated by a person of average skill in the art that nucleic acid sequences may be substituted conservatively or non-conservatively without affecting the function of the polypeptide. Conservatively modified nucleic acids are those substituted for nucleic acids which encode identical or functionally identical variants of the amino acid sequences. It will be appreciated by the skilled reader that each codon in a nucleic acid (except AUG and UGG; typically the only codons for methionine or tryptophan, respectively) can be modified to yield a functionally identical molecule. Accordingly, each silent variation (i.e. synonymous codon) of a polynucleotide or polypeptide, which encodes a polypeptide of the present disclosure, is implicit in each described polypeptide sequence.

The disclosure provides a transformed cell, having a target nucleic acid sequence in a double stranded target polynucleotide, said cell comprising a Cas protein or polypeptide as provided herein and at least one targeting RNA molecule as provided herein, and an expression vector comprising a nucleic acid encoding at least one of said Cas protein and said targeting RNA molecule. The Cas protein and targeting RNA molecule may enable or permit binding, cleaving, marking or modifying of the target sequence to occur in the transformed cell at a raised temperature, or at a range of temperatures, for example between 37 and 100°C, as disclosed herein. The disclosure further provides a method of binding, cleaving, marking or modifying a target nucleic acid in a cell comprising either 1) transforming, transfecting or transducing the cell with an expression vector comprising a nucleotide sequence encoding a Cas protein or polypeptide of the disclosure and a nucleotide sequence encoding a targeting RNA molecule of the disclosure; or 2) transforming, transfecting or transducing the cell with an expression vector comprising a nucleotide sequence encoding a Cas protein or polypeptide of the disclosure and a further expression vector comprising a nucleotide sequence encoding a targeting RNA molecule of the disclosure; or 3) transforming, transfecting or transducing the cell with an expression vector comprising a nucleotide sequence encoding a Cas protein or polypeptide of the disclosure, and delivering a targeting RNA molecule as provided herein to, or into the cell. The Cas protein or polypeptide may be expressed from the genome of the transformed cell, for example following stable integration into the genome of a nucleotide sequence encoding the Cas protein or polypeptide.

The disclosure also provides kits comprising one or more of the reagents for carrying out the uses and methods of the disclosure, or for generating the transformed cells or nucleoprotein complex of the disclosure, said kits including: a Cas protein or polypeptide of the disclosure or an expression vector comprising a nucleic acid sequence encoding a Cas protein or polypeptide of the disclosure; and/or a targeting RNA molecule of the disclosure or an expression vector comprising a nucleic acid sequence encoding a targeting RNA molecule of the disclosure. The kits may further include instructions for carrying out the disclosure, for example instructions for how to design a targeting RNA molecule in accordance with the disclosure.

RNA Guides and Target Sequences

Cas proteins of the invention allow for sequence-specific binding, cleavage, tagging, marking or modification of target nucleic acids at elevated temperatures. Target nucleic acids may be DNA (single-stranded or double-stranded), RNA or synthetic nucleic acids. A particularly useful application of the present invention is the sequence-specific targeting and modification of genomic DNA by one or more Cas proteins of the invention in complex with one or more guide RNAs (gRNAs) that complementarily bind to a targeted sequence of the genomic DNA. Consequently, the target nucleic acid is preferably double-stranded DNA. Such targeting may be performed *in vitro* or *in vivo*. Preferably such targeting is performed *in vivo*. In this way, Cas proteins of the invention may be used to target and modify specific DNA sequences located in the genomic DNA of a cell. It is envisaged that the Cas system may be used to modify genomes in a variety of cell types of and/or in different organisms.

The gRNAs, also called targeting RNA molecules, recognize the target nucleic acid sequence on the polynucleotide target strand. The RNA molecules may be designed to recognize a target sequence in a double stranded target polynucleotide, wherein the non-target strand comprises a protospacer adjacent motif (PAM) sequence directly adjacent the 3' end of the protospacer sequence. Disclosed herein are PAM sequences that work in an optimal manner with the Cas proteins and polypeptides of the invention. With knowledge of these PAM sequences, gRNAs may be designed for use with the Cas proteins and polypeptides of the invention across the temperature ranges and increased temperatures of the invention.

Accordingly, the present invention provides a ribonucleoprotein complex comprising a Cas protein or a polypeptide of the invention as hereinbefore described, and further comprising at least one RNA molecule which has a targeting function in that it recognizes a particular nucleotide sequence in a target polynucleotide. The present invention also provides use of at

CLAIMS

1. A method of modifying the genetic material of a eukaryotic cell, comprising (i) integrating a polynucleotide encoding a ThermoCas9 under the control of a first promoter into the genome of the cell, wherein the expressed ThermoCas9 comprises an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 77% identity therewith, or an active fragment thereof; (ii) transforming the cell with an expression vector comprising a polynucleotide sequence encoding a guide RNA and under the control of a second promoter, wherein the guide RNA has a nucleic acid sequence which recognizes a nucleic acid sequence comprised at a desired target locus in the genome of the cell, and (iii) transforming the cell with a repair oligonucleotide.

2. A method of modifying the genetic material of a eukaryotic cell, comprising (i) integrating a polynucleotide encoding a ThermoCas9 under the control of a first promoter into the genome of the cell, wherein the expressed ThermoCas9 comprises an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 77% identity therewith, or an active fragment thereof; and (ii) transforming the cell with an expression vector comprising a polynucleotide sequence encoding a guide RNA under the control of the first or of a separate second promoter, wherein the guide RNA has a nucleic acid sequence which recognizes a nucleic acid sequence comprised at a desired target locus in the genome of the cell, and a repair oligonucleotide also under the control of the first or the second promoter, or a separate third promoter.

3. A method as claimed in claim 1 or claim 2, wherein the repair oligonucleotide is a double-stranded DNA repair oligo; optionally comprising a polynucleotide sequence for insertion into the genome of the cell by way of homologous recombination following guide RNA directed ThermoCas9 endonuclease cutting; preferably wherein:

   (a) the first promoter is a constitutive promoter; and/or

   (b) the first promoter is an inducible promoter; and/or

   (c) the second promoter is a constitutive or inducible promoter; and/or

   (d) wherein the third promoter is a constitutive or inducible promoter.

4. A method as claimed in claim 1 or claim 2, wherein the cell is transformed with the expression plasmid(s) and/or the repair-oligo by heat-shock; and/or wherein the cell is transformed and/or grown following transformation at a temperature in the range 26 °C -

60 °C; preferably 31°C – 60 °C; more preferably 35 °C – 60 °C; even more preferably 34 °C – 41 °C, e.g. 37 °C.

5. A polynucleotide expression vector for modifying the genetic material at a target locus of a host organism which comprises the expression vector, the vector comprising:

    a) a polynucleotide sequence encoding a ThermoCas9 nuclease, wherein the ThermoCas9 nuclease comprises an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 77% identity therewith, or an active fragment thereof;
    b) a polynucleotide sequence encoding a guide RNA, wherein the guide RNA has a nucleic acid sequence which recognizes a nucleic acid sequence comprised in the target locus;
    c) a first promoter oriented with respect to polynucleotide sequences of (a) and (b) so as to drive expression thereof in the organism.

6. A vector as claimed in claim 5, wherein the sequence of (a) is 3' of the promoter and the sequence of (b) is 3' of the sequence of (b).

7. A vector as claimed in claim 5 or claim 6, wherein the first promoter is an inducible promoter; optionally wherein the inducible promoter is selected from a β-glucosidase promoter inducible with cellobiose or a Pm promoter inducible with 3-methylbenzoate.

8. A vector as claimed in any preceding claim, further comprising a polynucleotide sequence which encodes an homologous recombination (HR) fragment under the control of the first promoter or a separate second promoter.

9. A vector as claimed in claim 8, wherein the second promoter controlling the HR fragment is a constitutive promoter; optionally wherein the constitutive promoter is P3.

10. A vector as claimed in claim 8 or claim 9, wherein the arms of the HR fragment comprise a nucleic acid sequence which allowing recombination upstream and downstream, respectively, of a locus of interest in the host organism; preferably wherein the locus of interest comprises the target sequence; more preferably the locus of interest is a gene.

11. A vector as claimed in any of claims 8 to 10, wherein the HR fragment also comprises an insertion element between the upstream and downstream arms thereof; preferably wherein the insertion element is a gene of interest; optionally with a suitable promoter in

operational orientation to provide for expression of the gene of interest in the host organism.

12. A vector as claimed in claim 10 or claim 11, wherein the locus of interest comprises the PAM sequence 5'-NNNNCNN-3' located 3' of the target sequence; optionally at least 2, 3, 4, 5, 6 or more nucleotides from the target sequence.

13. A vector as claimed in any preceding claim, wherein the guide RNA is a single guide RNA (sgRNA).

14. An *in vitro* method of modifying the genetic material of an organism comprising transforming the organism with a first expression vector being a vector of any of claims 5 to 7, and a second expression vector comprising a polynucleotide sequence which encodes an homologous recombination (HR) fragment under the control of a promoter.

15. An *in vitro* method of modifying the genetic material of a prokaryotic organism comprising transforming the organism with an expression vector of any of claims 8 to 13.

16. A method as claimed in claim 14 or claim 15, wherein the transformed organism is cultured for a period at a first temperature, and then cultured at a second temperature before or during induction of the promoter of the ThermoCas9 sequence; preferably wherein the first temperature is 60 °C or less and the second temperature is a higher temperature which is at least 55 °C; more preferably wherein the first temperature is 55 °C or less and the second temperature is greater than 55 °C.

17. A method as claimed in any of claims 14 to 16, wherein the organism is a prokaryote; preferably wherein the prokaryote is:

    (a) a thermophilic bacterium, e.g. *Geobacillus thermoglucosidans*; or
    (b) a bacterium with growth temperature optimum of at least 40 °C; preferably at least 45 °C., e.g. *Bacillus coagulans* or *Pseudomonas putida*; or
    (c) a soil bacterium; preferably a saprotrophic soil bacterium, e.g. *Pseudomonas putida.*

18. A prokaryotic cell transformed with an expression vector of any of claims 5 to 13.