



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년04월19일
(11) 등록번호 10-2523543
(24) 등록일자 2023년04월14일

- (51) 국제특허분류(Int. Cl.)
C12N 15/63 (2006.01) C12N 15/113 (2010.01)
C12N 9/22 (2006.01)
- (52) CPC특허분류
C12N 15/63 (2013.01)
C12N 15/113 (2013.01)
- (21) 출원번호 10-2019-7020536
- (22) 출원일자(국제) 2017년08월16일
심사청구일자 2020년08월14일
- (85) 번역문제출일자 2019년07월15일
- (65) 공개번호 10-2019-0104344
- (43) 공개일자 2019년09월09일
- (86) 국제출원번호 PCT/EP2017/070797
- (87) 국제공개번호 WO 2018/108339
국제공개일자 2018년06월21일
- (30) 우선권주장
PCT/EP2016/081077 2016년12월14일
유럽특허청(EPO)(EP)
- (56) 선행기술조사문헌
WO2016179038 A1
BioRxiv. May 2017, XP55401514.*
*는 심사관에 의하여 인용된 문헌
- (73) 특허권자
바게닝겐 유니버시티
네덜란드, 엔엘-6708 피비 바게닝겐, 드로브달세
스티그 4
스티칭 부어 드 테크니셰 웨텐샤펜
네덜란드, 엔엘-3527 제이피 우트레흐, 반 볼렌호
벤란 661
- (72) 발명자
반 더 우스트, 존
네덜란드, 6871 에이치체트 렌컴, 브람 스트리프
랜드베그 116
반 크라넨부르크, 리차드
네덜란드, 4206 에이씨 고린캠, 아르켈세디스크
46
(뒷면에 계속)
- (74) 대리인
청운특허법인

전체 청구항 수 : 총 18 항

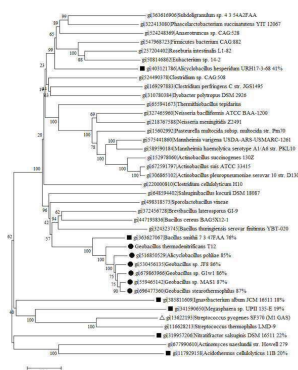
심사관 : 이현지

(54) 발명의 명칭 열안정성 CAS9 뉴클레아제

(57) 요약

열안정성 Cas9 뉴클레아제. 본 발명은 유전 공학의 분야에 관한 것으로, 좀 더 구체적으로, 핵산 편집 및 게놈 변형에 관한 것이다. 본 발명은, SEQ ID NO: 1의 아미노산 서열 또는 이와 적어도 77% 동일성의 서열을 갖는 단 리된 Cas 단백질 또는 이의 폴리펩티드 단편을 제공한다. 상기 Cas 단백질 또는 폴리펩티드는, 20℃ 내지 100℃ 범위의 온도에서 이중 가닥 표적 폴리뉴클레오티드를 결합, 절단, 표지 또는 변형시킬 수 있다. 본 발명은 상기 Cas9 뉴클레아제를 인코딩하는 단리된 핵산 분자, 발현 벡터 및 숙주 세포를 더욱 제공한다. 본 발명은 또한 Cas 단백질 또는 폴리펩티드에 의해 인식되는 PAM 서열을 제공한다. 여기에 개시된 Cas9 뉴클레아제는, 일반적으로 유전 공학을 위한, 특히 상승된 온도에서, 새로운 도구를 제공한다.

대표도 - 도1



(52) CPC특허분류

C12N 9/22 (2013.01)

C12N 2310/20 (2017.05)

C12N 2310/531 (2013.01)

C12N 2330/51 (2013.01)

C12N 2750/14143 (2013.01)

(72) 발명자

보스마, 엘리케 페나

덴마크, 2000 프레데릭스베르, 3.티비, 헬게스베르
28

모우기아코스, 이오아니스

네덜란드, 6706 제이이 바게닝겐, 돌더스트라트 4
8디

모한라주, 프라사나

네덜란드, 6701 디피 바게닝겐, 보우레스파크 12

명세서

청구범위

청구항 1

표적 핵산 서열을 포함하는 이중 가닥 표적 폴리뉴클레오티드를 갖는 시험관 내 형질전환된 인간 세포로서, 여기서, 상기 이중 가닥 표적 폴리뉴클레오티드는, 상기 표적 핵산 서열을 포함하는 표적 핵산 가닥, 및 상기 표적 핵산 서열에 상보적인 프로토스페이서 핵산 서열을 포함하는 비-표적 핵산 가닥을 포함하며, 상기 세포는:

SEQ ID NO: 1의 아미노산 서열을 갖는, 일정한 간격을 두고 주기적으로 분포하는 짧은 회문구조 반복서열 (CRISPR)-관련 (Cas) 단백질;

상기 표적 핵산 가닥 내에 표적 핵산 서열을 인식하는 적어도 하나의 표적화 RNA 분자로서, 여기서, 비-표적 가닥은, 프로토스페이서 서열의 3' 말단에 바로 인접한 프로토스페이서 인근 모티프 (PAM) 서열을 더욱 포함하며, 여기서, 상기 PAM 서열은 5'-NNNNCNN-3'을 포함하는, 적어도 하나의 표적화 RNA 분자; 및

상기 Cas 단백질 및 상기 표적화 RNA 분자 중 적어도 하나를 인코딩하는 핵산을 포함하는 발현 벡터를 포함하고; 여기서, 인간 세포는 배아 줄기 세포가 아닌, 시험관 내 형질전환된 세포.

청구항 2

청구항 1에 있어서,

상기 Cas 단백질은 발현 벡터로부터 발현되는, 시험관 내 형질전환된 세포.

청구항 3

Cas 단백질, 이중 가닥 표적 폴리뉴클레오티드에서 표적 핵산 서열을 인식하는 적어도 하나의 표적화 RNA 분자, 및 표적 폴리뉴클레오티드를 포함하는 핵산단백질 복합체에 있어서,

상기 Cas 단백질은, SEQ ID NO: 1의 아미노산 서열을 가지며;

상기 이중 가닥 표적 폴리뉴클레오티드는, 상기 표적 핵산 서열을 포함하는, 표적 핵산 가닥, 및 상기 표적 핵산 서열에 상보적인 프로토스페이서 핵산 서열 및 프로토스페이서 서열의 3' 말단에 바로 인접한 프로토스페이서 인근 모티프 (PAM) 서열을 포함하는, 비-표적 핵산 가닥을 포함하고; 여기서, 상기 PAM 서열은 5'-NNNNCNN-3'을 포함하며, 및 상기 핵산단백질 복합체는 시험관 내 인간 세포 내에 있고; 및 여기서, 상기 세포는 배아 줄기 세포가 아닌, 핵산단백질 복합체.

청구항 4

청구항 3에 따른 핵산단백질 복합체를 포함하는 시험관 내 형질전환된 세포.

청구항 5

청구항 1 또는 4에 있어서,

상기 Cas 단백질 및 표적화 RNA 분자에 의한 표적 폴리뉴클레오티드의 결합, 절단, 표지, 또는 변형은 20℃ 내지 100℃의 온도, 30℃ 내지 80℃의 온도, 37℃ 내지 78℃의 온도, 55℃ 이상의 온도, 55℃ 내지 80℃의 온도, 55℃ 내지 65℃의 온도, 또는 60℃ 내지 65℃의 온도에서 발생하는, 시험관 내 형질전환된 세포.

청구항 6

청구항 1 또는 4에 있어서,

상기 표적 핵산 서열을 포함하는 이중 가닥 표적 폴리뉴클레오티드는, Cas 단백질에 의해 절단되고, 상기 절단은 DNA 절단인, 시험관 내 형질전환된 세포.

청구항 7

청구항 1 또는 4에 있어서,

상기 표적 서열을 포함하는 표적 폴리뉴클레오티드는 이중 가닥 DNA이고,

Cas 단백질 및 표적화 RNA 분자에 의한 표적 DNA 서열의 결합, 절단, 표지, 또는 변형은 표적 폴리뉴클레오티드에서 이중 가닥 절단을 결과하는, 시험관 내 형질전환된 세포.

청구항 8

청구항 1 또는 4에 있어서,

상기 Cas 단백질은 이중 가닥 DNA를 절단하는 뉴클레아제 활성이 없으나, 표적 폴리뉴클레오티드의 유전자 침묵이 존재하는, 시험관 내 형질전환된 세포.

청구항 9

청구항 1 또는 4에 있어서,

상기 PAM 서열은, 5'-NNNNCNNA-3' [SEQ ID NO: 47]을 포함하거나; 상기 PAM 서열은, 5'-NNNNCSAA-3' [SEQ ID NO: 48]을 포함하거나; 또는 상기 PAM 서열은, 5'-NNNNCCAA-3' [SEQ ID NO: 50]을 포함하는, 시험관 내 형질전환된 세포.

청구항 10

청구항 1 또는 4에 있어서,

Cas 단백질 및 표적화 RNA 분자에 의한 표적 DNA 서열의 결합, 절단, 표지, 또는 변형은 20℃ 내지 70℃의 온도; 또는 25℃ 내지 65℃의 온도에서 발생하는, 시험관 내 형질전환된 세포.

청구항 11

청구항 1 또는 4에 있어서,

상기 Cas 단백질은 박테리아, 고세균, 바이러스, 호열성 박테리아, 지오바실러스 속, 또는 지오바실러스 씨모테니트리피칸스로부터 얻을 수 있는, 시험관 내 형질전환된 세포.

청구항 12

청구항 1 또는 4에 있어서,

상기 표적화 RNA 분자는 crRNA 및 tracrRNA를 포함하는, 시험관 내 형질전환된 세포.

청구항 13

청구항 1 또는 4에 있어서,

상기 적어도 하나의 표적화 RNA 분자의 길이는 35-200 뉴클레오티드 잔기의 범위이며; 또는 상기 표적 핵산 서열은 길이가 15 내지 32 뉴클레오티드 잔기인, 시험관 내 형질전환된 세포.

청구항 14

청구항 1 또는 4에 있어서,

상기 Cas 단백질은 적어도 하나의 기능성 모이어티 또는 단백질을 더욱 포함하는, 시험관 내 형질전환된 세포.

청구항 15

청구항 14에 있어서,

상기 Cas 단백질의 뉴클레아제 활성은 비활성인, 시험관 내 형질전환된 세포.

청구항 16

청구항 14에 있어서,

상기 단백질은 헬리카제, 뉴클레아제, 헬리카제-뉴클레아제, DNA 메틸라아제, 히스톤 메틸라아제, 아세틸라제, 포스파타아제, 키나아제, 전사 (공-)활성인자, 전사 억제인자, DNA 결합 단백질, DNA 구조 단백질, 마커 단백질, 리포터 단백질, 형광 단백질, 리간드 결합 단백질, 신호 펩티드, 세포내 위치정보 서열, 항체 에피토프 또는 친화성 정제 태그로부터 선택된 단백질인, 시험관 내 형질전환된 세포.

청구항 17

표적 핵산 서열을 포함하는 표적 핵산 가닥, 및 상기 표적 핵산 서열에 상보적인 프로토스페이서 핵산 서열을 포함하는 비-표적 핵산 가닥을 포함하는, 시험관 내 인간 세포에서 이중 가닥 표적 폴리뉴클레오티드를 결합, 절단, 표지, 또는 변형시키는 방법으로서, 상기 방법은:

- 적어도 하나의 표적화 RNA 분자를 디자인하는 단계로서, 여기서, 상기 표적화 RNA 분자는 표적 가닥에서 표적 서열을 인식하고, 상기 비-표적 가닥은, 프로토스페이서 핵산 서열의 3' 말단에 바로 인접한 프로토스페이서 인근 모티프 (PAM) 서열을 더욱 포함하며, 여기서, 상기 PAM 서열은 5'-NNNNCNN-3'을 포함하는, 디자인 단계;
- 상기 표적화 RNA 분자 및 Cas 단백질을 포함하는 리보핵산 단백질 복합체를 형성하는 단계로서, 여기서, 단 리된 Cas 단백질은 SEQ ID NO: 1의 아미노산 서열을 갖는, 형성 단계; 및
- 상기 리보핵산 단백질 복합체를 사용하여 상기 표적 폴리뉴클레오티드를 결합, 절단, 표지, 또는 변형시키는 단계를 포함하며, 여기서, 상기 인간 세포는 배아 줄기 세포가 아니며; 및 여기서, 상기 방법은 인간의 생식 계열 유전적 동일성을 변형시키지 않는, 이중 가닥 표적 폴리뉴클레오티드를 결합, 절단, 표지, 또는 변형시키는 방법.

청구항 18

표적 핵산 서열을 포함하는 이중 가닥 표적 폴리뉴클레오티드를 결합, 절단, 표지 또는 변형시키기 위한 적어도 하나의 표적화 RNA 분자 및 Cas 단백질의 사용 방법으로서, 여기서:

상기 이중 가닥 표적 폴리뉴클레오티드는, 상기 표적 핵산 서열을 포함하는, 표적 핵산 가닥, 및 상기 표적 핵산 서열에 상보적인 프로토스페이서 핵산 서열을 포함하는, 비-표적 핵산 가닥을 포함하고;

상기 Cas 단백질은, SEQ ID NO: 1의 아미노산 서열을 가지며;

상기 적어도 하나의 표적화 RNA 분자는 표적 서열을 인식하고;

상기 비-표적 핵산 가닥은, 프로토스페이서 핵산 서열의 3' 말단에 바로 인접한 프로토스페이서 인근 모티프 (PAM) 서열을 더욱 포함하며, 여기서, 상기 PAM 서열은 5'-NNNNCNN-3'을 포함하고; 여기서, 상기 사용은 시험관 내 인간 세포에 사용이고, 및 여기서, 상기 인간 세포는 배아 줄기 세포가 아닌, 적어도 하나의 표적화 RNA 분자 및 Cas 단백질의 사용 방법.

청구항 19

삭제

청구항 20

삭제

청구항 21

삭제

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

청구항 36

삭제

청구항 37

삭제

청구항 38

삭제

청구항 39

삭제

청구항 40

삭제

청구항 41

삭제

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

청구항 60

삭제

청구항 61

삭제

청구항 62

삭제

청구항 63

삭제

청구항 64

삭제

청구항 65

삭제

청구항 66

삭제

청구항 67

삭제

청구항 68

삭제

청구항 69

삭제

청구항 70

삭제

청구항 71

삭제

청구항 72

삭제

청구항 73

삭제

청구항 74

삭제

청구항 75

삭제

청구항 76

삭제

청구항 77

삭제

청구항 78

삭제

청구항 79

삭제

청구항 80

삭제

청구항 81

삭제

청구항 82

삭제

청구항 83

삭제

청구항 84

삭제

청구항 85

삭제

청구항 86

삭제

청구항 87

삭제

청구항 88

삭제

청구항 89

삭제

청구항 90

삭제

청구항 91

삭제

청구항 92

삭제

청구항 93

삭제

청구항 94

삭제

청구항 95

삭제

청구항 96

삭제

청구항 97

삭제

청구항 98

삭제

청구항 99

삭제

청구항 100

삭제

발명의 설명

기술 분야

[0001]

본 발명은 유전 공학 분야에 관한 것으로, 특히, 핵산 편집 (nucleic acid editing) 및 게놈 변형 (genome modification)에 관한 것이다. 본 발명은, 유전 물질의 서열-인도 부위-특이적 결합 (sequence-directed site-specific binding), 닉킹 (nicking), 절단 및 변형을 위해 구성될 수 있는 뉴클레아제 (nucleases); 또한, 유전 물질의 서열 특이적 부위에서 활성 (activity), 특히 뉴클레아제 활성을 작용하게 하는 리보핵산 단백질 (ribonucleoproteins), 및 마커 (markers)로서 사용하기 위한 변형된 뉴클레아제 및 리보핵산 단백질의 형태의 유전 공학 도구 (genetic engineering tools)에 관한 것이다. 따라서, 본 발명은 또한 세포 내에 가이드 RNAs (guide RNAs) 및 뉴클레아제의 전달 및 발현을 위한 관련 발현 구축물 (expression constructs)과 관련된다. 더욱이, 본 발명은 시험관 내 또는 생체 내에서 핵산의 서열-특이적 편집 및 이를 달성하는데 사용되는 방법과

관련된다.

배경 기술

- [0002] 2007년에, CRISPR-Cas가 많은 박테리아 및 대부분의 고세균류 (archaea)에서 적응 면역계 (adaptive immune system)인 것은 처음 입증되었다 (Barrangou et al., 2007, Science 315: 1709-1712, Brouns et al., 2008, Science 321: 960-964). 기능적 및 구조적 기준에 기초하여, 3 타입을 각각 포함하는 2 부류의 CRISPR-Cas 시스템은 지금까지 특징화되었으며 (characterized), 이들의 대부분은, 상보적인 DNA 서열을 표적으로 하기 위해 가이드로서 작은 RNA 분자를 사용한다 (Makarova et al., 2015, Nat Rev Microbiol 13: 722-736; Mahanraju et al., 2016, Science 353: aad5147).
- [0003] Doudna/Chargementier 실험실에 의한 최근의 연구에서, (특이적 스페이서 서열을 갖는) 디자인된 CRISPR RNA 가이드의 도입이, 플라스미드 (plasmids) 상에 상보적인 서열 (프로토스페이서 (protospacers))을 표적으로 삼아, 이 플라스미드의 이중 가닥 절단 (double strand breaks)을 유발하는 것의 입증을 포함하여, 부류 2/타입 II CRISPR-Cas 시스템 (Cas9)의 이펙터 효소 (effector enzyme)의 철저한 특징화 (thorough characterization)는, 수행되었다 (Jinek et al., 2012, Science 337: 816-821). Jinek et al., 2012 후에, Cas9는 게놈 편집을 위한 도구로 사용된다.
- [0004] Cas9는 다양한 진핵세포 (예를 들어, 어류, 식물, 인간)의 게놈을 조작하는데 사용되어 왔다 (Chargementier and Doudna, 2013, Nature 495: 50-51).
- [0005] 부가적으로, Cas9는, 전용 재조합 이벤트 (dedicated recombination events) 동안 선택에 의해 박테리아에서 상동 재조합 (homologous recombination)의 수율을 개선하는데 사용되어 왔다 (Jiang et al., 2013, Nature Biotechnol 31: 233-239). 이를 달성하기 위해, 독성 단편 (toxic fragment) (표적화 구축물 (Targeting construct))은, 원하는 변화 (alteration)를 가지는 구출 단편 (rescuing fragment) (점 돌연변이 (point mutation) 또는 결손 (deletions)을 가지는, 편집 구축물)으로 동시-형질감염된다 (co-transfected). 상기 표적화 구축물은, 숙주 염색체상에 원하는 재조합의 부위를 규정하는, 항생제 내성 마커 및 디자인 CRISPR와 조합한 Cas9로 이루어지고; 상응하는 항생제의 존재하에서, 숙주 염색체에서 표적화 구축물의 인테그레이션 (integration)은 선택된다. 숙주 염색체상에 CRISPR 표적 부위를 갖는 편집 구축물의 부가적인 재조합이 일어날 때에만, 숙주는 자가-면역 문제로부터 벗어날 수 있다. 그러므로, 항생제의 존재하에서, 오직 원하는 (마커-부재) 돌연변이만이 생존하고, 성장할 수 있다. 염색체로부터 인테그레이션된 표적화 구축물의 후속적 제거를 위한 선택에 관련된 전략은, 또한 제시되어, 진정한 마커 부재 돌연변이 (genuine marker free mutant)를 발생시킨다.
- [0006] 최근에, CRISPR-Cas 매개 게놈 편집은, 유전 공학을 위한 유용한 도구를 구성하는 것으로 확립되었다. 원핵생물 CRISPR 시스템은, 이들의 숙주를 적응 면역 시스템으로서 제공하고 (Jinek et al., 2012, Science 337:816-821), 및 신속하고 효과적인 유전 공학을 위해 사용될 수 있어 (예를 들어, Mali et al., 2013, Nat Methods 10:957-963), 관심의 서열을 표적화하기 위해 가이드 서열의 변형만을 필요로 한다는 것이 확립되었다.
- [0007] 그러나, 유전 연구 및 게놈 편집의 분야에 적용을 위해 다양한 실험 조건하에서 개선된 서열-특이적 핵산 검출, 절단 (cleavage) 및 조작이 가능한 작용제 (agents)의 개발에 대한 계속적인 요구가 있다. 특히, Cas9를 포함하는, 현재 이용 가능한 서열-특이적 게놈 편집 도구는, 모든 조건 또는 유기체에서 사용하는데 적용 가능하지 않은데, 예를 들어, 서열-특이적 뉴클레아제는 상대적으로 열-민감성이고, 따라서 절대적 호열성 미생물 (41℃ 내지 122℃에서 성장할 수 있고, >45℃ 내지 80℃의 온도 범위에서 최적으로 성장할 수 있으며, 초호열성은 80℃ 위에서 최적으로 성장할 수 있음), 예를 들어, 시험관 내에서 상승된 온도에서 수행된 실험실 공정 동안 또는 산업적 발효에 사용되는 미생물에 사용하는데 적용 가능하지 않다.
- [0008] 지금까지, 호열성 미생물에서 활성 Cas9 단백질에 대한 실험적 증거는 없다. 박테리아에서 Cas9의 존재하에 Chylinski et al.에 의한 비교 게놈 스크리닝 (comparative genome screening) (2014; Nucleic Acids Research 42: 6091-6105)에 기초하여, 타입 II-C CRISPR-Cas 시스템은, 모든 박테리아 게놈의 대략 3.3%에만 존재하는 것으로 밝혀졌다. 호열성 박테리아 가운데, 타입 II 시스템은, 통계 분석에 근거하여 불충분하게 나타난다 (P=0.0019). 부가적으로, 타입 II 시스템은, 아직 고세균류에서 발견되지 않았지만, 이는 아마도 고세균류에서 (타입 II 시스템과 관련된) RNase III 단백질의 부재에 기인할 수 있다. Chylinski, et al., (2014; Nucleic Acids Research 42: 6091-6105)은, 타입 II CRISPR-Cas 시스템의 분류 및 진화 (evolution)를 기재하고 있으며, 특히, 이들 시스템을 나타내는 두 개의 종들 (species)이 동정되었지만, 이들 종들은 55℃에서 최대로

성장하고, 80℃ 위에서 최적으로 성장할 수 있는 초호열성과 함께, 최적 성장 온도 60-80℃를 갖는 절대적 호열성 성장을 나타내지 않는다.

- [0009] 박테리아 계놈에서 CRISPR-Cas 시스템의 최소성 및, 특히 Cas9가 45℃ 아래에서 최적 성장 온도를 갖는 (고세균류가 아닌) 박테리아에서 단지 발견되었다는 사실에도 불구하고, 본 발명자들은 놀랍게도, 상승된 온도에서 계놈 편집이 수행되는 것을 가능하게 하는 몇 가지 열안정성 Cas9 변이체 (variants)를 발견했다. 본 발명자들은 또한 상승된 온도를 포함하는, 광범위한 온도 범위에 걸쳐 계놈 편집이 수행되는 것을 가능하게 하는 열안정성 Cas9 변이체와 함께 작동하는 최적화된 프로토스페이스 인근 모티프 (protospacer adjacent motif: PAM) 서열을 발견하였다. 이들 Cas9 뉴클레아제, 및 관련 PAM 서열의 지식으로 디자인된 RNA 분자는, 상승된 온도에서 유전 공학을 위한 새로운 도구를 제공하며, 및 호열성 유기체; 특히 미생물의 유전자 조작에서 특별한 가치가 있다.
- [0010] 호열성 속 지오바실러스 (thermophilic genus *Geobacillus*)의 계통학적 재-평가가 최근 수행되어, 새로운 속인, 파라지오바실러스 (*Parageobacillus*)를 결과하였다. 결과적으로, 이전에 지오바실러스 속의 몇몇 종들은, 파라지오바실러스로 계통적으로 재-할당되었고, 따라서 재-명명되었다 (Aliyu *et al.*, (2016) *Systematic and Applied Microbiology* 39:527-533).
- [0011] 일정한 간격을 두고 주기적으로 분포하는 짧은 회문구조 반복서열 (Clustered Regularly Interspaced Short Palindromic Repeats: CRISPR) 및 CRISPR-관련 (Cas) 단백질은, 침입하는 유전 인자에 대해 원핵생물에서 적응성 및 유전적 면역을 제공한다 (Brouns *et al.* *Science* 321, (2008); Barrangou *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, (2007); Wright *et al.* *Cell* 164, 29-44 (2016); Mohanraju *et al.* *Science* 353, aad5147 (2016)). CRISPR-Cas 시스템은, 이들의 복잡성 및 시그니처 단백질 (signature protein)에 의존하여, 두 부류 (1 및 2) 및 6개 타입 (I-VI)으로 세분된다 (Makarova *et al.* *Nat. Rev. Microbiol.* 13, 722-736 (2015)). 타입-II CRISPR-Cas9 및 타입 V CRISPR-Cas12a (이전에 CRISPR-Cpf1이라 함)를 포함하는, 부류 2 시스템은, 최근에 두 진핵생물에 대한 계놈 공학 도구 (genome engineering tools)로서 이용되어 왔다 (Komor *et al.* *Cell* 168, 20-36 (2017); Puchta, *Curr. Opin. Plant Biol.* 36, 1-8 (2017); Xu *et al.* *J. Genet. Genomics* 42, 141-149 (2015); Tang *et al.* *Nat. Plants* 3, 17018 (2017); Zetsche *et al.* *Nat. Biotechnol.* 35, 31-34 (2016)) 및 prokaryotes (Mougiakos, *et al.* *Trends Biotechnol.* 34, 575-587 (2016)). 이들 시스템은, 이들 시스템이 단일 Cas 엔도뉴클레아제 및 RNA 가이드에 의해 형성된 리보핵산 단백질 (RNP) 복합체에 기초한 표적화된 이중 가닥 DNA 절단 (DSB)을 도입한 것으로 알려진 가장 단순한 CRISPR-Cas 시스템 중 하나이다.
- [0012] 현재까지, *Streptococcus pyogenes* Cas9 (SpCas9)는, 계놈 공학에 대해 가장 잘 특징화되고, 가장 널리 사용되는 Cas9이다. 몇 가지 다른 타입-II 시스템이 특징화되었지만, 그 중 어느 것도 호열성 유기체로부터 유래되지 않았다 (Nakade, *et al.* *Bioengineered* 1-9 (2017). doi:10.1080/21655979.2017.1282018). CRISPR-Cas 시스템의 특징화는, 기본적 통찰 (fundamental insights)을 얻을 뿐만 아니라 새로운 적용을 개발하는데 흥미로운 것이다.
- [0013] 비록 기본적인 유전 도구가 다수의 호열성 미생물에 대해 이용 가능하지만 (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)), 이들 도구들의 효율은, 여전히 너무 낮아서 유기체의 관심의 군의 충분한 탐구 및 이용이 가능하지 않다. SpCas9가 $\geq 42^{\circ}\text{C}$ 의 생체 내에서 활성이 없다는 우리의 확인에 기초하여, 우리는 이전에 상승된 온도에서 상동 재조합 및 적당한 온도에서 SpCas9-계 반대-선택 (counter-selection)을 조합하여, 조건적 호열성 미생물 (facultative thermophiles)에 대한 SpCas9-계 공학 도구를 개발했다 (Mougiakos *et al.* *ACS Synth. Biol.* 6, 849-861 (2017)). 그러나, 절대 호열성 미생물에 대한 Cas9-계 편집 및 침묵 도구 (silencing tool)는, SpCas9가 42℃ 이상에서 활성이 없음에 따라, 아직 이용 가능하지 않고 (Mougiakos *et al.* *ACS Synth. Biol.* 6, 849-861 (2017)), 및 지금까지 호열성 Cas9는 특징화되지 않았다.

발명의 내용

해결하려는 과제

- [0014] 본 발명자들은 ThermoCas9: 호열성 박테리아 지오바실러스 썬모데니트리피칸스 (*Geobacillus thermodenitrificans*) T12의 CRISPR-Cas 타입-II C 시스템 유래의 RNA-가이드 DNA-엔도뉴클레아제를 발견하고 특징화하였다. 본 발명자들은 놀랍게도 광범위한 온도 범위에 걸쳐 이의 시험관내 활성을 밝혀냈고, 열안정성을

위한 sgRNA-구조의 중요성을 입증하였으며, ThermoCas9를 넓은 온도 범위에 걸쳐 생체 내 게놈 편집을 위해 적용했다.

과제의 해결 수단

- [0015] 따라서, 본 발명은, 하기를 포함하는 단리된 일정한 간격을 두고 주기적으로 분포하는 짧은 회문구조 반복서열 (CRISPR)-관련 (Cas) 단백질 또는 폴리펩티드 제공한다:
- [0016] a. 아미노산 모티프 EKDGYKYYC [SEQ ID NO: 2]; 및/또는
- [0017] b. 아미노산 모티프 $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], 여기서, X_1 은 이소류신, 메티오닌 또는 프롤린으로부터 독립적으로 선택되고, X_2 는 발린, 세린, 아스파라긴 또는 이소류신으로부터 독립적으로 선택되며, X_3 는 글루타메이트 또는 리신으로부터 독립적으로 선택되고, 및 X_4 는 알라닌, 글루타메이트 또는 아르기닌 중 하나임; 및/또는
- [0018] c. 아미노산 모티프 X_5LXX_6IE [SEQ ID NO: 4], 여기서, X_5 는 메티오닌 또는 페닐알라닌으로부터 독립적으로 선택되고, 및 X_6 는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택됨; 및/또는
- [0019] d. 아미노산 모티프 X_7VYSX_8K [SEQ ID NO: 5], 여기서, X_7 은 글루타메이트 또는 이소류신이고, X_8 은 트립토판, 세린 또는 리신 중 하나임; 및/또는
- [0020] e. 아미노산 모티프 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], 여기서, X_9 은 알라닌 또는 글루타메이트이고, X_{10} 은 글루타민 또는 라신이며, X_{11} 은 아르기닌 또는 알라닌이고, X_{12} 는 아스파라긴 또는 알라닌이며, X_{13} 은 리신 또는 세린이다.
- [0021] 의심의 소지를 없애기 위해, 본 발명의 Cas 단백질을 코딩하는 단백질, 폴리 펩타이드 또는 핵산은 또한 "GtCas9" 또는 "ThermoCas9"로 지칭될 수 있다. "GtCas9" 및 "ThermoCas9"는 본 명세서 전체에서 상호교환적으로 사용되고, 동일한 의미를 갖는다.
- [0022] 본 발명의 맥락에서 폴리펩티드는, 전체 길이 Cas 단백질의 단편으로 고려될 수 있다. 이러한 단편은, 비활성일 수 있고, 및 유전 물질의 결합, 편집 및/또는 절단과 직접적으로 관련이 없는 방식으로 및 목적을 위해, 예를 들어, 분석법의 표준을 위해 또는 항체를 늘리기 위해 또는 이와 유사한 것을 위해 사용될 수 있다.
- [0023] 그러나, 바람직한 구체 예에서, Cas 단백질 또는 폴리펩티드는, 적어도 하나의 표적화 RNA 분자 (targeting RNA molecule), 및 상기 표적화 RNA 분자에 의해 인지된 표적 핵산 서열을 포함하는 폴리뉴클레오타이드와 관련되는 경우, 20℃ 내지 100℃ 범위의 온도에서 절단, 결합, 표지 (marking) 또는 변형시킬 수 있고, 기능성이다. 바람직하게는, Cas 단백질 또는 폴리펩티드는, 50℃ 내지 70℃의 온도 범위, 예를 들어, 55℃ 또는 60℃에서 절단, 결합, 표지 또는 변형시킬 수 있고, 기능성이다.
- [0024] 특정 구체 예에서, 본 발명은 아미노산 모티프 EKDGYKYYC [SEQ ID NO: 2]를 포함하는 Cas 단백질 또는 폴리펩티드를 제공할 수 있다. 다른 구체 예에서, 상기 Cas 단백질 또는 폴리펩티드는, 아미노산 모티프 $X_1X_2CTX_3X_4$ [SEQ ID NO: 3]를 더욱 포함할 수 있고, 여기서 X_1 은 이소류신, 메티오닌 또는 프롤린으로부터 독립적으로 선택되며, X_2 는 발린, 세린, 아스파라긴 또는 이소류신으로부터 독립적으로 선택되고, X_3 는 글루타메이트 또는 리신으로부터 독립적으로 선택되며, 및 X_4 는 알라닌, 글루타메이트 또는 아르기닌 중 하나이다.
- [0025] 다른 구체 예에서, 여기서 정의된 Cas 단백질 또는 폴리펩티드는, 아미노산 모티프 X_5LXX_6IE [SEQ ID NO: 4]를 부가적으로 더욱 포함할 수 있으며, 여기서 X_5 는 메티오닌 또는 페닐알라닌으로부터 독립적으로 선택되고, 및 X_6 는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택된다.
- [0026] 다른 구체 예에서, 여기서 정의된 Cas 단백질 또는 폴리펩티드는, 아미노산 모티프 X_7VYSX_8K [SEQ ID NO: 5]를 부가적으로 더욱 포함할 수 있으며, 여기서, X_7 은 글루타메이트 또는 이소류신이고, 및 X_8 은 트립토판, 세린 또는 라신 중 하나이다.
- [0027] 다른 구체 예에서, 여기서 정의된 Cas 단백질 또는 폴리펩티드는, 아미노산 모티프 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6]을 부가적으로 더욱 포함할 수 있으며, 여기서 X_9 는 알라닌 또는 글루타메이트이고, X_{10} 은 글루타민 또는

리신이며, X_{11} 은 아르기닌 또는 알라닌이고, X_{12} 는 아스파라긴 또는 알라닌이며, 및 X_{13} 은 리신 또는 세린이다.

[0028] 본 발명에 따르면, 본 발명의 Cas 단백질 또는 폴리펩티드는, SEQ ID NO: 2 내지 6의 모티프 중 어느 하나를 단독으로 또는 조합하여 포함할 수 있는 것으로 인정될 수 있다. 본 발명의 Cas 단백질 또는 폴리펩티드를 특징으로 할 수 있는 모티프의 각각의 조합을 요약하면 다음과 같다:

[0029] EKDGKYYC [SEQ ID NO: 2].

[0030] EKDGKYYC [SEQ ID NO: 2]; 및 $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], 여기서, X_1 은, 이소류신, 메티오닌 또는 프롤린으로부터 독립적으로 선택되고, X_2 는 발린, 세린, 아스파라긴 또는 이소류신으로부터 독립적으로 선택되며, X_3 은 글루타메이트 또는 리신으로부터 독립적으로 선택되고, 및 X_4 는 알라닌, 글루타메이트 또는 아르기닌 중 하나임.

[0031] EKDGKYYC [SEQ ID NO: 2]; 및 $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], 여기서, X_1 은 이소류신, 메티오닌 또는 프롤린으로부터 독립적으로 선택되고, X_2 는 발린, 세린, 아스파라긴 또는 이소류신으로부터 독립적으로 선택되며, X_3 은 글루타메이트 또는 리신으로부터 독립적으로 선택되고, 및 X_4 는 알라닌, 글루타메이트 또는 아르기닌 중 하나임; 및 X_5LXX_6IE [SEQ ID NO: 4], 여기서, X_5 는 메티오닌 또는 페닐알라닌으로부터 독립적으로 선택되고, 및 X_6 는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택됨.

[0032] EKDGKYYC [SEQ ID NO: 2]; 및 $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], 여기서, X_1 은 이소류신, 메티오닌 또는 프롤린으로부터 독립적으로 선택되고, X_2 는 발린, 세린, 아스파라긴 또는 이소류신으로부터 독립적으로 선택되며, X_3 는 글루타메이트 또는 리신으로부터 독립적으로 선택되고, 및 X_4 는 알라닌, 글루타메이트 또는 아르기닌 중 하나임; 및 X_5LXX_6IE [SEQ ID NO: 4], 여기서, X_5 는 메티오닌 또는 페닐알라닌으로부터 독립적으로 선택되고, 및 X_6 는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택됨; 및 X_7VYSX_8K [SEQ ID NO: 5], 여기서, X_7 은 글루타메이트 또는 이소류신이고, 및 X_8 은 트립토판, 세린 또는 리신 중 하나임.

[0033] EKDGKYYC [SEQ ID NO: 2]; 및 $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], 여기서, X_1 은 이소류신, 메티오닌 또는 프롤린으로부터 독립적으로 선택되고, X_2 는 발린, 세린, 아스파라긴 또는 이소류신으로부터 독립적으로 선택되며, X_3 는 글루타메이트 또는 리신으로부터 독립적으로 선택되고, 및 X_4 는 알라닌, 글루타메이트 또는 아르기닌 중 하나임; 및 X_5LXX_6IE [SEQ ID NO: 4], 여기서, X_5 는 메티오닌 또는 페닐알라닌으로부터 독립적으로 선택되고, 및 X_6 는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택됨; 및 X_7VYSX_8K [SEQ ID NO: 5], 여기서, X_7 은 글루타메이트 또는 이소류신이고, 및 X_8 은 트립토판, 세린 또는 리신 중 하나임; 및 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], 여기서, X_9 는 알라닌 또는 글루타메이트이고, X_{10} 은 글루타민 또는 리신이며, X_{11} 은 아르기닌 또는 알라닌이고, X_{12} 는 아스파라긴 또는 알라닌이며, 및 X_{13} 은 리신 또는 세린임.

[0034] EKDGKYYC [SEQ ID NO: 2]; 및 $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], 여기서, X_1 은 이소류신, 메티오닌 또는 프롤린으로부터 독립적으로 선택되고, X_2 는 발린, 세린, 아스파라긴 또는 이소류신으로부터 독립적으로 선택되며, X_3 는 글루타메이트 또는 리신으로부터 독립적으로 선택되고, 및 X_4 는 알라닌, 글루타메이트 또는 아르기닌 중 하나임; 및 X_5LXX_6IE [SEQ ID NO: 4], 여기서, X_5 는 메티오닌 또는 페닐알라닌으로부터 독립적으로 선택되고, 및 X_6 는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택됨; 및 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], 여기서, X_9 은 알라닌 또는 글루타메이트이고, X_{10} 은 글루타민 또는 리신이며, X_{11} 은 아르기닌 또는 알라닌이고, X_{12} 는 아스파라긴 또는 알라닌이며, 및 X_{13} 은 리신 또는 세린임.

[0035] EKDGKYYC [SEQ ID NO: 2]; 및 $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], 여기서, X_1 은 이소류신, 메티오닌 또는 프롤린으로부터 독립적으로 선택되고, X_2 는 발린, 세린, 아스파라긴 또는 이소류신으로부터 독립적으로 선택되며, X_3 는 글루타메이트 또는 리신으로부터 독립적으로 선택되고, 및 X_4 는 알라닌, 글루타메이트 또는 아르기닌 중 하나임; 및 X_7VYSX_8K [SEQ ID NO: 5], 여기서, X_7 은 글루타메이트 또는 이소류신이고, X_8 은 트립토판, 세린 또는 리신 중 하

나임; 및 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], 여기서, X_9 은 알라닌 또는 글루타메이트이고, X_{10} 은 글루타민 또는 리신이고, X_{11} 은 아르기닌 또는 알라닌이며, X_{12} 는 아스파라긴 또는 알라닌이고, 및 X_{13} 은 리신 또는 세린임.

[0036] EKDGKYYC [SEQ ID NO: 2]; 및 X_5LXX_6IE [SEQ ID NO: 4], 여기서, X_5 는 메티오닌 또는 페닐알라닌으로부터 독립적으로 선택되고, 및 X_6 는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택됨; 및 X_7VYSX_8K [SEQ ID NO: 5], 여기서, X_7 은 글루타메이트 또는 이소류신이고, X_8 은 트립토판, 세린 또는 리신 중 하나임; 및 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], 여기서, X_9 은 알라닌 또는 글루타메이트이고, X_{10} 은 글루타민 또는 리신이며, X_{11} 은 아르기닌 또는 알라닌이고, X_{12} 는 아스파라긴 또는 알라닌이며, 및 X_{13} 은 리신 또는 세린임.

[0037] EKDGKYYC [SEQ ID NO: 2]; 및 X_5LXX_6IE [SEQ ID NO: 4], 여기서, X_5 는 메티오닌 또는 페닐알라닌으로부터 독립적으로 선택되고, 및 X_6 는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택됨.

[0038] EKDGKYYC [SEQ ID NO: 2]; 및 X_7VYSX_8K [SEQ ID NO: 5], 여기서, X_7 은 글루타메이트 또는 이소류신이고, X_8 은 트립토판, 세린 또는 리신 중 하나임.

[0039] EKDGKYYC [SEQ ID NO: 2]; 및 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], 여기서, X_9 은 알라닌 또는 글루타메이트이고, X_{10} 은 글루타민 또는 리신이며, X_{11} 은 아르기닌 또는 알라닌이고, X_{12} 는 아스파라긴 또는 알라닌이며, 및 X_{13} 은 리신 또는 세린임.

[0040] EKDGKYYC [SEQ ID NO: 2]; 및 X_5LXX_6IE [SEQ ID NO: 4] 여기서, X_5 는 메티오닌 또는 페닐알라닌으로부터 독립적으로 선택되고, 및 X_6 는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택됨; 및 X_7VYSX_8K [SEQ ID NO: 5], 여기서, X_7 은 글루타메이트 또는 이소류신이고, 및 X_8 은 트립토판, 세린 또는 리신 중 하나임;

[0041] EKDGKYYC [SEQ ID NO: 2]; 및 X_5LXX_6IE [SEQ ID NO: 4], 여기서, X_5 는 메티오닌 또는 페닐알라닌으로부터 독립적으로 선택되고, 및 X_6 는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택됨; 및 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], 여기서, X_9 은 알라닌 또는 글루타메이트이고, X_{10} 은 글루타민 또는 리신이며, X_{11} 은 아르기닌 또는 알라닌이고, X_{12} 는 아스파라긴 또는 알라닌이며, 및 X_{13} 은 리신 또는 세린임.

[0042] EKDGKYYC [SEQ ID NO: 2]; 및 X_7VYSX_8K [SEQ ID NO: 5], 여기서, X_7 은 글루타메이트 또는 이소류신이고, 및 X_8 은 트립토판, 세린 또는 리신 중 하나임; 및 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], 여기서, X_9 은 알라닌 또는 글루타메이트이고, X_{10} 은 글루타민 또는 리신이며, X_{11} 은 아르기닌 또는 알라닌이고, X_{12} 는 아스파라긴 또는 알라닌이며, 및 X_{13} 은 리신 또는 세린임.

[0043] $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], 여기서, X_1 은 이소류신, 메티오닌 또는 프롤린으로부터 독립적으로 선택되고, X_2 는 발린, 세린, 아스파라긴 또는 이소류신으로부터 독립적으로 선택되며, X_3 는 글루타메이트 또는 리신으로부터 독립적으로 선택되고, 및 X_4 는 알라닌, 글루타메이트 또는 아르기닌 중 하나임; 및 X_5LXX_6IE [SEQ ID NO: 4], 여기서, X_5 는 메티오닌 또는 페닐알라닌으로부터 독립적으로 선택되고, 및 X_6 는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택됨.

[0044] $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], 여기서, X_1 은 이소류신, 메티오닌 또는 프롤린으로부터 독립적으로 선택되고, X_2 는 발린, 세린, 아스파라긴 또는 이소류신으로부터 독립적으로 선택되며, X_3 는 글루타메이트 또는 리신으로부터 독립적으로 선택되고, 및 X_4 는 알라닌, 글루타메이트 또는 아르기닌 중 하나임; 및 X_5LXX_6IE [SEQ ID NO: 4], 여기서, X_5 는 메티오닌 또는 페닐알라닌로부터 독립적으로 선택되고, 및 X_6 는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택됨; 및 X_7VYSX_8K [SEQ ID NO: 5], 여기서, X_7 은 글루타메이트 또는 이소류신이고 및 X_8 은 트립토판, 세린 또는 리신 중 하나임.

[0045] $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], 여기서, X_1 은 이소류신, 메티오닌 또는 프롤린으로부터 독립적으로 선택되고, X_2 는

발린, 세린, 아스파라긴 또는 이소류신으로부터 독립적으로 선택되며, X_3 는 글루타메이트 또는 리신으로부터 독립적으로 선택되고, 및 X_4 는 알라닌, 글루타메이트 또는 아르기닌 중 하나임; 및 X_5 LKX₆IE [SEQ ID NO: 4], 여기서, X_5 는 메티오닌 또는 페닐알라닌으로부터 독립적으로 선택되고, 및 X_6 는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택됨; 및 X_7 VYSX₈K [SEQ ID NO: 5], 여기서, X_7 은 글루타메이트 또는 이소류신이고, 및 X_8 은 트립토판, 세린 또는 리신 중 하나임; 및 X_9 FYX₁₀X₁₁REQX₁₂KEX₁₃ [SEQ ID NO: 6], 여기서, X_9 는 알라닌 또는 글루타메이트이고, X_{10} 은 글루타민 또는 리신이며, X_{11} 은 아르기닌 또는 알라닌이고, X_{12} 는 아스파라긴 또는 알라닌이며 및 X_{13} 는 리신 또는 세린임.

[0046] X_1X_2 CTX₃X₄ [SEQ ID NO: 3], 여기서, X_1 은 이소류신, 메티오닌 또는 프롤린으로부터 독립적으로 선택되고, X_2 는 발린, 세린, 아스파라긴 또는 이소류신으로부터 독립적으로 선택되며, X_3 는 글루타메이트 또는 리신으로부터 독립적으로 선택되고, 및 X_4 는 알라닌, 글루타메이트 또는 아르기닌 중 하나임; 및 X_7 VYSX₈K [SEQ ID NO: 5], 여기서, X_7 은 글루타메이트 또는 이소류신이고, 및 X_8 은 트립토판, 세린 또는 리신 중 하나임; 및 X_9 FYX₁₀X₁₁REQX₁₂KEX₁₃ [SEQ ID NO: 6], 여기서, X_9 는 알라닌 또는 글루타메이트이고, X_{10} 은 글루타민 또는 리신이며, X_{11} 은 아르기닌 또는 알라닌이고, X_{12} 는 아스파라긴 또는 알라닌이며, 및 X_{13} 은 리신 또는 세린임.

[0047] X_1X_2 CTX₃X₄ [SEQ ID NO: 3], 여기서, X_1 은 이소류신, 메티오닌 또는 프롤린으로부터 독립적으로 선택되고, X_2 는 발린, 세린, 아스파라긴 또는 이소류신으로부터 독립적으로 선택되며, X_3 는 글루타메이트 또는 리신으로부터 독립적으로 선택되고, 및 X_4 는 알라닌, 글루타메이트 또는 아르기닌 중 하나임; 및 X_7 VYSX₈K [SEQ ID NO: 5], 여기서, X_7 은 글루타메이트 또는 이소류신이고, 및 X_8 은 트립토판, 세린 또는 리신 중 하나임.

[0048] X_1X_2 CTX₃X₄ [SEQ ID NO: 3], 여기서, X_1 은 이소류신, 메티오닌 또는 프롤린으로부터 독립적으로 선택되고, X_2 는 발린, 세린, 아스파라긴 또는 이소류신으로부터 독립적으로 선택되며, X_3 는 글루타메이트 또는 리신으로부터 독립적으로 선택되고, 및 X_4 는 알라닌, 글루타메이트 또는 아르기닌 중 하나임; 및 X_9 FYX₁₀X₁₁REQX₁₂KEX₁₃ [SEQ ID NO: 6], 여기서, X_9 는 알라닌 또는 글루타메이트이고, X_{10} 은 글루타민 또는 리신이며, X_{11} 은 아르기닌 또는 알라닌이고, X_{12} 는 아스파라긴 또는 알라닌이며 및 X_{13} 은 리신 또는 세린임.

[0049] X_5 LKX₆IE [SEQ ID NO: 4], 여기서, X_5 는 메티오닌 또는 페닐알라닌으로부터 독립적으로 선택되고, 및 X_6 는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택됨; 및 X_7 VYSX₈K [SEQ ID NO: 5], 여기서, X_7 은 글루타메이트 또는 이소류신이고 및 X_8 은 트립토판, 세린 또는 리신 중 하나임; 및 X_9 FYX₁₀X₁₁REQX₁₂KEX₁₃ [SEQ ID NO: 6], 여기서, X_9 는 알라닌 또는 글루타메이트이고, X_{10} 은 글루타민 또는 리신이며, X_{11} 은 아르기닌 또는 알라닌이고, X_{12} 는 아스파라긴 또는 알라닌이며 및 X_{13} 은 리신 또는 세린임.

[0050] X_5 LKX₆IE [SEQ ID NO: 4], 여기서, X_5 는 메티오닌 또는 페닐알라닌으로부터 독립적으로 선택되고, 및 X_6 는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택됨; 및 X_7 VYSX₈K [SEQ ID NO: 5], 여기서, X_7 은 글루타메이트 또는 이소류신이고 및 X_8 은 트립토판, 세린 또는 리신 중 하나임.

[0051] X_5 LKX₆IE [SEQ ID NO: 4], 여기서, X_5 는 메티오닌 또는 페닐알라닌으로부터 독립적으로 선택되고, 및 X_6 는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택됨; 및 X_9 FYX₁₀X₁₁REQX₁₂KEX₁₃ [SEQ ID NO: 6], 여기서, X_9 는 알라닌 또는 글루타메이트이고, X_{10} 은 글루타민 또는 리신이며, X_{11} 은 아르기닌 또는 알라닌이고, X_{12} 는 아스파라긴 또는 알라닌이며 및 X_{13} 은 리신 또는 세린임.

[0052] X_7 VYSX₈K [SEQ ID NO: 5], 여기서, X_7 은 글루타메이트 또는 이소류신이고 및 X_8 은 트립토판, 세린 또는 리신 중 하나임; 및 X_9 FYX₁₀X₁₁REQX₁₂KEX₁₃ [SEQ ID NO: 6], 여기서, X_9 는 알라닌 또는 글루타메이트이고, X_{10} 은 글루타민 또는 리신이며, X_{11} 은 아르기닌 또는 알라닌이고, X_{12} 는 아스파라긴 또는 알라닌이며 및 X_{13} 은 리신 또는 세린임.

- [0053] 또 다른 관점에서, 본 발명은, SEQ ID NO: 1의 아미노산 서열 또는 이와 적어도 77% 동일성의 서열을 갖는 단리된 Cas 단백질 또는 이의 폴리펩티드 단편을 제공하고, 여기서, 상기 Cas 단백질 또는 이의 폴리펩티드 단편은, 다음의 모티프 또는 아미노산 중 어느 하나를, 단독 또는 조합으로 포함할 수 있다:
- [0054] IGLDIGITSIG [SEQ ID NO: 23], 바람직하게는 IGLDIGITSIGWAVINLD [SEQ ID NO: 24]를 포함하는 RuvC-I 도메인.
- [0055] RSARR [SEQ ID NO: 25], 바람직하게는 PRRLARSARRRLRRRKHRLERIRRL [SEQ ID NO: 26]을 포함하는 브릿지 도메인 (Bridge domain); 및/또는
- [0056] WQLR [SEQ ID NO: 27]을 포함하는 α -나선형/인식 로브 도메인 (lobe domain); 및/또는
- [0057] HLAKRKG [SEQ ID NO: 28], 바람직하게는 LARILLHLAKRRG [SEQ ID NO: 29]를 포함하는 α -나선형/인식 로브 도메인; 및/또는
- [0058] IFAKQ [SEQ ID NO: 30], 바람직하게는 EIKLIFAKQ [SEQ ID NO: 31]을 포함하는 α -나선형/인식 로브 도메인; 및/또는
- [0059] IWASQR [SEQ ID NO: 32]를 포함하는 α -나선형/인식 로브 도메인; 및/또는
- [0060] KVGFCFEPKEKRAPK [SEQ ID NO: 33]; 및/또는
- [0061] FTVWEHINKLRL [SEQ ID NO: 34]; 및/또는
- [0062] 모티프 IANPVVMRALTQ [SEQ ID NO: 35], 바람직하게는 IANPVVMRALTQARKVVNAIIKKYG [SEQ ID NO: 36]을 포함하는 RuvC-II 도메인; 및/또는
- [0063] 모티프 ELAR [SEQ ID NO: 37], 바람직하게는 IHIELARE [SEQ ID NO: 38]을 포함하는 RuvC-II 도메인; 및/또는
- [0064] 모티프 QNGKCAY [SEQ ID NO: 39], 바람직하게는 IVKFKLWSEQNGKCAY [SEQ ID NO: 40]을 포함하는 HNH 도메인; 및/또는
- [0065] 모티프 VDHVIP [SEQ ID NO: 41], 바람직하게는 VDHVIPYRSRLDSDYTNKVL [SEQ ID NO: 42]를 포함하는 HNH 도메인; 및/또는
- [0066] 모티프 DTRYISRFLAN [SEQ ID NO: 43]을 포함하는 RuvC-III 도메인; 및/또는
- [0067] 모티프 VYTVNGRITAHLSRW [SEQ ID NO: 44]를 포함하는 RuvC-III 도메인; 및/또는
- [0068] 모티프 HHAVIDA [SEQ ID NO: 45], 바람직하게는 HHAVIDAIVA [SEQ ID NO: 46]를 포함하는 RuvC-III 도메인.
- [0069] 바람직하게는, 본 발명은 SEQ ID NO: 1의 아미노산 서열 또는 이와 적어도 77% 동일성의 서열을 갖는 단리된 Cas 단백질 또는 이의 폴리펩티드 단편을 제공하고, 여기서, 상기 Cas 단백질 또는 이의 폴리펩티드 단편은, 각각의 아미노산 모티프 [SEQ ID NO: 23] 내지 [SEQ ID NO: 46]을 조합하여 포함한다.
- [0070] 또 다른 관점에서, 본 발명은, SEQ ID NO: 1의 아미노산 서열 또는 이와 적어도 77% 동일성의 서열을 갖는 단리된 Cas 단백질 또는 이의 폴리펩티드 단편을 제공한다. 바람직하게는, 상기 Cas 단백질 또는 폴리펩티드는, 20℃ 내지 100℃ 범위의 온도에서 결합, 절단, 표지 또는 변형시킬 수 있다. 바람직하게는, 상기 Cas 단백질 또는 폴리펩티드는, 20℃ 내지 70℃ 범위의 온도, 예를 들어, 25℃, 55℃, 60℃ 또는 65℃에서 절단, 결합, 표지 또는 변형시킬 수 있다. 바람직하게는, Cas 단백질 또는 폴리펩티드는, 50℃ 내지 70℃ 범위의 온도, 예를 들어, 55℃ 또는 60℃에서 절단, 결합, 표지 또는 변형시킬 수 있다. 바람직하게는, Cas 단백질 또는 폴리펩티드는, 30℃ 내지 80℃ 범위의 온도, 37℃ 내지 78℃의 온도, 바람직하게는, 55℃ 이상의 온도; 좀 더 바람직하게는, 55℃ 내지 80℃의 온도; 더욱 더 바람직하게는, 55℃ 내지 65℃ 또는 60℃ 내지 65℃의 온도에서 절단, 결합, 표지 또는 변형시킬 수 있다.
- [0071] 본 발명은 또한 표적 핵산 서열을 포함하는 표적 폴리뉴클레오타이드의 결합, 절단, 표지 또는 변형을 위해, 여기서 제공되는 표적화 RNA 분자 및 Cas 단백질 또는 폴리펩티드를 사용하는 방법을 제공한다. 상기 표적화 RNA 분자는 폴리뉴클레오타이드의 표적 핵산 가닥 상에 표적 핵산 서열을 인식한다.
- [0072] 상기 표적 핵산 서열을 포함하는 표적 폴리뉴클레오타이드는 이중 가닥일 수 있으며, 및 상기 표적 핵산 서열을 포함하는 표적 핵산 가닥, 및 프로토스페이서 핵산 서열을 포함하는 비-표적 핵산 가닥을 포함한다. 프로토스페이서 핵산 서열은, 표적 핵산 서열에 실질적으로 상보적이고, 이중 가닥 표적 폴리뉴클레오타이드에서 이와 쌍을 이룬다. 상기 비-표적 핵산 가닥은 프로토스페이서 서열의 3' 말단에 바로 인접한 프로토스페이서 인근 모티프

(PAM) 서열을 더욱 포함할 수 있다. PAM 서열은 길이가 적어도 6, 7, 또는 8개의 핵산일 수 있다. 바람직하게는, PAM 서열은 제5 위치에 시토신을 갖는다. 바람직하게는, PAM 서열은, 5'-말단으로부터 PAM 서열이 5'-NNNNC-3'을 시작하도록, 서열 5'-NNNNC-3'을 포함한다. 부가적으로 또는 선택적으로, PAM 서열은, PAM 서열이 서열 5'-NNNNNNNA-3'을 포함하고, 5'-말단으로부터 PAM 서열이 5'-NNNNNNNA-3'을 시작하도록, 제8 위치(eighth position)에 아데닌을 가질 수 있다. 부가적으로 또는 선택적으로, PAM 서열은, 5'-말단으로부터 PAM 서열이 5'-CNNNN-3', 5'-NCNNN-3', 5'-NNCNN-3', 5'-NNNCN-3', 및/또는 5'-NNNNNC-3'을 시작하도록, 제1, 제2, 제3, 제4, 및 제6 위치 중 하나 이상에서 시토신을 가질 수 있다. 선택적으로, PAM 서열은, 5'-말단으로부터 PAM 서열이 5'-CCCCCAA-3' [SEQ ID NO: 10]을 시작하도록 포함하고, 및 더욱 바람직하게는, PAM 서열은, 5'-말단으로부터 PAM 서열이 5'-CCCCCAA-3' [SEQ ID NO: 11]을 시작하도록, 포함한다. 다른 바람직한 PAM 서열은, 5'-ATCCCCAA-3' [SEQ ID NO: 21] 및 5'-ACGGCCAA-3' [SEQ ID NO: 22]를 포함한다.

[0073] 바람직하게는, Cas 단백질 또는 폴리펩티드는, 40℃ 내지 80℃의 범위, 바람직하게는, 45℃ 내지 80℃의 범위, 및 더욱 바람직하게는, 50℃ 내지 80℃ 범위의 온도에서 결합, 절단, 표지 또는 변형시킬 수 있다. 예를 들어, 결합, 절단, 표지 또는 변형은, 45℃, 46℃, 47℃, 48℃, 49℃, 50℃, 51℃, 52℃, 53℃, 54℃, 55℃, 56℃, 57℃, 58℃, 59℃, 60℃, 61℃, 62℃, 63℃, 64℃, 65℃, 66℃, 67℃, 68℃, 69℃, 70℃, 71℃, 72℃, 73℃, 74℃, 75℃, 76℃, 77℃, 78℃, 79℃ 또는 80℃의 온도에서 발생한다. 좀 더 바람직하게, Cas 단백질 또는 폴리펩티드는, 55 내지 65℃의 온도에서 결합, 절단, 표지 또는 변형시킬 수 있다. 바람직한 관점에서, 본 발명의 Cas 단백질 또는 폴리펩티드 단편은, SEQ ID NO: 1과 적어도 75% 동일성; 바람직하게는, 적어도 85%; 좀 더 바람직하게는, 적어도 90%; 더욱더 바람직하게는, 적어도 95% 동일성의 아미노산 서열을 포함할 수 있다.

[0074] Cas 단백질 또는 폴리펩티드는, 표적 핵산 가닥 상에 표적 핵산 서열을 인식하는 표적화 RNA 분자와 조합하여 사용될 수 있으며, 여기서, 비-표적 핵산 서열은, 여기서 개시된 바와 같이, 비-표적 가닥 상에 프로토스페이서 서열의 3' 말단에 바로 인접한 PAM 서열을 갖는다. 따라서, PAM 서열은 서열 5'-NNNNC-3'을 포함할 수 있으며, Cas 단백질은, 20℃ 내지 100℃의 범위, 바람직하게는, 30℃ 내지 90℃의 범위, 37℃ 내지 78℃의 범위, 40℃ 내지 80℃의 범위, 50℃ 내지 70℃의 범위, 또는 55℃ 내지 65℃ 범위의 온도에서 표적 가닥을 결합, 절단, 표지 또는 변형할 수 있다. 바람직하게는, 5'-말단으로부터 PAM 서열은 5'-NNNNC-3'을 시작하고, Cas 단백질은, 20℃ 내지 100℃의 범위, 바람직하게는, 30℃ 내지 90℃의 범위, 37℃ 내지 78℃의 범위, 40℃ 내지 80℃의 범위, 50℃ 내지 70℃의 범위, 또는 55℃ 내지 65℃ 범위의 온도에서 표적 가닥을 결합, 절단, 표지 또는 변형할 수 있다. 더욱 바람직하게는, 5'-말단으로부터 PAM 서열은 5'-NNNNNCNA-3' [SEQ ID NO: 47]을 시작하고, Cas 단백질은, 20℃ 내지 100℃의 범위, 바람직하게는, 30℃ 내지 90℃의 범위, 37℃ 내지 78℃의 범위, 40℃ 내지 80℃의 범위, 50℃ 내지 70℃의 범위, 또는 55℃ 내지 65℃ 범위의 온도에서 표적 가닥을 결합, 절단, 표지 또는 변형할 수 있다.

[0075] 좀 더 구체적으로는, 본 발명의 Cas 단백질 또는 폴리펩티드는: 적어도 60%, 적어도 61%, 적어도 62%, 적어도 63%, 적어도 64%, 적어도 65%, 적어도 66%, 적어도 67%, 적어도 68%, 적어도 69%, 적어도 70%, 적어도 71%, 적어도 72%, 적어도 73%, 적어도 74%, 적어도 75%, 적어도 76%, 적어도 77%, 적어도 78%, 적어도 79%, 적어도 80%, 적어도 81%, 적어도 82%, 적어도 83%, 적어도 84%, 적어도 85%, 적어도 86%, 적어도 87%, 적어도 88%, 적어도 89%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 적어도 99.5%, 또는 적어도 99.8%와 같은, SEQ ID NO: 1과 퍼센트 동일성(percentage identity)을 갖는 아미노산 서열을 포함할 수 있다. 상기 퍼센트 동일성은 적어도 89%일 수 있다. 상기 퍼센트 동일성은 적어도 90%일 수 있다. 바람직하게는, 상기 퍼센트 동일성을 적어도 95%, 예를 들어, 98%일 것이다.

[0076] SEQ ID NO: 1과 퍼센트 아미노산 서열 동일성은, 두 서열의 최적 정렬을 위해 도입될 필요가 있는, 갭(gaps)의 수, 및 각 갭의 길이를 고려하여, 선택된 비교 창(comparison window)에서 서열에 의해 공유되는 동일한 위치의 수의 함수에 따라 결정 가능하다.

[0077] 본 발명의 Cas 단백질 또는 폴리펩티드 단편은, 참조 서열(reference sequences) SEQ ID NO: 1 및 퍼센트 서열 동일성에 의해 정의된 바와 같은 임의의 전술된 이의 퍼센트 변이체 모두의 관점에서, 필수적인 특성(essential features)으로서 전술한 아미노산 모티프들(즉, SEQ ID NOS 2 및/또는 3 및/또는 4 및/또는 5 및/

또는 6) 중 어느 하나와 조합하여 또는 단독을 특징으로 할 수 있다.

[0078] 본 발명은 표적 핵산 서열을 포함하는 표적 핵산 가닥을 결합, 절단, 표지 또는 변형시키기 위한 본 발명의 Cas 단백질 또는 폴리펩티드 및 여기에 제공된 바와 같은 표적화 RNA 분자의 사용 방법을 제공한다. 바람직하게는, 상기 결합, 절단, 표지 또는 변형은, 여기에 개시된 온도, 예를 들어, 20 내지 100℃의 온도에서 발생한다. 본 발명은 또한 여기에 제공된 바와 같은 표적화 RNA 분자를 디자인하는 단계 및 본 발명의 Cas 단백질 또는 폴리펩티드 및 상기 표적화 RNA 분자를 포함하는 리보핵산 단백질 복합체를 형성하는 단계를 포함하는 표적 핵산 가닥에서 표적 핵산 서열을 결합, 절단, 표지 또는 변형시키는 방법을 제공한다. 바람직하게는, 상기 리보핵산 단백질 복합체는, 여기서 개시된 온도, 예를 들어, 37 내지 100℃의 온도에서 표적 핵산 서열을 결합, 절단, 표지 또는 변형시킨다.

[0079] 본 발명의 사용 및 방법들은 수행될 수 있고, 본 발명의 핵산단백질은, 생체 내에서, 예를 들어, 인간 세포에서 형성되고 사용된다. 선택적으로, 본 발명의 사용 및 방법은 수행될 수 있고, 본 발명의 핵산단백질은, 생체내에서, 예를 들어, 배아 줄기 세포가 아닌, 인간 세포에서 형성되고 사용된다. 선택적으로, 본 발명의 사용 및 방법은 수행될 수 있으며, 본 발명의 핵산단백질은, 생체내에서, 예를 들어, 인간의 생식 계열 (germ line) 유전적 동일성을 변형시키는 것을 포함하지 않는 방법으로 인간 세포에서 형성되고 사용된다. 선택적으로, 본 발명의 사용 및 방법은 수행될 수 있으며, 본 발명의 핵산단백질은 시험관 내에서 형성되고 사용된다. 본 발명의 Cas 단백질은, 예를 들어, 시험관 내에서 사용되거나 또는 형질감염 (transfection)에 의해 세포에 부가된 경우, 단리된 형태로 제공될 수 있고, Cas 단백질은, 예를 들어, Cas 단백질을 인코딩하는 핵산에 의한 세포의 일시적 또는 안정한 형질전환 후에, 이종방식으로 (heterologously) 발현될 수 있으며, 표적화 RNA 분자는 RNA 분자를 인코딩하는 핵산에 의해 세포의 일시적 또는 안정한 형질전환 후에 발현 벡터로부터 전사될 수 있고, 및/또는 RNA 분자는, 예를 들어, 시험관 내에서 사용되거나 또는 형질감염에 의해 세포에 부가된 경우, 단리된 형태로 제공될 수 있다. 바람직한 구체 예에서, Cas 단백질 또는 폴리펩티드는, 숙주 세포의 게놈에서 Cas 단백질 또는 폴리펩티드를 인코딩하는 핵산의 안정한 인터그레이션 후에, 숙주 세포의 게놈으로부터 발현된다. 따라서, Cas 단백질 및/또는 RNA 분자는, 이것이 달리 존재하지 않는, 세포에 단백질 또는 핵산 분자를 첨가하기 위한 임의의 인공적 또는 인위적인 방법을 사용하여 생체 내 또는 시험관 내 환경에 첨가될 수 있다.

[0080] 표적 핵산 서열을 포함하는 폴리뉴클레오타이드는, Cas 단백질에 의해 절단될 수 있고, 선택적으로 절단은 DNA 절단일 수 있다. 표적 서열을 포함하는 표적 핵산 가닥은 이중 가닥 DNA일 수 있으며, 상기 방법 또는 사용은, 표적 핵산 서열을 포함하는 폴리뉴클레오타이드에서 이중 가닥 절단을 결과할 수 있다. 표적 핵산 서열을 포함하는 폴리뉴클레오타이드는, 이중 가닥 DNA일 수 있고, Cas 단백질은 이중 가닥 DNA를 절단하는 능력이 없을 수 있으며, 상기 사용 또는 방법은 폴리뉴클레오타이드의 유전자 침묵을 결과할 수 있다.

[0081] Cas 단백질 또는 폴리펩티드는, 250nM 이하의 농도, 예를 들어, 200nM 이하, 150nM 이하, 100nM 이하, 50nM 이하, 25nM 이하, 10nM 이하, 5nM 이하, 1nM 이하 또는 0.5nM 이하의 농도로 본 발명의 방법, 사용 및 핵산단백질에 대해 제공될 수 있다. 선택적으로, Cas 단백질 또는 폴리펩티드는, 적어도 0.5nM, 적어도 1nM, 적어도 5nM, 적어도 10nM, 적어도 25nM, 적어도 50nM, 적어도 100nM, 적어도 150nM, 적어도 200 nM, 또는 적어도 250 nM의 농도로 제공될 수 있다. 본 발명의 PAM 서열은, PAM 서열이 5'-NNNNNNA-3'서열을 포함하도록, 제8 위치에 아데닌을 가질 수 있고, Cas 단백질 또는 폴리펩티드의 농도는 100 nM 이하, 50 nM 이하, 25nM 이하, 10nM 이하, 5nM 이하, 1nM 이하 또는 0.5nM 이하일 수 있다. PAM 서열은 서열 5'-NNNNCNA-3' [SEQ ID NO: 47]을 포함할 수 있으며, Cas 단백질 또는 폴리펩티드의 농도는 100 nM 이하, 50 nM 이하, 25 nM 이하, 10 nM 이하, 5 nM 이하, 1nM 또는 0.5 nM 이하일 수 있다. PAM 서열은 서열 5'-CCCCCNA-3' [SEQ ID NO:10]을 포함할 수 있고, Cas 단백질 또는 폴리펩티드의 농도는 100 nM 이하, 50 nM 이하, 25 nM 이하, 10 nM 이하, 5nM 이하, 1nM 이하 또는 0.5nM 이하일 수 있다.

[0082] 또한, 본 발명은 본 발명의 전술된 단백질 또는 폴리펩티드 중 어느 하나를 인코딩하는 핵산을 제공한다. 상기 핵산은 단리될 수 있거나 또는 발현 구축물의 형태일 수 있다.

[0083] 본 발명의 모든 전술된 관점에서, 아미노산 잔기는 보존적으로 또는 비-보존적으로 치환될 수 있다. 보존적 아미노산 치환은, 아미노산 잔기가 유사한 화학적 특성 (예를 들어, 전하 또는 소수성)을 갖는 다른 아미노산 잔기로 치환되고, 따라서 그 결과로 생긴 폴리펩티드의 기능적 특성을 변경시키지 않는 것을 지칭한다.

[0084] 유사하게, 핵산 서열이 폴리펩티드의 기능에 영향을 미치지 않으면서 보존적으로 또는 비-보존적으로 치환될 수 있는 것은 기술분야의 당업자에 의해 인정될 것이다. 보존적으로 변형된 핵산은, 동일하거나 또는 기능적으로 동일한 아미노산 서열의 변이체를 인코딩하는 핵산에 대해 치환된다. 핵산에서 각 코돈 (codon) (AUG 및 UGG를

제외한; 통상적으로 메티오닌 또는 트립토판, 각각에 대한 단 하나의 코돈)이 기능적으로 동일한 분자를 산출하도록 변형될 수 있는 것은 당업자에 의해 인정될 것이다. 따라서, 본 발명의 폴리펩티드를 인코딩하는, 폴리뉴클레오티드 또는 폴리펩티드의 각 침묵 변이 (silent variation) (즉, 동의 코돈)는, 각 기재된 폴리펩티드 서열에서 내포된다.

[0085] 본 발명은 이중 가닥 표적 폴리뉴클레오티드에서 표적 핵산 서열을 갖는, 형질전환된 인간 세포 (transformed human cell)를 제공하며, 상기 세포는, 여기에 제공된 바와 같은 Cas 단백질 또는 폴리펩티드 및 여기에 제공된 바와 같은 적어도 하나의 표적화 RNA 분자, 및 상기 Cas 단백질 및 상기 표적화 RNA 분자 중 적어도 하나를 인코딩하는 핵산을 포함하는 발현 벡터를 포함한다. 선택적으로, 형질전환된 인간 세포는, 단리된 인간 세포이다. 선택적으로, 인간 세포는 배아 줄기 세포가 아니다. Cas 단백질 및 표적화 RNA 분자는, 표적 서열의 결합, 절단, 표지 또는 변형이 증가된 온도 또는 여기서 개시된 바와 같은, 다양한 온도, 예를 들어, 37℃ 내지 100℃에서 형질전환된 세포에서 발생하는 것을 가능하게 하거나 또는 허용할 수 있다. 본 발명은: 1) 본 발명의 Cas 단백질 또는 폴리펩티드를 인코딩하는 뉴클레오티드 서열 및 본 발명의 표적화 RNA 분자를 인코딩하는 뉴클레오티드 서열을 포함하는 발현 벡터로 세포를 형질전환, 형질감염 또는 형질도입하는 단계; 또는 2) 본 발명의 Cas 단백질 또는 폴리펩티드를 인코딩하는 뉴클레오티드 서열을 포함하는 발현 벡터 및 본 발명의 표적화 RNA 분자를 인코딩하는 뉴클레오티드 서열을 포함하는 추가의 발현 벡터로 세포를 형질전환, 형질감염 또는 형질도입하는 단계; 또는 3) 본 발명의 Cas 단백질 또는 폴리펩티드를 인코딩하는 뉴클레오티드 서열을 포함하는 발현 벡터로 세포를 형질전환, 형질감염 또는 형질도입 단계, 및 여기에 제공된 바와 같은 표적화 RNA 분자를 세포에, 또는 세포 내로 전달하는 단계를 포함하는, 인간 세포에서 표적 핵산을 결합, 절단, 표지 또는 변형시키는 방법을 더욱 제공한다. Cas 단백질 또는 폴리펩티드는, 예를 들어, Cas 단백질 또는 폴리펩티드를 인코딩하는 뉴클레오티드 서열의 게놈 내로 안정한 인테그레이션 후에, 형질전환된 세포의 게놈으로부터 발현될 수 있다. 선택적으로, 상기 방법은, 인간의 생식 계열 유전적 동일성을 변형시키는 것을 포함하지 않는다.

[0086] 본 발명은 또한 본 발명의 사용 및 방법을 수행하기 위한, 또는 본 발명의 형질전환된 인간 세포 또는 핵산단백질 복합체를 발생하기 위한, 하나 이상의 시약 (reagents)을 포함하는 키트를 제공하며, 상기 키트는: 본 발명의 Cas 단백질 또는 폴리펩티드 또는 본 발명의 Cas 단백질 또는 폴리펩티드를 인코딩하는 핵산 서열을 포함하는 발현 벡터; 및/또는 본 발명의 표적화 RNA 분자 또는 본 발명의 표적화 RNA 분자를 인코딩하는 핵산 서열을 포함하는 발현 벡터를 포함한다. 상기 키트는 본 발명을 수행하기 위한 설명서, 예를 들어, 본 발명에 따라 표적화 RNA 분자를 디자인하는 방법에 대한 설명서를 더욱 포함할 수 있다.

도면의 간단한 설명

[0087] 이하, 본 발명은 특정 구체 예 및 첨부된 도면을 참조하여 상세히 설명된다:

도 1은, Cas9 단백질 서열의 근린-결합 트리 (Neighbor-Joining tree)를 나타낸다. pBLAST 또는 PSI-BLAST에 기초한 군주 T12와 40%를 초과하는 서열 유사성을 갖는 모든 서열은 포함되며, 이에 더하여 현재 잘-특징화된 서열 (*S. pyogenes*, *S. thermophiles* 및 *A. naeslundii*)뿐만 아니라, 40% 아래의 동일성이 있는 경우 현재 동정된 모든 호열성 서열 또한 포함된다. 모든 호열성 서열에 대해, T12에 대한 퍼센트 동일성 (percentage identity)은 군주 명 뒤에 표시된다. 유전자 식별자 (Gene identifier: gi) 번호는 종명 (species name) 앞에 표시된다. 범례: 폐쇄 원: 호열성 (최적 60℃ 이상) Cas9 서열, 폐쇄 사각형: 내열성 (최적 <50℃) Cas9 서열, 개방 삼각형: 중온성 기원 (mesophilic origin)으로부터 게놈 편집 목적을 위해 현재 가장 많이 사용된 Cas9 서열; 부호 없음: 중온성 Cas9. 노드 (nodes)에서 값은 1000-복제 부트스트랩 값 (bootstrap values)을 나타내고; 스케일 바 (scale bar)는 부위 (site)에 대한 추정된 아미노산 치환을 나타낸다.

도 2는, Cas9 유전자 서열의 근린-결합 트리를 나타낸다. 유전자 수준에서 동일성은 극도로 열악하고; 단백질 정렬 (protein alignment)을 위해 사용된 것과 동일한 유기체 유래의 서열은, 유전자 정렬을 위해 사용된다. 유전자 식별자 (gi) 번호는 종명 앞에 표시된다. 범례: 폐쇄 원: 호열성 (최적 60℃ 이상) Cas9 서열, 폐쇄 사각형: 내열성 (최적 <50℃) Cas9 서열, 개방 삼각형: 중온성 기원으로부터 게놈 편집 목적을 위해 현재 가장 많이 사용된 Cas9 서열; 부호 없음: 중온성 Cas9. 노드에서 값은 1000-복제 부트스트랩 값을 나타낸다.

도 3은, 잘-특징화된 타입 (well-characterized Type) II-C (*A. naeslundii*/'ana'; SEQ ID NO: 8) 및 타입 II-A (*S. pyogenes*/'pyo'; SEQ ID NO: 9 및 *S. thermophilus*) Cas9 서열들과 함께 gCas9 (SEQ ID NO: 1) (타입 II-C)에 대한 단백질 서열 정렬을 나타낸다. 중요한 활성 부위 잔기 (active site residues)는 잘 보존되며, 및 검은색 화살표로 표시된다. Ana-Cas9 및 Pyo-Cas9에 대해 기재된 바와 같은 단백질 도메인 (Jinek, et al., 2014, Science 343: 1247997)은, 음영 박스 및 비슷하게 착색된 문자로 표시된다. PAM 인지 도메인

(recognition domain)은, 임의의 타입 II-C 시스템이 아닌, *S. pyogenes* 타입 II-A 시스템에 대해 결정되었고, 따라서, *S. pyogenes* 서열에서 오직 표시된다.

도 4는, *A. naeslundii* Cas9 (Cas9-Ana)의 단백질 아키텍처 (architecture)을 나타낸다 (Jinek et al., 2014). gtCas9는 동일한 타입 II-C CRISPR 시스템에 속하며, 및 활성 부위 잔기는 동정될 수 있다.

도 5는, 상보적인 dsDNA의 crRNA-가이드 표적화의 비교를 나타낸다. 염기쌍 (Base pairing)은 점선으로 표시된다. RNA는 검은색으로, DNA는 회색으로 묘사된다. crRNA 스페이서와 표적 프로토스페이서 사이에 염기쌍은, 두꺼운 검정 점선으로 표시되고, DNA 가닥들 사이 및 RNA 가닥들 사이의 염기쌍은, 두꺼운 회색 점선으로 표시된다. crRNA의 5' 말단은 표시된다. 타입 I에서 PAM (작은 흰색 상자)은 표적 가닥 (프로토스페이서)의 다운스트림에 있는 반면, 타입 II에서는, 치환된 가닥 (displaced strand) 상의 다른 말단에 존재한다. 마찬가지로, 시드 (seed) (표적 DNA 가닥과의 염기쌍이 시작되고, 및 불일치가 허용되지 않는 가이드의 예측된 서열)은, PAM에 가깝게 위치되고, 이로써 타입 I 및 II에서 다르다 (Van der Oost, 2014 *ibid.*). 패널 A는 *E. coli*의 타입 I 캐스케이드 시스템 (Cascade system)의 개략도를 나타낸다. crRNA는, 스템-루프 구조 (stem-loop structure) (머리핀)로 이루어진 29 뉴클레오타이드 (nt) 3' 핸들 (handle) 및 8 nt 5' 핸들에 의해 측면에 위치된, 내부 스페이서 (회색 상자, 표적 인지를 가능하게 하는 31-32 nt)을 갖는다 (Jore 2011 *ibid.*). 패널 B는 *S.pyogenes*의 타입 II Cas9 시스템의 개략도를 나타낸다. tracrRNA와의 crRNA 염기쌍은, RNaseIII에 의한 프로세싱을 가능하게 한다 (대립 검정 삼각형). 부가적으로, crRNA의 5' 말단은, RNase에 의해 잘라내어 (검정 삼각형), 통상적으로 20 nt의 스페이서를 결과한다. 합성 루프 (synthetic loop)가 crRNA 및 tracrRNA를 연결하기 위해 도입될 수 있어, 단일 가이드 RNA (sgRNA)를 결과한다는 점이 주목된다 (Jinek et al., 2012).

도 6은, *G. 썬모디니트리피칸스* (*G.thermodenitrificans*) T12 타입 IIc CRISPR 시스템의 서열의 정렬을 나타낸다.

도 7은, gtCas9에 대한 인실리코 (*in silico*) PAM 예측을 제공하기 위해 얻어진 6개의 단일 히트 (single hits)를 나타낸다.

도 8은, 도 7에 예시된 정렬의 결과를 조합한 웹로그 (weblogo)를 나타낸다. 상기 웹로그는 weblogo.berkeley.edu를 사용하여 발생된다.

도 9는, 정제된 gtCas9로 플라스미드를 표적화하는 60°C에서 시험관 내 절단 분석 (cleavage assay)의 결과를 나타낸다. 플라스미드는 PAM 서열의 특이적 8개 뉴클레오타이드-길이 서열 변이체를 포함한다.

도 10은, CCCCCCAA [SEQ ID NO: 11] PAM 서열로 표적화된 플라스미드를 사용하여, gtCas9 농도의 효과를 조사하기 위한 시험관 내 분석의 결과를 나타낸다.

도 11은, 다양한 온도에 걸쳐 CCCCCCAA [SEQ ID NO:11] PAM 서열로 표적화된 플라스미드를 사용한 시험관 내 분석의 결과를 나타낸다.

도 12는, 실시 예 9에서 설명된 바와 같은, 선택 플레이트 (selection plates) 상에 바실러스 스미시이 (*Bacillus smithii*) ET138 세포의 콜로니의 성장 또는 부재에 의한, gtCas9 및 8nt PAM 서열을 사용한 바실러스 스미시이 ET138 세포의 생체 내 게놈 편집의 결과를 나타낸다. 콜로니는 도 12에서 화살표로 나타낸다.

도 13은, pyrF 유전자가 결손된 콜로니에 대한 PCR 스크린의 결과를 나타낸다. 콜로니는 구축물 3 (음성 대조군)으로 바실러스 스미시이 ET138 세포의 형질전환 (transformation) 후에 발생된다. 15개의 콜로니는 스크리닝되지만, 실시 예 9에서 설명된 바와 같이, 결손 유전자형 -2.1kb 밴드 크기 (deletion genotype -2.1 kb band size)를 나타내지 않았으며, 대신에 야생형 -2.9kb 밴드 크기를 모두 나타낸다.

도 14는, pyrF 유전자가 결손된 콜로니에 대한 PCR 스크린의 결과를 나타낸다. 콜로니는 구축물 1 (PAM 서열 ATCCCCAA [SEQ ID NO: 21])로 바실러스 스미시이 ET138 세포의 형질전환 후에 발생된다. 20개의 콜로니는 스크리닝되고, 실시 예 9에서 설명된 바와 같이, 하나는 결손 유전자형 -2.1 kb 밴드 크기를 나타내는 반면, 나머지는 야생형 -2.9 kb 밴드 크기 및 결손 유전자형 -2.1 kb 밴드 크기 모두를 나타낸다. 야생형 만의 유전자형은 관찰되지 않았다.

도 15는, 지오바실러스 썬모데니트리피칸스 T12 타입-IIc CRISPR-Cas 유전자좌가 열안정성 Cas9 동족체 (homolog)인, ThermoCas9를 인코딩한 것을 나타낸다.

(A) ThermoCas9를 인코딩하는 게놈 유전자좌의 개략적 표현. ThermoCas9의 도메인 아키텍처는, 예상된 활성 부

위 잔기가 적색으로 강조된, 서열 비교에 기초한다. Phyre 2 (Kelley *et al. Nat. Protoc.* 10, 845-858 (2015))를 사용하여 발생된 ThermoCas9의 상동성 모델은, 도메인에 대해 다른 색상으로, 나타낸다.

(B) ThermoCas9와 매우 동일한 Cas9 상동유전자 (orthologue)의 계통수 (Phylogenetic tree). 진화적 분석은 MEGA7에서 수행된다 (Kumar *et al. Mol. Biol. Evol.* 33, 1870-1874 (2016)).

(C) 금속-친화성 크로마토그래피 및 겔 여과에 의한 정제 후 ThermoCas9의 SDS-PAGE. 얻어진 단일 밴드의 이동 (migration)은, apo-ThermoCas9의 이론적 분자량 126 kD와 일치한다.

도 16은, ThermoCas9 PAM 분석을 나타낸다.

(A) 프로토스페이서 인접 모티프 (PAM)의 위치 및 동일성 (5'-NNNNNN-3')을 밝혀내기 위한 시험관 내 절단 분석을 예시하는 개략도. 검은 삼각형은 절단 위치를 나타낸다.

(B) 표적 라이브러리의 ThermoCas9-계 절단의 비교 분석에 의해 얻어진, ThermoCas9의 공통 (consensus) 7nt 길이 PAM의 서열 로고. 각 위치에서의 문자 높이는 정보량 (information content)으로 측정된다.

(C) 시험관 내 절단 분석에 의해 8th 위치로 PAM 동일성의 확장. 별개의 5'-CCCCCAN-3' PAM을 각각 함유하는, 4개의 선형화된 플라스미드 표적은, ThermoCas9 및 sgRNA와 함께 55℃에서 1시간 동안 배양한 후, 아가로스 겔 전기영동에 의해 분석된다.

(D) 30℃ 및 55℃에서 다른 PAMs을 갖는 DNA 표적에 대한 시험관 내 절단 분석. 하나의 별개의 5'-CCCCNNA-3' [SEQ ID NO: 13] PAM을 각각 함유하는, 16개의 선형화된 플라스미드 표적은, ThermoCas9 및 sgRNA와 함께 배양한 후, 아가로스 겔 전기영동에 의해 절단 효율에 대해 분석된다. 또한, 도 21을 참조.

도 17은, ThermoCas9가 넓은 온도 범위에서 활성이고, sgRNA에 결합시 이의 열안정성이 증가되는 것을 나타낸다.

(A) sgRNA 및 메칭 표적 (matching target) DNA의 도식적 표현. 표적 DNA는 검정 외곽선의 직사각형으로 나타내며, PAM은, 후면 외곽선의, 어두운 회색의 수평 타원형으로 나타낸다. crRNA는 검정 외곽선의 어두운 회색 직사각형으로 나타나며, crRNA의 3'-말단이 tracrRNA의 5'-말단과 연결된 부위는, 검정의, 수직 타원형으로 나타낸다. 흰 문자가 있는 검정 박스 및 검은 문자가 있는 밝은 회색 박스는, 각각 tracrRNA의 3'-측에서 예측된 3개 및 2개의 루프를 나타낸다. crRNA의 상보적인 3'-말단 및 tracrRNA의 5'-말단에 의해 형성된 -반복/안티-반복 영역의 41-nt 절단 (truncation)은, 길고, 밝은 회색의, 수직, 점선으로 표시된다. 처음 tracrRNA 루프의 예견된 3' 위치는, 검정 삼각형 및 검정 점선으로 표시된다. 2번째 tracrRNA 루프의 예견된 3' 위치는, 흰색 삼각형 및 검정 점선으로 표시된다. 3번째 tracrRNA 루프의 예견된 3' 위치는, 흰색 삼각형 및 흰색 점선으로 표시된다.

(B) tracrRNA 스캐폴드 (scaffold)의 예견된 3개의 스템-루프의 중요성은, sgRNA의 절단된 변이체를 전사하고, 다양한 온도에서 표적 DNA를 절단하는데 ThermoCas9를 가이드하는 이들의 능력을 평가하여 시험된다. 적어도 둘의 생물학적 반복수 (replicates)의 평균값은, S.D를 나타내는 오차 막대 (error bars)와 함께, 나타낸다.

(C) 최대 온도를 확인하기 위해, ThermoCas9:sgRNA RNP 복합체의 엔도뉴클레아제 활성 (endonuclease activity)은, 60℃, 65℃ 및 70℃에서 5분 또는 10분 동안 배양한 후 분석된다. 사전-가열된 DNA 기질은 첨가되고, 반응은 상응하는 온도에서 1 시간 동안 배양된다.

(D) 활성 분석에 의한 ThermoCas9 및 SpCas9의 활성 온도 범위의 비교는, 지시된 온도에서 5분 배양 후 수행된다. 사전-가열된 DNA 기질은 첨가되고, 반응은 동일한 온도에서 1시간 동안 배양된다.

도 18은, 호열성 미생물에서 ThermoCas9-계 계놈 공학을 나타낸다.

(A) 기본 pThermoCas9_Δ 관심의 유전자 (gene-of-interest: goi) 구축물의 도식적 개요. thermocas9 유전자는, pNW33n (B. smithii) 또는 pEMG (P. putida) 벡터에 도입된다. 상동 재조합 플랭크는, 업스트림 thermocas9에 도입되고, 표적 계놈에서 관심의 유전자 (goi)의 1kb (B.smithii) 또는 0.5kb (P.putida) 업스트림 및 1kb 또는 0.5kb 다운스트림 영역을 포괄한다. sgRNA-발현 모델은, thermocas9 유전자의 다운스트림에 도입된다. 복제 기점 (ori), 복제 단백질 (rep), 항생제 내성 마커 (AB) 및 가능한 부수적 인자 (accessory elements: AE)가 특이적 백본이므로, 이들은 점의 외곽선으로 나타낸다.

(B) 아가로스 겔 전기영동은 B.smithii ET 138의 계놈으로부터 ThermoCas9-계 pyrF 결손 과정으로부터의 10개

의 콜로니에 대한 계능-특이 PCR로부터 결과하는 생성물을 나타낸다. 모든 10개의 콜로니는 Δ pyrF 유전자형을 함유하고, 하나의 콜로니는 야생형 생성물이 걸썩된, 세정 (clean) Δ pyrF 돌연변이체이다.

(C) 기본 pThermoCas9i_goi 구축물의 개략적 개요. 촉매적으로 비활성인 ThermoCas9 (Thermo-dCas9: D8A, H582A 돌연변이)의 발현을 목표로 하여, 상응하는 돌연변이는 도입되어, thermo-dcas9 유전자를 생성시킨다. thermo-dcas9 유전자는 pNW33n 벡터에 도입된다. sgRNA-발현 모듈은, thermo-dcas9의 다운스트림에 도입된다.

(D) 생산, 성장 및 RT-qPCR의 그래픽 표현은 Thermo-dCas9를 사용한 1dhL 침묵 실험으로부터 결과한다. 그래프는 대조군 배양물과 비교하여 억제된 배양물 (repressed cultures)에서 락테이트 생산, 600nm에서의 광학 밀도 및 1dhL 전사의 퍼센트를 나타낸다. 적어도 둘의 생물학적 반복수의 평균값은, S.D를 나타내는 오차 막대와 함께, 나타낸다.

도 19는, 타입 II-A, B 및 C Cas9 상동유전자의 다중 서열 정렬을 나타낸다. Streptococcus pyogenes (Sp), Streptococcus thermophilus (St), Wolinella succinogenes (Ws), Neisseria meningitidis (Nm), Actinomyces naeslundii (An) 및 지오바실러스 써모데니트리피칸스 (Thermo)의 Cas9 단백질 서열은, 기본 설정으로 MEGA7 2의 ClustalW1을 사용하여 정렬되고; ESPrpt3는 시각화를 발생하는데 사용된다. 절대 보존 잔기(strictly conserved residues)는, 회색 바탕에 흰색 텍스트로 나타내고; 유사한 잔기는, 검정 외곽선이 있는 흰색 수직 직사각형에서 검정 텍스트로 나타낸다. 피라미드 (Pyramids)는, 모든 서열에서 2개의 보존된 뉴클레아제 도메인을 나타낸다. 수평 검정 화살표 및 킴 (curls)은, SpCas9 2차 구조 (단백질 데이터베이스 nr 4CMP4)에서, 각각 β -가닥 및 α -헬릭스를 나타낸다. 구조적 도메인은 도 15A에서와 동일한 색채 조합 (colour scheme)을 사용하여 SpCas9 및 ThermoCas9에 대해 표시된다.

도 20은, 인실리코 PAM 결정 결과를 나타낸다. 패널 (A)는 CRISPRtarget6을 사용하여 파지 계능 (phage genomes)으로 얻은 2개의 히트 (hits)를 나타낸다. 패널 (B)는 인실리코 PAM 분석에 의해 얻어진, ThermoCas9의 공통 7nt 길이 PAM의 시퀀스 로고를 나타낸다. 각 위치에서 문자 높이는 정보량으로 측정된다.

도 21은, ThermoCas9 PAM 발견을 나타낸다. 20°C, 37°C, 45°C 및 60°C에서 다른 PAM을 갖는 DNA 표적에 대한 시험관 내 절단 분석. 별개의 5'-CCCCNNA-3' [SEQ ID NO: 13] PAM을 각각 함유하는, 7개 (20°C) 또는 16개 (37°C, 45°C, 60°C)의 선형화된 플라스미드 표적은, ThermoCas9 및 sgRNA와 함께 배양된 후, 아가로즈 겔 전기영동으로 분석된다.

도 22는, 하나의 루프를 함유하는 sgRNA를 사용하여 넓은 온도 범위에서 ThermoCas9의 활성을 나타낸다. tracrRNA 스캐폴드의 예견된 3개의 스템 루프의 중요성은, sgRNA의 절단된 변이를 전사하고, 다양한 온도에서 표적 DNA를 절단하는데 ThermoCas9를 가이드하는 능력을 평가하여 시험된다. 위에서 나타난 것은, 다양한 온도에서 ThermoCas9의 활성에 대한 하나의 루프의 효과이다. 적어도 둘의 생물학적 반복수의 평균값은, S.D를 나타내는 오차 막대 (error bars)와 함께, 나타낸다.

도 23은, ThermoCas9가 2가 양이온을 촉매로 사용하여 dsDNA 표적을 매개하고, ssDNA를 절단하지 않는 것을 나타낸다. 패널 (A)는 EDTA 및 다양한 금속 이온과 함께 ThermoCas9에 의한 시험관 내 플라스미드 DNA 절단을 나타낸다. M = 1kb DNA 래더 (DNA ladder). 패널 (B)는 ssDNA 기질 상에 ThermoCas9의 활성을 나타낸다. M = 10bp DNA 래더.

도 24는, 1dhL 침묵 실험을 위한 스페이서 선택을 나타낸다. 1dhL 침묵 과정 동안 스페이서 (sgRNA)-프로토스페이서 어닐링의 도식적 표현; 선택된 프로토스페이서는 1dhL 유전자의 개시 코돈 (start codon)의 39nt 다운스트림 및 비-주형 가닥 상에 존재한다.

도 25는, pEMG 백본, 슈도모나스 푸티다 (Pseudomonas putida) pyrF 플랭킹 영역 및 thermocas9 유전자 및 sgRNA를 표적으로 하는 슈도모나스 푸티다 pyrF로 이루어진 플라스미드 pThermoCas9_pp Δ pyrF의 지도를 나타낸다.

도 26은, 슈도모나스 푸티다의 계능으로부터 ThermoCas9-계 pyrF 결손 과정으로부터 얻어진 콜로니에 대한 계능 특이 PCR로부터 결과하는 생성물을 나타내는 모세관 겔 전기영동의 결과를 나타낸다. 1854 bp 밴드 및 1112 bp 밴드는, 각각, pyrF 및 Δ pyrF 유전자형에 상응한다.

발명을 실시하기 위한 구체적인 내용

RNA 가이드 및 표적 서열 (RNA Guides and Target Sequences)

[0088]

- [0089] 본 발명의 Cas 단백질은, 상승된 온도에서 표적 핵산의 서열-특이적 결합, 절단, 태깅 (tagging), 표지 또는 변형을 가능하게 한다. 표적 핵산은 DNA (단일-가닥 또는 이중-가닥), RNA 또는 합성 핵산일 수 있다. 본 발명의 특히 유용한 적용은, 게놈 DNA의 표적화된 서열에 상보적으로 결합하는 하나 이상의 가이드 RNA (gRNA)와의 복합체에서 본 발명의 하나 이상의 Cas 단백질에 의한 게놈 DNA의 서열-특이적 표적화 및 변형이다. 결과적으로, 표적 핵산은, 바람직하게는, 이중-가닥 DNA이다. 이러한 표적화는 시험관 내 또는 생체 내에서 수행될 수 있다. 바람직하게는, 이러한 표적화는 생체 내에서 수행된다. 이러한 방식에서, 본 발명의 Cas 단백질은, 인간 세포의 게놈 DNA에 위치한 특이적 DNA 서열을 표적화하고 변형시키는데 사용될 수 있다. Cas 시스템은, 다른 유기체에서 및/또는 다른 유기체의 다양한 세포 타입에서 게놈을 변형시키는데 사용될 수 있는 것으로 예상된다. 선택적으로, 본 발명의 Cas 단백질 및 시스템은, 다양한 인간 세포 타입, 특히 배아 줄기 세포를 제외한 단리된 인간 세포의 게놈을 변형시키는데 사용될 수 있다.
- [0090] 표적화 RNA 분자로 또한 불리는, gRNA는, 폴리뉴클레오티드 표적 가닥 상에 표적 핵산 서열을 인식한다. RNA 분자는 이중 가닥 표적 폴리뉴클레오티드에서 표적 서열을 인식하도록 디자인될 수 있으며, 여기서, 비-표적 가닥은 프로토스페이서 서열의 3' 말단에 바로 인접한 프로토스페이서 인근 모티프 (PAM) 서열을 포함한다. 본 발명의 Cas 단백질 및 폴리펩티드와 최적의 방식으로 작용하는 PAM 서열은 여기에 개시된다. 이들 PAM 서열에 대한 지식으로, gRNA는, 본 발명의 온도 범위 및 증가된 온도에 걸쳐서 본 발명의 Cas 단백질 및 폴리펩티드와 함께 사용하기 위해 디자인될 수 있다.
- [0091] 따라서, 본 발명은, 전술된 바와 같은 본 발명의 Cas 단백질 또는 폴리펩티드를 포함하고, 및 표적 폴리뉴클레오티드의 특정 뉴클레오티드 서열을 인지한다는 점에서 표적화 기능을 갖는 적어도 하나의 RNA 분자를 더욱 포함하는, 리보핵산 단백질 복합체를 제공한다. 본 발명은 또한 표적 핵산 가닥을 결합, 절단, 표지 또는 변형시키기 위한 적어도 하나의 표적화 RNA 분자 및 Cas 단백질 또는 폴리펩티드의 사용, 및 본 발명의 리보핵산 단백질 또는 핵산단백질을 사용하여 표적 핵산 가닥에서 표적 핵산 서열을 결합, 절단, 표지 또는 변형시키는 방법, 뿐만 아니라 Cas 단백질 또는 폴리펩티드 및 표적화 RNA 분자를 갖는 형질전환된 인간 세포를 제공한다. 상기 표적 폴리뉴클레오티드는, 여기에 제공된 PAM 서열에 따라, 프로토스페이서 서열의 3' 말단에 바로 인접한 한정된 PAM 서열을 더욱 포함할 수 있다. 상기 PAM 서열은 길이가 6, 7 또는 8개의 핵산, 또는 그 이상, 바람직하게는, 길이가 8개의 핵산일 수 있다. 바람직하게는, RNA 분자는, 단일-가닥 RNA 분자, 예를 들어, CRISPR RNA (crRNA)이며, 및 예를 들어, tracrRNA와 혼성화 (hybridization)에 의해 결합된다. 상기 표적화 RNA는 crRNA 및 tracrRNA의 키메라 (chimera)일 수 있다. 전술된 RNA 분자는, 표적 뉴클레오티드 서열과 적어도 90%의 동일성 또는 상보성 (complementarity)의 리보뉴클레오티드 서열을 가질 수 있다. 선택적으로, RNA 분자는, 표적 뉴클레오티드 서열과 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성 또는 상보성을 갖는다. 바람직한 표적 뉴클레오티드 서열은 DNA이다. 선택적으로, 본 발명의 적어도 하나의 표적화 RNA 분자 및 Cas 단백질 또는 폴리펩티드를 표적 핵산 가닥을 결합, 절단, 표지 또는 변형시키는데 사용, 및 본 발명의 리보핵산 단백질 또는 핵산단백질을 사용하여 표적 핵산 가닥에서 표적 핵산 서열을 결합, 절단, 표지 또는 변형시키는 방법은, 인간 세포에서 사용될 수 있다. 선택적으로, 인간 세포는 단리될 것이다. 선택적으로, 인간 세포는 배아 줄기 세포가 아니다. 선택적으로, 본 발명의 Cas 단백질 또는 폴리펩티드 또는 리보핵산 단백질 또는 핵산단백질은, 인간의 생식 계열 유전적 동일성을 변형시키는데 사용되지 않을 것이다.
- [0092] 바람직한 관점에서, 본 발명은 전술된 바와 같은 리보핵산 단백질 복합체를 제공하며, 여기서, 적어도 하나의 표적화 RNA 분자는, 이의 길이를 따라 표적 DNA 서열에 실질적으로 상보적이다.
- [0093] 비-표적 가닥 상에 표적 서열 및 PAM 서열을 포함하는, 표적 폴리뉴클레오티드가 본 발명의 핵산단백질 복합체와 결합하고 및 핵산단백질 복합체의 일부를 형성할 수 있도록, 표적화 RNA 분자는 핵산단백질 복합체 내의 표적 서열에 화합되거나 또는 결합될 수 있다.
- [0094] 따라서, 본 발명의 Cas 단백질과 결합하는 RNA 가이드의 서열의 변경 (Alteration)은, Cas 단백질이 가이드 RNA에 상보적인 부위에서 이중-가닥 DNA를 표지하거나 또는 절단하도록 프로그램되는 것을 가능하게 한다.
- [0095] 바람직하게는, 본 발명의 리보핵산 단백질 복합체에서 적어도 하나의 표적화 RNA 분자의 길이는, 35 내지 135 잔기의 범위, 선택적으로 35 내지 134 잔기, 35 내지 133 잔기, 35 내지 132 잔기, 35 내지 131 잔기, 35 내지 130 잔기, 35 내지 129 잔기, 35 내지 128 잔기, 35 내지 127 잔기, 35 내지 126 잔기, 35 내지 125 잔기, 35 내지 124 잔기, 35 내지 123 잔기, 35 내지 122 잔기, 35 내지 121 잔기, 35 내지 120 잔기, 35 내지 119 잔기, 35 내지 118 잔기, 35 내지 117 잔기, 35 내지 116 잔기, 35 내지 115 잔기, 35 내지 114 잔기, 35 내지

113 잔기, 35 내지 112 잔기, 35 내지 111 잔기, 35 내지 100 잔기, 35 내지 109 잔기, 35 내지 108 잔기, 35 내지 107 잔기, 35 내지 106 잔기, 35 내지 105 잔기, 35 내지 104 잔기, 35 내지 103 잔기, 35 내지 102 잔기, 35 내지 101 잔기, 35 내지 100 잔기, 35 내지 99 잔기, 35 내지 98 잔기, 35 내지 97 잔기, 35 내지 96 잔기, 35 내지 95 잔기, 35 내지 94 잔기, 35 내지 93 잔기, 35 내지 92 잔기, 35 내지 91 잔기, 35 내지 90 잔기, 35 내지 89 잔기, 35 내지 88 잔기, 35 내지 87 잔기, 35 내지 86 잔기, 35 내지 85 잔기, 35 내지 84 잔기, 35 내지 83 잔기, 35 내지 82 잔기, 35 내지 81 잔기, 35 내지 80 잔기, 35 내지 79 잔기, 35 내지 78 잔기, 35 내지 77 잔기, 35 내지 76 잔기, 35 내지 75 잔기, 35 내지 74 잔기, 35 내지 73 잔기, 35 내지 72 잔기, 35 내지 71 잔기, 35 내지 70 잔기, 35 내지 69 잔기, 35 내지 68 잔기, 35 내지 67 잔기, 35 내지 66 잔기, 35 내지 65 잔기, 35 내지 64 잔기, 35 내지 63 잔기, 35 내지 62 잔기, 35 내지 61 잔기, 35 내지 60 잔기, 35 내지 59 잔기, 35 내지 58 잔기, 35 내지 57 잔기, 35 내지 56 잔기, 35 내지 55 잔기, 35 내지 54 잔기, 35 내지 53 잔기, 35 내지 52 잔기, 35 내지 51 잔기, 35 내지 50 잔기, 35 내지 49 잔기, 35 내지 48 잔기, 35 내지 47 잔기, 35 내지 46 잔기, 35 내지 45 잔기, 35 내지 44 잔기, 35 내지 43 잔기, 35 내지 42 잔기, 35 내지 41 잔기, 35 내지 40 잔기, 35 내지 39 잔기, 35 내지 38 잔기, 35 내지 37 잔기, 35 내지 36 잔기의 범위 또는 35 잔기이다. 바람직하게는, 적어도 하나의 RNA 분자의 길이는, 36 내지 174 잔기, 37 내지 173 잔기, 38 내지 172 잔기, 39 내지 171 잔기, 40 내지 170 잔기, 41 내지 169 잔기, 42 내지 168 잔기, 43 내지 167 잔기, 44 내지 166 잔기, 45 내지 165 잔기, 46 내지 164 잔기, 47 내지 163 잔기, 48 내지 162 잔기, 49 내지 161 잔기, 50 내지 160 잔기, 51 내지 159 잔기, 52 내지 158 잔기, 53 내지 157 잔기, 54 내지 156 잔기, 36 내지 74 잔기, 37 내지 73 잔기, 38 내지 72 잔기, 39 내지 71 잔기, 40 내지 70 잔기, 41 내지 69 잔기, 42 내지 68 잔기, 43 내지 67 잔기, 44 내지 66 잔기, 45 내지 65 잔기, 46 내지 64 잔기, 47 내지 63 잔기, 48 내지 62 잔기, 49 내지 61 잔기, 50 내지 60 잔기, 51 내지 59 잔기, 52 내지 58 잔기, 53 내지 57 잔기, 54 내지 56 잔기의 범위이다.

[0096] 바람직한 관점에서, 본 발명은 리보핵산 단백질 복합체를 제공하며, 여기서, 적어도 하나의 RNA 분자의 상보적 부분은 길이가 적어도 30 잔기이다. 선택적으로, 적어도 하나의 RNA 분자의 상보적 부분은, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74 또는 75 잔기 길이일 수 있다.

[0097] 표적화 RNA 분자는, 바람직하게는, 표적 핵산 서열에 대해 높은 특이성 및 친화성을 요구할 것이다. 1 μ M 내지 1 pM, 바람직하게는, 1 nM 내지 1 pM; 좀 더 바람직하게는, 1 내지 100pM 범위의 해리 상수 (K_d)는, 비변성 겔 전기영동 (native gel electrophoresis), 또는 선택적으로 등은 적정 열량계, 표면 플라즈몬 공명, 또는 형광계 적정 방법에 의해 결정될 수 있음에 따라 바람직하다. 친화도 (Affinity)는, 겔 지연 분석 (gel retardation assay)으로 또한 불리는, 전기 이동성 교대 분석 (Electromoretic mobility shift assay, EMSA)을 사용하여 결정될 수 있다 (Semenova E et al. (2011) PNAS 108: 10098-10103 참조).

[0098] 표적화 RNA 분자는, 바람직하게는, 원핵생물 내에 자연물 (nature)로부터 CRISPR RNA (crRNA) 분자로 알려진 것으로 모델링된다. crRNA 분자의 구조는, 이미 Jore et al., 2011, Nature Structural & Molecular Biology 18: 529-537에서 보다 상세히 설명되고 확립되어 있다. 간단히 말해서, 타입 I-E의 성숙한 crRNA (mature crRNA)는, 종종 61 뉴클레오타이드 길이이고, 및 8개 뉴클레오타이드의 5' "핸들" 영역, 32개 뉴클레오타이드의 "스페이스" 서열, 및 테트라뉴클레오타이드 루프를 갖는 헤어핀을 형성하는 21개 뉴클레오타이드의 3' 서열로 이루어진다 (도 5). 타입 I 시스템은 타입 II (Cas9)와 다르며 및 다른 시스템의 상세는, Van der Oost 2014 Nat Rev Micr 12: 479-492에 기재되어 있다. 타입 II (Cas9) 시스템에서, 제2 RNA (tracrRNA) 및 두 개의 리보뉴클레아제를 활용하는, 다른 공정 메커니즘이 있다. 머리핀 외에, 타입 II에서 성숙한 crRNA는 tracrRNA의 단편에 부착되어 있다 (도 5). 그러나, 본 발명에서 사용되는 RNA는, 길이, 영역 또는 특이적 RNA 서열에 있어서도, 자연 발생형 crRNA의 디자인으로 절대적으로 디자인될 필요는 없다. 분명한 점은, 본 발명에 사용하기 위한 RNA 분자가, 공개된 데이터베이스에서 유전자 서열 정보에 기초하여 디자인될 수 있거나 또는 새로 발견될 수 있고, 그 다음, 예를 들어, 전체 또는 부분적으로 화학적 합성에 의해 인위적으로 만들어질 수 있다는 것이다. 본 발명의 RNA 분자는 또한 유전적으로 변형된 세포 또는 무세포 발현 시스템 (cell free expression)에서 발현하는 방식으로 디자인 및 생성될 수 있으며, 이 선택은, RNA 서열의 일부 또는 전부의 합성을 포함할 수 있다.

[0099] 타입 II (Cas9)에 crRNA의 구조 및 요건은 또한 Jinek et al., 2012 ibid에 기재되어 있다. 타입 I에서, 스페이스 서열의 5' 말단을 형성하고 및 8 뉴클레오타이드의 5' 핸들에 의해 거기의 5' 측면에 위치된, 소위 "시드 (SEED)" 부분이 있다. Semenova et al. (2011, PNAS 108: 10098-10103)은, 비록 위치 6에서 잔기에 대해, 불일치가 허용될 수 있을지라도, 시드 서열의 모든 잔기가 표적 서열에 상보적이어야 하는 것을 확인하였다 (도 5).

타입 II에서, 스페이스의 3' 말단에 위치한 10-12 뉴클레오타드의 시드가 있다 (도 5) (Van der Oost 2014 ids 에 의해 검토됨). 유사하게, 표적 유전자와 (target locus) (즉, 서열)를 향하는 본 발명의 리보핵산 단백질 복합체의 RNA 성분을 디자인하고, 만드는 경우, 타입 II 시트 서열에 대한 필요한 일치 및 불일치 규칙은 적용될 수 있다.

[0100] 따라서, 본 발명은, 핵산 샘플을 전술된 바와 같은 본 발명의 리보핵산 단백질 복합체와, 또는 전술된 바와 같은 본 발명의 Cas 단백질 또는 폴리펩티드 및 개별 표적화 RNA 성분과 접촉시키는 단계를 포함하는, 표적 핵산 분자에서 단일 염기 변화를 검출 및/또는 위치시키는 방법을 포함하고, 및 여기서 (리보핵산 단백질 복합체에서의 경우를 포함하는) 표적화 RNA의 서열은, 예를 들어, 8 뉴클레오타드 잔기의 연속 서열 (contiguous sequence)의 위치 6에서, 단일 염기 변화로 인해 정상적인 대립 유전자와 돌연변이 대립 유전자 사이에서 구별하도록 하는 것이다.

[0101] 특정 이론에 구속되는 것을 원하지는 않지만, 본 발명의 리보핵산 단백질 복합체의 표적화 RNA 성분을 제조하는데 사용될 수 있는 디자인 규칙은, 이중 가닥 표적 폴리뉴클레오타드에서 소위 "PAM" (protospacer adjacent motif) 서열을 포함한다. E. coli의 타입 I-E 시스템에서, PAM 서열은 5'-CTT-3', 5'-CAT-3', 5'-CCT-3', 5'-CAC-3', 5'-TTT-3', 5'-ATT-3', 및 5'-AWG-3'과 같은, 뉴클레오타드 잔기의 보존된 트리플렛 (conserved triplet)일 수 있고, 여기서 W는 A, T 또는 U이다. 타입 I에서, 표적화된 가닥에 위치한 PAM 서열은, 보통 시드의 5'에 상응하는 위치에 있다. 타입 II에서, 그러나, PAM은, 시드의 3'에 상응하는 위치에서, crRNA 스페이스의 3' 말단에 가깝게 변위된, 또는 비-표적, 가닥 상에, 다른 말단에 위치된다 (도 5) (Jinek et al., 2012, op. cit.). 화농성 연쇄상 구균 (Streptococcus pyogenes) Cas9의 경우, PAM 서열은, 5'-NGG-3'인, 뉴클레오타드 잔기의 보존된 쌍을 갖는다. 최근에, 다른 Cas9 변이체 (타입 IIA 및 타입 IIC) (Ran et al., 2015 Nature 520: 186-191)-Fig 1A)는, 특징화되었고, PAM은 밝혀졌다 (Ran et al., 2015, ibid.-Fig 1C). 현재 입증된 Cas9 PAMs은: 타입 IIA 5'-NGGNNN-3' (Streptococcus pyogenes), 5'-NNGTNN-3' (Streptococcus pasteurianus), 5'-NNGGAAN-3' (Streptococcus thermophilus), 5'-NNGGGNN-3' (Staphylococcus aureus), 및 타입 IIC 5'-NGGNNN-3' (Corynebacterium difteriae), 5'-NNGGGTN-3' (Campylobacter lari), 5'-NNNCATN-3' (Parvobaculum lavamentivorans), 5'-NNNNGTA-3' (Neisseria cinerea)을 포함한다. 지오바실러스 씨모데니트리피칸스 (지오바실러스 씨모데니트리피칸스) T12의 Cas9 (본 발명)는, 타입 IIC에 속한다 (Ran et al., 2015, ibid.). 본 발명자들은, 놀랍게도, 본 발명에 사용하기 위한 PAM 서열의 선택이 본 발명의 Cas 단백질 및 폴리펩티드가 표적 서열과 상호작용하는 온도(들)에 영향을 줄 수 있다는 것을 발견했다. 특히, 본 발명자들은, 표적 서열의 3' 말단 이후에 5th 위치에 시토신, 및/또는 8th 위치에 아데닌으로, 넓은 온도 범위에 걸쳐 활성을 부여하는데 8-mer PAM 서열이 바람직하다는 것을 확인했다. 또한, 프로토스페이스 서열의 3' 말단 이후에 PAM 서열의 1st, 2nd, 3rd, 4th 및/또는 6th 위치에 시토신이 바람직하다.

[0102] 특정 관점에서, 예를 들어, 20℃ 내지 100℃, 20℃ 내지 80℃, 30 내지 80℃, 20℃ 내지 70℃ 또는 25℃ 내지 65℃의 넓은 온도 범위 내에서 표적 서열과의 상호 작용은, 5'-NNNCVAA-3' [SEQ ID NO: 48]의 PAM 서열을 활용하여 달성될 수 있다. 처음 4 PAM 위치에 대해서는 특별히 선호하지 않는다. 따라서, 처음 4개의 뉴클레오타드는 편리하게 임의의 뉴클레오타드 (NNNN)일 수 있다. 바람직하게는, 이러한 넓은 온도 범위 내에서 표적 서열과의 상호 작용은, 5'-NNNCSAA-3' [SEQ ID NO: 49]의 PAM 서열을 활용하여 달성될 수 있다. 최적으로, PAM은 서열 5'-NNNCGAA-3' [SEQ ID NO: 50] 또는 5'-NNNCCAA-3' [SEQ ID NO: 51]일 수 있다.

[0103] 표적 서열과 상호작용이 ≥30℃, 예를 들어, 30℃ 내지 100℃, 바람직하게는 30℃ 내지 70℃, 30℃ 내지 65℃, 또는 45℃ 내지 65℃의 온도 범위에서 요구되는 경우, PAM 서열은 최적으로 서열 5'-NNNCCNAA-3' [SEQ ID NO: 52] 또는 5'-NNNCCMCA-3' [SEQ ID NO: 53]일 수 있다. 처음 4 PAM 위치에 대해서는 특별히 선호하지 않는다. 따라서, 처음 4개의 뉴클레오타드는 편리하게 임의의 뉴클레오타드 (NNNN)일 수 있다. 선택적으로, 예를 들어, PAM 서열은 5'-CCCCNAA-3' 또는 5'-CCCCMCA-3'일 수 있다. 선택적으로, 예를 들어, PAM 서열은, 5'-CCCCAAA-3', 5'-CCCCATA-3', 5'-CCCCAGA-3', 5'-CCCCACA-3', 5'-CCCCCTAA-3', 5'-CCCCCTTA-3', 5'-CCCCCTGA-3', 5'-CCCCCTCA-3', 5'-CCCCCGAA-3', 5'-CCCCCGTA-3', 5'-CCCCCGGA-3', 5'-CCCCCGCA-3', 5'-CCCCCAA-3' [SEQ ID NO: 11], 5'-CCCCCTA-3', 5'-CCCCCGA-3', 또는 5'-CCCCCCA-3'으로부터 선택될 수 있다.

[0104] 본 발명의 구체 예에서, 표적화 RNA 분자는 35-200 잔기의 범위에서 길이를 가질 수 있다. 바람직한 구체 예에서, 원하는 핵산 서열에 표적화를 위해 사용되고, 상보적인 RNA의 부분은, 15 내지 32 잔기 길이이다. 자연-발생 crRNA의 맥락에서, 이것은, 예를 들어, Semenova et al. (2011 ibid.)의 도 1에 나타난 바와 같은 스페이스

부분에 상응할 것이다.

- [0105] 본 발명의 리보핵산 단백질 복합체는, DNA 표적 서열에 실질적인 상보성을 갖는 RNA 서열에 CRISPR 반복 5'으로부터 유래된 8 잔기를 포함하는 표적화 성분을 가질 수 있다. DNA 표적 서열에 대해 상보성을 갖는 RNA 서열은, 스페이서 서열인 것으로서 crRNA의 맥락에서 상응하는 것으로 이해될 것이다. RNA의 5' 플랭킹 서열 (flanking sequence)은; 예를 들어, Semenova et al. (2011 *ibid.*)의 도 1에 나타난 바와 같이, crRNA의 5' 핸들에 상응하는 것으로 고려될 것이다.
- [0106] 본 발명의 리보핵산 단백질 복합체는, DNA 표적 서열에 상보성을 갖는 표적화 RNA 서열의 3'에, 즉, 예를 들어, Semenova et al. (2011 *ibid.*)의 도 1에 나타난 바와 같은, crRNA에서 스페이서 서열 측면에 위치하는 3' 핸들에 상응하는 것의 3'에 헤어핀 및 테트라뉴클레오티드 루프 형성 서열을 가질 수 있다.
- [0107] 특정 이론에 구속되는 것을 원하지는 않지만, 바람직한 리보핵산 단백질 복합체 및 이중 가닥 표적 폴리뉴클레오티드에서, 리보핵산 단백질 복합체의 표적화 RNA와 쌍을 이루지 않는 비-표적 핵산 가닥은, 5'-NNNNCNNA-3' [SEQ ID NO: 47], 5'-CNNCCN-3', 5'-NNCCN-3', 5'-NNCNCN-3', 5'-NNNNCCN-3', 및 5'-NCNCCN-3' 중 하나 이상으로부터 선택된 직접 3'에 인접 PAM 서열을 포함할 수 있다. 선택적으로, 예를 들어, PAM 서열은, 5'-NNNNC-3', 5'-NNNNCNNA-3' [SEQ ID NO: 47], 5'-CNNNC-3', 5'-CNNCCNNA-3', 5'-NCNNC-3', 5'-NCNCCNNA-3', 5'-NNCNC-3', 5'-NNCCNNA-3', 5'-NNNCC-3', 5'-NNCCNNA-3', 5'-NNNNCC-3', 5'-NNNNCCNA-3', 5'-CCNNC-3', 5'-CCNCCNNA-3', 5'-CNCNC-3', 5'-CNCNCCNA-3', 5'-CNNCCN-3', 5'-CNNCCNNA-3', 5'-CNNCC-3', 5'-CNNCCNA-3', 5'-CCCNCN-3', 5'-CCCNCNNA-3', 5'-CCNCCN-3', 5'-CCNCCNNA-3', 5'-CCNCC-3', 5'-CCNCCNA-3', 5'-CCCCC-3' [SEQ ID NO: 12], 5'-CCCCCNNA-3' [SEQ ID NO: 13], 5'-CCCCC-3' [SEQ ID NO: 14], 5'-CCCCCNA-3' [SEQ ID NO: 10], 5'-NCCNC-3', 5'-NCCNCCNA-3', 5'-NCCCC-3', 5'-NCCCCNNA-3', 5'-NCCCCC-3' [SEQ ID NO: 15], 5'-NCCCCCNA-3' [SEQ ID NO: 16], 5'-NNCCC-3', 5'-NNCCCNNA-3', 5'-NNCCCC-3', 5'-NNCCCCNA-3', 5'-NNNCCC-3', 및 5'-NNNCCNNA-3'으로부터 선택될 수 있다. PAM 서열은, 5'-CNCCCCAC-3' [SEQ ID NO: 17], 5'-CCCCCAG-3' [SEQ ID NO: 18], 5'-CCCCCAA-3' [SEQ ID NO: 11], 5'-CCCCCAT-3' [SEQ ID NO: 19], 5'-CCCCCAC-3' [SEQ ID NO: 20], 5'-ATCCCCAA-3' [SEQ ID NO: 21], 또는 5'-ACGGCCAA-3' [SEQ ID NO: 22]일 수 있다. 바람직하게는, PAM 서열은 서열 5'-NNNNCNNA-3' [SEQ ID NO: 47]일 것이다. 그러나, 뉴클레오티드의 다른 조합이 Cas 단백질 또는 폴리펩티드의 농도 및/또는 원하는 적용에 의존하여 사용될 수 있는 것으로 인정될 것이다. 특히, 처음 4 PAM 위치에 대한 특별한 선호도는 없다. 따라서, 처음 4개의 뉴클레오티드는, 편리하게 임의의 뉴클레오티드 (NNNN)일 수 있다. 이들 서열은, 자연적으로 발생하는 crRNA의 맥락에서 "프로토스페이서 인근 모티프" 또는 "PAM"으로 불리는 것에 상응한다. 타입 II CRISPR/Cas 시스템에서, 이들 PAM 서열은, 표적 서열에 대한 - 자연 시스템 (natural system)에서 표적, 및 따라서 바람직하게는, 또한 본 발명에 따른 RNAs의 표적 모두에서 - crRNA의 고도의 특이성을 보장하기 위해, 캐스케이드/crRNA 복합체와 이의 dsDNA 표적과의 안정한 상호작용을 가능하게 한다. 바람직하게는, 프로토스페이서에 바로 인접한 서열은, 5'-NNNCATN-3'이 아닐 것이다.
- [0108] 부가적으로, PAM 서열은, 서열 5'-NNNNCNNA-3' [SEQ ID NO: 47], 예를 들어 5'-NNNNCNAA-3' [SEQ ID NO: 52], 또는 5'-NNNNCMCA-3' [SEQ ID NO: 53]일 수 있다.
- [0109] 중온성 SpCas9의 한계 중 하나는, 25 내지 44℃에서 활성을 오직 나타낸다는 것이고; 이들 온도를 벗어나면, SpCas9 활성은 감지할 수 없는 수준으로 급격히 감소한다 (Mougiakos et al., 2017, *ACS Synth Biol.* 6:849-861). 이의 중온성 상동유전자 SpCas9의 25-44℃ 범위와 대조적으로, 본 발명의 ThermoCas9는, 시험관 내에서 20-70℃의 훨씬 넓은 온도 범위에서 활성이다. ThermoCas9의 확장된 활성 및 안정성은, 20-70℃의 온도에서 DNA 조작을 필요로 하는 분자 생물학 기술에서 이의 적용뿐만 아니라, 강력한 효소 활성을 필요로 하는 가혹한 환경에서 이의 활용도 가능하게 한다. 따라서, ThermoCas9는 또한 인간 세포에 대한 게놈 편집 도구로써 사용될 수 있다. 선택적으로, 상기 인간 세포는 배아 줄기 세포가 아닐 것이다.
- [0110] 넓은 기능성 온도 활성 범위를 갖는 것, 즉, 저온 및 고온 모두에서, 예를 들어, 20℃ 및 70℃, 또는 20℃ 및 65℃ 또는 25℃ 및 65℃ 모두에서, 기능성인 것에 부가하여, ThermoCas9 또는 (예를 들어, sgRNA 또는 tracrRNA와 같은) 관련 인자의 구조적 특색 (features)을 변형시켜, ThermoCas9이 표적화된 절단 또는 결합할 수 있거나 또는 표적화된 절단 또는 결합이 효율적으로 일어나는, 온도 범위를 조작하기 위한 능력은, 핵산 서열 조작을 넘어 발휘되는 더 큰 수준의 제어를 가능하게 할 것이다. 그러나, 지금까지, 특정 온도에서 Cas9 활성의 분자 결정요인 (molecular determinants)에 대해서 알려진 것은 거의 없다.
- [0111] 본 발명자들은 ThermoCas9의 열안정성을 부여하는데 중요한 몇 가지 요인을 확인하였으며, 그 하나는

ThermoCas9의 PAM 선호도 (PAM preferences)이다. ThermoCas9의 PAM 선호도는, 온도 범위의 낮은 부분 ($\leq 30^{\circ}\text{C}$)에서의 활성에 대해 매우 엄격한 반면, PAM의 더 많은 다양성은 보통 내지 최적 온도 (37°C 내지 60°C)에서 활성을 가능하게 한다. 이로써, PAM 서열은, 정해진 온도에서 표적의 가장 효율적인 결합, 절단, 표지 또는 변형을 얻기 위해 변경될 수 있다. 이는, 특정 적용에 의존하여, ThermoCas9의 적용에 상당한 유연성을 제공한다. 예를 들어, 몇몇 적용에서, 매우 넓은 온도 범위, 예를 들어, 20°C 내지 70°C , 바람직하게는, 20°C 내지 65°C 또는 25°C 내지 65°C 에서, 표적 결합, 절단, 표지 또는 변형은 바람직할 수 있다. 이러한 넓은 온도 범위 내에서 표적 서열의 결합, 절단, 표지 또는 변형은, 5'-NNNNCVAA-3' [SEQ ID NO: 48]의 PAM 서열을 활용하여 달성될 수 있다. 바람직하게는, 이러한 넓은 온도 범위 내에서 표적 서열의 결합, 절단, 표지 또는 변형은, 5'-NNNNCSAA-3' [SEQ ID NO: 49], 예를 들어, 5'-NNNNCGAA-3' [SEQ ID NO: 50] 또는 5'-NNNNCCAA-3' [SEQ ID NO: 51]의 PAM 서열을 활용하여 달성될 수 있다. 처음 4 PAM 위치에 대해서 특이적 선호도는 없다. 따라서, 처음 4개의 뉴클레오티드는, 편리하게 임의의 뉴클레오티드 (NNNN)일 수 있다. 선택적으로, 예를 들어, 5'-CCCCGAA-3' 또는 5'-CCCCCAA-3' [SEQ ID NO: 11]일 수 있다.

[0112] 표적의 결합, 절단, 표지 또는 변형이 $\geq 30^{\circ}\text{C}$, 예를 들어, 30°C 내지 100°C , 바람직하게는 30°C 내지 70°C , 30°C 내지 65°C , 또는 45°C 내지 65°C 의 범위에서 요구되는 경우, PAM 서열은 최적으로 서열 5'-NNNNCNAA-3' [SEQ ID NO: 52] 또는 5'-NNNNCMCA-3' [SEQ ID NO: 53]일 수 있다. 처음 4 PAM 위치에 대해서는 특별히 선호하지 않는다. 따라서, 처음 4개의 뉴클레오티드는 편리하게 임의의 뉴클레오티드 (NNNN)일 수 있다. 선택적으로, 예를 들어, PAM 서열은 5'-CCCCNAA-3' 또는 5'-CCCCMCA-3'일 수 있다. 선택적으로, 예를 들어, PAM 서열은 5'-CCCCAAA-3', 5'-CCCCATA-3', 5'-CCCCAGA-3', 5'-CCCCACA-3', 5'-CCCCCTAA-3', 5'-CCCCCTTA-3', 5'-CCCCCTGA-3', 5'-CCCCCTCA-3', 5'-CCCCCGAA-3', 5'-CCCCCGTA-3', 5'-CCCCCGGA-3', 5'-CCCCCGCA-3', 5'-CCCCCAA-3' [SEQ ID NO: 11], 5'-CCCCCTA-3', 5'-CCCCCGA-3', 또는 5'-CCCCCCA-3'으로부터 선택될 수 있다.

[0113] 여기에 제공된 본 발명의 PAM 서열은, 여기서 개시된 서열, 예를 들어, 6-mer, 7-mer 또는 8-mer 서열을 포함한다. 상기 6-mer, 7-mer 또는 8-mer 서열은, 표적화 RNA에 의해 결합된 것에 상보적인, 프로토스페이서 서열과, PAM 서열의 5' 말단 사이에 공간을 차지하는 부가적인 핵산이 없이, 비-표적 가닥 상에 프로토스페이서 서열의 3'을 바로 시달할 수 있다. 그러나, 상기 6-mer, 7-mer 또는 8-mer 서열의 3' 말단에서 PAM 서열의 일부를 형성하는 부가적인 핵산일 수 있는 것으로 인정될 것이다. 부가적으로 또는 선택적으로, 비-표적 가닥은, PAM 서열의 3'에 부가적인 핵산을 포함할 수 있다.

[0114] 본 발명의 핵산단백질 복합체는, 본 발명의 리보핵산 단백질 복합체 및 상기 리보핵산 단백질이 결합되는, 핵산의 표적 핵산 가닥을 포함할 수 있다.

[0115] 결합, 절단, 표지 및 변형 온도

[0116] 본 발명의 Cas 단백질의 활성, 예를 들어, 뉴클레아제 활성의 최적 온도 범위를 포함하는, 온도 범위는, 공지된 Cas9 단백질의 온도 범위보다 상당히 높다. 또한, 활성을 유지하는 상한 범위는 공지된 Cas9 단백질의 것보다 훨씬 높다. 더 높은 최적의 온도 및 기능적 범위는, 고온에서의 유전 공학에 상당한 장점을 제공한다. 따라서, 본 발명의 방법, 사용, 핵산단백질 및 형질전환된 인간 세포는, 상승된 온도에서 게놈 편집에 유용할 수 있다. 비-표적 가닥의 프로토스페이서 서열에 바로 인접한, 본 발명의 PAM 서열의 존재는, 표적 서열에 대한 Cas 단백질 및 폴리펩티드의 특이성을 개선하고, 및 더 높은 온도에서 및 더 넓은 기능적 온도 범위에서 걸쳐 Cas 단백질 및 폴리펩티드의 사용을 지원한다.

[0117] 현저히 더 큰 열안정성에 따라, 본 발명의 Cas 단백질은, 공지된 Cas9 단백질의 것보다 훨씬 더 넓은 온도 범위에 걸쳐서, 기능, 예를 들어, 뉴클레아제 활성을 보유한다. 또한, 이것이 활성을 보유하는 범위의 상한은, 공지된 Cas9 단백질의 것보다 훨씬 높다. 더 높은 최적의 온도 및 기능적 범위는, 고온에서의 유전 공학에서, 예를 들어, 상승된 온도에서 인간 세포의 게놈 편집에서 상당한 장점을 제공한다. 선택적으로, 배아 줄기 세포의 공학 기술을 포함하지 않는다. ThermoCas9의 확장된 활성 및 안정성은, 넓은 온도 범위, 예를 들어, $20-70^{\circ}\text{C}$ 에서의 DNA 조작을 필요로 하는 분자 생물학 기술에서 적용뿐만 아니라, 강력한 효소 활성을 필요로 하는 가혹한 환경에서 이의 활용을 가능하게 한다.

[0118] 유리하게는, 본 발명자들은 또한 본 발명의 Cas 단백질이 표적 서열의 전사적 제어, 예를 들어, 표적 서열에 대한 서열-특이적 결합에 의한 침묵 전사 (silencing transcription)을 지시하는데 사용될 수 있다는 것을 알아냈다. 따라서, ThermoCas9는 또한 인간 세포에서, 예를 들어, 표적 유전자의 전사를 침묵 또는 활성화하는데, 전사적 제어 도구로써 사용될 수 있다. 따라서, ThermoCas9는 또한 인간 세포에서 유전자-침묵 도구로써 사용될

수 있다. 선택적으로, 본 발명의 Cas 단백질은, 배아 줄기 세포에서 사용되지 않는다. 선택적으로, 본 발명의 Cas 단백질은, 인간의 생식 계열 유전적 동일성을 변경하는데 사용되지 않는다.

[0119] 유리하게는, 본 발명의 Cas 단백질 또는 폴리펩티드는, 20℃ 내지 100℃의 온도에서 핵산 결합, 절단, 표지 또는 변형할 수 있지만, 상승된 온도에서, 예를 들어, 41℃ 내지 122℃, 바람직하게는, 50℃ 내지 100℃의 온도에서 특히 유용하다. 본 발명의 Cas 단백질 및 폴리펩티드는, DNA, RNA 및 합성 핵산을 결합, 절단, 표지 또는 변형시킬 수 있다. 본 발명의 Cas 단백질 또는 폴리펩티드는 또한, 예를 들어, 20 내지 50℃ 범위의 온도에서 뉴클레아제 활성, 유전자 편집 및 핵산 표지 적용에 대한 조작성 (operability)을 제공할 수 있다.

[0120] 온도 범위가 여기에 포함되는 경우, 말단점이 개시된 온도 범위에 포함, 즉 범위가 "포괄하는" 것으로 의도된다. 예를 들어, 20℃ 내지 100℃ 범위의 온도에서 활성이 있는 것으로 명시된 경우, 20℃ 및 100℃의 온도는 상기 범위에 포함된다.

[0121] 바람직하게는, 본 발명의 Cas 단백질 또는 폴리펩티드는, 결합, 절단, 표지 또는 변형될 폴리뉴클레오티드 분자(들)에서 표적 서열을 인식하는 적절한 gRNA (가이드 RNA, 또한 표적화 RNA 분자라 함)와 결합되는 경우, 20℃ 내지 100℃, 선택적으로 20℃ 내지 70℃, 20℃ 내지 65℃, 25℃ 내지 70℃, 25℃ 내지 65℃, 55℃ 내지 100℃, 50℃ 내지 70℃, 50℃ 내지 70℃, 55℃ 내지 70℃, 또는 55℃ 내지 65℃ 범위의 온도에서 결합, 절단, 표지 또는 변형시킨다.

[0122] 바람직하게는, 본 발명의 Cas 단백질 또는 폴리펩티드는, 결합, 절단, 표지 또는 변형될 폴리뉴클레오티드 분자(들)에서 표적 서열을 인식하는 적절한 gRNA (가이드 RNA, 또한 표적화 RNA 분자라 함)와 결합되는 경우, 50℃ 내지 100℃, 선택적으로 55℃ 내지 100℃, 60℃ 내지 100℃, 65℃ 내지 100℃, 70℃ 내지 100℃, 75℃ 내지 100℃, 80℃ 내지 100℃, 85℃ 내지 100℃, 90℃ 내지 100℃, 95℃ 내지 100℃ 범위의 온도에서 결합, 절단, 표지 또는 변형시킨다. 좀 더 바람직하게는, 본 발명의 Cas 단백질은, 51℃ 내지 99℃, 52℃ 내지 98℃, 53℃ 내지 97℃, 54℃ 내지 96℃, 55℃ 내지 95℃, 56℃ 내지 94℃, 57℃ 내지 93℃, 58℃ 내지 92℃, 59℃ 내지 91℃, 60℃ 내지 90℃, 61℃ 내지 89℃, 62℃ 내지 88℃, 63℃ 내지 87℃, 64℃ 내지 86℃, 65℃ 내지 85℃, 66℃ 내지 84℃, 67℃ 내지 83℃, 68℃ 내지 82℃, 69℃ 내지 81℃, 70℃ 내지 80℃, 71℃ 내지 79℃, 72℃ 내지 78℃, 73℃ 내지 77℃, 74℃ 내지 76℃ 범위의 온도, 또는 75℃의 온도에서 핵산을 결합, 절단, 표지 또는 변형시킨다. 바람직하게는, 본 발명의 Cas 단백질은, 60℃ 내지 80℃, 61℃ 내지 79℃, 62℃ 내지 78℃, 63℃ 내지 77℃, 64℃ 내지 76℃, 60℃ 내지 75℃, 60℃ 내지 70℃의 온도 범위에서 핵산을 결합, 절단, 표지 또는 변형한다. 최적으로, 본 발명의 Cas 단백질은, 60℃ 내지 65℃의 범위, 바람직하게는, 65℃의 온도에서 핵산을 결합, 절단, 표지 또는 변형시킨다.

[0123] 표적화 RNA 분자는, 본 발명의 Cas 단백질 및 폴리펩티드와 함께 사용하도록 디자인될 수 있고, 여기서, 상기 표적화 RNA 분자는 표적 가닥에서 표적 서열에 결합되며, 비-표적 가닥은 프로토스페이서 서열의 3' 바로 옆에 여기에 제공된 PAM 서열을 더욱 포함한다. 상기 PAM 서열은, 5'-NNNNNNA-3', 바람직하게는, 5'-NNNNCNNA-3' [SEQ ID NO: 47], 선택적으로, 예를 들어, 5'-CCCCCNA-3' [SEQ ID NO: 10] 또는 5'-CCCCCAA-3' [SEQ ID NO: 11]을 포함할 수 있고, 및 본 발명의 사용, 방법, 형질전환된 세포, 및 핵산단백질은, 55℃ 내지 65℃의 온도 범위에 걸쳐, 바람직하게는, 50℃ 내지 70℃, 40℃ 내지 65℃, 45℃ 내지 75℃, 37℃ 내지 78℃ 및/또는 20℃ 내지 80℃의 온도 범위에 걸쳐, 표적 가닥의 결합, 절단, 표지 및/또는 변형을 제공할 수 있다.

[0124] PAM 서열은, 주어진 온도에서 표적의 가장 효율적인 절단을 얻기 위해 변경될 수 있다. 이는, 특정 적용에 의존하여, 본 발명의 Cas 단백질의 적용에 큰 폭의 유연성을 제공한다. 20℃ 내지 100℃, 바람직하게는, 20℃ 내지 70℃, 또는 20℃ 내지 65℃, 또는 25℃ 내지 65℃의 온도 범위 내에서 결합, 절단, 표지 또는 변형 활성, 예를 들어 절단 활성이 요구되는 경우, 그 다음 활성은, 5'-NNNNVAA-3' [SEQ ID NO: 48]의 PAM서열을 활용하여 달성될 수 있고, 바람직하게는, 이러한 온도 범위 내에서 활성은, 5'-NNNNCSAA-3' [SEQ ID NO: 49], 예를 들어, 5'-NNNNCGAA-3' [SEQ ID NO: 50] 또는 5'-NNNNCCAA-3' [SEQ ID NO: 51]의 PAM 서열을 활용하여 달성될 수 있다. 선택적으로, 활성은, 예를 들어, 5'-CCCCGAA-3' [SEQ ID NO: 52] 또는 5'-CCCCCAA-3' [SEQ ID NO: 11]의 PAM 서열을 활용하여 달성될 수 있다.

[0125] 본 발명자들은 ThermoCas9의 열안정성이 가이드 (sgRNA)의 결합과 함께 증가하여 리보핵단백질 복합체를 형성한다는 것을 확인했다. 가이드 (sgRNA)는, tracrRNA 및 crRNA를 적절하게 포함할 수 있다. 이러한 배열에서, 가이드는, 뉴클레오티드 스페이서-단편 및 반복-단편을 포함하는 crRNA를 적절하게 포함할 수 있다. crRNA는, 17-20nt의 길이가 적절할 수 있다. 선택적으로, crRNA는 길이가 17nt일 수 있다. 선택적으로, crRNA는 길이가 18nt, 길이가 19nt 또는 길이가 20nt일 수 있다. 가이드는 또한 tracrRNA (안티-반복 단편 (crRNA의 반복 단편

과 염기쌍))을 포함할 수 있다. tracrRNA 및 crRNA는 합성 링커 (synthetic linker)에 의해 분리될 수 있다. 다음의 가이드는 바람직한 배열을 나타낸다: 5'-[crRNA (17-20 뉴클레오티드 스페이서-단편 및 반복-단편) - (선택적: 2개의 RNAs를 연결하는 합성 루프) - tracrRNA (안티-반복 단편 (crRNA의 반복 단편과 염기쌍) 및 일부 가변 스템-루프 구조 (아래 참조), 이는 몇몇 시스템에서는 어느 정도 절단될 수 있음)]-3'.

[0126] 일반적으로, tracrRNA는, 예를 들어, crRNA 및 tracrRNA 포함하는, 키메라 단일-가이드 RNA (sgRNA)의 일부로서 제공될 것이다. tracrRNA는 하나 이상의 헤어핀 구조, 바람직하게는 둘 이상의 헤어핀 구조 또는 좀 더 바람직하게는 셋 이상의 헤어핀 구조가 수반되는 안티-반복 영역으로 이루어질 수 있다. 스페이서 원위 단부 (spacer distal end)에서 (4-뉴클레오티드 링커, 예를 들어, 5'-GAAA-3'에 의해 융합된 합성 sgRNA 키메라에서 상보적인 tracrRNA 부분 (안티-반복)의 5'말단 및 crRNA 부분 (반복)의 3'말단에 의해 형성된) 전체-길이의 반복/안티-반복 헤어핀의 존재는, 뉴클레아제에 앵커 (anchor)로써 기능하지만, 표적 선택 및 절단 활성화에 필수적인 것은 아니다. 예를 들어, tracrRNA 부분의 최대 50-nt 길이의 결손의 스페이서 원위 단부에서 결손은, DNA 절단 효율에 거의 또는 전혀 영향을 없이 허용될 수 있다. 따라서, 예를 들어, 전체-길이 반복-안티-반복 헤어핀의 스페이서 원위 단부의 결손은, 표적 DNA 절단 효율의 면에서 타협 없이, 최대 50nt, 최대 45nt, 최대 40nt, 최대 35nt, 최대 30nt, 최대 25nt, 최대 20nt, 최대 15nt, 최대 10nt, 또는 최대 5nt로 이루어질 수 있다.

[0127] 놀랍게도, 본 발명자들은 또한 tracrRNA의 구조가 ThermoCas9의 열안정성 및 활성, 특히 절단 활성의 효율에 영향을 미친다는 것을 확인했다. 구체적으로, tracrRNA 또는 sgRNA에서 헤어핀 (또는 스템-루프) 구조의 수는, 주어진 온도에서 표적의 가장 효율적인 결합, 절단, 표지 또는 변형을 얻기 위해 변경될 수 있다. 이는, 특정 적용에 의존하여, 본 발명의 Cas 단백질의 적용에 큰 폭의 유연성을 제공한다. 선택적으로, tracrRNA 또는 sgRNA는, 하나 이상의 스템-루프 구조, 둘 이상의 스템-루프 구조 또는 셋 이상의 스템-루프 구조를 형성할 수 있는 핵산 서열과 함께 제공될 수 있다. 선택적으로, tracrRNA 또는 sgRNA는, 하나 이상의 스템-루프 구조, 둘 이상의 스템-루프 구조 또는 셋 이상의 스템-루프 구조를 형성하도록 배열된 핵산 서열과 함께 제공될 수 있다. 바람직하게는, sgRNA는, 적어도 3개의 스템-루프 구조를 형성할 수 있는 핵산 서열과 함께 제공될 것이다.

[0128] 선택적으로, 20℃ 내지 60℃, 바람직하게는 37℃ 내지 60℃의 온도 범위, 또는 37℃, 40℃, 45℃, 50℃, 55℃ 또는 60℃에서 결합, 절단, 표지 또는 변형 활성화, 예를 들어, 절단 활성이 요구되는 경우, 그 다음 활성은, 하나 이상의 스템-루프 구조를 형성할 수 있는 sgRNA 서열을 활용하여 달성될 수 있다.

[0129] 선택적으로, 20℃ 내지 65℃, 바람직하게는 37℃ 내지 65℃, 좀 더 바람직하게는 45℃ 내지 55℃의 온도 범위 또는 37℃, 40℃, 45℃, 50℃, 55℃ 또는 60℃에서 결합, 절단, 표지 또는 변형 활성화, 예를 들어, 절단 활성이 요구되는 경우, 그 다음 활성은, 2 이상의 스템-루프 구조를 형성할 수 있는 sgRNA 서열을 활용하여 달성될 수 있다.

[0130] 선택적으로, 20℃ 내지 100℃, 바람직하게는 20℃ 내지 70℃, 좀 더 바람직하게는 37℃ 내지 65℃의 온도 범위 또는 37℃, 40℃, 45℃, 50℃, 55℃, 60℃ 또는 65℃에서 결합, 절단, 표지 또는 변형, 예를 들어, 절단 활성이 요구되는 경우, 그 다음 활성은, 3 이상의 스템-루프 구조를 형성할 수 있는 sgRNA 서열을 활용하여 달성될 수 있다.

[0131] 바람직하게는, tracrRNA에 상응하는 sgRNA의 부분은, 5' 헤어핀을 예시하는 서열; AAGGGCUUCUGCCUAUAGGCAGACUGCCC [SEQ ID NO: 54]을 포함할 것이다. 바람직하게는, tracrRNA에 상응하는 sgRNA 부분은, '중간' 헤어핀을 예시하는, 서열; GUGGCGUUGGGGAUCGCCUAUCGCC [SEQ ID NO: 55]을 더욱 포함할 것이다. 바람직하게는, tracrRNA에 상응하는 sgRNA 부분은, 3' 헤어핀을 예시하는, 서열; CGCUUUCUUCGGGAUUCUCCCCACUCUUAAGGCGUUU [SEQ ID NO: 56]을 더욱 포함할 것이다.

[0132] 바람직하게는, tracrRNA에 상응하는 sgRNA 부분은, 5' 헤어핀 및 중간 헤어핀을 포함하는, 서열; AAGGGCUUCUGCCUAUAGGCAGACUGCCCUGGCGUUGGGGAUCGCCUAUCGCC [SEQ ID NO: 57]을 포함할 것이다.

[0133] 바람직하게는, tracrRNA에 상응하는 sgRNA 부분은, 5' 헤어핀, 중간 헤어핀 및 3' 헤어핀을 포함하는, 서열; AAGGGCUUCUGCCUAUAGGCAGACUGCCCUGGCGUUGGGGAUCGCCUAUCGCCCGCUUUCUUCGGGAUUCUCCCCACUCUUAAGGCGUUU [SEQ ID NO: 58]을 포함할 수 있다.

[0134] 본 발명자들은, tracrRNA 스캐폴드의 예측된 스템-루프의 수가, 특히 상승된 온도에서, DNA 절단에 결정적인 역할을 한다는 것을 발견했다. 그들은, tracrRNA 스캐폴드의 3개의 스템-루프의 존재가 절단 활성화에 필수적이지는 않을지라도, 3개의 모든 루프가 존재할 때, 전체 길이의 tracrRNA가 상승된 온도에서 최적의 ThermoCas9-계 DNA

절단을 위해 요구되는 것을 나타내는 범위의 모든 온도에서 절단이 가장 효율적이라는 것을 밝혀냈다. 대조적으로, 3' 헤어핀의 제거는, 절단 효율의 감소를 결과한다. 게다가, 본 발명자들은, 중간 및 3' 헤어핀 모두의 제거가 ThermoCas9의 절단 효율에서, 특히 기능적 온도 범위의 상한 및 하한에서, 심각한 저하를 결과한다는 것을 확인했다. 바람직하게는, 표적 서열의 결합, 절단, 표지 또는 변형은, 상승된 온도, 예를 들어, 45℃ 내지 100℃, 50℃ 내지 100℃, 50℃ 내지 70℃, 50℃ 내지 65℃, 55℃ 내지 65℃ 또는 20℃ 내지 100℃, 20℃ 내지 70℃, 20℃ 내지 65℃와 같은 넓은 온도 범위 내에서 요구된다. 바람직하게는, 3개의 스템-루프 구조를 갖는 sgRNA와 결합된 ThermoCas9는, 20℃ 내지 100℃, 20℃ 내지 70℃, 20℃ 내지 65℃, 45℃ 내지 100℃, 50℃ 내지 100℃, 50℃ 내지 70℃, 50℃ 내지 65℃ 또는 55℃ 내지 65℃ 범위에서 선택된 온도에서 적어도 1분, 적어도 2분, 적어도 3분, 적어도 4분, 또는 적어도 5분, 바람직하게는 5분 동안 표적 서열의 결합, 절단, 표지 또는 변형이 가능할 것이고, 안정하게 유지될 것이다.

[0135] 부가적으로, 본 발명자들은 ThermoCas9 활성화, 예를 들어, 결합, 절단, 표지 또는 변형 활성화의 효율을 조종하기 위해 sgRNA의 스페이서 서열의 길이가 변할 수 있음을 발견하였다. 통상적으로, 스페이서 서열은, 길이가 18nt 내지 25nt의 범위일 것이다. 선택적으로, 스페이서 서열은 18nt, 19nt, 20nt, 21nt, 22nt, 23nt, 24nt 또는 25nt의 길이일 것이다. 바람직하게는, 19nt, 20nt, 21nt 또는 23nt의 스페이서 길이는 사용될 것인데, 이는 본 발명의 Cas9 단백질이 이들 스페이서 길이를 갖는 sgRNA와 결합하는 경우 가장 높은 효율로 표적 서열을 절단하기 때문이다. 18nt의 스페이서가 사용되는 경우, 절단 효율은 현저히 떨어진다. 바람직하게는, 스페이서의 길이는 23nt일 것이다.

[0136] 본 발명의 모든 관점에서, Cas 단백질 또는 폴리펩티드는 박테리아, 고세균류 또는 바이러스로부터 얻거나 유래될 수 있거나; 또는 선택적으로 새로 합성될 수 있다. 바람직한 구체 예에서, 본 발명의 Cas 단백질 또는 폴리펩티드는, 고세균 또는 박테리아로 분류될 수 있지만, 바람직하게는, 박테리아인, 호열성 원핵 유기체로부터 유래된다. 좀 더 바람직하게는, 본 발명의 Cas 단백질 또는 폴리펩티드는 호열성 박테리아로부터 유래될 것이다. 여기서, 용어 "호열성"은 비교적 높은 온도에서 생존 및 성장할 수 있는, 예를 들어, 본 발명의 맥락에서, 41 내지 122℃ (106 내지 252 °F)의 온도에서 핵산을 절단, 결합 또는 변형시킬 수 있는 것을 의미하는 것으로 이해되어야 한다. 바람직하게는, 본 발명의 Cas 단백질 또는 폴리펩티드는, 하나 이상의 호열성 박테리아로부터 단리될 수 있고, 60℃ 초과에서 기능할 것이다. 바람직하게는, 본 발명의 Cas 단백질 또는 폴리펩티드는, 하나 이상의 호열성 박테리아로부터 단리될 수 있고, 60℃ 내지 80℃ 범위 및 최적으로 60℃ 내지 65℃ 범위에서 기능할 것이다. 바람직한 구체 예에서, 본 발명의 Cas 단백질 또는 폴리펩티드는 지오바실러스 속 (*Geobacillus* sp.)으로부터 유래된다. 좀 더 바람직하게는, 본 발명의 Cas 단백질은 지오바실러스 씨모데니트리피칸스로부터 유래된다. 더욱 더 바람직하게는, 본 발명의 Cas 단백질은 지오바실러스 씨모데니트리피칸스 T12로부터 유래된다. 본 발명의 Cas 단백질 또는 폴리펩티드는 바이러스로부터 유래될 수 있다.

[0137] **기능성 모이어티 (Functional Moieties)**

[0138] 유리하게는, 서열-특이적 방식으로 임의의 폴리뉴클레오티드 서열을 표적화하는 본 발명의 Cas 단백질, 폴리펩티드 및 리보핵산 단백질 복합체의 능력은, 어떻게든 표적 핵산을 변형시키기 위해, 예를 들어, 이를 절단 및/또는 이를 표지 및/또는 이를 변형시키는데 활용될 수 있다. 따라서, 부가적인 단백질이 Cas 단백질 또는 폴리펩티드와 함께 제공될 수 있어 이를 달성하는 것으로 인정될 것이다. 따라서, 본 발명의 Cas 단백질 또는 폴리펩티드는, 적어도 하나의 기능성 모이어티를 더욱 포함할 수 있고, 및/또는 본 발명의 Cas 단백질, 폴리펩티드 또는 리보핵산 단백질 복합체는, 적어도 하나의 추가 단백질을 포함하는 단백질 복합체의 일부로서 제공될 수 있다. 바람직한 관점에서, 본 발명은 Cas 단백질, 폴리펩티드 또는 리보핵산 단백질 복합체를 제공하며, 여기서, 상기 Cas 단백질 또는 적어도 하나의 추가 단백질은, 적어도 하나의 기능성 모이어티를 더욱 포함한다. 상기 적어도 하나의 기능성 모이어티는 Cas 단백질에 융합 또는 연결될 수 있다. 바람직하게는, 상기 적어도 하나의 기능성 모이어티는, 천연 또는 인공 단백질 발현 시스템에서 발현을 통해 Cas 단백질에 번역적으로 융합될 수 있다. 선택적으로, 상기 적어도 하나의 기능성 모이어티는, 화학적 합성 단계에 의해 Cas 단백질에 공유결합으로 연결될 수 있다. 바람직하게는, 상기 적어도 하나의 기능성 모이어티는, Cas 단백질의 N-말단 및/또는 C-말단; 바람직하게는, C-말단에 융합 또는 연결된다.

[0139] 바람직하게는, 상기 적어도 하나의 기능성 모이어티는 단백질일 것이다. 이는 이중 단백질일 수 있거나, 또는 선택적으로 Cas 단백질이 유래된 박테리아 종에 기원할 수 있다. 적어도 하나의 기능성 모이어티는; 헬리카제, 뉴클레아제, 헬리카제-뉴클레아제, DNA 메틸라아제, 히스톤 메틸라아제, 아세틸라아제, 포스페타아제, 키아나제, 전사 (공-)활성인자, 전사 억제인자, DNA 결합 단백질, DNA 구조 단백질, 마커 단백질, 리포터 단백질, 형광 단백질, 리간드 결합 단백질, 신호 펩티드, 세포내 위치 측정 서열, 항체 에피토프 또는 친화성 정제

태그 (affinity purification tag)로부터 선택적으로 선택된 단백질일 수 있다.

[0140] 특히 바람직한 관점에서, 본 발명은 Cas 단백질, 폴리펩티드, 또는 리보핵산 단백질 복합체를 제공하고, 여기서 적어도 하나의 기능성 모이어티는, 마커 단백질, 예를 들어, GFP이다.

[0141] 뉴클레아제 활성 (Nuclease Activity)

[0142] 본 발명의 Cas 리보핵산 단백질은, 온도, 바람직하게는, 여기에 개시된 상승된 온도, 예를 들어, 50°C 내지 100°C의 온도에서, 핵산 결합, 절단, 표지 또는 변형 활성을 갖는다. 본 발명의 리보핵산 단백질은, DNA, RNA 또는 합성 핵산을 결합, 절단, 표지 또는 변형할 수 있다. 바람직한 관점에서, 본 발명의 Cas 리보핵산 단백질은, 서열-특이적 방식으로 DNA, 특히 이중-가닥 DNA를 절단할 수 있다.

[0143] 본 발명의 Cas 단백질, 폴리펩티드 또는 리보핵산 단백질은, 하나 이상의 뉴클레아제 도메인을 가질 수 있다. 부위-특이적 뉴클레아제는, DNA의 가닥을 따라 선택된 위치에서 이중 가닥 절단 (DSBs)의 발생을 가능하게 할 수 있다. 표적 숙주 세포에서, 이것은 게놈 내에 특이적 미리-선택된 위치에서 DSBs가 만들어지는 것을 가능하게 한다. 부위-특이적 뉴클레아제에 의한 이러한 절단의 생성은, 관심의 게놈 내에 원하는 위치에서 DNA를 삽입, 결손 또는 변형시키기 위해 내인성 세포내 복구 기작 (endogenous cellular repair machinery)이 용도 변경되는 것을 촉발한다.

[0144] 단백질 또는 폴리펩티드 분자의 하나 이상의 뉴클레아제 활성 부위는, 예를 들어, 단백질 또는 폴리펩티드에 연결된 또는 융합된 또 다른 기능성 모이어티, 예를 들어, FokI 뉴클레아제와 같은 뉴클레아제 도메인의 활성을 가능하게 하기 위해, 비활성화될 수 있다.

[0145] 따라서, 본 발명의 Cas 단백질, 폴리펩티드 및 리보핵산 단백질이 내인성 뉴클레아제 활성을 가질 수 있다는 사실에도 불구하고, 어떤 적용의 경우, Cas 단백질의 비변성 뉴클레아제 활성을 비활성화시키고, Cas 단백질 또는 리보핵산 단백질 복합체를 제공하는 것이 바람직할 수 있으며, 여기서, 비변성 Cas9 뉴클레아제 활성은 비활성화되고, 및 Cas 단백질은 적어도 하나의 기능성 모이어티에 연결된다. 비변성 Cas9 뉴클레아제 활성의 상보성에 의한 불일치-표적화 사건의 발생률을 감소시키는 것은, 이러한 적용 중 하나이다. 이는, 바람직하게는, Cas 단백질 또는 리보핵산 단백질 복합체의 비변성 Cas9 뉴클레아제 활성의 비활성화 및 Cas 단백질에 바람직하게 융합된, 이중 뉴클레아제의 제공에 의해 달성될 수 있다. 따라서, 본 발명은 Cas 단백질 또는 리보핵산 단백질 복합체를 제공하고, 여기서, 적어도 하나의 기능성 모이어티는 뉴클레아제 도메인, 바람직하게는, FokI 뉴클레아제 도메인이다. 특히 바람직한 관점에서, FokI 뉴클레아제 도메인에 융합된 본 발명의 Cas 단백질 또는 리보핵산 단백질 복합체는, 바람직하게는, FokI 뉴클레아제 도메인에 융합된 본 발명의 또 다른 Cas 단백질 또는 리보핵산 단백질 복합체를 포함하는, 단백질 복합체의 일부로서 제공되며, 여기서 2개의 복합체는 표적 게놈 DNA의 대립 가닥 (opposite strands)을 표적으로 한다.

[0146] 몇몇 적용의 경우, 예를 들어, Cas 단백질 또는 리보핵산 단백질 복합체가, 예를 들어, 진단 시험의 일부로 이를 표지하기 위해, 핵산에서 특이적 표적 서열을 인지하고, 변형시키는데 활용되는 적용에서, Cas 단백질, 폴리펩티드 또는 리보핵산 단백질의 뉴클레아제 활성을 완전히 약화시키는 것이 바람직할 수 있다. 이러한 적용에서, Cas 단백질의 뉴클레아제 활성은 비활성화될 수 있고, 및 Cas 단백질에 융합된 기능성 모이어티는; 헬리카제, 뉴클레아제, 헬리카제-뉴클레아제, DNA 메틸라아제, 히스톤 메틸라아제, 아세틸라아제, 포스페타아제, 키아나제, 전사 (공-)활성인자, 전사 억제인자, DNA 결합 단백질, DNA 구조 단백질, 마커 단백질, 리포터 단백질, 형광 단백질, 리간드 결합 단백질, 신호 펩티드, 세포내 위치 측정 서열, 항체 에피토프 또는 친화성 정제 태그로부터 선택적으로 선택된, 단백질일 수 있다.

[0147] 바람직한 관점에서, 뉴클레아제 활성이 결핍된 효소의 기능이 없는, 또는 "죽은" Cas 단백질 또는 폴리펩티드 (dCas)는, 표적 핵산 서열에 결합될 수 있고, 이에 의해 그 서열의 활성을 입체구조로 억제할 수 있다. 예를 들어, 표적화 RNA는 유전자의 프로모터 또는 엑손 서열에 상보적인 것으로 디자인될 수 있어서, 유전자에 dCas 및 표적화 RNA의 결합이 유전자 서열의 전사적 개시 (transcriptional initiation) 또는 신장 (elongation)을 입체구조로 억제하고, 이에 의해 유전자의 발현을 억제한다. 선택적으로, 여기에 기재된 방법 및 사용은, 니카제 (nickases)인 gtCas9의 변형된 뉴클레아제 변이체를 사용할 수 있다. 니카제는 gtCas9 뉴클레아제의 HNH 또는 RuvC 촉매 도메인 중 어느 하나에서 돌연변이를 통해 생성될 수 있다. 이것은, 각각, 비활성 RuvC 또는 HNH 뉴클레아제 도메인을 갖는, spCas9-돌연변이체 D10A 및 H840A를 갖는 *S. pyogenes* Cas9 (spCas)에 대해 나타낸다. 이들 두 개의 돌연변이의 조합은, 촉매적으로 죽은 Cas9 변이체 (dead Cas9 variant)로 이어진다 (Standage-Beier, K. et al., 2015, ACS Synth. Biol. 4, 1217-1225; Jinek, M. et al., 2012, Science 337,

816-821, Xu, T. et al., 2015, Appl. Environ. Microbiol. 81, 4423-4431). 서열 상동성 (도 3)에 기초하여, 이들 잔기는 gtCas9에서 D8 (도 3의 D17) 및 D581 또는 H582 (도 3)일 수 있다.

[0148] 바람직하게는, gtCas9에서 돌연변이 D8A 및 H582A (ThermoCas9)는, 뉴클레아제 활성이 결핍된, ThermoCas9의 촉매적으로 비활성, 또는 "죽은" Cas 단백질 또는 폴리펩티드 변이체 (dCas)를 생성하는데 사용될 수 있다. 이러한 dCas는, 예를 들어, dsDNA 절단을 도입하지 않고 정상적이고 명확히 DNA 인자 (DNA elements)에 결합할 수 있는, 효과적인 열활성 전사 침묵 CRISPRi 도구로서, 적용을 유용하게 발견할 수 있다. 유리하게도, 이러한 시스템은, 다른 것들 중에서, 인간 세포의 대사 연구를 크게 촉진시킬 수 있다.

[0149] 특히 바람직한 관점에서, 본 발명은 Cas 단백질 또는 리보핵산 단백질 복합체를 제공하고, 여기서, 상기 Cas 단백질의 뉴클레아제 활성은 비활성화되고, 적어도 하나의 기능성 모이어티는 마커 단백질, 예를 들어, GFP이다. 이러한 방식에서, 관심의 핵산 서열을 구체적으로 표적화하고, 및 광학 신호를 발생하는 마커를 사용하여 이를 시각화하는 것은 가능할 수 있다. 적절한 마커는, 예를 들어, 형광 리포터 단백질, 예를 들어, 녹색 형광 단백질 (GFP), 황색 형광 단백질 (YFP), 적색 형광 단백질 (RFP), 시안 형광 단백질 (CFP) 또는 mCherry를 포함할 수 있다. 이러한 형광 리포터 유전자는, 이의 발현이 형광 측정에 의해 간단하고 즉시 분석될 수 있기 때문에, 단백질 발현의 시각화를 위한 적절한 마커를 제공한다. 선택적으로, 리포터 핵산은 루시페라제 (예를 들어, 반딧불이 루시페라제 (firefly luciferase))와 같은, 발광 단백질을 인코딩할 수 있다. 선택적으로, 리포터 유전자는 광학 신호를 발생하는데 사용될 수 있는, 발색 효소 (chromogenic enzyme), 예를 들어, (베타-갈락토시다아제 (LacZ) 또는 베타-글루쿠로니다아제 (Gus)와 같은) 발색 효소일 수 있다. 발현의 측정에 사용되는 리포터는 또한 항원 펩티드 태그될 수 있다. 다른 리포터 또는 마커는 당업계에 공지되어 있으며, 이들은 적절하게 사용될 수 있다.

[0150] 마커가 시각화될 수 있기 때문에, 표적 핵산이 RNA, 특히 mRNA인 어떤 구체 예에서, 특히 마커에 의해 발생된 광학 신호가 발현 산물의 양에 직접적으로 비례하는 경우, 마커에 의해 제공된 광학 신호의 검출 및 정량화에 의해, 유전자의 전사 활성을 정량화하는 것은 가능하다. 따라서, 본 발명의 바람직한 구체 예에서, 본 발명의 Cas 단백질 또는 리보핵산 단백질은, 관심의 유전자의 발현 산물을 분석하는데 사용될 수 있다.

[0151] 하나의 관점에서, 여기에 기재된 gtCas9는, 상동 재조합 (HR) 매개된 게놈 변형 방법에 사용될 수 있다. 이러한 방법은 HR 및 부위-인도 gtCas9 활성을 포함하며, 이에 의해 반대 선택 (counter selection)은 HR에 의해 도입된 원하는 변형을 갖지 않는 세포를 제거하는 gtCas9 활성에 의해 일어난다.

[0152] 따라서, 여기에 제공된 방법 및 사용은 게놈이 원하는 돌연변이로 변형될 수 있는 제1 단계 및 미변형 세포가 gtCas9 리보뉴클레아제 복합체에 의해 표적화되어 미변형 세포의 게놈으로 DSDB를 도입할 수 있는 제2 단계 동안 선택되는 상동 재조합의 공정을 가능하게 한다. 이들 방법 및 사용은 미변형 세포를 제거하면서 원하는 돌연변이를 갖는 세포의 개체군을 전반적으로 증가시킨다. 바람직하게는, 이러한 방법 및 사용은 실질적으로 내인성 NHEJ 복구 메카니즘을 갖지 않은 미생물에서 사용된다. 선택적으로, 상기 방법 및 사용은 내인성 NHEJ 복구 메카니즘을 갖는 세포에 적용될 수 있다. 여기에 기재된 방법 및 사용은 내인성 NHEJ 복구 메카니즘을 갖는 세포에 적용될 수 있지만, 여기서, 상기 NHEJ 복구 메카니즘은 조건부로 감소되거나 또는 NHEJ 활성은 녹아웃된다 (knocked out).

[0153] 여기에 제공된 방법 및 사용은, 가이드 RNA가 변형된 게놈을 더 이상 인식할 수 없도록, 가이드 RNA와 적어도 하나의 불-일치를 갖는 상동 재조합 폴리뉴클레오티드의 서열을 활용할 수 있다. 이것은 gtCas9 리보뉴클레아제 복합체가 변형된 게놈을 인식하지 못한다는 것을 의미한다. 따라서, DSDB는 gtCas9 리보뉴클레아제 복합체에 의해 도입될 수 없으므로, 변형된 세포는 생존할 것이다. 그러나, 미변형 게놈을 갖는 세포는, 여전히 가이드 RNA에 실질적인 상보성을 가지며, 결과적으로 gtCas9 리보뉴클레아제 복합체에 의해 부위-특이적으로 절단될 수 있다.

[0154] 본 발명의 방법 및 사용의 또 다른 관점에서, gtCas9 리보뉴클레아제 복합체가 게놈을 절단하는 작용을 방해하는 방식은, 가이드에 의해 표적화된 서열을 변형시키거나 제거하는 것과 큰 관련이 없고, 오히려 gtCas9 리보뉴클레아제 복합체가 필요로 하는 PAM을 변형시키거나 제거하는 것이다. PAM은 특이적 절단 부위에 gtCas9 리보뉴클레아제 복합체를 블라인드 (blind)하기 위해 변형되거나 제거된다. 따라서, 본 발명의 방법 및 사용은 gtCas9 리보뉴클레아제 복합체에 의해 인식되는 PAM 서열을 포함하지 않는 상동 재조합 폴리뉴클레오티드의 서열을 사용하는 방법을 포함할 수 있다. 따라서, DSDB는 gtCas9 리보뉴클레아제 복합체에 의해 도입될 수 없으므로, HR 변형된 세포는 생존할 것이다. 그러나, 미변형 세포는, gtCas9 리보뉴클레아제 복합체 및 이의 가이드에 의해 여전히 인식될 것이고, 결과적으로 부위-특이적으로 절단될 것이다.

- [0155] 따라서, 세포의 게놈을 변형시키기 위해 HR에 의존하는 방법 및 사용은 여기에 제공된다. 바람직하게는, 업스트림 플랜크 (upstream flank) 및 다운스트림 플랜크는 길이가 각각 0.5kb (kilobases) 내지 1.0 kb이다. 그러나, 더 크거나 더 짧은 단편을 사용하는 재조합은 또한 가능하다. 상동 재조합 폴리뉴클레오티드는 업스트림과 다운스트림 플랜킹 영역들 사이에 폴리뉴클레오티드 서열을 더욱 포함할 수 있다. 이 폴리뉴클레오티드 서열은, 예를 들어, 게놈으로 도입될 변형을 함유할 수 있다.
- [0156] 상동 재조합은 표적 영역에 실질적인 상보성을 갖는 업스트림 및 다운스트림 플랜크에 의존하지만, 불일치도 또한 수용될 수 있다. 따라서, 몇몇 구체 예에서, 상동 재조합은 업스트림 및 다운스트림 플랜크에 광범위한 상동성을 갖는 DNA 세그먼트 (segments) 사이에 일어나는 것으로 알려져 있다. 선택적인 구체 예에서, 업스트림 및 다운스트림 플랜크는 표적 영역에 완전한 상보성을 갖는다. 업스트림 및 다운스트림 플랜크는 크기가 동일할 필요는 없다. 그러나, 몇몇 사례에서, 업스트림 및 다운스트림 플랜크는 크기가 동일하다. 상동 재조합의 효율은 플랜크의 가장 작은 단편 길이의 상동 재조합의 가능성에 매우 의존할 것이다. 그러나, 상동 재조합 공정이 비효율적일지라도, 유리하게는, 여기서 기재된 방법은 미변형 세포에 대해 원하는 변형을 갖는 임의의 세포를 선택할 것이다. 상동 재조합은 또한 완전한 유전자 클러스터 (gene clusters)를 둘러싸게 만드는 큰 결손 (예를 들어, 50kb 이상)을 가능하게 한다. 상동 재조합은 재조합공학 (recombineering)에도 사용되며, 이것은 더 작은 단편 (45-100 nt)에 대한 재조합을 가능하게 하는 잘-알려진 방법이다. 여기에 기재된 방법 및 사용은 선택적으로 게놈에서 표적을 함유하는 제2 표적 영역에 실질적으로 상보적인 서열을 갖는 상동 재조합 폴리뉴클레오티드를 인코딩하는 서열을 포함하는 폴리뉴클레오티드 또는 적어도 또 다른 상동 재조합 폴리뉴클레오티드를 더욱 포함할 수 있다.
- [0157] 바람직한 구체 예에서, 여기에 기재된 방법 및 사용은, DNA인 상동 재조합 폴리뉴클레오티드를 활용한다. 몇몇 구체 예에서, DNA는 단일 가닥이다. 다른 구체 예에서, DNA는 이중 가닥이다. 또 다른 구체 예에서, DNA는 이중 가닥 및 플라스미드 유래 (plasmid borne)이다.
- [0158] 여기에 제공된 방법 및 사용에서 HR은 게놈으로부터 폴리뉴클레오티드 서열을 제거하는데 사용될 수 있다. 선택적으로, 여기에 제공된 방법 및 사용에서 HR은 게놈에, 하나 이상의 유전자(들), 또는 이의 단편(들)을 삽입하는데 사용될 수 있다. 또 다른 대안으로서, 여기에 제공된 방법 및 사용에서 HR은, 게놈에서 적어도 하나의 뉴클레오티드를 변형 또는 치환하는데 사용될 수 있다. 결과적으로, 여기에 제공된 방법 및 사용은, 임의의 원하는 종류의 게놈 변형을 위해 사용될 수 있다.
- [0159] 선택적으로, 여기서 기재된 gCas9는, 세포에서 HR 매개된 게놈 변형 방법에 사용될 수 있고, 이에 의해, gCas9 활성은, spCas9에 대해 나타낸 바와 같이, DSB를 도입하고, 세포에서 세포의 HR (cellular HR)을 유도할 수 있다 (Jiang *et al.* (2013) *Nature Biotech.*, 31, 233-239; Xu *et al.* (2015) *Appl Environ Microbiol.*, 81, 4423-4431; Huang *et al.* (2015) *Acta Biochimica et Biophysica Sinica*, 47, 231-243).
- [0160] 선택적으로, 상동 재조합은 재조합공학을 통해, 예를 들어, Mougikakos 등 ((2016), *Trends Biotechnol.* 34: 575-587)에 의해 검토된 바와 같은, RecT 또는 베타 단백질을 인코딩하는 유전자를 발현하는 세포 내로 올리고 뉴클레오티드를 도입시켜, 가능하게 할 수 있다. 또 다른 구체 예에서, Cas9는 Ronda 등 ((2016), *Sci. Rep.*, 6: 19452)에 의해 예증된 바와 같이, Multiplex Automated Genome Engineering (MAGE)과 조합될 수 있다.
- [0161] 모든 부분에서, 본 발명의 Cas 단백질의 참조 서열은, 아미노산 서열을 인코딩하는 뉴클레오티드 서열로서 정의될 수 있다. 예를 들어, SEQ ID NO: 2 내지 6에 정의된 모티프의 아미노산 서열은 또한 그 아미노산 서열을 인코딩하는 모든 핵산 서열을 포함한다.
- [0162] 따라서, 본 발명은 또한 다음을 포함하는 Cas 단백질을 인코딩하는 단리된 핵산 분자를 제공한다:
- [0163] a. 아미노산 모티프 EKDGKYYC [SEQ ID NO: 2]; 및/또는
- [0164] b. 아미노산 모티프 $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], 여기서, X_1 은 이소류신, 메티오닌 또는 프롤린으로부터 독립적으로 선택되고, X_2 는 발린, 세린, 아스파라긴 또는 이소류신으로부터 독립적으로 선택되며, X_3 는 글루타메이트 또는 리신으로부터 독립적으로 선택되고, 또는 X_4 는 알라닌, 글루타메이트 또는 아르기닌 중 하나임; 및/또는
- [0165] c. 아미노산 모티프 X_5LXX_6IE [SEQ ID NO: 4], 여기서, X_5 는 메티오닌 또는 페닐알라닌으로부터 독립적으로 선택되고, 및 X_6 는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택됨; 및/또는

- [0166] d. 아미노산 모티프 X_7VYSX_8K [SEQ ID NO: 5], 여기서, X_7 은 글루타메이트 또는 이소류신이고, 및 X_8 은 트립토판, 세린 또는 리신 중 하나임; 및/또는
- [0167] e. 아미노산 모티프 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], 여기서, X_9 는 알라닌 또는 글루타메이트이고, X_{10} 은 글루타민 또는 리신이며, X_{11} 은 아르기닌 또는 알라닌이고, X_{12} 는 아스파라긴 또는 알라닌이며, 및 X_{13} 은 리신 또는 세린이고;
- [0168] 여기서, Cas 단백질은, 적어도 하나의 표적화 RNA 분자, 및 상기 표적화 RNA 분자에 의해 인지되는 표적 핵산 서열을 포함하는 폴리뉴클레오타이드와 결합되는 경우, 50℃ 내지 100℃에서 DNA 결합, 절단, 표지 또는 변형할 수 있다. 선택적으로, Cas 단백질은, 인간 세포의 절단, 표지 또는 변형을 할 수 있다. 선택적으로, 여기서, 상기 인간 세포는 배아 줄기 세포가 아니다. 선택적으로, 여기서, 상기 절단, 표지 또는 변형은, 인간의 생식 계열 유전적 동일성을 변형시키지 않는다.
- [0169] 다른 관점에서, 본 발명은 또한, SEQ ID NO: 1의 아미노산 서열 또는 이와 적어도 77% 동일성의 서열을 갖는 일정한 간격을 두고 주기적으로 분포하는 짧은 회문구조 반복서열 (CRISPR)-관련 (Cas) 단백질을 인코딩하는 단리된 핵산 분자를 제공한다.
- [0170] 또 다른 관점에서, 본 발명은 또한, 번역시에 Cas 단백질에 융합되는 펩티드를 인코딩하는 적어도 하나의 핵산 서열을 더욱 포함하는, 단리된 핵산 분자를 제공한다.
- [0171] 또 다른 관점에서, 본 발명은 또한 단리된 핵산 분자를 제공하며, 여기서, Cas 단백질을 인코딩하는 핵산 분자에 융합된 적어도 하나의 핵산 서열은, 헬리카제, 뉴클레아제, 헬리카제-뉴클레아제, DNA 메틸라아제, 히스톤 메틸라아제, 아세틸라제, 포스포타아제, 키나아제, 전사 (공-)활성인자, 전사 억제인자, DNA 결합 단백질, DNA 구조 단백질, 마커 단백질, 리포터 단백질, 형광 단백질, 리간드 결합 단백질, 신호 펩티드, 세포내 위치정보 서열, 항체 에피토프 또는 친화성 정제 태그로부터 선택된 단백질을 인코딩한다.
- [0172] **ThermoCas9 뉴클레아제 활성: 2가 양이온**
- [0173] 사전에 특징화된, 중온성 Cas9 엔도뉴클레아제는, 표적 DNA에서 DSBs의 발생을 촉진하기 위해 2가 양이온을 사용한다. 본 발명자들은, ThermoCas9가 다음의 이가 양이온: Mg^{2+} , Ca^{2+} , Mn^{2+} , Co^{2+} , Ni^{2+} , 및 Cu^{2+} 중 어느 하나의 존재하에서 dsDNA 절단을 매개할 수 있다는 것을 밝혀냈다.
- [0174] **ThermoCas9 뉴클레아제 활성: 기질**
- [0175] 본 발명자들은 또한, 놀랍게도, 어떤 타입-II C 시스템이 효율적인 단일 가닥 DNA 가위 (DNA cutter)라는 보고 ((Ma, *et al.*, *Mol. Cell* 60, 398-407 (2015); Zhang, *et al.*, *Mol. Cell* 60, 242-255 (2015))에도 불구하고, ThermoCas9가 ssDNA의 절단을 지시할 수 없음을 밝혀냈다. ThermoCas9의 뉴클레아제 활성은 dsDNA 기질에 제한된다.
- [0176] **발현 벡터 (Expression Vectors)**
- [0177] 본 발명의 핵산은 단리될 수 있다. 그러나, 핵산 센싱 구조물의 발현이 선택 세포에서 수행될 수 있게 하기 위해, Cas 단백질 또는 리보핵산 단백질을 인코딩하는 폴리뉴클레오타이드 서열은, 바람직하게는, 발현 구조물에 제공될 것이다. 몇몇 구체 예에서, Cas 단백질 또는 리보핵산 단백질을 인코딩하는 폴리뉴클레오타이드는, 적절한 발현 벡터의 일부로서 제공될 것이다. 어떤 구체 예에서, (발현시 Cas 단백질에 융합되는 아미노산 잔기를 인코딩하는 뉴클레오타이드 서열을 갖거나 또는 갖지 않는) 본 발명의 발현 벡터는, 전술된 바와 같은 표적화 RNA 분자를 인코딩하는 뉴클레오타이드 서열을 더욱 포함할 수 있다. 결과적으로, 이러한 발현 벡터는, 적절한 숙주에서 사용되어, 원하는 뉴클레오타이드 서열을 표적으로 할 수 있는 본 발명의 리보핵산 단백질 복합체를 발생시킬 수 있다. 선택적으로, 전술된 바와 같은 표적화 RNA 분자를 인코딩하는 뉴클레오타이드 서열은, 개별 발현 벡터에 제공될 수 있거나 또는 선택적으로 다른 수단에 의해 표적 세포로 전달될 수 있다.
- [0178] 적절한 발현 벡터는, 수용 세포 (recipient cell)에 따라 변할 수 있으며, 및 표적 세포에서 발현이 가능하고 및 바람직하게는, 높은-수준의 발현을 용이하게 하는 조절 인자 (regulatory element)를 적절하게 혼입할 수 있다. 이러한 조절 서열 (regulatory sequences)은, 예를 들어, 개시 (initiation), 정확도, 속도, 안정성, 다운스트림 공정 및 이동성의 관점에서, 유전자 또는 유전자 산물의 전사 또는 번역에 영향을 미칠 수 있다.
- [0179] 이러한 인자는, 예를 들어, 강한 및/또는 구성적 프로모터 (constitutive promoters), 5' 및 3' UTR's, 전사

및/또는 번역 인핸서 (enhancers), 전사 인자 또는 단백질 결합 서열, 시작 부위 및 종결 서열, 리보솜 결합 부위, 재조합 부위, 폴리아데닐화 서열, 센스 또는 안티센스 서열, 전사의 정확한 개시를 보장하는 서열 및 선택적으로 숙주 세포에서 전사 및 전사 안정화의 종결을 보장하는 폴리-A 신호를 포함할 수 있다. 조절 서열은 식물-, 동물-, 박테리아-, 곰팡이- 또는 바이러스 유래일 수 있다. 분명히, 적절한 조절 인자는, 관심의 숙주 세포에 따라 변할 것이다. 인간 숙주 세포에서 높은-수준의 발현을 용이하게 하는 조절 인자는, 효모에서 AOX1 또는 GAL1 프로모터 또는 CMV- 또는 SV40-프로모터, CMV-인핸서, SV40-인핸서, Herpes simplex virus VIP16 전사 활성인자 또는 인간 세포에서 글로빈 인트론 (globin intron)의 함유물 (inclusion)을 포함할 수 있다.

[0180] 적절한 조절 인자는 구성적 (constitutive)일 수 있고, 이에 의해 이들은, 대부분의 환경 조건 또는 발생 단계 (developmental stages), 발생 단계 특이성 또는 유도성 하에서 발현을 지시한다. 바람직하게는, 프로모터는 유도성이어서, 온도, 빛, 화학제, 건조 (drought) 및 다른 자극과 같은, 환경적, 화학적 또는 발생적 신호 (cues)에 반응하여 발현을 지시한다. 적절하게, 프로모터는, 특정 발생 단계에서 또는 세포-외 또는 세포-내 조건, 신호 또는 외부에서 가해진 자극에 반응하여, 관심의 단백질의 발현을 가능하게 선택될 수 있다. 예를 들어, 특정 성장 단계 (예를 들어, osmY 정지상 프로모터)에서 또는 특정 자극 (예를 들어, HtpG Heat Shock Promoter)에 반응하여, 높은-수준의 발현을 제공하는 E. coli에서 사용하기 위한 다양한 프로모터는 존재한다.

[0181] 적절한 발현 벡터는 적절한 숙주 세포에서 및/또는 특정 조건하에서 상기 벡터의 선택을 가능하게 하는 선택 가능한 마커를 인코딩하는 부가적인 서열을 포함할 수 있다.

[0182] 본 발명은 또한 세포를 전수된 바와 같은 발현 벡터 중 어느 하나로 형질주입 (transfecting), 형질전환 (transforming) 또는 형질도입 (transducing)하는 것을 포함하는, 인간 세포에서 표적 핵산을 변형시키는 방법을 포함한다. 형질주입, 형질전환, 또는 형질도입의 방법은 기술분야의 당업자에게 잘 알려진 타입들이다. 본 발명의 리보핵산 단백질 복합체의 발현을 발생하는데 사용되는 하나의 발현 벡터가 존재하고, 표적화 RNA가 세포에 직접 첨가되는 경우, 그 다음, 형질주입, 형질전환 또는 형질도입의 같거나 또는 다른 방법은 사용될 수 있다. 유사하게, 본 발명의 리보핵산 단백질 복합체의 발현을 발생하기 위해 사용되는 하나의 발현 벡터가 존재하고, 및 또 다른 발현 벡터가 발현을 통해 인시튜 (in situ)로 표적화 RNA를 발생시키는데 사용되는 경우, 그 다음, 형질주입, 형질전환, 또는 형질도입의 같거나 또는 다른 방법은 사용될 수 있다.

[0183] 다른 구체 예에서, Cas 단백질 또는 폴리펩티드를 인코딩하는 mRNA는, 캐스케이드 복합체 (Cascade complex)가 세포에서 발현되도록, 세포 내로 도입된다. Cas 단백질 복합체를 원하는 표적 서열로 안내하는 표적화 RNA는 또한, 필요한 리보핵산 단백질 복합체가 세포에서 형성되도록, mRNA와 동시에, 개별적으로 또는 순차적으로, 세포 내로 도입된다.

[0184] 따라서, 본 발명은 또한 전수된 바와 같은 리보핵산 단백질 복합체와 핵산을 접촉시키는 단계를 포함하는, 표적 핵산을, 변형, 즉, 절단, 태깅, 변형, 표지 또는 결합하는 방법을 제공한다. 선택적으로, 상기 핵산은 인간 세포 내에 있다. 선택적으로, 상기 인간 세포는 배아 줄기 세포가 아니다.

[0185] 부가적으로, 본 발명은 또한, 전수된 바와 같은 표적화 RNA 분자에 부가하여, 전수된 바와 같은 Cas 단백질 또는 폴리펩티드와 핵산을 접촉시키는 단계를 포함하는 표적 핵산을 변형시키는 방법을 포함한다.

[0186] 상기 방법들에 따르면, 표적 핵산의 변형은, 따라서, 시험관에서 및 무-세포 환경에서 수행될 수 있다. 무-세포 환경에서, 표적 핵산, Cas 단백질 및 표적화 RNA 분자의 각각의 첨가는, 동시에, 순차적으로 (원하는 대로의 임의의 순서로) 또는 개별적으로 이루어질 수 있다. 따라서, 표적 핵산 및 표적화 RNA를 반응 혼합물에 동시에 첨가한 다음, 본 발명의 Cas 단백질 또는 폴리펩티드를 후속 단계에서 개별적으로 첨가하는 것이 가능하다.

[0187] 마찬가지로, 표적 핵산의 변형은, 단리된 세포이든 또는 다세포 조직, 기관 또는 유기체의 일부이든 간에, 생체 내에서, 즉, 인간 세포에서 인시튜로 이루어질 수 있다. 전체 조직 및 기관의 맥락에서, 및 유기체의 맥락에서, 상기 방법은 바람직하게는, 생체 내에서 수행될 수 있거나, 또는 선택적으로, 전체 조직, 기관 또는 유기체 유래의 인간 세포를 단리하는 단계, 상기 방법에 따라 리보핵산 단백질 복합체로 세포를 처리하는 단계, 및 그 후에 리보핵산 단백질 복합체로 처리된 세포를, 같거나 또는 다른 유기체 내에서, 이의 이전 위치 또는 다른 위치로 복귀시키는 단계에 의해 수행될 수 있다. 선택적으로, 상기 인간 세포는 배아 줄기 세포가 아니다.

[0188] 이들 구체 예에서, 리보핵산 단백질 복합체 또는 Cas 단백질 또는 폴리펩티드는, 세포 내로 적절한 형태의 전달을 필요로 한다. 이러한 적절한 전달 시스템 및 방법은, 당업자에게 잘 알려져 있으며, 및 세포질 또는 핵 미세 주입 (nuclear microinjection)을 포함하지만, 이에 제한되는 것은 아니다. 바람직한 전달 방식에서, 아데노-연관 바이러스 (AAV)는 사용되고; 이 전달 시스템은, 사람에게 질병을 일으키지 않으며, 유럽에서 임상적 사용이

승인되었다.

[0189] 따라서, 본 발명은:

[0190] a. 전술된 바와 같은 리보핵산 단백질 복합체; 또는

[0191] b. 전술된 바와 같은 단백질 또는 단백질 복합체 및 전술된 바와 같은 RNA 분자와 핵산을 접촉시키는 단계를 포함하는 표적 핵산을 변형시키는 방법을 제공한다.

[0192] 또 다른 관점에서, 본 발명은 전술된 바와 같은 리보핵산 단백질 복합체를 인코딩하는 뉴클레오티드 서열을 포함하는 발현 벡터로 세포를 형질전환, 형질주입, 또는 형질도입하는 단계를 포함하는, 인간 세포에서 표적 핵산을 변형시키는 방법을 제공하거나; 또는 선택적으로 전술된 바와 같은 단백질 또는 단백질 복합체를 인코딩하는 뉴클레오티드 서열을 포함하는 발현 벡터 및 전술된 바와 같은 표적화 RNA 분자를 인코딩하는 뉴클레오티드 서열을 포함하는 추가 발현 벡터로 세포를 형질전환, 형질주입 또는 형질도입시키는 단계를 포함한다. 선택적으로, 상기 표적 핵산은 인간 세포 내에 있고, 여기서, 상기 인간 세포는 배아 줄기 세포가 아니다.

[0193] 또 다른 관점에서, 본 발명은, 전술된 바와 같은 단백질 또는 단백질 복합체를 인코딩하는 뉴클레오티드 서열을 포함하는 발현 벡터로 세포를 형질전환, 형질주입 또는 형질도입하는 단계, 및 그 다음 전술된 바와 같이 표적화 RNA 분자를 세포 내로 전달하는 단계를 포함하는 인간 세포에서 표적 핵산을 변형시키는 방법을 제공한다. 선택적으로, 상기 표적 핵산은 인간 세포 내에 있고, 여기서, 상기 인간 세포는 배아 줄기 세포가 아니다.

[0194] 가이드 (즉, 표적화) RNA (gRNA) 분자 및 Cas 단백질 또는 폴리펩티드가 리보핵산 단백질 복합체의 일부로서가 아닌 별도로 제공되는 구체 예에서, gRNA 분자는, Cas 단백질 또는 단백질 복합체와 동시에, 개별적으로 또는 순차적으로, 인간 세포 내로 적절한 형태의 전달을 필요로 한다. 세포 내로 RNA를 도입시키는 이러한 형태는, 기술분야의 당업자에게 잘 알려져 있으며, 및 종래의 형질주입 방법을 통해 시험관 내 또는 생체 외 전달을 포함할 수 있다. 미세주입 및 전기천공법 (electroporation)과 같은, 물리적 방법뿐만 아니라 칼슘 공-침전, 및 상업적으로 이용 가능한 양이온성 중합체 및 지질, 및 세포-침투 펩티드, 세포-침투 (유전자총 (biolistic)) 입자는, 각각 사용될 수 있다. 예를 들어, 바이러스, 특히 바람직하게는, AAV는, 예를 들어, 바이러스성 입자에 본 발명의 Cas 단백질 복합체 또는 본 발명의 리보핵산 단백질 복합체의 (가역성) 융합을 통해, 세포질 내 및/또는 핵 내, 전달 비히클 (delivery vehicles)로서 사용될 수 있다.

[0195] 또 다른 관점에서, 본 발명은 표적 핵산을 변형시키는 방법을 제공하며, 여기서, 적어도 하나의 기능성 모이어티는 마커 단백질 또는 리포터 단백질이고, 상기 마커 단백질 또는 리포터 단백질은 표적 핵산과 결합하며; 바람직하게는, 여기서, 상기 마커는 형광 단백질, 예를 들어, 녹색 형광 단백질 (GFP)이다.

[0196] 표적 핵산을 변형시키는 전술된 방법에서, 기능성 모이어티는 마커일 수 있고, 상기 마커는 표적 핵산과 결합하며; 바람직하게는, 여기서 상기 마커는 단백질; 선택적으로 형광 단백질, 예를 들면, 녹색 형광 단백질 (GFP), 황색 형광 단백질 (YFP), 적색 형광 단백질 (RFP) 또는 mCherry이다. 시험관 내, 생체 외 또는 생체 내를 불문하고, 본 발명의 방법은, 바람직하게는, 슈퍼코일화된 플라스미드 (supercoiled plasmid) 또는 염색체와 같은, 고차 구조, 또는 mRNA와 같은 단일 가닥 표적 핵산의 형태로, 핵산 분자 내에 표적 유전자좌를 직접적으로 시각화하는데 사용될 수 있다. 표적 유전자좌의 직접 시각화는 전자 현미경, 또는 형광 현미경을 사용할 수 있다. 그러나, 본 발명의 방법의 맥락에서, 다른 종류의 라벨 (label)이, 유기 염료 분자, 동정용 방사성 동위원소 (radiolabels) 및 소분자일 수 있는 스핀 라벨 (spin labels)을 포함하는 마커로서 사용될 수 있는 것으로 인정될 것이다.

[0197] 표적 핵산이 dsDNA인 표적 핵산을 변형시키는 본 발명의 방법에서, 기능성 모이어티는 뉴클레아제 또는 헬리카제-뉴클레아제일 수 있고, 및 상기 변형은 바람직하게는, 원하는 유전자좌에서 단일 가닥 또는 이중 가닥 절단이다. 이러한 방식에서, DNA의 독특한 서열 특이적 절단은, 리보핵산 단백질 복합체에 융합된 적절한 기능성 모이어티를 사용하여 조작될 수 있다. 최종 리보핵산 단백질 복합체의 RNA 성분의 선택 서열은, 기능성 모이어티의 작용을 위한 원하는 서열 특이성을 제공한다.

[0198] 따라서, 본 발명은 또한 dsDNA 분자로부터 뉴클레오티드 서열의 적어도 일부를 제거하기 위해; 선택적으로 유전자 또는 유전자들의 기능을 녹아웃되게 하기 위해, 원하는 유전자좌에 인간 세포에서 dsDNA 분자의 비-상동 말단 연결 (non-homologous end joining) 방법을 제공하며; 여기서, 상기 방법은 전술된 바와 같은 표적 핵산을 변형시키는 방법들 중 어느 하나를 사용하여 이중 가닥 절단을 만드는 단계를 포함한다. 선택적으로, 상기 방법은 인간 세포에서 수행되고, 여기서, 상기 인간 세포는 배아 줄기 세포가 아니다.

- [0199] 본 발명은 기존의 뉴클레오티드 서열을 변형시키거나 또는 원하는 뉴클레오티드 서열을 삽입하기 위해 원하는 유전자좌에 세포에서 dsDNA 분자 내로 핵산의 상동 재조합시키는 방법을 더욱 제공하며, 여기서, 상기 방법은 전술된 바와 같은 표적 핵산을 변형시키는 방법들 중 어느 하나를 사용하여 원하는 유전자좌에 이중 가닥 절단을 만드는 단계를 포함한다. 선택적으로, 상기 방법은 인간 세포에서 수행되고, 여기서, 상기 인간 세포는 배아 줄기 세포가 아니다.
- [0200] 따라서, 본 발명은 또한 전술된 방법들 중 어느 하나에 따라 표적 핵산 서열을 변형시키는 단계를 포함하는 유기체에서 유전자 발현을 변형시키는 방법을 제공하며, 여기서, 상기 핵산은 dsDNA이고, 기능성 모이어티는 DNA 변형 효소 (예를 들어, 메틸라아제 또는 아세틸라제), 전사 활성인자 또는 전사 억제인자로부터 선택되고, 및 여기서, 상기 유기체는 인간이다.
- [0201] 본 발명은 부가적으로 전술된 방법들 중 어느 하나에 따라 표적 핵산 서열을 변형시키는 단계를 포함하는 유기체에서 유전자 발현을 변형시키는 방법을 제공하며, 여기서, 상기 핵산은 mRNA이고, 기능성 모이어티는 리보뉴클레아제이며; 선택적으로 엔도뉴클레아제, 3' 엑소뉴클레아제 또는 5' 엑소뉴클레아제로부터 선택되고, 여기서, 상기 유기체는 인간이다.
- [0202] 표적 핵산은 DNA, RNA 또는 합성 핵산일 수 있다. 바람직하게는, 표적 핵산은 DNA; 바람직하게는, dsDNA이다.
- [0203] 그러나, 표적 핵산은 RNA; 바람직하게는, mRNA일 수 있다. 따라서, 선택적으로, 본 발명은 또한 표적 핵산을 변형시키는 방법을 제공하며, 여기서 상기 표적 핵산은 RNA이다.
- [0204] 또 다른 관점에서, 본 발명은 인간 세포에서 표적 핵산을 변형시키는 방법을 제공하며, 여기서, 핵산은 dsDNA이고, 적어도 하나의 기능성 모이어티는 뉴클레아제 또는 헬리카제-뉴클레아제이며, 및 상기 변형은 원하는 유전자좌에서 단일-가닥 또는 이중-가닥 절단이다. 선택적으로, 상기 인간 세포는 배아 줄기 세포가 아니다. 선택적으로, 여기서, 상기 방법은 인간의 생식 계열 유전적 동일성을 변형시키지 않는다.
- [0205] 또 다른 관점에서, 본 발명은, 인간 세포에서 표적 핵산을 변형시키는 방법을 제공하며, 여기서 변형은 원하는 유전자좌에서 유전자 발현의 침묵을 결과하고; 및 여기서 상기 방법은:
- [0206] a. dsDNA 분자에서 이중-가닥 절단을 만드는 단계; 및
- [0207] b. 비-상동 말단 재조합 (NHEJ)에 의해 세포에서 dsDNA 분자의 복구 단계를 포함하고; 선택적으로, 여기서, 상기 인간 세포는 배아 줄기 세포가 아니다. 선택적으로, 여기서, 상기 방법은, 인간의 생식 계열 유전적 동일성을 변형시키지 않는다.
- [0208] 또 다른 관점에서, 본 발명은 인간 세포에서 표적 핵산을 변형시키는 방법을 제공하며; 여기서, 기존의 뉴클레오티드 서열이 변형 또는 결손되고 및/또는 원하는 뉴클레오티드 서열이 원하는 위치에 삽입되며, 여기서 상기 방법은:
- [0209] a. 원하는 유전자좌에서 이중 가닥 절단을 만드는 단계; 및
- [0210] b. 상동 재조합에 의해 세포에서 dsDNA 분자의 복구 단계를 포함하고;
- [0211] 선택적으로, 여기서, 상기 인간 세포는 배아 줄기 세포가 아니다. 선택적으로, 여기서, 상기 방법은 인간의 생식 계열 유전적 동일성을 변형시키지 않는다.
- [0212] 또 다른 관점에서, 본 발명은 전술된 바와 같은 표적 핵산 서열을 변형시키는 단계를 포함하는 인간 세포에서 유전자 발현을 변형시키는 방법을 제공하고; 여기서, 상기 핵산은 dsDNA이며, 및 상기 기능성 모이어티는 DNA 변형 효소 (예를 들어, 메틸라아제 또는 아세틸라제), 전사 활성인자 또는 전사 억제인자로부터 선택되고; 선택적으로 여기서 상기 인간 세포는 배아 줄기 세포가 아니다. 선택적으로, 여기서, 상기 방법은 인간의 생식 계열 유전적 동일성을 변형시키지 않는다.
- [0213] 또 다른 관점에서, 본 발명은 전술된 바와 같은 표적 핵산 서열을 변형시키는 단계를 포함하는 인간 세포에서 유전자 발현을 변형시키는 방법을 제공하며, 여기서, 상기 핵산은 mRNA이고, 상기 기능성 모이어티는 리보뉴클레아제이며; 선택적으로 엔도뉴클레아제, 3' 엑소뉴클레아제 또는 5' 엑소뉴클레아제로부터 선택되고; 선택적으로, 여기서, 상기 인간 세포는 배아 줄기 세포가 아니다. 선택적으로, 여기서, 상기 방법은, 인간의 생식 계열 유전적 동일성을 변형시키지 않는다.
- [0214] 또 다른 관점에서, 본 발명은, 전술된 바와 같은 표적 핵산을 변형시키는 방법을 제공하며, 여기서 상기 방법은

45℃ 내지 100℃의 온도에서 수행된다. 바람직하게는, 상기 방법은 50℃ 이상의 온도에서 수행된다. 좀 더 바람직하게는, 상기 방법은 55℃ 내지 80℃의 온도에서 수행된다. 최적으로, 상기 방법은 60℃ 내지 65℃의 온도에서 수행된다. 선택적으로, 상기 방법은, 20℃ 내지 45℃의 온도에서 수행될 수 있다. 좀 더 바람직하게는, 30℃ 내지 45℃의 온도에서 수행된다. 더욱 바람직하게는, 37℃ 내지 45℃의 온도에서 수행되고; 선택적으로, 여기서, 표적 핵산은 인간 세포내 있다. 선택적으로, 상기 인간 세포는 배아 줄기 세포가 아니다. 선택적으로, 여기서, 상기 방법은, 인간의 생식 계열 유전적 동일성을 변형시키지 않는다.

[0215] 전술된 표적 핵산을 변형시키는 방법 중 어느 하나에서, 세포는 인간 세포일 수 있다. 선택적으로, 전술된 표적 핵산을 변형시키는 방법 중 어느 하나에서, 상기 인간 세포는, 배아 줄기 세포가 아닐 수 있다. 선택적으로, 상기 방법은, 인간의 생식 계열 유전적 동일성을 변형시키지 않는다.

[0216] 숙주 세포 (Host Cells)

[0217] 유리하게는, 여기서 개시된 방법은 광범위한 적용성이 있고, 본 발명의 숙주 세포는, 배양될 수 있는 임의의 인간 세포로부터 유래될 수 있다. 따라서, 본 발명은 전술된 바와 같은 방법에 의해 형질전환된 인간 세포를 제공한다. 본 발명은 이중 가닥 표적 폴리뉴클레오티드에서 표적 핵산 서열을 갖는 형질전환된 인간 세포를 제공하며, 상기 세포는 여기에 제공된 바와 같은 Cas 단백질 또는 폴리펩티드 및 여기에 제공된 바와 같은 적어도 하나의 표적화 RNA 분자를 포함하고, 및 발현 벡터는 상기 Cas 단백질 및 상기 표적화 RNA 분자 중 적어도 하나를 인코딩하는 핵산을 포함한다. 선택적으로, 상기 인간 세포는, 배아 줄기 세포가 아닐 수 있다.

[0218] 흔히 사용되는 숙주 세포는, 유전적으로 접근하기 쉽고, 배양될 수 있는 인간 세포를 포함하는, 예를 들어, (배아 줄기 세포가 아닌) 인간 세포를 포함하는 본 발명에 따른 사용을 위해 선택될 수 있다. 본 발명에 따른 사용에 바람직한 숙주 세포는, 높은 성장 속도를 통상적으로 나타내고, 쉽게 배양되거나 및/또는 형질전환되며, 짧은 세대 시간을 나타내고, 그들과 관련된 확립된 유전자원 (genetic resources)을 갖는 일반적인 것들, 또는 특정 조건하에서 이중 단백질의 최적 발현을 위해 선택, 변형 또는 합성되는 세포들이다. 관심의 단백질이 궁극적으로 특정 산업적, 또는 치료적 환경에 사용되는 본 발명의 바람직한 구체 예에서, 적절한 숙주 세포는, 관심의 단백질이 효율적으로 활용되는 원하는 특별한 조건 또는 세포내 환경에 기초하여 선택될 수 있다. 바람직하게는, 숙주 세포는 인간 세포일 것이다. 여기서 기재된 본 발명의 방법 및 사용은, 인간 세포의 게놈을 변형시키는데 사용될 수 있다.

[0219] 이하는 본 발명에 따라 사용된 Cas 단백질의 폴리뉴클레오티드 및 아미노산 서열이다.

[0220] [SEQ ID NO: 1] 지오바실러스 썬모테니트리피칸스 T12 Cas9 단백질 AA 서열

[0221] MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRLARSARRRLRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRV
EALDRKLNDELARILLHLAKRRGFRSNRKSERTNKENSTMLKHIEENQSISSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVARDDLEREIKLIFAKQREY
GNIVCTEAFEHEYISIWASQRPFAKDDIEKKVGFCTFEPKEKRAPKATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFHDVRTLL
NLPDDTRFKGLLYDRNTTLKENEKVRFLELGAYHKIRKAIDSVYKGGAAKSFRPIDFDTFGYALTMFKDDTDIRSYLNEYEQNGKRMENLADKVYDEELIE
ELLNLSFSKFGHLCLKALRNILPYMEQGEVYSTACERAGYTFTGPKKKQKTVLLPNIPPIANPVVMRALQARKVNAIIKKYGPSVSIHIELARELSQSF
ERRKMKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIVKFLWSEQNGKCAYSLQPIEIERLLEPGYTEVDHVIPYSRSLDDSYTNKVLVTOKENREKGNRT
PAEYLGLSERWQQFETVLTNKFSSKKRDRLLRLHYDENEENEFKNRNLNDTRYISRFLANFIREHLKFADSDDKQKVYTVNGRITAHLSRWNFNKNRE
ESNLHHAVDAAIVACTTPSDIARVTAIFYQRREQNELSKTDPQFPQWPHFADELQARLSKNPKESIKALNLGNYDNEKLESQPVFVSRMPKRSITGA
AHQETLRRYIGIDERSGKIQTIVVKKLSEIQLDKTGHPMYGKESDPRTYEAIRQLLEHNNDPKKAFQEPLYKPKKNGELGPIIRTIKIIDTTNQVIPLNDGK
TVAYNSNIVRVDFEKGDKYCVPIYITIDMMKGILPNKAIENPKPYSEWKEMTEDYTFRSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYYQITDSSN
GGLSLVSHDNNFSLRSIGSRTLKRFKYQVDVLGNIYKVRGEKRVGVASSSHSKAGETIRPL*

[0222] [SEQ ID NO: 7] 지오바실러스 썬모테니트리피칸스 T12 Cas9 DNA 서열

[0223] ATGAAGTATAAAATCGGTCTTGATATCGGCATTACGTCTATCGGTTGGGCTGTCATTAATTTGGACATTCCTCGCATCGAAGATTTAGGTGTCCGCATTTT
GACAGAGCGGAAAACCGGAAAACCGGGAGTCACTAGCTCTTCCACGTCGCCTCGCCCGTCCGCCGACGTCGTCTGCGCGTTCGAAACATCGACTGGAG
CGCATTCGCCGCTGTTTCGTCCGCAAGGAATTTTAAACGAAGGAAGAGCTGAACAAGCTGTTGAAAAAAGCACGAAATCGACGCTCGCAGCTTCGTGTT
GAAGCACTGGATCGAAAATAAATAACGATGAATTAGCCGCATCCTTCTTCATCTGGCTAAACGGCGTGGATTAGATCCAACCGCAAGAGTGAGCGCACC
AACAAGAAAAAGTACGATGCTCAACATATTGAAGAAAAACCAATCCATTCTTTCAAGTTACCGAACGGTTCGAGAAATGGTTGTCAAGGATCCGAAATTT
TCCCTGCACAAGCGTAATAAAGAGGATAATTACACCAACACTGTGCCCCGACGATCTTGAACGGGAAATCAAACGATTTTCGCCAAACAGCGCAATAT
GGGAACATCGTTTGACAGAAGCATTTGAACACGAGTATATTTCCATTTGGGCATCGCAACGCCCTTTTGCTTCTAAGGATGATATCGAGAAAAAGTCGGT
TTCTGTACGTTTGAGCCTAAAGAAAAACGCGCGCCAAAAGCAACATACATTCAGTCCTTACCGTCTGGGAACATATTAACAACTTCGTCTGTCTCC

CCGGGAGGCATCCGGGCGCTAACCAGATGATGAACGTCGCTTATATACAAGCAAGCATTTTCATAAAAAATAAAATCACCTTCCATGATGTTGCAACATTGCTTAACTTGCTGACGACACCCGTTTAAAGGTCTTTTATATGACCGAAACACCACGCTGAAGGAAAATGAGAAAGTTCGCTTCCTTGAACTCGGCGCTATCATAAAAACGAAAGCGATCGACAGCGTCTATGGCAAAGGAGCAGCAAAATCATTTCGTCGATTGATTTTGATACATTTGGCTACGCATTAACGATGTTTAAAGACGACACCGATTCGCACTTACTTGCGAAACGAATACGAACAAAATGGAAACGAATGGAAAATCTAGCGGATAAAGTCTATGATGAAGAATTGATTGAAGAACTTTTAACTTATCGTTTCTAAGTTTGGTCATCTATCCCTTAAAGCGCTTCGCAACATCCTTCCATATATGGAACAAGGCGAAGTCTACTCAACCGCTTGTGAACGAGCAGGATATACATTTACAGGGCCAAAGAAAAACAGAAAACGGTATTGCTGCCGAACATTCGCGCGATCGCCAATCCGGTCGTCATGCGCGCATGACACAGGACGCAAGTGGTCAATGCCATTATCAAAAAGTACGGCTCACCAGTCTCCATCCATATCGAACTGGCCCGGAACTATCACAATCCTTTGATGAACGAGTAAAAATGCAGAAAGAACGGAAGGAAACCGAAAGAAAAACGAAACTGCCATTGCGCAACTTGTGAATATGGGTGACGCTCAATCCAACCTGGGCTTGACATTGTGAAATTCAACTATGGAGCGAACAAAACGGAATGTGCTATTACTCCAACCGATCGAAATCGAGCGGTGCTCGAACCAGGCTATACAGAAGTCGACCATGTGATTCCATACAGCCGAGCTTGAGCAGATAGCTATACCAATAAAGTCTTGTGTTGACAAAGGAGAACCGTGAAGGAAACCGCACCACAGCTGAATATTTAGGATTAGGCTCAGAACGTTGGCAACAGTTCGAGACGTTTGTCTTGACAAATAAGCAGTTTTCGAAAAAGAGCGGGATCGACTCCTTCGGCTTCATTACGATGAAAACGAAGAAAATGAGTTTAAAAATCGTAATCTAAATGATACCGCTTATATCTCACGCTTCTTGCTAACTTTATTCGCGAACATCTCAAAATTCGCGACAGCGATGACAAACAAAAGTATACAGGTCAACGGCGTATTACCGCCATTTACGCAGCGTTGGAATTTTAAACAAAACCGGGAAATCGAATTGTCATCATGCCGTCGATGCTGCCATCGTCGCTGCACAACGCGAGCGATATCGCCGAGTCACCGCCTTCTATCAACGCGCGAACAAAACAAAGAACTGTCCAAAAAGACGGATCCGAGTTTCCGACGCTTGCCGCACTTTGCTGATGAACTGCAGGCGGTTTATCAAAAAATCCAAAGGAGAGTATAAAGCTCTCAATCTTGAAATTTATGATAACGAGAACTCGAATCGTTGAGCGGTTTGTCTCCCGAATGCCGAAGCGGAGCATAACAGGAGCGGCTCATCAAGAAACATTGCGCGTTATATCGGCATCGACGAACGAGCGGAAAAATACAGACGGTCGTCAAAAAGAACTATCCGAGATCCAACTGGATAAAACAGGTCAATTTCCCAATGTACGGAAAGAAAGCGATCCAAGGACATATGAAGCCATTGCGCAACGGTTGCTGAACATAACAATGACCCAAAAAGCGGTTTCAAGAGCCTCTGTATAAACCGAAGAAGAACGAGAACTAGGTCTATCATCCGAACAATCAAAATCATCGATACGACAAATCAAGTTATTCCGCTCAACGATGGCAAAACAGTCGCCTACAACAGCAACATCGTGCAGGTCGACGCTTTGAGAAAGATGGCAAATATTATTGTGTCCTATCTATACAATAGATATGATGAAAGGGATCTTGCCAAACAAAGCGATCGAGCCGAACAAACCGTACTCTGAGTGAAGGAAATGACGGAGGACTATACATTCCGATTGAGTCTATACCCAAATGATCTTATCCGTATCGAATTTCCCGAGAAAAACAATAAAGACTGTGTGGGGGAAGAAATCAAAATTAAGGATCTGTTCGCCTATTATCAAACCATCGACTCCTCCAAATGGAGGTTAAGTTTGGTTAGCCATGATAACAACCTTTTCGCTCCGAGCATCGGTTCAAGAACCTCAAACGATTGAGAAATACCAAGTAGATGTGCTAGGCAACATCTACAAAGTGAGAGGGGAAAGAGAGTTGGGTGGCGTCATCTTCTCATTCGAAAGCCGGGAACTATCCGTCGGTTATAA

[0224] 실시 예 1: 지오바실러스 썬모테니트리피칸스의 단리

[0225] G. 썬모테니트리피칸스는, 놀랍게도 혐기성 조건하에서 리그노셀룰로스 기질 (lignocellulosic substrates)을 분해할 수 있는 호열균에 대한 ± 500 단리군 (isolates)의 라이브러리의 검색 동안에 발견되었다. 처음에, ± 500 단리군의 라이브러리는 확립되고, 셀룰로오스 및 크실란 상에 단리에 의한 여러 차례의 선택 라운드 후에 110 단리군까지 줄인다. 110 단리군의 라이브러리는, 라이브러리의 79%를 나타내는 G. 썬모테니트리피칸스로, 지오바실러스 단리군 만으로 이루어진다.

[0226] 단리된 G. 썬모테니트리피칸스 균주는 "T12"로 명명된다. G. 썬모테니트리피칸스 T12 유래의 Cas9 단백질은 "gtCas9"로 명명된다.

[0227] 실시 예 2: 지오바실러스 썬모테니트리피칸스에서 Cas9에 대한 필수적인 공통 서열 (consensus sequences)을 정의

[0228] 다음의 데이터베이스 검색 및 정렬은 수행된다:

[0229] pBLAST 및 nBLAST는 G. 썬모테니트리피칸스 T12의 단백질 또는 유전자 서열이 질의 서열 (query sequence)로서 사용되는, 사내의 BLAST 서버상에서 수행된다. 이 데이터베이스는 2014년 5월에 최종 업데이트되었으며, 따라서 가장 최근에 추가된 지오바실러스 게놈을 함유하지 않지만, 보통 온라인 BLAST는 T12 서열의 공개를 방지하는데 사용되지 않았다. BLAST 검색에서 40%를 초과하는 것으로 동정된 서열 동일성은 도 1에 포함된다.

[0230] 더 최근의 서열 데이터를 포함하기 위해, (gtCas9에 가장 밀접하게 관련된) 지오바실러스 MAS1의 서열은, NCBI 웹사이트 상에서 PSI-BLAST를 수행하는데 사용된다 (Johnson et al., 2008 Nucleic Acids Res. 36(Web Server issue): W5-9). 2 연속 라운드의 PSI-BLAST는 수행되며, 여기서, 하기의 기준에 부합하는 서열만이 다음 라운드에 사용된다: 제1 라운드에서 96% 및 제2 및 제3 라운드에서 97%의 최소 서열 커버리지, 최소 동일성 40%, 종에 대하여 오직 하나의 가닥.

[0231] PSI-BLAST로부터 결과하는 서열뿐만 아니라, PSI-BLAST에 나타나지 않은 내부 서버 pBLAST로부터 T12와 40% 초과 동일성을 갖는 서열은, 현재 잘-특징화된 중온성 서열 및 현재 동정된 모든 호열성 서열도, 이들이 더 먼 관계인 경우와 함께, 정렬되며, 이로부터 근린-결합 트리는 구성된다 (도 1 참조). 정렬은 ClustalW를 사용하여

Mega6에서 수행되고, 그 이후 트리는 근린-결합 방법을 사용하여 구축되며, 부트스트랩 분석 (bootstrap analysis)은 1000 반복수 (replicates)를 사용하여 수행된다.

[0232] BLASTn이 질의 서열로서 지오바실러스 속 MAS1을 사용하여 수행된 경우, 오직 지오바실러스 속 JF8 Cas9는 88%의 동일성으로 동정되어, 유전자 수준에서 매우 적은 상동성을 나타낸다. 도 2는 클러스탈-정렬된 (Clustal-aligned) Cas9 유전자 서열의 근린-결합 트리이다.

[0233] G. 써모데니트리피칸스 T12, *A. naeslundii* 및 *S. pyogenes*의 단백질 서열은, BLOSUM62를 기본 설정 (default settings)으로 사용하여 CloneManager에서 이들을 정렬하여 단백질 도메인 상동성을 위해 더욱 분석된다 (도 3 참조).

[0234] **실시 예 3: CAS9의 기능에 필수적인 코어 아미노산 모티프 및 호열성 Cas9 뉴클레아제에서 열안정성을 부여하는 아미노산 모티프의 동정**

[0235] 전술된 정렬된 단백질 서열의 퍼센트 동일성은 도 1에 제공된다. gtCas9는 타입 II-C에 속한다. 타입 II-C 시스템의 가장 자주-연구되고, 최근에 결정화된 구조는, *Actinomyces naeslundii*에서 유래된다 (Jinek et al., 2014, Science 343: 1247997). 이 단백질 서열은 gtCas9에 대해 단지 20%의 동일성을 나타내지만, 고도로 보존된 잔기 (conserved residues)를 추정하는데 사용될 수 있다. 두 개의 잘-특징화된 타입 II-A 시스템 (*S. pyogenes* 및 *S. thermophilus*)은 또한 분석에 포함된다 (Jinek et al., 2014, Science 343: 1247997; Nishimasu et al., 2014, Cell 156: 935-949). 이들 4개의 단백질 서열의 정렬은, 도 3에 나타내고; 도 4는 *A. naeslundii* ('Ana-Cas9')에 대해 결정된 바와 같은 단백질 아키텍처 (protein architecture)를 나타낸다 (Jinek et al., 2014, Science 343: 1247997). t12 (gtCas9) 및 *Actinomyces naeslundii* 유래의 Cas9의 길이는, 매우 유사하고 (*A. naeslundii* 1101 aa, T12 1082 aa) 및 gtCas9는 유사한 단백질 아키텍처를 가질 것으로 예상되지만, cas9-Ana에 대한 전체 서열 동일성이 단지 20%이기 때문에, 이것은 아직 결정되지 않는다. *A. naeslundii* 및 *S. pyogenes* 유래의 Cas9에서 Jinek et al. (Jinek et al., 2014, Science 343: 1247997)에 의해 기재된 모든 활성 부위 잔기는 gtCas9에서 동정될 수 있다 (도 3 참조). PAM-결합 도메인은, *S. pyogenes* 타입 II-A 시스템에 대해 결정되어 있지만, 임의의 타입 II-C 시스템에 대해서는 결정되지 않고, 따라서 *S. pyogenes* 서열에서만 나타난다. 게다가, PAM-인식 부위는, CRISPR 시스템들 사이뿐만 아니라 동일한 시스템을 함유하는 종들 사이에서도, 강하게 변한다.

[0236] **실시 예 4: G. 써모데니트리피칸스 gtCas9의 PAM 서열의 결정**

[0237] 원핵생물 CRISPR 시스템이 적응 면역 시스템으로서 이들의 숙주에 도움이 되고 (Jinek et al., 2012, Science 337: 816-821), 및 신속하고 효과적인 유전 공학을 위해 사용될 수 있는 것으로 확립되었다 (Mali et al., 2013, Nat Methods 10: 957-963.).

[0238] Cas9 단백질은 타입 II CRISPR 시스템에 대한 서열-특이적 뉴클레아제로서 기능한다 (Makarova et al., 2011, Nat Rev Micro 9: 467-477). 반복 영역에 연결된 "스페이서" (표적)로 이루어진, 작은 crRNA 분자는, CRISPR 유전자좌의 전사 및 공정 산물 (processing products)이다. "스페이서"는 박테리오파지 및 유동적 유전 인자 (mobile genetic elements)의 게놈에서 자연적으로 기원하지만, 이들은 또한 유전 공학 공정 동안, 특이적 뉴클레오티드 서열을 표적으로 하도록 디자인될 수도 있다 (Bikard et al., 2013, Nucleic Acids Research 41: 7429-7437). crRNA 분자는 이들의 DNA 표적의 동정을 위한 가이드로서 Cas9에 의해 이용된다. 상기 스페이서 영역은 절단 DNA 영역에 대한 표적물인, "프로토스페이서"와 동일하다 (Brouns et al., 2012, Science 337: 808-809). 프로토스페이서 다음의, PAM (Protospacer Adjacent Motif)은, Cas9에 의한 표적의 인지를 위해 요구된다 (Jinek et al., 2012, Science 337: 816-821).

[0239] 타입 II 시스템에 대한 시험관 내 또는 생체 내 PAM-결정 연구를 수행하기 위해, 시스템의 CRISPR 어레이 (CRISPR array), tracrRNA-발현 모듈을 인실리코 예측하는 것은 필요하다. CRISPR 어레이는 crRNA 모듈의 동정을 위해 사용된다. tracrRNA-발현 서열은, 500 bp-윈도우 플랜킹 Cas9 (500 bp-window flanking Cas9) 내에 또는 Cas 유전자와 CRISPR 유전자좌 사이에 위치된다 (Chylinski, K., et al. (2014) Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* 42, 6091-6105). tracrRNA는 CRISPR 어레이의 직접 반복에 대한 높은 수준의 상보성을 갖는 5'-서열, 뒤이어 2개 이상의 스템-루프 구조 및 Rho-비의존 전사 종결 신호의 예측된 구조로 이루어져야 한다 (Ran, F.A., et al. (2015) In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* 520, 186-191). crRNA 및 tracrRNA 분자는 그 다음 키메라 sgRNA 모듈을 디자인하는데 사용될 수 있다. sgRNA의 5'-말단은, CRISPR 어레이의 절단된 (truncated) 20nt 길이의 스페이서

뒤이어 16-20nt 길이의 절단된 반복으로 이루어진다. 이 반복은 상응하는 절단된 anti-반복 및 tracrRNA 모듈의 스템 루프가 뒤따른다. sgRNA의 반복 및 anti-반복 부분은, 일반적으로 GAAA 링커에 연결된다 (Karvelis, T., *et al.* (2015) Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biol.* 16, 253).

[0240] G. 써모데니트리피칸스 T12 타입 IIc CRISPR 시스템의 cas 유전자들 (cas9 뒤이어 cas1 및 cas2 유전자)은, T12 염색체의 안티센스 가닥 (antisense strand)을 사용하여 전사된다. cas2 유전자는 100bp 길이의 DNA 단편이 되따르고, 전사시 다중 루프를 갖는 RNA 구조를 형성한다. 이 구조는 분명히 전사 종결인자 (transcriptional terminator)의 역할을 한다.

[0241] 11개의 반복 및 10개의 스페이서 서열을 갖는 CRISPR 어레이는, 전사 종결 서열의 업스트림에 위치되며, 어레이의 리더는 어레이의 5' 말단에 위치된다. tracrRNA로 전사되는 DNA 유전자좌는, cas9 유전자의 다운스트림에 존재할 것으로 예상된다. CRISPR 어레이로부터 36bp 길이의 반복을 갖는 cas9 유전자의 다운스트림 우측에 325bp 길이의 서열의 정렬은 (도 6에 나타낸) 반복과 거의 동일한 tracrRNA 유전자좌에 36bp 길이의 서열이 있음이 밝혀졌다. 이 결과는 tracrRNA 유전자좌의 전사의 방향이 CRISPR 어레이의 전사의 방향과 반대이어야 한다는 결론을 이끌어냈다. 결과적으로, tracrRNA의 5'말단은 crRNA의 3'-말단에 상보적일 것이어서, -Cas9가 필요로 하는- 이중-RNA 분자의 형성으로 이어진다.

[0242] 실시 예 5: 무작위화된 PAM을 갖는 표적 발생

[0243] G. 써모데니트리피칸스 T12 균주의 CRISPR II 유전자좌로부터 두 개의 다른 스페이서는, 주형 (template)으로서 G. 써모데니트리피칸스 T12 게놈 DNA를 사용하는 PCR에 의해 증폭된다. 2쌍의 축퇴 프라이머 (degenerate primers)는 각 스페이서의 증폭을 위해 사용된다:

[0244] 먼저, "프로토스페이서" 단편의 업스트림에 6개의 무작위 뉴클레오티드의 도입을 일으키는 쌍은 사용되어, 무작위화된 PAM 서열을 갖는 프로토스페이서의 풀 (pool)의 생성으로 이어진다.

[0245] 둘째로, "프로토스페이서" 단편의 다운스트림에 6개의 무작위 뉴클레오티드의 도입을 일으키는 쌍은 사용되어, 무작위화된 PAM 서열을 갖는 프로토스페이서의 풀의 생성으로 이어진다.

[0246] 생성된 단편은 pNW33n 벡터에 결합되어, 6-뉴클레오티드 길이의 각 PAMs의 모든 가능한 4096의 다른 조합을 갖는, "프로토스페이서" 구축물의 4 풀을 생성한다. 조립된 DNA는 G. 써모데니트리피칸스 T12 세포의 형질전환을 위해 사용된다. 세포는 클로람페니콜 선택 (chloramphenicol selection) 상에 평판 배양되고 및 각 프로토스페이서 풀 유래의 2×10^6 이상의 세포는 풀이 될 (pooled) 것이다. 플라스미드 DNA는 풀로부터 추출되고, 표적 영역은 PCR 증폭될 것이며, 생성물은 딥 시퀀싱 (deep sequencing)을 위해 보내진다. 가장 적은 리드 (reads)를 갖는 PAMs은 활성인 것으로 고려될 것이고, 상기 공정은 이들 PAMs을 갖는 스페이서를 함유하는 pNW33n 구축물을 이용하여 오직 반복될 것이다. G. 써모데니트리피칸스 T12의 감소된 형질전환 효율은, PAMs의 활성을 확인하는 것이다.

[0247] 실시 예 6. gtCas9에 대한 PAM 서열의 체외 결정

[0248] pRham:cas9gt 벡터의 구축

[0249] cas69gt 유전자는 BG6927 및 BG6928 프라이머를 사용하여, G. 써모데니트리피칸스 T12 게놈으로부터 PCR 증폭되고, pRham C-His Kan Vector (Lucigen)와 하나의 혼합물로 조합된다. 상기 혼합물은 제공된 프로토콜에 따라 E. cloni 열-수용성 세포 (thermo-competent cells)를 형질전환시키는데 사용된다. 100 μ l의 형질전환 혼합물은 LB+50카나마이신 플레이트 (LB+50kanamycin plates)에 평판 배양하여 37 $^{\circ}$ C에서 하룻밤 동안 성장된다. 형성된 E. cloni::pRham:cas9gt 단일 콜로니 중에서 무작위로 3개를 선택하고, 50 μ g/ml 카나마이신을 함유하는 10ml LB 배지에 접종하였다. 글리세롤 스톱 (Glycerol stocks)은, 20%(v/v)의 최종 농도까지 각 배양물 (culture) 1ml에 멸균 글리세롤을 첨가하여 배양물로부터 준비된다. 글리세롤 스톱은 -80 $^{\circ}$ C에서 저장된다. 각 배양물에서 나머지 9ml는 "GeneJET Plasmid Miniprep Kit" (Thermoscientific) 프로토콜에 따라 플라스미드 단리를 위해 사용된다. 플라스미드는 cas9gt의 서열 검증 (verification)을 위해 보내지고, 플라스미드 중 하나는 올바른 서열을 가진 유전자를 포함하는 것으로 검증되었다. 상응하는 배양물은, gtCas9의 이중 발현 및 정제를 위해 더욱 사용된다.

[0250] E. cloni::pRham:cas9gt 벡터에서 gtCas9의 이종숙주 발현 (Heterologous Expression)

- [0251] E. cloni::pRham:cas9gt 전배양물 (preculture)은 상응하는 글리세롤 스톱으로 10ml LB+50카나마이신을 접종하여 준비된다. 37℃ 및 180rpm에서 하룻밤 동안 성장시킨 후, 전배양물 2ml는, 200ml의 LB+50카나마이신 배지를 접종하기 위해 사용된다. E. cloni::pRham:cas9gt 배양물은, 37℃, 180rpm에서, OD₆₀₀이 0.7이 될때까지 배양된다. gtCas9 발현은 그 다음 L-람노오스를 0.2% w/v의 최종 농도로 첨가하여 유도된다. 발현은 8시간 동안 진행된 후, 배양물을 4700rpm, 10분 동안 4℃로 원심분리하여 세포를 얻었다. 배지는 폐기되고, 펠렛화된 세포는 -20℃에서 보관하거나 또는 다음의 프로토콜에 따라 무세포 추출물 (CFE)의 제조를 위해 사용된다:
- [0252] 1. 20ml 초음파분해 버퍼 (20mM 인산 나트륨 버퍼 (pH=7.5), 100mM NaCl, 5mM MgCl₂, 5%(v/v) 글리세롤, 1mM DTT)에 펠릿을 재현탁.
- [0253] 2. 초음파분해로 1ml의 세포를 분쇄 (30초의 8 펄스, 그 사이에서 얼음에서 20초 동안 냉각).
- [0254] 3. 불용성 부분을 침전시키기 위하여, 4℃, 35000g에서 15분 동안 원심분리.
- [0255] 4. 상등액을 제거하고 4℃ 또는 얼음에 보관.
- [0256] gtCas9를 위한 sgRNA 모듈을 표적화하는 PAM 라이브러리의 디자인 및 구축
- [0257] G. 써모데나트리피칸스 T12 균주의 게놈에서 tracrRNA 발현 DNA 모듈의 인실리코로 결정 후에 (상기 실시 예 4 참조), 단일 분자에서 CRISP/Cas9 시스템의 crRNA 및 tracrRNA 모듈들을 조합한 단일 가이드 (sg)RNA 발현 DNA 모듈은 디자인된다. sgRNA의 5'-말단의 스페이서는, 플라스미드 라이브러리의 프로토스페이서에 상보적으로 디자인되고, 모듈은 T7 프로모터의 전사 조절하에 놓여진다. pT7_sgRNA DNA 모듈은 베이스클리어 (Baseclear)에 의해 합성되고, pUC57 벡터에 수용되어, pUC57:pT7_sgRNA 벡터를 형성한다. DH5α 수용능 대장균 세포 (NEB)는 벡터로 형질전환되고, 형질전환 혼합물은 100μg/ml 암피실린을 함유하는 LB-한천 플레이트 상에 평판 배양된다. 플레이트 (plates)는 37℃에서 밤새 배양된다. 형성된 단일 콜로니 중 3개는 100μg/ml 암피실린을 함유하는 10 ml LB 배지에 접종된다. 글리세롤 스톱은, 20%(v/v)의 최종 농도까지 각 배양물 1ml에 멸균 글리세롤을 첨가하여 배양물로부터 준비된다. 글리세롤 스톱은 -80℃에서 보관된다. 각 배양물에서 나머지 9ml는 "GeneJET Plasmid Miniprep Kit" (Thermoscientific) 프로토콜에 따라 플라스미드 단리를 위해 사용된다. 단리된 플라스미드는, pT7_sgRNA 모듈의 증폭을 위한 PCR 주형으로 사용된다. 218bp 길이의 pT7_sgRNA DNA 모듈 (이의 처음 18bp는 pT7에 상응함)은, 프라이머 BG6574 및 BG6575를 사용하여 얻어진다. 완전한 PCR 혼합물은, 1.5% 아가로스 겔에서 실행된다. 원하는 크기의 밴드는 잘라내고, "Zymoclean™ Gel DNA Recovery Kit" 프로토콜에 따라 정제된다.
- [0258] 시험관 내 전사 (IVT)는, "HiScribe™ T7 고 수율 RNA 합성 키트" (NEB)를 사용하여 수행된다. 정제된 pT7_sgRNA DNA 모듈은, 주형으로 사용된다. IVT 혼합물은, 동일 부피의 RNA 로딩 염료 (RNA loading dye) (NEB)와 혼합되고, 70℃에서 15분 동안 가열되어 2차 구조를 파괴하였다. 열처리된 IVT 혼합물은 변성 우레아-PAGE 상에서 실행되고, 그 결과로 생긴 폴리아크릴아미드 겔은 염색을 목적으로 10μl의 SYBR Gold (Invitrogen)를 함유하는 100ml 0.5x TBE 버퍼에서 10분 동안 처리된다. 원하는 크기 (200nt)의 밴드는 잘라내고, sgRNA는 다음의 RNA 정제 프로토콜에 따라 정제된다:
- [0259] 1. 메스 (scalpel)로 RNA 겔 단편을 절단하고, RNA 용출 버퍼 1ml를 첨가하여, 밤새 실온에서 놓음.
- [0260] 2. 새로운 1.5ml 튜브에 330μl 앨리쿼트 (aliquots)로 나눔.
- [0261] 3. 3 부피 (990μl)의 사전-냉각된 (-20℃) 100% EtOH을 첨가.
- [0262] 4. -20℃에서 60분 동안 배양.
- [0263] 5. 실온에서 소형원심분리기로 13000rpm에서 20분 동안 원심분리.
- [0264] 6. EtOH를 제거하고 1ml 70% EtOH로 펠릿을 세척.
- [0265] 7. 실온에서 소형원심분리기로 13000rpm으로 5분 동안 원심분리.
- [0266] 8. 990μl의 상등액을 제거.
- [0267] 9. 남은 EtOH를 열혼합기 (thermomixer)에서 55℃로 15-20분 동안 증발.
- [0268] 10. 20μl MQ에 재현탁된 펠릿을 -20℃에서 보관.

- [0269] 7nt 길이의 PAM 라이브러리의 디자인 및 구축, 및 라이브러리의 선형화 (linearization)
- [0270] PAM 라이브러리의 디자인 및 구축은 pNW33n 벡터에 기초한다. 20bp 길이의 프로토스페이서는 벡터에 도입되고, 7 축퇴 뉴클레오타이드 길이의 서열에 의해 이의 3' 측의 옆에 위치되며; 상기 축퇴 서열은 PAM으로 역할을 하고, 프로토스페이서가 우측 PAM에 의해 측면에 위치되는 경우, 이것은 Cas9로 로딩되고 절단된 sgRNA에 의해 표적으로서 인지될 수 있다. PAM 라이브러리는 다음의 프로토콜에 따라 제조된다:
- [0271] 1. 단일 가닥 DNA 올리고 1 (BG6494) 및 2 (BG6495)를 어닐링하여 SpPAM 이중 가닥 DNA 인서트 (DNA insert)를 준비.
- [0272] I. 10 μ l 10x NEBuffer 2.1
- [0273] II. 1 μ l 50 μ M 올리고 1 (~1.125 μ g)
- [0274] III. 1 μ l 50 μ M 올리고 2 (~1.125 μ g)
- [0275] IV. 85 μ l MQ
- [0276] V. 혼합물을 94 $^{\circ}$ C에서 5분 동안 배양하고, 37 $^{\circ}$ C에서 0.03 $^{\circ}$ C/sec의 속도로 냉각.
- [0277] 2. 어닐링된 올리고 혼합물에 1 μ l Klenow 3' \rightarrow 5' 엑소⁻ 폴리머라제 (NEB)를 첨가한 후, 2.5 μ l의 10 μ M dNTPs를 첨가. 37 $^{\circ}$ C에서 1시간, 그 다음 75 $^{\circ}$ C에서 20분 동안 배양.
- [0278] 3. 46 μ l의 어닐링 혼합물에 2 μ l의 HF-BamHI 및 2 μ l의 BspHI 제한 효소를 첨가. 37 $^{\circ}$ C에서 1시간 동안 배양. 이 공정은 점착성 말단 (sticky ends)을 갖는 SpPAMbb 인서트로 이어질 것임. Zymo DNA 세정 및 농축기 키트 (Zymo Research)를 사용하여, 생성된 인서트를 세정.
- [0279] 4. HF-BamHI 및 BspHI (NEB)로 pNW33n을 소화 (Digest), 및 Zymo DNA 세정 및 농축기 키트 (Zymo Research)를 사용하여, 점착성 말단을 갖는 3.400bp 길이의 선형 pNW33nbb 단편을 정제.
- [0280] 5. 제공된 프로토콜에 따라 NEB T4 리가아제를 사용하여 11ng의 SPPAMbb 인서트로 50ng의 pNW33nBB를 절찰. Zymo DNA 세정 및 농축기 키트 (Zymo Research)를 사용하여 절찰 혼합물 (ligation mixture)을 정제.
- [0281] 6. DH10b 전기-수용성 세포 (500ng의 DNA를 가진 200 μ l의 세포)를 형질전환. SOC 배지에 세포 (800 μ l SOC에 200 μ l 세포)를 1시간 동안 회수한 다음, 회수된 세포로 50ml의 LB+12.5 μ g/ml 크로람페니콜에 접종. 배양물을 37 $^{\circ}$ C에서 180rpm으로 밤새 배양.
- [0282] 7. JetStar 2.0 maxiprep kit (GENOMED)를 사용하여 배양물로부터 플라스미드 DNA를 단리.
- [0283] 8. 단리된 플라스미드를 선형화하기 위해 제공된 프로토콜에 따라 SapI (NEB) 제한효소를 사용.
- [0284] PAM 결정 반응의 디자인 및 실행
- [0285] 다음의 절단 반응은, 표적화된 프로토스페이서의 3' 말단의 다운스트림에 우측 PAM을 함유하는 PAM 라이브러리 멤버로 dsDNA 절단 (dsDNA breaks)의 gtCas9-유도 도입을 위해 설정된다:
- [0286] 1. 반응 당 2.5 μ g의 E. cloni::pRham:cas9gt CFE
- [0287] 2. 30nM 최종 농도로 sgRNA
- [0288] 3. 반응 당 200ng의 선형화된 PAM 라이브러리
- [0289] 4. 2 μ l의 절단 버퍼 (100mM 인산 나트륨 버퍼 (pH=7.5), 500mM NaCl, 25mM MgCl₂, 25%(v/v) 글리세롤, 5mM DTT)
- [0290] 5. 20 μ l 최종 부피까지 MQ 수 (MQ water)
- [0291] 반응은 60 $^{\circ}$ C에서 1시간 동안 배양되고, 4 μ l의 6x 겔 로딩 염료 (NEB)를 첨가 후에 중단된다. 반응 혼합물은 그 다음 1% 아가로즈 겔에 로딩된다. 상기 겔은 100V에서 1시간 15분간 전기영동한 후, 10 μ l의 SYBR Gold 염료 (ThermoFisher)를 함유하는 100ml 0.5xTAE 버퍼에서 30분 동안 배양된다. DNA 밴드를 청색광으로 가시화한 후에, 성공적으로 절단 및 PAM을 함유하는 DNA 단편에 상응하는 밴드는 겔에서 잘라내고, 겔은 제공된 프로토콜에 따라 "ZymocleanTM Gel DNA Recovery Kit"을 사용하여 정제된다.

[0292] 시퀀싱을 위한 PAM-함유 gtCas9 절단된 DNA 단편의 태깅 (Tagging)

[0293] Cas9-유도된 DNA 절단은, 보통 PAM 서열에 가까운 쪽에, 프로토스페이서의 3rd와 4th 뉴클레오타이드 사이에 도입된다. 결과적으로, 서열을 추가하고, PAM 서열을 결정하기 위해, 절단된 DNA 단편의 PAM-함유 부분을 PCR 증폭할 수 있는 한 쌍의 프라이머를 디자인하는 것은 가능하지 않다. 이 목적을 위해, 5-단계 공정은 사용된다:

[0294] **단계 1:** Taq 폴리머라제로 A-테일링 (A-Tailing)

[0295] A-테일링은, Taq 폴리머라제를 사용하여 무딘, 이중-가닥 DNA 분자의 3' 말단에 주형이 없는 아데닌 (non-templated adenine)을 추가하는 과정이다.

[0296] 반응 성분:

[0297] · gtCas9-절단 및 PAM-함유 DNA 단편 - 200ng

[0298] · 10X ThermoPol® Buffer (NEB) - 5 μ l

[0299] · 1 mM dATP - 10 μ l

[0300] · Taq DNA 폴리머라제 (NEB) - 0.2 μ l

[0301] · H₂O - 50 μ l까지의 최종 반응 부피

[0302] · 배양 시간 - 20분

[0303] · 배양 온도 - 72°C

[0304] **단계 2:** 시퀀싱 어댑터 (sequencing adaptors)의 구축

[0305] 2개의 상보적인 짧은 ssDNA 올리고뉴클레오타이드는 인산화되고, 어닐링되어 단계 1 유래의 DNA 단편의 PAM-근위 부위 (PAM-proximal site)에 대한 시퀀싱 어댑터를 형성한다. 올리고뉴클레오타이드 중 하나는, A-테일화된 단편에 어댑터의 결합을 촉진하기 위해 이의 3' 말단에 부가적인 티민을 갖는다.

[0306] 어댑터 올리고뉴클레오타이드 인산화 (각 올리고에 대한 개별 인산화 반응)

[0307] · 100 μ M 올리고뉴클레오타이드 스톡 - 2 μ l

[0308] · 10X T4 DNA 리가아제 버퍼 (NEB) - 2 μ l

[0309] · 멸균 MQ 수 - 15 μ l

[0310] · T4 폴리뉴클레오타이드 키나제 (NEB) - 1 μ l

[0311] · 배양 시간 - 60분

[0312] · 배양 온도 - 37°C

[0313] · T4 PNK 비활성화 - 65°C에서 20분

[0314] 인산화된 올리고뉴클레오타이드의 어닐링

[0315] · 상응하는 인산화 혼합물 유래의 올리고뉴클레오타이드 1 - 5 μ l

[0316] · 상응하는 인산화 혼합물 유래의 올리고뉴클레오타이드 1 - 5 μ l

[0317] · 멸균 MQ 수 - 90 μ l

[0318] · 인산화된 올리고를 95°C에서 3분 동안 배양. ~30분 내지 1시간 동안 실온에서 천천히 반응을 냉각.

[0319] **단계 3:** 시퀀싱 어댑터로 gtCas9-절단된, A-테일화된 단편의 결합

[0320] 단계 1 및 2의 생성물은 다음의 프로토콜에 따라 결합된다:

[0321] · 10X T4 DNA 리가아제 버퍼 - 2 μ l

[0322] · 생성물 단계 1 - 50 ng

[0323] · 생성물 단계 2 - 4 ng

- [0324] · T4 DNA 리가아제 - $1\mu\text{l}$
- [0325] · 멸균 MQ 수 - $20\mu\text{l}$ 까지
- [0326] · 배양 시간 - 10분
- [0327] · 배양 온도 - $20\sim 25^{\circ}\text{C}$
- [0328] · 65°C 에서 10분 동안 가열 비활성화

[0329] **단계 4:** 150-뉴클레오타이드 길이의 PAM-함유 단편의 PCR 증폭

[0330] 단계 4의 결합 혼합물 $5\mu\text{l}$ 는, Q5 DNA 폴리머라제 (NEB)를 사용하여 PCR 증폭을 위한 주형으로 사용된다. 단계 2 유래의 티민 연장을 갖는 올리고뉴클레오타이드는, 정방향 프라이머로 사용되고, 역방향 프라이머는 PAM 서열의 다운스트림에 150 뉴클레오타이드를 어닐링하도록 디자인된다.

[0331] 동일한 서열은 주형으로서 비-gtCas9 처리된 PAM-라이브러리 DNA를 사용하여 증폭된다. PCR 생성물 모두는 겔 정제되고, Illumina HiSeq 2500 쌍-말단 시퀀싱 (paired-end sequencing) (Baseclear)을 위해 보내진다.

[0332] 시퀀싱 결과의 분석 및 후보 PAM 서열 (candidate PAM sequence)의 결정

[0333] 시퀀싱 결과를 분석한 후, 다음의 빈도 행렬 (frequency matrices)은 구축된다. 행렬은 gtCas9 소화된 (digested) 및 비-소화된 라이브러리의 모든 PAM 위치에서 각 뉴클레오타이드의 상대 존재비 (abundance)를 묘사한다:

표 1

[0334]

비-소화된	pos1	pos2	pos3	pos4	pos5	pos6	pos7
A	19.22	20.83	19.12	24.43	24.59	21.75	18.22
C	34.75	30	31.9	30.54	25.96	27.9	27.17
T	19.16	22.19	25.34	21.28	26.09	26	21.56
G	26.87	26.98	23.64	23.75	23.36	24.35	33.05
소화된	pos1	pos2	pos3	pos4	pos5	pos6	pos7
A	10.63	18.65	14.6	14.49	3.36	8.66	27.54
C	66.22	49.59	56.82	60.35	92.4	62.26	34.94
T	8.09	11.21	19.12	12.15	2.35	14.66	5.58
G	15.05	20.54	9.45	13.01	1.89	14.43	31.94

[0335] 이들 결과는 5th PAM 위치에서 시토신을 가진 표적에 대한 명확한 선호도 및 처음 4 PAM 위치에서 시토신을 갖는 표적에 대한 선호도를 나타낸다.

[0336] **실시 예 7: gtCas9에 대한 인실리코 PAM 예측**

[0337] PAMs의 인실리코 예측은, 충분한 프로토스페이스 서열이 게놈 데이터베이스에서 이용 가능하다면 가능하다. gtCas9 PAM의 인실리코 예측은, GenBank와 같은 게놈 데이터베이스의 서열과 비교하여 G. 썬모데니트리피칸스 T12 균주의 게놈에서 CRISPR 어레이 유래의 스페이스의 히트 (hits)의 동정으로 시작한다. "CRISPR 화인더" (<http://crispr.u-psud.fr/Server/>) 도구는 T12에서 후보 CRISPR 유전자좌를 동정하는데 사용된다. 동정된 CRISPR 유전자좌 출력은 그 다음, 선택된 데이터베이스를 검색하고, 일치하는 프로스페이스로 출력을 제공하는, "CRISPR 표적" (http://bioanalysis.otago.ac.nz/CRISPRTarget/crispr_analysis.html) 도구에 로딩된다. 이들 프로토스페이스 서열은 그 다음 유일무이한 히트 (unique hits) 및 스페이스에 대한 상보성에 대해 스크리닝된다 - 예를 들어, 시드 서열에서 불일치는 유망한 거짓 양성 (false positive) 히트인 것으로 고려되고, 추후 분석에서 배제된다. 프로파지 서열 (prophage sequences) 및 (인테그레이션된) 플라스미드에 대해 동일성을 갖는 히트는 얻어진 히트가 참 양성 (true positives)인 것을 입증한다. 전반적으로, 이 과정은 6개 단일 히트를 산출한다 (도 7). 그 결과로서, 남아있는, 유일무이한 프로토스페이스 히트의 플랜킹 영역 (Type II gtCas 뉴클레아제에 대해 3')은 WebLogo (<http://weblogo.berkeley.edu/logo.cgi>) (Crooks GE, Hon G, Chandonia JM, Brenner SE WebLogo: A sequence logo generator, *Genome Research*, 14:1188-1190, (2004)) 도구를 사용하여

컨센서스 서열 (consensus sequences)에 대해 정렬 및 비교된다 (도 8).

[0338] 인실리코 결과는, PAM 서열의 5th 잔기의 동일성에 대한 편중 (bias)이 시토신인 것으로 시험관 내 PAM 동일성 실험 결과 (실시 예 6 참조)와 비슷하다.

[0339] **실시 예 8: gtCas9에 대한 8 뉴클레오타이드 길이의 PAM 서열의 결정**

[0340] 실시 예 8의 인실리코 데이터는, gtCas9가 8th 위치에서 아데노신에 대해 약간의 선호도가 있음을 시사하고, 따라서 추가의 PAM 결정 실험은 수행되며, 여기서 PAM 서열의 8th 위치는 또한 시험된다. 이것은 프로토스페이서의 3' 말단에서 5th 및 8th 위치들 사이에서 연장된 것으로 확인된 중온성 *Brevibacillus laterosporus* SSP360D4 (Karvelis et al., 2015) Cas9 PAM 서열의 특징과 일치한다.

[0341] PAMs의 특이적 8 뉴클레오타이드-길이 변이체는, gtCas9로 시험된다:

[0342] 1) CNCCCCAC [SEQ ID NO: 17],

[0343] 2) CCCCCAG [SEQ ID NO: 18],

[0344] 3) CCCCCAA [SEQ ID NO: 11],

[0345] 4) CCCCCAT [SEQ ID NO: 19],

[0346] 5) CCCCCAC [SEQ ID NO: 20],

[0347] 6) NNNNTNNC (음의 대조군 PAM)

[0348] 정제된 gtCas9 및 이전과 동일한 sgRNA (실시 예 6 참조)로 이들 (비-선형화된) 플라스미드를 표적화하는 60°C에서 시험관 내 절단 분석을 수행한 후에, CCCCCAA [SEQ ID NO: 11] 서열이 PAM으로 사용된 경우 증가된 gtCas9 절단 활성은 관찰된다 (도 9). 그러나, 절단 활성은 모든 시험된 PAM 서열에 대해 명확하게 검출 가능하였으며, 음성 대조군 PAM 서열에 대해서도 희미한 절단 밴드가 관찰되었다. 특정 이론에 구속되는 것을 원하지는 않지만, 높은 gtCas9 농도의 사용은, 음성 대조군에서 관찰된 절단에 기여할 가능성이 있다. 시험관 내 분석에서 높은 Cas9 농도는, 절대 PAM 요건 없이 Cas9-유도된 DNA 절단으로 이어지는 것으로 일반적으로 관찰된다.

[0349] Cas9 농도는 일반적으로 Cas9 유도된 DNA 절단의 효율에 영향을 미치는 것으로 알려져 있다 (높은 Cas9 농도는, 더 높은 Cas9 활성을 결과한다). 이것은, CCCCCAA [SEQ ID NO: 11] PAM 서열 및 다른 gtCas9 농도로 표적화된 플라스미드를 사용하여 시험관 내 분석을 수행하는 경우 또한 관찰된다 (도 10).

[0350] 전술된 바와 같은 시험관 내 분석을 위한 CCCCCAA [SEQ ID NO: 11] PAM 서열로 표적화된 플라스미드는 38 내지 78°C의 넓은 온도 범위에서 수행된다 (도 11). 놀랍게도, gtCas9는 모든 온도에서 활성이었고, 40.1 내지 64.9°C에서 가장 높은 활성을 보인다.

[0351] 따라서, 지오바실러스 중 유래의 Cas9의 최적 온도 범위는, 현재까지 특징화된 Cas9 단백질의 온도 범위보다 훨씬 높다. 유사하게, 이것이 뉴클레아제 활성을 보유하는 범위의 상한은, 공지된 Cas9 단백질의 것보다 훨씬 높다. 더 높은 최적 온도 및 기능적 범위는, 고온의 유전 공학에서 상당한 장점을 제공하며, 따라서 상승된 온도에서 수행된 다양한 산업, 농업 및 제약 공정에서 유용성을 갖는, 호열성 유기체의 게놈을 편집하는데, 상당한 장점을 제공한다.

[0352] **실시 예 9: gtCas9 및 8개의 뉴클레오타이드 길이 PAM 서열로 바실러스 스미시이 ET138의 생체 내 게놈 편집**

[0353] 8개의 뉴클레오타이드 PAM이 또한 생체 내에서 gtCas9에 의해 인지되는지를 확인하기 위해, 실험은 55°C에서 바실러스 스미시이 ET138의 게놈에서 pyrF 유전자가 결손되도록 디자인된다.

[0354] 이 방법은 표적 (pyrF) 유전자의 업스트림 및 다운스트림에 상보적인 영역이 B. 스미시이 ET 138 세포에 제공되는 상동 재조합 주형 구축물을 제공하는 것에 의존한다. 주형의 도입은, 상동 재조합의 과정이 (pyrF 유전자가 없는) 상동 재조합 주형을 게놈으로 도입하는데 사용되는 것을 허용하여 세포의 게놈에서 WT pyrF 유전자를 대체한다.

[0355] 상동 재조합 구축물에 gtCas9 및 sgRNA의 포함은, WT pyrF를 함유하는 박테리아 게놈으로 이중 가닥 DNA 절단 (DSDBs)를 도입하는데 사용될 수 있다. 박테리아 게놈의 DSDBs는 통상적으로 세포 죽음을 결과한다. 따라서, WT pyrF에서 서열을 인식하는 sgRNA는, DSDB 및 단지 WT pyrF를 함유하는 세포의 죽음을 결과할 수 있다. DSDB의

도입은 또한 gtCas9에 의해 인식되는 프로토스페이서의 3' 말단에서 다운스트림에 위치하는 적절한 PAM 서열에 의존한다.

- [0356] pNW33n 플라스미드는 클론 (clone)을 만들기 위한 백본으로 사용된다:
- [0357] i) 사내에서 개발된 글루코오스 억제성 프로모터의 제어하에 있는 cas9gt 유전자; 및
- [0358] ii) B. 스미시이 ET138의 게놈으로부터 pyrF 유전자의 결손을 결과하는 상동 재조합을 위한 주형으로서 B. 스미시이 ET138의 게놈에서 pyrF 유전자의 1kb 업스트림 및 1kb 다운스트림 영역; 및
- [0359] iii) 구성적 프로모터의 전사 조절하에 있는 단일 가이드 RNA (sgRNA) 발현 모듈.
- [0360] 3개의 개별 구축물은 생성되고, 여기서, gtCas9를 게놈 내에 이의 특이적 DNA 표적 (또한, 스페이서로 알려짐)으로 안내하는 서열에 상응하는, 처음 20개 뉴클레오티드에서 단일 가이드 RNA의 서열은 다르다. 3개의 다른 스페이서 서열은, B. 스미시이 ET138의 pyrF 유전자에서 3개의 다른 후보 프로토스페이서를 표적으로 하도록 디자인된다. 상기 구축물은 여기서 구축물 1, 2 및 3으로 각각 지칭된다.
- [0361] 3개의 다른 표적화된 프로토스페이서는, 이들의 3'-말단에서 다음의 후보 PAM 서열을 갖는다:
- [0362] 1. TCCATTCC (시험관 내 분석의 결과에 따른 음성 대조군; 구축물 번호 3에 인코딩된 sgRNA에 의해 표적화된 프로토스페이서의 3'-말단)
- [0363] 2. ATCCCCAA (구축물 번호 1에 인코딩된 sgRNA에 의해 표적화된 프로토스페이서의 3'-말단 [SEQ ID NO: 21])
- [0364] 3. ACGGCCAA (구축물 번호 2에 인코딩된 sgRNA에 의해 표적화된 프로토스페이서의 3'-말단, [SEQ ID NO: 22])
- [0365] B. 스미시이 ET 138 세포를 3개의 구축물 중 하나로 형질전환 및 선택 플레이트 상에 도말한 후에, 다음의 결과는 얻어진다:
- [0366] 1. 세포가 3' 말단에서 음성 대조군 TCCATTCC PAM 서열을 갖는 프로토스페이서를 표적으로 하는 구축물로 형질전환된 경우 (구축물 번호 3), 형질전환 효율은 영향을 받지 않았다 (도 12 A). 콜로니의 수는, pNW33n 양성 대조군 구축물로 형질전환 후에 콜로니의 수와 같은 범위였다 (도 12 B). pyrF 유전자가 결손된 콜로니를 스크리닝하기 위해 콜로니 PCR에 적용된 15개의 콜로니 중에서, 결손 유전자형 -2.1kb 예상 밴드 크기-을 나타내지 않았고, 모두 야생형 -2.9kb 예상된 밴드 크기-이다 (도 13). 이것은 시험된 PAM이 실제로 생체 내에서 gtCas9에 의해 인지되지 않음을 나타낸다.
- [0367] 2. 세포가 구축물 번호 1로 형질전환된 경우, 양성 대조군 (pNW33n으로 형질전환된 세포)과 비교했을 때 소수의 콜로니는 얻어진다 (도 12 C). 20개의 콜로니는 콜로니 PCR에 적용되어 pyrF 유전자가 결손된 콜로니를 스크린한다. 대다수 (19개)의 콜로니는, 야생형 및 pyrF 결손 유전자형 모두를 함유하는 반면, 하나의 콜로니는 pyrF 결손 유전자형을 갖는다 (도 14). 이 결과는, WT 만의 유전자형이 관찰되지 않았기 때문에 PAM 서열 ATCCCCAA [SEQ ID NO: 21]이 gtCas9에 의해 생체 내에서 인지된다는 것을 나타낸다. 감소된 형질전환 효율은 또한 세포군의 비율이 감소되었음을 나타내고, 이는 gtCas9에 의한 성공적인 표적화로 인해 DSDB에 의한 WT 만의 유전자형 세포로 유발된 세포 죽음에 기인할 수 있다.
- [0368] 3. 세포가 구축물 번호 2로 형질전환된 경우, 콜로니는 얻어지지 않는다 (도 12 D). 콜로니의 결핍은 모든 세포군이 gtCas9에 의해 성공적으로 표적화되어, DSDB에 의한 세포 죽음으로 이어졌다는 것을 나타낸다. 이것은 ACGGCCAA [SEQ ID NO: 22] PAM 서열이 gtCas9에 의해 인지된 것을 시사한다.
- [0369] 이들 결과는, 시험관 내 PAM 결정 결과와 일치하는 결과인, gtCas9가 전달된 PAM 서열을 갖는 생체 내 55°C에서 활성인 것을 나타낸다. 게다가, 이것은 플라스미드 유래 상동 재조합 주형과 조합하여 동일한 온도에서 게놈 편집 도구로 사용될 수 있다.

[0370] **실시 예 10: ThermoCas9 동정 및 정제**

[0371] 최근에 우리는 65°C에서 최적의 성장 온도를 갖는 그람 양성, 중간 정도의 호열성 박테리아인, 지오바실러스 씨모데니트리피칸스 T12 균주를 단리하고, 염기서열을 결정했다 (Daas et al. Biotechnol. Biofuels 9, 210 (2016)). 타입 II CRISPR-Cas 시스템이 호열성 박테리아에 존재하지 않는다는 이전의 주장 (Li et al, Nucleic Acids Res. 44, e34-e34 (2016))과는 달리, 시퀀싱 결과는, G. 씨모데니트리피칸스 T12의 게놈에서 타입-II CRISPR-Cas 시스템의 존재를 밝혔다 (도 15의 A). 이 시스템의 Cas9 엔도뉴클레아제 (ThermoCas9)는, SpCas9 (1368 아미노산)와 같은, 다른 Cas9 상동유전자와 비교하여 상대적으로 작은 (1082 아미노산) 것으로 예견된다.

크기 차이는, 다른 작은 Cas9 상동유전자에 대해 입증된 것처럼, 잘린 REC 로브에 대부분 기인한다 (도 19) (Ran et al. Nature 520, 186-191 (2015)). 더군다나, ThermoCas9는, G. 썬모테니트리피칸스 T12의 적어도 최적 온도 부근에서 활성인 것으로 예상된다 (Daas et al. Biotechnol. Biofuels 9, 210 (2016)). 질의 서열로서 ThermoCas9 서열을 사용하여, 우리는 NCBI/비-중복 단백질 서열 (non-redundant protein sequences) 데이터세트에서 BLAST-P 검색을 수행하였고, ThermoCas9가 호열성 박테리아의 고도로 보존된 방어 시스템의 일부라는 생각을 뒷받침하는 (도 15의 B), 대부분 지오바실러스 속으로, 다수의 매우 동일한 Cas9 상동유전자 (단백질 수준에서 87-99%의 동일성, 표 2)를 확인했다. 이들 특징은, 이것이 호열성 미생물에 대한 게놈 편집 및 침묵 도구로서 이용을 위한, 및 향상된 단백질 견고성 (protein robustness)이 요구되는 조건을 위한, 가능한 후보물질일 수 있음을 시사한다.

[0372] 우리는 처음에 전술된 접근법을 사용하여 G. 썬모테니트리피칸스 T12 CRISPR-Cas 시스템의 crRNA 및 tracrRNA 모듈의 인실리코 예측을 수행하였다 (Mougiakos et al. Trends Biotechnol. 34, 575-587 (2016), Ran et al. Nature 520, 186-191 (2015)). 이 예측에 기초하여, 190nt sgRNA 키메라는, 예측된 전체 크기의 crRNA (30nt 길이의 스페이서 다음에 36nt 길이의 반복) 및 tracrRNA (36nt 길이의 안티-반복 다음에, 3개의 예측된 헤어핀 구조를 갖는 88nt 서열)를 연결하여 디자인된다. ThermoCas9는, E.coli에서 이중방식으로 발현되고, 동질성으로 정제된다. ThermoCas9에 sgRNA의 로딩이 단백질을 안정화시킬 것으로 가정하면서, 우리는, 60℃ 및 65℃에서 15분 및 30분 동안 시험관 내 전사된 sgRNA로 로딩된, 정제된 apo-ThermoCas9 및 ThermoCas9를 배양했다. SDS-PAGE 분석은, 정제된 ThermoCas9가 65℃에서 변성되지만, 60℃에서는 변성되지 않는 반면, ThermoCas9-sgRNA 복합체의 변성 온도가 65℃ 이상임을 보여준다 (도 15의 C). ThermoCas9의 입증된 열안정성은, 내열성 CRISPR-Cas9 게놈 편집 도구로서 이의 잠재성을 암시하며, 몇몇 관련 분자 특색을 좀 더 상세히 분석하는데 도움이 되었다.

표 2

[0373] ThermoCas9와 비교한 도 1의 Cas9 단백질 서열의 pBLAST 결과.

종	% 동일성 ^a
<i>Geobacillus</i> 47C-IIb	99
<i>Geobacillus</i> 46C-IIa	89
<i>Geobacillus</i> LC300	89
<i>Geobacillus</i> <i>jurassicus</i>	89
<i>Geobacillus</i> MAS1	88
<i>Geobacillus</i> <i>stearothermophilus</i>	88
<i>Geobacillus</i> <i>stearothermophilus</i> ATCC 12980	88
<i>Geobacillus</i> Sah69	88
<i>Geobacillus</i> <i>stearothermophilus</i>	88
<i>Geobacillus</i> <i>kaustophilus</i>	88
<i>Geobacillus</i> <i>stearothermophilus</i>	88
<i>Geobacillus</i> <i>genomosp.</i> 3	87
<i>Geobacillus</i> <i>genomosp.</i> 3	87
<i>Geobacillus</i> <i>subterraneus</i>	87
<i>Effusibacillus</i> <i>pohliae</i>	86

[0374] 실시 예 11: ThermoCas9 PAM 결정

[0375] ThermoCas9의 특징화를 향한 제1 단계는, DNA 표적의 성공적인 절단을 위한 이의 PAM 선호도의 인실리코 예측이다. 우리는, CRISPRtarget (Biswas et al, RNA Biol. 10, 817-827 (2013))을 사용하여 바이러스 및 플라스미드 서열에서 잠재적인 프로토스페이서를 조사하기 위해 G. 썬모테니트리피칸스 T12 CRISPR 유전자좌의 10개의 스페이서를 사용했다. 단지 2개의 히트가 파지 게놈으로 얻어짐에 따라 (도 20의 A), 이것은 시험관 내 PAM 결정법으로 진행하기로 결정했다. 우리는, 매칭 프로토스페이서로 선형 dsDNA 기질을 표적으로 하는 ThermoCas9-계에 대한 스페이서를 함유하는 예측된 sgRNA 서열을 시험관 내에서 전사하였다. 프로토스페이서는 무작위화된 7-염기 쌍 (bp) 서열에 의해 이의 3'-말단에서 측면에 위치된다. 55℃에서 ThermoCas9-계 절단 분석을 수행한 후, (대조군로서 비-표적화된 라이브러리 샘플과 함께) 라이브러리의 절단된 멤버 (cleaved members)는, 딥-시퀀싱

하고, ThermoCas9 PAM 선호도를 확인하기 위해 비교된다 (도 16의 A). 시퀀싱 결과는, ThermoCas9가 중온성 Cas9 변이체와 유사하게, 3rd 및 4th PAM 근위 뉴클레오타이드 사이에 대부분 위치하는 이중 가닥 DNA 절단을 도입한다는 것을 밝혀냈다. 게다가, 절단된 서열은, ThermoCas9가, 1st, 3rd, 4th 및 6th PAM 위치에서 시토신에 대한 미묘한 선호도 (subtle preference)로, 5'-NNNNCNR-3' PAM을 인식한다는 것을 밝혀냈다 (도 16의 B). 최근의 연구는, 특정 타입 II Cas9 상동유전자의 표적 인식을 위한 8th PAM 위치의 중요성을 밝혀냈다 (Karvelis *et al. Genome Biol.* 16, 253 (2015); Kim *et al. Genome Res.* 24, 1012-9 (2014)). 이러한 목적을 위해, 인실리코 ThermoCas9 PAM 예측으로부터의 결과를 고려하여, 우리는 추가적인 PAM 결정 분석을 수행했다. 이것은 8th PAM 위치에서 아데닌의 존재하에서 최적의 표적화 효율을 나타낸다 (도 16의 C). 흥미롭게도, 히트 수가 제한되어 있음에도 불구하고, 전술된 인실리코 PAM 예측 (도 20의 B)은 또한, 8th PAM 위치에서 아데닌 및 5th PAM 위치에서 시토신의 중요성을 제시한다.

[0376]

6th 및 7th PAM 위치에서 PAM의 모호성을 더욱 명확히하기 위해, 우리는, 매칭 프로토스페이스가 5'-CCCCNNA-3' [SEQ ID NO: 13] PAMs 옆에 위치하는 한 세트의 16개의 다른 표적 DNA 단편을 발생시켰다. (각각 6th 및 7th 뉴클레오타이드의 독특한 조합으로) 이들 단편의 절단 분석은, 다른 구성요소 (ThermoCas9, sgRNA 가이드, dsDNA 표적)들이, 이들을 조합하고 상응하는 분석 온도에서 1시간 동안 배양하기 전에, 다른 온도 (20, 30, 37, 45, 55 및 60°C)에서 10분 동안 개별적으로 사전-가열되어 수행된다. 분석이 37°C 내지 60°C의 온도에서 수행되는 경우, 모든 다른 DNA 기질은, 절단된다 (도 16의 D, 도 21). 그러나, 대부분 소화된 표적 단편은, PAM 서열 (5th 내지 8th PAM 위치) 5'-CNAA-3' 및 5'-CMCA-3'로 이루어진 반면, 최소 소화된 표적은, 5'-CAKA-3' PAM을 함유한다. 30°C에서, 최적의 PAM 서열 (5th 내지 8th PAM 위치) 5'-CNAA-3' 및 5'-CMCA-3'을 갖는 DNA 기질의 절단만이 관찰되었다 (도 16의 D). 마지막으로, 20°C에서, (5th 내지 8th PAM 위치) 5'-CVAA-3' 및 5'-CCCA PAM 서열을 갖는 DNA 기질만이 표적이 되어 (도 21), 이들 서열을 가장 바람직한 PAMs으로 만든다. 이들 연구 결과들은, 이의 저온 한계에서 ThermoCas9가 바람직한 PAM으로 단편을 오직 절단한다는 것을 입증한다. 이러한 특징은, 생체내 편집 과정 동안, 예를 들어, 오프-타겟 효과 (off-target effects)를 피하기 위해 이용될 수 있다.

[0377]

실시 예 12: 열안정성 및 절단 (truncations)

[0378]

예측된 tracrRNA는, 안티-반복 영역 다음에 3개의 헤어핀 구조로 이루어진다 (도 17의 A). sgRNA 키메라를 형성하기 위해 crRNA와 함께 tracrRNA의 사용은, DNA 기질의 성공적인 가이드된 절단을 결과한다. 이중 가이드의 비변성 상태와 가장 유사할 가능성이 가장 큰, 전체-길이 반복-안티-반복 헤어핀의 스페이서 원위 단부의 41-nt 길이의 절손 (도 17의 A)은, DNA 절단 효율에 거의 영향이 없는 것으로 관찰되었다. ThermoCas9의 절단 효율에 대한 예측된 헤어핀 (도 17의 A)의 추가 절단의 효과는, 모든 구성요소 (sgRNA, ThermoCas9, 기질 DNA)들이, 이들을 조합하고 다양한 분석 온도 (37-65°C)에서 1시간 동안 배양하기 전에, 다른 온도 (37-65°C)에서 1, 2, 및 5분 동안 개별적으로 사전-가열되는, 절단 시계-열 (cleavage time-series)을 수행하여 평가된다. tracrRNA 스키펴의 예견된 스템-루프의 수는, DNA 절단에서 결정적인 역할을 하는 것으로 보이고; 3개의 루프가 모두 존재하는 경우, 절단 효율은 모든 시험 온도에서 가장 높은 반면, 효율은 3' 헤어핀의 제거시 감소하였다 (도 17의 B). 게다가, 중간 및 3' 헤어핀 모두 제거시 절단 효율은 급격히 떨어졌다 (도 22). 65°C에서 1분 또는 2분 동안 ThermoCas9의 사전-가열이 검출 가능한 절단을 결과하는 반면, 절단 활성은 5분 배양 후 완전히 없어졌다. 열안정성 분석은, 3' 스템-루프가 없는 sgRNA 변이체가 65°C에서 ThermoCas9 단백질의 감소된 안정성을 결과하여, 전체 길이 tracrRNA가 상승된 온도에서 최적의 ThermoCas9-계 DNA 절단을 위해 필요하다는 것을 나타낸다. 추가적으로, 우리는 또한 스페이서 서열의 길이를 (25nt에서 18nt로) 변화시켰고, 23, 21, 20 및 19nt의 스페이서 길이가 가장 높은 효율로 표적을 절단한다는 것을 확인했다. 18nt의 스페이서가 사용되는 경우, 절단 효율은 현저히 떨어진다.

[0379]

생체 내에서 ThermoCas9:sgRNA RNP 복합체는, 아마도 수분 내에 형성될 것이다. 위의 연구 결과와 함께, 이것은 우리에게 RNP의 활성 및 열안정성을 평가할 동기를 부여했다. 미리-조립된 RNP 복합체는, 사전-가열된 DNA를 첨가하기 전 60, 65 및 70°C에서 5분 및 10분 동안 가열되고, 그 후에 60, 65 및 70°C에서 1시간 동안 배양된다. 놀랍게도, 우리는 ThermoCas9 RNP가 70°C에서 5분 동안 이의 사전-가열에도 불구하고, 70°C까지 활성인 것을 관찰했다 (도 17의 C). 이 연구 결과는, ThermoCas9 안정성이 적절한 sgRNA 가이드 (Ma *et al.*, *Mol. Cell* 60, 398-407 (2015))와의 연관성과 밀접한 상관관계가 있다는 우리의 추정을 확인했다.

[0380] ThermoCas9가 넓은 온도 활성 범위를 갖는 것, 즉, 저온 및 고온 모두에서 기능적인 것으로, 몇몇 적용에서, 장 점일 수 있다. 또한, 몇몇 상황에서, ThermoCas9의 활성이 더 좁은 온도 범위, 예를 들어, 오직 저온 또는 오직 고온에서 활성으로 제한될 수 있는 경우, 유리하다. 결과적으로, ThermoCas9 또는 (sgRNA와 같은) 관련 인자의 구조적 특색을 변형하여, ThermoCas9가 표적화된 절단 또는 결합이 가능하거나 또는 표적화된 절단 또는 결합이 효율적으로 일어나는, 온도 범위를 조작하기 위한 능력은, 핵산 서열 조작을 넘어 발휘되는 더 큰 수준의 제어를 가능하게 할 것이다. 그러므로, 우리는 ThermoCas9 온도 범위를 *Streptococcus pyogenes* Cas9 (SpCas9)의 온도 범위와 비교하는 것을 시작했다. 두 Cas9 동족체는 20 내지 65°C에서 시험관 내 활성 분석에 적용된다. 두 단백질은, sgRNA 및 표적 DNA 분자의 첨가 전에, 상응하는 분석 온도에서 5분 동안 배양된다. 이전의 분석²⁶과 일치하여, 중온성 SpCas9는, 25 내지 44°C에서 오직 활성이었고 (도 17의 D); 이 온도 위에서, SpCas9 활성은 감지할 수 없는 수준으로 급격히 감소했다. 대조적으로, ThermoCas9 절단 활성은, 25 내지 65°C에서 감지될 수 있다 (도 17의 D). 이것은, ThermoCas9를 호열성 및 중온성 유기체 모두에 대한 게놈 편집 도구로 사용할 수 있는 가능성을 나타낸다.

[0381] 이전에 특징화된, 중온성 Cas9 엔도뉴클레아제는, 표적 DNA에서 DSBs의 발생을 촉매하는 2가 양이온을 사용한다 (Jinek *et al. Science* 337, 816-821 (2012); Chen *et al. J. Biol. Chem.* 289, 13284-13294 (2014)). 어떤 양이온이 ThermoCas9에 의한 DNA 절단에 기여하는지 평가하기 위해, 플라스미드 절단 분석은, 다음의 2가 양이온: Mg^{2+} , Ca^{2+} , Mn^{2+} , Co^{2+} , Ni^{2+} , 및 Cu^{2+} 중 하나의 존재하에서 수행되고; 양이온-킬레이트제 (cation-chelating agent) EDTA를 이용한 분석은, 음성 대조군으로 포함된다. 예상된 바와 같이, 표적 dsDNA는, 2가 양이온의 존재하에서 절단되고, EDTA의 존재하에서 온전히 유지된다 (도 23의 A). 특정 타입-IIIC 시스템이 효율적인 단일 가닥 DNA 가위라는 보고 (Ma *et al. Mol. Cell* 60, 398-407 (2015); Zhang *et al. Mol. Cell* 60, 242-255 (2015))에 기초하여, 우리는 ssDNA 기질에 대한 ThermoCas9의 활성을 시험했다. 그러나, 절단은 관찰되지 않았으며, 이는 ThermoCas9가 dsDNA 뉴클레아제임을 나타낸다 (도 23의 B).

[0382] 실시 예 13: 호열성 박테리아 *B. smithii*의 ThermoCas9-계 유전자 결손

[0383] 우리는 호열성 박테리아에 대한 ThermoCas9-계 게놈 편집 도구를 개발하기 시작했다. 여기서, 우리는 55°C에서 배양된 *Bacillus smithii* ET 138을 사용하여 원리 검증 (proof of principle)을 보여준다. 최소한의 유전자 부분 (genetic parts)을 사용하기 위해, 우리는 단일 플라스미드 접근법을 따랐다. 우리는, 비변형 *xyiL* 프로모터 (P_{xyiL})의 제어하에 *thermocas9* 유전자, 관심의 유전자 내에 Cas9-유도된 이중 가닥 DNA 절단을 복구하기 위한 상동 재조합 주형, 및 *Bacillus coagulans* 유래의 구성적 *pta* 프로모터 (P_{pta})의 제어하의 sgRNA 발현 모듈을 함유하는 한 세트의 pNW33n-계 pThermoCas9 플라스미드를 구축했다 (도 18의 A).

[0384] 제1 목표는 *B. smithii* ET 138의 게놈으로부터 전체 길이 *pyrF* 유전자를 결손시키는 것이다. pNW33n-유래 플라스미드 pThermoCas9_bs $\Delta pyrF1$ 및 pThermoCas9_bs $\Delta pyrF2$ 는, *pyrF* 유전자의 다른 부위를 표적으로 하는 스페이서를 갖는 다른 ThermoCas9 가이드의 발현에 사용되는 반면, 제3 플라스미드 (pThermoCas9_ctrl)는, sgRNA 발현 모듈에 무작위 비-표적화 스페이서를 함유한다. 대조군 플라스미드 pNW33n (가이드 없음) 및 pThermoCas9_ctrl로 55°C에서 *B. smithii* ET 138 수용능 세포의 형질전환은, 각 ~200 콜로니의 형성을 결과한다. 10개의 스크리닝된 pThermoCas9_ctrl 콜로니 중에서, 어느 것도 $\Delta pyrF$ 유전자형을 함유하지 않아, 이전의 연구에서 *B. smithii* ET 138의 상동 재조합이 세정 돌연변이체를 얻는데 충분하지 않다는 연구결과를 확인하였다 (Mougiakos *et al. ACS Synth. Biol.* 6, 849-861 (2017); Bosma *et al. Microb. Cell Fact.* 14, 99 (2015)). 대조적으로, pThermoCas9_bs $\Delta pyrF1$ 및 pThermoCas9_bs $\Delta pyrF2$ 플라스미드로의 형질전환은, 각각, 20 및 0 콜로니를 결과하여, 55°C에서 ThermoCas9의 생체 내 활성을 확인하고, 단백질의 전술된 넓은 시험관 내 온도 범위를 입증한다. 스크리닝된 10개의 pThermoCas9_ $\Delta pyrF1$ 콜로니 중에서, 하나는 세정 $\Delta pyrF$ 돌연변이체인 반면, 나머지는 혼합된 야생형/ $\Delta pyrF$ 유전자형을 나타내어 (도 18의 B), 표적화된 *pyrF* 유전자의 디자인된 상동성 인도 복구 (homology directed repair)이 성공적인 것으로, 시스템의 적용가능성을 증명한다. 그럼에도 불구하고, 우리가 이전에 개발한 엄격히 제어된 SpCas9-계 반대-선택 시스템에서, *pyrF* 결손 효율은 더 높았다 (Olson *et al., Curr. Opin. Biotechnol.* 33, 130-141 (2015)). ThermoCas9-계 도구에서 얻어진 형질전환체 및 세정 돌연변이체의 적은 수는, 고도의 활성 ThermoCas9의 구성적 발현과 조합된 *B. smithii* (Olson *et al., Curr. Opin. Biotechnol.* 33, 130-141 (2015))에서 낮은 상동 재조합 효율에 의해 설명될 수 있다. 정확히 제어가능한 프로모터의 사용은, 효율을 증가시킬 것으로 예상된다.

[0385] 실시 예 14: 중온성 슈도모나스 퓨티다에서 ThermoCas9-계 유전자 결손

[0386] ThermoCas9-계 게놈 편집 도구의 적용가능성을 넓히고, 및 시험관 내 결과가 생체 내에서 확인될 수 있는지를 평가하기 위해, 중온성 그람-음성 박테리아 *P. putida* KT2440에서 이의 활성은, 상동 재조합 및 ThermoCas9-계 반대-선택을 조합하여 평가된다. 이 유기체의 경우, Cas9-계 도구는, 현재까지 보고되지 않았다. 한 번 더, 우리는 단일 플라스미드 접근법을 따랐다. 우리는, 3-메틸벤조에이트-유도성 Pm-프로모터의 제어하에 thermocas9 유전자, pyrF 유전자의 결손을 위한 상동 재조합 주형 및 구성적 P3 프로모터의 제어하에 sgRNA 발현 모듈을 함유하는 pEMG-계 pThermoCas9_ppΔpyrF 플라스미드를 구축했다. *P. putida* KT2440 세포의 형질전환 및 플라스미드 인터그레이션의 PCR 확인 후, 콜로니는, 37°C에서 밤새 배양을 위한 선택적 액체 배지에 접종된다. 오버나이트 배양물 (overnight culture)은, 선택 배지의 접종을 위해 사용하고, ThermoCas9 발현은 3-메틸벤조에이트로 유도된다. 뒤이어, 희석액은 3-메틸벤조에이트가 보충된, 비-선택성 배지 상에 도말된다. 비교를 위해, 3-메틸벤조에이트로 ThermoCas9 발현을 유도하지 않은 병행 실험은 수행된다. 이 과정은, 유도 배양 (induced culture)에 대한 76개의 콜로니 및 비-유도 대조군 배양에 대한 52개의 콜로니를 결과한다. 유도 배양의 경우, 38개의 콜로니 (50%)는, 세정 결손 유전자형을 가지며, 6개의 콜로니는 혼합된 야생-형/결손 유전자형을 갖는다. 반대로, 비-유도 배양의 오직 1개의 콜로니 (2%)는, 결손 유전자형을 가지며, 혼합된 야생-형/결손 유전자형을 가진 콜로니는 검색되지 않았다 (도 24). 이들 결과는, ThermoCas9가, 37°C에서 성장된 경우, 중온성 *P. putida* KT2440에서 효과적인 반대-선택 도구로 사용될 수 있음을 나타낸다.

[0387] **실시 예 15: ThermoCas9-계 유전자 침묵**

[0388] 효율적인 열활성 전사적 침묵 CRISPRi 도구는 현재 이용할 수 없다. 이러한 시스템은, 다수의 적용들에서 유용할 수 있다. 예를 들어, 이러한 시스템은, 호열성의 대사 연구를 크게 촉진할 것이다. ThermoCas9의 촉매적으로 죽은 변이체는, dsDNA 절단을 도입하지 않고 DNA 인자에 지속적으로 결합하여 이 목적을 만족시킬 수 있다. 이를 위해, 우리는, ThermoCas9의 RuvC 및 HNH 촉매 도메인을 확인하고, 죽은 (d) ThermoCas9를 생성하기 위해 상응하는 D8A 및 H582A 돌연변이를 도입했다. 디자인된 서열의 확인 후, Thermo-dCas9는, 이중방식으로 생산되며, 정제되고, 전술한 ThermoCas9 분석에서 사용된 것과 동일한 DNA 표적으로 시험관 내 절단 분석을 위해 사용되며; 뉴클레아제의 촉매 비활성화를 확인하는 절단은 관찰되지 않았다.

[0389] Thermo-dCas9-계 CRISPRi 도구의 개발을 위해, 우리는 *B. smithii* ET138의 게놈에서 고도로 발현된 ldhL 유전자의 전사적 침묵을 목표로 했다. 우리는, pNW33n-계 벡터들인, pThermoCas9i_ldhL 및 pThermoCas9i_ctrl를 구축했다. 벡터들 모두는, P_{xyL} 프로모터의 제어하에 thermo-dCas9 유전자 및 구성적 P_{pta} 프로모터의 제어하에 sgRNA 발현 모듈을 함유한다 (도 18의 C). pThermoCas9i_ldhL 플라스미드는 *B. smithii* ET 138에서 138 ldhL 유전자의 5' 말단에 비-주형 DNA 가닥을 표적화하기 위한 스페이서를 함유한다. 위치 및 표적화된 가닥 선택은, 이전의 연구 (Bikard et al. Nucleic Acids Res. 41, 7429-7437 (2013), Larson et al. Nat. Protoc. 8, 2180-2196 (2013))에 기초하여, ldhL 유전자의 효율적인 하향-조절 (down-regulation)을 목표로 한다. pThermoCas9i_ctrl 플라스미드는, sgRNA-발현 모듈에 무작위 비-표적화 스페이서를 함유한다. 구축물은, 55°C에서 *B. smithii* ET 138 수용능 세포를 형질전환시키는데 사용된 다음, LB2 한천 플레이트에 도말하여, 동일한 양의 콜로니를 결과한다. 구축물 당 대략 700개의 콜로니 중에서 2개는, 전술된 바와 같이 (Bosma et al. Appl. Environ. Microbiol. 81, 1874-1883 (2015)), 24시간 동안 미호기성 (microaerobic) 락테이트-생성 조건하에서 배양하기 위해 선택된다. pThermoCas9i_ldhL 배양물의 성장은, pThermoCas9i_ctrl 배양물의 성장보다 50% 적다 (도 18의 D). 우리는, ldhL 유전자의 결손이 미-호기성 조건하에서 Ldh-계 NAD⁺-재생 능력의 결핍으로 인해 *B. smithii* ET 138에서 심각한 성장 지연으로 이어진다는 것을 이전에 보여주었다 (Bosma et al. Microb. Cell Fact 14, 99 (2015)). 따라서, 성장에서 관찰된 감소는, ldhL 유전자의 전사적 억제 및 뒤이은 NAD⁺-재생 능력의 상실로 인한 산화환원 불균형 때문일 수 있다. 실제로, HPLC 분석은, ldhL 침묵 배양물의 락테이트 생산에서 40% 감소를 나타냈고, RT-qPCR 분석은, pThermoCas9i_ctrl 배양물과 비교하여 pThermoCas9i_ldhL 배양물에서 ldhL 유전자의 전사 수준이 상당히 감소되었음을 보여준다 (도 18의 D).

[0390] **실시 예 16: 요약**

[0391] 대부분의 CRISPR-Cas 적용은, Cas9 및 Cas12a와 같은, 부류 2 CRISPR-Cas 단백질에 의한 RNA-가이드된 DNA 간섭에 기초한다 (Komor et al., Cell 168, 20-36 (2017); Puchta, Curr. Opin. Plant Biol. 36, 1-8 (2017); Xu et al. J. Genet. Genomics 42, 141-149 (2015); Tang et al. Nat. Plants 3, 17018 (2017); Zetsche et al. Nat. Biotechnol. 35, 31-34 (2016); Mougiakos et al., Trends Biotechnol. 34, 575-587 (2016)). 이 연구 이전에, 호열성 박테리아 및 고세균에 존재하는 매우 풍부한 부류 1 CRISPR-Cas 시스템과는 달리, 부류 2 CRISPR-Cas 면역 시스템은 호열성 미생물에서 동정 및 특징화되지 않았고 (Makarova et al., Nat. Rev.

Microbiol. 13, 722-736 (2015); Weinberger et al., *MBio* 3, e00456-12 (2012)), 이의 몇몇은, 호열성 미생물의 게놈 편집을 위해 사용되었다 (Li et al. *Nucleic Acids Res.* 44, e34-e34 (2016)). 결과적으로, CRISPR-Cas 기술의 적용은, 사용된 Cas-엔도뉴클레아제의 중온성 성질에 기인하여, 42℃ 이하의 온도로 주로 제한된다. 그러므로, 이는 상승된 온도 및/또는 개선된 단백질 안정성을 요구하는 절대 호열성 미생물에서 및 실험적 접근법에서 이들 기술의 적용을 배제한다.

[0392] 본 발명자들은 이전에 퇴비 (compost)에서 단리된 균주인, 호열성 박테리아 *G. 썬모테니트리피칸스* T12의 Cas9 상동유전자인, ThermoCas9를 특징화하였다 (Daas et al., *Biotechnol. Biofuels* 9, 210 (2016)). 데이터 마이닝 (Data mining)은, ThermoCas9와 거의 동일한 다른 호열성 미생물의 게놈에 부가적인 Cas9 상동유전자를 보여주는데, 이는 CRISPR-Cas 타입-II 시스템이, 호열성 미생물, 적어도 바실러스 (*Bacillus*) 및 지오바실러스 (*Geobacillus*) 속의 몇몇 분과 (branches)에 존재하는 것을 처음으로 보여주는 것이다. 본 발명자들은 ThermoCas9가 20-70℃의 넓은 온도 범위에서 시험관 내에서 활성인 것을 알아냈는데, 이는 이의 중온성 상동유전자 SpCas9의 25-44℃ 범위보다 훨씬 더 넓은 것이다. ThermoCas9의 확장된 활성 및 안정성은, 20-70℃의 온도에서 DNA 조작을 필요로 하는 분자 생물학 기술에서 이의 적용뿐만 아니라 강력한 효소 활성을 필요로 하는 가혹한 환경에서 이의 활용을 가능하게 한다. 더군다나, 본 발명자들은, ThermoCas9의 열안정성을 부여하는데 중요한 몇몇 요인들을 확인하였다. 먼저, 본 발명자들은, ThermoCas9의 PAM 선호도가 온도 범위의 하부 ($\leq 30^{\circ}\text{C}$)에서 활성에 대해 매우 엄격한 반면, PAM에서 더 많은 다양성이 보통 내지 최적 온도 (37-60℃)에서 활성을 가능하다는 것을 입증하였다. 둘째, 본 발명자들은, ThermoCas9 활성 및 열안정성이 적절한 sgRNA 가이드와의 연관성에 강하게 의존한다는 것을 입증하였다. 어떤 특정 이론에 구속되는 것을 원하지는 않지만, 본 발명자들은, 다중-도메인 Cas9 단백질의 이러한 안정화가, 가이드 결합시 SpCas9에 대해 기재된 바와 같이, 개방/유연한 상태에서 다소 콤팩트한 상태 (compact state)로의 주요 형태 변화 (conformational change)의 결과일 가능성이 크다고 추측한다 (Jinek et al. *Science* 343, 1247997-1247997 (2014)).

[0393] 신규 ThermoCas9의 여기서 기재된 특징화에 기초하여, 본 발명자들은 절대 호열성 원핵생물을 위한 게놈 공학 도구를 성공적으로 개발하였다. 우리는, ThermoCas9가 55℃ 및 37℃에서 생체 내 활성인 것을 보여주었으며, 우리는, 호열성 *B. smithii* ET 138 및 중온성 *P. putida* KT2440에 대한 현재 Cas9-계 공학 기술을 채택했다. ThermoCas9의 넓은 온도 범위로 인해, 간단하고, 효과적인, 단일 플라스미드-계 ThermoCas9 접근법은 37℃에서 70℃까지의 온도에서 성장할 수 있는 넓은 범위의 호열성 및 중온성 미생물에 대해 적합할 것으로 예상된다. 이것은, 기존의 중온성 기술을 보완하며, 따라서 이러한 효율적인 도구를 지금까지 이용할 수 없었던 많은 유기체의 군에 대해 이들의 사용을 가능하게 한다.

[0394] 원하는 특성을 지닌 신규 효소에 대한 천연 자원의 스크린은, 의문의 여지없이 가치가 있다. 이전의 연구는, 유도 진화 (directed evolution) 및 단백질 공학과 함께, 더 높은 온도에 중온성 Cas9 상동유전자의 적응이 호열성 Cas9 단백질²⁹의 구축에 대한 최선의 접근법일 것이라고 제안했다. 대신에, 우리는, 몇몇 호열성 박테리아에서 Cas9의 계통 분기 (clade)를 확인했고, 이들 열안정성 ThermoCas9 변이체 중 하나를 호열성 및 중온성 유기체 모두에 대한 강력한 게놈 공학 도구로 전환시켰다. 이 연구를 통해, 우리는, Cas9-계 게놈 편집 기술의 잠재력을 더욱 넓히고, 가혹한 조건에서 또는 넓은 온도 범위에 걸쳐 활성을 필요로 하는 신규한 적용에서 Cas9 기술을 사용하기 위한 새로운 가능성을 열었다.

[0395] 실시 예 17: 물질 및 방법

[0396] a. 박테리아 균주 및 성장 조건

[0397] ThermoCas9를 이용한 유전자 편집 및 침묵 실험을 위해 중간 정도의 호열성 균 *B. smithii* ET 138 $\Delta\text{sigF } \Delta\text{hsdR}$ (Mougiakos, et al., (2017) *ACS Synth.Biol.* 6, 849-861)는 사용된다. 이것은, 55℃에서 LB2 배지 (Bosma, et al. *Microb. Cell Fact.* 14, 99 (2015))에서 성장된다. 플레이트 (plates)의 경우, 배지 1L 당 30g의 한천 (Difco)은 모든 실험에 사용된다. 필요한 경우, 클로람페니콜은 7 $\mu\text{g/mL}$ 의 농도로 첨가된다. 단백질 발현을 위해, *E. coli* Rosetta (DE3)는, 온도가 16℃로 전환된 후, OD_{600nm}가 0.5에 도달할 때까지 120rpm의 셰이커 배양기에서 37℃의 플라스크 내의 LB 배지에서 성장된다. 30분 후, 발현은, 이소프로필-1-티오- β -D-갈락토피라노시드 (IPTG)를 0.5mM의 최종 농도로 첨가하여 유도되고, 그 후 배양은 16℃에서 계속된다. 6th 및 7th, 및 8th 위치에 대한 PAM 구축물을 클로닝하기 위해, 제조업자가 제공한 설명서에 따라, DH5-알파 수용능 *E. coli* (NEB)는 형질전환되고, 37℃에서 LB 한천 플레이트에서 밤새 성장된다. 축퇴 7-nt 길이의 PAM 라이브러리를 클로닝하기 위해, 전기-수용능 DH10B 대장균 세포는, 표준 절차 (Sambrook, Fritsch & Maniatis, T. *Molecular*

cloning: a laboratory manual. (Cold Spring Harbor Laboratory, 1989)에 따라 형질전환되고, 37℃에서 밤새 LB 한천 플레이트 상에서 성장된다. 대장균 DH5α λpir (Invitrogen)는, Ausubel *et al.*에 기재된 형질전환 절차 (*Current Protocols in Molecular Biology*. (John Wiley & Sons, Inc., 2001). doi:10.1002/0471142727)를 사용하여 *P. putida* 플라스미드 구축을 위해 사용된다. 모든 *E. coli* 균주에 대해, 필요하다면, 클로람페니콜은 25 mg/L의 농도로, 카나마이신은 50 mg/L로 사용된다. 슈도모나스 퓨티다 KT2440 (DSM 6125) 균주는, 별도로 명시하지 않는 한, LB 배지에서 37℃로 배양된다. 필요한 경우, 카나마이신은 50 mg/L의 농도로, 3-메틸벤조에이트는 3 mM의 농도로 첨가된다.

[0398] **b. ThermoCas9 발현 및 정제**

[0399] ThermoCas9는 G. 씨모데니트리피칸스 T12의 게놈으로부터 PCR 증폭된 다음, *E. coli* Rosetta (DE3)에서 복제되고, 이종 방식으로 (heterologously) 발현되며, Ni²⁺-친화성, 음이온 교환 및 겔 여과 크로마토그래피 단계들의 조합에 의한 FPLC를 사용하여 정제된다. 유전자는, Tobacco Etch Virus (TEV) 프로테아제 절단 부위 및 헥사히스티딘 서열을 포함하는 N-말단 태그와 융합된 ThermoCas9 폴리펩티드 서열 (잔기 1-1082)을 인코딩하는 단백질 발현 구축물을 발생시키기 위해 올리고뉴클레오티드 (표 3)를 사용하여 결찰-독립 클로닝 (ligation-independent cloning)에 의해 (UC Berkeley MacroLab, Addgene #29653으로부터 얻어진) 플라스미드 pML-1B으로 삽입된다. 촉매적으로 비활성인 ThermoCas9 단백질 (Thermo-dCas9)을 발현시키기 위해, D8A 및 H582A 점 돌연변이는, PCR을 사용하여 삽입되고, DNA 시퀀싱으로 확인된다.

표 3

[0400] 본 연구에 사용된 올리고뉴클레오티드.

	올리고	서열	설명	SEQ ID
--	-----	----	----	--------

PAM 라이브러리 구축	BG6494	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACNNN</u> <u>NNNNCTAGATCCTTTTAAATTAAAAATGAAGTTTTAA</u> ATCAATC	7-nt 길이의 무작위 PAM 서열을 갖는 체외 표적 DNA의 구축을 위한 FW	59
	BG6495	TATGCC <u>GGATCC</u> TCAGACCAAGTTACTCATATATAC TTTAGATTGATTTAAACTTCATTTTAAATTTAAAG GATCTAG	체외 표적 DNA 서열의 구축을 위한 RV	60
	BG7356	TCGTCCGCAGCGTCAGATGTGTATAAGAGACAG-T-	A-tailed ThermoCas9 절단된 단편에 연 결하는, BG7357로 어닐링된 경우 어댑 터	61
	BG7357	CTGTCTCTTATACACATCTGACGCTGCCGACGA	A-tailed ThermoCas9 절단된 단편에 연 결하는, BG7356으로 어닐링된 경우 어 댑터	62
	BG7358	TCGTCCGCAGCGTCAG	ThermoCas9 절단된 단편의 PCR 증폭을 위한 FW 시퀀싱 아답터	63
	BG7359	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAC CATGATTACGCCAAGC	ThermoCas9 절단된 단편의 PCR 증폭을 위한 RV 시퀀싱 아답터	64
	BG7616	TCGTCCGCAGCGTCAGATGTGTATAAGAGACAGGGTC ATGAGATTATCAAAAAGGATCTTC	대조군 단편의 PCR 증폭을 위한 RV 시 퀀싱 어댑터	65
	BG8157	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACCCC</u> <u>CCCAGCTAGATCCTTTTAAATTAAAAATGAAGTTTAA</u> AATCAATC	PAM "CCCCCAG"로 체외 표적 DNA의 구 축을 위한 FW	66
	BG8158	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACCCC</u> <u>CCCAACTAGATCCTTTTAAATTAAAAATGAAGTTTAA</u> AATCAATC	PAM "CCCCCAA"로 체외 표적 DNA의 구 축을 위한 FW	67
	BG8159	<u>TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCC</u> <u>CCCATCTAGATCCTTTTAAATTAAAAATGAAGTTTAA</u> AATCAATC	PAM "CCCCCAT"로 체외 표적 DNA의 구 축을 위한 FW	68
	BG8160	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACCCC</u> <u>CCCACTAGATCCTTTTAAATTAAAAATGAAGTTTAA</u> AATCAATC	PAM "CCCCCAC"로 체외 표적 DNA의 구 축을 위한 FW	69
	BG8161	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACNNN</u> <u>NTNNCTAGATCCTTTTAAATTAAAAATGAAGTTTAA</u> ATCAATC	PAM "NNNNTNN"로 체외 표적 DNA의 구축 을 위한 FW	70
	BG8363	ACGGTTATCCACAGAATCAG	PAM 동정 라이브러리의 PCR 선형화를 위한 FW	71
	BG8364	CGGGATTGACTTTTAAAAAAGG	PAM 동정 라이브러리의 PCR 선형화를 위한 RV	72
	BG8763	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACCCC</u> <u>CCAACTAGATCCTTTTAAATTAAAAATGAAGTTTAA</u> AATCAATC	PAM 위치 6&7 "AA"로 체외 표적 DNA의 구축을 위한 FW	73
	BG8764	<u>TATGCCTCATGAGATTATCAAAAAGGATCTTCACCCC</u> <u>CCATACTAGATCCTTTTAAATTAAAAATGAAGTTTAA</u> AATCAATC	PAM 위치 6&7 "AT"로 체외 표적 DNA의 구축을 위한 FW	74
	BG8765	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACCCC</u> <u>CCAGACTAGATCCTTTTAAATTAAAAATGAAGTTTAA</u> AATCAATC	PAM 위치 6&7 "AG"로 체외 표적 DNA의 구축을 위한 FW	75
	BG8766	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACCCC</u> <u>CCACACTAGATCCTTTTAAATTAAAAATGAAGTTTAA</u> AATCAATC	PAM 위치 6&7 "AC"로 체외 표적 DNA의 구축을 위한 FW	76
	BG8767	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACCCC</u> <u>CCTAACTAGATCCTTTTAAATTAAAAATGAAGTTTAA</u> AATCAATC	PAM 위치 6&7 "TA"로 체외 표적 DNA의 구축을 위한 FW	77
	BG8768	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACCCC</u> <u>CCTTACTAGATCCTTTTAAATTAAAAATGAAGTTTAA</u> AATCAATC	PAM 위치 6&7 "TT"로 체외 표적 DNA의 구축을 위한 FW	78
	BG8769	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACCCC</u> <u>CCTGACTAGATCCTTTTAAATTAAAAATGAAGTTTAA</u> AATCAATC	PAM 위치 6&7 "TG"로 체외 표적 DNA의 구축을 위한 FW	79
	BG8770	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACCCC</u> <u>CCTCACTAGATCCTTTTAAATTAAAAATGAAGTTTAA</u> AATCAATC	PAM 위치 6&7 "TC"로 체외 표적 DNA의 구축을 위한 FW	80
	BG8771	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACCCC</u> <u>CCGAAGTATAGATCCTTTTAAATTAAAAATGAAGTTTAA</u> AATCAATC	PAM 위치 6&7 "GA"로 체외 표적 DNA의 구축을 위한 FW	81

PAM 라이 브러리 구축	BG8772	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACCCC</u> CCGTA CTAGATCCTTTTAAATTA AAAATGAAGTTTTA AATCAATC	PAM 위치 6&7 "GT"로 제외 표적 DNA의 구축을 위한 FW	82
	BG8773	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACCCC</u> CCGGACTAGATCCTTTTAAATTA AAAATGAAGTTTTA AATCAATC	FW for construction of <i>in vitro</i> target DNA with PAM 위치 6&7 "GG"	83
	BG8774	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACCCC</u> CCGCA CTAGATCCTTTTAAATTA AAAATGAAGTTTTA AATCAATC	PAM 위치 6&7 "GC"로 제외 표적 DNA의 구축을 위한 FW	84
	BG8775	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACCCC</u> CCCAACTAGATCCTTTTAAATTA AAAATGAAGTTTTA AATCAATC	PAM 위치 6&7 "CA"로 제외 표적 DNA의 구축을 위한 FW	85
	BG8776	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACCCC</u> CCCTA CTAGATCCTTTTAAATTA AAAATGAAGTTTTA AATCAATC	PAM 위치 6&7 "CT"로 제외 표적 DNA의 구축을 위한 FW	86
	BG8777	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACCCC</u> CCCGACTAGATCCTTTTAAATTA AAAATGAAGTTTTA AATCAATC	PAM 위치 6&7 "CG"로 제외 표적 DNA의 구축을 위한 FW	87
	BG8778	TATGCC <u>TCATGAGATTATCAAAAAGGATCTTCACCCC</u> CCCCA CTAGATCCTTTTAAATTA AAAATGAAGTTTTA AATCAATC	PAM 위치 6&7 "CC"로 제외 표적 DNA의 구축을 위한 FW	88
체외 전사 를 위한 gRNA 모듈	BG6574	AAGCTTGAAATAATACGACTCACTATAGG	1차 PAM 동정 과정을 위한 sgRNA 주형 의 PCR 증폭을 위한 FW (30nt 길이 스 페이서)	89
	BG6576	AAAAAGACCTTGACGTTTTCC	1차 PAM 동정 과정을 위한 sgRNA 주형 의 PCR 증폭을 위한 FW	90
	BG9307	AAGCTTGAAATAATACGACTCACTATAGGTGAGATTA TCAAAAAGGATCTTCACGTC	1차를 제외한 모든 PAM 동정 과정을 위 한 sgRNA 주형의 PCR 증폭을 위한 RV (25nt 길이 스페이서)	91
	BG9309	AAAACGCCTAAGAGTGGGGAATG	1차를 제외한 모든 PAM 동정 과정을 위한 3-헤어핀 길이 sgRNA 주형의 PCR 증폭을 위 한 RV	92
	BG9310	AAAAGGCGATAGCGATCC	1차를 제외한 모든 PAM 동정 과정을 위한 2-헤어핀 길이 sgRNA 주형의 PCR 증폭을 위 한 RV	93
	BG9311	AAAACGGTCAGTCTGCCTATAG	1차를 제외한 모든 PAM 동정 과정을 위한 1-헤어핀 길이 sgRNA 주형의 PCR 증폭을 위 한 RV	94
	BG9308	AAGCTTGAAATAATACGACTCACTATAGGTGAGATTA TCAAAAAGGATCTTCACGTC	pT7 및 25nt 스페이서 sgRNA Fw	95
	BG10118	AAGCTTGAAATAATACGACTCACTATAGGAGATTATC AAAAAGGATCTTCACGTCA	pT7 및 24nt 스페이서 sgRNA Fw	96
	BG10119	AAGCTTGAAATAATACGACTCACTATAGGAAGATTAT CAAAAAGGATCTTCACGTCATAG	pT7 및 23nt 스페이서 sgRNA Fw	97
	BG10120	AAGCTTGAAATAATACGACTCACTATAGGATTATCAA AAAGGATCTTCACGTCATAGT	pT7 및 22nt 스페이서 sgRNA Fw	98
	BG10121	AAGCTTGAAATAATACGACTCACTATAGGAATTATCA AAAAGGATCTTCACGTCATAGTT	pT7 및 21nt 스페이서 sgRNA Fw	99
	BG10122	AAGCTTGAAATAATACGACTCACTATAGGTTATCAAA AAGGATCTTCACGTCATAGTT	pT7 및 20nt 스페이서 sgRNA Fw	100
	BG10123	AAGCTTGAAATAATACGACTCACTATAGGTATCAAAA AGGATCTTCACGTCATAGTTC	pT7 및 19nt 스페이서 sgRNA Fw	101
	BG10124	AAGCTTGAAATAATACGACTCACTATAGGATCAAAAA GGATCTTCACGTCATAGTTC	pT7 및 18nt 스페이서 sgRNA Fw	102
	BG9312	AAAACGCCTAAGAGTGGGGAATGCCGAAGAAAGCGG GCGATAGCGATCC	3 루프 sgRNA OH Rv	103
	BG8191	AAGCTTGGCGTAATCATGGTC	pThermoCas9_ctrl 플라스미드 pThermoCas9_bs ΔyrF1/2의 구축을 위해	& 104
	BG8192	TCATGAGTTCCCATGTTGTG	pThermoCas9_ctrl 플라스미드 pThermoCas9_bs ΔyrF1/2의 구축을 위해	& 105

편집 및 첨목 구축	BG8194	tatggcgaatcacaacatgggaactcatgaGAACATCCTCTTTCTTAG	pThermoCas9_ctrl 플라스미드 & 10 pThermoCas9_bs ΔyrF1/2의 구축을 위해 6
	BG8195	gccgatatacaagaccgattttatacttcatTTAAGTTACCTCCTCGATTG	pThermoCas9_ctrl 플라스미드 & 10 pThermoCas9_bs ΔyrF1/2의 구축을 위해 7
	BG8196	ATGAAGTATAAAATCGGTCTTG	pThermoCas9_ctrl 플라스미드 & 10 pThermoCas9_bs ΔyrF1/2의 구축을 위해 8
	BG8197	TAACGGACGGATAGTTTC	pThermoCas9_ctrl 플라스미드 & 10 pThermoCas9_bs ΔyrF1/2의 구축을 위해 9
	BG8198	gaaagccggggaaactatccgtccgttataAATCAGACAAAATGGCCTGCTTATG	pThermoCas9_ctrl 플라스미드 & 11 pThermoCas9_bs ΔyrF1/2의 구축을 위해 0
	BG8263	gaactatgacactttattttcagaatggacGTATAACGGTATCCATTTTAAGAATAATCC	pThermoCas9_ctrl 플라스미드의 구축을 위 11 해 1
	BG8268	accgttatactccattctgaaaataaagtGTCATAGTTCCCCTGAGAT	pThermoCas9_ctrl 플라스미드의 구축을 위 11 해 2
	BG8210	aacagctatgaccatgattacgccaagcttCCCTCCCATGCACAATAG	pThermoCas9_ctrl 플라스미드 & 11 pThermoCas9_bs ΔyrF1/2의 구축을 위해 3
	BG8261	gaactatgacatcatggagttttaaatccaGTATAACGGTATCCATTTTAAGAATAATCC	pThermoCas9_bs ΔyrF1의 구축을 위해 11 4
	BG8266	accgttatactggatttaaaactccatgatGTCATAGTTCCCCTGAGAT	pThermoCas9_bs ΔyrF2의 구축을 위해 11 5
	BG8317	gaactatgaccaccagcttacatcaacaaGTATAACGGTATCCATTTTAAGAATAATCC	pThermoCas9_ΔspyrF2의 구축을 위해 11 6
	BG8320	accgttatacttggtgatgaagctgggtgGTCATAGTTCCCCTGAGAT	pThermoCas9_bs ΔyrF2의 구축을 위해 11 7
	BG9075	CTATCGGCATTACGTCTATC	pThermoCas9i_ctrl의 구축을 위해 11 8
	BG9076	GCGTCGACTTCTGTATAGC	pThermoCas9i_ctrl의 구축을 위해 11 9
	BG9091	TGAAGTATAAAATCGGTCTTGCTATCGGCATTACGTCATATC	pThermoCas9i_ctrl의 구축을 위해 12 0
	BG9092	CAAGCTTCGGCTGTATGGAATCACAGCGTCGACTTCTGTATAGC	pThermoCas9i_ctrl의 구축을 위해 12 1
	BG9077	GCTGTGATTCCATACAG	pThermoCas9i_ctrl의 구축을 위해 12 2
	BG9267	GGTGCAGTAGGTTGCAGCTATGCTTGATAACGGTATCCAT	pThermoCas9i_ctrl의 구축을 위해 12 3
	BG9263	AAGCATAGCTGCAACCTACTGCACCGTCATAGTTCCTGAGATTATCG	pThermoCas9i_ctrl의 구축을 위해 12 4
	BG9088	TCATGACCAAAATCCCTTAACG	pThermoCas9i_ctrl의 구축을 위해 12 5
	BG9089	TTAAGGGATTTTGGTCATGAGAATCCTCTTTCTTAG	pThermoCas9i_ctrl의 구축을 위해 12 6
	BG9090	GCAAGACCGATTTTATACTTCATTTAAG	pThermoCas9i_ctrl의 구축을 위해 12 7
	BG9548	GGATCCCATGACGCTAGTATCCAGCTGGGTCATAGTTCCCCTGAGATTATCG	pThermoCas9i_ldhL의 구축을 위해 12 8
	BG9601	TTCAATATTTTTTTTGAATAAAAAATACGATACAATAAAAATGTCTAGAAAAAGATAAAAAATG	pThermoCas9i_ldhL의 구축을 위해 12 9
	BG9600	TTTTTTATTCAAAAAAATATTGAATTTTAAAAATGATGGTGTAGTATGAAG	pThermoCas9i_ldhL의 구축을 위해 13 0
	BG9549	CCAGCTGGATACTAGCGTCATGGGATCCGTATAACGGTATCCATTTTAAGAATAATCC	pThermoCas9i_ldhL의 구축을 위해 13 1
	BG8552	TCGGGGGTTCGTTTCCCTTG	게놈 pyrF 결손 KO 검사를 검사하기 위한 13 FW 2
	BG8553	CTTACACAGCCAGTGACGGAAC	게놈 pyrF 결손 KO 검사를 검사하기 위한 13 RV 3
	BG2365	GCCGGCGTCCCGAAAAACGA	pThermoCas9_pp ΔpyrF의 구축을 위해 13 4
	BG2366	GCAGGTCGGGTTCCTCGCATCCATGCCCCGAAC	pThermoCas9_pp ΔpyrF의 구축을 위해 13 5
	BG2367	ggcttcggaatcggtttccgggacgccgcACGGCATGGCAAGCCAAG	pThermoCas9_pp ΔpyrF의 구축을 위해 13 6
	BG2368	gacacagcatcggtGCAGGTCTCTTGGAAGTC	pThermoCas9_pp ΔpyrF의 구축을 위해 13 7

	BG2369	gccaagagaccctgCACCGATGCCTGTGTCGAACC	pThermoCas9_pp Δ pyrF의 구축을 위해	138
	BG2370	cttggcggaacgtcaaggtctttttacACGCGCATCAACTTCAAGGC	pThermoCas9_pp Δ pyrF의 구축을 위해	139
	BG2371	atgac gagctgttcaccagcagcgc TATTATTGAAGCATTTATCAGGG	pThermoCas9_pp Δ pyrF의 구축을 위해	140
	BG2372	GTAAAAAAGACCTTGACGTTTTTC	pThermoCas9_pp Δ pyrF의 구축을 위해	141
	BG2373	tatgaagcgggcatTTGAAGACGAAAGGCCTC	pThermoCas9_pp Δ pyrF의 구축을 위해	142
	BG2374	taata gcgctgctggtgaacagctc GTTCATAGTTCCCTGAGATTATCG	pThermoCas9_pp Δ pyrF의 구축을 위해	143
	BG2375	tggagtcataaacatATGAAGTATAAAATCGGTCTTG	pThermoCas9_pp Δ pyrF의 구축을 위해	144
	BG2376	ccctttcgtcttcAAATGGCCCGCTTCATAAGCAG	pThermoCas9_pp Δ pyrF의 구축을 위해	145
	BG2377	gattttatacTTCATATGTTTCATGACTCCATTATTATG	pThermoCas9_pp Δ pyrF의 구축을 위해	146
	BG2378	gggggcatggatgCGAGGAACCCGACCTGCATTGG	pThermoCas9_pp Δ pyrF의 구축을 위해	147
	BG2381	ACACGGCGGATGCACTTACC	P. putida에서 플라스미드 인터그레이션 및 pyrF 결손을 확인하기 위한 FW	148
	BG2382	TGGACGTGTACTTCGACAAC	P. putida에서 pyrF 결손을 확인하기 위한 RV	149
	BG2135	ACACGGCGGATGCACTTACC	P. putida에서 플라스미드 인터그레이션 확인을 위한 RV	150
시퀀싱 프레임 라이머	BG8196	TGGACGTGTACTTCGACAAC	<i>thermocas9</i> seq. 1	151
	BG8197	TAACGGACGGATAGTTTC	<i>thermocas9</i> seq. 2	152
	BG6850	GCCTCATGAATGCAGCGATGGTCCGGTGTTTC	<i>pyrF</i> US	153
	BG6849	GCCTCATGAGTTCCTATGTTGTGATTC	<i>pyrF</i> DS	154
	BG6769	CAATCCAACCTGGCTTGAC	<i>thermocas9</i> seq. 3	155
	BG6841	CAAGAACTTTATTGGTATAG	<i>thermocas9</i> seq. 4	156
	BG6840	TTGCAGAAATGGTTGTCAAG	<i>thermocas9</i> seq. 5	157
	BG9215	GAGATAATGCCGACTGTAC	pNW33n 백본 seq. 1	158
	BG9216	AGGGCTCGCCTTTGGGAAG	pNW33n 백본 seq. 2	159
	BG9505	GTTGCCAACGTTCTGAG	<i>thermocas9</i> seq. 6	160
	BG9506	AATCCACGCCGTTTAG	<i>thermocas9</i> seq. 7	161

절단 분석	BG8363	ACGGTTATCCACAGAATCAG	DNA 표적의 PCR 선형화를 위한 FW	162
	BG8364	CGGGATTGACTTTTAAAAAAGG	DNA 표적의 PCR 선형화를 위한 RV	163
	BG9302	AAACTTCATTTTAAATTTAAAGGATCTAGAACCCCC CGTGAAGATCCTTTTGGATAATCTCATGACCAAAATC CCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACC CCGTAGAAA	ssDNA 절단 분석용 비-주형 가닥 올리고뉴클레오티드	164
	BG9303	TTTCTACGGGGTCTGACGCTCAGTGAACGAAAACTC ACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGG ATCTTCACCCCCCAACTAGATCCTTTTAAATTAAAA ATGAAGTTT	ssDNA 절단 분석을 위한 주형 가닥 올리고뉴클레오티드	165
	BG9304	TTTCTACGGGGTCTGACGCTCAGTGAACGAAAACTC ACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGG ATCTTCACGGGGGTCTAGATCCTTTTAAATTAAAA ATGAAGTTT	ssDNA 절단 분석을 위한 주형 가닥 올리고뉴클레오티드	166
ThermoCas9 발현 및 RT-qPCR	BG7886	TACTTCCAATCCAATGCAAAGTATAAAATCGGTCTTG ATATCG	FW LIC_thermocas9	167
	BG7887	TTATCCCACTTCCAATGTTATTATAACGGACGGATAGT TTCCCCGGCTTTC	RV LIC_thermocas9	168
ThermoCas9 발현	BG9665	ATGACGAAAGGAGTTTCTTATTATG	RV qPCR 체크 <i>ldh1</i>	169
	BG9666	AACGGTATTCCGTGATTAAG	FW qPCR 체크 <i>ldh1</i>	170

[0401] 제한 부위는 이텔리체로 표시된다. PAMs는 밑줄로 표시된다. 스페이서 영역은 굵게 표시된다. 소문자로 된 뉴클레오티드는, HiFi DNA Assembly에 대한 프라이머 오버행 (primer overhangs)에 상응한다. LIC: 리가아제 독립 클로닝 (Ligase Independent cloning); FW: 포워드 프라이머; RV: 역 프라이머.

[0402] 단백질은 E. coli Rosetta 2 (DE3) 균주에서 발현된다. 배양물은 0.5-0.6의 OD_{600nm}까지 성장된다. 발현은, IPTG를 0.5mM의 최종 농도로 첨가하여 유도되고, 배양은 16°C에서 밤새 계속된다. 세포는 원심분리에 의해 수확되고, 세포 펠릿은, 프로테아제 억제제 (Roche cOmplete, EDTA-free) 및 리소자임이 보충된, 20 mL의 용해 버퍼 (Lysis Buffer) (50mM 인산 나트륨 pH 8, 500mM NaCl, 1mM DTT, 10mM 이미다졸)에서 재현탁된다. 균질화하자마자, 세포는 초음파 MS72 마이크로팁 프로브 (Bandelin)를 사용하고, 5-8분 동안 30% 진폭에서 2초 펄스 및 2.5초 정지로 이루어진, 초음파처리 (Sonopuls, Bandelin)에 의해 용해된 다음, 불용성 물질을 제거하기 위해 1시간 동안 4°C에서 16,000×g로 원심분리된다. 정화된 용해물은, 0.22 미크론 필터 (Mdi 멤브레인 기술)로 여과되고, 니켈 컬럼 (Hisrap HP, GE Lifesciences)에 적용되며, 세척한 후, 250 mM 이미다졸로 용출된다. ThermoCas9를 함유하는 분획은, 모아 투석 버퍼 (250mM KCl, 20mM HEPES/ KOH 및 1mM DTT, pH 7.5)로 밤샘 투석된다. 투석 후, 샘플은, 10 mM HEPES/KOH pH 8에서 1:1로 희석되고, IEX-A 버퍼 (150 mM KCl, 20 mM HEPES/KOH pH 8)로 사전-평형화된 헤파린 FF 컬럼에 로딩된다. 컬럼은, IEX-A로 세척되고, 그 다음 IEX-C (2M KCl, 20mM HEPES/KOH pH 8)의 구배 (gradient)로 용출된다. 샘플은, FPLC (AKTA Pure)를 통해 겔 여과 컬럼 (HiLoad 16/600 Superdex 200)에 로딩하기 전에 700μl로 농축된다. 겔 여과로부터의 분획은 SDS-PAGE에 의해 분석된다; ThermoCas9를 함유하는 분획은 모아, 200μl (50mM 인산나트륨 pH 8, 2mM DTT, 5% 글리세롤, 500mM NaCl)로 농축되고, 생화학적 분석을 위해 직접 사용하거나 보관을 위해 -80°C에서 동결된다.

[0403] c. sgRNA의 체외 합성

[0404] sgRNA 모듈은, 예측된 crRNA 및 tracrRNA 서열을 5'-GAAA-3' 링커와 융합시켜 디자인된다. sgRNA-발현 DNA 서열은, T7 프로모터의 전사적 조절하에 놓인다. 이것은 합성되고 (Baseclear, Leiden, The Netherlands), pUC57 백본에 제공된다. 생화학 반응에 사용된 모든 sgRNA는, HiScribe™ T7 고수율 RNA 합성 키트 (NEB)를 사용하여 합성된다. 5' 말단 상에 T7 서열을 갖는, sgRNA를 코딩하는 PCR 단편은, 체외 전사 반응에서 주형으로 활용된다. T7 전사는, 4시간 동안 수행된다. sgRNA는 실행되고, 우레아-PAGE 겔에서 절제되고, 에탄올 침전을 사용하여 정제된다.

[0405] d. 체외 절단 분석

[0406] 체외 절단 분석은, 정제된 재조합 ThermoCas9로 수행된다. ThermoCas9 단백질, 체외 전사된 sgRNA 및 DNA 기질 (표 3에 기재된 프라이머를 사용하여 PCR 증폭을 사용하여 발생됨)은, 명시된 온도에서 10분 동안 (별도의 언급

이 없는 한) 개별적으로 배양된 후, 성분들을 함께 조합하고, 1시간 동안 절단 버퍼 (100mM 인산 나트륨 버퍼 (pH=7), 500mM NaCl, 25mM MgCl₂, 25 (V/V%) 글리세롤, 5mM 디티오프레이트 (DTT))에서 다양한 분석 온도로 이 들은 배양한다. 각 절단 반응은, 160 nM의 ThermoCas9 단백질, 4 nM의 기질 DNA, 및 150 nM의 합성된 sgRNA를 함유한다. 반응은, 6x 로딩 염료 (NEB)를 첨가하고, 1.5% 아가로스 겔 상에서 수행하여 중단된다. 겔은 SYBR safe DNA stain (Life Technologies)으로 염색되고, Gel Doc™ EZ 겔 이미징 시스템 (Bio-rad)으로 이미지화 된다.

[0407] **e. 체외 PAM 스크린을 위한 라이브러리 구축**

[0408] PAM 라이브러리의 구축을 위해, 프로토스페이서 및 이의 3' 말단에서 7-bp 길이 축퇴 서열을 함유하는, 122-bp 길이의 DNA 단편은, 프라이머 어닐링 및 Klenow 단편 (exo-) (NEB) 기반 확장에 의해 구축된다. PAM-라이브러리 단편 및 pNW33n 벡터는, BspHI 및 BamHI (NEB)에 의해 소화된 다음, 절찰된다 (T4 리가제, NEB). 절찰 혼합물 은 전기-수용능 대장균 DH10B 세포로 형질전환되고, 플라스미드는 액체 배양물로부터 단리된다. 7nt-길이의 PAM 결정 과정의 경우, 플라스미드 라이브러리는, SapI (NEB)에 의해 선형화되고, 표적으로 사용된다. 나머지 분석 의 경우, DNA 기질은 PCR 증폭에 의해 선형화된다.

[0409] **f. PAM 스크리닝 분석**

[0410] ThermoCas9의 PAM 스크리닝은, (반응 당): 160 nM의 ThermoCas9, 150 nM의 체외 전사된 sgRNA, 4 nM의 DNA 표 적, 4μl의 절단 버퍼 (100 mM 인산 나트륨 버퍼 pH 7.5, 500 mM NaCl, 5 mM DTT, 25% 글리세롤) 및 최대 20μl 의 최종 반응 부피의 MQ 수로 이루어진, 체외 절단 분석을 사용하여 수행된다. 55°C 반응으로부터 절단 단편을 함유하는 PAM은 겔 정제되고, Illumina 시퀀싱 어댑터로 절찰되며, Illumina HiSeq 2500 시퀀싱 (Baseclear)을 요청한다. 등물 양의 비-thermoCas9로 처리된 PAM 라이브러리는, 동일한 과정에 적용되고, 참조로서 Illumina HiSeq 2500 시퀀싱을 요청한다. 참조 서열에 완벽한 서열 매칭으로 HiSeq 판독 (HiSeq reads)은 추가 분석을 위 해 선택된다. 선택된 판독으로부터, ThermoCas9 처리된 라이브러리에서 1000회 이상 존재하고, 대조군 라이브러 리와 비교하여 ThermoCas9 처리된 라이브러리에서 적어도 10배 이상 존재하는 것은, WebLogo 분석 (Crooks *et al.*, *Genome Res.* 14, 1188-1190 (2004))을 위해 사용된다.

[0411] **g. B. smithii 및 P. putida에 대한 편집 및 침묵 구축물**

[0412] 플라스미드 구축을 위해 사용된 모든 프라이머 및 플라스미드는, NEBuilder HiFi DNA 어셈블리 (NEB)를 수행하 기 위한 적절한 오버행으로 디자인되며, 이들은 표 3 및 표 4에 각각 열거된다. 플라스미드 조립하기 위한 단편 은, Q5 Polymerase (NEB) 또는 Phusion Flash High-Fidelity PCR Master Mix (ThermoFisher Scientific)로 PCR을 통해 얻어지고, PCR 생성물은, 1% 아가로스 겔 전기영동에 적용되며, 그들은 Zymogen gel DNA 회복 키트 (Zymo Research)를 사용하여 정제된다. 조립화된 플라스미드는, P. putida 구축물의 경우에서 화학적 수용능 E. coli DH5α 세포 (NEB), 또는 E. coli DH5α λpir (Invitrogen)로 형질전환되고, 후자는 직접 벡터 인터그레 이션을 용이하게 한다. 단일 콜로니는 LB 배지에 접종되고, 플라스미드 물질은, GeneJet 플라스미드 미니프랩 키트 (ThermoFisher Scientific) 및 입증된 서열 (GATC-바이오테크) 및 이전에 기재된 프로토콜 (Bosma, *et al. Microb. Cell Fact.* 14, 99 (2015))에 따라 제조된, B.smithii ET 138 전기-수용능 세포의 형질전환된 1μg 의 각 구축물을 사용하여 단리된다. Masterpure™ 그램 양성 DNA 정제 키트 (Epicentre)는, B. smithii 및 P. putida 액체 배양물로부터 게놈 DNA를 단리를 위해 사용된다.

[0413] pThermoCas9_ctrl, pThermoCas9_bsΔpyrF1 및 pThermoCas9_bsΔpyrF2 벡터의 구축을 위해, ΔpyrF 상동 재조합 플랭크와 함께 pNW33n 백본은, pWUR_Cas9sp1_hr 벡터 (Mougiakos, *et al. ACS Synth. Biol.* 6, 849-861 (2017)) (BG8191 및 BG8192)로부터 PCR 증폭된다. 비변성 P_{xyIA} 프로모터는, B. smithii ET 138의 게놈 (BG8194 및 BG8195)으로부터 PCR 증폭된다. thermocas9 유전자는, G. 써모데니트리피칸스 T12의 게놈 (BG8196 및 BG8197)으로부터 PCR 증폭된다. P_{pta} 프로모터는, pWUR_Cas9sp1_hr 벡터 (Mougiakos, *et al.*, *ACS Synth. Biol.* 6, 849-861 (2017)) (BG8198 및 BG8261_2/BG8263_nc2/BG8317_3)로부터 PCR 증폭된다. sgRNA 스캐폴드가 수반되는 스페이서는, pUC57_T7t12sgRNA 벡터 (BG8266_2/BG8268_nc2/8320_3 및 BG8210)로부터 PCR 증폭된다.

[0414] 4-단편 어셈블리는, pThermoCas9i_1dhL 벡터의 구축을 위해 디자인되고 실행된다. 처음에, 표적화된 점 돌연변 이는, 주형으로 pThermoCas9_ctrl을 사용하는 2-단계 PCR 접근법을 통해, thermocas9 촉매 잔기의 코돈에 도입 된다 (돌연변이 D8A 및 H582A). 제1 PCR 단계 (BG9075, BG9076) 동안, 원하는 돌연변이는 생성된 PCR 단편의 말단에 도입되고, 제2 단계 (BG9091, BG9092) 동안, 생성된 단편은, 적절한 어셈블리-오버행 (assembly-overhangs)의 도입을 위한 PCR 주형으로 사용된다. 1dhL 침묵 스페이서와 함께 제2 돌연변이 다운스트림에

thermocas9의 부분은, pThermoCas9_ctrl을 주형 (BG9077 및 BG9267)으로 사용하여 PCR 증폭된다. pNW33n 백본과 함께 sgRNA 스캐폴드는, pThermoCas9_ctrl을 주형 (BG9263 및 BG9088)으로 사용하여 PCR 증폭된다. 제1 돌연변이 업스트림에 thermocas9의 부분을 포함하는 프로모터는, pThermoCas9_ctrl을 주형 (BG9089, BG9090)으로 사용하여 PCR 증폭된다.

[0415] 2-단편 어셈블리는, pThermoCas9i_ctrl 벡터의 구축을 위해 디자인되고, 실행된다. pThermoCas9i_ldhL 벡터에서 스페이서 서열은, 양 말단에서 BaeI 제한 부위를 함유하는 무작위 서열로 대체된다. pNW33n 백본과 함께 sgRNA 스캐폴드는, pThermoCas9_ctrl을 주형 (BG9548, BG9601)으로 사용하여 PCR 증폭된다. 구축물의 다른 절반은, Thermo-dCas9로 이루어지고, 프로모터는, pThermoCas9i_ldhL을 주형 (BG9600, BG9549)으로 사용하여 증폭된다.

[0416] 5-단편 어셈블리는 P. putida KT2440 벡터 pThermoCas9_ppΔpyrF의 구축을 위해 디자인되고 실행된다. 자살 벡터 (suicide vector) pEMG 유래의 레플리콘 (replicon)은 PCR 증폭된다 (BG2365, BG2366). pyrF의 플랭킹 영역은, KT2440 게놈 DNA (576-bp 업스트림 플랭크에 대해 BG2367, BG2368, 및 540-bp 다운스트림 플랭크에 대해 BG2369, BG2370)로부터 증폭된다. 플랭크는 프라이머들인 BG2368 및 BG2369의 오버랩 (overlaps)을 이용하여 프라이머인 BG2367 및 BG2370을 사용하는 오버랩 확장 PCR (overlap extension PCR)에서 융합된다. sgRNA는 pThermoCas9_ctrl 플라스미드 (BG2371, BG2372)로부터 증폭된다. 구성적 P3 프로모터는, pSW_I-SceI (BG2373, BG2374)로부터 증폭된다. 이 프로모터 단편은, 프라이머인 BG2371 및 BG2374의 오버랩을 이용하여 프라이머인 BG2372 및 BG2373을 사용하는 오버랩 확장 PCR에서 sgRNA 단편에 융합된다. ThermoCas9는, pThermoCas9_ctrl 플라스미드 (BG2375, BG2376)로부터 증폭된다. ThermoCas9의 3-메틸벤조이트 유도에 사용될, 유도성 Pm-XylS 시스템은, pSW_I-SceI (BG2377, BG2378)로부터 증폭된다.

표 4

[0417] 이 연구에서 사용된 플라스미드

플라스미드	설명	사용된 제한 부위	프라이머	공급원
pNW33n	<i>E. coli-Bacillus</i> 서틀 벡터, 클로닝 벡터, Cam ^R	-	-	BGSC
pUC57_T7sgRNA full	T7 프로모터의 제어하에 sgRNA를 인코딩하는 DNA를 함유하는 pUC57 벡터; 전체 길이 반복/안티반복 sgRNAs의 체외 전사를 위한 주형으로 역할			Baseclear
pMA2_T7sgRNAtruncated R/AR	T7 프로모터의 제어하에 sgRNA의 절단된 반복/안티반복 부분을 인코딩하는 DNA를 함유하는 벡터; 절단된 반복/안티반복 sgRNA의 체외 전사를 위한 주형으로 역할	-	-	Gen9
pRARE	T7 RNA 중합효소 기반 발현 벡터, Kan ^R	-	-	EMD Millipore
pML-1B	희귀한 tRNAs를 인코딩하는, <i>E. coli</i> Rosetta TM (DE3) 플라스미드, Cam ^R	-	-	Macrolab, Addgene
pEMG	Kan ^R 및 레플리콘용 주형으로 사용된, <i>P. putida</i> 자살 벡터		표 3 참조	1
pSW_I-SceI	xyIS 및 P _{Pm} 용 주형으로 사용된, I-SceI를 함유하는 <i>P. putida</i> 벡터		표 3 참조	1
pWUR_Cas9sp1_hr	pyrF 유전자를 표적으로 하는 spCas9-모듈 함유 스페이서를 갖는 pNW33n. 이 플라스미드는 ThermoCas9 기반 구축물을 구축하기 위한 주형으로 사용됨	-	-	²
pThermo_Cas9	N-말단을 갖는 <i>thermocas9</i> . pML-1에서 BHis-tag 및 TEV 절단 부위. ThermoCas9용 발현 벡터	SspI 및 Ligase Independent Cloning	BG7886 및 BG7887	본 연구
pThermo_dCas9	N-말단을 갖는 <i>cas9dthermocas9</i> . pML-1에서 His-tag 및 TEV 절단 부위. 촉매적으로 비활성인 (죽은) dThermoCas9용 발현 벡터	SspI 및 Ligase Independent Cloning	BG7886 및 BG7888	본 연구
pNW-PAM7nt	체외 PAM 결정 분석을 위한 7-nt 변성 PAM을 함유하는 pNW33n 벡터에서 표적 서열	BamHI 및 BspHI	표 3 참조	본 연구
pNW63-pNW78	PAM (CCCCNNA)의 6th 및 7th 위치에 별개의 뉴클레오타이드를 함유하는 pNW33n 벡터에서 표적 서열	BamHI 및 BspHI	표 3 참조	본 연구

pThermoCas9_ctrl	비-표적화 스페이서를 함유하는 ThermoCas9-모듈 ¹ 을 갖는 pNW33n. 음성 대조군으로 사용됨	-	표 3 참조	본 연구
pThermoCas9_bsΔyrF1	pyrF 유전자 및 융합된 us+ds pyrF-flanks를 표적화하는 스페이서 1을 함유하는 ThermoCas9-모듈 ¹ 을 갖는 pNW33n	-	표 3 참조	본 연구
pThermoCas9_bsΔyrF2	pyrF 유전자 및 융합된 us+ds pyrF-flanks를 표적화하는 스페이서 2를 함유하는 ThermoCas9-모듈 ¹ 을 갖는 pNW33n	-	표 3 참조	본 연구
pThermoCas9i_ctrl	비-표적화 스페이서를 함유하는 Thermo-dCas9-모듈 ² 을 갖는 pNW33n. 야생형 대조군으로 사용됨	-	표 3 참조	본 연구
pThermoCas9i_ldhL	ldhL 유전자를 표적화하는 스페이서 2를 함유하는 Thermo-dCas9-모듈 ² 을 갖는 pNW33n	-	표 3 참조	본 연구
pThermoCas9_ppΔpyrF	pyrF 유전자 및 융합된 us+ds pyrF-flanks를 표적화하는 스페이서를 함유하는 슈도모나스 퓨티다용 ThermoCas9-module ³ 을 갖는 pEMG	-	표 3 참조	본 연구

[0418] ¹ ThermoCas9 모듈은, 비변성 P_{xytL} 프로모터 하에 thermocas9 다음에, B. coagulans P_{pta} 프로모터 하에 sgRNA를 함유한다 (도 4).

[0419] ² ThermoCas9 모듈과 유사하지만, thermocas9 대신 thermo-dCas9를 이용한다 (도 4).

[0420] ³ 슈도모나스 퓨티다용 ThermoCas9 모듈은, 유도성 Pm-XylS 시스템의 전사 제어 하에 thermocas9 다음에, 구성적 P3 프로모터 하에 sgRNA를 함유한다.

[0421] h. P. putida를 위한 편집 프로토콜

[0422] P. putida로의 플라스미드의 형질전환은 Choi et al. (Choi et al., *J. Microbiol. Methods* 64, 391-397 (2006))에 따라 수행된다. 성분의 선택 및 형질전환 후에, 오버나이트 배양물은 접종된다. 10μl의 오버나이트 배양물은 3ml의 새로운 선별 배지에 접종을 위해 사용되고, 37℃에서 2시간 동안 성장시킨 후, ThermoCas9는 3-메틸벤조에이트로 유도된다. 추가적인 6시간 후, 배양물의 희석액은, 3-메틸벤조에이트가 보충된 비-선택 배지 상에 도말된다. 대조군 배양물의 경우, 모든 단계에서 3-메틸벤조에이트의 첨가는 생략된다. P. putida 염색체에서 플라스미드 인터그레이션의 확인은, 프라이머 BG2381 및 BG2135로 콜로니 PCR에 의해 수행된다. pyrF 결손의 확인은, 프라이머 BG2381 및 BG2382로 콜로니 PCR에 의해 수행된다.

[0423] i. RNA 단리

[0424] RNA 단리는, 이전에 기재된 프로토콜 (van Hijum et al. *BMC Genomics* 6, 77 (2005))에 기초한 페놀 추출에 의해 수행된다. 오버나이트 10 mL의 배양물은, 4℃ 및 4816×g에서 15분 동안 원심분리되고, 즉시 RNA 단리를 위해 사용된다. 배지를 제거한 후, 세포는, 0.5 mL의 얼음같이 찬 TE 버퍼 (pH 8.0)에 현탁되고, 얼음에서 유지된다. 모든 샘플은, 0.5g의 지르코늄 비드, 30μl의 10% SDS, 30μl의 3M 아세트산 나트륨 (pH 5.2), 및 500μl의 로티-페놀 (Roti-Phenol) (pH 4.5-5.0, Carl Roth GmbH)을 함유하는 두 개의 2mL 스크류-캡 튜브로 나누어진다. 세포는, FastPrep-24 장치 (MP Biomedicals)를 사용하여 5500rpm에서 45초 동안 파괴되고, 4℃에서 10000rpm으로 5분간 원심분리된다. 각 튜브로부터 400μl의 수상 (water phase)은, 새로운 튜브로 옮기고, 400μl의 클로로포름-이소아밀 알코올 (Carl Roth GmbH)을 첨가한 후, 샘플을 4℃ 및 18,400×g로 3분동안 원심분리하였다. 300μl의 수상 상 (aqueous phase)은, 새로운 튜브로 옮기고, 고순도 RNA 단리 키트 (Roche)로부터의 300μl의 용해 버퍼와 혼합된다. 뒤이어, 이 키트의 나머지의 절차는, 45분 동안 수행된 DNase 배양 단계를 제외하고, 제조사의 프로토콜에 따라 수행하였다. cDNA의 농도 및 무결성은 Nanodrop-1000 Integrity를 사용하여 결정되고, 단리된 RNA의 농도는 NanoDrop 1000에서 검사된다.

[0425] j. RT-qPCR에 의한 mRNA의 정량

[0426] 1차-가닥 cDNA 합성은, 제조사의 프로토콜에 따라 SuperScript™ III Reverse Transcriptase (Invitrogen)를 사용하여 수행된다. qPCR은 Quanta Biosciences의 iQ 용 PerfeCTa SYBR Green Supermix를 사용하여 수행된다. 40 ng의 각 cDNA 라이브러리는 qPCR용 주형으로 사용된다. 2세트의 프라이머는 사용된다; ldhL 유전자의 150-nt 길이의 영역을 증폭하는 BG9665:BG9666 및 qPCR용 대조군으로서 사용된, rpoD (RNA 폴리머라제 시그마 인자) 유

전자의 150-nt 길이의 서열을 증폭시키는 BG9889:BG9890. qPCR은 Bio-Rad C1000 Thermal Cycler에서 실행된다.

- [0427] **k. HPLC**
- [0428] 락테이트 정량화를 위해 고-압 액체 크로마토그래피 (HPLC) 시스템 ICS-5000은 사용된다. 이 시스템은, Bio-Rad Laboratories의 Aminex HPX 87H 컬럼으로 작동되고, RI-150 40℃의 굴절률 검출기 및 210nm에서 작동하는 UV1000 검출기가 장착된다. 이동 상 (mobile phase)은, 0.16 N H₂SO₄로 이루어지고, 컬럼은 0.8 mL/min으로 작동된다. 모든 샘플은 0.01N H₂SO₄에서 10 mM DMSO로 4:1로 희석된다.
- [0429] 본 상세한 설명의 다음의 섹션은 여기서 이미 기재된 본 발명의 주장을 단순히 제공하는 번호가 매겨진 단락으로 이루어진다. 이 섹션에서 번호가 매겨진 단락은 청구항이 아니다. 청구항은 그 다음 섹션의 "청구 범위"에서 서술된다.
- [0430] 1. 하기를 포함하는 단리된 일정한 간격을 두고 주기적으로 분포하는 짧은 회문구조 반복서열 (CRISPR)-관련 (Cas) 단백질 또는 폴리펩티드:
- [0431] a. 아미노산 모티프 EKDGKYYC [SEQ ID NO: 2]; 및/또는
- [0432] b. 아미노산 모티프 X₁X₂CTX₃X₄ [SEQ ID NO: 3], 여기서, X₁은 이소류신, 메티오닌 또는 프롤린으로부터 독립적으로 선택되고, X₂는 발린, 세린, 아스파라긴 또는 이소류신으로부터 독립적으로 선택되며, X₃는 글루타메이트 또는 리신으로부터 독립적으로 선택되고, 및 X₄는 알라닌, 글루타메이트 또는 아르기닌 중 하나임; 및/또는
- [0433] c. 아미노산 모티프 X₅LKX₆IE [SEQ ID NO: 4], 여기서, X₅는 메티오닌 또는 페닐알라닌으로부터 독립적으로 선택되고, 및 X₆는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택됨; 및/또는
- [0434] d. 아미노산 모티프 X₇VYSX₈K [SEQ ID NO: 5], 여기서, X₇은 글루타메이트 또는 이소류신이고, X₈은 트립토판, 세린 또는 리신 중 하나임; 및/또는
- [0435] e. 아미노산 모티프 X₉FYX₁₀X₁₁REQX₁₂KEX₁₃ [SEQ ID NO: 6], 여기서, X₉은 알라닌 또는 글루타메이트이고, X₁₀은 글루타민 또는 라신이며, X₁₁은 아르기닌 또는 알라닌이고, X₁₂는 아스파라긴 또는 알라닌이며, X₁₃은 리신 또는 세린이고;
- [0436] 여기서, Cas 단백질은, 적어도 하나의 표적화 RNA 분자, 및 상기 표적화 RNA 분자에 의해 인지된 표적 핵산 서열을 포함하는 폴리뉴클레오티드와 결합하는 경우, 50℃ 내지 100℃에서 DNA 절단을 할 수 있다.
- [0437] 2. SEQ ID NO: 1의 아미노산 서열 또는 이와 적어도 77% 동일성의 서열을 갖는 단리된 Cas 단백질 또는 이의 폴리펩티드 단편, 여기서, 상기 Cas 단백질은, 표적 서열을 인지하는 적어도 하나의 RNA 분자와 결합하는 경우, 50℃ 내지 100℃의 온도에서 표적 핵산 서열을 포함하는 폴리뉴클레오티드를 결합, 절단, 변형 또는 표지를 할 수 있다.
- [0438] 3. 단락 번호 1 또는 2의 Cas 단백질 또는 폴리펩티드 단편에서, 상기 Cas 단백질 또는 폴리펩티드 단편은, 50℃ 내지 70℃ 범위의 온도, 바람직하게는, 60℃ 이상의 온도; 좀 더 바람직하게는, 60℃ 내지 80℃의 온도; 더 욱 더 바람직하게는, 60℃ 내지 65℃의 온도에서 핵산 절단, 결합, 표지 또는 변형시킬 수 있다.
- [0439] 4. 전술된 단락 번호 1 내지 3 중 어느 하나의 Cas 단백질 또는 폴리펩티드 단편에서, 상기 핵산 결합, 절단, 표지 또는 변형은 DNA 절단이다.
- [0440] 5. 전술된 단락 번호 중 어느 하나의 Cas 단백질 또는 폴리펩티드 단편에서, 상기 아미노산 서열은, SEQ ID NO: 1의 아미노산 서열 또는 이와 적어도 77% 동일성의 서열을 포함한다.
- [0441] 6. 전술된 단락 번호 중 어느 하나의 Cas 단백질 또는 폴리펩티드 단편에서, 상기 Cas 단백질은 박테리아, 고세균 또는 바이러스로부터 얻을 수 있다.
- [0442] 7. 전술된 단락 번호 중 어느 하나의 Cas 단백질 또는 폴리펩티드 단편에서, 상기 Cas 단백질은 지오바실러스 속, 바람직하게는, 지오바실러스 쉐모데니트리피칸스로부터 얻을 수 있다.
- [0443] 8. 전술된 단락 번호 중 어느 하나의 Cas 단백질을 포함하고, 및 표적 폴리뉴클레오티드에서 서열을 인식하는 적어도 하나의 표적화 RNA 분자를 포함하는 리보핵산 단백질 복합체.

- [0444] 9. 단락 번호 8의 리보핵산 단백질 복합체에서, 상기 표적화 RNA 분자는 crRNA 및 선택적으로 tracrRNA를 포함한다.
- [0445] 10. 단락 번호 8 내지 9 중 어느 하나의 리보핵산 단백질 복합체에서, 상기 적어도 하나의 RNA 분자의 길이는 35-135 뉴클레오타이드 잔기의 범위이다.
- [0446] 11. 단락 번호 8 내지 9 중 어느 하나의 리보핵산 단백질 복합체에서, 상기 표적 서열은 길이가 31 또는 32 뉴클레오타이드 잔기이다.
- [0447] 12. 단락 번호 1 내지 7 중 어느 하나의 Cas 단백질 또는 폴리펩티드에서, 또는 단락 번호 8 내지 11 중 어느 하나의 리보핵산 단백질 복합체에서, 상기 단백질 또는 폴리펩티드는, 적어도 하나의 추가의 기능성 또는 비-기능성 단백질을 포함하는 단백질 복합체의 일부로서 제공된다.
- [0448] 13. 단락 번호 12의 Cas 단백질, 폴리펩티드, 또는 리보핵산 단백질 복합체에서, 상기 Cas 단백질 또는 폴리펩티드는, 및/또는 적어도 하나의 추가의 단백질은, 적어도 하나의 기능성 모이어티를 더욱 포함한다.
- [0449] 14. 단락 번호 13의 Cas 단백질, 폴리펩티드, 또는 리보핵산 단백질 복합체에서, 상기 적어도 하나의 기능성 모이어티는, Cas 단백질, 폴리펩티드, 또는 리보핵산 단백질 복합체의 N-말단 및/또는 C-말단; 바람직하게는, N-말단에 융합 또는 연결된다.
- [0450] 15. 단락 번호 13 또는 14의 Cas 단백질, 폴리펩티드, 또는 리보핵산 단백질 복합체에서, 상기 적어도 하나의 기능성 모이어티는: 헬리카제, 뉴클레아제, 헬리카제-뉴클레아제, DNA 메틸라아제, 히스톤 메틸라아제, 아세틸라제, 포스파타아제, 키나아제, 전사 (공-)활성인자, 전사 억제인자, DNA 결합 단백질, DNA 구조 단백질, 마커 단백질, 리포터 단백질, 형광 단백질, 리간드 결합 단백질, 신호 펩티드, 세포내 위치정보 서열, 항체 에피토프 또는 친화성 정제 태그로부터 선택적으로 선택된 단백질이다.
- [0451] 16. 단락 번호 15의 Cas 단백질 또는 폴리펩티드, 또는 리보핵산 단백질 복합체에서, 상기 Cas9 뉴클레아제 활성의 고유 활성은 비활성화되고, 상기 Cas 단백질은 적어도 하나의 기능성 모이어티에 연결된다.
- [0452] 17. 단락 번호 15 또는 16의 Cas 단백질 또는 폴리펩티드, 또는 리보핵산 단백질 복합체에서, 상기 적어도 하나의 기능성 모이어티는 뉴클레아제 도메인; 바람직하게는, FokI 뉴클레아제 도메인이다.
- [0453] 18. 단락 번호 15 내지 17 중 어느 하나의 Cas 단백질 또는 폴리펩티드, 또는 리보핵산 단백질 복합체에서, 상기 적어도 하나의 기능성 모이어티는 마커 단백질, 예를 들어, GFP이다.
- [0454] 19. Cas 단백질 또는 폴리펩티드를 인코딩하는 단리된 핵산 분자로서, 상기 단리된 핵산 분자는:
- [0455] a. 아미노산 모티프 EKDGYK [SEQ ID NO: 2]; 및/또는
- [0456] b. 아미노산 모티프 $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], 여기서, X_1 은 이소류신, 메티오닌 또는 프롤린으로부터 독립적으로 선택되고, X_2 는 발린, 세린, 아스파라긴 또는 이소류신으로부터 독립적으로 선택되며, X_3 는 글루타메이트 또는 리신으로부터 독립적으로 선택되고, 및 X_4 는 알라닌, 글루타메이트 또는 아르기닌 중 하나임; 및/또는
- [0457] c. 아미노산 모티프 X_5LXX_6IE [SEQ ID NO: 4], 여기서, X_5 는 메티오닌 또는 페닐알라닌으로부터 독립적으로 선택되고, 및 X_6 는 히스티딘 또는 아스파라긴으로부터 독립적으로 선택됨; 및/또는
- [0458] d. 아미노산 모티프 X_7VYSX_8K [SEQ ID NO: 5], 여기서, X_7 은 글루타메이트 또는 이소류신이고, X_8 은 트립토판, 세린 또는 리신 중 하나임; 및/또는
- [0459] e. 아미노산 모티프 $X_9FYX_{10}X_{11}REX_{12}KEX_{13}$ [SEQ ID NO: 6]을 포함하며, 여기서, X_9 은 알라닌 또는 글루타메이트이고, X_{10} 은 글루타민 또는 라신이며, X_{11} 은 아르기닌 또는 알라닌이고, X_{12} 는 아스파라긴 또는 알라닌이며, X_{13} 은 리신 또는 세린이고;
- [0460] 여기서, Cas 단백질 또는 폴리펩티드는, 적어도 하나의 표적화 RNA 분자, 및 상기 표적화 RNA 분자에 의해 인지된 표적 핵산 서열을 포함하는 폴리뉴클레오타이드와 결합하는 경우, 50°C 내지 100°C에서 DNA 결합, 절단, 표지 또는 변형시킬 수 있다.
- [0461] 20. SEQ ID NO: 1의 아미노산 서열 또는 이와 적어도 77% 동일성의 서열을 갖는 일정한 간격을 두고 주기적으로

분포하는 짧은 회문구조 반복서열 (CRISPR)-관련 (Cas) 단백질; 또는 이의 폴리펩티드 단편을 인코딩하는 단리된 핵산 분자.

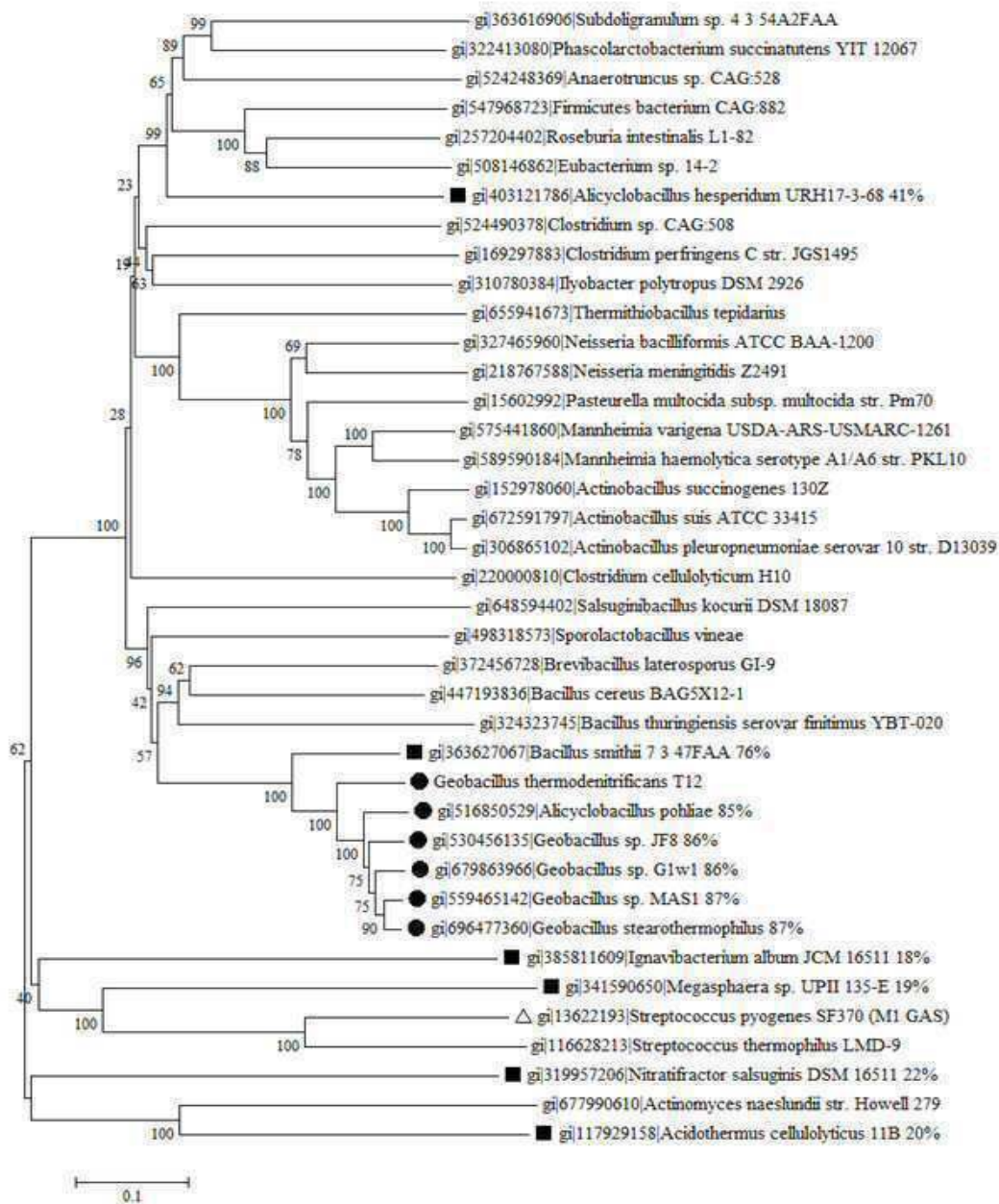
- [0462] 21. 단락 번호 19 또는 20의 단리된 핵산 분자에서, 번역시에 Cas 단백질 또는 폴리펩티드와 융합되는 아미노산 서열을 인코딩하는 적어도 하나의 핵산 서열을 더욱 포함한다.
- [0463] 22. 단락 번호 21의 단리된 핵산 분자에서, 상기 Cas 단백질 또는 폴리펩티드를 인코딩하는 핵산 분자에 융합된 적어도 하나의 핵산 서열은, 헬리카제, 뉴클레아제, 헬리카제-뉴클레아제, DNA 메틸라아제, 히스톤 메틸라아제, 아세틸라제, 포스포타아제, 키나아제, 전사 (공-)활성인자, 전사 억제인자, DNA 결합 단백질, DNA 구조 단백질, 마커 단백질, 리포터 단백질, 형광 단백질, 리간드 결합 단백질, 신호 펩티드, 세포내 위치정보 서열, 항체 에피토프 또는 친화성 정제 태그로부터 선택된 단백질을 인코딩한다.
- [0464] 23. 단락 번호 19 내지 22 중 어느 하나의 핵산 분자를 포함하는 발현 벡터.
- [0465] 24. 단락 번호 23의 발현 벡터에서, 적어도 하나의 표적화 RNA 분자를 인코딩하는 뉴클레오티드 서열을 더욱 포함한다.
- [0466] 25. 하기와 핵산을 접촉시키는 단계를 포함하는, 인간 세포에서 표적 핵산을 변형시키는 방법:
- [0467] a. 단락 번호 6 내지 11 중 어느 하나의 리보핵산 단백질 복합체; 또는
- [0468] b. 단락 번호 12 내지 18 중 어느 하나의 단백질 또는 단백질 복합체 및 단락 번호 6 내지 11 중 어느 하나에서 정의된 바와 같은 적어도 하나의 표적화 RNA 분자; 선택적으로, 여기서, 상기 인간 세포는 배아 줄기 세포가 아니며; 및 선택적으로, 여기서, 상기 방법은 인간의 생식 계열 유전적 동일성을 변형시키지 않는다.
- [0469] 26. 단락 번호 24의 발현 벡터로 세포를 형질전환, 형질감염 또는 형질도입하는 단계; 또는 선택적으로, 단락 번호 23의 발현 벡터 및 단락 번호 6 내지 11 중 어느 하나에 정의된 바와 같은 표적화 RNA 분자를 인코딩하는 뉴클레오티드 서열을 포함하는 추가의 발현 벡터로 세포를 형질전환, 형질감염 또는 형질도입하는 단계를 포함하는, 인간 세포에서 표적 핵산을 변형시키는 방법으로서; 선택적으로, 여기서, 상기 인간 세포는 배아 줄기 세포가 아니며; 및 선택적으로, 여기서, 상기 방법은 인간의 생식 계열 유전적 동일성을 변형시키지 않는다.
- [0470] 27. 단락 번호 23의 발현 벡터로 세포를 형질전환, 형질감염 또는 형질도입하는 단계, 및 그 다음 단락 번호 6 내지 11 중 어느 하나에 정의된 바와 같은 표적화 RNA 분자를 세포에 또는 세포 내로 전달하는 단계를 포함하는, 인간 세포에서 표적 핵산을 변형시키는 방법으로서; 선택적으로, 상기 인간 세포는 배아 줄기 세포가 아니며; 및 선택적으로, 여기서, 상기 방법은 인간의 생식 계열 유전적 동일성을 변형시키지 않는다.
- [0471] 28. 단락 번호 25 내지 28 중 어느 한 항의 표적 핵산을 변형시키는 방법에서, 상기 적어도 하나의 기능성 모이어티는, 마커 단백질 또는 리포터 단백질이고, 상기 마커 단백질 또는 리포터 단백질은 표적 핵산과 결합하고; 바람직하게는 상기 마커는 형광 단백질, 예를 들어, 녹색 형광 단백질 (GFP)이다.
- [0472] 29. 단락 번호 25 내지 28 중 어느 한 항의 표적 핵산을 변형시키는 방법에서, 상기 표적 핵산은 DNA; 바람직하게는 dsDNA이다.
- [0473] 30. 단락 번호 25 내지 28 중 어느 한 항의 표적 핵산을 변형시키는 방법에서, 상기 표적 핵산은 RNA이다.
- [0474] 31. 단락 번호 29의 표적 핵산을 변형시키는 방법에서, 상기 핵산은 dsDNA이고, 상기 적어도 하나의 기능성 모이어티는 뉴클레아제 또는 헬리카제-뉴클레아제이며, 및 상기 변형은 원하는 유전자좌에서 단일-가닥 또는 이중-가닥 절단이다.
- [0475] 32. 단락 번호 26, 27, 29 또는 31 중 어느 하나의 방법 중 어느 하나에 따라 원하는 유전자좌에서 유전자 발현을 침묵시키는 방법.
- [0476] 33. 단락 번호 26, 27, 29 또는 31 중 어느 하나의 방법 중 어느 하나에 따라 원하는 위치에 원하는 뉴클레오티드 서열을 변형 또는 결손 및/또는 삽입하는 방법.
- [0477] 34. 단락 번호 25 내지 29 중 어느 하나의 방법에서와 같이 표적 핵산 서열을 변형시키는 단계를 포함하는, 세포에서 유전자 발현을 변형시키는 방법에서, 상기 핵산은 dsDNA이고, 상기 기능성 모이어티는 DNA 변형 효소 (예를 들어, 메틸라아제 또는 아세틸라제), 전사 활성인자 또는 전사 억제인자로부터 선택된다.
- [0478] 35. 단락 번호 30의 방법에서와 같이 표적 핵산 서열을 변형시키는 단계를 포함하는, 세포에서 유전자 발현을 변형시키는 방법에서, 상기 핵산은 mRNA이고, 상기 기능성 모이어티는 리보뉴클레아제이며; 선택적으로 엔도뉴

클레아제, 3' 엑소뉴클레아제 또는 5' 엑소뉴클레아제로부터 선택된다.

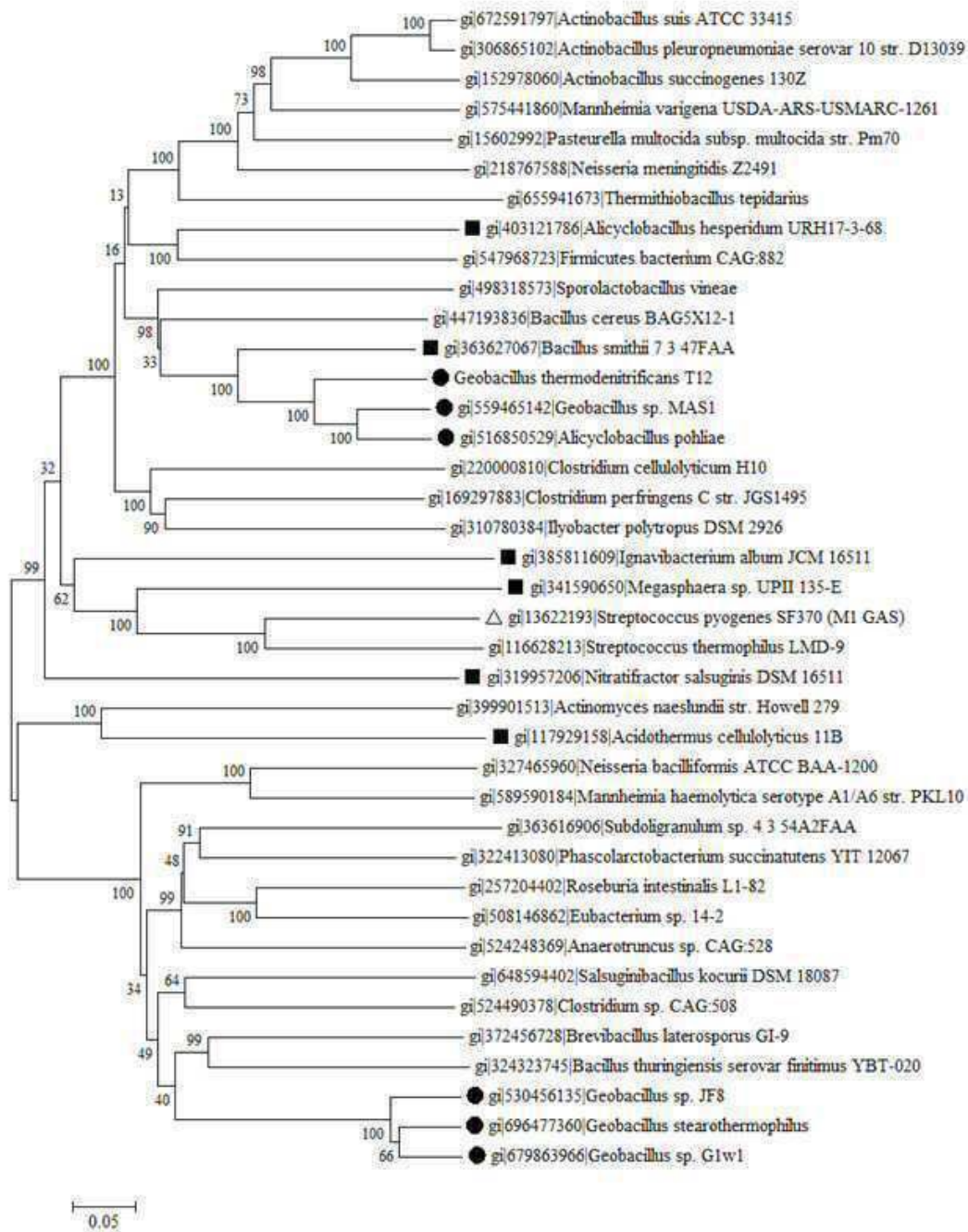
- [0479] 36. 단락 번호 25 내지 35 중 어느 하나에서와 같이 표적 핵산을 변형시키는 방법에서, 상기 방법은, 50℃ 내지 100℃의 온도에서 수행된다.
- [0480] 37. 단락 번호 36에서와 같이 표적 핵산을 변형시키는 방법에서, 상기 방법은, 60℃ 이상, 바람직하게는 60℃ 내지 80℃, 보다 바람직하게는 60℃ 내지 65℃의 온도에서 수행된다.
- [0481] 단락 번호 22 내지 36 중 어느 하나의 방법에 의해 형질전환된 숙주 세포로서; 여기서, 상기 세포는 인간 세포이고, 선택적으로, 상기 인간 세포는, 배아 줄기 세포가 아니며; 및 선택적으로, 여기서, 상기 방법은 인간의 생식 계열 유전적 동일성을 변형시키지 않는다.

도면

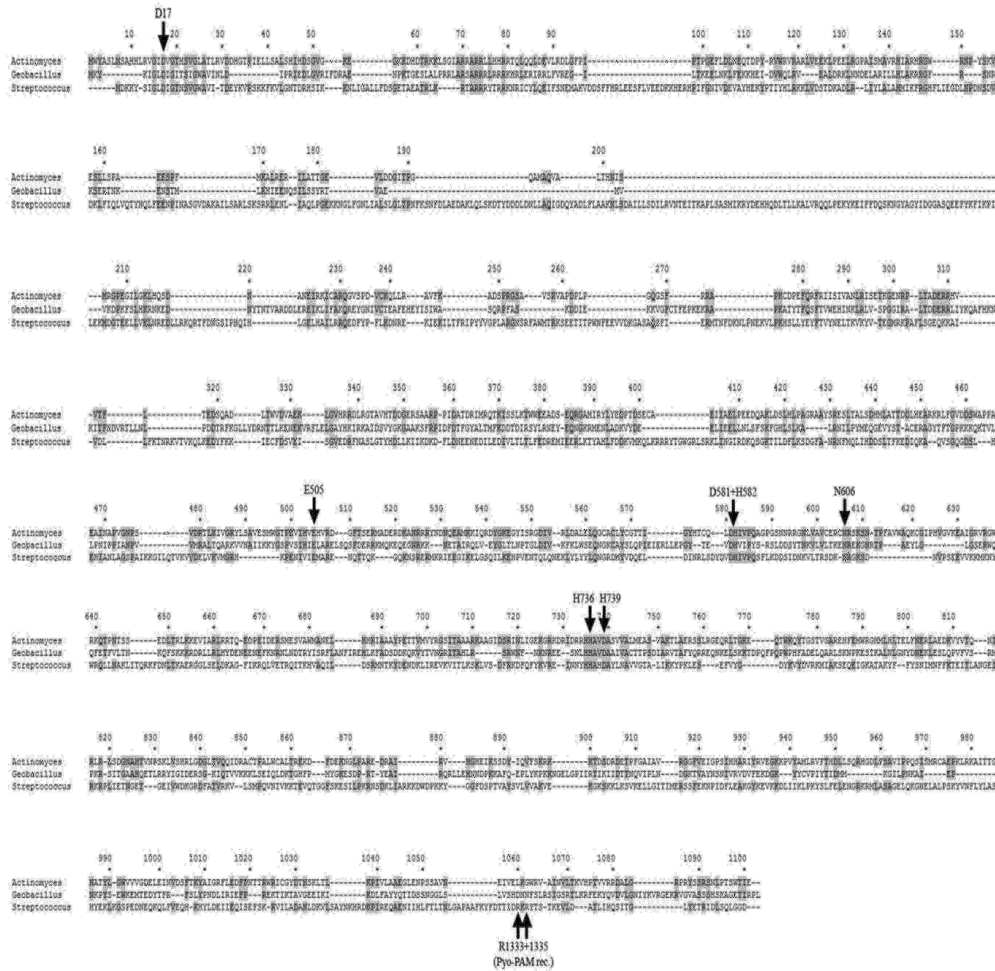
도면1



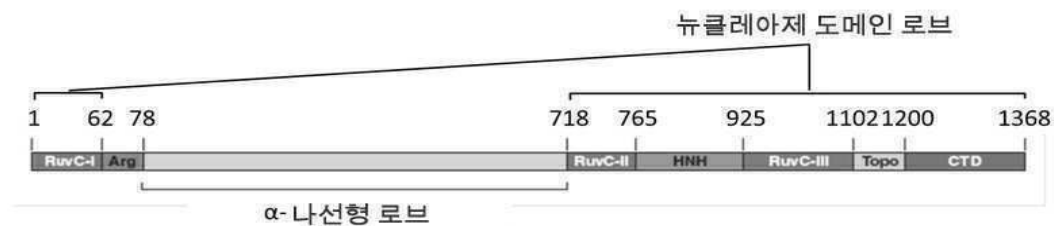
도면2



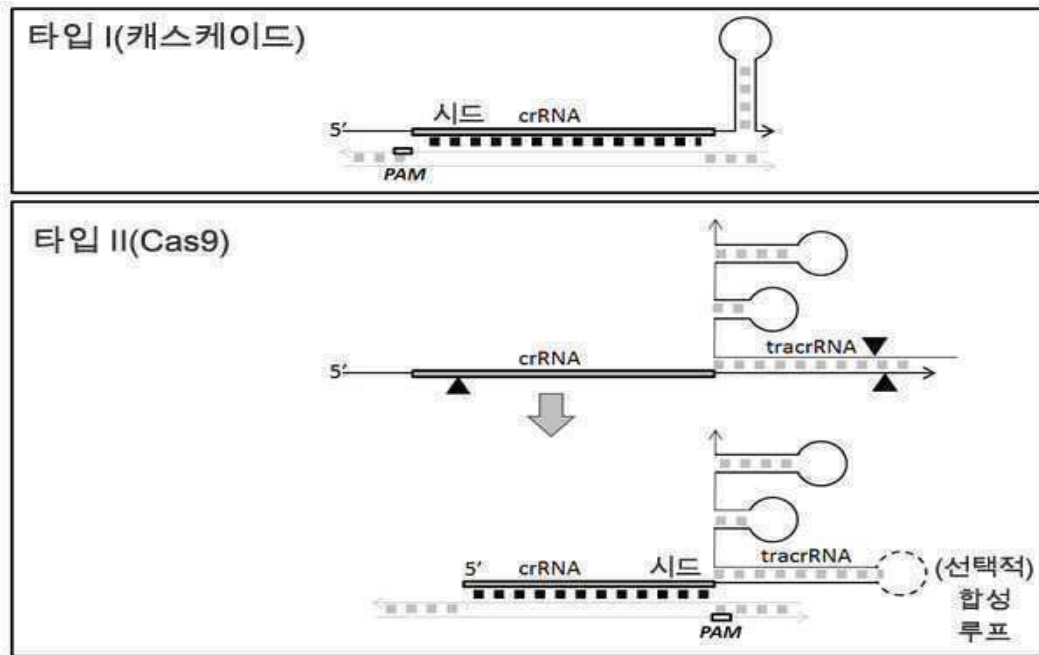
도면3



도면4



도면5



도면6

```

tracrRNA t12      1 attgttttccctcccatgcacaatagttttatagtaaaaaagaccttgaogttttccgccaaggtcttc
repeat gb      -----

tracrRNA t12      71 gttgcctaagagtgggaatgccgaagaaagcggcgataggcgatccccaacgccacgggtcagttc
repeat gb      -----

tracrRNA t12      141 gcctataggcagaaaagcccttatcatagtaaccctgagatcattgctgtgtgtataaacctattactataa
repeat gb      1 -----gtcatagttccctgagattatcgctgtgtgtataat-----

tracrRNA t12      211 taatgttttatatttgggaaaatcaagtcctttttctatatattttttatactttcatttcttcttgcattat
repeat gb      -----

tracrRNA t12      281 gatgatgtgagggaggatagatttctgacaggaggtttcacatcg
repeat gb      -----
    
```


도면7a

1. Match to: [DQ453159](#) [DQ453159](#) [Geobacillus virus E2, complete genome\(DQ453159\)](#) position: 39980-39951, with: [spacer4](#) CRISPR No.4 Spacer No.1 (spacer4_4_1) position: 1-30, Strand: +

5'	-----UUGGCGGUGCGAAUUCUAACCGUCCCGGA-----	3'	<- CRISPR RNA
spacer			
3'	TAGCGTACAAACGCCACGCTTAAGATTGGCAGGGCCTTCGTGGTGT	5'	<- Sequence
Protospacer			
5'	ATCGCATGTTGGCGGTGCGAATTCTAACCGTCCCGGAAGCACCACA	3'	<- [Entrez Nucleotide]

Score: 30

2. Match to: [Bacillus alveayuensis strain 24KAM51 LG50_053, whole genome shotgun sequence\(NZ_JYCE01000053\)](#) position: 110918-110889, with: [spacer9](#) CRISPR No.9 Spacer No.1 (spacer9_9_1) position: 1-30, Strand: +

5'	-----UUCUACCUCUACUCUGAUUCACGAAUCGG-----	3'	<- CRISPR RNA
spacer			
3'	CGACGAGAAGATGGAGATGAGAGCTAAGTGCTTAGCCAAAAGGTT	5'	<- Sequence
Protospacer			
5'	GCTGCTCTTCTACCTCTACTCTCGATTACGAATCGGTTTTCCAA	3'	<- [Entrez Nucleotide]

Score: 30

3. Match to: [Anoxybacillus flavithermus WK1, complete genome\(CP000922\)](#) position: 701422-701450, with: [spacer5](#) CRISPR No.5 Spacer No.1 (spacer5_5_1) position: 1-29, Strand: -

5'	-----UCACGGAGCUUUACACAAUAAAGCCGGA-----	3'	<- CRISPR RNA
spacer			
3'	GATGTTCTAGTGCTCGAAATGTGTTTATTTCGGCCTCGTGCTT	5'	<- Sequence
Protospacer			
5'	CTACAAGATCACGGAGCTTTACACAAATAAAGCCGGAAGCGACGAA	3'	<- [Entrez Nucleotide]

Score: 29

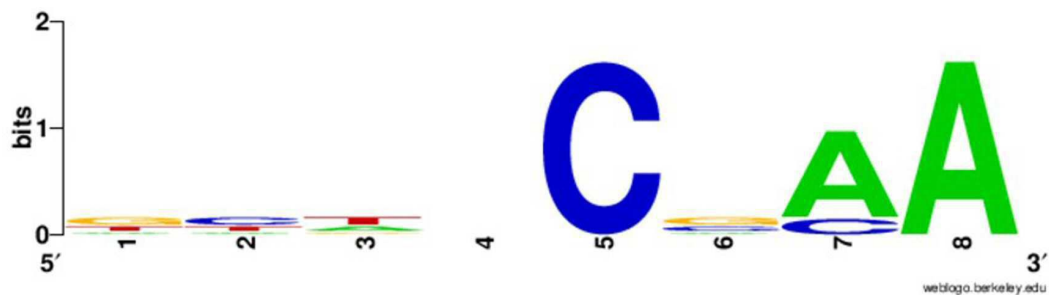
4. Match to: [Geobacillus kaustophilus strain Et23 LG51_086, whole genome shotgun sequence\(NZ_JYCF01000086\)](#) position: 15565-15537, with: [spacer5](#) CRISPR No.5 Spacer No.1 (spacer5_5_1) position: 1-29, Strand: +

5'	-----UCACGGAGCUUUACACAAUAAAGCCGGA-----	3'	<- CRISPR RNA
spacer			
3'	CATGTTCTAGTGCTCGAAATGTGTTTATTTCGGCCTCGTGCTT	5'	<- Sequence
Protospacer			

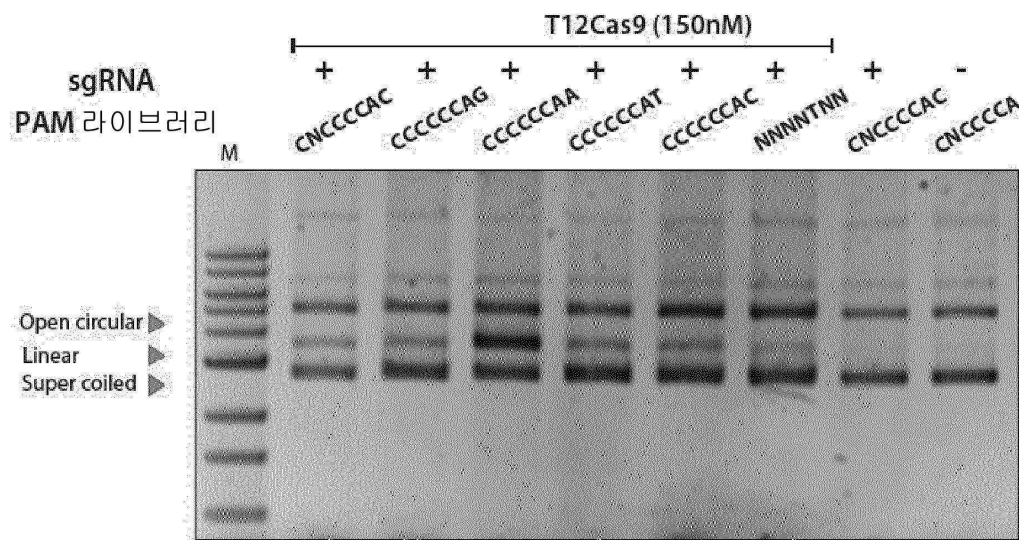
도면7b



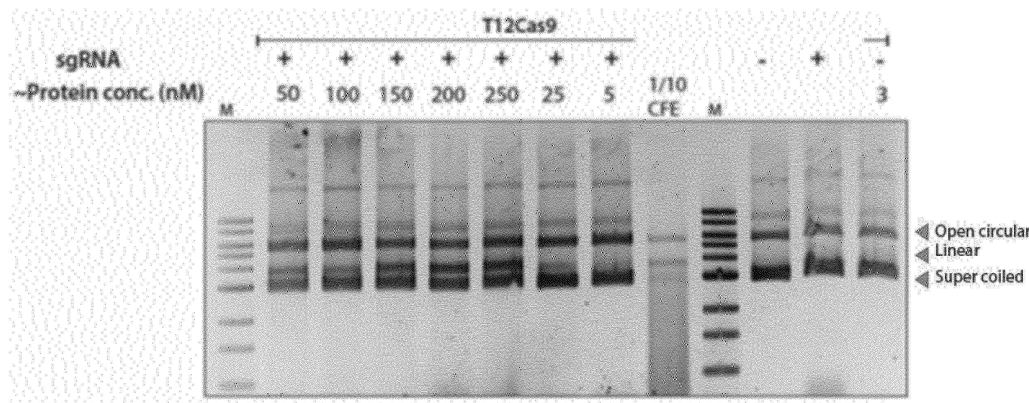
도면8



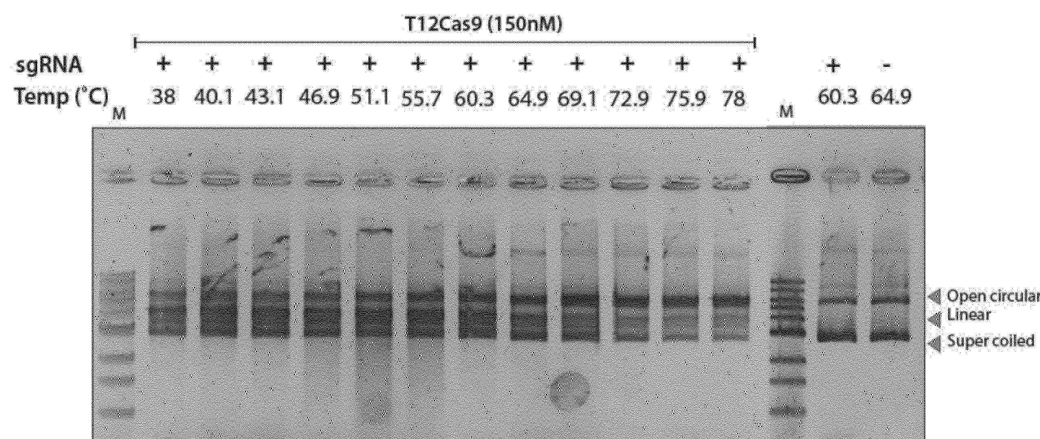
도면9



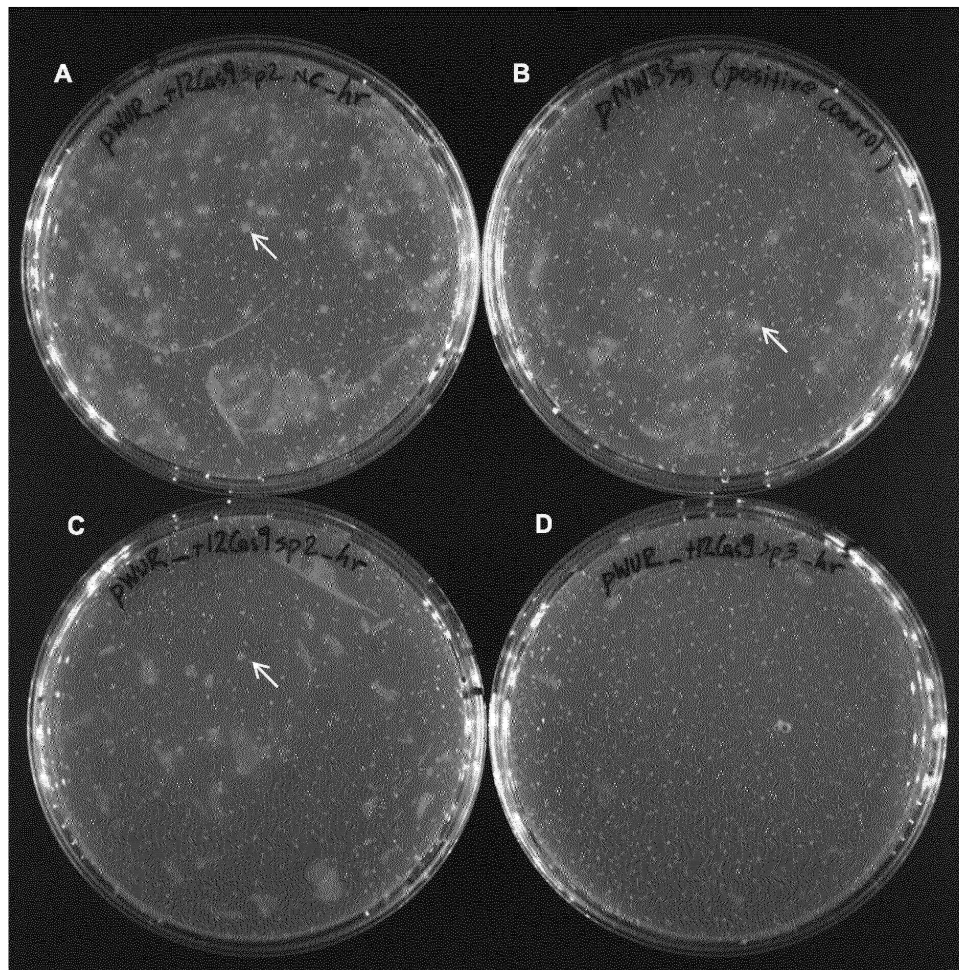
도면10



도면11

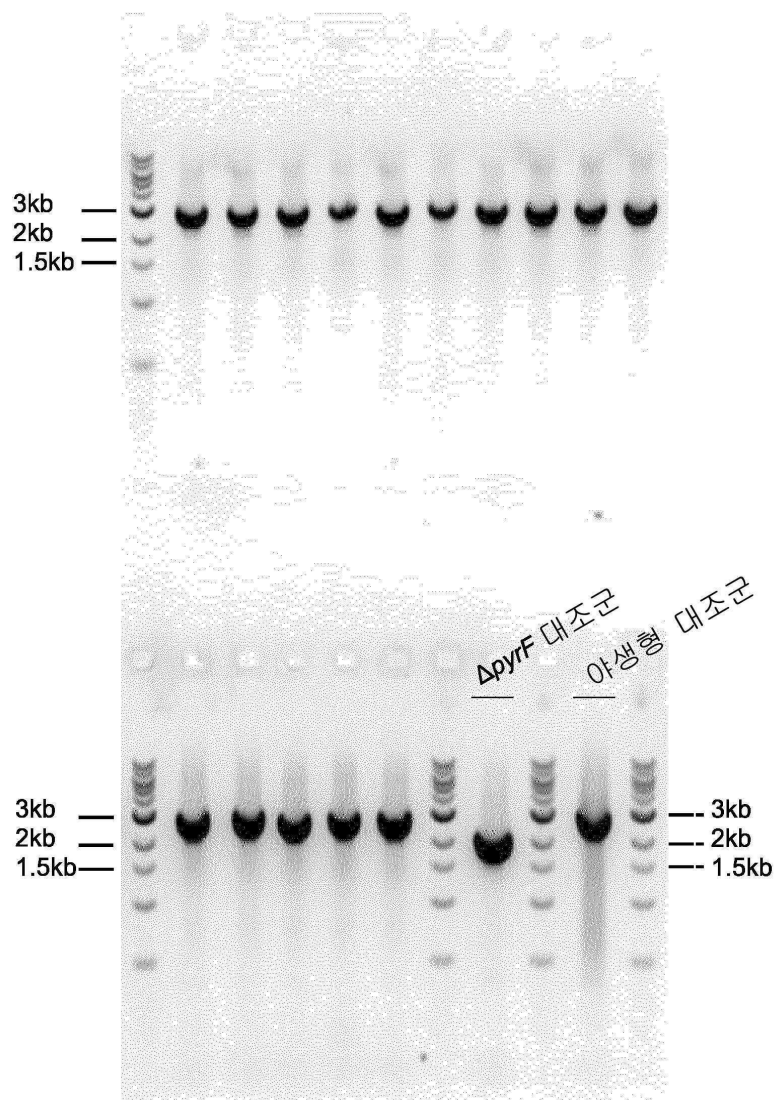


도면12

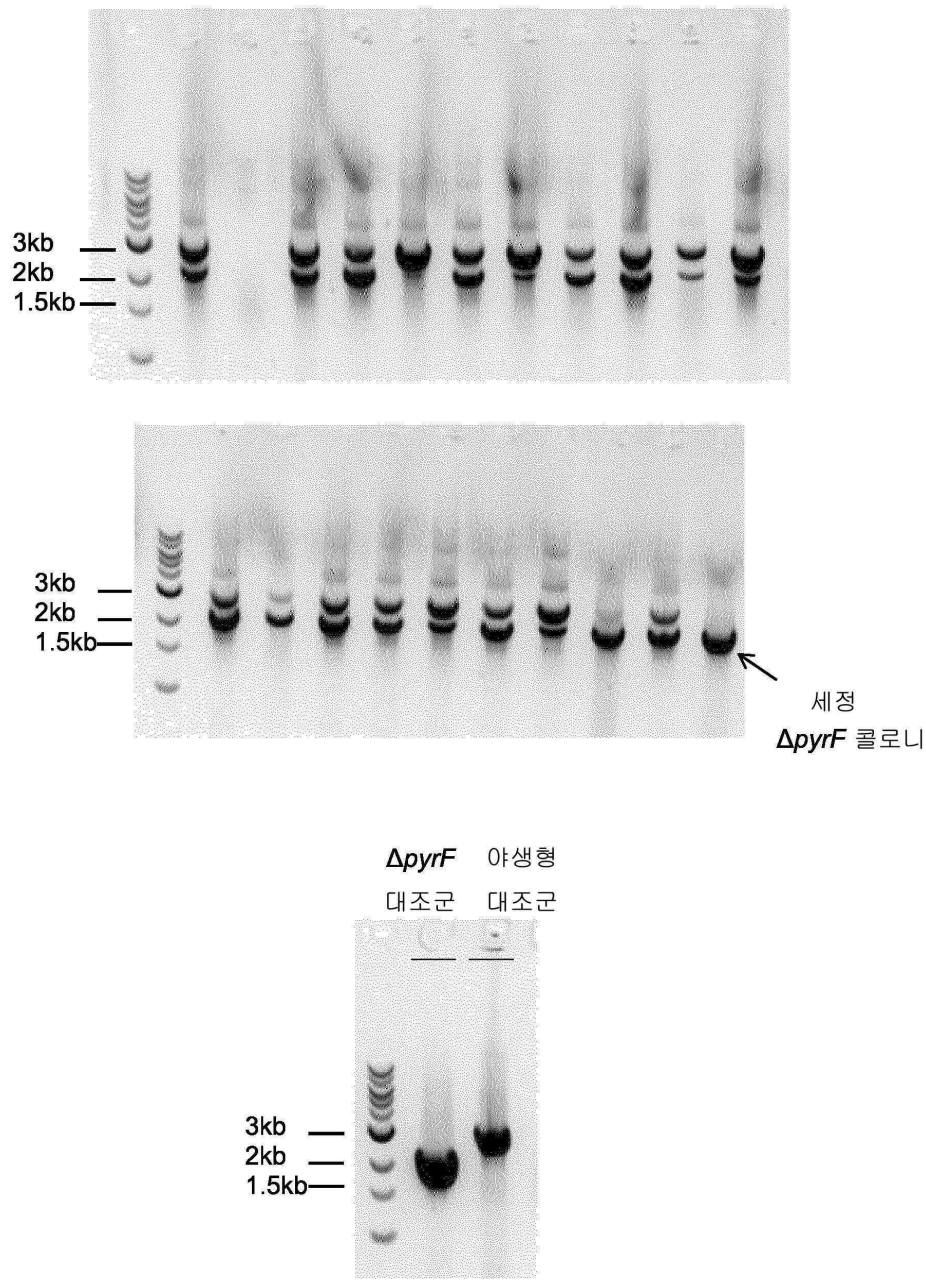


콜로니는 화살표로 나타냄.

도면13

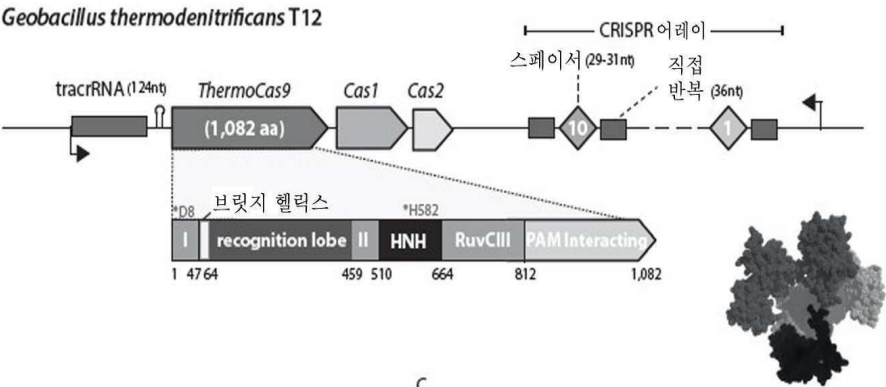


도면14

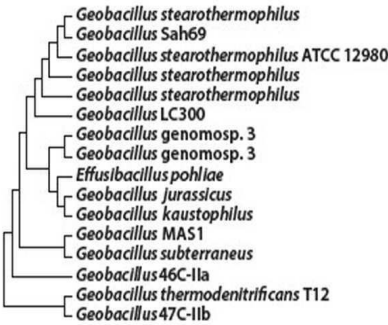


도면15

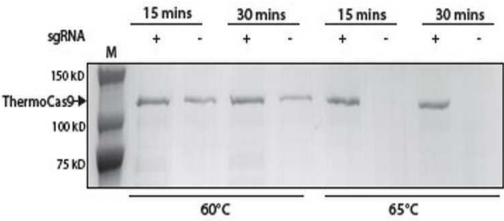
A.



B.



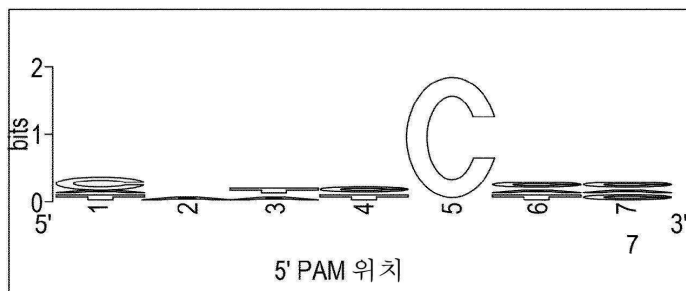
C.



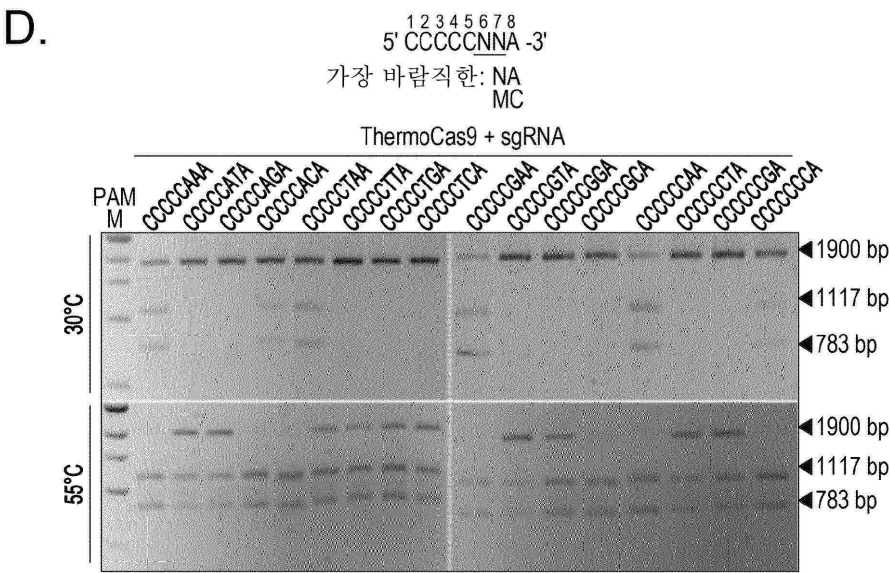
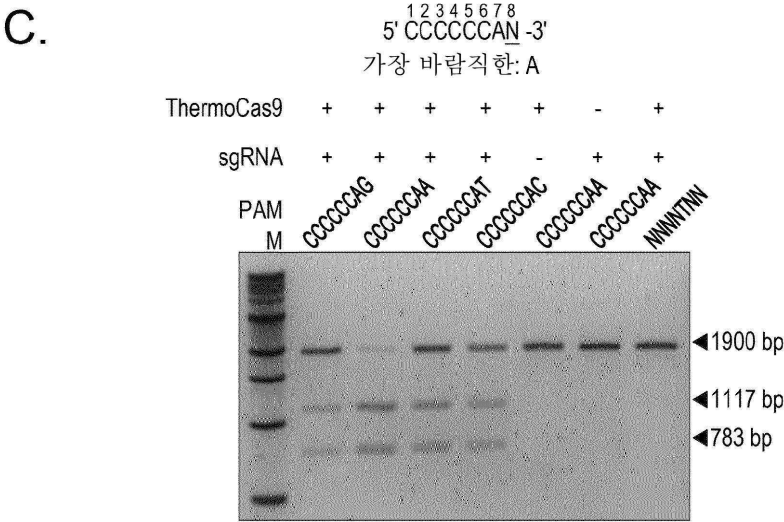
도면 16a



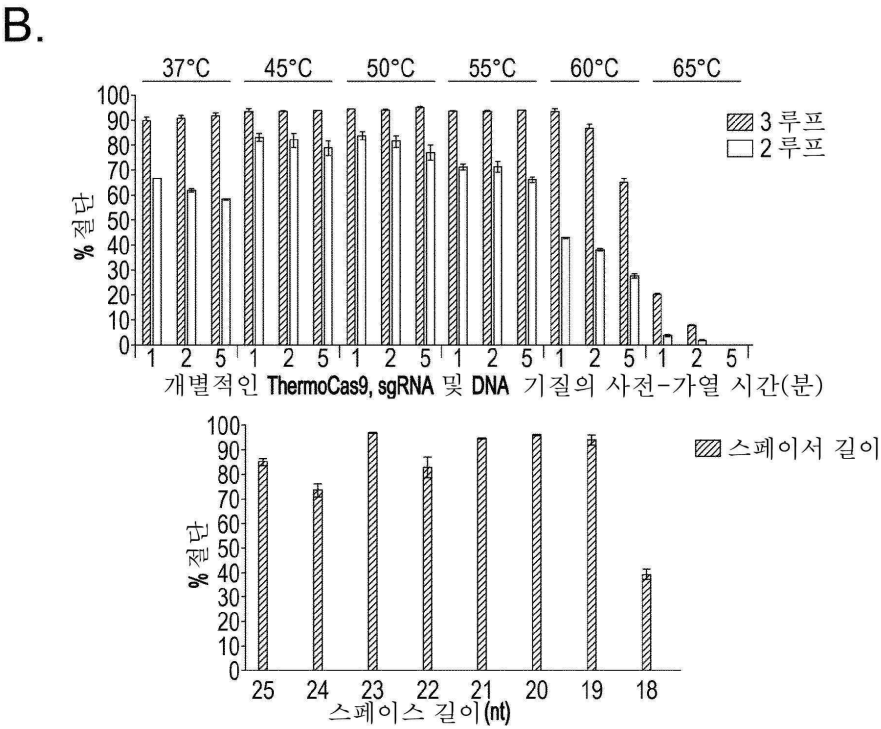
B.



도면 16b

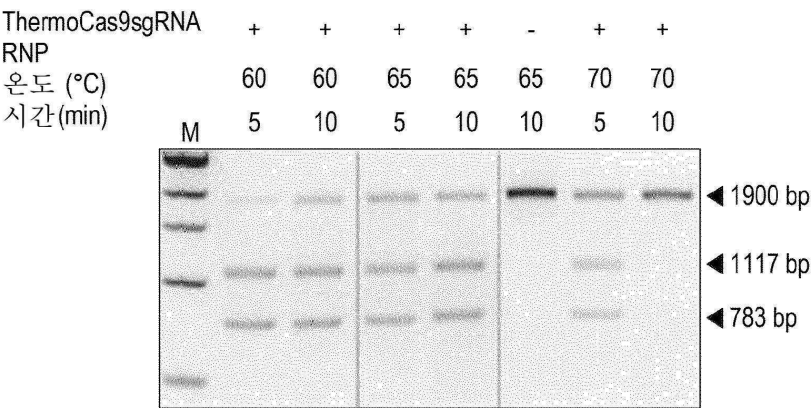


도면17a

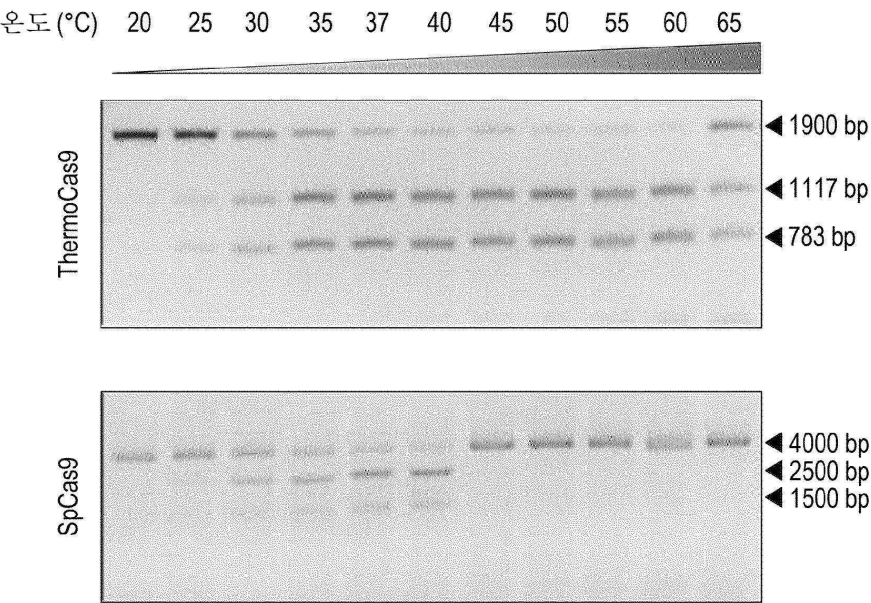


도면17b

C.

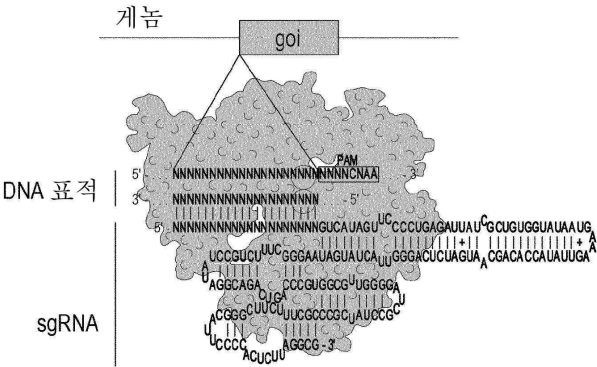
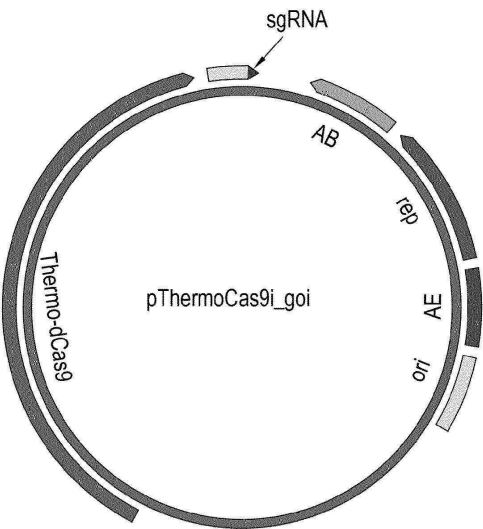


D.

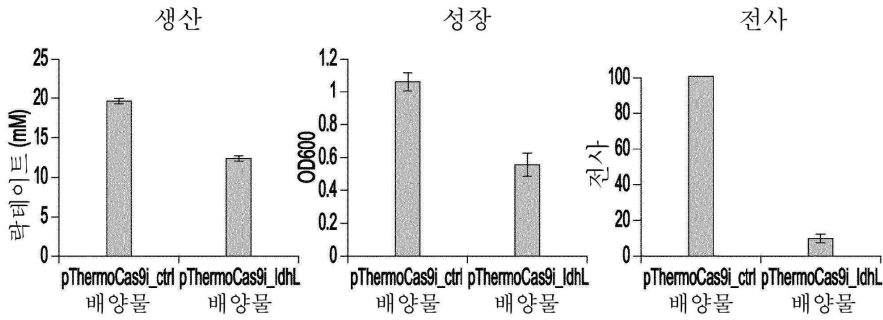


도면18b

C.



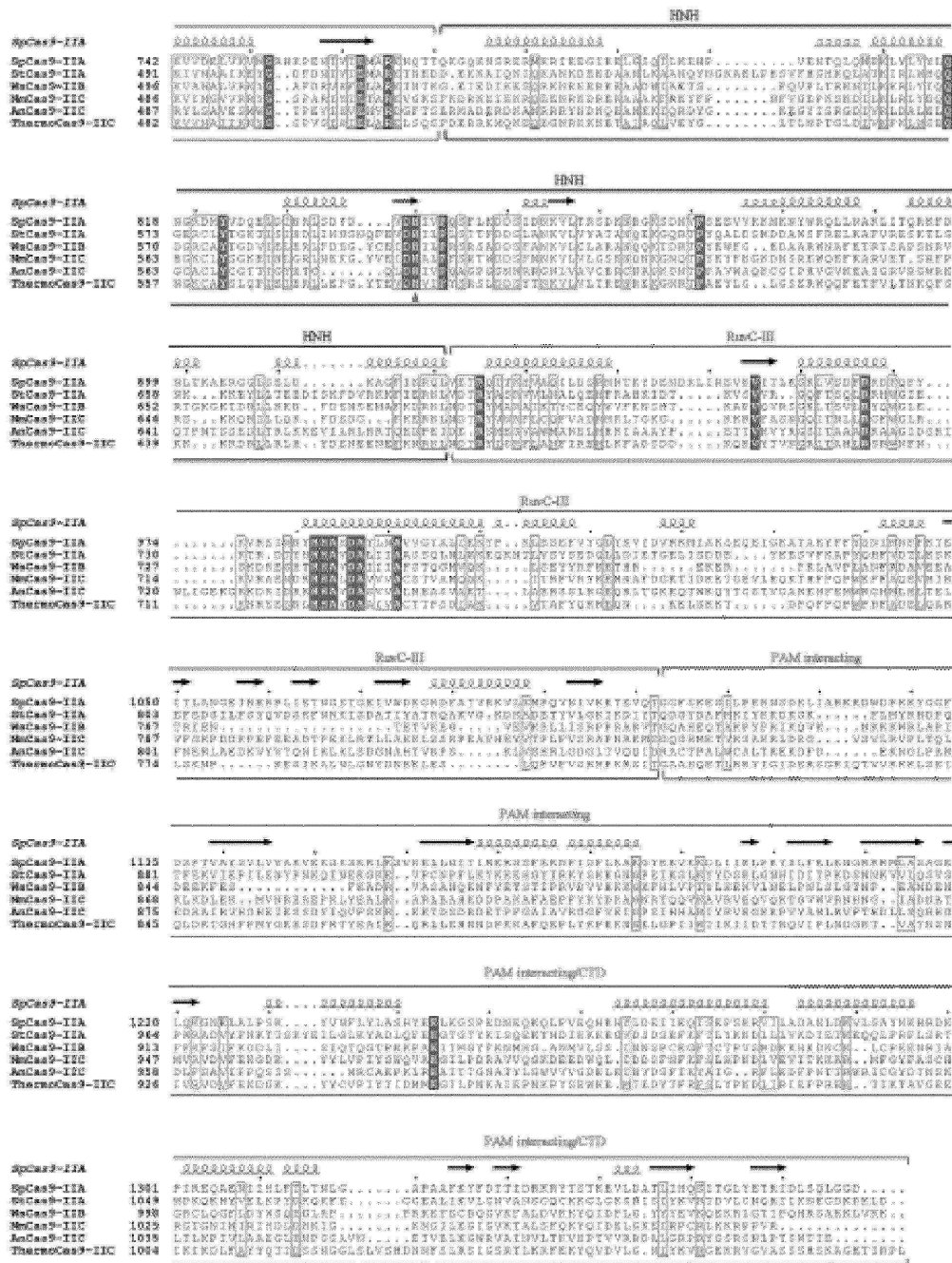
D.



도면 19a

Figure 1 displays the multiple sequence alignment of RunC domain proteins. The figure is organized into 12 blocks, each corresponding to a specific RunC domain (RunC-1 through RunC-12). Each block contains the amino acid sequences of four proteins: SpCas9-11A, SpCas9-11B, SpCas9-11C, and ThermaCas9-11C. The sequences are aligned to show conserved residues, with some positions highlighted in grey. Above each block, a schematic diagram shows the domain structure with arrows indicating the direction of the sequence. The alignment is presented in a tabular format with sequence identifiers on the left and the corresponding amino acid sequences on the right.

도면19b



도면20

A.

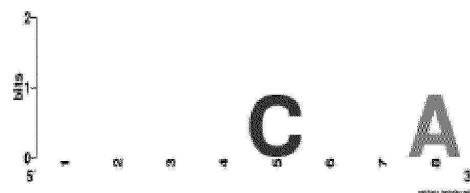
Match to: **DQ453159** **DQ453159** **Geobacillus virus E2, complete genome(DQ453159)**
 position: 39980-39951, with: **spacer4** CRISPR No.4 Spacer No.1 (spacer4_4_1) position: 1-30,
 Strand: +

5' -----UUGGCGGUGCGAAUUCUAACCGVCCCGGAA----- 3'	<- CRISPR
spacer	RNA
3' TAGCGTACAAACCGCCACGCTTAAGATTGGCAGGGCCTTCGTGGTGT 5'	<-
Protospacer	Sequence
5' ATCGCATGTTGGCGGTGCGAATTCTAACCGTCCCGGAAGCACCACA 3'	<-
[Entrez Nucleotide]	

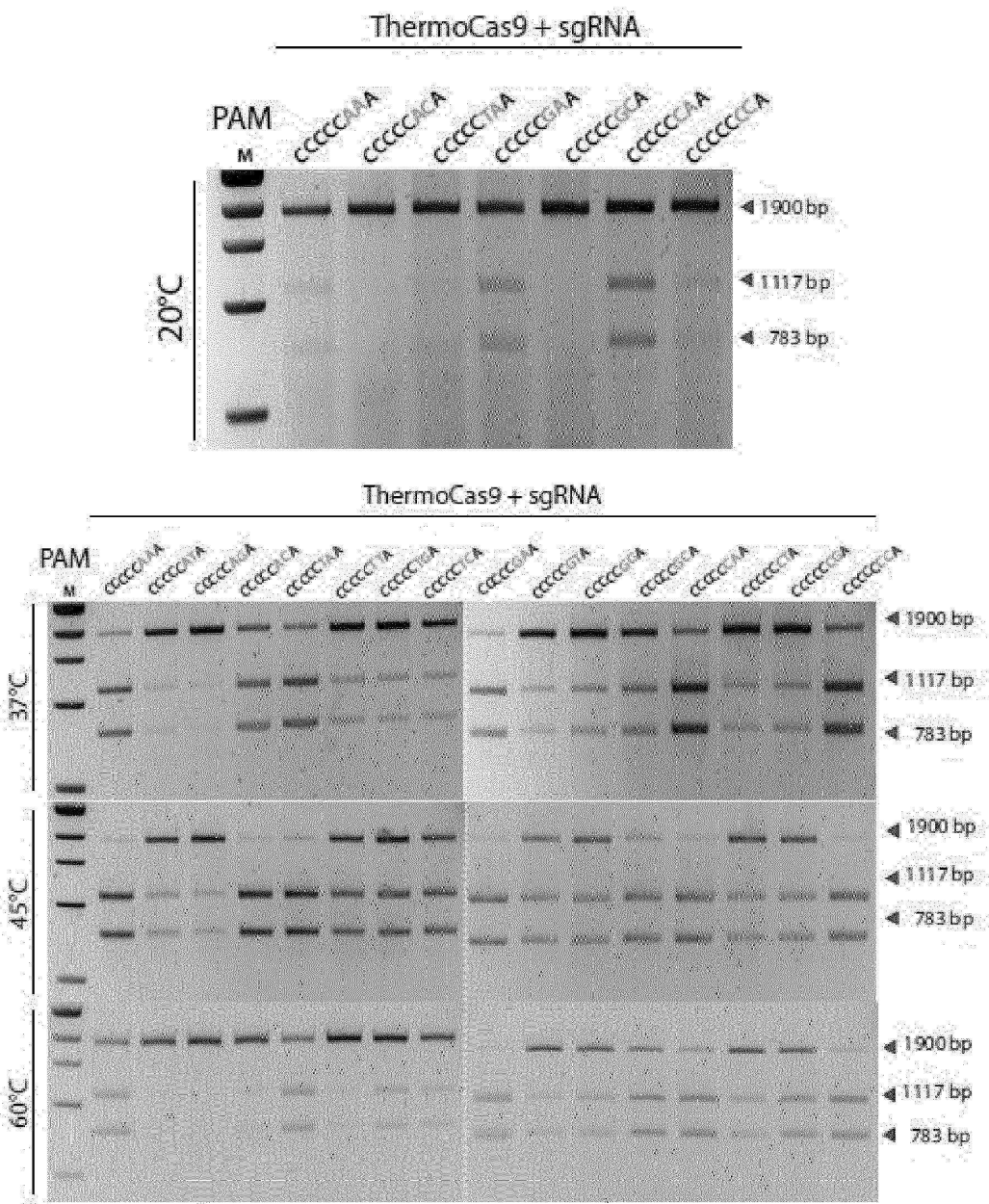
Match to: **DQ453159** **DQ453159** **Geobacillus virus E2, complete genome(DQ453159)**
 position: 6492-6463, with: **spacer6** CRISPR No.6 Spacer No.1 (spacer6_6_1) position: 1-30,
 Strand: +

5' -----CAACACCUUCCGCGCUGUCUCGUCUACUUU----- 3'	<- CRISPR
spacer	RNA
3' CGAAGAAAGTTTGGGAGGCACGACAGAGTAGATGAAAAAATGCTT 5'	<-
Protospacer	Sequence
5' GCTTCCTTCAACACCTTCCGTGCTGTCTCATCTTCTTTACGAA 3'	<-
[Entrez Nucleotide]	

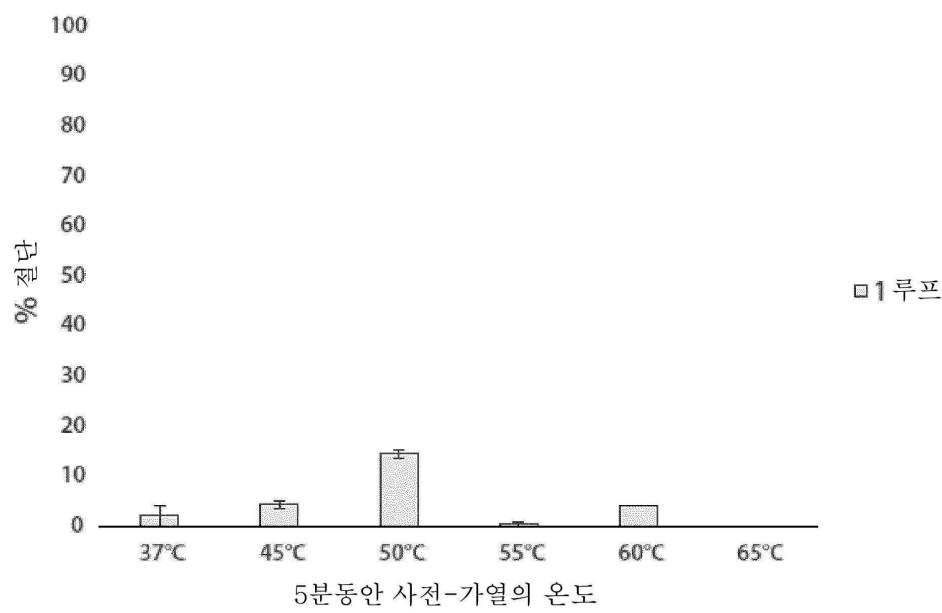
B.



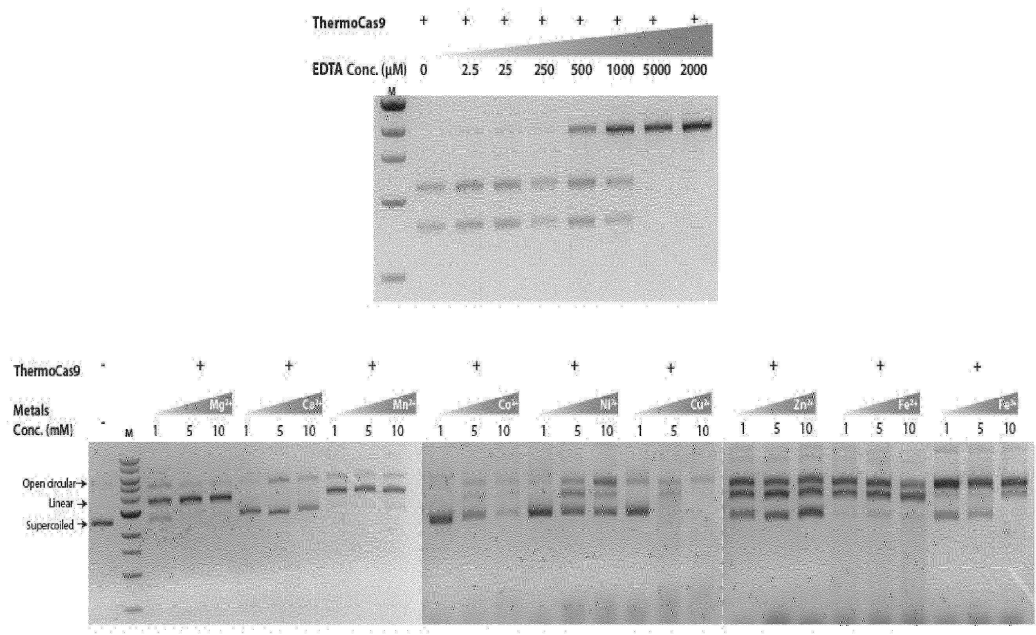
도면21



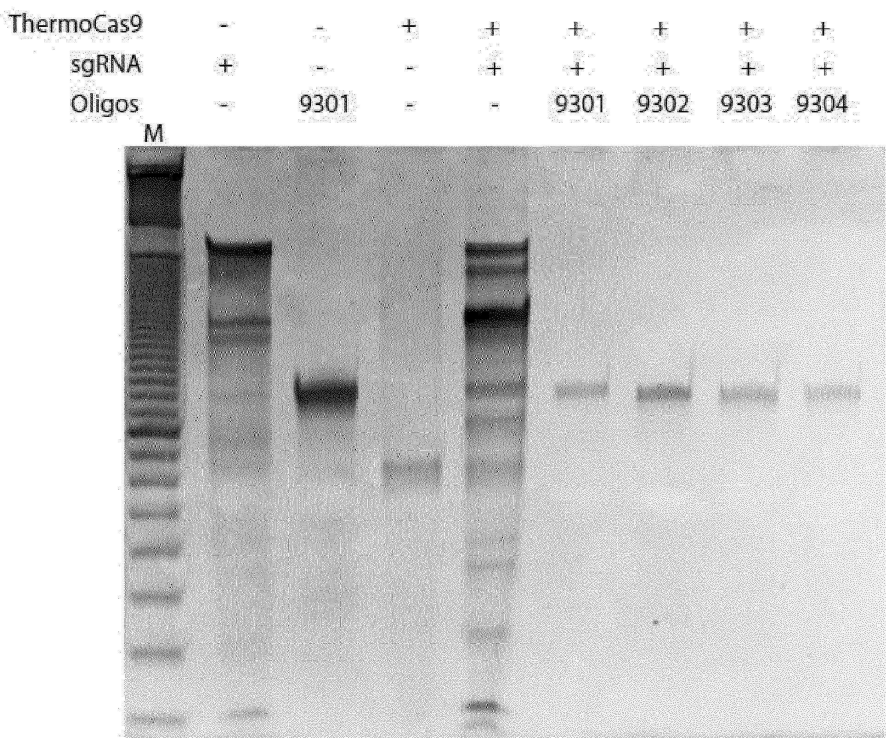
도면22



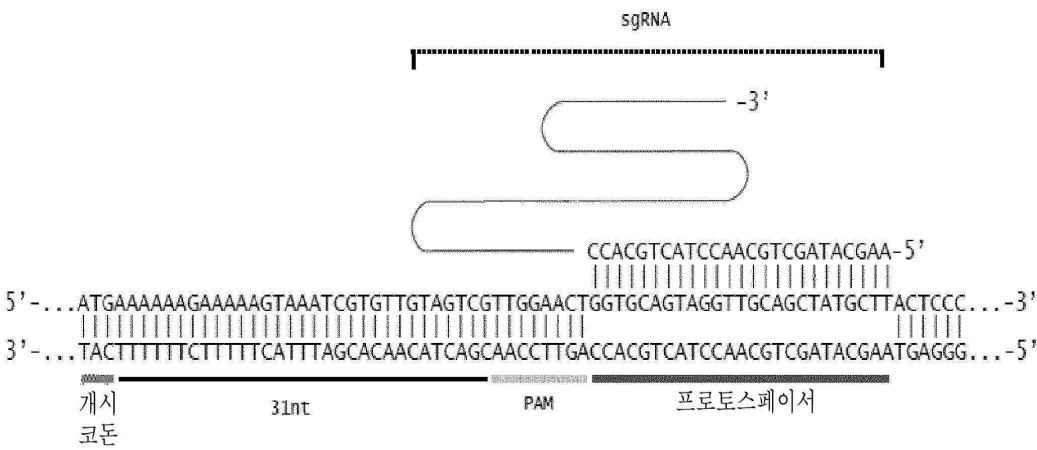
도면23a



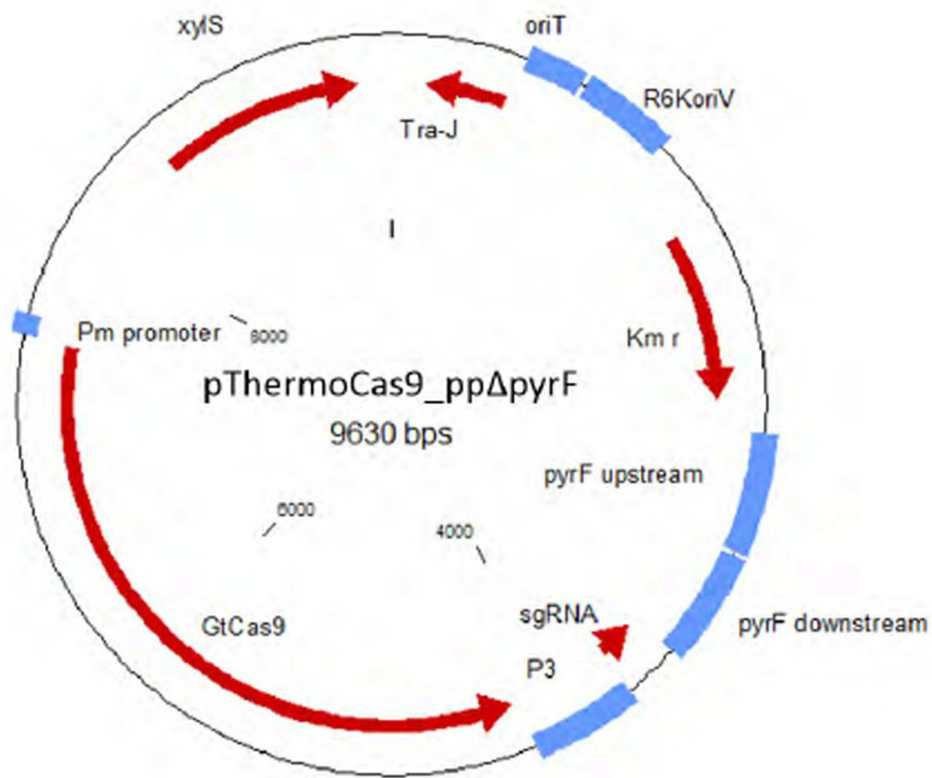
도면23b



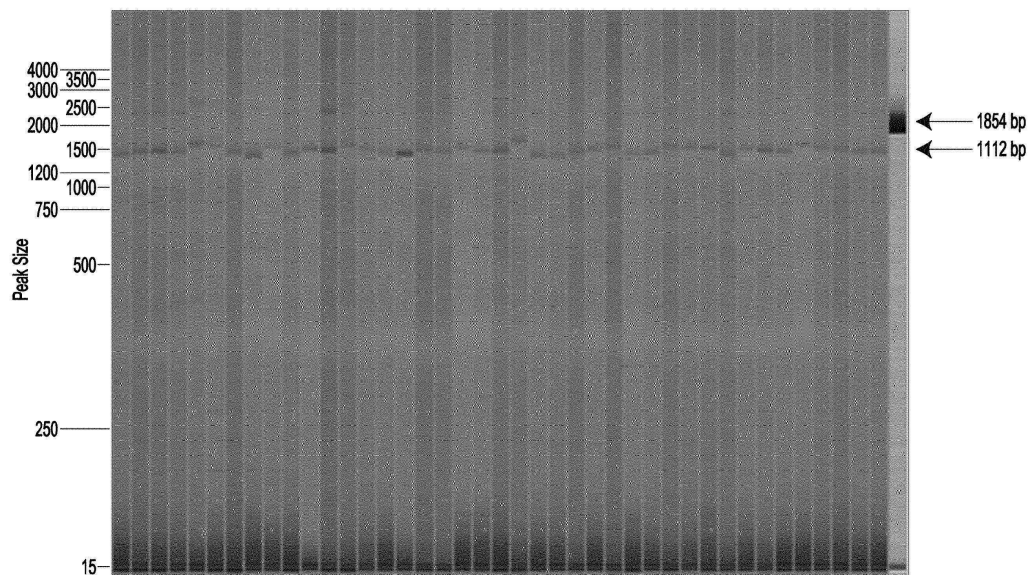
도면24



도면25



도면26



서열 목록

SEQUENCE LISTING

<110> Wageningen Universiteit & Stichting voor de Technische Wetenschappen

<120> THERMOSTABLE CAS9 NUCLEASES

<130> P220294W03

<150> PCT/EP2016/081077

<151> 2016-12-14

<160> 170

<170> PatentIn version 3.5

<210> 1

<211> 1082

<212> PRT

<213> Geobacillus thermodenitrificans T12

<400> 1

Met Lys Tyr Lys Ile Gly Leu Asp Ile Gly Ile Thr Ser Ile Gly Trp

1 5 10 15

Ala Val Ile Asn Leu Asp Ile Pro Arg Ile Glu Asp Leu Gly Val Arg

20 25 30

Ile Phe Asp Arg Ala Glu Asn Pro Lys Thr Gly Glu Ser Leu Ala Leu

35 40 45

Pro Arg Arg Leu Ala Arg Ser Ala Arg Arg Arg Leu Arg Arg Arg Lys

50 55 60

His Arg Leu Glu Arg Ile Arg Arg Leu Phe Val Arg Glu Gly Ile Leu

65 70 75 80

Thr Lys Glu Glu Leu Asn Lys Leu Phe Glu Lys Lys His Glu Ile Asp

85 90 95

Val Trp Gln Leu Arg Val Glu Ala Leu Asp Arg Lys Leu Asn Asn Asp

100 105 110

Glu Leu Ala Arg Ile Leu Leu His Leu Ala Lys Arg Arg Gly Phe Arg

115 120 125

Ser Asn Arg Lys Ser Glu Arg Thr Asn Lys Glu Asn Ser Thr Met Leu

130 135 140

Lys His Ile Glu Glu Asn Gln Ser Ile Leu Ser Ser Tyr Arg Thr Val

145 150 155 160

Ala Glu Met Val Val Lys Asp Pro Lys Phe Ser Leu His Lys Arg Asn

165 170 175
 Lys Glu Asp Asn Tyr Thr Asn Thr Val Ala Arg Asp Asp Leu Glu Arg
 180 185 190
 Glu Ile Lys Leu Ile Phe Ala Lys Gln Arg Glu Tyr Gly Asn Ile Val
 195 200 205
 Cys Thr Glu Ala Phe Glu His Glu Tyr Ile Ser Ile Trp Ala Ser Gln

 210 215 220
 Arg Pro Phe Ala Ser Lys Asp Asp Ile Glu Lys Lys Val Gly Phe Cys
 225 230 235 240
 Thr Phe Glu Pro Lys Glu Lys Arg Ala Pro Lys Ala Thr Tyr Thr Phe
 245 250 255
 Gln Ser Phe Thr Val Trp Glu His Ile Asn Lys Leu Arg Leu Val Ser
 260 265 270
 Pro Gly Gly Ile Arg Ala Leu Thr Asp Asp Glu Arg Arg Leu Ile Tyr

 275 280 285
 Lys Gln Ala Phe His Lys Asn Lys Ile Thr Phe His Asp Val Arg Thr
 290 295 300
 Leu Leu Asn Leu Pro Asp Asp Thr Arg Phe Lys Gly Leu Leu Tyr Asp
 305 310 315 320
 Arg Asn Thr Thr Leu Lys Glu Asn Glu Lys Val Arg Phe Leu Glu Leu
 325 330 335
 Gly Ala Tyr His Lys Ile Arg Lys Ala Ile Asp Ser Val Tyr Gly Lys

 340 345 350
 Gly Ala Ala Lys Ser Phe Arg Pro Ile Asp Phe Asp Thr Phe Gly Tyr
 355 360 365
 Ala Leu Thr Met Phe Lys Asp Asp Thr Asp Ile Arg Ser Tyr Leu Arg
 370 375 380
 Asn Glu Tyr Glu Gln Asn Gly Lys Arg Met Glu Asn Leu Ala Asp Lys
 385 390 395 400
 Val Tyr Asp Glu Glu Leu Ile Glu Glu Leu Leu Asn Leu Ser Phe Ser

 405 410 415

Lys Phe Gly His Leu Ser Leu Lys Ala Leu Arg Asn Ile Leu Pro Tyr
 420 425 430
 Met Glu Gln Gly Glu Val Tyr Ser Thr Ala Cys Glu Arg Ala Gly Tyr
 435 440 445
 Thr Phe Thr Gly Pro Lys Lys Lys Gln Lys Thr Val Leu Leu Pro Asn
 450 455 460
 Ile Pro Pro Ile Ala Asn Pro Val Val Met Arg Ala Leu Thr Gln Ala

 465 470 475 480
 Arg Lys Val Val Asn Ala Ile Ile Lys Lys Tyr Gly Ser Pro Val Ser
 485 490 495
 Ile His Ile Glu Leu Ala Arg Glu Leu Ser Gln Ser Phe Asp Glu Arg
 500 505 510
 Arg Lys Met Gln Lys Glu Gln Glu Gly Asn Arg Lys Lys Asn Glu Thr
 515 520 525
 Ala Ile Arg Gln Leu Val Glu Tyr Gly Leu Thr Leu Asn Pro Thr Gly

 530 535 540
 Leu Asp Ile Val Lys Phe Lys Leu Trp Ser Glu Gln Asn Gly Lys Cys
 545 550 555 560
 Ala Tyr Ser Leu Gln Pro Ile Glu Ile Glu Arg Leu Leu Glu Pro Gly
 565 570 575
 Tyr Thr Glu Val Asp His Val Ile Pro Tyr Ser Arg Ser Leu Asp Asp
 580 585 590
 Ser Tyr Thr Asn Lys Val Leu Val Leu Thr Lys Glu Asn Arg Glu Lys

 595 600 605
 Gly Asn Arg Thr Pro Ala Glu Tyr Leu Gly Leu Gly Ser Glu Arg Trp
 610 615 620
 Gln Gln Phe Glu Thr Phe Val Leu Thr Asn Lys Gln Phe Ser Lys Lys
 625 630 635 640
 Lys Arg Asp Arg Leu Leu Arg Leu His Tyr Asp Glu Asn Glu Glu Asn
 645 650 655
 Glu Phe Lys Asn Arg Asn Leu Asn Asp Thr Arg Tyr Ile Ser Arg Phe

660 665 670
 Leu Ala Asn Phe Ile Arg Glu His Leu Lys Phe Ala Asp Ser Asp Asp
 675 680 685
 Lys Gln Lys Val Tyr Thr Val Asn Gly Arg Ile Thr Ala His Leu Arg
 690 695 700
 Ser Arg Trp Asn Phe Asn Lys Asn Arg Glu Glu Ser Asn Leu His His
 705 710 715 720
 Ala Val Asp Ala Ala Ile Val Ala Cys Thr Thr Pro Ser Asp Ile Ala

 725 730 735
 Arg Val Thr Ala Phe Tyr Gln Arg Arg Glu Gln Asn Lys Glu Leu Ser
 740 745 750
 Lys Lys Thr Asp Pro Gln Phe Pro Gln Pro Trp Pro His Phe Ala Asp
 755 760 765
 Glu Leu Gln Ala Arg Leu Ser Lys Asn Pro Lys Glu Ser Ile Lys Ala
 770 775 780
 Leu Asn Leu Gly Asn Tyr Asp Asn Glu Lys Leu Glu Ser Leu Gln Pro

 785 790 795 800
 Val Phe Val Ser Arg Met Pro Lys Arg Ser Ile Thr Gly Ala Ala His
 805 810 815
 Gln Glu Thr Leu Arg Arg Tyr Ile Gly Ile Asp Glu Arg Ser Gly Lys
 820 825 830
 Ile Gln Thr Val Val Lys Lys Lys Leu Ser Glu Ile Gln Leu Asp Lys
 835 840 845
 Thr Gly His Phe Pro Met Tyr Gly Lys Glu Ser Asp Pro Arg Thr Tyr

 850 855 860
 Glu Ala Ile Arg Gln Arg Leu Leu Glu His Asn Asn Asp Pro Lys Lys
 865 870 875 880
 Ala Phe Gln Glu Pro Leu Tyr Lys Pro Lys Lys Asn Gly Glu Leu Gly
 885 890 895
 Pro Ile Ile Arg Thr Ile Lys Ile Ile Asp Thr Thr Asn Gln Val Ile
 900 905 910

Pro Leu Asn Asp Gly Lys Thr Val Ala Tyr Asn Ser Asn Ile Val Arg

915 920 925
Val Asp Val Phe Glu Lys Asp Gly Lys Tyr Tyr Cys Val Pro Ile Tyr

930 935 940
Thr Ile Asp Met Met Lys Gly Ile Leu Pro Asn Lys Ala Ile Glu Pro

945 950 955 960
Asn Lys Pro Tyr Ser Glu Trp Lys Glu Met Thr Glu Asp Tyr Thr Phe

965 970 975
Arg Phe Ser Leu Tyr Pro Asn Asp Leu Ile Arg Ile Glu Phe Pro Arg

980 985 990
Glu Lys Thr Ile Lys Thr Ala Val Gly Glu Glu Ile Lys Ile Lys Asp

995 1000 1005
Leu Phe Ala Tyr Tyr Gln Thr Ile Asp Ser Ser Asn Gly Gly Leu

1010 1015 1020
Ser Leu Val Ser His Asp Asn Asn Phe Ser Leu Arg Ser Ile Gly

1025 1030 1035
Ser Arg Thr Leu Lys Arg Phe Glu Lys Tyr Gln Val Asp Val Leu

1040 1045 1050
Gly Asn Ile Tyr Lys Val Arg Gly Glu Lys Arg Val Gly Val Ala

1055 1060 1065
Ser Ser Ser His Ser Lys Ala Gly Glu Thr Ile Arg Pro Leu

1070 1075 1080

<210> 2

<211> 8

<212> PRT

<213> Geobacillus thermodenitrificans T12

<400> 2

Glu Lys Asp Gly Lys Tyr Tyr Cys

1 5

<210> 3

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> amino acid motif from Cas9

<220><221> X

<222> (1)..(1)

<223> Isoleucine, Methionine or Proline

<220><221> X

<222> (2)..(2)

<223> Valine, Serine, Asparagine or Isoleucine

<220><221> X

<222> (5)..(5)

<223> Glutamate or Lysine

<220><221> X

<222> (6)..(6)

<223> Alanine, Glutamate or Arginine

<400> 3

Xaa Xaa Cys Thr Xaa Xaa

1 5

<210> 4

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> amino acid motif from G. thermodenitrificans T12 Cas9

<220><221> X

<222> (1)..(1)

<223> Methionine or Phenylalanine

<220><221> X

<222> (4)..(4)

<223> Histidine or Asparagine

<400> 4

Xaa Leu Lys Xaa Ile Glu

1 5

<210> 5

<211> 6
 <212> PRT
 <213> Artificial Sequence
 <220><223> amino acid motif from G. thermodenitrificans T12
 <220><221> X
 <222> (1)..(1)
 <223> Glutamate or Isoleucine
 <220><221> X
 <222> (5)..(5)
 <223> Tryptophan, Serine or Lysine
 <400> 5
 Xaa Val Tyr Ser Xaa Lys
 1 5

<210> 6
 <211> 12
 <212> PRT
 <213> Artificial sequence
 <220><223> amino acid motif from thermophilic Cas9
 <220><221> X
 <222> (1)..(1)
 <223> Alanine or Glutamate
 <220><221> X
 <222> (4)..(4)
 <223> Glutamine or Lysine
 <220><221> X
 <222> (5)..(5)
 <223> Arginine or Alanine
 <220><221> X
 <222> (9)..(9)
 <223> Asparagine or Alanine
 <220><221> X
 <222> (12)..(12)
 <223> Lysine or Serine
 <400> 6

Xaa Phe Tyr Xaa Xaa Arg Glu Gln Xaa Lys Glu Xaa

1 5 10

<210> 7

<211> 3249

<212> DNA

<213> Geobacillus thermodenitrificans T12

<400> 7

atgaagtata aaatcggctc tgatcgcgc attacgtcta tcggttgggc tgcattaat	60
ttggacattc ctgcacga agatttaggt gtccgcattt ttgacagagc ggaaaacccg	120
aaaaccgggg agtcactagc tcttcacgt cgctcgccc gctccgccg acgtcgtctg	180
cggcgtcgca aacatcgact ggagcgcat cgccgcctgt tcgtccgcga aggaatttta	240
acgaaggaag agctgaacaa gctgtttgaa aaaaagcacg aaatcgacgt ctggcagctt	300
cgtgttgaag cactggatcg aaaactaat aacgatgaat tagcccgcat ctttctcat	360
ctggctaaac ggctggatt tagatccaac cgcaagagtg agcgaccaa caaagaaaac	420
agtacgatgc tcaaacatat tgaagaaaac caatccattc tttcaagtta ccgaacggtt	480
gcagaaatgg ttgtcaagga tccgaaattt tccctgcaca agcgtataa agaggataat	540
tacaccaaca ctgttgcgc cgacgatctt gaacgggaaa tcaaaactgat tttcgccaaa	600
cagcgcaat atgggaacat cgtttgcaca gaagcatttg aacacgagta ttttccatt	660
tgggcatcgc aacgccctt tgccttaag gatgatatcg agaaaaagt cggtttctgt	720
acgtttgagc ctaaagaaaa acgcgcgcca aaagcaacat acacattcca gtccttcacc	780
gtctgggaac atattaacaa acttcgtctt gtctccccgg gaggcacccg ggcactaacc	840
gatgatgaac gtcgtcttat atacaagcaa gcatttcata aaaataaaat caccttccat	900
gatgttcgaa cattgcttaa cttgctgac gacaccggt ttaaaggtct tttatatgac	960
cgaacacca cgtgaagga aaatgagaaa gtctgcttcc ttgaactcgg cgcctatcat	1020
aaaatacggg aagcgatcga cagcgtctat ggcaaaggag cagcaaaatc atttcgtccg	1080
attgattttg atacatttgg ctacgatta acgatgttta aagacgacac cgacattcgc	1140
agttacttgc gaaacgaata cgaacaaaat ggaaaacgaa tggaaaatct agcggataaa	1200
gtctatgatg aagaattgat tgaagaactt ttaaacttat cgttttctaa gtttggtcat	1260
ctatccctta aagcgcttcg caacatcctt ccatatatgg aacaaggcga agtctactca	1320
accgcttgtg aacgagcagg atatacattt acagggccaa agaaaaaaca gaaaacggta	1380
ttgctgccga acattccgcc gatcgccaat ccggtcgtca tgcgcgcact gacacaggca	1440

cgcaaagtgg tcaatgccat tatcaaaaag tacggctcac cggctccat ccataatcgaa 1500
ctggcccgga aactatcaca atcctttgat gaacgacgta aaatgcagaa agaacaggaa 1560
ggaaaccgaa agaaaaacga aactgccatt cgccaacttg ttgaatatgg gctgacgctc 1620

aatccaactg ggcttgacat tgtgaaattc aaactatgga gcgaacaaaa cggaaaatgt 1680
gcctattcac tccaaccgat cgaaatcgag cggttgctcg aaccaggcta tacagaagtc 1740
gacatgtga ttccatacag ccgaagcttg gacgatagct ataccaataa agttcttctg 1800
ttgacaaagg agaaccgtga aaaaggaaac cgcacccag ctgaatattt aggattaggc 1860
tcagaacgtt ggcaacagtt cgagacgttt gtcttgacaa ataagcagtt ttcgaaaaag 1920
aagcgggacg gactccttcg gcttcattac gatgaaaacg aagaaaatga gtttaaaaat 1980
cgtaatctaa atgatacccg ttatatctca cgcttcttgg ctaactttat tcgcgaacat 2040

ctcaaattcg ccgacagcga tgacaacaaa aaagtataca cggccaacgg ccgtattacc 2100
gcccatttac gcagccgttg gaattttaac aaaaaccggg aagaatcgaa ttgcatcat 2160
gccgtcgatg ctgccatcgt cgcctgcaca acgccgagcg atatcgcccg agtcaccgcc 2220
ttctatcaac ggcggaaca aaacaaagaa ctgtccaaaa agacggatcc gcagtttccg 2280
cagccttggc cgcaatttgc tgatgaactg caggcgctgt tatcaaaaaa tccaaaggag 2340
agtataaaag ctctcaatct tggaaattat gataacgaga aactcgaatc gttgcagccg 2400
gtttttgtct cccgaatgcc gaagcggagc ataacaggag cggctcatca agaaacattg 2460

cggcgttata tcggcatcga cgaacggagc ggaaaaatac agacggctcg caaaaagaaa 2520
ctatccgaga tccaactgga taaaacaggt catttcccaa tgtacgggaa agaaagcgat 2580
ccaaggacat atgaagccat tcgccaacgg ttgcttgaac ataacaatga cccaaaaaag 2640
gcgtttcaag agcctctgta taaaccgaag aagaacggag aactaggtcc tatcatccga 2700
acaatcaaaa tcatcgatc gacaaatcaa gttattccgc tcaacgatgg caaaacagtc 2760
gcctacaaca gcaacatcgt gcgggtcgac gtctttgaga aagatggcaa atattattgt 2820
gtccctatct atacaataga tatgatgaaa gggatcttgc caaacaagc gatcgagccg 2880

aacaaaccgt actctgagtg gaaggaaatg acggaggact atacattccg attcagtcta 2940
tacccaaatg atcttatccg tatcgaattt ccccgagaaa aaacaataaa gactgctgtg 3000
ggggaagaaa tcaaaattaa ggatctgttc gcctattatc aaaccatcga ctctccaat 3060
ggagggttaa gtttggttag ccatgataac aacttttcgc tccgcagcat cggttcaaga 3120
accctcaaac gattcgagaa ataccaagta gatgtgctag gcaacatcta caaagtgaga 3180
ggggaaaaga gagtgggggt ggcgtcatct tctcattcga aagccgggga aactatccgt 3240

ccgttataa

3249

<210> 8

<211> 1045

<212> PRT

<213> Actinomyces naeslundii

<400> 8

Met Trp Tyr Ala Ser Leu Met Ser Ala His His Leu Arg Val Gly Ile

1 5 10 15

Asp Val Gly Thr His Ser Val Gly Leu Ala Thr Leu Arg Val Asp Asp

20 25 30

His Gly Thr Pro Ile Glu Leu Leu Ser Ala Leu Ser His Ile His Asp

35 40 45

Ser Gly Val Gly Lys Glu Gly Lys Lys Asp His Asp Thr Arg Lys Lys

50 55 60

Leu Ser Gly Ile Ala Arg Arg Ala Arg Arg Leu Leu His His Arg Arg

65 70 75 80

Thr Gln Leu Gln Gln Leu Asp Glu Val Leu Arg Asp Leu Gly Phe Pro

85 90 95

Ile Pro Thr Pro Gly Glu Phe Leu Asp Leu Asn Glu Gln Thr Asp Pro

100 105 110

Tyr Arg Val Trp Arg Val Arg Ala Arg Leu Val Glu Glu Lys Leu Pro

115 120 125

Glu Glu Leu Arg Gly Pro Ala Ile Ser Met Ala Val Arg His Ile Ala

130 135 140

Arg His Arg Gly Trp Arg Asn Pro Tyr Ser Lys Val Glu Ser Leu Leu

145 150 155 160

Ser Pro Ala Asn Ala Asn Glu Ile Arg Lys Ile Cys Ala Arg Gln Gly

165 170 175

Val Ser Pro Asp Val Cys Lys Gln Leu Leu Arg Ala Val Phe Lys Ala

180 185 190

Asp Ser Pro Arg Gly Ser Ala Val Ser Arg Val Ala Pro Asp Pro Leu

195	200	205	
Pro Gly Gln Gly Ser Phe Arg Arg Ala Pro Lys Cys Asp Pro Glu Phe			
210	215	220	
Gln Arg Phe Arg Ile Ile Ser Ile Val Ala Asn Leu Arg Ile Ser Glu			
225	230	235	240
Thr Lys Gly Glu Asn Arg Pro Leu Thr Ala Asp Glu Arg Arg His Val			
	245	250	255
Val Thr Phe Leu Thr Glu Asp Ser Gln Ala Asp Leu Thr Trp Val Asp			
260	265	270	
Val Ala Glu Lys Leu Gly Val His Arg Arg Asp Leu Arg Gly Thr Ala			
275	280	285	
Val His Thr Asp Asp Gly Glu Arg Ser Ala Ala Arg Pro Pro Ile Asp			
290	295	300	
Ala Thr Asp Arg Ile Met Arg Gln Thr Lys Ile Ser Ser Leu Lys Thr			
305	310	315	320
Trp Trp Glu Glu Ala Asp Ser Glu Gln Arg Gly Ala Met Ile Arg Tyr			
325	330	335	
Leu Tyr Glu Asp Pro Thr Asp Ser Glu Cys Ala Glu Ile Ile Ala Glu			
340	345	350	
Leu Pro Glu Glu Asp Gln Ala Lys Leu Asp Ser Leu His Leu Pro Ala			
355	360	365	
Gly Arg Ala Ala Tyr Ser Arg Glu Ser Leu Thr Ala Leu Ser Asp His			
370	375	380	
Met Leu Ala Thr Thr Asp Asp Leu His Glu Ala Arg Lys Arg Leu Phe			
385	390	395	400
Gly Val Asp Asp Ser Trp Ala Pro Pro Ala Glu Ala Ile Asn Ala Pro			
405	410	415	
Val Gly Asn Pro Ser Val Asp Arg Thr Leu Lys Ile Val Gly Arg Tyr			
420	425	430	
Leu Ser Ala Val Glu Ser Met Trp Gly Thr Pro Glu Val Ile His Val			
435	440	445	

Glu His Val Arg Asp Gly Phe Thr Ser Glu Arg Met Ala Asp Glu Arg
 450 455 460
 Asp Lys Ala Asn Arg Arg Arg Tyr Asn Asp Asn Gln Glu Ala Met Lys
 465 470 475 480
 Lys Ile Gln Arg Asp Tyr Gly Lys Glu Gly Tyr Ile Ser Arg Gly Asp
 485 490 495
 Ile Val Arg Leu Asp Ala Leu Glu Leu Gln Gly Cys Ala Cys Leu Tyr
 500 505 510
 Cys Gly Thr Thr Ile Gly Tyr His Thr Cys Gln Leu Asp His Ile Val
 515 520 525
 Pro Gln Ala Gly Pro Gly Ser Asn Asn Arg Arg Gly Asn Leu Val Ala
 530 535 540
 Val Cys Glu Arg Cys Asn Arg Ser Lys Ser Asn Thr Pro Phe Ala Val
 545 550 555 560
 Trp Ala Gln Lys Cys Gly Ile Pro His Val Gly Val Lys Glu Ala Ile
 565 570 575
 Gly Arg Val Arg Gly Trp Arg Lys Gln Thr Pro Asn Thr Ser Ser Glu
 580 585 590
 Asp Leu Thr Arg Leu Lys Lys Glu Val Ile Ala Arg Leu Arg Arg Thr
 595 600 605
 Gln Glu Asp Pro Glu Ile Asp Glu Arg Ser Met Glu Ser Val Ala Trp
 610 615 620
 Met Ala Asn Glu Leu His His Arg Ile Ala Ala Ala Tyr Pro Glu Thr
 625 630 635 640
 Thr Val Met Val Tyr Arg Gly Ser Ile Thr Ala Ala Ala Arg Lys Ala
 645 650 655
 Ala Gly Ile Asp Ser Arg Ile Asn Leu Ile Gly Glu Lys Gly Arg Lys
 660 665 670
 Asp Arg Ile Asp Arg Arg His His Ala Val Asp Ala Ser Val Val Ala
 675 680 685
 Leu Met Glu Ala Ser Val Ala Lys Thr Leu Ala Glu Arg Ser Ser Leu

690 695 700
 Arg Gly Glu Gln Arg Leu Thr Gly Lys Glu Gln Thr Trp Lys Gln Tyr
 705 710 715 720
 Thr Gly Ser Thr Val Gly Ala Arg Glu His Phe Glu Met Trp Arg Gly
 725 730 735
 His Met Leu His Leu Thr Glu Leu Phe Asn Glu Arg Leu Ala Glu Asp
 740 745 750
 Lys Val Tyr Val Thr Gln Asn Ile Arg Leu Arg Leu Ser Asp Gly Asn

 755 760 765
 Ala His Thr Val Asn Pro Ser Lys Leu Val Ser His Arg Leu Gly Asp
 770 775 780
 Gly Leu Thr Val Gln Gln Ile Asp Arg Ala Cys Thr Pro Ala Leu Trp
 785 790 795 800
 Cys Ala Leu Thr Arg Glu Lys Asp Phe Asp Glu Lys Asn Gly Leu Pro
 805 810 815
 Ala Arg Glu Asp Arg Ala Ile Arg Val His Gly His Glu Ile Lys Ser

 820 825 830
 Ser Asp Tyr Ile Gln Val Phe Ser Lys Arg Lys Lys Thr Asp Ser Asp
 835 840 845
 Arg Asp Glu Thr Pro Phe Gly Ala Ile Ala Val Arg Gly Gly Phe Val
 850 855 860
 Glu Ile Gly Pro Ser Ile His His Ala Arg Ile Tyr Arg Val Glu Gly
 865 870 875 880
 Lys Lys Pro Val Tyr Ala Met Leu Arg Val Phe Thr His Asp Leu Leu

 885 890 895
 Ser Gln Arg His Gly Asp Leu Phe Ser Ala Val Ile Pro Pro Gln Ser
 900 905 910
 Ile Ser Met Arg Cys Ala Glu Pro Lys Leu Arg Lys Ala Ile Thr Thr
 915 920 925
 Gly Asn Ala Thr Tyr Leu Gly Trp Val Val Val Gly Asp Glu Leu Glu
 930 935 940

Ile Asn Val Asp Ser Phe Thr Lys Tyr Ala Ile Gly Arg Phe Leu Glu

945 950 955 960

Asp Phe Pro Asn Thr Thr Arg Trp Arg Ile Cys Gly Tyr Asp Thr Asn

965 970 975

Ser Lys Leu Thr Leu Lys Pro Ile Val Leu Ala Ala Glu Gly Leu Glu

980 985 990

Asn Pro Ser Ser Ala Val Asn Glu Ile Val Glu Leu Lys Gly Trp Arg

995 1000 1005

Val Ala Ile Asn Val Leu Thr Lys Val His Pro Thr Val Val Arg

1010 1015 1020

Arg Asp Ala Leu Gly Arg Pro Arg Tyr Ser Ser Arg Ser Asn Leu

1025 1030 1035

Pro Thr Ser Trp Thr Ile Glu

1040 1045

<210> 9

<211> 1160

<212> PRT

<213> Streptococcus pyogenes

<400> 9

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
 100 105 110
 His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
 115 120 125
 His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
 130 135 140
 Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

 145 150 155 160
 Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
 165 170 175
 Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Arg
 180 185 190
 Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
 195 200 205
 Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

 210 215 220
 Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
 225 230 235 240
 Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
 245 250 255
 Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
 260 265 270
 Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

 275 280 285
 Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
 290 295 300
 Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
 305 310 315 320
 Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
 325 330 335
 Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

340 345 350
 Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
 355 360 365
 Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
 370 375 380
 Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
 385 390 395 400
 Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

 405 410 415
 Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
 420 425 430
 His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
 435 440 445
 Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
 450 455 460
 Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

 465 470 475 480
 Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
 485 490 495
 Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
 500 505 510
 His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
 515 520 525
 Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

 530 535 540
 Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
 545 550 555 560
 Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
 565 570 575
 Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
 580 585 590

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

595 600 605

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

610 615 620

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

625 630 635 640

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

645 650 655

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

660 665 670

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

675 680 685

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

690 695 700

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

705 710 715 720

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

725 730 735

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

740 745 750

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

755 760 765

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

770 775 780

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

785 790 795 800

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys

805 810 815

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser

820 825 830

Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu

835	840	845	
Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile			
850	855	860	
Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser			
865	870	875	880
Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly			
885	890	895	
Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile			
900	905	910	
Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser			
915	920	925	
Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly			
930	935	940	
Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile			
945	950	955	960
Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala			
965	970	975	
Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys			
980	985	990	
Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser			
995	1000	1005	
Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys			
1010	1015	1020	
Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys			
1025	1030	1035	
Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln			
1040	1045	1050	
His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe			
1055	1060	1065	
Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu			
1070	1075	1080	

Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala
 1085 1090 1095

Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro

1100 1105 1110
 Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr

1115 1120 1125

Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser

1130 1135 1140

Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly

1145 1150 1155

Gly Asp

1160

<210> 10

<211> 8

<212> DNA

<213> Artificial sequence

<220><223> PAM

<220><221> misc_feature

<222> (7)..(7)

<223> n is a, c, g, or t

<400> 10

ccccccna 8

<210> 11

<211> 8

<212> DNA

<213> Artificial Sequence

<220><223> PAM

<400> 11

ccccccaa 8

<210> 12

<211> 5

<212> DNA

<213> Artificial Sequence

<220><223> PAM	
<400> 12	
cccccc	5
<210> 13	
<211> 8	
<212> DNA	
<213> Artificial Sequence	
<220><223> PAM	
<220><221> misc_feature	
<222> (6)..(7)	
<223> n is a, c, g, or t	
<400> 13	
ccccccna	8
<210> 14	
<211> 6	
<212> DNA	
<213> Artificial Sequence	
<220><223> PAM	
<400> 14	
ccccccc	6
<210> 15	
<211> 6	
<212> DNA	
<213> Artificial Sequence	
<220><223> PAM	
<220><221> misc_feature	
<222> (1)..(1)	
<223> n is a, c, g, or t	
<400> 15	
ncccccc	6
<210> 16	
<211> 8	
<212> DNA	

<213> Artificial Sequence
 <220><223> PAM
 <220><221> misc_feature
 <222> (1)..(1)
 <223> n is a, c, g, or t
 <220><221> misc_feature
 <222> (7)..(7)
 <223> n is a, c, g, or t
 <400> 16
 nccccna
 <210> 17
 <211> 8
 <212> DNA
 <213> Artificial Sequence

 <220><223> PAM
 <220><221> misc_feature
 <222> (2)..(2)
 <223> n is a, c, g, or t
 <400> 17
 cnccccac
 <210> 18
 <211> 8
 <212> DNA
 <213> Artificial Sequence
 <220><223> PAM
 <400> 18
 cccccag
 <210> 19
 <211> 8
 <212> DNA
 <213> Artificial Sequence
 <220><223> PAM
 <400> 19

8

8

8

ccccccat 8

<210> 20

<211> 8

<212> DNA

<213> Artificial Sequence

<220><223> PAM

<400> 20

ccccccac 8

<210> 21

<211> 8

<212> DNA

<213> Artificial Sequence

<220><223> PAM

<400> 21

atccccaa 8

<210> 22

<211> 8

<212> DNA

<213> Artificial Sequence

<220><223> PAM

<400> 22

acggccaa 8

<210> 23

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> RuvC-1 domain motif

<400> 23

Ile Gly Leu Asp Ile Gly Ile Thr Ser Ile Gly

1 5 10

<210> 24

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> RuvC-I domain motif

<400> 24

Ile Gly Leu Asp Ile Gly Ile Thr Ser Ile Gly Trp Ala Val Ile Asn

1 5 10 15

Leu Asp

<210> 25

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Bridge domain motif

<400> 25

Arg Ser Ala Arg Arg

1 5

<210> 26

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Bridge domain motif

<400> 26

Pro Arg Arg Leu Ala Arg Ser Ala Arg Arg Arg Leu Arg Arg Arg Lys

1 5 10 15

His Arg Leu Glu Arg Ile Arg Arg Leu

20 25

<210> 27

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> alpha-helical/recognition lobe domain motif

<400> 27

Trp Gln Leu Arg

1

<210> 28

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> alpha-helical/recognition lobe domain

<400> 28

His Leu Ala Lys Arg Arg Gly

1 5

<210> 29

<211> 13

<212> PRT

<213> Artificial Sequence

<220><223> alpha-helical/recognition lobe domain

<400> 29

Leu Ala Arg Ile Leu Leu His Leu Ala Lys Arg Arg Gly

1 5 10

<210> 30

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> alpha-helical/recognition lobe domain

<400> 30

Ile Phe Ala Lys Gln

1 5

<210> 31

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> alpha-helical/recognition lobe domain

<400> 31

Glu Ile Lys Leu Ile Phe Ala Lys Gln

1 5

<210> 32

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> alpha-helical/recognition lobe domain

<400> 32

Ile Trp Ala Ser Gln Arg

1 5

<210> 33

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> alpha-helical/recognition lobe domain motif

<400> 33

Lys Val Gly Phe Cys Thr Phe Glu Pro Lys Glu Lys Arg Ala Pro Lys

1 5 10 15

<210> 34

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> alpha-helical/recognition lobe domain motif

<400> 34

Phe Thr Val Trp Glu His Ile Asn Lys Leu Arg Leu

1 5 10

<210> 35

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> RuvC-II domain motif

<400> 35

Ile Ala Asn Pro Val Val Met Arg Ala Leu Thr Gln

1 5 10

<210> 36

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> RuvC-II domain motif

<400> 36

Ile Ala Asn Pro Val Val Met Arg Ala Leu Thr Gln Ala Arg Lys Val

1 5 10 15

Val Asn Ala Ile Ile Lys Lys Tyr Gly

20 25

<210> 37

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> RuvC-II domain motif

<400> 37

Glu Leu Ala Arg

1

<210> 38

<211> 8

<212> PRT

<213> Artificial Sequence

<220><223> RuvC-II domain motif

<400> 38

Ile His Ile Glu Leu Ala Arg Glu

1 5

<210> 39

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> HNH domain motif

<400> 39

Gln Asn Gly Lys Cys Ala Tyr

1 5

<210> 40

<211

> 16

<212> PRT

<213> Artificial Sequence

<220><223> HNH domain motif

<400> 40

Ile Val Lys Phe Lys Leu Trp Ser Glu Gln Asn Gly Lys Cys Ala Tyr

1 5 10 15

<210> 41

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> HNH domain motif

<400> 41

Val Asp His Val Ile Pro

1 5

<210> 42

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> HNH domain motif

<400> 42

Val Asp His Val Ile Pro Tyr Ser Arg Ser Leu Asp Asp Ser Tyr Thr

1 5 10 15

Asn Lys Val Leu

20

<210> 43

<211> 11

<212> PRT

<213> Artificial Sequence

<220><223> RuvC-III domain motif

<400> 43

Asp Thr Arg Tyr Ile Ser Arg Phe Leu Ala Asn

1 5 10

<210> 44

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> RuvC-III domain motif

<400> 44

Val Tyr Thr Val Asn Gly Arg Ile Thr Ala His Leu Arg Ser Arg Trp

1 5 10 15

<210> 45

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> RuvC-III domain motif

<400> 45

His His Ala Val Asp Ala

1 5

<210> 46

<211> 10

<212> PRT

<213> Artificial Sequence

<220><223> RuvC-III domain motif

<400> 46

His His Ala Val Asp Ala Ala Ile Val Ala

1 5 10

<210> 47

<211> 8

<212> DNA

<213> Artificial Sequence

<220><223> PAM

<220><221> misc_feature

<222> (1)..(4)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (6)..(7)

<223> n is a, c, g, or t

<400> 47
 nnnncnna 8
 <210> 48
 <211> 8
 <212> DNA
 <213> Artificial Sequence
 <220><223> PAM
 <220><221> misc_feature
 <222> (1)..(4)
 <223> n is a, c, g, or t
 <400> 48
 nnnncvaa 8
 <210> 49
 <211> 8
 <212> DNA
 <213> Artificial Sequence
 <220><223> PAM
 <220><221> misc_feature
 <222> (1)..(4)
 <223> n is a, c, g, or t
 <400> 49
 nnnncsaa 8
 <210> 50
 <211> 8
 <212> DNA
 <213> Artificial Sequence
 <220><223> PAM
 <220><221> misc_feature
 <222> (1)..(4)
 <223> n is a, c, g, or t
 <400> 50
 nnnncgaa 8
 <210> 51
 <211> 8

<212> DNA
 <213> Artificial Sequence
 <220><223> PAM
 <220><221> misc_feature
 <222> (1)..(4)

<223>

n is a, c, g, or t

<400> 51

nnnnccaa

8

<210> 52

<211> 8

<212> DNA

<213> Artificial Sequence

<220><223> PAM

<220><221> misc_feature

<222> (1)..(4)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (6)..(6)

<223> n is a, c, g, or t

<400> 52

nnnncnaa

8

<210> 53

<211> 8

<212> DNA

<213> Artificial Sequence

<220><223> PAM

<220><221> misc_feature

<222> (1)..(4)

<223> n is a, c, g, or t

<400> 53

nnnncmca

8

<210> 54

<211> 30

<212> RNA

<213> Artificial Sequence

<220><223> tracrRNA 5' hairpin

<400> 54

aagggcuuuc ugccuauagg cagacugccc 30

<210> 55

<211> 25

<212> RNA

<213> Artificial Sequence

<220><223> tracrRNA 'middle' hairpin

<400> 55

guggcguugg ggaucgccua ucgcc 25

<210> 56

<211> 36

<212> RNA

<213> Artificial Sequence

<220><223> tracrRNA 3' hairpin

<400> 56

cgcuuucuuuc gggcAUUccc cAcucuuagg cguuuu 36

<210> 57

<211> 55

<212> RNA

<213> Artificial Sequence

<220><223> tracrRNA 5' hairpin and middle hairpin

<400> 57

aagggcuuuc ugccuauagg cagacugccc guggcguugg ggaucgccua ucgcc 55

<210> 58

<211> 91

<212> RNA

<213> Artificial Sequence

<220><223> tracrRNA 5' hairpin, middle hairpin and 3' hairpin.

<400> 58

aagggcuuuc ugccuauagg cagacugccc guggcguugg ggaucgccua ucgccgcuu 60

ucuucgggca uuccccacuc uuaggcguuu u 91

<210> 59

<211> 81

<212> DNA

<213> Artificial Sequence

<220><223> BG6494

<220><221> misc_feature

<222> (35)..(41)

<223> n is a, c, g, or t

<400> 59

tatgcctcat gagattatca aaaaggatct tcacnnnnnn nctagatcct tttaaattaa 60

aatgaagtt ttaaataaat c 81

<210> 60

<211> 81

<212> DNA

<213> Artificial Sequence

<220><223> BG6495

<400> 60

tatgccggat cctcagacca agtttactca tatatacttt agattgattt aaaacttcat 60

ttttaattta aaaggatcta g 81

<210> 61

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> BG7356

<400> 61

tcgtcggcag cgtcagatgt gtataagaga cagt 34

<210> 62

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> BG7357

<400> 62

ctgtctcttta tacacatctg acgctgccga cga	33
<210> 63	
<211> 16	
<212> DNA	
<213> Artificial Sequence	
<220><223> BG7358	
<400> 63	
tcgtcggcag cgtcag	16
<210> 64	
<211> 53	
<212> DNA	
<213> Artificial Sequence	
<220><223> BG7359	
<400> 64	
gtctcgtggg ctcgagatg tgtataagag acaggacat gattacgcca agc	53
<210> 65	
<211> 61	
<212> DNA	
<213> Artificial Sequence	
<220><223> BG7616	
<400> 65	
tcgtcggcag cgtcagatgt gtataagaga cagggtcatg agattatcaa aaaggatctt	60
c	61
<210> 66	
<211> 82	
<212> DNA	
<213> Artificial Sequence	
<220><223> BG8157	
<400> 66	
tatgcctcat gagattatca aaaaggatct tcaccccccc agctagatcc ttttaaatta	60
aaaatgaagt tttaaatcaa tc	82
<210> 67	
<211> 82	

<212> DNA

<213> Artificial Sequence

<220><223> BG8158

<400> 67

tatgcctcat gagattatca aaaaggatct tcaccccccc aactagatcc ttttaaatta 60

aaaatgaagt tttaaatcaa tc 82

<210> 68

<211> 82

<212> DNA

<213> Artificial Sequence

<220><223> BG8159

<400> 68

tatgcctcat gagattatca aaaaggatct tcaccccccc atctagatcc ttttaaatta 60

aaaatgaagt tttaaatcaa tc 82

<210> 69

<211> 82

<212> DNA

<213> Artificial Sequence

<220><223> BG8160

<400> 69

tatgcctcat gagattatca aaaaggatct tcaccccccc acctagatcc ttttaaatta 60

aaaatgaagt tttaaatcaa tc 82

<210> 70

<211> 81

<212> DNA

<213> Artificial Sequence

<220><223> BG8161

<220><221> misc_feature

<222> (35)..(38)

<223> n is a, c, g, or t

<220><221> misc_feature

<222> (40)..(41)

<223> n is a, c, g, or t

<400> 70
tatgcctcat gagattatca aaaaggatct tcacnnmntn nctagatcct tttaaattaa 60
aaatgaagtt ttaaatacaat c 81
<210> 71
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> BG8363
<400> 71
acggttatcc acagaatcag 20
<210> 72
<211> 22
<212> DNA
<213> Artificial Sequence
<220><223> BG8364
<400> 72
cgggattgac ttttaaaaaa gg 22
<210> 73
<211> 82
<212> DNA
<213> Artificial Sequence
<220><223> BG8763
<400> 73
tatgcctcat gagattatca aaaaggatct tcacccccca aactagatcc ttttaaatta 60
aaaatgaagt tttaatcaa tc 82
<210> 74
<211> 82
<212> DNA
<213> Artificial Sequence
<220><223> BG8764
<400> 74
tatgcctcat gagattatca aaaaggatct tcacccccca tactagatcc ttttaaatta 60
aaaatgaagt tttaatcaa tc 82

<210> 75
 <211> 82
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG8765
 <400> 75
 tatgcctcat gagattatca aaaaggatct tcacccccca gactagatcc ttttaaatta 60
 aaaatgaagt tttaaatcaa tc 82
 <210> 76
 <211> 82
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG8766
 <400> 76
 tatgcctcat gagattatca aaaaggatct tcacccccca cactagatcc ttttaaatta 60
 aaaatgaagt tttaaatcaa tc 82
 <210> 77
 <211> 82
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG8767
 <400> 77
 tatgcctcat gagattatca aaaaggatct tcacccccct aactagatcc ttttaaatta 60
 aaaatgaagt tttaaatcaa tc 82
 <210> 78
 <211> 82
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG8768
 <400> 78
 tatgcctcat gagattatca aaaaggatct tcacccccct tactagatcc ttttaaatta 60
 aaaatgaagt tttaaatcaa tc 82
 <210> 79

<211> 82

<212> DNA

<213> Artificial Sequence

<220><223> BG8769

<400> 79

tatgcctcat gagattatca aaaaggatct tcacccccct gactagatcc ttttaaatta 60

aaaatgaagt tttaaatcaa tc 82

<210> 80

<211> 82

<212> DNA

<213> Artificial Sequence

<220><223> BG8770

<400> 80

tatgcctcat gagattatca aaaaggatct tcacccccct cactagatcc ttttaaatta 60

aaaatgaagt tttaaatcaa tc 82

<210> 81

<211> 82

<212> DNA

<213> Artificial Sequence

<220><223> BG8771

<400> 81

tatgcctcat gagattatca aaaaggatct tcaccccccg aactagatcc ttttaaatta 60

aaaatgaagt tttaaatcaa tc 82

<210> 82

<211> 82

<212> DNA

<213> Artificial Sequence

<220><223> BG8772

<400> 82

tatgcctcat gagattatca aaaaggatct tcaccccccg tactagatcc ttttaaatta 60

aaaatgaagt tttaaatcaa tc 82

<210> 83

<211> 82

<212> DNA
 <213> Artificial Sequence
 <220><223> BG8773
 <400> 83
 tatgcctcat gagattatca aaaaggatct tcacccccg gactagatcc ttttaaatta 60
 aaaatgaagt tttaaatcaa tc 82
 <210> 84
 <211> 82
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG8774
 <400> 84
 tatgcctcat gagattatca aaaaggatct tcacccccg cactagatcc ttttaaatta 60
 aaaatgaagt tttaaatcaa tc 82
 <210> 85
 <211> 82
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG8775
 <400> 85
 tatgcctcat gagattatca aaaaggatct tcaccccccc aactagatcc ttttaaatta 60
 aaaatgaagt tttaaatcaa tc 82
 <210> 86
 <211> 82
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG8776
 <400> 86
 tatgcctcat gagattatca aaaaggatct tcaccccccc tactagatcc ttttaaatta 60
 aaaatgaagt tttaaatcaa tc 82
 <210> 87
 <211> 82
 <212> DNA

<213> Artificial Sequence

<220><223> BG8777

<400> 87

tatgcctcat gagattatca aaaaggatct tcacccccc gactagatcc ttttaaatta 60

aaaatgaagt tttaaatcaa tc 82

<210> 88

<211> 82

<212> DNA

<213> Artificial Sequence

<220><223> BG8778

<400> 88

tatgcctcat gagattatca aaaaggatct tcacccccc cactagatcc ttttaaatta 60

aaaatgaagt tttaaatcaa tc 82

<210> 89

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> BG6574

<400> 89

aagcttgaaa taatacgact cactatagg 29

<210> 90

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> BG6576

<400> 90

aaaaaagacc ttgacgtttt cc 22

<210> 91

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> BG9307

<400> 91

aagcttgaaa taatacgact cactatagg gagattatca aaaaggatct tcacgtc 57

<210> 92
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220><223> BG9309
 <400> 92
 aaaacgccta agagtgggga atg 23
 <210> 93
 <211> 19
 <212> DNA
 <213> Artificial Sequence

<220><223> BG9310
 <400> 93
 aaaagcgcat agcgcatcc 19
 <210> 94
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220><223> BG9311
 <400> 94
 aaaacgggtc agtctgccta tag 23
 <210> 95
 <211> 57
 <212> DNA
 <213> Artificial Sequence

<220><223> BG9308
 <400> 95
 aagcttgaaa taatacgact cactataggt gagattatca aaaaggatct tcacgtc 57
 <210> 96
 <211> 56
 <212> DNA
 <213> Artificial Sequence

<220><223> BG10118

<400> 96
aagcttgaaa taatacgact cactatagga gattatcaaa aaggatcttc acgtca 56
<210> 97
<211> 60
<212> DNA
<213> Artificial Sequence
<220><223> BG10119
<400> 97
aagcttgaaa taatacgact cactatagga agattatcaa aaaggatctt cacgtcatag 60
<210> 98
<211> 58
<212> DNA
<213> Artificial Sequence
<220><223> BG10120
<400> 98
aagcttgaaa taatacgact cactatagga ttatcaaaaa ggatcttcac gtcatagt 58
<210> 99
<211> 60
<212> DNA
<213> Artificial Sequence
<220><223> BG10121
<400> 99
aagcttgaaa taatacgact cactatagga attatcaaaa aggatcttca cgtcatagtt 60
<210> 100
<211> 58
<212> DNA
<213> Artificial Sequence
<220><223> BG10122
<400> 100
aagcttgaaa taatacgact cactataggt tatcaaaaag gatcttcacg tcatagtt 58
<210> 101
<211> 58
<212> DNA

<213> Artificial Sequence

<220><223> BG10123

<400> 101

aagcttgaaa taatacgact cactataggt atcaaaaagg atcttcacgt catagtgc 58

<210> 102

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> BG10124

<400> 102

aagcttgaaa taatacgact cactatagga tcaaaaagga tcttcacgtc atagtgc 57

<210> 103

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> BG9312

<400> 103

aaaacgccta agagtgggga atgcccgaag aaagcgggcg ataggcgatc c 51

<210> 104

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> BG8191

<400> 104

aagcttggcg taatcatggt c 21

<210> 105

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> BG8192

<400> 105

tcatgagttc ccatgttgtg 20

<210> 106

<211> 48
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG8194
 <400> 106
 tatggcgaat cacaacatgg gaactcatga gaacatcctc tttcttag 48
 <210> 107
 <211> 50
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG8195
 <400> 107
 gccgatatca agaccgattt tatacttcat ttaagttacc tcctcgattg 50
 <210> 108
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG8196
 <400> 108
 atgaagtata aaatcggtct tg 22
 <210> 109
 <211> 18
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG8197
 <400> 109
 taacggacgg atagtttc 18
 <210> 110
 <211> 55
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG8198
 <400> 110

gaaagccggg gaaactatcc gtccgttata aatcagacaa aatggcctgc ttatg 55

<210> 111

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> BG8263

<400> 111

gaactatgac actttatgtt cagaatggac gtataacggt atccatttta agaataatcc 60

<210> 112

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> BG8268

<400> 112

accgttatac gtccattctg aaaataaagt gtcatagttc ccctgagat 49

<210> 113

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> BG8210

<400> 113

aacagctatg accatgatta cgccaagctt cctcccatg cacaatag 48

<210> 114

<211> 60

<212> DNA

<213> Artificial Sequence

<220><223> BG8261

<400> 114

gaactatgac atcatggagt tttaaatacca gtataacggt atccatttta agaataatcc 60

<210> 115

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> BG8266

<400> 115
accgttatac tggatttaaa actccatgat gtcatagttc ccctgagat 49
<210> 116
<211> 60
<212> DNA
<213> Artificial Sequence

<220><223> BG8317
<400> 116
gaactatgac caccagctt acatcaacaa gtataacggt atccatttta agaataatcc 60
<210> 117
<211> 49
<212> DNA
<213> Artificial Sequence

<220><223> BG8320
<400> 117
accgttatac ttgttgatgt aagctgggtg gtcatagttc ccctgagat 49
<210> 118
<211> 20
<212> DNA
<213> Artificial Sequence

<220><223> BG9075
<400> 118
ctatcggcat tacgtctatc 20
<210> 119
<211> 19
<212> DNA
<213> Artificial Sequence

<220><223> BG9091
<400> 119
gcgtcgactt ctgtatagc 19
<210> 120
<211> 41
<212> DNA

<213> Artificial Sequence

<220><223> BG9091

<400> 120

tgaagtataa aatcgggtctt gctatcggca ttacgtctat c 41

<210> 121

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> BG9092

<400> 121

caagcttcgg ctgtatggaa tcacagcgtc gacttctgta tagc 44

<210> 122

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> BG9077

<400> 122

gctgtgattc catacag 17

<210> 123

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> BG9267

<400> 123

ggtgcagtag gttgcagcta tgcttgata acggtatcca t 41

<210> 124

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> BG9263

<400> 124

aagcatagct gcaacctact gcaccgtcat agttccctg agattatcg 49

<210> 125

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> BG9088

<400> 125

tcatgaccaa aatcccttaa cg 22

<210> 126

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> BG9089

<400> 126

ttaagggatt ttggtcatga gaacatcctc tttcttag 38

<210> 127

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> BG9090

<400> 127

gcaagaccga ttttatactt catttaag 28

<210> 128

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> BG9548

<400> 128

ggatcccatg acgctagtat ccagctgggt catagttccc ctgagattat cg 52

<210> 129

<211> 63

<212> DNA

<213> Artificial Sequence

<220><223> BG9601

<400> 129

ttcaatattt tttttgaata aaaaatacga tacaataaaa atgtctagaa aaagataaaa 60

atg 63

<210> 130

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> BG9600

<400> 130

ttttttattc aaaaaaata ttgaatttta aaaatgatgg tgctagtatg aag 53

<210> 131

<211> 58

<212> DNA

<213> Artificial Sequence

<220><223> BG9549

<400> 131

ccagctggat actagcgtca tgggatccgt ataacggtat ccattttaag aataatcc 58

<210> 132

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> BG8552

<400> 132

tcgggggttc gtttccttg 20

<210> 133

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> BG8553

<400> 133

cttacacagc cagtgcgga ac 22

<210> 134

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> BG2365

<400> 134

gccggcggtcc cggaaaacga 20

<210> 135

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> BG2366

<400> 135

gcaggtcggg ttctctgcat ccatgcccc gaact 35

<210> 136

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> BG2367

<400> 136

ggcttcggaa tcgttttccg ggacgccgc acggcattgg caaggccaag 50

<210> 137

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> BG2368

<400> 137

gacacaggca tcggtgcagg gtctcttggc aagtc 35

<210> 138

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> BG2369

<400> 138

gccaagagac cctgcaccga tgctgtgtc gaacc 35

<210> 139

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> BG2370

<400> 139

cttggcggaa aacgtcaagg tcttttttac acgcgcatca acttcaaggc 50

<210> 140

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> BG2371

<400> 140

atgacgagct gttcaccagc agcgcctatta ttgaagcatt tatcaggg 48

<210> 141

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> BG2372

<400> 141

gtaaaaaaga ccttgacgtt ttc 23

<210> 142

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> BG2373

<400> 142

tatgaagcgg gccatttgaa gacgaaaggg cctc 34

<210> 143

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> BG2374

<400> 143

taatagcgct gctggtgaac agctcgcat agtccctg agattatcg 49

<210> 144

<211> 37
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG2375
 <400> 144
 tggagtcatg aacatatgaa gtataaaatc ggtcttg 37
 <210> 145
 <211> 35
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG2376
 <400> 145
 ccctttcgtc ttcaaatggc ccgcttcata agcag 35
 <210> 146
 <211> 39
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG2377
 <400> 146
 gattttatac ttcatatgtt catgactcca ttattattg 39
 <210> 147
 <211> 35
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG2378
 <400> 147
 gggggcatgg atgcgaggaa cccgacctgc attgg 35
 <210> 148
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG2381
 <400> 148

acacggcgga tgcattacc	20
<210> 149	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> BG2382	
<400> 149	
tggacgtgta cttcgacaac	20
<210> 150	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> BG2135	
<400> 150	
acacggcgga tgcattacc	20
<210> 151	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> BG8196	
<400> 151	
tggacgtgta cttcgacaac	20
<210> 152	
<211> 18	
<212> DNA	
<213> Artificial Sequence	
<220><223> BG8197	
<400> 152	
taacggacgg atagtttc	18
<210> 153	
<211> 31	
<212> DNA	
<213> Artificial Sequence	

<220><223> BG6850
 <400> 153
 gcctcatgaa tgcagcgatg gtccggtgtt c 31
 <210> 154
 <211> 27
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG6849
 <400> 154
 gcctcatgag ttcccatgtt gtgattc 27

 <210> 155
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG6769
 <400> 155
 caatccaact gggcttgac 19
 <210> 156
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG6841
 <400> 156
 caagaacttt attggtatag 20
 <210> 157
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG6840
 <400> 157
 ttgcagaaat ggttgtcaag 20

 <210> 158
 <211> 19

<212> DNA
 <213> Artificial Sequence
 <220><223> BG9215
 <400> 158
 gagataatgc cgactgtac 19
 <210> 159
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG9216
 <400> 159
 agggctcgcc tttgggaag 19
 <210> 160
 <211> 17
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG9505
 <400> 160
 gttgccaacg ttctgag 17

 <210> 161
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG9506
 <400> 161
 aatccacgcc gtttag 16
 <210> 162
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG8363
 <400> 162
 acggttatcc acagaatcag 20
 <210> 163

<211> 22
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG8364
 <400> 163
 cgggattgac ttttaaaaaa gg 22

<210> 164
 <211> 120
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG9302
 <400> 164
 aaacttcatt ttaatttaa aaggatctag aacccccgt gaagatcctt ttgataatc 60
 tcatgaccaa aatcccttaa cgtgagtttt cgttccactg agcgtcagac cccgtagaaa 120

<210> 165
 <211> 120
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG9303
 <400> 165
 tttctacggg gtctgacgct cagtggaaacg aaaactcacg ttaagggatt ttggatcatga 60
 gattatcaaa aaggatcttc acccccccaa ctagatcctt ttaaattaaa aatgaagttt 120

<
 210> 166
 <211> 120
 <212> DNA
 <213> Artificial Sequence
 <220><223> BG9304
 <400> 166
 tttctacggg gtctgacgct cagtggaaacg aaaactcacg ttaagggatt ttggatcatga 60
 gattatcaaa aaggatcttc acgggggggtt ctagatcctt ttaaattaaa aatgaagttt 120

<210> 167
 <211> 43
 <212> DNA

<213> Artificial Sequence

<220><223> BG7886

<400> 167

tacttccaat ccaatgcaaa gtataaaatc ggtcttgata tcg 43

<210> 168

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> BG7887

<400> 168

ttatccactt ccaatgttat tataacggac ggatagtttc cccggctttc 50

<210> 169

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> BG9665

<400> 169

atgacgaaag gagtttctta ttatg 25

<210> 170

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> BG9666

<400> 170

aacggtattc cgtgattaag 20