



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0145324
(43) 공개일자 2022년10월28일

- (51) 국제특허분류(Int. Cl.)
C12N 9/22 (2006.01) C12N 15/10 (2017.01)
C12N 15/113 (2010.01)
- (52) CPC특허분류
C12N 9/22 (2013.01)
C12N 15/102 (2013.01)
- (21) 출원번호 10-2022-7019785
- (22) 출원일자(국제) 2020년07월02일
심사청구일자 없음
- (85) 번역문제출일자 2022년06월10일
- (86) 국제출원번호 PCT/RU2020/050145
- (87) 국제공개번호 WO 2021/096391
국제공개일자 2021년05월20일
- (30) 우선권주장
2019136164 2019년11월11일 러시아(RU)

- (71) 출원인
조인트 스타크 컴퍼니 "바이오펜드"
러시아 198515 상트페테르부르크 인트라시티 유니
시펠리티 더 세틀먼트 오브 스텔레나 더 세틀먼트
오브 스텔레나 유엘. 스비아지 스트리트 1 디. 38
포메슈. 89
- (72) 발명자
세베리노프 콘스탄틴 빅토로비치
러시아 188301 모스크바 케이브이. 368 유엘. 드
미트리야 울리아노바 4/2
쉬마코프 세르게이 아나톨레비치
러시아 140200 보스크레첸스크 케이브이. 15. 유
엘. 피오너스카야 14 모스크프스카야 오비엘.
(뒷면에 계속)
- (74) 대리인
유미특허법인

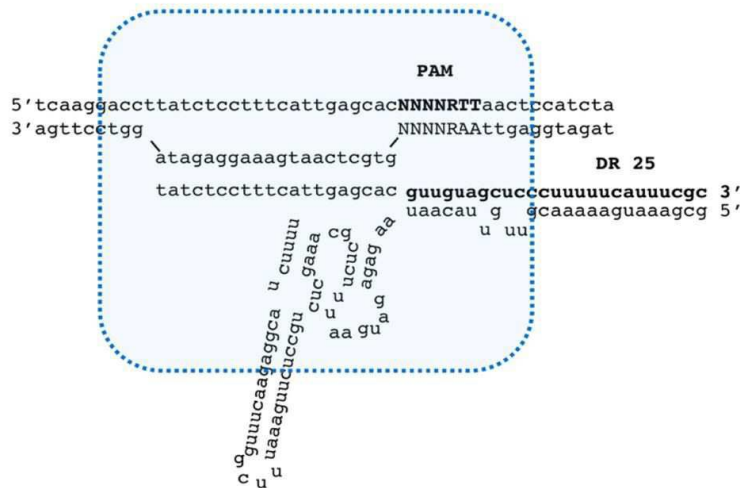
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 세균 파스퇴렐라 뉴모트로피카 유래 CAS9 단백질의 용도

(57) 요약

본 발명은 세균 *P. 뉴모트로피카(pneumotropica)* 유래 CRISPR-Cas9 시스템의 신규한 세균 뉴클레아제, 및 DNA 분자 내 정확히 특정 이중 가닥 절단을 형성하기 위한 그의 용도에 관한 것이다. 상기 뉴클레아제는 특이한 특성을 가지고, 단세포 생물 또는 다세포 생물의 세포 내 게놈 DNA 서열을 변형하기 위한 도구로서 사용될 수 있다. 따라서, 이용 가능한 CRISPR-Cas9 시스템의 다양성이 증가되며, 이러한 사실은 많은 특정 부위에서 및/또는 다양한 조건 하에, 다양한 생물 내 게놈 또는 플라스미드 DNA의 절단을 위한 Cas9 뉴클레아제의 다양한 변이의 사용을 가능케 할 것이다.

대표도 - 도11



(52) CPC특허분류

C12N 15/113 (2013.01)

C12Y 301/00 (2013.01)

C12N 2310/20 (2017.05)

(72) 발명자

아르타모노바 다리아 니콜레브나

러시아 121359 모스크바 9-59. 유엘. 마르샬라 티 모셴코

고리야닌 이그나티 이고레비치

러시아 117587 모스크바 케이브이. 351. 제다나 1 유엘. 키로보그라드스카야 16

무샤로바 올가 세르게이브나

러시아 123098 모스크바 케이브이. 289. 에스티알. 2 유엘. 로고바 7

안드레바 줄리아 발레레브나

러시아 142784 모스크바 유엘. 모스크비티나 3-2-108 모스코브스키

지유브코 타니아나 이고레브나

러시아 197374 상트-페테르부르크 에스티알. 1 유엘. 야테나야 3

페도로바 이아나 비탈레브나

러시아 188301 가트치나 14 유엘. 쿠프리나 40 레 닌그라드스카야 오비엘.

코도르코브스키 미카일 알렉세이비치

러시아 195426 상트-페테르부르크 케이브이. 268 케이. 1 디. 23 피알. 코시기나

포베갈로프 게오르기 에브게네비치

러시아 198205 상트-페테르부르크 케이브이. 317 디. 12 유엘. 파르티자나 게르마나

아르세니예프 아나톨리 니콜라예비치

러시아 195112 상트-페테르부르크 케이브이. 36 케이. 1 피알. 노보체르카스키 25

셀코바 폴리나 아나톨레브나

러시아 427439 붓킨스크 케이브이. 72 돔 15 유엘. 베르크나야 레스. 우드무르츠키야

마실리에바 알렉산드라 안드레브나

러시아 198334 상트-페테르부르크 케이브이. 144 디. 41 유엘. 파르티자나 게르마나

아르타모노바 타티아나 올레고브나

러시아 199406 상트-페테르부르크 케이브이. 7 디. 27 유엘. 날리치나야

아브라모바 마리나 빅토르브나

러시아 194021 상트-페테르부르크 코르페스 2 디. 8 피알. 네포코르예니

명세서

청구범위

청구항 1

DNA 분자 내에서, 상기 DNA 분자 내 뉴클레오티드 서열 5'-NNNN(A/G)TT-3' 바로 앞에 위치하는 이중 가닥 절단(double-strand break)을 형성하기 위한, 서열번호 1의 아미노산 서열을 포함하거나 또는 서열번호 1의 아미노산 서열과 적어도 95% 동일하고 서열번호 1과 비-보존 아미노산 잔기에서만 다른 아미노산 서열을 포함하는 단백질의 용도.

청구항 2

제1항에 있어서,

상기 DNA 분자 내 이중 가닥 절단은 35 내지 45°C의 온도에서 형성되는 것을 특징으로 하는, 용도.

청구항 3

제1항에 있어서,

상기 단백질은 서열번호 1의 아미노산 서열을 포함하는 것인, 용도.

청구항 4

제1항에 있어서,

상기 DNA 분자 내 이중 가닥 절단은 포유동물 세포의 게놈 DNA 내에서 형성되는 것을 특징으로 하는, 용도.

청구항 5

제4항에 있어서,

상기 DNA 분자 내 이중 가닥 절단은 상기 포유동물 세포의 게놈 DNA의 변형을 초래하는 것을 특징으로 하는, 용도.

청구항 6

게놈 DNA를 포함하는 단세포 또는 다세포 생물의 세포 내 게놈 DNA 서열을 변형하는 방법으로서, 상기 방법은 상기 생물의 세포 내로, 효과적인 양의: a) 서열번호 1의 아미노산 서열을 포함하는 단백질, 또는 상기 서열번호 1의 아미노산 서열을 포함하는 단백질을 암호화하는 핵산, 및 b) 뉴클레오티드 서열 5'-NNNN(A/G)TT-3'과 직접 인접하고 이중나선(duplex) 형성 후 상기 단백질과 상호 작용하는, 생물의 게놈 DNA 영역의 뉴클레오티드 서열과 이중 나선을 형성하는 서열을 포함하는 가이드 RNA, 또는 상기 가이드 RNA를 암호화하는 DNA 서열을 도입하는 단계를 포함하고;

상기 단백질과 가이드 RNA 및 뉴클레오티드 서열 5'-NNNN(A/G)TT-3'의 상호 작용은 상기 서열 5'-NNNN(A/G)TT-3'에 바로 인접하는 게놈 DNA 서열 내 이중 가닥 절단의 형성을 초래하는, 방법.

청구항 7

제6항에 있어서,

상기 가이드 RNA와 동시에 외래(exogenous) DNA 을 도입하는 단계를 더 포함하는, 방법.

청구항 8

제6항에 있어서,

상기 세포는 포유동물 세포인, 방법.

발명의 설명

기술 분야

[0001] 본 발명은 생명 공학, 구체적으로 DNA 절단 및 다양한 생물의 게놈 편집(editing)에 사용되는 신규한 효소인 CRISPR-Cas 시스템의 Cas 뉴클레아제에 관한 것이다. 이러한 기술은 장차 인간 유전성 질병의 유전자 치료 및 다른 생물의 게놈 편집을 위하여 사용될 수 있다.

배경 기술

[0002] DNA 서열 변형(modification)은 현재 생명 공학 분야에서 화제의 문제 중 하나이다. 시험관 내 DNA 조작뿐 아니라, 진핵 및 원핵 생물의 게놈의 편집 및 변형 또한 DNA 서열 내 이중 가닥 절단(double-strand breaks)의 표적화된 도입을 필요로 한다.

[0003] 이러한 문제를 해결하기 위하여, 다음 기술이 현재 사용되고 있다: 징크 핑거(zinc finger) 타입의 도메인을 함유하는 인공 뉴클레아제 시스템, TALEN 시스템, 및 세균 CRISPR-Cas 시스템. 첫번째 두 기술은 특정 DNA 서열의 인식을 위하여 뉴클레아제 아미노산 서열의 힘든 최적화를 필요로 한다. 이와 대조적으로, CRISPR-Cas 시스템에 관한 한, DNA 표적을 인식하는 구조가 문제가 아니라, 짧은 가이드 RNAs가 문제이다. 특정 DNA 표적의 절단은 뉴클레아제 또는 그 유전자의 드노보(de novo) 합성을 요하는 것이 아니라, 표적 서열에 상보적인 사이드 RNAs를 사용하는 방식으로 이루어진다. 이는 CRISPR Cas 시스템이 다양한 DNA 서열 절단을 위한 편리하고 효율적인 수단이 되게 한다. 이러한 기술은 상이한 서열의 가이드 RNAs를 사용하여 몇몇 영역에서 DNA의 동시 절단을 허용한다. 이러한 접근은 또한 진핵 생물 내 몇몇 유전자를 동시에 변형하는데도 사용된다.

[0004] 그들의 본질상, CRISPR-Cas 시스템은 절단(breaks)을 바이러스 유전 물질 내로 매우 특이적으로 도입할 수 있는 원핵 면역 시스템이다 (Mojica F. J. M. et al. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements //Journal of molecular evolution. - 2005. - Vol. 60. - Issue 2. -pp. 174-182). 약어 CRISPR-Cas는 "Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR associated Genes"을 의미한다 (Jansen R. et al. Identification of genes that are associated with DNA repeats in prokaryotes //Molecular microbiology. - 2002. - Vol. 43. - Issue 6. - pp. 1565-1575). 모든 CRISPR-Cas 시스템은 CRISPR 카세트 및 다양한 Cas 단백질을 암호화하는 유전자로 구성된다 (Jansen R. et al., Molecular microbiology. - 2002. - Vol. 43. - Issue 6. - pp. 1565-1575). CRISPR 카세트는 각각 고유의 뉴클레오티드 서열을 가지는 스페이서(spacers) 및 반복적인 palindromic repeats로 구성된다 (Jansen R. et al., Molecular microbiology. - 2002. - Vol. 43. - Issue 6. - pp. 1565-1575). CRISPR 카세트의 전사에 이은 그 프로세싱은 가이드 crRNAs의 형성을 초래하며, 이는 Cas 단백질과 함께 이펙터 복합체(effector complex)를 형성한다 (Brouns S. J. J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes //Science. - 2008. - Vol. 321. - Issue 5891. - pp. 960-964). crRNA와 프로토스페이서(protospacer)라 불리는 표적 DNA 부위 간의 상보적 페어링으로 인하여, Cas 뉴클레아제는 DNA 표적을 인식하고 그 안에 절단을 매우 특이적으로 도입한다.

[0005] 단일 이펙터 단백질을 가지는 CRISPR-Cas 시스템은 그 시스템 내 포함되는 Cas 단백질에 따라 6개의 상이한 타입으로 분류된다 (타입 I-VI). 2013년에, 인간 세포의 게놈 DNA 편집을 위하여 타입 II CRISPR-Cas9 시스템을 사용하는 것이 최초로 제안되었다 (Cong L, et al., Multiplex genome engineering using CRISPR/Cas systems. Science. 2013 Feb 15;339(6121):819-23). 타입 II CRISPR-Cas9 시스템은 단순한 조성 및 활성 메커니즘을 특징으로 한다, 즉 그 작용은 하나의 Cas9 단백질 및 crRNA 및 트래이서 RNA(tracrRNA)의 두 개의 짧은 RNAs로만 구성되는 이펙터 복합체의 형성을 필요로 한다. 상기 트래이서 RNA는 CRISPR 반복으로부터 유래되는 crRNA 영역과 상보적으로 페어링하여, 가이드 RNAs의 Cas 이펙터에 결합에 필요한 2차 구조를 형성한다. 가이드 RNAs 서열 결정은 이전에 연구되지 않은 Cas 상동 유전자(orthologues)의 규명에 중요한 단계이다. Cas 이펙터 단백질은, 표적 DNA의 상보 가닥에 절단을 도입하여, 이중 가닥 DNA 절단을 생산하는, 두 개의 뉴클레아제 도메인(HNH 및 RuvC)을 가지는 RNA-의존적 DNA 엔도뉴클레아제이다 (Deltcheva E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III //Nature. - 2011. - Volume 471. - Issue 7340. - p. 602).

[0006] 따라서, 지금까지, 몇몇 CRISPR-Cas 뉴클레아제가 이중 가닥 절단을 DNA 내로 표적 및 특이적으로 도입할 수 있는 것으로 알려졌다. 이러한 CRISPR-Cas9 기술은, 시험관내 적용 또한 제공하고, 세균 균주에서 인간 세포에 걸치는, 다양한 생물의 DNA 내 절단을 도입하기 위한 가장 최신의 신속히 발전하고 있는 기술 중 하나이다 (Song

M. The CRISPR/Cas9 system: Their delivery, in vivo and ex vivo applications and clinical development by startups. *Biotechnol Prog.* 2017 Jul;33(4):1035-1045).

[0007] 상기 Cas9 및 crRNA/tracrRNA 이중가닥(duplex)으로 구성되는 이펙터 리보핵산 복합체는, crRNA 스페이서-프로토스페이서 상보성 이외에, DNA의 인식 및 그 후의 가수분해를 위하여 DNA 표적 상에 PAM(protospacer adjusted motif)의 존재를 필요로 한다 (Mojica F. J.M. et al. 2009). PAM은 오프-타겟 사슬 상에 프로토스페이서의 3' 말단에 인접하거나 몇개의 뉴클레오티드 떨어진 타입 II 시스템 내에 위치한 몇개의 뉴클레오티드의 엄격히 정의된 서열이다. PAM의 부재 하에, DNA 결합의 가수분해 후 이중 가닥 절단의 형성은 일어나지 않는다. 표적 상에 PAM 서열의 존재에 대한 요구는 인식 특이성을 증가시키나, 이와 동시에 절단을 도입하기 위한 표적 DNA 영역의 선택에 대하여 제한을 부과한다. 따라서, DNA 표적을 3'-말단으로부터 플랭킹하는 원하는 PAM 서열의 존재는 DNA 부위에서 CRISPR-Cas 시스템의 사용을 제한하는 특징이다.

[0008] 상이한 CRISPR-Cas 단백질은 그의 활성을 위하여 상이한 고유의 PAM 서열을 사용한다. 신규한 다양한 PAM 서열을 가지는 CRISPR-Cas 단백질의 사용은 생물의 시험관내 및 게놈 내, DNA 영역의 변형을 가능케 하기 위하여 필요하다. 진핵 생물 게놈의 변형 또한 CRISPR-Cas 시스템의 세포 내로 AAV-개체 전달을 제공하기 위하여 작은 크기의 뉴클레아제의 사용을 필요로 한다.

[0009] DNA 절단 및 게놈 DNA 서열 변형을 위한 많은 기술이 공지되어 있으나, 다양한 생물 내에서 DNA 서열의 정확히 특정 부위에서 DNA를 변형하기 위한 신규한 효과적인 수단이 여전히 요구된다.

발명의 내용

해결하려는 과제

[0010] 본 발명의 목적은 CRISPR-Cas9 시스템을 사용하여 단세포 또는 다세포 생물의 게놈 DNA 서열을 변형하기 위한 신규한 수단을 제공하는 것이다. 현재 존재하는 시스템은 변형될 DNA 영역의 3' 말단에 존재하여야 하는 특정 PAM 서열로 인하여 사용이 제한된다. 다른 PAM 서열을 가지는 신규한 Cas9 효소에 대한 서치는 다양한 생물의 DNA 분자 내 원하는, 정확히 특정 부위에서 이중 가닥 절단의 형성을 위한 이용 가능한 수단 범위를 확대시킬 것이다. 이러한 문제를 해결하기 위하여, 본 발명자들은 파스퇴렐라 뉴모트로피카(*Pasteurella pneumotropica* (*P. pneumotropica*))에 대하여 이전에 예측된 타입 II CRISPR 뉴클레아제 PpCas9가 상기 및 다른 생물 모두의 게놈 내로 지시된 변형을 도입하는데 사용될 수 있음을 규명하였다. 본 발명은 다음의 필수적인 특징을 가진다: (a) 다른 공지된 PAM 서열들과 다른 짧은 PAM 서열; (b) 1055 아미노산 잔기(a.a.r)의, 비교적 작은 크기의 규명된 PpCas9 단백질.

과제의 해결 수단

[0011] 상기 문제점은, DNA 분자 내에서, 상기 DNA 분자 내 뉴클레오티드 서열 5'-NNNN(A/G)TT-3' 바로 앞에 위치하는 이중 가닥 절단(double-strand break)을 형성하기 위한, 서열번호 1의 아미노산 서열을 포함하거나, 또는 서열번호 1의 아미노산 서열과 적어도 95% 동일하고 서열번호 1과 비-보존 아미노산 잔기에서만 다른 아미노산 서열을 포함하는 단백질의 용도에 의하여 해결된다. 일부 구현예에서, 상기 용도는 상기 DNA 분자 내 이중 가닥 절단이 35 내지 45°C의 온도에서 형성됨을 특징으로 한다. 본 발명의 일부 구현예에서, 상기 용도는 상기 이중 가닥 절단이 포유동물 세포의 게놈 DNA 내에서 형성됨을 특징으로 한다. 본 발명의 일부 구현예에서, 상기 용도는 상기 DNA 분자 내 이중 가닥 절단의 형성이 상기 포유동물 세포의 게놈 DNA의 변형을 초래함을 특징으로 한다.

[0012] 상기 문제점은, 단세포 또는 다세포 생물의 세포 내 게놈 DNA 서열을 변형하는(modifying) 방법으로서, 상기 방법은 상기 생물의 세포 내로, 효과적인 양의: a) 서열번호 1의 아미노산 서열을 포함하는 단백질, 또는 상기 서열번호 1의 아미노산 서열을 포함하는 단백질을 암호화하는 핵산, 및 b) 뉴클레오티드 서열 5'-NNNN(A/G)TT-3'과 직접 인접하고 이중나선(duplex) 형성 후 상기 단백질과 상호 작용하는, 생물의 게놈 DNA 영역의 뉴클레오티드 서열과 이중나선을 형성하는 서열을 포함하는 가이드 RNA, 또는 상기 가이드 RNA를 암호화하는 DNA 서열을 도입하는 단계를 포함하고; 상기 단백질과 가이드 RNA 및 뉴클레오티드 서열 5'-NNNN(A/G)TT-3'의 상호 작용은 상기 서열 5'-NNNN(A/G)TT-3'에 바로 인접하는 게놈 DNA 서열 내 이중 가닥 절단의 형성을 초래하는, 방법에 의하여 더욱 해결된다.

[0013] 본 발명의 일부 구현예에서, 상기 방법은 상기 가이드 RNA와 동시에 외래(exogenous) DNA의 도입을 더 포함하는 것을 특징으로 한다.

[0014] 표적 DNA 영역 및 PpCas9 단백질과 복합체를 형성할 수 있는, crRNA 및 트레이서 RNA (tracrRNA)의 혼합물을 가이드 RNA로서 사용할 수 있다. 본 발명의 바람직한 구현예에서, crRNA 및 트레이서 RNA를 기초로 제작되는 하이브리드 RNA를 가이드 RNA로서 사용할 수 있다. 하이브리드 가이드 RNA의 제작 방법은 당업자에게 공지되어 있다 (Hsu PD, et al., DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol. 2013 Sep;31(9):827-32). 하이브리드 RNA 제작을 위한 접근법 중 한가지가 후술하는 실시예에 개시된다.

[0015] 본 발명은 시험관내 표적 DNA 절단을 위하여, 및 일부 생물의 게놈 변형을 위하여 모두 사용될 수 있다. 상기 게놈 DNA는 직접적인 방식으로, 즉 게놈 DNA를 해당 부위에서 절단하고, 외래 DNA 서열을 상동 복구(homologous repair)를 통하여 삽입함으로써, 변형될 수 있다.

[0016] 투여를 위하여 사용되는 것 이외의 생물의 게놈으로부터의 이중 가닥 또는 단일 가닥 DNA의 임의의 영역 (또는 그러한 영역과 다른 DNA 단편과의 구성)을 외래 DNA 서열로서 사용할 수 있으며, 상기 영역(또는 영역들의 구성)은 PpCas9 뉴클레아제에 의하여 유도되는 표적 DNA 내 이중 가닥 절단 위치 내로 통합이 의도된다. 본 발명의 일부 구현예에서, 하나 이상의 뉴클레오티드의 삽입 또는 결실에 의해서뿐 아니라, 변이(뉴클레오티드 치환)에 의하여 더욱 변형된, PpCas9 단백질의 도입을 위하여 사용되는 생물의 게놈 DNA로부터의 이중 가닥 DNA 영역을 외래 DNA 서열로서 사용할 수 있다.

발명의 효과

[0017] 본 발명의 기술적 결과는 더 많은 수의 특정 부위 및 특정 조건 하에 게놈 또는 플라스미드 DNA를 절단하기 위한 Cas9 뉴클레아제의 사용을 가능케 하기 위하여, 이용 가능한 CRISPR-Cas9 시스템의 다양성을 증가시키는 것이다. 상기 신규한 뉴클레아제는 세균, 포유동물 또는 기타 생물의 세포 내에 사용될 수 있다.

도면의 간단한 설명

[0018] 도 1. CRISPR PpCas9 시스템의 유전자좌(locus)에 대한 계획. DR(direct repeat)은 CRISPR 카세트의 일부인 규칙적으로 반복되는 영역이다.

도 2. 시험관내 PAM 스크리닝. 실험 계획

도 3. 상이한 반응 온도에서 7N 라이브러리 단편의 PpCas9 뉴클레아제 절단

도 4. (a) 각각의 PAM (FC) 위치에서 각각의 특정 뉴클레오티드에 대한 비율 변화 로그의 계산을 이용한 시험관내 PpCas9 뉴클레아제 스크리닝 결과의 분석. (b) PpCas9 뉴클레아제의 PAM Logo. 아데닌, 시토신, 티민 및 구아닌 존재를 각각의 위치에 대하여 나타낸다. 문자의 높이는 PAM 서열의 주어진 위치에서 뉴클레오티드 존재에 상응한다.

도 5. PpCas9 뉴클레아제에 의한 DNA 표적 절단의 효율성에 대한 PAM 포지션 1에서 단일 뉴클레오티드 치환의 영향의 검증

도 6. PpCas9 PAM 서열 내 뉴클레오티드 위치의 중요성의 검증

도 7. PpCas9 뉴클레아제에 의한 DNA 표적의 절단의 효율성에 대한 PAM 포지션 5에서 A 내지 G 치환의 영향의 검증.

도 8. PpCas9 뉴클레아제에 의한 DNA 표적의 절단의 효율성에 대한 PAM 포지션 7에서 단일 뉴클레오티드 치환의 영향의 검증.

도 9. PpCas9 단백질을 사용하는 다양한 DNA 부위 절단. 레인 1 및 2는 양성 대조군이다.

도 10. PpCas9 뉴클레아제에 의한 PAM 서열 CAGCATT의 인식의 검증. 레인 1 및 2는 양성 대조군이다.

도 11. DNA 절단 도구 PpCas9의 도면

도 12. DNA 표적의 절단에 대한 실험. 상이한 길이의 하이브리드 가이드 RNAs를 사용하였다.

도 13. NCBI BLASTp 소프트웨어를 사용하는 (디폴트 파라미터) 스태필로코커스 아우레우스(*Staphylococcus aureus*)로부터의 PpCas9 및 Cas9 단백질의 아미노산 서열 정렬.

도 14. PpCas9를 사용하는 인간 세포의 게놈 DNA의 변형. (a)는 PpCas9 함유 플라스미드를 사용하는 인간 세포의 게놈 DNA 변형의 효율성을 결정하기 위한 실험 계획이다. (b)는 인간 세포의 게놈 DNA의 표적 부위의 서열

내로 뉴클레오티드의 삽입 및 결실의 분석 결과이다 (상부 - T7 엔도뉴클레아제 I과의 반응 생성물을 아가로스 겔 전기영동 상에 적용, 하부 - 대용량 시퀀싱(high throughput sequencing)에 의하여 결정된 EMX1 유전자 내 PpCas9에 의하여 형성되는 삽입 및 결실의 예).

발명을 실시하기 위한 구체적인 내용

- [0019] 본원 명세서에 사용되는 용어 "포함하다" 및 "포함하는"은 "다른 것들 중에서 포함하다"를 의미하는 것으로 해석될 것이다. 상기 용어는 "~만으로 구성되는"으로 해석되는 것을 의도하지 않는다. 달리 정의하지 않으면, 본원에서 기술적 및 과학적 용어들은 과학 및 기술 문헌에서 일반적으로 받아들여지는 전형적인 의미를 가진다.
- [0020] 본원에서 사용되는 용어 "두 서열의 상동성 백분율(percent homology of two sequences)"은 "두 서열의 동일성 백분율(percent identity of two sequences)"과 동등하다. 서열 동일성은 참조 서열(reference sequence)에 근거하여 결정된다. 서열 분석을 위한 알고리즘은 Altschul et al., J. Mol. Biol., 215, pp. 403-10 (1990)에 기재된 BLAST와 같이 당업계에 공지되어 있다. 본 발명의 목적을 위하여, 뉴클레오티드 서열 및 아미노산 서열 간의 동일성 및 유사성 수준을 결정하기 위하여, 뉴클레오티드 및 아미노산 서열의 비교를 이용할 수 있으며, 이는 표준 파라미터로 갭 정렬(gapped alignment)을 사용하여 the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/blast>)에 의하여 제공되는 BLAST 소프트웨어 패키지에 의하여 수행된다. 두 서열의 상동성 백분율은 정렬에 의한 두 서열의 최적 비교를 위한 갭의 수 및 들어갈 각각의 갭의 길이를 고려하여, 그 두 서열 내 동일한 아미노산 포지션의 수에 의하여 결정된다. 동일성 백분율은 서열 정렬을 고려한 주어진 포지션에서 동일한 아미노산의 수를 총 포지션 수로 나눈 값에 100을 곱한 값과 같다.
- [0021] 용어 "특이적으로 혼성화하는(specifically hybridizes)"은, 당업계에 전형적으로 사용되는 소정의 조건 하에 그러한 혼성화를 허용하는, 두 단일 가닥 핵산 분자 또는 충분히 상보적인 서열들 간에 관계를 나타낸다.
- [0022] 문구 "뉴클레오티드 PAM 서열 바로 앞에 위치한 이중 가닥 절단"은 표적 DNA 서열 내 이중 가닥 절단이 뉴클레오티드 PAM 서열 앞 0 내지 25 뉴클레오티드 거리에서 이루어질 것임을 의미한다.
- [0023] 가이드 RNA와 동시에 도입되는 외래 DNA 서열은 가이드 RNA의 특이성에 의하여 결정되는 절단 부위에서 이중 가닥 표적 DNA의 특이적 변형을 위하여 특이적으로 제조된 DNA 서열을 의미하기 위한 것이다. 이러한 변형은, 예를 들어, 표적 DNA 내 절단 부위에서 특정 뉴클레오티드의 삽입 또는 결실일 수 있다. 외래 DNA는 다른 생물로부터의 DNA 영역, 또는 표적 DNA의 것과 동일한 생물로부터의 DNA 영역일 수 있다.
- [0024] 특정 아미노산 서열을 포함하는 단백질은 상기 아미노산 서열 및 가능하다면 상기 아미노산 서열에 펩타이드 결합에 의하여 연결되는 기타 서열들로 구성되는 아미노산 서열을 가지는 단백질을 의미한다. 기타 서열의 예는 핵 위치 신호(nuclear localization signal(NLS)), 또는 상기 아미노산 서열에 증가된 기능을 제공하는 기타 서열일 수 있다.
- [0025] 가이드 RNA와 동시에 도입되는 외래 DNA 서열은 가이드 RNA의 특이성에 의하여 결정되는 절단 부위에서 이중 가닥 표적 DNA의 특이적 변형을 위하여 특이적으로 제조된 DNA 서열을 의미하기 위한 것이다. 이러한 변형은, 예를 들어, 표적 DNA 내 절단 부위에서 특정 뉴클레오티드의 삽입 또는 결실일 수 있다. 외래 DNA는 다른 생물로부터의 DNA 영역, 또는 표적 DNA의 것과 동일한 생물로부터의 DNA 영역일 수 있다.
- [0026] 세포 내로 도입되는 효과적인 양의 단백질 및 RNA는, 상기 세포 내로 도입될 때, 기능적 복합체, 즉 표적 DNA에 특이적으로 결합하여 그 안에서 DNA 상에 가이드 RNA 및 PAM 서열에 의하여 결정되는 부위에서 이중 가닥 절단을 생산할 복합체를 형성할 수 있는, 단백질 및 RNA의 양을 의미한다. 이러한 과정의 효율성은 당업자들에게 공지된 전형적인 기법을 사용하여 상기 세포로부터 분리된 표적 DNA를 분석함으로써 평가될 수 있다.
- [0027] 단백질 및 RNA는 다양한 기법에 의하여 세포에 전달될 수 있다. 예를 들어, 단백질은 그 단백질의 유전자를 암호화하는 DNA 플라스미드로서, 세포질 내 그 단백질의 번역을 위한 mRNA로서, 또는 그 단백질 및 가이드 RNA를 포함하는 리보뉴클레오단백질 복합체로서 전달될 수 있다.
- [0028] 핵산 암호화 시스템의 성분은 다음과 같이 세포 내로 직접 또는 간접적으로 도입될 수 있다: 당업자에게 공지된 방법에 의한 세포의 트랜스펙션(transfection) 또는 형질전환(transformation)에 의하여, 재조합 바이러스 사용에 의하여, DNA 미량주입과 같이 세포 상에서 조작에 의하여, 등.
- [0029] 뉴클레아제 및 가이드 RNAs 및 외래 DNA(필요하다면)로 구성되는 리보핵산 복합체는 상기 복합체를 세포 내로 트랜스펙션함으로써 또는 상기 복합체를 세포 내로 예를 들어 미량주입에 의하여 기계적으로 도입함으로써 전달

될 수 있다.

[0030] 세포 내로 도입될 단백질을 암호화하는 핵산 분자는 염색체 내로 통합되거나 또는 염색체외 복제 (extrachromosomally replicating) DNA일 수 있다. 일부 구현예에서, 세포 내로 도입된 DNA를 가지는 단백질 유전자의 효과적인 발현을 보증하기 위하여, 발현을 위한 코돈을 최적화하도록 상기 DNA 서열을 세포 유형에 따라 변형하는 것이 필요하며, 이는 다양한 생물의 계놈의 코딩 영역 내 동의어 코돈(synonymous codons) 존재의 빈도가 다르기 때문이다. 코돈 최적화는 동물, 식물, 균류 또는 미생물 세포 내 발현을 증가시키기 위하여 필요하다.

[0031] 서열번호 1의 아미노산 서열에 적어도 95% 동일한 서열을 가지는 단백질이 진핵 세포 내에서 작용하기 위해서는, 이 단백질이 그 세포의 핵 내에서 끝나는 것이 필요하다. 따라서, 본 발명의 일부 구현예에서, 서열번호 1의 아미노산 서열과 적어도 95% 동일하고 일 말단 또는 양 말단에서 하나 이상의 핵 위치 신호의 부가 (addition)에 의하여 더 변형된 단백질을 사용하여 표적 DNA 내에 이중 가닥 절단을 형성한다. 예를 들어, SV40 바이러스로부터의 핵 위치 신호를 사용할 수 있다. 효율적인 핵 전달을 제공하기 위하여, 상기 핵 위치 신호는 예를 들어 Shen B, et al. "Generation of gene-modified mice via Cas9/RNA-mediated gene targeting", Cell Res. 2013 May;23(5):720-3에 기재된 스페이서 서열에 의하여 주 단백질 서열로부터 분리될 수 있다. 나아가, 다른 구현예에서, 다른 핵 위치 신호 또는 상기 단백질을 세포핵 내로 전달하기 위한 대안적인 방법을 사용할 수 있다.

[0032] 본 발명은 이중 가닥 절단을 DNA 분자 내에 정확히 특정 포지션에 도입하기 위한, 이전에 규명된 Cas9 단백질과 상동인, P. 뉴모트로피카(*P. pneumotropica*)로부터의 단백질의 용도에 관한 것이다. 표적화된 변형을 계놈에 도입하기 위한 CRISPR 뉴클레아제의 사용은 많은 이점을 가진다. 첫번째로, 상기 시스템의 활성의 특이성이 crRNA 서열에 의하여 결정되며, 이는 모든 표적 유전자좌(loci)에 대하여 하나의 유형의 뉴클레아제의 사용을 허용한다. 두번째로, 상기 기법은 상이한 유전자 표적에 상보적인 몇몇 가이드 RNAs를 세포 내로 동시에 전달하는 것을 가능케 하여, 몇몇 유전자를 한번에 동시에 변형하는 것을 가능케 한다.

[0033] PpCas9는 동물 폐 내에 사는 설치류 병원균인, 파스퇴렐라 뉴모트로피카(*Pasteurella pneumotropica*) ATCC 35149에서 발견되는 Cas 뉴클레아제이다. 상기 파스퇴렐라 뉴모트로피카(*Pasteurella pneumotropica* (*P. pneumotropica*)) CRISPR Cas9 시스템(이하 CRISPR PpCas9로 언급)은 타입 II-C CRISPR Cas 시스템에 속하고, 고유의 스페이서 서열에 의하여 이격되는, 서열 5' ATTATAGCACTGCGAAATGAAAAAGGGAGCTACAAC3 을 가지는 4 개의 직접 반복(direct repeats(DR))을 가지는 CRISPR 카세트로 구성된다. 상기 시스템의 스페이서 중 어떠한 것도 현재 알려진 박테리오파지 또는 플라스미드와 서열에 있어서 일치하지 않으며, 이러한 사실은 생물 정보학적 분석에 의하여 관심 PpCas9 PAM을 결정하는 것을 불가능하게 한다. 상기 CRISPR 카세트에는, 새로운 스페이서의 조정 및 통합에 수반되는 Cas1 및 Cas2 단백질 유전자뿐만 아니라, 이펙터 Cas9 단백질 PpCas9의 유전자가 인접한다. 상기 Cas 유전자 가까이에, 직접 반복과 부분적으로 상보적이고 특징적인 2차 구조로 폴딩되는 서열이 발견되었으며, 이는 트레이서 RNA (tracrRNA)인 것으로 생각된다 (도 1).

[0034] 타입 II-C 시스템의 RNA-Cas 단백질 복합체의 특징적인 구조에 대한 지식은 CRISPR 카세트의 전사 방향을 예측하는 것을 가능케 하였다: pre-crRNA가 Cas 유전자와 반대 방향으로 전사된다 (도 1).

[0035] 따라서, PpCas9 유전자좌의 서열 분석은 트레이서 및 가이드 RNAs의 서열의 예측을 가능케 하였다 (표 1).

표 1

[0036] 생물 정보학적 방법에 의하여 결정된, CRISPR PpCas9 시스템의 가이드 RNAs의 서열. 굵은 글씨는 직접 반복 D R의 서열을 나타낸다.

명칭	서열
PpCas9 trRNA	5'GCGAAATGAAAAACGUUGUUACAUAAGAGAUGAAUUUCUCGCAAAGCTCUGCCUCUUGAAAUUUCGGUUUCAAGAGGCAUCUUUUU-3'
PpCas9 crRNA	5'-XXXXXXXXXXXXXXXXXXXX GUUGUAGCUCUUUUUCAUUUCGC -3'

[0037] PpCas9 뉴클레아제의 활성을 검증하고 관심 PpCas9 PAM을 결정하기 위하여, 시험관내 DNA 절단 반응을 재현하는 실험을 수행하였다. PpCas9 단백질의 PAM 서열을 결정하기 위하여, 이중 가닥 PAM 라이브러리의 시험관내 절단을 이용하였다. 이를 위하여, 다음과 같은 PpCas9 이펙터 복합체의 모든 성분을 얻는 것이 필요하였다: 제조합 형태의 가이드 RNAs 및 뉴클레아제. 가이드 RNA 서열의 결정은 crRNA 및 tracrRNA 분자의 시험관내 합성을 가능

케 하였다. 상기 합성은 NEB HiScribe T7 RNA 합성 키트를 사용하여 수행하였다. 이중 가닥 DNA 라이브러리는 3' 말단으로부터 무작위 7개 뉴클레오티드(5'-NNNNNN-3')에 의하여 플랭킹되는 프로토스페이서 서열을 포함하는, 374 염기쌍(bp) 단편이었다: 5'-
 cccggggtaccacggagagatggtggaatcatcttctcgtgggcatccttgatggccacctcgtcggaagtccccacgaggatgacagcaatgccaatgctgggggggctcttctgagaacgagctctgctgcctgacacggccaggacggccaacaccaaccagaactgggagaacagcactccgctctgggcttcatctcaactcgtcgactccctgcaaacacaaagaagagcatgttaaataggatctacatcacgtaacctgtcttagaagaggctagatactgcaatcaaggaccttatctctcttattgagcacNNNNNNNaactccatctacagcctactctcttatctctgtgatt -3'

- [0038] 이러한 표적을 절단하기 위하여, 다음 서열의 가이드 RNA를 사용하였다:
- [0039] tracrRNA:
- [0040] 5'GCGAAATGAAAAACGUUGUACAAUAGAGAUGAAUUCUCGCAAAGCTCUGCCUCUUGAAAUUCGGUUUCAAGAGGCAUCUUUUU
- [0041] 및 crRNA: 5'
- [0042] **uaucuccuuucauugagcac**GUUGUAGCUCUUUUUCAUUUCGC.
- [0043] 짧은 글씨는 상기 프로토스페이서에 상보적인 crRNA 서열을 나타낸다 (표적 DNA 서열).
- [0044] 제조합 PpCas9 단백질을 생산하기 위하여, 그 유전자를 플라스미드 pET21a 내로 클로닝하였다. 통합 DNA 기술 (Integrated DNA Technologies (IDT))에 의하여 합성된 DNA를 상기 유전자를 암호화하는 DNA로서 사용하였다. 상기 서열을 코돈-최적화하여 P. 뉴코트로피카 계통 내에서 발견되는 레어 코돈을 제외시켰다. E. coli Rosetta 세포를 결과의 플라스미드 pET21a-6xHis-PpCas9로 형질전환시켰다.
- [0045] 밤새 배양물 500 μ l를 LB 배지 500ml 내에 희석하고, 0.6Ru의 광학 밀도가 얻어질 때까지 세포를 37 $^{\circ}$ C에서 증식시켰다. IPTG를 1 mM의 농도로 첨가함으로써 표적 단백질의 합성이 유도되었고, 그 후 상기 세포를 20 $^{\circ}$ C에서 6 시간 동안 인큐베이션하였다. 그 다음, 상기 세포를 5,000g에서 30분 동안 원심분리하고, 결과의 세포 침전물을 -20 $^{\circ}$ C에서 냉동시켰다.
- [0046] 상기 침전물을 얼음에서 30분 동안 해동시키고, 15mg의 라이소자임이 보충된 15ml의 용해 완충액(Tris-HCl 50 mM pH 8, 500 mM NaCl, β -머캅토에탄올 1 mM, 이미다졸 10 mM) 내에 재현탁시키고, 얼음에서 30분 동안 재-인큐베이션하였다. 그 후, 상기 세포를 30분 동안 초음파에 의하여 파괴시키고, 16,000g에서 40분 동안 원심분리하였다. 결과의 상청액을 0.2 μ m 필터를 통과시키고 1ml/min으로 HisTrap HP 1 mL 컬럼(GE Healthcare) 상에 가하였다.
- [0047] AKTA FPLC 크로마토그래프(GE Healthcare)를 사용하여 1ml/min으로 크로마토그래피를 수행하였다. 단백질을 가한 컬럼을 30mM 이미다졸이 보충된 20ml의 용해 완충액으로 세척한 후, 상기 단백질을 300mM 이미다졸이 보충된 용해 완충액으로 씻어 냈다.
- [0048] 그 다음, 친화성 크로마토그래피 과정에서 얻어진 단백질 분획을 다음 완충액으로 평형화된 Superdex 200 10/300 GL 겔 여과 컬럼 (24ml)을 통과시켰다: Tris-HCl 50 mM pH 8, 500 mM NaCl, 1 mM DTT. Amicon 농축기 (30 kDa 필터를 가지는)를 사용하여, PpCas9 단백질의 모노머 형태에 해당하는 분획을 3 mg/ml로 농축시킨 다음, 정제된 단백질을 10% 글리세롤 함유 완충액 내에 -80 $^{\circ}$ C에서 저장하였다.
- [0049] 선형 PAM 라이브러리를 절단하는 시험관내 반응을 다음 조건 하에 20 μ l 부피로 수행하였다: 반응 혼합물은: 1X CutSmart 완충액(NEB), 5 mM DTT, 100 nM PAM 라이브러리, 2 μ M trRNA/crRNA, 400 nM PpCas9 단백질로 구성되었다. 대조군으로서, RNA를 함유하지 않은 샘플을 유사한 방식으로 제조하였다. 상기 샘플을 상이한 온도에서 인큐베이션하고 2% 아가로스 겔 내에서 겔 전기영동에 의하여 분석하였다. PpCas9 단백질에 의하여 DNA를 정확한 인식하고 특이적으로 절단하면, 약 326 및 48 염기쌍의 두 개의 DNA 단편이 생성될 것이다 (도 2).
- [0050] 상기 실험 결과는 PpCas9가 뉴클레아제 활성을 가지며 PAM 라이브러리 단편 부분을 절단함을 보였다. 온도 구배 (도 3)는 상기 단백질이 35-45 $^{\circ}$ C 범위 온도에서 활성임을 나타냈다. 그 후의 연구는 작업 온도로서 42 $^{\circ}$ C의 온도를 사용하였다.
- [0051] 라이브러리 절단 반응을 선택된 조건 하에 반복하였다. 반응 생성물을 1.5% 아가로스 겔 상에 가하고 전기영동시켰다. 374 bp 길이의 미절단 DNA 단편을 겔로부터 추출하고, NEB NextUltra II 키트를 사용하여 대용량 시퀀싱을 준비하였다. 상기 샘플을 Illumina 플랫폼 상에서 시퀀싱한 다음, 서열 분석을 생물 정보학적 방법을 사용하여 수행하였다: Maxwell CS, et al., A detailed cell-free transcription-translation-based assay to

decipher CRISPR protospacer-adjacent motifs. *Methods*. 2018 Jul 1;143:48-57에 기재된 접근법을 사용하여, 대조 샘플에 비하여, PAM (NNNNNN)의 개별 포지션에서 뉴클레오티드 존재의 차이를 결정하였다. 나아가, PAM 로고를 만들어 결과를 분석하였다 (도 4).

[0052] 데이터 분석을 위한 두 접근법 모두(도 4) PAM 포지션 5, 6 및 7의 중요성을 나타낸다. 따라서, 시험관내 분석은 PpCas9에 대한 추정 PAM 서열을 다음과 같이 확립한다: NNNNATT. 그러나, 이 서열은 PAM 결정을 위한 스크리닝 접근법에 의하여 얻어지는 부정확한 결과를 고려하면 단지 잠정적일뿐이다.

[0053] 이와 관련하여, 서열의 더 정확한 결정을 위하여 개별 PAM 서열 포지션의 중요성을 검증하였다. 이를 위하여, PAM 서열 CAACATT(또는 그 유도체)에 의하여 플랭킹된 DNA 표적 5'-atctcctttcattgagcac-3'을 함유하는 DNA 단편의 시험관내 절단 반응을 수행하였다: 5'-
 cccggggtaccacggagagatggtggaatcatcttctcgtgggcatccttgatggccacctcgtcggaagtccccacgaggatgacagcaatgccaatgctgggggggctcttctgagaacgagctctgctgcctgacacggccaggacggccaaccaaccagaacttgggagaacagcactccgctctgggcttcatctcaactcgtcgactccctgcaaacacaagaagagcatgttaaataggatctacatcacgtaacctgtcttagaagaggctagatactgcaattcaaggacctttatctcctttcattgagcacCAACATTaactccatctaccagectactctcttatctctggtatt- 3'

[0054] 모든 DNA 절단 반응을 다음 조건 하에 수행하였다:

[0055] 1xCutSmart 완충액

[0056] 400 nM PpCas9

[0057] 20 nM DNA

[0058] 2 μM crRNA

[0059] 2 μM tracrRNA

[0060] 인큐베이션 시간 - 30분, 반응 온도 - 42°C.

[0061] PAM 포지션 1의 네 개의 가능한 뉴클레오티드 변이로 치환은 단백질 활성 효능에 영향을 미치지 않았다 (도 5).

[0062] 포지션 5 및 6의 예측된 중요성을 각각의 PAM 포지션에서 단일 뉴클레오티드 치환(퓨린을 피리미딘으로 및 반대로)에 의하여 실험적으로 확인하였다. 상기 치환이 포지션 5 및 6에서 일어났을 때, 단백질은 그 활성이 실제로 중단되었다. 상기 치환이 포지션 7에서 일어났을 때, PpCas9 활성 효율성은 2배 감소하였으며, 이러한 사실은 이 포지션에서 뉴클레오티드에 대하여 감소된 요건을 반영한다 (도 6). 따라서, 시험관내 PpCas9 뉴클레아제의 PAM 스크리닝 결과에 따르면, PAM 포지션 5에서 가장 가능성 있는 뉴클레오티드는 아데닌 또는 구아닌이고, 이러한 사실은 실험적으로 확인되었다 (도 7). A - G 치환은 단편 절단의 효율성을 감소시키지 않았다.

[0063] 시험관내 스크리닝 결과에 따르면, 포지션 7에서 "T" 또는 "S"를 가지는 단편은 더 효율적으로 인식된다. 추가적인 실험을 수행하여 이 포지션에서 뉴클레오티드의 중요성을 명확히 검증하였다. 시험관내 시험 결과는 포지션 7에서 뉴클레오티드 "T"의 A 또는 G로 치환은 절단 효율성을 40-50% 감소시켰다 (도 8). 따라서, PAM 포지션 7은 포지션 5 및 6에 비하여 덜 보존된다: 포지션 7의 퓨린은 인식 효율성을 감소시키나 PpCas9 단백질이 이중가닥 절단을 DNA 내로 도입하는 것을 방지하지 않는다.

[0064] 연구의 결과는 다음과 같았다: PpCas9 뉴클레아제에 의하여 인식되는 PAM은 다음 식 5'- NNNN(A/G)TT-3'에 해당된다. 포지션 7은 덜 보존된다.

[0065] 다음 방법의 실시 구현에는 본 발명의 특징을 개시할 목적으로 제공되며, 어떠한 방식으로든 본 발명의 범위를 제한하는 것으로 해석되지 않아야 한다.

[0066] **실시예 1:** 다양한 DNA 표적의 절단에서 PpCas9 단백질의 활성 시험

[0067] 서열 5'-NNNN(A/G)TT-3'로 플랭킹되는 다양한 DNA 서열을 인식하는 PpCas9 단백질의 능력을 확인하기 위하여, 인간 grin2b 유전자 서열로부터의 DNA 표적의 시험관내 절단 실험을 수행하였다 (표 2 참조).

표 2

인간 GRIN2B 유전자로부터의 DNA 표적

서열	PAM						
TATCTCCTTTCATTGAGCAC	C	A	A	A	C	C	C

CAGCTGAAGTAATGTTAGAG	C	C	A	C	A	T	T
AATAAGAAAAACATTATTAT	C	A	C	C	A	T	T
GGGGCTATAAGTACACAAGC	C	C	T	G	C	A	T
CGTTCTCAGAAGAGCCCCC	C	A	G	C	A	T	T
CCCACGAGAAAGATGATTTC	C	A	C	C	A	T	C

[0069] 짐작건대 PAM 공통 서열(consensus sequence) 5'-NNNN(A/G)TT-3'에 따라 PpCas9에 의하여 인식 가능한 인식 부위를 가지는 *grin2b* 유전자의 PCR 단편을 (표 2) 절단 반응에서 표적으로 사용하였다. PpCas9를 이러한 부위로 디렉팅하는 CrRNAs를 합성하여 상기 서열들을 인식하였다.

[0070] 절단 반응을 PpCas9에 대하여 선택된 조건 하에 수행하였고; 그 결과를 도 9에 나타낸다. 도 9는 PpCas9 효소가 적합한 PAM으로 네 개의 타겟 중 세 개를 성공적으로 절단함을 보인다.

[0071] 레인 6의 표적은 PAM 서열 CAGCATT을 가졌으며, 이는 감소 분석(depletion analysis) 결과에 기초한 예측에 따르면, 상기 단백질에 의하여 효율적으로 인식될 것이다. 그러나, 상기 단편의 인식은 이 실험에서 일어나지 않았다.

[0072] 따라서, PAM CAGCATT를 동일 PAM으로 제한된 다른 프로토스페이서 표적 상에서 추가로 검증하였다 (도 10). 이 경우, PAM이 효과적으로 인식되었으며, 이는 DNA 절단을 초래하였다. 따라서, 상기 단백질은 DNA 표적 서열에 대한 추가의 선호도를 가진다. 이러한 선호도는 아마도 DNA의 2차 구조와 관련이 있을 것이다.

[0073] 따라서, 상기 연구는 PpCas9 내 뉴클레아제 활성의 존재를 보이며, 그 PAM 서열의 결정 및 가이드 RNAs의 서열 검증을 허용하였다.

[0074] PpCas9 리보핵산단백질 복합체는 PAM 5'-NNNN(A/G)TT -3'로 제한된 표적 내에 프로토스페이서의 5' 말단으로부터 표적 내 절단을 특이적으로 도입한다. PpCas9/RNA 복합체의 계획을 도 11에 도시한다.

[0075] **실시예 2: DNA 표적 절단을 위한 하이브리드 가이드 RNA의 사용**

[0076] sgRNA는 융합된 tracrRNA (트레이서 RNA) 및 crRNA인, 가이드 RNAs의 형태이다. 최적의 sgRNA를 선택하기 위하여, tracrRNA-crRNA 이중나선의 길이가 다른, 세계의 서열 변이를 제작하였다. RNAs를 시험관내 합성하고, 이를 수반하는 DNA 표적 절단 실험을 수행하였다 (도 12).

[0077] 다음 RNA 서열을 하이브리드 RNAs로서 사용하였다:

[0078] 1 - sgRNA1 25DR:
UAUCUCCUUCAUUGAGCACGUUGUAGCUCCUUUUUCAUUUCGCGAAAGCGAAAUGAAAAACGUUGUACAAUAGAGAUGAAUUUCUCGCAAAGCTCTGC
 CUCUUGAAAUUCGGUUUCAAGAGGCAUCUUUUU

[0079] 2 - sgRNA2 36DR
UAUCUCCUUCAUUGAGCACGUUGUAGCUCCUUUUUCAUUUCGAGUGCUAUAUAGAAAAUUAUAGCACUGCGAAAUGAAAAACGUUGUACAAUAGAG
 AUGAAUUUCGCAAAGCUCUGCCUCUUGAAAUUCGGUUUCAAGAGGCAUCUUUUU

[0080] 짧은 글씨는 DNA 표적과 페어링을 제공하는 20-뉴클레오타이드 서열을 나타낸다 (sgRNA의 가변부). 나아가, 이 실험은 RNA없는 대조 샘플, 및 crRNA+trRNA를 사용하는 표적 절단인 양성 대조군을 사용하였다.

[0081] 해당 공통 서열 PAMCAACATT과 인식 부위 5' **tatctcctttcattgagcac** 3'를 함유하는 서열을 DNA 표적으로 사용하였다:
 5'-
 cccggggtaccacggagagatggtggaatcatcttctcgtgggcatccttgatggccacctcgtcggaagtcccacgaggatgacagcaatgccaatgc
 tgggggggctcttctgagaacgagctctgctgcctgacacggccaggacggccaacaccaaccagaactgggagaacagcactccgctctgggcttcatct
 tcaactcgtcgactccctgcaaacacaaagaagagcatgttaaataaggatctacatcacgtaacctgtcttagaagaggctagatactgcaattcaagga
 cct**tatctcctttcattgagcac**CAACATTcaactccatctaccagcctactctcttatctctggtatt - 3'

[0082] 짧은 글씨는 인식 부위를 나타내고, 대문자는 PAM을 나타낸다.

[0083] 반응을 다음 조건 하에 수행하였다: PAM (CAACATT)을 함유하는 DNA 서열의 농도는 20nM였고, 단백질 농도는 400nM였고, RNA 농도는 2 μM였고; 인큐베이션 시간은 30분 이었고, 인큐베이션 온도는 37°C였다.

[0084] 선택된 sgRNA1 및 sgRNA2는 천연 tracrRNA 및 crRNA 서열로서 효율적인 것으로 밝혀졌다: DNA 표적의 80% 이상

에서 절단이 일어났다 (도 12).

[0085] 이러한 하이브리드 RNA 변이는, DNA 표적과 직접 페어링하는 서열을 변형한 후, 임의의 다른 표적 DNA를 절단하는데 사용될 수 있다.

[0086] **실시예 3:** P. 뉴모트로피카에 속하는 밀접히 관련된 생물로부터의 Cas9 단백질

[0087] 지금까지, 어떠한 CRISPR-Cas9 효소도 P. 뉴모트로피카(*P. pneumotropica*) 내에서 규명되지 않았다. 크기가 유사한 스타필로코커스 아우레우스(*Staphylococcus aureus*)로부터의 Cas9 단백질은 PpCas9와 28% 동일하다 (도 13, 동일성 정도는 BLASTp 소프트웨어, 디폴트 변수)에 의하여 계산되었다). 유사한 정도의 동일성이 다른 공지의 Cas9 단백질 내에 존재한다 (도시하지 않음).

[0088] 따라서, PpCas9 단백질은 그 아미노산 서열에 있어서 지금까지 연구된 다른 Cas9 단백질과 유의하게 다르다.

[0089] 유전공학 분야의 당업자는 본 발명에서 출원인에 의하여 얻어지고 규명된 PpCas9 단백질 서열 변이가 단백질 자체의 기능을 변화시키지 않고 변형될 수 있음을 (예를 들어, 기능적 활성화에 직접적인 영향을 미치지 않는 아미노산 잔기의 돌연변이 유도(directed mutagenesis)에 의하여 (Sambrook et al., Molecular Cloning: A Laboratory Manual, (1989), CSH Press, pp. 15.3-15.108)) 이해할 것이다. 특히, 당업자는 단백질 기능에 책임 있는 (단백질 기능 또는 구조를 결정하는) 잔기에 영향을 미치지 않고, 비-보존된 아미노산 잔기가 변형될 수 있음을 인식할 것이다. 이러한 변형의 예는 비-보존된 아미노산 잔기를 상동의 것으로 치환하는 것을 포함한다. 비-보존 아미노산 잔기를 함유하는 영역 중 일부를 도 12에 도시한다. 본 발명의 일부 구현예에서, 서열번호 1의 아미노산 서열과 적어도 95% 동일하고 비-보존 아미노산 잔기에서만 서열번호 1와 다른 아미노산 서열을 포함하는 단백질을 사용하여, DNA 분자 내에서, 상기 DNA 분자 내 뉴클레오티드 서열 5'-NNNN(A/G)TT-3' 바로 앞에 위치한 이중 가닥 절단을 형성하는 것이 가능하다. 상동 단백질은 해당 핵산 분자의 돌연변이 유도(예를 들어, 부위-특이적(site-directed) 또는 PCR-매개 돌연변이(PCR-mediated mutagenesis)) 후, 본원에 기재된 기능 분석에 따라 암호화된 변형된 Cas9 단백질을 그 기능의 보존에 대하여 시험함으로써 얻어질 수 있다.

[0090] **실시예 4:** PpCas9를 사용하는 인간 세포의 게놈 DNA의 변형

[0091] 인간 세포의 게놈 DNA를 변형하기 위하여, PpCas9 뉴클레아제 유전자를 CMV 프로모터의 조절 하에 진핵 생물 플라스미드 벡터 내로 클로닝하였다. 세포핵으로 뉴클레아제 전달을 보증하는 핵 위치 신호를 암호화하는 서열을 PpCas9 유전자의 5' 및 3' 말단에 첨가하였다. sgRNA 서열을 U6 프로모터의 조절 하에 벡터 내로 클로닝하였다. 상기 시스템의 활성을 시험하기 위하여, 20 내지 24 뉴클레오티드 길이의 표적 DNA와 상보적인 서열을 가지는 sgRNA를 사용하였다. 당업계에 알려진 SpCas9-기반 게놈 DNA 변형 시스템을 함유하는 유사한 플라스미드를 양성 대조군으로서 사용하였다. 트랜스펙션의 효능을 평가하기 위하여, 상기 플라스미드는 GFP (녹색 형광 단백질) 유전자를 더 함유하였다. 다음의 인간 게놈 DNA 영역을 DNA 표적으로 사용하였다 (표 3).

표 3

인간 EMX1 및 GRIN2B 유전자의 DNA 표적

[0092]

뉴클레아제	부위 명칭	표적 서열	PAM
PpCas9	EMX1.1 sg20	GCCCTTCCTCCTCCAGCTTC	GTT
PpCas9	EMX1.1 sg24	TCAGGCCCTTCCTCCTCCAGCTTC	GTT
FpCas9	EMX1.2 sg20	GGAGGTGACATCGATGTCCT	ATT
FpCas9	EMX1.2 sg24	CATTGGAGGTGACATCGATGTCCT	ATT
PpCas9	GRIN2B1.1 sg20	CAGCTGAAGTAATGTTAGAG	ATT
PpCas9	GRIN2B1.1 sg24	TTAGCAGCTGAAGTAATGTTAGAG	ATT
PpCas9	GRIN2B1.2 sg20	AATAAGAAAAAATTATTAT	ATT
PpCas9	GRIN2B1.2 sg24	ATAAAATAAGAAAAAATTATTAT	ATT
SpCas9	EMX1 sg20	GAGTCCGAGCAGAAGAAGAA	GGG
SpCas9	GRIN2B sg20	ACCTTTATTGCCTTGTCA	AGG

[0093] EMX1.1 및 EMX1.2는 EMX1 유전자 내 두개의 다른 변형 부위이고; 유사하게, GRIN2B1.1 및 GRIN2B1.2는 GRIN2B 유전자 내 두개의 다른 변형 부위이다. DNA 표적을 PpCas9 5'-NNNNRRTT -3' 또는 SpCas9 5'- NGG -3'의 PAM 서열에 의하여 3' 말단으로부터 플랭킹하였다.

[0094] 진핵 세포 내 PpCas9 뉴클레아제의 효과적 활성을 위하여, 상기 단백질을 진핵 세포의 핵 내로 불러오는 것이

필요하다. 이는 Shen B, et al. "Generation of gene-modified mice via Cas9/RNA-mediated gene targeting", Cell Res. 2013 May;23(5):720-3에 기재된 스페이서 서열을 통하여 또는 스페이서 서열 없이 PpCas9 서열에 연결되는 SV40 T-항원(Lanford et al., Cell, 1986, 46: 575-582)으로부터의 핵 위치 신호를 사용하는 방식으로 수행될 수 있다.

[0095] 제공되는 실시예에서, 인간 세포 핵 내에 수송된 뉴클레아제의 완전한 아미노산 서열은 다음 서열이었다:

[0096] MAPKKKRVGTHGVPAAEQNNPLNYILGLDLGIASIGWAVVEIDEESSPIRLIDVGVRTFERAEVAKTGESLALSRRLARSSRRLIKRRAERLKKAKRLLKAEKILHSIDEKLPINWQLRVKGLKEKLERQEWAAVLLHLSKHRGYLSQRKNEGKSDNKELGALLSGIASNHQMLQSSEYRTPAEIAVKKFQVEEGHIRNQRGSYTHTF SRLDLLAEMELLFQRQAELGNSYTSSTLLENLTALLMWQKPALAGDAILKMLGKCTFEPSEYKAAKNSYSAERFVWLTKLNNLRIENGLTERALNDNERFALLEQPYEKSCLTYAQRAMLALS DNAIFKGVRYLGEDKKTVESKTTL IEMKFYHQIRKTLGSAELKKEWNEKGNSDLLDEIGTAFSLYKTDDDICRYLEGLPERVLNALLENLNFDKFIQLSLKALHQILPLMLQGQRYDEAVSAIYGDHYGKSKSTETRLPTIPADEIRNPVVLRLTLQARKVINAVVRLYGSPEARIHIE TAREVGSYQDRKLEKQEDNRKQRESAVKFKEMFPHFVGEKPKGDKILMRLYELQQAKCLYSGKSELEHRLLEKGYVEVDHALPFSRTWDDSFNNKVLVLANENQNKGNLTPYEWLDGKNSERWQHFVVRVQTSFGFSYAKKQIRLNHKLDEKGFIERNLNDRYVARFLCNFIADNMLVGKGRNVFASNGQITALLRHRWGLQKVRQENDRHHALDAVVVACSTVAMQKQITRFVRYNEGNVFSGERIDRETGEIIPHFPSPWAFKENVIRIFSENPKLELENRLPDYPQYNHEWVQPLFVSRMPTRKMTGQGHMETVKSARKLNEGLSVLKVPLTQLKLSDLERMVNDRDIEALYESLKRLEQFGNDPAKFAEFPYKGGALVKA VRLEQTQKSGVLVRDNGVADNASMVRVDVFTKGGKYFLVP IYTWQVAKGILPNRAATQKGDENDWDIMDEMATFQFSLCQNDLIKLVTKKKTIFGYFNGLNRATSNINIKEHDLDKSKGLGIYLEVGVKLAISLEKYQVDELGKNIRPCRPTKRQHVRFKRPAAATKKAGQAKKKK

[0097] 이 실험에서 사용된 플라스미드는 다음 서열을 가졌다:

[0098] gagggcctatttcccatgattccttcatttgcataatgcatatagcacaagagctgttagagagataattggaattaatgtgactgtaaacacaaagatattagta caaaatcgtgacgtagaaagtaataatttcttgggtagttgagcttttaaaattatgttttaaatggactatcatatgcttaccgtaacttgaagatattt cgatttcttggctttatataatcttggaaaggacgaacaccgXXXXXXXXXXXXXXXXXXXXGTTGTAGCTCCCTTTTTCATTCGCGAAAGCGA AATGAAAACGTTGTTACAATAAGAGATGAATTTCTCGCAAAGCTCGCCTTGAATTTCCGGTTTCAAGAGGCATCTTTTTgctTCTCATGTCCAATAT GACCGCCATGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGGTTacataactta cggtaaatggccccctggctgaccgcccacgacccccccattgacgtcaataatgacgtatgttccatagtacgccaataggactttccattgac gtcaatgggtggagtattacggtaaacgcccacttggcagtagatcaatgtagtcatatgccaagtcgccccctattgacgtcaatgacggtaaatggc cgcctggcattatgccagtagacacttaccggactttcctacttggcagtagatctacgtattagtcatcgctattaccatggtagcgggtttggc agtagacacaaatggcgtggatagcggttgactcacgggatttccaagctccacccattgacgtcaatgggagttgtttggcaccacaaatcaacggg actttccaaaatgtagtaaacccccccccctgtagcgaatgggggtagcgtgtacggtaggaggtctatataagcAGAGCTCGTTTAGTGAACCGTC AGAATTAATTCAGATCGATCTACCaccgccaccATGATGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGAACAGAATAATCCG CTTAACTACATCTCTGGGCTGGATTGGGAATTGCGAGTATAGGCTGGGCGGTGGTTGAGATCGATGAAGAGAGTAGTCCGATACGCCCTTATCGAGTGGG AATTAGGACGTTGAGAGGGCGGAGGTCGCCAAGACCGGTGAGAGCTGGCCCTCAGCCGGCGGCTCGCTCGATCTAGTCGAGGCTTATAAAGAGGAGGGCT GAGCGCCTAAGAAAGCTAAGAGGCTCCTAAGGCAGAAAAAATCTGCATAGTATCGACGAAAAGCTGCGGATAAATGTTTGGCAGCTCCGAGTAAAAGGG CTGAAGGAAAAATGGAAAGGCAGGAGTGGGCGCGGACTGCTTCACTCTCCAAGCACCGGGCTATCTGTCTCAGCGAAAAACGAAGGTAAGTCAGAC AACAGGAGCTGGGCGCACTTTTGTCCGGATAGCGTCAATATCAGATGCTCCAATCAAGTAGTATCGGACCCCTGCGGAGATCGCCGTTAAAAAGTT CAAGTTGAGGAGGGCCACATCAGAAATCAGAGGGGCTTACACCCATACGTTCTCTAGACTCGACCTCCTTGGGAAATGGAACCTCTGTTTCAGCGCCAG GCGGAGCTTGGTAACTCTACAGTCCACTACCTCCTGAAAACTGACAGCCCTGCTGATGTGCAGAAAGCCGCTTGGCGGGGATGCCATCTGAAG ATGCTGGGTAATGCACCTTTGAGCCGTCAGAAATATAAAGCCGCAAGAATAGTTACTCTCGGAGCGATTTGTTTGGTTGACAAAGTTGAATAACCTGCGC ATCTGGAGAACGTTACCGAGCGCACTCAATGATAATGAGCGCTTCGCCCTCTGGAACAGCCCTACGAGAAGTCCAAGTCCACTACGCCAAGTCAGA GCCATGCTGGCTCTTAGTGACAACCGATTTTAAAGGGCGTGCATACTTGGGCGAGGATAAGAAAACCGTAGAGTCAAAAACGACTCTGATCGAGATGAAA TTCTATCACCAAATTAGAAAAGCCCTCGGTTCTGCCGAGCTGAAAAAGGAATGGAACGAACCTAAGGGTAACAGCGACCTGCTCGATGAAATCGGTACC GCA TTTAGCCTTTATAAAAACGACGACACTCTGCCGATATTTGGAGGGGAAGCTCCCAGAGCGAGTATTGAATGCACCTCTTGAACCTTAATTTTGACAAG TTCATTCAGTGTCCCTCAAAGCACTGCATCAAATCCTCCACTTATGCTGCAAGGACAACGATACGACGAAGCCGTCAGCGGATATATGGAGATCATTAC GGAAAAAAGTCCACCAGACACAGACTGCTTCTACGATCCCGCGGATGAGATCAGAAATCCCGTAGTCTTCAACACTTACTCAGGCTAGGAAGGTG ATTAATGCGGTAGTTAGTTGTATGGATCTCCGGCACGGATACATATAGAAAACAGCTCGCGAAGTGGGTAATCTTACCAAGACCGCAAGAAATGGAGAAA CAACAGGAGGATAACCGAAAAGCAACGAGAATCTGCCGTTAAAAAGTTTAAAGAAATGTTTCTCACTTTGTAGGAGAACCGAAGGTAAGATATCTTGAAA ATGCGGTTGTACGAGTTGAGCAAGCTAAGTGTCTATAGCGCAAGAGTTTGAATGACCCGCTCTGGAGAAAGGCTACGTGGAAGTAGACCATGCG CTCCGTTTTCCGAACCTGGGATGATTCTTCAATAACAAAGTCTTGTGCTGGCAATGAGAACCAGAACAAAGAAATCTGACTCCTTATGAGTGGTTG GATGGCAAGAATAATCTGAGCGGTGGCAACATTTGTTGTCGCGTCCAAAGCTCAGGGTTCAGCTATGCTAAGAAAACAAAGGATCTCAATCACAAGCTC GACGAGAAAGGATTCATAGAACGAAATTTGAATGACACTAGGATGTGGCTCGATTTCTCTGCAATTTTATTGCTGACAAATATGCTCCTCGTTGGGAAGGGA AAGCGGAATGTTTTGTCATAAATGGCAGATAACGGCGCTCTTGAGACATAGATGGGGCTGCAAAAGGTGAGAGAGCAAAATGATAGACATACGCCCTG

GATGCCGTTGTAGTCGCCTGTTCAACGGTTGCGATGCAGCAAAGATCACTCGGTTTCGTTAGGTATAACGAAGGGAACGTTTTTAGTGAGAGCGCATAGAT
 CGGGAAACAGGCGAAATCATCCCTTTGCATTCCCAAGTCTTGGGCTTTTTTCAAAGAGAAATGTGAAATAAGGATATTAGTGAAAAACCTAAGTTGGAG
 CTTGAGAATCGGTTGCCGATTATCCCAGTACAATCATGAGTGGGTTCAACCGCTGTTTCGATCCCGCATGCCAACCGAAAGATGACCGGGCAGGGTACAC
 ATGGAGACTGTGAAATCTGCAAAGAGACTTAATGAGGGCCTGTCACTGTTGAAGGTGCCCTTGACTCAACTGAAATGAGCGACCTCGAGCGCATGGTAAAC
 CGCGATAGAGAAATCGCACTTTATGAGAGTCTGAAGGCGCGATTGGAACAATTCGGTAATGATCCGGCAAAGGCTTTCGCTGAGCCATTCTACAAGAAGGGT
 GGAGCGCTGGTTAAGGCTGTCCGACTCGAACAGACAAAAAGTCAGGGGTCTTGGTCAGAGATGGTAACGGGGTTGCCGACAACCGCTCCATGGTACGAGTA
 GATGTTTTACGAAAGGAGGAAATACTTTCTGGTACCTATCTATACCTGGCAAGTTGCCAAGGGAATACTCCCGAATAGGGCGGCGACCCAGGAAAGGAT
 GAAAAAGACTGGGATATAATGGATGAAATGGCTACGTTTTCAGTTTTCAGTTTGTGCCAGAATGACCTCATAAACTGGTAACCAAAAAAAGACTATATTCGGG
 TATTTCAATGGCCTTAATCGGGCAACTTCCAATATCAACATCAAGGAACATGATCTGGATAAGAGCAAGGAAAGCTTGGTATCTATCTCGAAGTTGGAGTC
 AAGCTCGCTATTTCCCTCGAGAAATATCAAGTAGATGAACTGGGAAAGAAATATACGGCCATGCCGGCCCAAAAAAGACAACCGTACGGTTCAAAGGCCG
 GCGGCCACGAAAAAGGCCGCGCCAGGCAAAAAAGAAAAAGGGATCCTACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATAC
 CCATATGATGTCCCGACTATGCCGGCGCAACAACTTCTCTCTGCTGAAACAAGCCGAGATGTCGAAGAGAATCCTGGACCGgtgagcaagggcgaggag
 ctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaacggccacaagttcagcgtgtccggcgagggcgaggcgcatgccacctacggc
 aagctgacctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccacctcgtgaccacctgacctacggcgtgcagtgttcagccgctac
 cccgaccacatgaagcagcagcactcttcaagtccgcatgccgaaggtacgtccaggagcgccacctctctcaaggacgacggcaactacaagacc
 cgcccgaggtgaagttcgagggcgacacctggtgaaccgcatcgagctgaaggcatcgacttcaaggaggacggcaacatcctggggcacaagctggag
 tacaactacaacagccacaacgtctatcatggccgacaagcagaagaacggcatcaagtgacttcaagatccgccaacatcgaggacggcagcgtg
 cagctcgccgacctaccagcagaacacccccatcgccgacggccccgtgctgctgcccacaaccactacctgagcaccagtcgcccctgagcaaaagac
 cccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgcccgggatcactctcgcatggacgagctgtacaagTAA

- [0099] 상기 플라스미드 서열에서 다음 부분들이 구별되었다: U6 프로모터 (첫번째 영역, 대문자), 프로토스페이스에 상보적인 서열 ("XXX-XXX"), sgRNA의 보존된 부위 (세번째 영역, 대문자), PpCas9 유전자(굵은 글씨로 강조), GFP 유전자 (마지막 영역, 대문자).
- [0100] PpCas9 또는 SpCas9를 가지는 플라스미드를 리포펙타민 2000 시약을 사용하여 인간 HEK293T 세포 배양액 내로 트랜스펙션하였다. 트랜스펙션 후 72시간에, 세포를 용해하고, 결과의 용해물을 PCR 처리하여 게놈 DNA의 표적 변형 부위를 포함하는 영역을 생성하였다. 결과의 PCR 단편들을 T7 엔도뉴클레아제 I과 시험관내 반응시켜 게놈 DNA의 표적 부위 내 삽입 및 결실 빈도를 결정하였다. 반응 생성물을 아가로스 겔 상에 가하고 전기영동시켰다. 도 14a는 PpCas9가 종래 기술에 기재된 SpCas9 뉴클레아제와 유사한 효율성으로, EMX1 및 GRIN2b 유전자에 변형을 활발히 도입함을 보인다.
- [0101] 이 실험은, 게놈 DNA를 효과적으로 변형시키기 위하여, PpCas9는 SpCas9에 비하여 연장된 sgRNAs를 필요로 하며; 제공된 실시예에서, 유전자 변형의 효율성은 (20 뉴클레오티드 길이에 비하여) 24 뉴클레오티드 길이를 가지는 DNA 표적에 상보적인 서열을 가지는 sgRNA를 사용할 때 더 크다는 것을 보였다.
- [0102] 대용량 시퀀싱(high-throughput sequencing)은 표적 DNA 부위 내 도입된 변형을 확인하였다. 도 14b는 EMX1 유전자의 뉴클레오티드 서열에 대한 검출 가능한 변형의 예를 도시한다.
- [0103] NLS_PpCas9-NLS를 인간 세포에 전달하기 위하여 리보핵산 복합체 형태로 전달 또한 사용될 수 있다. 이는 CutSmart 완충액 (NEB) 내에서 가이드 RNAs와 PpCas9-NLS의 제조합 형태를 인큐베이션함으로써 수행된다. 상기 제조합 단백질은 친화성 크로마토그래피(NiNTA, Qiagen)와 크기 배제(Superdex 200)에 의하여 정제함으로써 세균 생산 세포로부터 생산된다.
- [0104] 상기 단백질을 RNAs와 1:2(PpCas9-NLS : sgRNA) 비율로 혼합하고, 상기 혼합물을 실온에서 10분 동안 인큐베이션한 다음, 세포 내로 트랜스펙션한다.
- [0105] 다음, 그로부터 추출된 DNA를 표적 DNA 부위에서 삽입/결실에 대하여 분석한다(상기한 바와 같이).
- [0106] 본 발명에서 규명된 세균 파스퇴렐라 뉴모트로피카로부터의 PpCas9 뉴클레아제는, DNA 변형을 위하여, 당업자에 게 공지된 표준 접근 및 방법을 사용하여 다양한 기원의 세포에 전달될 수 있다. PpCas9는 이전에 규명된 Cas9 단백질에 비하여 많은 이점들을 가진다.
- [0107] PpCas9는 시스템이 작동하는데 요구되는, 다른 공지의 Cas 뉴클레아제와 구분되는, 짧은 2-문자 PAM을 가진다. 본 발명은 프로토스페이스로부터 4 뉴클레오티드 떨어져 위치하는 짧은 PAM (RTT)의 존재로 PpCas9가 시험관내 성공적으로 기능하기에 충분함을 보였다.
- [0108] 이중 가닥 절단을 DNA 내로 도입할 수 있는, 지금까지 알려진 많은 작은 크기의 Cas 뉴클레아제는 복잡한 복수-

문자 PAM 서열을 가져, 절단에 적합한 서열의 선택을 제한한다. 짧은 PAMs을 인식하는, 지금까지 연구된 Cas 뉴클레아제 중, PpCas9만이 RTT 모티프에 의하여 플랭킹된 서열을 인식할 수 있다.

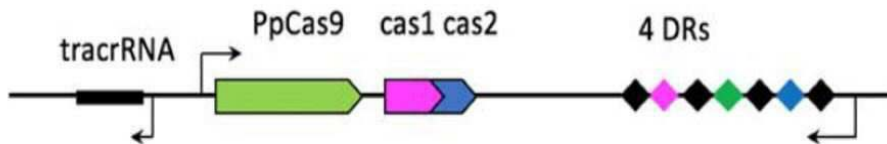
[0109] PpCas9의 두번째 이점은 작은 단백질 크기이다(1055 aar). 지금까지 연구된 작은 크기 단백질은 3-문자 RTT PAM 서열을 가지는 것이다.

[0110] PpCas9는 현재까지 알려진 다른 뉴클레아제의 PAM 서열과 다른, 짧고 사용이 용이한 PAM을 가지는 작은 크기의 Cas 뉴클레아제이다. 상기 PpCas9 단백질은 37°C에서 인간 세포 내 게놈 DNA를 포함하는, 다양한 DNA 표적을 고 효율로 절단하고, 새로운 게놈 편집 도구를 위한 기초가 될 수 있다.

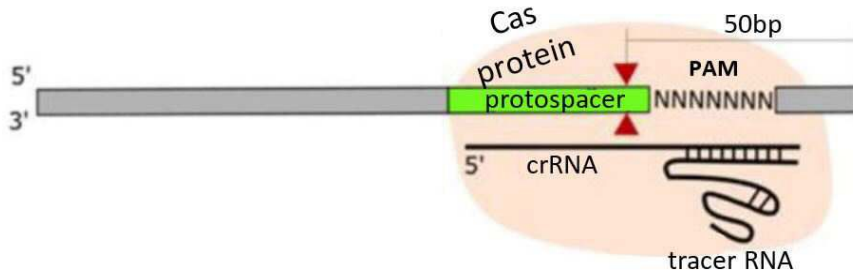
[0111] 본 발명을 개시된 구현예를 참조로 기재하였으나, 당업자는 상세히 기재된 특정 구현예들은 본 발명을 예시할 목적으로 제공된 것이며 어떠한 방식으로든 본 발명의 범위를 제한하는 것으로 해석되지 않음을 이해할 것이다. 본 발명의 사상으로부터 이탈됨이 없이 다양한 변형이 이루어질 수 있는 것으로 이해될 것이다.

도면

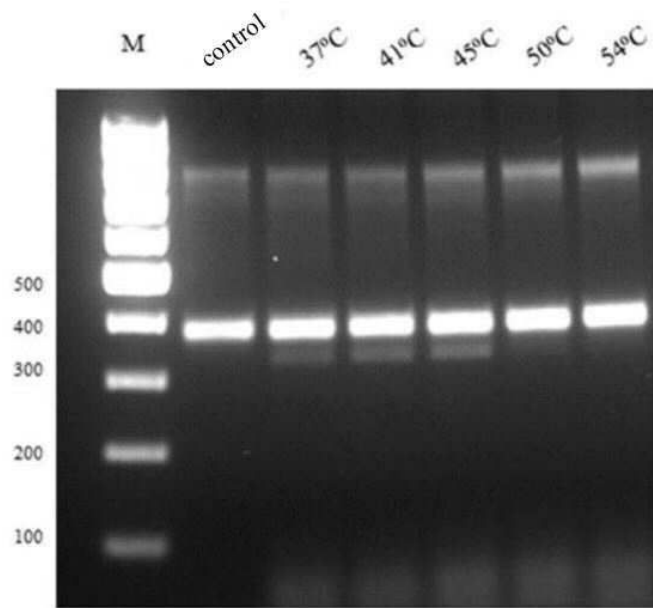
도면1



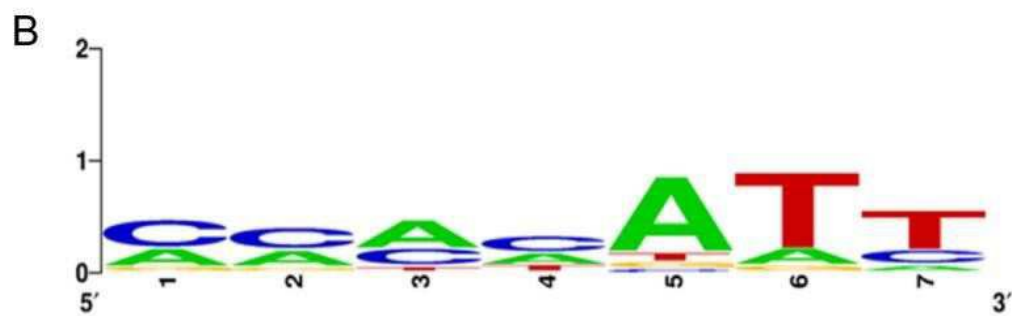
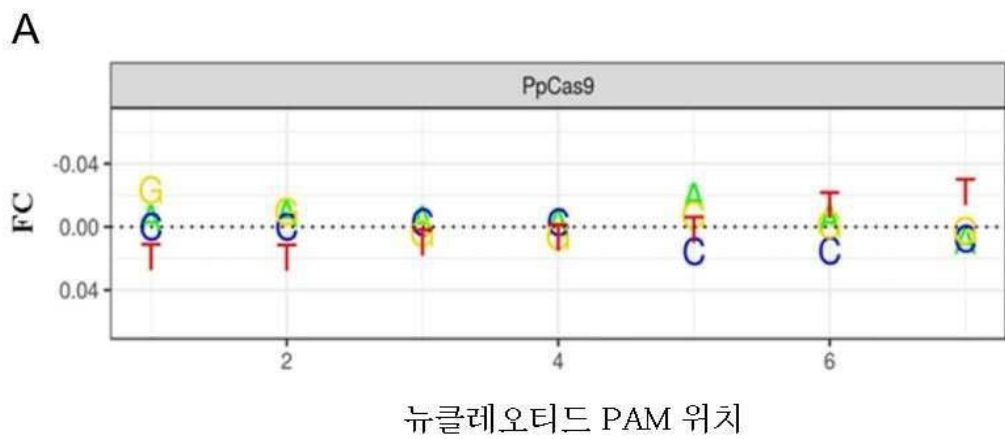
도면2



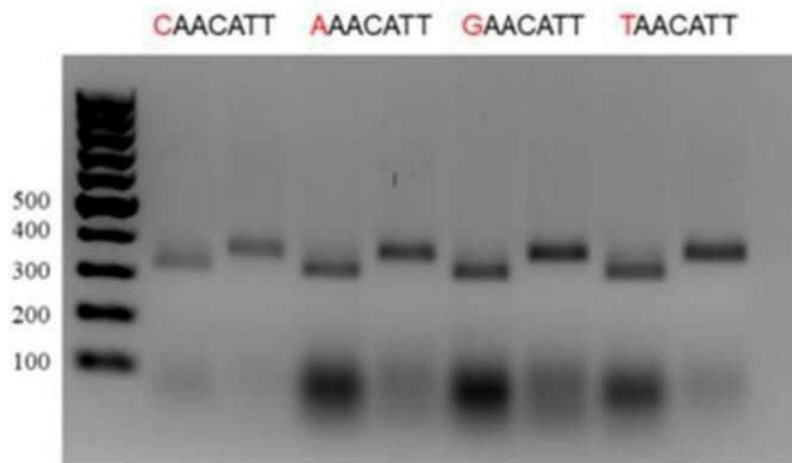
도면3



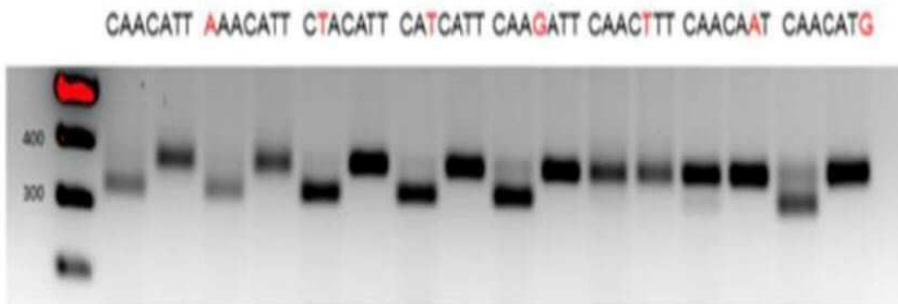
도면4



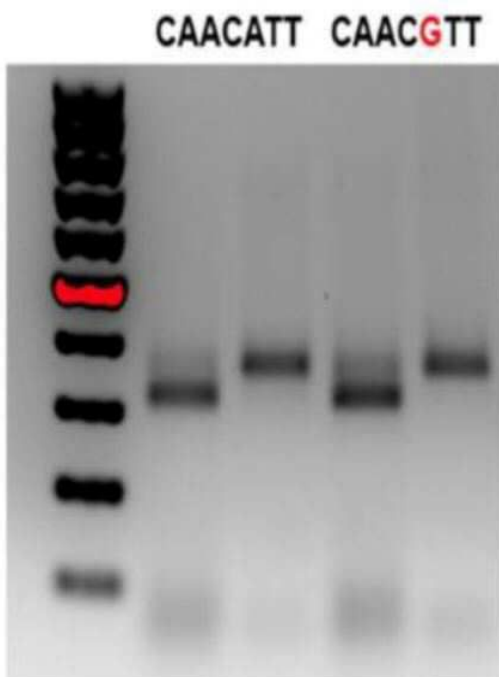
도면5



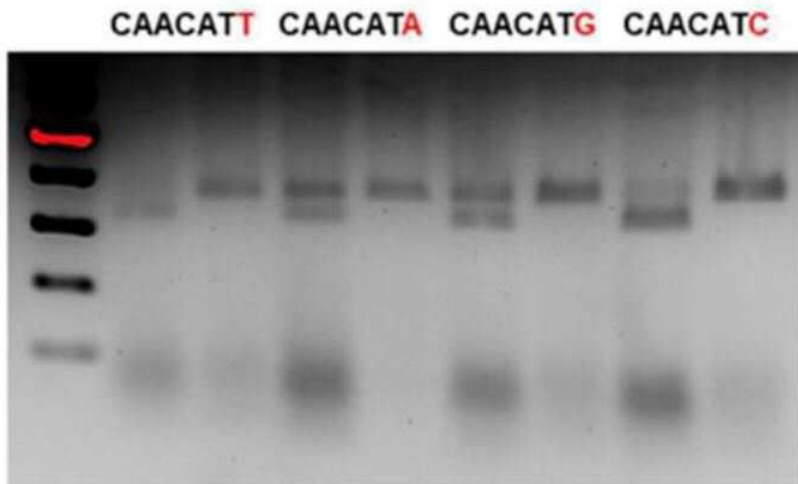
도면6



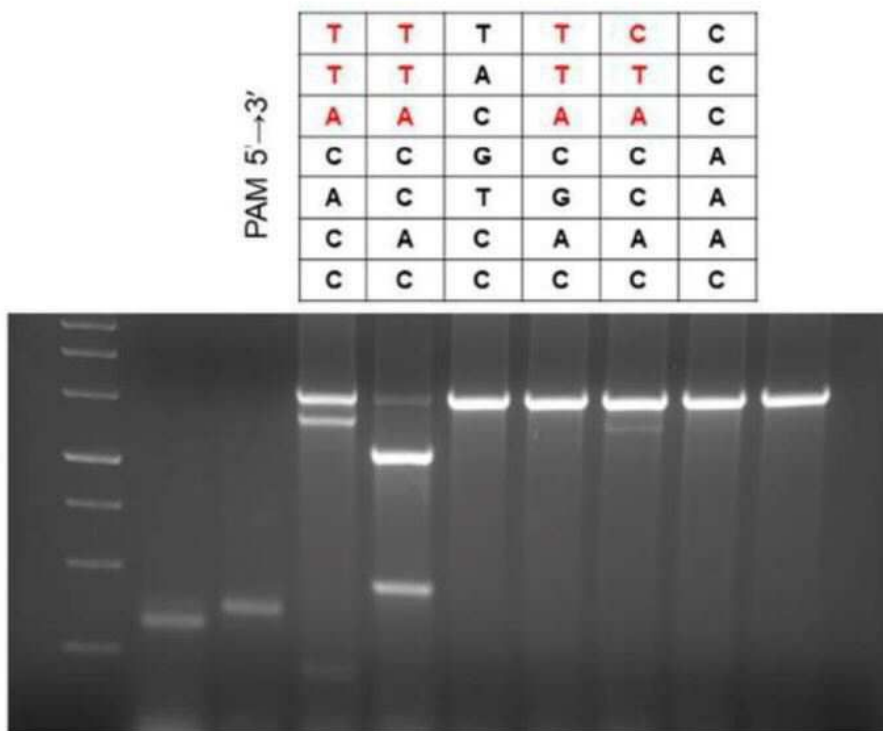
도면7



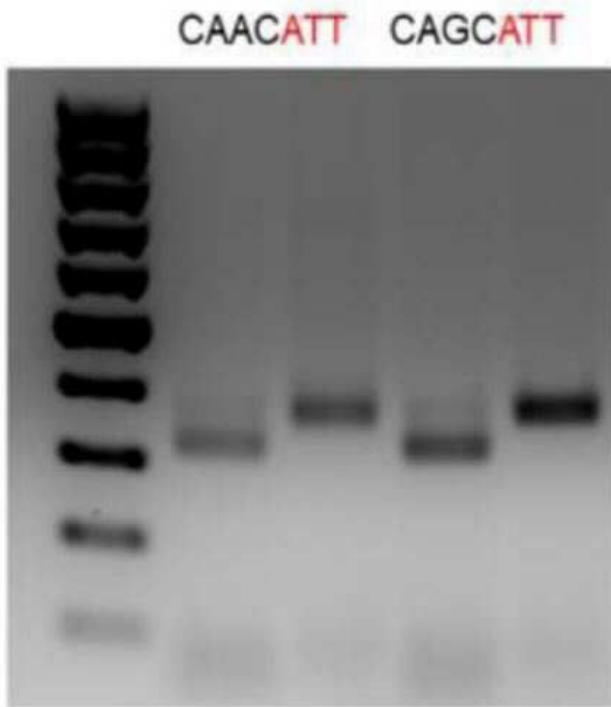
도면8



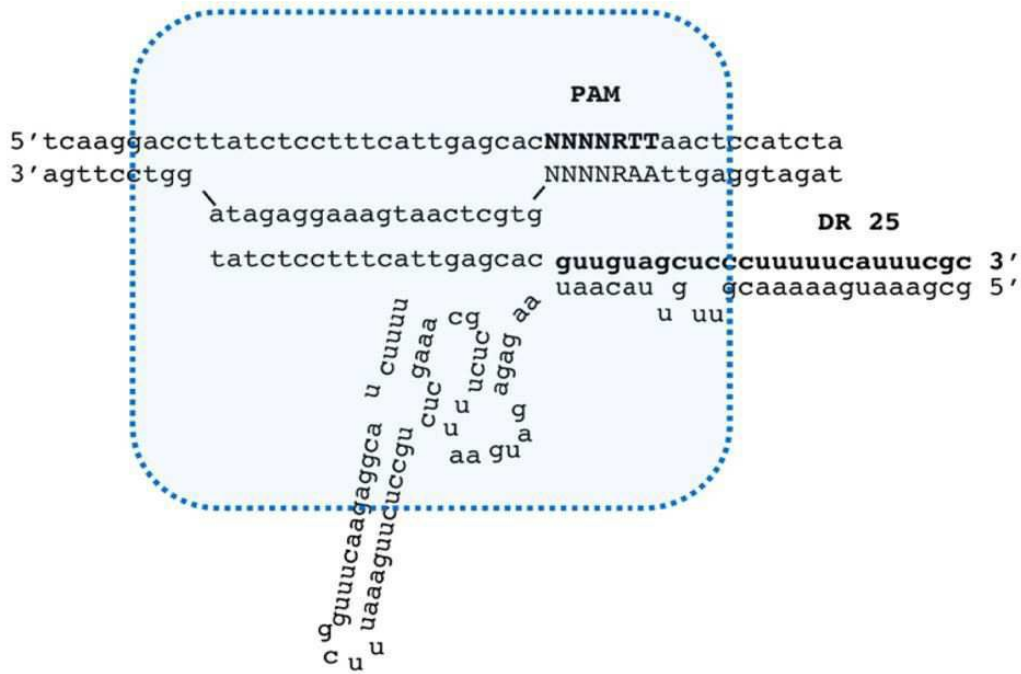
도면9



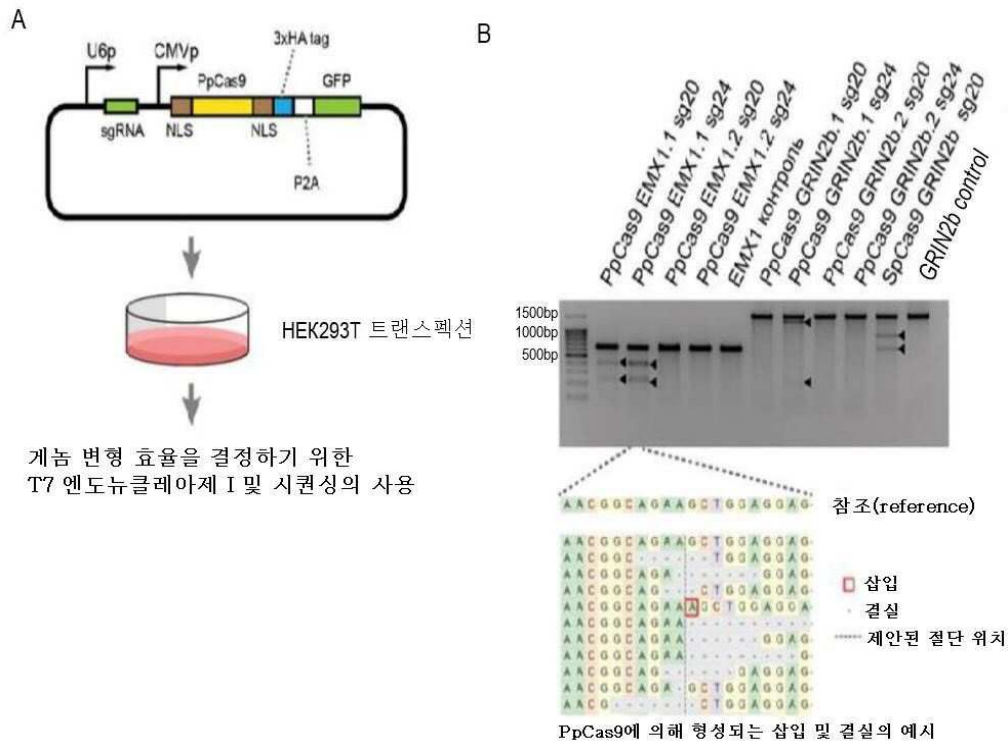
도면10



도면11



도면14



서열 목록

Sequence listing

<110> JSC BIOCAD

<120> USE OF CAS9 PROTEIN FROM THE BACTERIUM PASTEURELLA PNEUMOTROPICA

<150> RU 2019136164

<151> 2019-11-11

<160> 2

<210> 1

<211> 1055

<212> PRT

<213> Pasteurella pneumotropica

<220>

<223> protein homologous to Cas9

<400> 1

Met Gln Asn Asn Pro Leu Asn Tyr Ile Leu Gly Leu Asp Leu Gly Ile

1 5 10 15

Ala Ser Ile Gly Trp Ala Val Val Glu Ile Asp Glu Glu Ser Ser Pro

20 25 30

Ile Arg Leu Ile Asp Val Gly Val Arg Thr Phe Glu Arg Ala Glu Val
 35 40 45
 Ala Lys Thr Gly Glu Ser Leu Ala Leu Ser Arg Arg Leu Ala Arg Ser
 50 55 60
 Ser Arg Arg Leu Ile Lys Arg Arg Ala Glu Arg Leu Lys Lys Ala Lys
 65 70 75 80
 Arg Leu Leu Lys Ala Glu Lys Ile Leu His Ser Ile Asp Glu Lys Leu
 85 90 95

 Pro Ile Asn Val Trp Gln Leu Arg Val Lys Gly Leu Lys Glu Lys Leu
 100 105 110
 Glu Arg Gln Glu Trp Ala Ala Val Leu Leu His Leu Ser Lys His Arg
 115 120 125
 Gly Tyr Leu Ser Gln Arg Lys Asn Glu Gly Lys Ser Asp Asn Lys Glu
 130 135 140
 Leu Gly Ala Leu Leu Ser Gly Ile Ala Ser Asn His Gln Met Leu Gln
 145 150 155 160

 Ser Ser Glu Tyr Arg Thr Pro Ala Glu Ile Ala Val Lys Lys Phe Gln
 165 170 175
 Val Glu Glu Gly His Ile Arg Asn Gln Arg Gly Ser Tyr Thr His Thr
 180 185 190
 Phe Ser Arg Leu Asp Leu Leu Ala Glu Met Glu Leu Leu Phe Gln Arg
 195 200 205
 Gln Ala Glu Leu Gly Asn Ser Tyr Thr Ser Thr Thr Leu Leu Glu Asn
 210 215 220

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

Lys Lys Phe Lys Glu Met Phe Pro His Phe Val Gly Glu Pro Lys Gly
 530 535 540

 Lys Asp Ile Leu Lys Met Arg Leu Tyr Glu Leu Gln Gln Ala Lys Cys
 545 550 555 560
 Leu Tyr Ser Gly Lys Ser Leu Glu Leu His Arg Leu Leu Glu Lys Gly
 565 570 575
 Tyr Val Glu Val Asp His Ala Leu Pro Phe Ser Arg Thr Trp Asp Asp
 580 585 590
 Ser Phe Asn Asn Lys Val Leu Val Leu Ala Asn Glu Asn Gln Asn Lys
 595 600 605

 Gly Asn Leu Thr Pro Tyr Glu Trp Leu Asp Gly Lys Asn Asn Ser Glu
 610 615 620
 Arg Trp Gln His Phe Val Val Arg Val Gln Thr Ser Gly Phe Ser Tyr
 625 630 635 640
 Ala Lys Lys Gln Arg Ile Leu Asn His Lys Leu Asp Glu Lys Gly Phe
 645 650 655
 Ile Glu Arg Asn Leu Asn Asp Thr Arg Tyr Val Ala Arg Phe Leu Cys
 660 665 670

 Asn Phe Ile Ala Asp Asn Met Leu Leu Val Gly Lys Gly Lys Arg Asn
 675 680 685
 Val Phe Ala Ser Asn Gly Gln Ile Thr Ala Leu Leu Arg His Arg Trp
 690 695 700
 Gly Leu Gln Lys Val Arg Glu Gln Asn Asp Arg His His Ala Leu Asp
 705 710 715 720
 Ala Val Val Val Ala Cys Ser Thr Val Ala Met Gln Gln Lys Ile Thr
 725 730 735

 Arg Phe Val Arg Tyr Asn Glu Gly Asn Val Phe Ser Gly Glu Arg Ile
 740 745 750
 Asp Arg Glu Thr Gly Glu Ile Ile Pro Leu His Phe Pro Ser Pro Trp
 755 760 765
 Ala Phe Phe Lys Glu Asn Val Glu Ile Arg Ile Phe Ser Glu Asn Pro

Val Lys Leu Ala Ile Ser Leu Glu Lys Tyr Gln Val Asp Glu Leu Gly
 1025 1030 1035 1040
 Lys Asn Ile Arg Pro Cys Arg Pro Thr Lys Arg Gln His Val Arg
 1045 1050 1055

<210> 2

<211> 3168

<212> DNA

<213> Pasteurella pneumotropica

<220>

<223> protein homologous to Cas9

<400> 2

```

atgcaaaata atcattaaa ttacatttta gggtagatt taggcattgc ttctattggt      60
tgggcggttg tggaaattga tgaggagagt tcacctattc gcttaattga tgtggcgtc    120
cgtacatttg aacgggctga agtcgctaaa accggcgaaa gtttagcatt gtctcgtcgt    180
ttagctcgtt catcacggcg attaattaaa cgccgagcag agcgattaaa aaaagcaaaa    240
cgtttattaa aagcagaaaa gattttacat tctattgatg aaaaattacc cattaatggt    300
tggcagcttc gagtaaaagg attgaaggaa aaactcgaac gtcaggagtg ggcagcggtt    360

ttattacatt tgtcaaagca tcgtggctat ttatcacaac gtaaaaatga gggtaaaagt    420
gataataaag agctgggggc attactttca ggtatcgcaa gtaaccacca aatggtgcaa    480
tcctccgaat atcgtacccc tgcagaaatt gcagtcaaaa aatttcaagt agaagaagga    540
catattcgta atcaacgtgg atcttatacc cataccttta gccgtttga tttgttggca    600
gaaatggaat tattatttca acgccaagct gagttaggca attcttacac gtccaccaca    660
ttattagaaa atttgacggc gttactaatg tggcaaaagc cagctcttgc gggatgatgcg    720
attttaaaaa tgttgggcaa gtgtaccttc gaaccagcg aatataaagc cgcaaaaaat    780

agttattctg ctgaacgttt tgtgtgggta accaagetga ataatttagc cattttagaa    840
aatggcacgg aaagagcttt aatgacaat gaacgttttg ctttgcttga gcaaccgtat    900
gagaaatcaa aattaactta tgctcaagtg agagcaatgc ttgcgttatc tgataatgct    960
attttcaaag gggttcgtta tttaggcgaa gataaaaaaa cagtagagag caaaactacg   1020
ttgatagaaa tgaagtttta tcatcaaatc cgcaaacat taggcagtgc agaattaaaa   1080
aaggaatgga atgagttaaa aggcaattcc gatttattag atgagattgg cacggcattt   1140
tcgttgataa aaacggatga tgatatttgc cgttatttag agggaaaact accagaaagg   1200
    
```

gtattaaatg cgttattgga aaatttaaat ttcgataaat ttattcaact ttcacttaaa 1260
gccttacacc aaatthtacc attgatgctg caagggcaac gttatgatga ggcggtttct 1320
gcgatttatg gtgatcatta tggtaaaaaa tcgacagaaa caaccgctt gttgcccact 1380
attcctgccg atgaaatccg aaatcctgtg gtattacgca ccctgaccca agcccgtaaa 1440
gtgatcaatg cggtggtgcg gttatatggt tcgctgccc gtattcatat tgaacacgcg 1500
agagaagtgc gcaaatctta ccaagatcgt aaaaaacttg aaaaacagca agaagataat 1560
cgtaagcaac gtgaaagtgc ggtcaaaaaa ttaaaagaaa tgtttccgca ttttgtgggg 1620

gagccgaaag gtaaagatat ttaaaaaatg cgattgtatg agttacaaca agcgaatgt 1680
ttatatctg gaaaatcttt agaacttcat cgtttgcttg agaaggggta tgtagaagtg 1740
gatcacgctt tgccattttc tcgcacgtgg gatgatagct ttaataataa agtactggtg 1800
cttgccaacg agaacaaaaa taaaggcaat ttaacgcctt atgaatggtt agatggtaaa 1860
aataacagtg agcgttgga acattttgtt gtacgagtac aaaccagcgg tttctcttat 1920
gctaaaaaac aacgcatttt gaaccataaa ttggatgaaa aagggtttat cgaacgtaat 1980
ttaaacgata ctgctatgt agctcgtttc ttatgtaact ttattgccga taatatgttg 2040

ttggttggtg aaggcaagcg aaactgtttt gcttcaaacg ggcaaatcac ggcgttattg 2100
cggcatcgtt ggggcttaca aaaagtgcgt gaacagaatg atcgccacca cgcaactggac 2160
gcggttggg tggtctgctc tactgtggca atgcaacaaa aatcactcg atttgtgaga 2220
tatacgaag gaaatgtctt tagcggtgaa cgtatcgatc gtgaaactgg cgagattatt 2280
ccattacatt ttccaagccc ttgggctttt ttcaaagaga atgtggaat tcgcattttt 2340
agtgaaaatc caaaattgga attagaaaat cgctgcctg attatccgca atataatcac 2400
gaatgggtgc aaccattgtt tgtttcgaga atgccaaccc gaaaaatgac agggcaaggg 2460

catatggaaa cggtaaaatc cgcaaaacga ttaaatgaag gtttaagtgt gttaaaagtc 2520
cctttaacac aacttaaat gagtgattta gaacgaatgg ttaatcgtga tcgtgaaatt 2580
gcattgtatg aatccttaaa agcacgttta gagcaatttg gtaacgacce agccaaagcc 2640
tttgccgaac cattctataa aaagggtggg gcattagtca aagcagtccg attggaacaa 2700
acacaaaaat cgggggtatt agtacgtgat ggtaacggtg ttgaggataa tgcttcaatg 2760
gtacgggttg atgtttttac taaagggtga aaatatttct tagtgccgat ttatacttgg 2820
caggtagcga aagggtttt accgaatagg gctgcgacac aaggtaaaga tgaatatgat 2880

tgggatatta tggatgaaat ggctactttc caatthtctc tatgtcaaaa tgatctaatt 2940
aaattagtta ccaaaaagaa aacaatcttt ggatatttta atggattaaa tagagctact 3000
agcaataaa atattaaaga gcatgatcta gataagtcta aagggaatt aggtatttac 3060

ttagaagttg gtgtaaaact agctatttcc cttgaaaagt accaagtcga cgaactcggc 3120
aaaaatatcc gtccttgctg tccgactaaa cgacagcacg tgcgttaa 3168