



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0097408
(43) 공개일자 2022년07월07일

- | | |
|---|--|
| <p>(51) 국제특허분류(Int. Cl.)
C12N 9/22 (2006.01) C07K 14/20 (2006.01)
C12N 15/10 (2017.01) C12N 15/113 (2010.01)
C12N 9/78 (2006.01)</p> <p>(52) CPC특허분류
C12N 9/22 (2013.01)
C07K 14/20 (2013.01)</p> <p>(21) 출원번호 10-2022-7015955</p> <p>(22) 출원일자(국제) 2020년10월15일
심사청구일자 없음</p> <p>(85) 번역문제출일자 2022년05월12일</p> <p>(86) 국제출원번호 PCT/US2020/055659</p> <p>(87) 국제공개번호 WO 2021/076682
국제공개일자 2021년04월22일</p> <p>(30) 우선권주장
62/916,392 2019년10월17일 미국(US)</p> | <p>(71) 출원인
페어와이즈 플랜츠 서비시즈, 인크.
미국 27709 노쓰 캐롤라이나주 리서치 트라이앵글
파크 티더블유 알렉산더 드라이브 110</p> <p>(72) 발명자
와츠, 조셉 매튜
미국 27519 노쓰 캐롤라이나주 캐리 리터 코트
115</p> <p>잘리, 사티아 쉼라
미국 27519 노쓰 캐롤라이나주 캐리 오버컵 오크
레인 4121</p> <p>(74) 대리인
양영준, 이상남</p> |
|---|--|

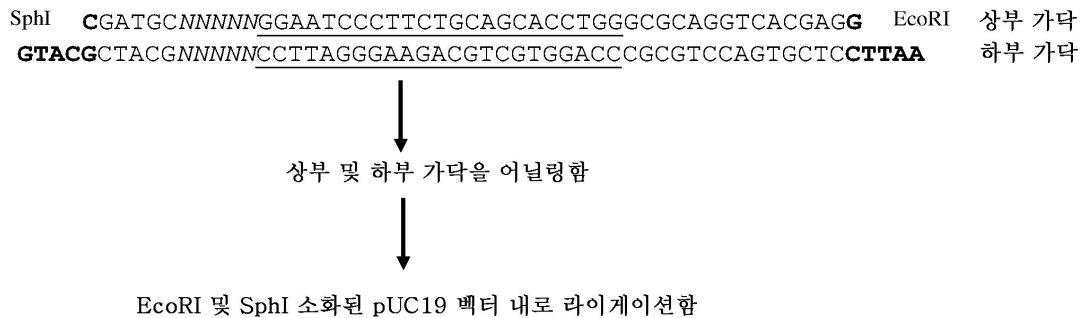
전체 청구항 수 : 총 60 항

(54) 발명의 명칭 CAS12A 뉴클레아제의 변이체 및 그의 제조 방법 및 용도

(57) 요약

본 발명은 변경된 프로토스페이서 인접 모티프 인식 특이성을 갖는 Cas12a 뉴클레아제의 변이체에 관한 것이다. 본 발명은 추가로 CRISPR-CAS 뉴클레아제 변이체를 제조하는 방법 및 변이체를 사용하여 핵산을 변형시키는 방법에 관한 것이다.

대표도



(52) CPC특허분류

C12N 15/102 (2013.01)

C12N 15/1093 (2013.01)

C12N 15/113 (2013.01)

C12N 9/78 (2013.01)

C12Y 305/04004 (2013.01)

C12Y 305/04005 (2013.01)

C07K 2319/70 (2013.01)

C12N 2310/20 (2017.05)

명세서

청구범위

청구항 1

변형된 *라크노스피라세아에*(*Lachnospiraceae*) 박테리아 CRISPR (클러스터링된 규칙적 간격의 짧은 회문식 반복부) Cas12a (LbCas12a) 폴리펩티드이며, **서열식별번호:1** (LbCas12a)의 아미노산 서열과 적어도 80% 동일성을 갖는 아미노산 서열 및 **서열식별번호:1**의 위치 넘버링에 관하여 하기 위치 K116, K120, K121, D122, E125, T148, T149, T152, D156, E159, Q529, D535, K538, D541, Y542, L585, K591, M592, K595, V596, S599, K600, K601, Y616, Y646, W649 중 하나 또는 하나 초과에 돌연변이, 임의로 **서열식별번호:1**의 위치 넘버링에 관하여 하기 위치 K116, K120, K121, D122, E125, T152, D156, E159, G532, D535, K538, D541, 및/또는 K595 중 하나 또는 하나 초과에 돌연변이를 포함하는 변형된 LbCas12a 폴리펩티드.

청구항 2

제1항에 있어서, 변형된 LbCas12a 폴리펩티드가 **서열식별번호:1**의 위치 넘버링에 관하여 K116R, K116N, K120R, K120H, K120N, K120T, K120Y, K120Q, K121S, K121T, K121H, K121R, K121G, K121D, K121Q, D122R, D122K, D122H, D122E, D122N, E125R, E125K, E125Q, E125Y, T148H, T148S, T148A, T148C, T149A, T149C, T149S, T149G, T149H, T149P, T149F, T149N, T149D, T149V, T152R, T152K, T152W, T152Y, T152H, T152Q, T152E, T152L, T152F, D156R, D156K, D156Y, D156W, D156Q, D156H, D156I, D156V, D156L, D156E, E159K, E159R, E159H, E159Y, E159Q, Q529N, Q529T, Q529H, Q529A, Q529F, Q529G, Q529Q, Q529S, Q529P, Q529W, Q529D, G532D, G532N, G532S, G532H, G532F, G532K, G532R, G532Q, G532A, G532L, G532C, D535N, D535H, D535V, D535T, D535,S D535A, D535W, D535K, K538R K538V, K538Q, K538W, K538Y, K538F, K538H, K538L, K538M, K538C, K538G, K538A, K538P, D541N, D541H, D541R, D541K, D541Y, D541I, D541A, D541S, D541E, Y542R, Y542K, Y542H, Y542Q, Y542F, Y542L, Y542M, Y542P, Y542V, Y542N, Y542T, L585G, L585H, L585F, K591W, K591F, K591Y, K591H, K591R, K591S, K591A, K591G, K591P, M592R, M592K, M592Q, M592E, M592A, K595R, K595Q, K595Y, K595L, K595W, K595H, K595E, K595S, K595D, K595M, V596T, V596H, V596G, V596A, S599G, S599H, S599N, S599D, K600R, K600H, K600G, K601R, K601H, K601Q, K601T, Y616K, Y616R, Y616E, Y616F, Y616H, Y646R, Y646E, Y646K, Y646H, Y646Q, Y646W, Y646N, W649H, W649K, W649Y, W649R, W649E, W649S, W649V, 및/또는 W649T의 아미노산 돌연변이 중 하나 또는 하나 초과를 포함하고, 임의로 **서열식별번호:1**의 위치 넘버링에 관하여 K116R, K116N, K120Y, K121S, K121R, D122H, D122N, E125K, T152R, T152K, T152Y, T152Q, T152E, T152F, D156R, D156W, D156Q, D156H, D156I, D156V, D156L, D156E, E159K, E159R, G532N, G532S, G532H, G532K, G532R, G532L, D535N, D535H, D535T, D535,S D535A, D535W, K538R K538V, K538Q, K538W, K538Y, K538F, K538H, K538L, K538M, K538C, K538G, K538A, D541E, K595R, K595Q, K595Y, K595W, K595H, K595S, 및/또는 K595M의 아미노산 돌연변이 중 하나 또는 하나 초과를 포함하는 것인 변형된 LbCas12a 폴리펩티드.

청구항 3

제1항 또는 제2항에 있어서, 변형된 LbCas12a 폴리펩티드가 변경된 PAM (프로토스페이서 인접 모티프) 특이성을 갖는 것인 변형된 LbCas12a 폴리펩티드.

청구항 4

제1항 또는 제2항에 있어서, 뉴클레아제 활성 부위 (예를 들어, RuvC 도메인)에 돌연변이를 추가로 포함하는 (예를 들어, 죽은LbCas12a, dLbCas12a) 변형된 LbCas12a 폴리펩티드.

청구항 5

제1항 내지 제4항 중 어느 한 항의 변형된 LbCas12a 폴리펩티드, 및 링커, 임의로 펩티드 링커를 포함하는 융합 단백질.

청구항 6

제5항에 있어서, 펩티드 링커가 변형된 LbCas12a 폴리펩티드의 C-말단 및/또는 N-말단에 연결된 것인 융합 단백질.

청구항 7

제1항 내지 제4항 중 어느 한 항의 변형된 LbCas12a 폴리펩티드, 및 관심의 폴리펩티드를 포함하는 융합 단백질.

청구항 8

제7항에 있어서, 관심의 폴리펩티드가 변형된 LbCas12a 폴리펩티드에 연결된 것인 융합 단백질.

청구항 9

제7항 또는 제8항에 있어서, 관심의 폴리펩티드가 변형된 LbCas12a 폴리펩티드의 C-말단 및/또는 N-말단에 연결된 것인 융합 단백질.

청구항 10

제7항 내지 제9항 중 어느 한 항에 있어서, 관심의 폴리펩티드가 링커, 임의로 펩티드 링커를 통해 변형된 LbCas12a 폴리펩티드에 연결된 것인 융합 단백질.

청구항 11

제7항 내지 제9항 중 어느 한 항에 있어서, 관심의 폴리펩티드가 데아미나제 (탈아미노화) 활성, 니카제 활성, 리코비나제 활성, 트랜스포사제 활성, 메틸라제 활성, 글리코실라제 (DNA 글리코실라제) 활성, 글리코실라제 억제제 활성 (예를 들어, 우라실-DNA 글리코실라제 억제제 (UGI)), 데메틸라제 활성, 전사 활성화 활성, 전사 억제 활성, 전사 방출 인자 활성, 히스톤 변형 활성, 뉴클레아제 활성, 단일-가닥 RNA 절단 활성, 이중-가닥 RNA 절단 활성, 제한 엔도뉴클레아제 활성 (예를 들어, Fok1), 핵산 결합 활성, 메틸트랜스퍼라제 활성, DNA 복구 활성, DNA 손상 활성, 디스뮤타제 활성, 알킬화 활성, 탈퓨린화 활성, 산화 활성, 피리미딘 이량체 형성 활성, 인테그라제 활성, 트랜스포사제 활성, 폴리머라제 활성, 리가제 활성, 헬리카제 활성, 및/또는 포토리아제 활성을 갖는 적어도 1종의 폴리펩티드 또는 단백질을 포함하는 것인 융합 단백질.

청구항 12

제7항 내지 제11항 중 어느 한 항에 있어서, 관심의 폴리펩티드가 데아미나제 활성을 갖는 적어도 1종의 폴리펩티드 또는 단백질 도메인을 포함하는 것인 융합 단백질.

청구항 13

제12항에 있어서, 데아미나제 활성을 갖는 적어도 1종의 폴리펩티드 또는 단백질 도메인이 시토신 데아미나제 도메인 또는 아데닌 데아미나제 도메인인 융합 단백질.

청구항 14

제13항에 있어서, 시토신 데아미나제 도메인이 아포지질단백질 B mRNA 편집 촉매 폴리펩티드-유사 (APOBEC) 도메인인 융합 단백질.

청구항 15

제13항에 있어서, 아데닌 데아미나제 도메인이 TadA (tRNA-특이적 아데노신 데아미나제) 및/또는 TadA* (진화된 tRNA-특이적 아데노신 데아미나제)인 융합 단백질.

청구항 16

제7항 내지 제14항 중 어느 한 항에 있어서, 적어도 1종의 폴리펩티드가 글리코실라제 억제제 활성을 갖고, 임의로 적어도 1종의 폴리펩티드가 우라실-DNA 글리코실라제 억제제 (UGI)인 융합 단백질.

청구항 17

제1항 내지 제4항 중 어느 한 항의 변형된 LbCas12a 폴리펩티드, 또는 제5항 내지 제16항 중 어느 한 항의 융합 단백질을 코딩하는 폴리뉴클레오티드.

청구항 18

제17항에 있어서, 변형된 LbCas12a 폴리펩티드를 코딩하는 폴리뉴클레오티드 또는 융합 단백질을 코딩하는 폴리뉴클레오티드가 프로모터와 작동적으로 회합되고, 임의로 프로모터가 인트론을 포함하는 프로모터 영역인 폴리뉴클레오티드.

청구항 19

제17항 또는 제18항에 있어서, 폴리뉴클레오티드가 유기체에서의 발현을 위해 코돈 최적화된 것인 폴리뉴클레오티드.

청구항 20

제17항 내지 제19항 중 어느 한 항에 있어서, 유기체가 동물, 식물, 진균, 고세균, 또는 박테리아인 폴리뉴클레오티드.

청구항 21

제1항 내지 제4항 중 어느 한 항의 변형된 LbCas12a 폴리펩티드, 또는 제5항 내지 제16항 중 어느 한 항의 융합 단백질, 및 가이드 핵산 (예를 들어, CRISPR RNA, CRISPR DNA, crRNA, crDNA)을 포함하는 복합체.

청구항 22

제21항의 복합체를 코딩하는 핵산 구축물.

청구항 23

(a) 제1항 내지 제4항 중 어느 한 항의 변형된 LbCas12a 폴리펩티드, 또는 제5항 내지 제16항 중 어느 한 항의 융합 단백질 및 (b) 가이드 핵산을 포함하는 조성물.

청구항 24

제17항 내지 제20항 중 어느 한 항의 폴리뉴클레오티드, 또는 제22항의 핵산 구축물을 포함하는 발현 카세트 또는 벡터.

청구항 25

(a) (i) 제1항 내지 제4항 중 어느 한 항의 변형된 LbCas12a 폴리펩티드 또는 제1항 내지 제4항 중 어느 한 항의 변형된 LbCas12a 폴리펩티드를 코딩하는 핵산, 및 (ii) 관심의 폴리펩티드 또는 관심의 폴리펩티드를 코딩하는 핵산을 포함하는 융합 단백질; 및

(b) 스페이스 서열 및 반복 서열을 포함하는 가이드 핵산 (CRISPR RNA, CRISPR DNA, crRNA, crDNA)을 포함하며, 여기서 가이드 핵산은 변형된 LbCas12a 폴리펩티드 또는 융합 단백질과 복합체를 형성할 수 있고, 스페이스 서열은 표적 핵산에 혼성화하여, 변형된 LbCas12a 폴리펩티드 및 관심의 폴리펩티드를 표적 핵산에 가이드할 수 있고, 그에 의해 표적 핵산은 변형되거나 (예를 들어, 절단 또는 편집) 또는 조정되는 (예를 들어, 전사를 조정함) 것인

유형 V 클러스터링된 규칙적 간격의 짧은 회문식 반복부 (CRISPR) - 연관된 (Cas) (CRISPR-Cas) 시스템.

청구항 26

제25항에 있어서, 변형된 LbCas12a 폴리펩티드가 뉴클레아제 활성 부위 (예를 들어, RuvC 도메인)에 돌연변이를 포함하는 (예를 들어, 죽은Cas12a, dCas12a) 것인 시스템.

청구항 27

제25항 또는 제26항에 있어서, 관심의 폴리펩티드가 변형된 LbCas12a 폴리펩티드의 C-말단 및/또는 N-말단에 연

결된 것인 시스템.

청구항 28

제25항 내지 제27항 중 어느 한 항에 있어서, 관심의 폴리펩티드가 링커, 임의로 펩티드 링커를 통해 변형된 LbCas12a 폴리펩티드의 C-말단 및/또는 N-말단에 연결된 것인 시스템.

청구항 29

제25항 내지 제28항 중 어느 한 항에 있어서, 관심의 폴리펩티드가 데아미나제 (탈아미노화) 활성, 니카제 활성, 리코비나제 활성, 트랜스포사제 활성, 메틸라제 활성, 글리코실라제 (DNA 글리코실라제) 활성, 글리코실라제 억제제 활성 (예를 들어, 우라실-DNA 글리코실라제 억제제 (UGI)), 데메틸라제 활성, 전사 활성화 활성, 전사 억제 활성, 전사 방출 인자 활성, 히스톤 변형 활성, 뉴클레아제 활성, 단일-가닥 RNA 절단 활성, 이중-가닥 RNA 절단 활성, 제한 엔도뉴클레아제 활성 (예를 들어, FokI), 핵산 결합 활성, 메틸트랜스퍼라제 활성, DNA 복구 활성, DNA 손상 활성, 디스뮤타제 활성, 알킬화 활성, 탈퓨린화 활성, 산화 활성, 피리미딘 이량체 형성 활성, 인테그라제 활성, 트랜스포사제 활성, 폴리머라제 활성, 리가제 활성, 헬리카제 활성, 및/또는 포토리아제 활성을 갖는 적어도 1종의 폴리펩티드 또는 단백질 도메인을 포함하는 것인 시스템.

청구항 30

제25항 내지 제29항 중 어느 한 항에 있어서, 관심의 폴리펩티드가 데아미나제 활성을 갖는 적어도 1종의 폴리펩티드 또는 단백질 도메인을 포함하는 것인 시스템.

청구항 31

제30항에 있어서, 데아미나제 활성을 갖는 적어도 1종의 폴리펩티드 또는 단백질 도메인이 시토신 데아미나제 도메인 또는 아데닌 데아미나제 도메인인 시스템.

청구항 32

제31항에 있어서, 시토신 데아미나제 도메인이 아포지질단백질 B mRNA 편집 촉매 폴리펩티드-유사 (APOBEC) 도메인인 시스템.

청구항 33

제31항에 있어서, 아데닌 데아미나제 도메인이 TadA (tRNA-특이적 아데노신 데아미나제) 및/또는 TadA* (진화된 tRNA-특이적 아데노신 데아미나제)인 시스템.

청구항 34

제25항 내지 제32항 중 어느 한 항에 있어서, 관심의 폴리펩티드가 우라실-DNA 글리코실라제 억제제 (UGI)를 포함하는 것인 시스템.

청구항 35

제25항 내지 제34항 중 어느 한 항에 있어서, (a) 및 (b) 중 하나 또는 둘 다가 1종 이상의 발현 카세트 및/또는 벡터에 포함되는 것인 시스템.

청구항 36

제17항 내지 제20항 중 어느 한 항의 폴리뉴클레오티드, 제22항의 핵산 구축물, 제24항의 발현 카세트 또는 벡터, 또는 제25항 내지 제35항 중 어느 한 항의 시스템을 포함하는 세포.

청구항 37

표적 핵산을

(a) (i) 제1항 내지 제4항 중 어느 한 항의 변형된 LbCas12a 폴리펩티드, 또는 제5항 내지 제16항 중 어느 한 항의 융합 단백질, 및 (ii) 가이드 핵산 (예를 들어, CRISPR RNA, CRISPR DNA, crRNA, crDNA);

(b) 제21항의 복합체 및 가이드 핵산;

(c) (i) 제1항 내지 제4항 중 어느 한 항의 변형된 LbCas12a 폴리펩티드, 또는 제5항 내지 제16항 중 어느 한 항의 융합 단백질, 및 (ii) 가이드 핵산을 포함하는 조성물; 및/또는

(d) 제25항 내지 제35항 중 어느 한 항의 시스템

과 접촉시키고, 그에 의해 표적 핵산을 변형시키는 것을 포함하는, 표적 핵산을 변형시키는 방법.

청구항 38

표적 핵산을 포함하는 세포 또는 무세포 시스템을

(a) (i) 제17항 내지 제20항 중 어느 한 항의 폴리뉴클레오티드, 또는 이를 포함하는 발현 카세트 또는 벡터, 및 (ii) 가이드 핵산, 또는 이를 포함하는 발현 카세트 또는 벡터; 및/또는

(b) 제22항의 핵산 구축물, 또는 이를 포함하는 발현 카세트 또는 벡터

와 접촉시키고, 그에 의해 표적 핵산을 변형시키는 것을 포함하는, 표적 핵산을 변형시키는 방법.

청구항 39

표적 핵산을

(a) (i) 제7항 내지 제16항 중 어느 한 항의 융합 단백질, 및 (a) (ii) 가이드 핵산;

(b) 제7항 내지 제16항 중 어느 한 항의 융합 단백질, 및 가이드 핵산을 포함하는 복합체;

(c) 제7항 내지 제16항 중 어느 한 항의 융합 단백질 및 가이드 핵산을 포함하는 조성물; 및/또는

(d) 제25항 내지 제35항 중 어느 한 항의 시스템

과 접촉시키고, 그에 의해 표적 핵산을 편집하는 것을 포함하는, 표적 핵산을 편집하는 방법.

청구항 40

표적 핵산을 포함하는 세포 또는 무세포 시스템을

(a) (i) 제7항 내지 제16항 중 어느 한 항의 융합 단백질을 코딩하는 폴리뉴클레오티드, 또는 이를 포함하는 발현 카세트 또는 벡터, 및 (a) (ii) 가이드 핵산, 또는 이를 포함하는 발현 카세트 또는 벡터; 및/또는

(b) 제7항 내지 제16항 중 어느 한 항의 융합 단백질, 및 가이드 핵산을 포함하는 복합체를 코딩하는 핵산 구축물, 또는 이를 포함하는 발현 카세트 또는 벡터; 및/또는

(c) 제25항 내지 제35항 중 어느 한 항의 시스템

과 접촉시키고, 그에 의해 표적 핵산을 편집하는 것을 포함하는, 표적 핵산을 편집하는 방법.

청구항 41

하기 단계를 포함하는 2개 이상의 이중 가닥 핵산 분자를 제조하는 것, 그에 의해 이중 가닥 핵산 분자를 포함하는 무작위적 DNA 라이브러리를 구축하는 것을 포함하는, 프로토스페이서의 5' 단부에 프로토스페이서 인접 모티프 (PAM) 인식 부위를 갖는 CRISPR-Cas 뉴클레아제의 PAM 요건/특이성을 결정하기 위한 이중 가닥 핵산 분자를 포함하는 무작위적 DNA 라이브러리를 구축하는 방법:

(a) 비-표적 올리고뉴클레오티드 (제1) 가닥 및 표적 올리고뉴클레오티드 (제2) 가닥을 2개 이상의 이중 가닥 핵산 분자의 각각에 대해 합성하는 단계이며, 여기서 비-표적 올리고뉴클레오티드 가닥은 5'에서 3'로

(i) 약 5 내지 약 15개의 뉴클레오티드를 갖는 제1 서열,

(ii) 적어도 4개의 무작위적 뉴클레오티드를 갖는 제2 서열,

(iii) 약 16 내지 약 25개의 뉴클레오티드를 포함하는 프로토스페이서 서열, 및

(iv) 약 5 내지 약 20개의 뉴클레오티드를 갖는 제3 서열

을 포함하고, (i)의 약 5 내지 15개의 뉴클레오타이드를 갖는 제1 서열은 (ii)의 제2 서열의 5' 단부에 바로 인접하고, (ii)의 제2 서열은 (iii)의 프로토스페이서 서열의 5' 단부에 바로 인접하고, 프로토스페이서 서열은 (iv)의 제3 서열의 5' 단부에 바로 인접하고; 표적 올리고뉴클레오타이드 (제2) 가닥은 비-표적 올리고뉴클레오타이드 가닥에 상보적인 것인 단계; 및

(b) 비-표적 올리고뉴클레오타이드 가닥을 상보적 표적 올리고뉴클레오타이드 가닥에 어닐링하여 이중 가닥 핵산 분자를 생산하는 단계이며, 여기서 제1 서열은 제한 부위를 (그의 5' 단부에) 포함하고, 제3 서열은 제한 부위를 (그의 3' 단부에) 포함하고, 2개 이상의 이중 가닥 핵산 분자의 각각의 제1 서열 (i), 프로토스페이서 서열 (iii) 및 제3 서열 (iv)은 동일한 것인 단계.

청구항 42

하기 단계를 포함하는 2개 이상의 이중 가닥 핵산 분자를 제조하는 것, 그에 의해 이중 가닥 핵산 분자를 포함하는 무작위적 DNA 라이브러리를 구축하는 것을 포함하는, 프로토스페이서의 3' 단부에 프로토스페이서 인접 모티프 (PAM) 인식 부위를 갖는 CRISPR-Cas 뉴클레아제의 PAM 요건/특이성을 결정하기 위한 이중 가닥 핵산 분자를 포함하는 무작위적 DNA 라이브러리를 구축하는 방법:

(a) 비-표적 올리고뉴클레오타이드 (제1) 가닥 및 표적 올리고뉴클레오타이드 (제2) 가닥을 2개 이상의 이중 가닥 핵산 분자의 각각에 대해 합성하는 단계이며, 여기서 비-표적 올리고뉴클레오타이드 가닥은 5'에서 3'로

(i) 약 5 내지 약 20개의 뉴클레오타이드를 갖는 제1 서열,

(ii) 약 16 내지 약 25개의 뉴클레오타이드를 포함하는 프로토스페이서 서열,

(iii) 적어도 4개의 무작위적 뉴클레오타이드를 갖는 제2 서열, 및

(iv) 약 5 내지 약 15개의 뉴클레오타이드를 갖는 제3 서열

을 포함하고, (i)의 약 5 내지 20개의 뉴클레오타이드를 갖는 제1 서열은 (ii)의 프로토스페이서 서열의 5' 단부에 바로 인접하고, (iii)의 제2 서열은 (iii)의 프로토스페이서 서열의 3' 단부에 바로 인접하고, (iv)의 제3 서열은 (iii)의 제2 서열의 3' 단부에 바로 인접하고; 표적 올리고뉴클레오타이드 (제2) 가닥은 비-표적 올리고뉴클레오타이드 가닥에 상보적인 것인 단계; 및

(b) 비-표적 올리고뉴클레오타이드 가닥을 상보적 표적 올리고뉴클레오타이드 가닥에 어닐링하여 이중 가닥 핵산 분자를 생산하는 단계이며, 여기서 제1 서열 (i)은 제한 부위를 (그의 5' 단부에) 포함하고, 제3 서열 (iv)은 제한 부위를 (그의 3' 단부에) 포함하고, 2개 이상의 이중 가닥 핵산 분자의 각각의 제1 서열 (i), 프로토스페이서 서열 (ii) 및 제3 서열 (iv)은 동일한 것인 단계.

청구항 43

제41항 또는 제42항에 있어서, 이중 가닥 핵산 분자를 벡터 내로 라이게이션하여 무작위적 DNA 라이브러리를 포함하는 벡터를 생산하는 것을 추가로 포함하는 방법.

청구항 44

제41항 내지 제43항 중 어느 한 항에 있어서, 무작위적 DNA 라이브러리를 증폭시키는 것을 추가로 포함하는 방법.

청구항 45

제44항에 있어서, 증폭이 무작위적 DNA 라이브러리를 포함하는 벡터를 1개 이상의 박테리아 세포 내로 도입하고, 1개 이상의 박테리아 세포를 배양하는 것을 포함하는 것인 방법.

청구항 46

제45항에 있어서, 무작위적 DNA 라이브러리를 포함하는 벡터를 1개 이상의 박테리아 세포로부터 분리하는 것을 추가로 포함하는 방법.

청구항 47

제46항에 있어서, 벡터를 선형화하여 무작위적 DNA 라이브러리를 제공하는 것을 추가로 포함하는 방법.

청구항 48

제47항에 있어서, 선형화가 벡터를 제한 효소와 접촉시키는 것을 포함하는 것인 방법.

청구항 49

제48항에 있어서, 제한 효소가 ScaI 또는 PfoI인 방법.

청구항 50

제41항 내지 제49항 중 어느 한 항의 방법에 의해 생산된 무작위적 DNA 라이브러리.

청구항 51

프로토스페이서의 5' 단부 상에 프로토스페이스 인접 모티프 (PAM) 인식 부위를 갖는 CRISPR-Cas 뉴클레아제의 PAM 요건/특이성을 결정하기 위한 무작위적 DNA 라이브러리이며, 2개 이상의 이중 가닥 핵산 분자를 포함하고, 그의 각각은

(a) 비-표적 올리고뉴클레오타이드 (제1) 가닥 및 표적 올리고뉴클레오타이드 (제2) 가닥 (여기서 비-표적 올리고뉴클레오타이드 가닥은 5'에서 3'로

(i) 약 5 내지 약 15개의 뉴클레오타이드 (예를 들어, 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 또는 15개의 뉴클레오타이드, 및 그 안의 임의의 범위 또는 값)를 갖는 제1 서열,

(ii) 적어도 4개의 무작위적 뉴클레오타이드 (예를 들어, 적어도 4, 5, 6, 7, 8, 9, 10개, 또는 그 초과, 및 그 안의 임의의 범위 또는 값)를 갖는 제2 서열,

(iii) 약 16 내지 약 25개의 뉴클레오타이드, 예를 들어, 약 16, 17, 18, 19, 20, 21, 22, 23, 24, 또는 25개의 뉴클레오타이드를 포함하는 프로토스페이스 서열, 및

(iv) 약 5 내지 약 20개의 뉴클레오타이드 (예를 들어, 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개의 뉴클레오타이드, 및 그 안의 임의의 범위 또는 값)를 갖는 제3 서열을 포함하고, (i)의 약 5 내지 15개의 뉴클레오타이드를 갖는 제1 서열은 (ii)의 제2 서열의 5' 단부에 바로 인접하고, (ii)의 제2 서열은 (iii)의 프로토스페이스 서열의 5' 단부에 바로 인접하고, 프로토스페이스 서열은 (iv)의 제3 서열의 5' 단부에 바로 인접하고; 표적 올리고뉴클레오타이드 (제2) 가닥은 비-표적 올리고뉴클레오타이드 가닥에 상보적임); 및

(b) 상보적 표적 올리고뉴클레오타이드 가닥에 어닐링되어 이중 가닥 핵산 분자를 생산하는 비-표적 올리고뉴클레오타이드 가닥 (여기서 제1 서열은 제한 부위를 (그의 5' 단부에) 포함하고, 제3 서열은 제한 부위를 (그의 3' 단부에) 포함하고, 2개 이상의 이중 가닥 핵산 분자의 각각의 제1 서열 (i), 프로토스페이스 서열 (iii) 및 제3 서열 (iv)은 동일함)

을 포함하는 것인 무작위적 DNA 라이브러리.

청구항 52

프로토스페이서의 3' 단부 상에 프로토스페이스 인접 모티프 (PAM) 인식 부위를 갖는 CRISPR-Cas 뉴클레아제의 PAM 요건/특이성을 결정하기 위한 무작위적 DNA 라이브러리이며, 2개 이상의 이중 가닥 핵산 분자를 포함하고, 그의 각각은

(a) 비-표적 올리고뉴클레오타이드 (제1) 가닥 및 표적 올리고뉴클레오타이드 (제2) 가닥 (여기서 비-표적 올리고뉴클레오타이드 가닥은 5'에서 3'로

(i) 약 5 내지 약 20개의 뉴클레오타이드 (예를 들어, 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개의 뉴클레오타이드, 및 그 안의 임의의 범위 또는 값)를 갖는 제1 서열,

(ii) 약 16 내지 약 25개의 뉴클레오타이드, 예를 들어, 약 16, 17, 18, 19, 20, 21, 22, 23, 24, 또는 25개의 뉴클레오타이드를 포함하는 프로토스페이스 서열,

(iii) 적어도 4개의 무작위적 뉴클레오타이드 (예를 들어, 적어도 4, 5, 6, 7, 8, 9, 10개, 또는 그 초과, 및 그

안의 임의의 범위 또는 값)를 갖는 제2 서열, 및

(iv) 약 5 내지 약 15개의 뉴클레오타이드 (예를 들어, 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15개의 뉴클레오타이드, 및 그 안의 임의의 범위 또는 값)를 갖는 제3 서열을 포함하고, (i)의 약 5 내지 20개의 뉴클레오타이드를 갖는 제1 서열은 (ii)의 프로토스페이서 서열의 5' 단부에 바로 인접하고, (iii)의 제2 서열은 (iii)의 프로토스페이서 서열의 3' 단부에 바로 인접하고, (iv)의 제3 서열은 (iii)의 제2 서열의 3' 단부에 바로 인접하고; 표적 올리고뉴클레오타이드 (제2) 가닥은 비-표적 올리고뉴클레오타이드 가닥에 상보적임); 및

(b) 상보적 표적 올리고뉴클레오타이드 가닥에 어닐링되어 이중 가닥 핵산 분자를 생산하는 비-표적 올리고뉴클레오타이드 가닥 (여기서 제1 서열은 제한 부위를 (그의 5' 단부에) 포함하고, 제3 서열은 제한 부위를 (그의 3' 단부에) 포함하고, 2개 이상의 이중 가닥 핵산 분자의 각각의 제1 서열 (i), 프로토스페이서 서열 (ii) 및 제3 서열 (iv)은 동일함)

을 포함하는 것인 무작위적 DNA 라이브러리.

청구항 53

CRISPR-Cas 뉴클레아제를 제51항 또는 제52항의 무작위적 DNA 라이브러리와 접촉시키는 단계; 및

무작위적 DNA 라이브러리의 이중 가닥 핵산 분자를 뉴클레아제와의 접촉 전에 (대조군) 및 후에 시퀀싱하는 단계이며, 여기서 뉴클레아제와의 접촉 전에 무작위적 DNA 라이브러리에 존재하지만, 뉴클레아제와의 접촉 후에 무작위적 DNA 라이브러리에 부재하는 이중 가닥 핵산 분자는 CRISPR-Cas 뉴클레아제의 프로토스페이서 인접 모티프 (PAM) 인식 서열을 확인하고, 그에 의해 CRISPR-Cas 뉴클레아제의 PAM 특이성을 결정하는 것인 단계

를 포함하는, CRISPR-Cas 뉴클레아제의 PAM 특이성을 결정하는 방법.

청구항 54

제53항에 있어서, 시퀀싱이 차세대 시퀀싱을 포함하는 것인 방법.

청구항 55

제17항 내지 제20항 및 제22항 중 어느 한 항의 폴리뉴클레오타이드, 및/또는 이를 포함하는 발현 카세트 또는 벡터, 및 임의로 이의 사용을 위한 지시서를 포함하는 키트.

청구항 56

제55항에 있어서, CRISPR-Cas12a 가이드 핵산 및/또는 이를 포함하는 발현 카세트 또는 벡터를 추가로 포함하는 키트.

청구항 57

제56항에 있어서, 가이드 핵산이 가이드 핵산의 백본 내로의 표적 핵산 서열과 동일하거나 그에 상보적인 핵산 서열의 클로닝을 위한 클로닝 부위를 포함하는 것인 키트.

청구항 58

제55항 내지 제57항 중 어느 한 항에 있어서, 폴리뉴클레오타이드가 1개 이상의 핵 국제화 신호를 추가로 코딩하고, 1개 이상의 핵 국제화 신호가 CRISPR-Cas12a 뉴클레아제에 융합된 것인 키트.

청구항 59

제55항 내지 제58항 중 어느 한 항에 있어서, 폴리뉴클레오타이드, 발현 카세트 또는 벡터가 1개 이상의 선택가능한 마커를 추가로 코딩하는 것인 키트.

청구항 60

제55항 내지 제59항 중 어느 한 항에 있어서, 폴리뉴클레오타이드가 mRNA이고, 코딩된 CRISPR-Cas12a 뉴클레아제 내의 1개 이상의 인트론을 코딩하는 것인 키트.

발명의 설명

기술 분야

[0001] 서열 목록의 전자적 출원에 관한 진술

[0002] 37 C.F.R. § 1.821 하에 제출되고, 명칭이 1499.7WO_ST25.txt이고, 크기가 257,774 바이트이고, 2020년 10월 13일에 생성되고, EFS-웹을 통해 출원된 ASCII 텍스트 형식의 서열 목록은 종이 사본 대신 제공된다. 이 서열 목록은 그의 개시내용에 대해 본 명세서에 참조로 포함된다.

[0003] 우선권의 진술

[0004] 본 출원은 2019년 10월 17일에 출원된 미국 가출원 번호 62/916,392의 35 U.S.C. § 119 (e) 하의 우선권을 주장하며, 그의 전체 내용은 본원에 참조로 포함된다.

[0005] 발명의 분야

[0006] 본 발명은 변경된 프로토스페이서 인접 모티프 인식 특이성을 갖는 Cas12a CRISPR-Cas 뉴클레아제의 변이체에 관한 것이다. 본 발명은 추가로 CRISPR-CAS 뉴클레아제 변이체를 제조하는 방법 및 변이체를 사용하여 핵산을 변형시키는 방법에 관한 것이다.

배경 기술

[0007] 게놈 편집/변형은 부위-지정된 뉴클레아제, 예를 들어, CRISPR-Cas 뉴클레아제를 이용하여, 표적화된 게놈 위치에 변이를 도입하는 프로세스이다. 게놈 변형을 위해 가장 폭넓게 이용되는 뉴클레아제인 Cas9는 NGG 모티프 (예를 들어, 프로토스페이서 인접 모티프 (PAM))의 상류의 게놈 영역에 돌연변이를 도입할 수 있다. 다른 Cas 뉴클레아제는 상이한 PAM 인식 특이성을 갖는다. 이들 뉴클레아제의 PAM 특이성이 특히 엄격한 경우, 이들은 그 뉴클레아제에 의한 변형을 위해 이용가능한 게놈 표적 부위의 수를 제한함으로써 게놈 변형을 위한 뉴클레아제의 유용성을 감소시킬 수 있다.

[0008] 관련 기술분야의 결점을 다루기 위해, 본 발명은 개선된 PAM 특이성을 갖는 변형된 CRISPR-Cas 뉴클레아제 및 이러한 CRISPR-Cas 뉴클레아제를 디자인하고, 확인하고, 선택하는 방법을 제공한다.

발명의 내용

[0009] 발명의 요약

[0010] 본 발명의 한 측면은 변형된 라크노스피라세아에(*Lachnospiraceae*) 박테리아 CRISPR (클러스터링된 규칙적 간격의 짧은 회문식 반복부) Cas12a (LbCas12a) 폴리펩티드를 제공하며 여기서 변형된 LbCas12a 폴리펩티드는 **서열 식별번호(SEQ ID NO):1** (LbCas12a)의 아미노산 서열과 적어도 80% 동일성을 갖는 아미노산 서열 및 임의의 조합으로 **서열식별번호:1**의 위치 넘버링에 관하여 하기 위치 K116, K120, K121, D122, E125, T148, T149, T152, D156, E159, Q529, G532, D535, K538, D541, Y542, L585, K591, M592, K595, V596, S599, K600, K601, Y616, Y646, 및/또는 W649 중 하나 이상 (예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13개 또는 그 초과)에 돌연변이, 임의로 임의의 조합으로 **서열식별번호:1**의 위치 넘버링에 관하여 하기 위치 K116, K120, K121, D122, E125, T152, D156, E159, G532, D535, K538, D541, 및/또는 K595 중 하나 이상에 돌연변이를 포함하거나, 이로 본질적으로 이루어지거나, 또는 이로 이루어진다.

[0011] 본 발명의 제2 측면은 (a) (i) 본 발명의 변형된 LbCas12a 폴리펩티드 또는 본 발명의 변형된 LbCas12a 폴리펩티드를 코딩하는 핵산, 및 (ii) 관심의 폴리펩티드 또는 관심의 폴리펩티드를 코딩하는 핵산; 및 (b) 스페이서 서열 및 반복 서열을 포함하는 가이드 핵산 (CRISPR RNA, CRISPR DNA, crRNA, crDNA)을 포함하며, 여기서 가이드 핵산은 변형된 LbCas12a 폴리펩티드 또는 융합 단백질과 복합체를 형성할 수 있고, 스페이서 서열은 표적 핵산에 혼성화하여, 변형된 LbCas12a 폴리펩티드 및 관심의 폴리펩티드를 표적 핵산에 가이드할 수 있고, 그에 의해 표적 핵산은 변형되거나 조정되는 것인 유형 V 클러스터링된 규칙적 간격의 짧은 회문식 반복부 (CRISPR) - 연관된 (Cas) (CRISPR-Cas) 시스템을 제공한다.

[0012] 본 발명의 제3 측면은 표적 핵산을 (a) (i) 본 발명의 변형된 LbCas12a 폴리펩티드, 또는 본 발명의 변형된 LbCas12a 폴리펩티드를 포함하는 융합 단백질, 및 (ii) 가이드 핵산 (예를 들어, CRISPR RNA, CRISPR DNA, crRNA, crDNA); (b) 본 발명의 변형된 LbCas12a 폴리펩티드 및 가이드 핵산을 포함하는 복합체; (c) (i) 본 발명의 변형된 LbCas12a 폴리펩티드, 또는 본 발명의 융합 단백질, 및 (ii) 가이드 핵산을 포함하는 조성물; 및/또는 (d) 본 발명의 시스템과 접촉시키고, 그에 의해 표적 핵산을 변형시키는 것을 포함하는, 표적 핵산을 변형

시키는 방법을 제공한다.

- [0013] 본 발명의 제4 측면은 표적 핵산을 포함하는 세포 또는 무세포 시스템을 (a) (i) 본 발명의 변형된 LbCas12a 폴리펩티드를 코딩하는 폴리뉴클레오티드, 또는 이를 포함하는 발현 카세트 또는 벡터, 및 (ii) 가이드 핵산, 또는 이를 포함하는 발현 카세트 또는 벡터; 및/또는 (b) (i) 본 발명의 변형된 LbCas12a 폴리펩티드 또는 융합 단백질을 포함하는 복합체, 및 (ii) 가이드 핵산을 코딩하는 핵산 구축물, 또는 이를 포함하는 발현 카세트 또는 벡터와 접촉시키고, 그에 의해 표적 핵산을 변형시키는 것을 포함하는, 표적 핵산을 변형시키는 방법을 제공한다.
- [0014] 본 발명의 제5 측면은 표적 핵산을 (a) (i) 본 발명의 변형된 LbCas12a 폴리펩티드를 포함하는 융합 단백질 및 (a) (ii) 가이드 핵산; (b) 본 발명의 융합 단백질, 및 가이드 핵산을 포함하는 복합체; (c) 본 발명의 융합 단백질 및 가이드 핵산을 포함하는 조성물; 및/또는 (d) 본 발명의 시스템과 접촉시키고, 그에 의해 표적 핵산을 편집하는 것을 포함하는, 표적 핵산을 편집하는 방법을 제공한다.
- [0015] 본 발명의 제6 측면은 표적 핵산을 포함하는 세포 또는 무세포 시스템을 (a) (i) 본 발명의 변형된 LbCas12a 폴리펩티드를 포함하는 융합 단백질을 코딩하는 폴리뉴클레오티드, 또는 이를 포함하는 발현 카세트 또는 벡터, 및 (a) (ii) 가이드 핵산, 또는 이를 포함하는 발현 카세트 또는 벡터; 및/또는 (b) 본 발명의 변형된 LbCas12a 폴리펩티드를 포함하는 융합 단백질, 및 가이드 핵산을 포함하는 복합체를 코딩하는 핵산 구축물, 또는 이를 포함하는 발현 카세트 또는 벡터; 및/또는 (c) 본 발명의 시스템과 접촉시키고, 그에 의해 표적 핵산을 편집하는 것을 포함하는, 표적 핵산을 편집하는 방법을 제공한다.
- [0016] 본 발명의 제7 측면은 하기 단계를 포함하는 2개 이상의 이중 가닥 핵산 분자를 제조하는 것, 그에 의해 이중 가닥 핵산 분자를 포함하는 무작위적 DNA 라이브러리를 구축하는 것을 포함하는, 프로토스페이서의 5' 단부에 프로토스페이스 인접 모티프 (PAM) 인식 부위를 갖는 CRISPR-Cas 뉴클레아제의 PAM 요건/특이성을 결정하기 위한 이중 가닥 핵산 분자를 포함하는 무작위적 DNA 라이브러리를 구축하는 방법을 제공한다: (a) 비-표적 올리고뉴클레오티드 (제1) 가닥 및 표적 올리고뉴클레오티드 (제2) 가닥을 2개 이상의 이중 가닥 핵산 분자의 각각에 대해 합성하는 단계이며, 여기서 비-표적 올리고뉴클레오티드 가닥은 5'에서 3'로 (i) 약 5 내지 약 15개의 뉴클레오티드를 갖는 제1 서열, (ii) 적어도 4개의 무작위적 뉴클레오티드를 갖는 제2 서열, (iii) 약 16 내지 약 25개의 뉴클레오티드를 포함하는 프로토스페이스 서열, 및 (iv) 약 5 내지 약 20개의 뉴클레오티드를 갖는 제3 서열을 포함하고, (i)의 약 5 내지 15개의 뉴클레오티드를 갖는 제1 서열은 (ii)의 제2 서열의 5' 단부에 바로 인접하고, (ii)의 제2 서열은 (iii)의 프로토스페이스 서열의 5' 단부에 바로 인접하고, 프로토스페이스 서열은 (iv)의 제3 서열의 5' 단부에 바로 인접하고; 표적 올리고뉴클레오티드 (제2) 가닥은 비-표적 올리고뉴클레오티드 가닥에 상보적인 것인 단계; 및 (b) 비-표적 올리고뉴클레오티드 가닥을 상보적 표적 올리고뉴클레오티드 가닥에 어닐링하여 이중 가닥 핵산 분자를 생산하는 단계이며, 여기서 제1 서열은 제한 부위를 (그의 5' 단부에) 포함하고, 제3 서열은 제한 부위를 (그의 3' 단부에) 포함하고, 2개 이상의 이중 가닥 핵산 분자의 각각의 제1 서열 (i), 프로토스페이스 서열 (iii) 및 제3 서열 (iv)은 동일한 것인 단계.
- [0017] 본 발명의 제8 측면은 하기 단계를 포함하는 2개 이상의 이중 가닥 핵산 분자를 제조하는 것, 그에 의해 이중 가닥 핵산 분자를 포함하는 무작위적 DNA 라이브러리를 구축하는 것을 포함하는, 프로토스페이서의 3' 단부에 프로토스페이스 인접 모티프 (PAM) 인식 부위를 갖는 CRISPR-Cas 뉴클레아제의 PAM 요건/특이성을 결정하기 위한 이중 가닥 핵산 분자를 포함하는 무작위적 DNA 라이브러리를 구축하는 방법을 제공한다: (a) 비-표적 올리고뉴클레오티드 (제1) 가닥 및 표적 올리고뉴클레오티드 (제2) 가닥을 2개 이상의 이중 가닥 핵산 분자의 각각에 대해 합성하는 단계이며, 여기서 비-표적 올리고뉴클레오티드 가닥은 5'에서 3'로 (i) 약 5 내지 약 20개의 뉴클레오티드를 갖는 제1 서열, (ii) 약 16 내지 약 25개의 뉴클레오티드를 포함하는 프로토스페이스 서열, (iii) 적어도 4개의 무작위적 뉴클레오티드를 갖는 제2 서열, 및 (iv) 약 5 내지 약 15개의 뉴클레오티드를 갖는 제3 서열을 포함하고, (i)의 약 5 내지 20개의 뉴클레오티드를 갖는 제1 서열은 (ii)의 프로토스페이스 서열의 5' 단부에 바로 인접하고, (iii)의 제2 서열은 (iii)의 프로토스페이스 서열의 3' 단부에 바로 인접하고, (iv)의 제3 서열은 (iii)의 제2 서열의 3' 단부에 바로 인접하고; 표적 올리고뉴클레오티드 (제2) 가닥은 비-표적 올리고뉴클레오티드 가닥에 상보적인 것인 단계; 및 (b) 비-표적 올리고뉴클레오티드 가닥을 상보적 표적 올리고뉴클레오티드 가닥에 어닐링하여 이중 가닥 핵산 분자를 생산하는 단계이며, 여기서 제1 서열 (i)은 제한 부위를 (그의 5' 단부에) 포함하고, 제3 서열 (iv)은 제한 부위를 (그의 3' 단부에) 포함하고, 2개 이상의 이중 가닥 핵산 분자의 각각의 제1 서열 (i), 프로토스페이스 서열 (iii) 및 제3 서열 (iv)은 동일한 것인 단계.
- [0018] 본 발명의 제9 측면은 프로토스페이서의 5' 단부 상에 프로토스페이스 인접 모티프 (PAM) 인식 부위를 갖는

CRISPR-Cas 뉴클레아제의 PAM 요건/특이성을 결정하기 위한 무작위적 DNA 라이브러리를 제공하며, 여기서 무작위적 DNA 라이브러리는 2개 이상의 이중 가닥 핵산 분자를 포함하고, 그의 각각은 (a) 비-표적 올리고뉴클레오타이드 (제1) 가닥 및 표적 올리고뉴클레오타이드 (제2) 가닥 (여기서 비-표적 올리고뉴클레오타이드 가닥은 5'에서 3'로 (i) 약 5 내지 약 15개의 뉴클레오타이드 (예를 들어, 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 또는 15개의 뉴클레오타이드, 및 그 안의 임의의 범위 또는 값)를 갖는 제1 서열, (ii) 적어도 4개의 무작위적 뉴클레오타이드 (예를 들어, 적어도 4, 5, 6, 7, 8, 9, 10개, 또는 그 초과, 및 그 안의 임의의 범위 또는 값)를 갖는 제2 서열, (iii) 약 16 내지 약 25개의 뉴클레오타이드, 예를 들어, 약 16, 17, 18, 19, 20, 21, 22, 23, 24, 또는 25개의 뉴클레오타이드를 포함하는 프로토스페이서 서열, 및 (iv) 약 5 내지 약 20개의 뉴클레오타이드 (예를 들어, 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개의 뉴클레오타이드, 및 그 안의 임의의 범위 또는 값)를 갖는 제3 서열을 포함하고, (i)의 약 5 내지 15개의 뉴클레오타이드를 갖는 제1 서열은 (ii)의 제2 서열의 5' 단부에 바로 인접하고, (ii)의 제2 서열은 (iii)의 프로토스페이서 서열의 5' 단부에 바로 인접하고, 프로토스페이서 서열은 (iv)의 제3 서열의 5' 단부에 바로 인접하고; 표적 올리고뉴클레오타이드 (제2) 가닥은 비-표적 올리고뉴클레오타이드 가닥에 상보적임); 및 (b) 상보적 표적 올리고뉴클레오타이드 가닥에 어닐링되어 이중 가닥 핵산 분자를 생산하는 비-표적 올리고뉴클레오타이드 가닥 (여기서 제1 서열은 제한 부위를 (그의 5' 단부에) 포함하고, 제3 서열은 제한 부위를 (그의 3' 단부에) 포함하고, 2개 이상의 이중 가닥 핵산 분자의 각각의 제1 서열 (i), 프로토스페이서 서열 (iii) 및 제3 서열 (iv)은 동일함)을 포함한다.

[0019] 본 발명의 제10 측면은 프로토스페이서의 3' 단부 상에 프로토스페이서 인접 모티프 (PAM) 인식 부위를 갖는 CRISPR-Cas 뉴클레아제의 PAM 요건/특이성을 결정하기 위한 무작위적 DNA 라이브러리를 제공하며, 여기서 무작위적 DNA 라이브러리는 2개 이상의 이중 가닥 핵산 분자를 포함하고, 그의 각각은 (a) 비-표적 올리고뉴클레오타이드 (제1) 가닥 및 표적 올리고뉴클레오타이드 (제2) 가닥 (여기서 비-표적 올리고뉴클레오타이드 가닥은 5'에서 3'로 (i) 약 5 내지 약 20개의 뉴클레오타이드 (예를 들어, 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개의 뉴클레오타이드, 및 그 안의 임의의 범위 또는 값)를 갖는 제1 서열, (ii) 약 16 내지 약 25개의 뉴클레오타이드, 예를 들어, 약 16, 17, 18, 19, 20, 21, 22, 23, 24, 또는 25개의 뉴클레오타이드를 포함하는 프로토스페이서 서열, (iii) 적어도 4개의 무작위적 뉴클레오타이드 (예를 들어, 적어도 4, 5, 6, 7, 8, 9, 10개, 또는 그 초과, 및 그 안의 임의의 범위 또는 값)를 갖는 제2 서열, 및 (iv) 약 5 내지 약 15개의 뉴클레오타이드 (예를 들어, 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15개의 뉴클레오타이드, 및 그 안의 임의의 범위 또는 값)를 갖는 제3 서열을 포함하고, (i)의 약 5 내지 20개의 뉴클레오타이드를 갖는 제1 서열은 (ii)의 프로토스페이서 서열의 5' 단부에 바로 인접하고, (iii)의 제2 서열은 (iii)의 프로토스페이서 서열의 3' 단부에 바로 인접하고, (iv)의 제3 서열은 (iii)의 제2 서열의 3' 단부에 바로 인접하고; 표적 올리고뉴클레오타이드 (제2) 가닥은 비-표적 올리고뉴클레오타이드 가닥에 상보적임); 및 (b) 상보적 표적 올리고뉴클레오타이드 가닥에 어닐링되어 이중 가닥 핵산 분자를 생산하는 비-표적 올리고뉴클레오타이드 가닥 (여기서 제1 서열은 제한 부위를 (그의 5' 단부에) 포함하고, 제3 서열은 제한 부위를 (그의 3' 단부에) 포함하고, 2개 이상의 이중 가닥 핵산 분자의 각각의 제1 서열 (i), 프로토스페이서 서열 (ii) 및 제3 서열 (iv)은 동일함)을 포함한다.

[0020] 본 발명은 추가로 본 발명의 CRISPR-Cas 뉴클레아제 및/또는 융합 단백질을 코딩하는 폴리뉴클레오타이드를 포함하는 발현 카세트 및/또는 벡터 및/또는 본 발명의 폴리뉴클레오타이드, 폴리펩티드 및/또는 융합 단백질을 포함하는 세포 및/또는 이를 포함하는 키트를 제공한다.

[0021] 본 발명의 이들 및 다른 측면은 하기 발명의 상세한 설명에서 보다 상세하게 제시된다.

[0022] **서열의 간단한 설명**

[0023] **서열식별번호:1 내지 17, 49, 50 및 51**은 Cas12a 뉴클레아제를 코딩하는 예시적인 뉴클레오타이드 서열이다.

[0024] **서열식별번호:18 내지 22**는 예시적인 아데노신 데아미나제이다.

[0025] **서열식별번호:23 내지 25 및 서열식별번호:42 내지 48**은 예시적인 시토신 데아미나제이다.

[0026] **서열식별번호:26**은 우라실-DNA 글리코실라제 억제제 (UGI)를 코딩하는 예시적인 뉴클레오타이드 서열이다.

[0027] **서열식별번호: 27 내지 29**는 유형 V CRISPR-Cas12a 뉴클레아제에 대한 프로토스페이서 인접 모티프 위치의 예를 제공한다.

[0028] **서열식별번호:30 내지 39**는 예를 들어, 시험관내 절단 검정에 사용하기 위한 본 발명의 무작위적 라이브러리를 생성하는데 유용한 예시 뉴클레오타이드 서열을 나타낸다.

- [0029] 서열식별번호:40 내지 41은 프로모터 및 인트론을 코딩하는 예시적인 조절 서열이다.
- [0030] 서열식별번호:52는 예시 발현 카세트의 뉴클레오타이드 서열을 제공한다.
- [0031] 서열식별번호:53은 예시 벡터의 뉴클레오타이드 서열을 제공한다.
- [0032] 서열식별번호:54 내지 61은 예시 스페이서 서열을 제공한다.
- [0033] 서열식별번호:62는 예시 CRISPR RNA를 제공한다.

도면의 간단한 설명

- [0034] 도 1은 본 발명의 예시 PAM 라이브러리 제조의 도해를 나타낸다. 이 방법에서, 5' 인산화된 올리고뉴클레오타이드를 어닐링하고, EcoRI 및 SphI 소화된 pUC19 벡터 내로 클로닝한다. Sca I을 사용하여 벡터를 선형화한다 (AGTACT 서열은 Lbcpf1에 의해 인식되지 않음). 상부 가닥 (서열식별번호:32); 하부 가닥 (서열식별번호:33).
- 도 2는 코딩 서열 내의 옥수수 (상부 패널) 및 대두 (하부 패널)에 대한 유전자 당 평균 프로토스페이서 인접 모티프 (PAM)를 나타낸다. Lbcpf1 유전자는 Cas9 변이체보다 훨씬 더 적은 유전자 서열에 접근할 수 있다.
- 도 3은 옥수수 (패널 a, c) 및 대두 (패널 b, d)에서 PAM에 의해 제한된 평균 접근가능한 시토신 (패널 a, b) 및 아데닌 (패널 c, d)을 나타낸다. 나타내어진 바와 같이, Lbcpf1 시토신 및 아데닌은 Cas9 변이체보다 훨씬 더 적은 시토신 및 아데닌에 접근할 수 있다.
- 도 4는 문헌 [Gao et al. (*Nat Biotechnol* 35(8):789-792 (2017))]의 단순화된 PAM 결정 검정을 나타내는 도해를 제공한다. 증폭된 단편은 CRISPR-Cas 뉴클레아제에 의해 커팅되지 않은 서열을 나타내는 반면, 증폭되지 않은 단편은 CRISPR-Cas 뉴클레아제에 의해 커팅되었다. 시퀀싱 및 무효소 대조군과의 비교는 증폭되지 않은 (즉, 대조군 집단에 존재하지만, 편집된 집단에는 존재하지 않는) 핵산 서열을 확인하고, 따라서, 뉴클레아제에 의해 인식되고 커팅되는 서열을 확인한다. 상부 패널: 상부 서열 (서열식별번호:36), 중간 서열 (서열식별번호:37), 하부 서열 (서열식별번호:38); 중간 패널: 상부 서열 (서열식별번호:39), 하부 서열 (서열식별번호:37); 하부 패널: 상부 서열 (서열식별번호:36), 하부 서열 (서열식별번호:37).
- 도 5. 최고 대 최저 수의 카운트로 플롯팅된 PAMDA 라이브러리의 3개의 일루미나(Illumina) 마이섹(MiSeq) NGS 판독물의 평균. NNNNN을 함유하는 1024개의 라이브러리 구성원은 39개의 판독물의 평균을 갖는 정규 분포를 따른다.
- 도 6. wtLbCas12a 및 플라스미드 스페이서를 표적화하지 않은 crRNA를 함유하는 음성 대조군의 세포 분류 결과.
- 도 7. wtLbCas12a 및 플라스미드 스페이서를 표적화하는 crRNA의 세포 분류 결과.
- 도 8. LbCas12a-K595Y 및 플라스미드 스페이서를 표적화하는 crRNA의 세포 분류 결과.
- 도 9. LbCas12a-G532R-K595R 이중 돌연변이 대조군 및 플라스미드 스페이서를 표적화하는 crRNA의 세포 분류 결과.
- 도 10. 플라스미드 스페이서를 표적화하는 crRNA를 갖는, 이 연구에서 점 돌연변이 중 2개의 조합인 LbCas12a-T152R-K595Y 이중 돌연변이의 세포 분류 결과.
- 도 11. 플라스미드 스페이서를 표적화하는 crRNA를 갖는 점 돌연변이 중 3개의 조합인 LbCas12a-T152R-K538W-K595Y 삼중 돌연변이의 세포 분류 결과.
- 도 12. 2개의 별개의 crRNA 없는 대조군 및 야생형 dLbCas12a 및 리포터 라이브러리에 대한 총 정규화된 NGS 카운트.
- 도 13. 256개의 4 뉴클레오타이드 PAM의 각각에 대한 단일 점 돌연변이 정규화된 PAM-SCANR 점수. 그래프를 관통하는 선은 2개의 음성 대조군 중 어느 하나에 대한 최고 관찰된 점수, 1.67을 나타낸다.
- 도 14. 256개의 4 뉴클레오타이드 PAM의 각각에 대한 조합적 돌연변이 정규화된 PAM-SCANR 점수.
- 도 15는 돌연변이체 K538W 및 K595Y를 조합하는 것이 공유한 PAM 인식 서열을 갖는 효소 LbCas12a-K538W-K595Y를 발생시킴을 나타낸다. 일부 경우에, K538W (수직 해칭됨 또는 K595Y (수평 해칭됨)로부터의 공유된 PAM 인식 모티프는 조합 돌연변이체에 의해 인식되지만, 종종 조합은 완전히 신규한 PAM 인식 서열 (태칭됨)을 발생시킨다.

도 16은 다수의 확장된 PAM 돌연변이를 조합하는 것이 때때로 부가적이지만 종종 고유한 PAM 인식 서열을 생성할 수 있음을 나타낸다.

도 17은 PAN-SCANR (회색 박스)로부터의 1.67 점수 초과를 나타낸 모든 비-TTTV PAM을 K595Y (좌측) 및 T152R (우측)과 비교한다. 1.67 컷오프 초과 PAM-SCANR 양성 PAM의 하나를 제외한 전부는 시험관내에서 9.2 컷오프 초과 PAM 고갈 점수를 가졌다.

도 18은 각각의 TTTV-함유 스페이서에 대한 HEK293T 세포에서 형성된 백분율 인텔을 나타낸다. 개별적 인텔 백분율은 TTTC, TTTA, 및 TTTG에 대해 각각 원, 정사각형, 또는 삼각형으로서 나타내어진다. 평균 선 및 가장 가까운 정수에 대해 반올림된 값은 또한 각각의 스페이서에 대해 나타내어진다.

도 19는 시험된 PAM 당 LbCas12a_K595Y HEK293T 최대 관찰된 인텔 백분율을 나타낸다. 0.1% 초과 값은 시퀀싱의 잡음의 외부이고, 진정한 인텔을 나타낸다.

도 20은 시험된 PAM 당 LbCas12a_T152R HEK293T 최대 관찰된 인텔 백분율을 나타낸다. 0.1% 초과 값은 시퀀싱의 잡음의 외부이고, 진정한 인텔을 나타낸다.

도 21은 시험된 PAM 당 LbCas12a_K538W HEK293T 최대 관찰된 인텔 백분율을 나타낸다. 0.1% 초과 값은 시퀀싱의 잡음의 외부이고, 진정한 인텔을 나타낸다.

도 22a 내지 22b. LbCas12a-T152R (도 22a) 및 LbCas12a-K595Y (도 22b)에 대한 % 인텔 (최대) 및 정규화된 박테리아 PAM-SCANR 점수 사이의 선형 상관관계.

발명을 실시하기 위한 구체적인 내용

- [0035] 본 발명은 이제 본 발명의 실시양태가 나타내어진 첨부된 도면 및 실시예를 참고로 이하 기재될 것이다. 이 설명은 본 발명이 실행될 수 있는 모든 상이한 방식, 또는 본 발명에 첨가될 수 있는 모든 특색의 상세한 목록인 것으로 의도되지 않는다. 예를 들어, 한 실시양태에 관하여 예시된 특색은 다른 실시양태에 포함될 수 있고, 특정 실시양태에 관하여 예시된 특색은 그 실시양태로부터 삭제될 수 있다. 따라서, 본 발명은 본 발명의 일부 실시양태에서, 본원에 제시된 임의의 특색 또는 특색의 조합이 배제되거나 생략될 수 있음을 고려한다. 또한, 본원에서 제안된 다양한 실시양태에 대한 다수의 변동 및 첨가는 본 발명으로부터 벗어나지 않는 본 개시내용의 관점에서 관련 기술분야의 통상의 기술자에게 명백할 것이다. 따라서, 하기 설명은 본 발명의 일부 특정 실시양태를 예시하는 것으로 의도되며, 그의 모든 순열, 조합 및 변동을 철저하게 특징하는 것으로 의도되지 않는다.
- [0036] 달리 정의되지 않는 한, 본원에 사용된 모든 기술 및 과학 용어는 본 발명이 속하는 관련 기술분야의 통상의 기술자에 의해 통상적으로 이해되는 것과 동일한 의미를 갖는다. 본원에서 본 발명의 설명에 사용된 용어는 단지 특정 실시양태를 기재하는 목적을 위한 것이며, 본 발명의 제한인 것으로 의도되지 않는다.
- [0037] 본원에 인용된 모든 공보, 특허 출원, 특허 및 다른 참고문헌은 참고문헌이 제시된 문장 및/또는 단락에 관한 교시내용에 대해 그 전문이 참조로 포함된다.
- [0038] 맥락이 달리 지시하지 않는 한, 본원에 기재된 본 발명의 다양한 특색은 임의의 조합으로 사용될 수 있음이 구체적으로 의도된다. 더욱이, 본 발명은 또한 본 발명의 일부 실시양태에서, 본원에 제시된 임의의 특색 또는 특색의 조합이 배제되거나 생략될 수 있음을 고려한다. 예시하기 위해, 본 명세서가 조성물이 성분 A, B 및 C를 포함한다고 진술하는 경우, 임의의 A, B 또는 C, 또는 이들의 조합이 단독으로 또는 임의의 조합으로 생략되고 청구배제될 수 있음이 구체적으로 의도된다.
- [0039] 본 발명의 설명 및 첨부된 청구항에 사용된 단수 형태는 맥락이 명백하게 달리 지시하지 않는 한, 복수 형태를 또한 포함하는 것으로 의도된다.
- [0040] 또한 본원에 사용된 "및/또는"은 연관된 열거된 항목 중 하나 이상의 임의의 및 모든 가능한 조합, 뿐만 아니라 대안 ("또는")으로 해석되는 경우 조합의 결여를 지칭하고 포괄한다.
- [0041] 측정가능한 값, 예컨대 양 또는 농도 등을 언급하는 경우 본원에 사용된 용어 "약"은 특정된 값의 $\pm 10\%$, $\pm 5\%$, $\pm 1\%$, $\pm 0.5\%$, 또는 심지어 $\pm 0.1\%$ 의 변동 뿐만 아니라 특정된 값을 포괄하는 것으로 의미된다. 예를 들어, "약 X" (여기서 X는 측정가능한 값임)는 X 뿐만 아니라 X의 $\pm 10\%$, $\pm 5\%$, $\pm 1\%$, $\pm 0.5\%$, 또는 심지어 $\pm 0.1\%$ 의 변동을 포함하는 것으로 의미된다. 측정가능한 값에 대해 본원에서 제공된 범위는 임의의 다른 범위

및/또는 그 안의 개별적 값을 포함할 수 있다.

- [0042] 본원에 사용된 "X 내지 Y" 및 "약 X 내지 Y"와 같은 어구는 X 및 Y를 포함하는 것으로 해석되어야 한다. 본원에 사용된 "약 X 내지 Y"와 같은 어구는 "약 X 내지 약 Y"를 의미하고, "약 X 내지 Y"와 같은 어구는 "약 X 내지 약 Y"를 의미한다.
- [0043] 본원에서 값의 범위의 나열은 본원에서 달리 지시되지 않는 한, 단지, 개별적으로 범위 내에 해당하는 각각의 별개의 값을 지칭하는 약칭 방법으로서 역할을 하는 것으로 의도되며, 각각의 별개의 값은 그것이 본원에 개별적으로 나열된 것처럼 명세서에 포함된다. 예를 들어, 범위 10 내지 15가 개시되는 경우, 11, 12, 13, 및 14가 또한 개시된다.
- [0044] 본원에 사용된 용어 "포함하다(comprise)", "포함하다(comprises)" 및 "포함하는"은 진술된 특색, 정수, 단계, 작동, 요소, 및/또는 성분의 존재를 특정하지만, 하나 이상의 다른 특색, 정수, 단계, 작동, 요소, 성분, 및/또는 그의 군의 존재 또는 첨가를 불가능하게 하지 않는다.
- [0045] 본원에 사용된 전이적 어구 "본질적으로 이루어진"은 청구항의 범위가 청구항에 나열된 특정된 물질 또는 단계 및 청구된 발명의 기본적인이고 신규한 특징(들)에 실질적으로 영향을 미치지 않는 것들을 포괄하는 것으로 해석되어야 한다. 따라서, 본 발명의 청구항에 사용되는 경우 용어 "본질적으로 이루어진"은 "포함하는"과 등가인 것으로 해석되는 것으로 의도되지 않는다.
- [0046] 본원에 사용된 용어 "증가시키다", "증가시키는", "증진시키다", "증진시키는", "개선시키다" 및 "개선시키는" (및 그의 문법적 파생어)은 대조군과 비교하여 적어도 약 25%, 50%, 75%, 100%, 150%, 200%, 300%, 400%, 500% 또는 그 초과를 상승을 기재한다.
- [0047] 본원에 사용된 용어 "감소시키다(reduce)", "감소된", "감소시키는", "감소", "감소시키다(diminish)" 및 "감소시키다(decrease)" (및 그의 문법적 파생어)는 예를 들어, 대조군과 비교하여 적어도 약 5%, 10%, 15%, 20%, 25%, 35%, 50%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, 또는 100%의 감소를 기재한다. 특정 실시양태에서, 감소는 검출가능한 활성 또는 양이 없거나 본질적으로 없는 것 (즉, 비유의한 양, 예를 들어, 약 10% 또는 심지어 5% 미만)을 발생시킬 수 있다.
- [0048] "이중" 또는 "재조합" 뉴클레오타이드 서열은 천연 발생 뉴클레오타이드 서열의 비-천연 발생 다중 카피를 포함한, 그것이 도입되는 숙주 세포와 천연적으로 연관되지 않은 뉴클레오타이드 서열이다.
- [0049] "천연" 또는 "야생형" 핵산, 뉴클레오타이드 서열, 폴리펩티드 또는 아미노산 서열은 천연 발생 또는 내인성 핵산, 뉴클레오타이드 서열, 폴리펩티드 또는 아미노산 서열을 지칭한다. 따라서, 예를 들어, "야생형 mRNA"는 유기체에서 천연 발생이거나 그에 대해 내인성인 mRNA이다. "상동" 핵산 서열은 그것이 도입되는 숙주 세포와 천연적으로 연관된 뉴클레오타이드 서열이다.
- [0050] 본원에 사용된 용어 "핵산", "핵산 분자", "뉴클레오타이드 서열" 및 "폴리뉴클레오타이드"는 선형 또는 분지형, 단일 또는 이중 가닥, 또는 그의 하이브리드인 RNA 또는 DNA를 지칭한다. 상기 용어는 또한 RNA/DNA 하이브리드를 포괄한다. dsRNA가 합성적으로 생산되는 경우, 덜 통상적인 염기, 예컨대 이노신, 5-메틸시토신, 6-메틸아데닌, 히포크산틴 등은 또한 안티센스, dsRNA, 및 리보자임 쌍형성에 사용될 수 있다. 예를 들어, 우리딘 및 시티딘의 C-5 프로핀 유사체를 함유하는 폴리뉴클레오타이드는 RNA에 고 친화도로 결합하고, 유전자 발현의 강력한 안티센스 억제제인 것으로 나타났다. 다른 변형, 예컨대 포스포디에스테르 백본, 또는 RNA의 리보스 당 기에서 2'-히드록시에 대한 변형은 또한 이루어질 수 있다.
- [0051] 본원에 사용된 용어 "뉴클레오타이드 서열"은 뉴클레오타이드의 이중중합체 또는 핵산 분자의 5'에서 3' 단부로 이들 뉴클레오타이드의 서열을 지칭하며, cDNA, DNA 단편 또는 부분, 게놈 DNA, 합성 (예를 들어, 화학적으로 합성된) DNA, 플라스미드 DNA, mRNA, 및 안티-센스 RNA (이들 중 임의의 것은 단일 가닥 또는 이중 가닥일 수 있음)를 포함한 DNA 또는 RNA 분자를 포함한다. 용어 "뉴클레오타이드 서열", "핵산", "핵산 분자", "핵산 구축물", "올리고뉴클레오타이드" 및 "폴리뉴클레오타이드"는 또한 뉴클레오타이드의 이중중합체를 지칭하기 위해 본원에서 상호교환가능하게 사용된다. 본원에서 제공된 핵산 분자 및/또는 뉴클레오타이드 서열은 본원에서 5'에서 3' 방향으로, 좌측에서 우측으로 제시되며, 미국 서열 규칙, 37 CFR § § 1.821 내지 1.825 및 국제 지적 재산권 기구 (WIPO) 표준 ST.25에 제시된 바와 같은 뉴클레오타이드 문자를 나타내기 위한 표준 부호를 사용하여 나타내어진다. 본원에 사용된 "5' 영역"은 폴리뉴클레오타이드의 5' 단부에 가장 가까운 폴리뉴클레오타이드의 영역을 의미할 수 있다. 따라서, 예를 들어, 폴리뉴클레오타이드의 5' 영역에 있는 요소는 폴리뉴클레오타이드의 5' 단부에 위치한 제1 뉴클레오타이드에서 폴리뉴클레오타이드를 통해 중간에 위치한 뉴클레오타이드까지 어디에나 위치할 수

있다. 본원에 사용된 "3' 영역"은 폴리뉴클레오타이드의 3' 단부에 가장 가까운 폴리뉴클레오타이드의 영역을 의미할 수 있다. 따라서, 예를 들어, 폴리뉴클레오타이드의 3' 영역에 있는 요소는 폴리뉴클레오타이드의 3' 단부에 위치한 제1 뉴클레오타이드에서 폴리뉴클레오타이드를 통해 중간에 위치한 뉴클레오타이드까지 어디에나 위치할 수 있다.

[0052] 본원에 사용된 용어 "유전자"는 mRNA, 안티센스 RNA, miRNA, 안티-마이크로RNA 안티센스 올리고데옥시리보뉴클레오타이드 (AMO) 등을 생산하는데 사용될 수 있는 핵산 분자를 지칭한다. 유전자는 기능적 단백질 또는 유전자 생성물을 생산하는데 사용될 수 있거나 그렇지 않을 수 있다. 유전자는 코딩 및 비-코딩 영역 (예를 들어, 인트론, 조절 요소, 프로모터, 인핸서, 종결 서열 및/또는 5' 및 3' 비번역된 영역) 둘 다를 포함할 수 있다. 유전자는 "단리될" 수 있으며, 이는 그의 자연 상태에서 핵산과 관련하여 통상적으로 발견되는 성분이 실질적으로 또는 본질적으로 없는 핵산을 의미한다. 이러한 성분은 세포 물질, 재조합 생산으로부터의 배양 배지, 및/또는 핵산을 화학적으로 합성하는데 사용되는 다양한 화학물질을 포함한다.

[0053] 용어 "돌연변이"는 점 돌연변이 (예를 들어, 미스센스, 또는 넌센스, 또는 격자 이동을 발생시키는 단일 염기 쌍의 삽입 또는 결실), 삽입, 결실, 및/또는 말단절단을 지칭한다. 돌연변이가 아미노산 서열 내의 잔기의 또 다른 잔기로의 치환, 또는 서열 내의 1개 이상의 잔기의 결실 또는 삽입인 경우, 돌연변이는 전형적으로 원래 잔기, 이어서 서열 내의 잔기의 위치 및 새롭게 치환된 잔기의 신원을 확인함으로써 기재된다.

[0054] 본원에 사용된 용어 "상보적" 또는 "상보성"은 염기-쌍형성에 의한 허용적인 염 및 온도 조건 하에서의 폴리뉴클레오타이드의 천연 결합을 지칭한다. 예를 들어, 서열 "A-G-T" (5'에서 3'로)는 상보적 서열 "T-C-A" (3'에서 5'로)에 결합한다. 2개의 단일-가닥 분자 사이의 상보성은 뉴클레오타이드의 단지 일부가 결합하는 "부분적"일 수 있거나, 또는 이는 총 상보성이 단일 가닥 분자 사이에 존재하는 경우 완전할 수 있다. 핵산 가닥 사이의 상보성의 정도는 핵산 가닥 사이의 혼성화의 효율 및 강도에 대한 유의한 효과를 갖는다.

[0055] 본원에 사용된 "상보체"는 비교자 뉴클레오타이드 서열과의 100% 상보성을 의미할 수 있거나, 또는 이는 100% 미만의 상보성 (예를 들어, 약 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 등의 상보성)을 의미할 수 있다.

[0056] 본 발명의 뉴클레오타이드 서열의 "부분" 또는 "단편"은 참조 핵산 또는 뉴클레오타이드 서열에 비해 감소된 (예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개 또는 그 초과 뉴클레오타이드만큼 감소된) 길이의 및 참조 핵산 또는 뉴클레오타이드 서열과 동일한 또는 거의 동일한 (예를 들어, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 동일한) 인접한 뉴클레오타이드의 뉴클레오타이드 서열을 포함하고/거나, 이로 본질적으로 이루어지고/거나 이로 이루어진 뉴클레오타이드 서열을 의미하는 것으로 이해될 것이다. 본 발명에 따른 이러한 핵산 단편 또는 부분은, 적절한 경우, 그것이 구성요소인 보다 큰 폴리뉴클레오타이드에 포함될 수 있다. 예로서, 본 발명의 가이드 핵산의 반복 서열은 야생형 CRISPR-Cas 반복 서열 (예를 들어, 야생형 Cas9 반복부, 야생형 Cas12a 반복부 등)의 부분을 포함할 수 있다.

[0057] 상동성을 갖는 상이한 핵산 또는 단백질은 본원에서 "동족체"로 지칭된다. 용어 동족체는 동일한 및 다른 종으로부터의 상동 서열 및 동일한 및 다른 종으로부터의 이종상동 서열을 포함한다. "상동성"은 위치적 동일성 (즉, 서열 유사성 또는 동일성)의 퍼센트의 관점에서 2개 이상의 핵산 및/또는 아미노산 서열 사이의 유사성의 수준을 지칭한다. 상동성은 또한 상이한 핵산 또는 단백질 중에서 유사한 기능적 특성의 개념을 지칭한다. 따라서, 본 발명의 조성물 및 방법은 본 발명의 뉴클레오타이드 서열 및 폴리펩티드 서열에 대한 동족체를 추가로 포함한다. 본원에 사용된 "이종상동"은 종분화 동안 공통 조상 유전자로부터 발생한 상이한 종에서의 상동 뉴클레오타이드 서열 및/또는 아미노산 서열을 지칭한다. 본 발명의 뉴클레오타이드 서열의 동족체는 본 발명의 상기 뉴클레오타이드 서열과 실질적 서열 동일성 (예를 들어, 적어도 약 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% 또는 100%)을 갖는다.

[0058] 본원에 사용된 "서열 동일성"은 2개의 최적으로 정렬된 폴리뉴클레오타이드 또는 폴리펩티드 서열이 성분, 예를 들어, 뉴클레오타이드 또는 아미노산의 정렬의 창 전반에 걸쳐 불변하는 정도를 지칭한다. "동일성"은 문헌 [Computational Molecular Biology (Lesk, A. M., ed.) Oxford University Press, New York (1988)]; [Biocomputing: Informatics and Genome Projects (Smith, D. W., ed.) Academic Press, New York (1993)]; [Computer Analysis of Sequence Data, Part I (Griffin, A. M., and Griffin, H. G., eds.) Humana Press,

New Jersey (1994)]; [*Sequence Analysis in Molecular Biology* (von Heinje, G., ed.) Academic Press (1987)]; 및 [*Sequence Analysis Primer* (Gribskov, M. and Devereux, J., eds.) Stockton Press, New York (1991)]에 기재된 것들을 포함하나 이에 제한되지는 않는 공지된 방법에 의해 용이하게 계산될 수 있다.

[0059] 본원에 사용된 용어 "퍼센트 서열 동일성" 또는 "퍼센트 동일성"은 2개의 서열이 최적으로 정렬된 경우 시험 ("대상") 폴리뉴클레오티드 분자 (또는 그의 상보적 가닥)와 비교하여 참조 ("의문") 폴리뉴클레오티드 분자 (또는 그의 상보적 가닥)의 선형 폴리뉴클레오티드 서열에서 동일한 뉴클레오티드의 백분율을 지칭한다. 일부 실시양태에서, "퍼센트 동일성"은 참조 폴리펩티드와 비교하여 아미노산 서열에서 동일한 아미노산의 백분율을 지칭할 수 있다.

[0060] 2개의 핵산 분자, 뉴클레오티드 서열 또는 단백질 서열의 맥락에서 본원에 사용된 어구 "실질적으로 동일한", 또는 "실질적 동일성"은 하기 서열 비교 알고리즘 중 하나를 사용하여 또는 육안 검사에 의해 측정된 바와 같은, 최대 상응성에 대해 비교되고 정렬된 경우, 적어도 약 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% 또는 100% 뉴클레오티드 또는 아미노산 잔기 동일성을 갖는 2개 이상의 서열 또는 하위서열을 지칭한다. 본 발명의 일부 실시양태에서, 실질적 동일성은 약 10개의 뉴클레오티드 내지 약 20개의 뉴클레오티드, 약 10개의 뉴클레오티드 내지 약 25개의 뉴클레오티드, 약 10개의 뉴클레오티드 내지 약 30개의 뉴클레오티드, 약 15개의 뉴클레오티드 내지 약 25개의 뉴클레오티드, 약 30개의 뉴클레오티드 내지 약 40개의 뉴클레오티드, 약 50개의 뉴클레오티드 내지 약 60개의 뉴클레오티드, 약 70개의 뉴클레오티드 내지 약 80개의 뉴클레오티드, 약 90개의 뉴클레오티드 내지 약 100개의 뉴클레오티드, 또는 그 초과 뉴클레오티드의 길이, 및 그 안의 임의의 범위 내지 서열의 전장인 본 발명의 뉴클레오티드 서열의 연속적 뉴클레오티드의 영역에 걸쳐 존재한다. 일부 실시양태에서, 뉴클레오티드 서열은 적어도 약 20개의 뉴클레오티드 (예를 들어, 약 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40개의 뉴클레오티드)에 걸쳐 실질적으로 동일할 수 있다. 일부 실시양태에서, 실질적으로 동일한 뉴클레오티드 또는 단백질 서열은 그것이 실질적으로 동일한 뉴클레오티드 (또는 코딩된 단백질 서열)과 실질적으로 동일한 기능을 수행한다.

[0061] 서열 비교를 위해, 전형적으로 한 서열은 시험 서열이 비교되는 참조 서열로서 작용한다. 서열 비교 알고리즘을 사용하는 경우, 시험 및 참조 서열은 컴퓨터 내로 들어가고, 하위서열 좌표가 필요에 따라 지정되고, 서열 알고리즘 프로그램 파라미터가 지정된다. 이어서 서열 비교 알고리즘은 지정된 프로그램 파라미터에 기반하여 참조 서열에 비해 시험 서열(들)에 대한 퍼센트 서열 동일성을 계산한다.

[0062] 비교 창을 정렬하기 위한 서열의 최적 정렬은 관련 기술분야의 통상의 기술자에게 널리 공지되어 있으며, 스미스 및 와터만(Smith and Waterman)의 국소 상동성 알고리즘, 니들만 및 운쉬(Needleman and Wunsch)의 상동성 정렬 알고리즘, 피어슨 및 리프만(Pearson and Lipman)의 유사성에 대한 검색 방법과 같은 틀에 의해, 및 임의로 이들 알고리즘의 컴퓨터화된 실행, 예컨대 GCG® 위스콘신 패키지(Wisconsin Package)® (악셀리스 인크.(Accelrys Inc.)), 미국 캘리포니아주 샌 디에고의 일부로서 이용가능한 GAP, BESTFIT, FASTA, 및 TFASTA에 의해 수행될 수 있다. 시험 서열 및 참조 서열의 정렬된 절편에 대한 "동일성 비율"은 2개의 정렬된 서열에 의해 공유된 동일한 성분의 수 나누기 참조 서열 절편, 즉, 전체 참조 서열 또는 참조 서열의 보다 작은 한정된 부분에서의 성분의 총 수이다. 퍼센트 서열 동일성은 동일성 비율 곱하기 100으로서 나타내어진다. 1개 이상의 폴리뉴클레오티드 서열의 비교는 전장 폴리뉴클레오티드 서열 또는 그의 부분에 대한 것, 또는 보다 긴 폴리뉴클레오티드 서열에 대한 것일 수 있다. 본 발명의 목적을 위해, "퍼센트 동일성"은 또한 번역된 뉴클레오티드 서열에 대해 BLASTX 버전 2.0 및 폴리뉴클레오티드 서열에 대해 BLASTN 버전 2.0을 사용하여 결정될 수 있다.

[0063] 2개의 뉴클레오티드 서열은 또한 2개의 서열이 엄격한 조건 하에서 서로에 혼성화하는 경우 실질적으로 상보적인 것으로 간주될 수 있다. 일부 대표적인 실시양태에서, 실질적으로 상보적인 것으로 간주되는 2개의 뉴클레오티드 서열은 고도로 엄격한 조건 하에서 서로에 혼성화한다.

[0064] 핵산 혼성화 실험, 예컨대 서던 및 노던 혼성화의 맥락에서 "엄격한 혼성화 조건" 및 "엄격한 혼성화 세척 조건"은 서열 의존적이며, 상이한 환경적 파라미터 하에서 상이하다. 핵산의 혼성화에 대한 광범위한 지침은 문헌 [*Tijssen Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes* part I chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays" Elsevier, New York (1993)]에서 발견된다. 일반적으로, 고도로 엄격한 혼성화 및 세척 조건은 한정된 이온 강도 및 pH에서 특이적 서열에 대한 열 융점 (T_m)보다 약 5°C 더 낮도록 선택된다.

[0065] T_m 은 표적 서열의 50%가 완벽하게 매칭된 프로브에 혼성화하는 온도 (한정된 이온 강도 및 pH 하에서)이다. 매우 엄격한 조건은 특정 프로브에 대해 T_m 과 동등하도록 선택된다. 서던 또는 노던 블롯에서 필터 상에 100개 초과 상보적 잔기를 갖는 상보적 뉴클레오타이드 서열의 혼성화를 위한 엄격한 혼성화 조건의 예는 42°C에서 1 mg의 헤파린을 갖는 50% 포름아미드이며, 혼성화는 밤새 수행된다. 고도로 엄격한 세척 조건의 예는 72°C에서 약 15분 동안 0.1 M NaCl이다. 엄격한 세척 조건의 예는 65°C에서 15분 동안 0.2x SSC 세척이다 (SSC 완충제의 설명에 대해서는, 상기 문헌 [Sambrook] 참조). 종종, 배경 프로브 신호를 제거하기 위해 저 엄격성 세척은 고 엄격성 세척에 선행한다. 예를 들어, 100개 초과 뉴클레오타이드의 두가닥에 대한 중간 엄격성 세척의 예는 45°C에서 15분 동안 1x SSC이다. 예를 들어, 100개 초과 뉴클레오타이드의 두가닥에 대한 저 엄격성 세척의 예는 40°C에서 15분 동안 4 내지 6x SSC이다. 짧은 프로브 (예를 들어, 약 10 내지 50개의 뉴클레오타이드)에 대해, 엄격한 조건은 전형적으로 pH 7.0 내지 8.3에서 약 1.0 M 미만의 Na 이온, 전형적으로 약 0.01 내지 1.0 M Na 이온 농도 (또는 다른 염)의 염 농도를 수반하고, 온도는 전형적으로 적어도 약 30°C이다. 엄격한 조건은 또한 불안정화제, 예컨대 포름아미드의 첨가로 달성될 수 있다. 일반적으로, 특정 혼성화 검정에서 비관련된 프로브에 대해 관찰된 것보다 2x (이상)의 신호 대 잡음 비는 특이적 혼성화의 검출을 지시한다. 엄격한 조건 하에서 서로에 혼성화하지 않는 뉴클레오타이드 서열은 이들이 코딩하는 단백질이 실질적으로 동일한 경우 여전히 실질적으로 동일하다. 이는 예를 들어, 뉴클레오타이드 서열의 카피가 유전 암호에 의해 허용되는 최대 코돈 축퇴성을 사용하여 생성되는 경우 일어날 수 있다.

[0066] 본 발명의 임의의 뉴클레오타이드 서열, 폴리뉴클레오타이드 및/또는 재조합 핵산 구축물은 관심의 임의의 유기체에서의 발현을 위해 코돈 최적화될 수 있다. 코돈 최적화는 관련 기술분야에 널리 공지되어 있으며, 종 특이적 코돈 사용빈도 표를 사용한 코돈 사용빈도 편향에 대한 뉴클레오타이드 서열의 변형을 수반한다. 코돈 사용빈도 표는 관심의 유기체/종에 대해 가장 고도로 발현된 유전자의 서열 분석에 기반하여 생성된다. 뉴클레오타이드 서열이 핵에서 발현되는 경우, 코돈 사용빈도 표는 관심의 종에 대해 고도로 발현된 핵 유전자의 서열 분석에 기반하여 생성된다. 뉴클레오타이드 서열의 변형은 종 특이적 코돈 사용빈도 표를 천연 폴리뉴클레오타이드 서열에 존재하는 코돈과 비교함으로써 결정된다. 관련 기술분야에 이해되어 있는 바와 같이, 뉴클레오타이드 서열의 코돈 최적화는 천연 뉴클레오타이드 서열과 100% 미만의 동일성 (예를 들어, 약 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 또는 99.9%, 및 그 안의 임의의 범위 또는 값)을 갖지만, 원래 천연 뉴클레오타이드 서열에 의해 코딩되는 것과 동일한 기능을 갖는 폴리펩티드를 여전히 코딩하는 뉴클레오타이드 서열을 발생시킨다. 따라서, 본 발명의 일부 실시양태에서, 본 발명의 폴리뉴클레오타이드, 핵산 구축물, 발현 카세트, 및/또는 벡터 (예를 들어, 본 발명의 폴리펩티드, 융합 단백질, 복합체, 예를 들어, 변형된 CRISPR-Cas 뉴클레아제를 포함함/코딩함)는 관심의 특정 종, 예를 들어, 특정 식물 종, 특정 박테리아 종, 특정 동물 종 등에서의 발현을 위해 코돈 최적화된다. 일부 실시양태에서, 본 발명의 코돈 최적화된 핵산 구축물, 폴리뉴클레오타이드, 발현 카세트, 및/또는 벡터는 코돈 최적화되지 않은 본 발명의 폴리뉴클레오타이드, 핵산 구축물, 발현 카세트, 및/또는 벡터와 약 70% 내지 약 99.9% (예를 들어, 약 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9% 또는 100%) 또는 그 초과 동일성을 갖는다.

[0067] 본원에 기재된 임의의 실시양태에서, 본 발명의 폴리뉴클레오타이드 또는 핵산 구축물은 식물 및/또는 식물의 세포에서의 발현을 위해 다양한 프로모터 및/또는 다른 조절 요소와 작동적으로 회합될 수 있다. 따라서, 일부 실시양태에서, 본 발명의 폴리뉴클레오타이드 또는 핵산 구축물은 1개 이상의 뉴클레오타이드 서열에 작동가능하게 연결된 1종 이상의 프로모터, 인트론, 인핸서, 및/또는 종결자를 추가로 포함할 수 있다. 일부 실시양태에서, 프로모터는 인트론 (예를 들어, Ubi1 프로모터 및 인트론)과 작동가능하게 회합될 수 있다. 일부 실시양태에서, 인트론과 회합된 프로모터는 "프로모터 영역" (예를 들어, Ubi1 프로모터 및 인트론)으로 지칭될 수 있다.

[0068] 폴리뉴클레오타이드에 관하여 본원에 사용된 "작동가능하게 연결된" 또는 "작동가능하게 회합된"이란, 지시된 요소가 서로와 기능적으로 관련되고, 또한 일반적으로 물리적으로 관련됨을 의미한다. 따라서, 본원에 사용된 용어 "작동가능하게 연결된" 또는 "작동가능하게 회합된"은 기능적으로 연관된 단일 핵산 분자 상의 뉴클레오타이드 서열을 지칭한다. 따라서, 제2 뉴클레오타이드 서열에 작동가능하게 연결된 제1 뉴클레오타이드 서열은 제1 뉴클레오타이드 서열이 제2 뉴클레오타이드 서열과 기능적 관계에 놓인 상황을 의미한다. 예를 들어, 프로모터는 프로모터가 뉴클레오타이드 서열의 전사 또는 발현을 실행하는 경우 상기 뉴클레오타이드 서열과 작동가능하게 회합된다. 관련 기술분야의 통상의 기술자는 제어 서열 (예를 들어, 프로모터)은, 제어 서열이 그의 발현을 지정하도록 기

능하는 한, 그것이 작동가능하게 회합된 뉴클레오티드 서열과 인접할 필요는 없음을 인정할 것이다. 따라서, 예를 들어, 개재하는 비번역된, 그러나 전사된 핵산 서열은 프로모터 및 뉴클레오티드 서열 사이에 존재할 수 있고, 프로모터는 뉴클레오티드 서열과 "작동가능하게 연결된" 것으로 여전히 간주될 수 있다.

[0069] 폴리펩티드에 관하여 본원에 사용된 용어 "연결된"은 한 폴리펩티드의 또 다른 것의 부착을 지칭한다. 폴리펩티드는 직접적으로 (예를 들어, 펩티드 결합을 통해) 또는 링커를 통해 (N-말단 또는 C-말단에서) 또 다른 폴리펩티드에 연결될 수 있다.

[0070] 용어 "링커"는 관련 기술분야에 인식되어 있으며, 2개의 분자 또는 모이어티, 예를 들어, 융합 단백질의 2개의 도메인, 예컨대, 예를 들어, LbCas12a CRISPR-Cas 뉴클레아제 도메인 및 관심의 폴리펩티드 (예를 들어, 핵산-편집 도메인, 데아미나제 도메인, 아데노신 데아미나제, 시토신 데아미나제)를 연결하는 화학기, 또는 분자를 지칭한다. 링커는 단일 연결 분자로 구성될 수 있거나, 또는 1개 초과 연결 분자를 포함할 수 있다. 일부 실시양태에서, 링커는 유기 분자, 기, 중합체, 또는 화학 모이어티, 예컨대 2가 유기 모이어티일 수 있다. 일부 실시양태에서, 링커는 아미노산 또는 펩티드일 수 있다. 일부 실시양태에서, 펩티드 링커는 약 4 내지 약 100개 또는 그 초과 아미노산의 길이, 예를 들어, 약 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100개 또는 그 초과 아미노산의 길이 (예를 들어, 약 4 내지 약 40, 약 4 내지 약 50, 약 4 내지 약 60, 약 5 내지 약 40, 약 5 내지 약 50, 약 5 내지 약 60, 약 9 내지 약 40, 약 9 내지 약 50, 약 9 내지 약 60, 약 10 내지 약 40, 약 10 내지 약 50, 약 10 내지 약 60개, 또는 약 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25개의 아미노산 내지 약 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100개 또는 그 초과 아미노산의 길이일 수 있다. 일부 실시양태에서, 펩티드 링커는 GS 링커일 수 있다.

[0071] "프로모터"는 프로모터와 작동가능하게 회합된 뉴클레오티드 서열 (예를 들어, 코딩 서열)의 전사를 제어하거나 조절하는 뉴클레오티드 서열이다. 프로모터에 의해 제어되거나 조절되는 코딩 서열은 폴리펩티드 및/또는 기능적 RNA를 코딩할 수 있다. 전형적으로, "프로모터"는 RNA 폴리머라제 II에 대한 결합 부위를 함유하고 전사의 개시를 지정하는 뉴클레오티드 서열을 지칭한다. 일반적으로, 프로모터는 상응하는 코딩 서열의 코딩 영역의 시작부에 비해 5', 또는 상류에서 발견된다. 프로모터는 유전자 발현의 조절제로서 작용하는 다른 요소; 예를 들어, 프로모터 영역을 포함할 수 있다. 이들은 TATA 박스 컨센서스 서열, 및 종종 CAAT 박스 컨센서스 서열을 포함한다 (Breathnach and Chambon, (1981) *Annu. Rev. Biochem.* 50:349). 식물에서, CAAT 박스는 AGGA 박스에 의해 치환될 수 있다 (Messing *et al.*, (1983) in *Genetic Engineering of Plants*, T. Kosuge, C. Meredith and A. Hollaender (eds.), Plenum Press, pp. 211-227). 일부 실시양태에서, 프로모터 영역은 적어도 1개의 인트론 (예를 들어, **서열식별번호:40** 또는 **서열식별번호:41**)을 포함할 수 있다.

[0072] 본 발명에 유용한 프로모터는 예를 들어, 재조합 핵산 분자, 예를 들어, "합성 핵산 구축물" 또는 "단백질-RNA 복합체"의 제조에 사용하기 위한 구성적, 유도성, 시간적으로 조절된, 발달적으로 조절된, 화학적으로 조절된, 조직-바람직한 및/또는 조직-특이적 프로모터를 포함할 수 있다. 이들 다양한 유형의 프로모터는 관련 기술분야에 공지되어 있다.

[0073] 프로모터의 선택은 발현을 위한 시간적 및 공간적 요건에 따라 다양할 수 있으며, 또한 형질전환되는 숙주 세포에 기반하여 다양할 수 있다. 많은 상이한 유기체에 대한 프로모터는 관련 기술분야에 널리 공지되어 있다. 관련 기술분야에 존재하는 광범위한 지식에 기반하여, 적절한 프로모터는 관심의 특정 숙주 유기체에 대해 선택될 수 있다. 따라서, 예를 들어, 모델 유기체에서 고도로 구성적으로 발현된 유전자의 상류의 프로모터에 관하여 많은 것이 공지되어 있으며, 이러한 지식은 용이하게 접근되고, 적절한 경우 다른 시스템에서 실행될 수 있다.

[0074] 일부 실시양태에서, 식물에서 기능적인 프로모터는 본 발명의 구축물에 사용될 수 있다. 식물에서 발현을 유도하는데 유용한 프로모터의 비-제한적 예는 루비스코(RubisCo) 소형 서브유닛 유전자 1의 프로모터 (PrbcS1), 액틴 유전자의 프로모터 (액틴(Pactin)), 니트레이트 리덕타제 유전자의 프로모터 (Pnr) 및 중복된 카르보닉 안히드라제 유전자 1의 프로모터 (Pdca1)를 포함한다 (문헌 [Walker *et al.* *Plant Cell Rep.* 23:727-735 (2005)];

[Li et al. *Gene* 403:132-142 (2007)]; [Li et al. *Mol Biol. Rep.* 37:1143-1154 (2010)] 참조). PrbcS1 및 팩틴은 구성적 프로모터이고, Pnr 및 Pdca1은 유도성 프로모터이다. Pnr은 니트레이트에 의해 유도되고, 암모늄에 의해 억제되며 (Li et al. *Gene* 403:132-142 (2007)), Pdca1은 염에 의해 유도된다 (Li et al. *Mol Biol. Rep.* 37:1143-1154 (2010)).

[0075]

식물에 유용한 구성적 프로모터의 예는 세스트룸 바이러스 프로모터 (cmp) (미국 특허 번호 7,166,770), 벼 액틴 1 프로모터 (Wang et al. (1992) *Mol. Cell. Biol.* 12:3399-3406; 뿐만 아니라 미국 특허 번호 5,641,876), CaMV 35S 프로모터 (Ode11 et al. (1985) *Nature* 313:810-812), CaMV 19S 프로모터 (Lawton et al. (1987) *Plant Mol. Biol.* 9:315-324), nos 프로모터 (Ebert et al. (1987) *Proc. Natl. Acad. Sci USA* 84:5745-5749), Adh 프로모터 (Walker et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:6624-6629), 수크로스 신타제 프로모터 (Yang & Russell (1990) *Proc. Natl. Acad. Sci. USA* 87:4144-4148), 및 유비퀴틴 프로모터를 포함하나 이에 제한되지는 않는다. 유비퀴틴으로부터 유래된 구성적 프로모터는 많은 세포 유형에서 축적된다. 유비퀴틴 프로모터는 트랜스제닉 식물에 사용하기 위한 몇몇 식물 중, 예를 들어, 해바라기 (Binet et al., 1991. *Plant Science* 79: 87-94), 메이즈 (Christensen et al., 1989. *Plant Molec. Biol.* 12: 619-632), 및 아라비도시스 (Norris et al. 1993. *Plant Molec. Biol.* 21:895-906)로부터 클로닝되었다. 메이즈 유비퀴틴 프로모터 (UbiP)는 트랜스제닉 단자엽식물 시스템에서 개발되었으며, 그의 서열 및 단자엽식물 형질전환을 위해 구축된 벡터는 특허 공개 EP 0 342 926에 개시되어 있다. 유비퀴틴 프로모터는 트랜스제닉 식물, 특히 단자엽식물에서의 본 발명의 뉴클레오티드 서열의 발현에 적합하다. 추가로, 문헌 [McElroy et al. (*Mol. Gen. Genet.* 231: 150-160 (1991))]에 의해 기재된 프로모터 발현 카세트는 본 발명의 뉴클레오티드 서열의 발현을 위해 용이하게 변형될 수 있으며, 단자엽식물 숙주에 사용하는데 특히 적합하다.

[0076]

일부 실시양태에서, 조직 특이적/조직 바람직한 프로모터는 식물 세포에서 이중 폴리뉴클레오티드의 발현에 사용될 수 있다. 조직 특이적 또는 바람직한 발현 패턴은 녹색 조직 특이적 또는 바람직한, 뿌리 특이적 또는 바람직한, 줄기 특이적 또는 바람직한, 꽃 특이적 또는 바람직한 또는 화분 특이적 또는 바람직한 것을 포함하나 이에 제한되지는 않는다. 녹색 조직에서의 발현에 적합한 프로모터는 광합성에 관여하는 유전자를 조절하는 많은 것을 포함하며, 이들 중 많은 것은 단자엽식물 및 쌍자엽식물 둘 다로부터 클로닝되었다. 한 실시양태에서, 본 발명에 유용한 프로모터는 포스포에놀 카르복실라제 유전자로부터의 메이즈 PEPC 프로모터이다 (Hudspeth & Guala, *Plant Molec. Biol.* 12:579-589 (1989)). 조직-특이적 프로모터의 비-제한적 예는 종자 저장 단백질 (예컨대 β -콘글리시닌, 크루시페린, 나핀 및 파세올린), 제인 또는 오일 바디 단백질 (예컨대 올레오신), 또는 지방산 생합성에 관여하는 단백질 (아실 운반체 단백질, 스테아로일-ACP 데사투라제 및 지방산 데사투라제 (fad2-1)를 포함함)을 코딩하는 유전자, 및 배 발생 동안 발현되는 다른 핵산 (예컨대 Bce4, 예를 들어, 문헌 [Kridl et al. (1991) *Seed Sci. Res.* 1:209-219]; 뿐만 아니라 유럽 특허 번호 255378 참조)과 연관된 것들을 포함한다. 식물, 특히 메이즈에서 본 발명의 뉴클레오티드 서열의 발현에 유용한 조직-특이적 또는 조직-우선적 프로모터는 뿌리, 종과피, 잎 또는 화분에서 발현을 지정하는 것들을 포함하나 이에 제한되지는 않는다. 이러한 프로모터는 예를 들어, 그 전문이 본원에 참조로 포함되는 WO 93/07278에 개시되어 있다. 본 발명에 유용한 조직 특이적 또는 조직 바람직한 프로모터의 다른 비-제한적 예는 미국 특허 6,040,504에 개시된 면화 루비스코 프로모터; 미국 특허 5,604,121에 개시된 벼 수크로스 신타제 프로모터; 문헌 [de Framond (FEBS 290:103-106 (1991))], 시바-가이키(Ciba-Geigy)의 EP 0 452 269에 의해 기재된 뿌리 특이적 프로모터; 미국 특허 5,625,136 (시바-가이키)에 기재되고, 메이즈 trpA 유전자의 발현을 유도하는 줄기 특이적 프로모터; WO 01/73087에 개시된 세스트룸 황화 및 말린 바이러스 프로모터; 및 벼로부터의 ProOsLPS10 및 ProOsLPS11 (Nguyen et al. *Plant Biotechnol. Reports* 9(5):297-306 (2015)), 메이즈로부터의 ZmSTK2_USP (Wang et al. *Genome* 60(6):485-495 (2017)), 토마토로부터의 LAT52 및 LAT59 (Twell et al. *Development* 109(3):705-713 (1990)), Zm13 (미국 특허 번호 10,421,972), 아라비도시스로부터의 PLA₂- δ 프로모터 (미국 특허 번호 7,141,424), 및/또는 메이즈로부터의 ZmC5 프로모터 (국제 PCT 공개 번호 W01999/042587을 포함하나 이에 제한되지는 않는 화분 특이적 또는 바람직한 프로모터를 포함한다.

[0077]

식물 조직-특이적/조직 바람직한 프로모터의 추가의 예는 뿌리 텔-특이적 **시스**-요소 (RHE) (Kim et al. *The Plant Cell* 18:2958-2970 (2006)), 뿌리-특이적 프로모터 RCc3 (Jeong et al. *Plant Physiol.* 153:185-197 (2010)) 및 RB7 (미국 특허 번호 5459252), 락틴 프로모터 (Lindstrom et al. (1990) *Der. Genet.* 11:160-167; 및 Vodkin (1983) *Prog. Clin. Biol. Res.* 138:87-98), 옥수수 알콜 데히드로게나제 1 프로모터 (Dennis et al. (1984) *Nucleic Acids Res.* 12:3983-4000), S-아테노실-L-메티오닌 신테타제 (SAMS) (Vander Mijnsbrugge et al. (1996) *Plant and Cell Physiology*, 37(8):1108-1115), 옥수수 광 수확 복합체 프로모터 (Bansal et

al. (1992) *Proc. Natl. Acad. Sci. USA* 89:3654-3658), 옥수수 열 충격 단백질 프로모터 (O'Dell *et al.* (1985) *EMBO J.* 5:451-458; 및 Rochester *et al.* (1986) *EMBO J.* 5:451-458), 완두콩 소형 서브유닛 RuBP 카르복실라제 프로모터 (Cashmore, "Nuclear genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase" pp. 29-39 In: *Genetic Engineering of Plants* (Hollaender ed., Plenum Press 1983; 및 Poulsen *et al.* (1986) *Mol. Gen. Genet.* 205:193-200), Ti 플라스미드 만노핀 신타제 프로모터 (Langridge *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:3219-3223), Ti 플라스미드 노팔린 신타제 프로모터 (Langridge *et al.* (1989), 상기 문헌), 페투니아 칼콘 이소머라제 프로모터 (van Tunen *et al.* (1988) *EMBO J.* 7:1257-1263), 콩 글리신 풍부 단백질 1 프로모터 (Keller *et al.* (1989) *Genes Dev.* 3:1639-1646), 말단절단된 CaMV 35S 프로모터 (O'Dell *et al.* (1985) *Nature* 313:810-812), 감자 파타틴 프로모터 (Wenzler *et al.* (1989) *Plant Mol. Biol.* 13:347-354), 뿌리 세포 프로모터 (Yamamoto *et al.* (1990) *Nucleic Acids Res.* 18:7449), 메이즈 제인 프로모터 (Kriz *et al.* (1987) *Mol. Gen. Genet.* 207:90-98; Langridge *et al.* (1983) *Cell* 34:1015-1022; Reina *et al.* (1990) *Nucleic Acids Res.* 18:6425; Reina *et al.* (1990) *Nucleic Acids Res.* 18:7449; 및 Wandelt *et al.* (1989) *Nucleic Acids Res.* 17:2354), 글로불린-1 프로모터 (Belanger *et al.* (1991) *Genetics* 129:863-872), α -튜불린 cab 프로모터 (Sullivan *et al.* (1989) *Mol. Gen. Genet.* 215:431-440), PEPC아제 프로모터 (Hudspeth & Grula (1989) *Plant Mol. Biol.* 12:579-589), R 유전자 복합체-연관된 프로모터 (Chandler *et al.* (1989) *Plant Cell* 1:1175-1183), 및 칼콘 신타제 프로모터 (Franken *et al.* (1991) *EMBO J.* 10:2605-2612)를 포함하나 이에 제한되지는 않는다.

[0078] 종자-특이적 발현에 유용한 것은 완두콩 비실린 프로모터 (Czako *et al.* (1992) *Mol. Gen. Genet.* 235:33-40; 뿐만 아니라 미국 특허 번호 5,625,136에 개시된 종자-특이적 프로모터이다. 성숙한 잎에서의 발현을 위한 유용한 프로모터는 노쇠의 개시에서 스위칭된 것들, 예컨대 아라비도시스으로부터의 SAG 프로모터 (Gan *et al.* (1995) *Science* 270:1986-1988)이다.

[0079] 또한, 엽록체에서 기능적인 프로모터가 사용될 수 있다. 이러한 프로모터의 비-제한적 예는 박테리오파지 T3 유전자 9 5' UTR 및 미국 특허 번호 7,579,516에 개시된 다른 프로모터를 포함한다. 본 발명에 유용한 다른 프로모터는 S-E9 소형 서브유닛 RuBP 카르복실라제 프로모터 및 쿠니츠(Kunitz) 트립신 억제제 유전자 프로모터 (Kti3)를 포함하나 이에 제한되지는 않는다.

[0080] 본 발명에 유용한 추가의 조절 요소는 인트론, 인핸서, 종결 서열 및/또는 5' 및 3' 비번역된 영역을 포함하나 이에 제한되지는 않는다.

[0081] 본 발명에 유용한 인트론은 식물에서 확인되고 그로부터 단리되며, 이어서 식물의 형질전환에 사용되도록 발현 카세트 내로 삽입된 인트론일 수 있다. 관련 기술분야의 통상의 기술자에 의해 이해될 것인 바와 같이, 인트론은 자기-삭제에 요구되는 서열을 포함할 수 있으며, 핵산 구축물/발현 카세트 내로 프레임 내에서 혼입된다. 인트론은 한 핵산 구축물에서 다수의 단백질-코딩 서열을 분리하는 스페이서로서 사용될 수 있거나, 또는 인트론은 한 단백질-코딩 서열 내부에서, 예를 들어, mRNA를 안정화시키기 위해 사용될 수 있다. 이들이 단백질-코딩 서열 내에서 사용되는 경우, 이들은 포함된 삭제 부위와 "프레임 내에서" 삽입된다. 인트론은 또한 발현을 개선시키거나 변형시키기 위해 프로모터와 회합될 수 있다. 예로서, 본 발명에 유용한 프로모터/인트론 조합은 메이즈 Ubi1 프로모터 및 인트론의 그것을 포함하나 이에 제한되지는 않는다.

[0082] 본 발명에 유용한 인트론의 비-제한적 예는 ADHI 유전자 (예를 들어, *Adh1-S* 인트론 1, 2 및 6), 유비퀴틴 유전자 (Ubi1), 루비스코 소형 서브유닛 (rbcS) 유전자, 루비스코 대형 서브유닛 (rbcL) 유전자, 액틴 유전자 (예를 들어, *액틴-1* 인트론), 피루베이트 데하이드로게나제 키나제 유전자 (pdk), 니트레이트 리덕타제 유전자 (nr), 중복된 카르보닉 안하이드라제 유전자 1 (Tdca1), psbA 유전자, atpA 유전자로부터의 인트론, 또는 이들의 임의의 조합을 포함한다. 비-제한적 예로서, 본 발명의 핵산 구축물은 최적화된 CRISPR-Cas 뉴클레아제 (예를 들어, **서열식별번호:1 내지 11 또는 23 내지 25**) 및 데아미나제를 포함하는 염기 편집제를 코딩할 수 있고, 여기서 핵산 구축물은 인트론을 포함하는/그와 회합된 프로모터를 추가로 포함한다. 추가의 비-제한적 예로서, 본 발명의 핵산 구축물은 최적화된 CRISPR-Cas 뉴클레아제 (예를 들어, **서열식별번호:1 내지 11 또는 23 내지 25**) 및 데아미나제를 포함하는 염기 편집제를 코딩할 수 있고, 여기서 뉴클레아제 및/또는 데아미나제는 1개 이상의 인트론을 포함하고, 임의로, 핵산 구축물은 인트론을 포함하는/그와 회합된 프로모터를 추가로 포함한다.

[0083] 일부 실시양태에서, 본 발명의 폴리뉴클레오타이드 및/또는 핵산 구축물은 "발현 카세트"일 수 있거나, 또는 발현 카세트 내에 포함될 수 있다. 본원에 사용된 "발현 카세트"는 예를 들어, 본 발명의 핵산 구축물 (예를 들어, 본 발명의 변형된 LbCas12a를 코딩함)을 포함하는 제조합 핵산 분자를 의미하고, 여기서 핵산 구축물은 적어도

제어 서열 (예를 들어, 프로모터)에 작동가능하게 회합된다. 따라서, 본 발명의 일부 실시양태는 예를 들어, 본 발명의 핵산 구축물 (예를 들어, 본 발명의 변형된 LbCas12a를 코딩하는 본 발명의 핵산 구축물)을 발현하도록 디자인된 발현 카세트를 제공한다.

[0084] 본 발명의 핵산 구축물을 포함하는 발현 카세트는 키메라일 수 있으며, 이는 그의 성분 중 적어도 하나가 그의 다른 성분 중 적어도 하나에 관하여 이중임을 의미한다 (예를 들어, 숙주 유기체에서 발현되는 관심의 폴리뉴클레오티드에 작동가능하게 연결된 숙주 유기체로부터의 프로모터, 여기서 관심의 폴리뉴클레오티드는 숙주와는 상이한 유기체로부터의 것이거나, 또는 통상적으로 그 프로모터와 연관되어 발견되지 않음). 발현 카세트는 또한 천연 발생이지만 이중 발현에 유용한 재조합 형태로 얻어진 것일 수 있다.

[0085] 발현 카세트는 임의로 선택된 숙주 세포에서 기능적인 전사 및/또는 번역 종결 영역 (즉, 종결 영역) 및/또는 인핸서 영역을 포함할 수 있다. 다양한 전사 종결자 및 인핸서는 관련 기술분야에 공지되어 있으며, 발현 카세트에 사용하기 위해 이용가능하다. 전사 종결자는 전사의 종결 및 정확한 mRNA 폴리아데닐화를 담당한다. 종결 영역 및/또는 인핸서 영역은 전사 개시 영역에 대해 천연일 수 있거나, 예를 들어, 본 발명의 핵산 구축물에 의해 코딩되는 LbCas12a 뉴클레아제를 코딩하는 유전자에 대해 천연일 수 있거나, 숙주 세포에 대해 천연일 수 있거나, 또는 또 다른 공급원에 대해 천연일 수 있다 (예를 들어, 프로모터에 대해, 본 발명의 핵산 구축물에 의해 코딩되는 LbCas12a 뉴클레아제를 코딩하는 유전자에 대해, 숙주 세포에 대해 외래 또는 이중, 또는 이들의 임의의 조합). 인핸서 영역은 본 발명의 핵산 구축물에 의해 코딩되는 LbCas12a 뉴클레아제를 코딩하는 유전자에 대해 천연일 수 있거나, 숙주 세포에 대해 천연일 수 있거나, 또는 또 다른 공급원으로부터의 것일 수 있다 (예를 들어, 프로모터에 대해, 본 발명의 핵산 구축물에 의해 코딩되는 LbCas12a 뉴클레아제를 코딩하는 유전자에 대해, 숙주 세포에 대해 외래 또는 이중, 또는 이들의 임의의 조합).

[0086] 본 발명의 발현 카세트는 또한 형질전환된 숙주 세포를 선택하는데 사용될 수 있는 선택가능한 마커를 코딩하는 뉴클레오티드 서열을 포함할 수 있다. 본원에 사용된 "선택가능한 마커"는 발현되는 경우 마커를 발현하는 숙주 세포에 대해 별개의 표현형을 부여하고, 따라서 이러한 형질전환된 세포가 마커를 갖지 않는 것들로부터 구별되는 것을 허용하는 뉴클레오티드 서열을 의미한다. 이러한 뉴클레오티드 서열은 마커가 화학적 수단에 의해, 예컨대 선택적 작용제 (예를 들어, 항생제 등)를 사용함으로써 선택될 수 있는 소질을 부여하는지 여부, 또는 마커가 단순히 관찰 또는 시험을 통해, 예컨대 스크리닝 (예를 들어, 형광)에 의해 확인할 수 있는 소질인지 여부에 따라, 선택가능한 또는 스크리닝가능한 마커 중 어느 하나를 코딩할 수 있다. 적합한 선택가능한 마커의 많은 예는 관련 기술분야에 공지되어 있으며, 본원에 기재된 발현 카세트에 사용될 수 있다.

[0087] 발현 카세트 외에도, 본원에 기재된 핵산 분자/구축물 및 폴리뉴클레오티드 서열은 벡터와 관련되어 사용될 수 있다. 용어 "벡터"는 핵산 (또는 핵산들)을 세포 내로 전달하거나(transferring), 전달하거나(delivering), 도입하기 위한 조성물을 지칭한다. 벡터는 전달되거나, 전달되거나 도입되는 뉴클레오티드 서열(들)을 포함하는 핵산 구축물을 포함한다. 숙주 유기체의 형질전환에 사용하기 위한 벡터는 관련 기술분야에 널리 공지되어 있다. 벡터의 일반적 부류의 비-제한적 예는 자기 전파가능하거나 동원가능할 수 있거나 그렇지 않을 수 있는 이중 또는 단일 가닥 선형 또는 원형 형태의 바이러스 벡터, 플라스미드 벡터, 파지 벡터, 파지미드 벡터, 코스미드 벡터, 포스미드 벡터, 박테리오파지, 인공 염색체, 미니서클, 또는 아그로박테리움(Agrobacterium) 2원 벡터를 포함한다. 일부 실시양태에서, 바이러스 벡터는 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연관된, 또는 단순 포진 바이러스 벡터를 포함할 수 있으나 이에 제한되지는 않는다. 본원에서 정의된 바와 같은 벡터는 세포 게놈 내로의 통합에 의해 원핵생물 또는 진핵생물 숙주를 형질전환시키거나, 염색체외적으로 존재할 수 있다 (예를 들어 복제 원점을 갖는 자율 복제 플라스미드). 추가로 포함되는 것은 서틀 벡터이며, 이는 방선균류 및 관련된 종, 박테리아 및 진핵생물 (예를 들어 고등 식물, 포유동물, 효모 또는 진균 세포)로부터 선택될 수 있는 2종의 상이한 숙주 유기체에서 자연적으로 또는 디자인에 의해 복제 가능한 DNA 비히클을 의미한다. 일부 실시양태에서, 벡터 중의 핵산은 숙주 세포에서의 전사를 위한 적절한 프로모터 또는 다른 조절 요소의 제어 하에 있고, 그에 작동가능하게 연결된다. 벡터는 다수의 숙주에서 기능하는 이중-기능적 발현 벡터일 수 있다. 게놈 DNA의 경우에, 이는 그 자신의 프로모터 및/또는 다른 조절 요소를 함유할 수 있고, cDNA의 경우에, 이는 숙주 세포에서의 발현을 위한 적절한 프로모터 및/또는 다른 조절 요소의 제어 하에 있을 수 있다. 따라서, 본 발명의 핵산 구축물 및/또는 이를 포함하는 발현 카세트는 본원에 기재된 바와 같은 및 관련 기술분야에 공지된 바와 같은 벡터에 포함될 수 있다. 일부 실시양태에서, 벡터는 고 카피 수 벡터 (예를 들어, 고 카피 수 *E. coli* 벡터; 예를 들어, pUC, p블루스크립트(pBluescript), pGEM 등)일 수 있다. 따라서, 예를 들어, 본 발명의 라이브러리는 고 카피 수 벡터를 사용하여 구축될 수 있다.

[0088] 본원에 사용된 "접촉시키다", "접촉시키는", "접촉된", 및 그의 문법적 파생어는 바람직한 반응의 성분을 바람

직한 반응 (예를 들어, 형질전환, 전사 제어, 게놈 편집, 니킹, 및/또는 절단)을 수행하는데 적합한 조건 하에서 함께 놓는 것을 지칭한다. 따라서, 예를 들어, 표적 핵산은 폴리뉴클레오티드/핵산 구축물이 발현되고 변형된 LbCas12a 뉴클레아제가 생산되는 조건 하에서, (a) 본 발명의 변형된 LbCas12a 뉴클레아제를 코딩하는 본 발명의 폴리뉴클레오티드 및/또는 핵산 구축물 및 (b) 가이드 핵산과 접촉될 수 있고, 여기서 뉴클레아제는 가이드 핵산과의 복합체를 형성하고, 복합체는 표적 핵산에 혼성화하고, 그에 의해 표적 핵산을 변형시킨다. 일부 실시양태에서, 표적 핵산은 (a) 본 발명의 변형된 LbCas12a 뉴클레아제 및/또는 이를 포함하는 융합 단백질 (예를 들어, 본 발명의 변형된 LbCas12a 뉴클레아제 및 관심의 폴리펩티드 (예를 들어, 데아미나제)) 및 (b) 가이드 핵산과 접촉될 수 있고, 여기서 변형된 LbCas12a 뉴클레아제는 가이드 핵산과의 복합체를 형성하고, 복합체는 표적 핵산에 혼성화하고, 그에 의해 표적 핵산을 변형시킨다. 본원에 기재된 바와 같이, 표적 핵산은 가이드 핵산과의 접촉 전에, 그와 공동으로, 또는 후에 본 발명의 폴리뉴클레오티드/핵산 구축물/폴리펩티드와 접촉될 수 있다.

[0089] 표적 핵산에 관하여 본원에 사용된 "변형시키는" 또는 "변형"은 표적 핵산의 편집 (예를 들어, 돌연변이화), 공유 변형, 핵산/뉴클레오티드 염기의 교환/치환, 결실, 절단, 니킹, 및/또는 전사 제어를 포함한다.

[0090] 관심의 폴리뉴클레오티드의 맥락에서 "도입하는", "도입하다", "도입된" (및 그의 문법적 파생어)은 뉴클레오티드 서열이 세포의 내부에의 접근을 얻도록 하는 방식으로 관심의 뉴클레오티드 서열 (예를 들어, 폴리뉴클레오티드, 핵산 구축물, 및/또는 가이드 핵산)을 숙주 유기체 또는 상기 유기체의 세포 (예를 들어, 숙주 세포; 예를 들어, 식물 세포)에 제시하는 것을 의미한다. 따라서, 예를 들어, 본원에 기재된 바와 같은 변형된 LbCas12a 뉴클레아제를 코딩하는 본 발명의 폴리뉴클레오티드 및 가이드 핵산은 유기체의 세포 내로 도입되고, 그에 의해 세포를 변형된 LbCas12a 뉴클레아제 및 가이드 핵산으로 형질전환시킬 수 있다.

[0091] 본원에 사용된 용어 "형질전환"은 세포 내로의 이중 핵산의 도입을 지칭한다. 세포의 형질전환은 안정하거나 일시적일 수 있다. 따라서, 일부 실시양태에서, 숙주 세포 또는 숙주 유기체는 본 발명의 폴리뉴클레오티드/핵산 분자로 안정하게 형질전환될 수 있다. 일부 실시양태에서, 숙주 세포 또는 숙주 유기체는 본 발명의 폴리뉴클레오티드/핵산 구축물로 일시적으로 형질전환될 수 있다.

[0092] 폴리뉴클레오티드의 맥락에서 "일시적 형질전환"은 폴리뉴클레오티드가 세포 내로 도입되고, 세포의 게놈 내로 통합되지 않는 것을 의미한다.

[0093] 세포 내로 도입된 폴리뉴클레오티드의 맥락에서 "안정하게 도입하는" 또는 "안정하게 도입된"이란, 도입된 폴리뉴클레오티드가 세포의 게놈 내로 안정하게 혼입되고, 따라서 세포가 폴리뉴클레오티드로 안정하게 형질전환되는 것으로 의도된다.

[0094] 본원에 사용된 "안정한 형질전환" 또는 "안정하게 형질전환된"은 핵산 분자가 세포 내로 도입되고, 세포의 게놈 내로 통합되는 것을 의미한다. 따라서, 통합된 핵산 분자는 그의 자손에 의해, 보다 특히, 다수의 연속적 세대의 자손에 의해 유전될 수 있다. 본원에 사용된 "게놈"은 핵 및 색소체 게놈을 포함하고, 따라서 예를 들어, 엽록체 또는 미토콘드리아 게놈 내로의 핵산의 통합을 포함한다. 본원에 사용된 안정한 형질전환은 또한 염색체외적으로, 예를 들어, 미니염색체 또는 플라스미드로서 유지되는 트랜스진을 지칭할 수 있다.

[0095] 일시적 형질전환은 예를 들어, 유기체 내로 도입된 1개 이상의 트랜스진에 의해 코딩된 펩티드 또는 폴리펩티드의 존재를 검출할 수 있는 효소-연결 면역흡착 검정 (ELISA) 또는 웨스턴 블롯에 의해 검출될 수 있다. 세포의 안정한 형질전환은 예를 들어, 유기체 (예를 들어, 식물) 내로 도입된 트랜스진의 뉴클레오티드 서열과 특이적으로 혼성화하는 핵산 서열을 갖는 세포의 게놈 DNA의 서던 블롯 혼성화 검정에 의해 검출될 수 있다. 세포의 안정한 형질전환은 예를 들어, 숙주 유기체 내로 도입된 트랜스진의 뉴클레오티드 서열과 특이적으로 혼성화하는 핵산 서열을 갖는 세포의 RNA의 노던 블롯 혼성화 검정에 의해 검출될 수 있다. 세포의 안정한 형질전환은 또한 예를 들어, 표준 방법에 따라 검출될 수 있는 트랜스진의 표적 서열(들)과 혼성화하여, 트랜스진 서열의 증폭을 발생시키는 특이적 프라이머 서열을 채용하는, 관련 기술분야에 널리 공지된 바와 같은 폴리머라제 연쇄 반응 (PCR) 또는 다른 증폭 반응에 의해 검출될 수 있다. 형질전환은 또한 관련 기술분야에 널리 공지된 직접적 시퀀싱 및/또는 혼성화 프로토콜에 의해 검출될 수 있다.

[0096] 따라서, 일부 실시양태에서, 본 발명의 뉴클레오티드 서열, 폴리뉴클레오티드, 및/또는 핵산 구축물 및/또는 이를 포함하는 발현 카세트 및/또는 벡터는 일시적으로 발현될 수 있고/거나, 이들은 숙주 유기체의 게놈 내로 안정하게 혼입될 수 있다. 따라서, 일부 실시양태에서, 본 발명의 핵산 구축물 (예를 들어, 본 발명의 변형된 LbCas12a 뉴클레아제, 또는 그의 융합 단백질; 예를 들어, 예를 들어, 관심의 폴리뉴클레오티드, 예를 들어, 데

아미나제 도메인에 연결된 변형된 LbCas12a 뉴클레아제를 포함하는 융합 단백질을 코딩함) (여기서 변형된 LbCas12a 뉴클레아제를 코딩하는 핵산 구축물은 유기체 (예를 들어, 식물, 포유동물, 진균, 박테리아 등)에서의 발현을 위해 코돈 최적화됨)은 가이드 핵산과 함께 및 따라서 세포에서 유지된 DNA 없이 유기체의 세포 내로 일시적으로 도입될 수 있다.

[0097] 본 발명의 폴리뉴클레오티드/핵산 구축물은 관련 기술분야의 통상의 기술자에게 공지된 임의의 방법에 의해 세포 내로 도입될 수 있다. 본 발명의 일부 실시양태에서, 세포의 형질전환은 핵 형질전환을 포함한다. 다른 실시양태에서, 세포의 형질전환은 색소체 형질전환 (예를 들어, 엽록체 형질전환)을 포함한다. 추가의 실시양태에서, 본 발명의 폴리뉴클레오티드/핵산 구축물은 통상적인 육종 기술을 통해 세포 내로 도입될 수 있다.

[0098] 진핵생물 및 원핵생물 유기체 둘 다를 형질전환시키는 절차는 관련 기술분야에 널리 공지되어 있고 통상적이며, 문헌 전반에 걸쳐 기재되어 있다 (예를 들어, 문헌 [Jiang et al. 2013. *Nat. Biotechnol.* 31:233-239]; [Ran et al. *Nature Protocols* 8:2281-2308 (2013)] 참조).

[0099] 따라서 뉴클레오티드 서열은 관련 기술분야에 널리 공지된 임의의 수의 방식으로 숙주 유기체 또는 그의 세포 내로 도입될 수 있다. 본 발명의 방법은 이들이 유기체의 적어도 하나의 세포의 내부에의 접근을 얻는 한, 유기체 내로 1개 이상의 뉴클레오티드 서열을 도입하는 특정 방법에 의존하지 않는다. 1개 초과 뉴클레오티드 서열이 도입되어야 하는 경우, 이들은 단일 핵산 구축물의 일부로서, 또는 별개의 핵산 구축물로서 어셈블리될 수 있고, 동일한 또는 상이한 핵산 구축물 상에 위치할 수 있다. 따라서, 뉴클레오티드 서열은 단일 형질전환 사건에서, 및/또는 별개의 형질전환 사건에서 관심의 세포 내로 도입될 수 있거나, 또는, 대안적으로, 관련되는 경우, 뉴클레오티드 서열은 예를 들어, 육종 프로토콜의 일부로서 식물 내로 혼입될 수 있다.

[0100] 본 발명은 비-천연 PAM 인식 부위/서열을 포함하도록 변형된 Cas12a 뉴클레아제 (예를 들어, 그 특정 Cas12a 뉴클레아제에 대한 천연 PAM 인식 특이성에 추가로 또는 그 대신에 비-천연 PAM 인식 특이성을 포함하는 Cas12a 뉴클레아제)에 관한 것이다. 또한, 본 발명은 개선된 PAM 인식 특이성을 포함한 바람직한 특징을 갖는 Cas12a 뉴클레아제를 디자인하고, 확인하고, 선택하는 방법에 관한 것이다.

[0101] 변형된 Cas12a 폴리펩티드에 관하여 본원에 사용된 "변경된 PAM 특이성"은 뉴클레아제의 PAM 특이성이 야생형 뉴클레아제의 그것으로부터 변경된 것을 의미한다 (예를 들어, 비-천연 PAM 서열은 천연 PAM 서열에 추가로 및/또는 그 대신에 인식된다. 예를 들어, 변형된 Cas12a 뉴클레아제는 그것이 TTV (여기서 V는 A, C 또는 G임)의 천연 Cas12a PAM 서열 이외의 및/또는 그에 추가의 PAM 서열을 인식하는 경우, 그의 PAM 특이성에 있어서 변경될 것이다.

[0102] 본 발명은 변형된 PAM 인식 특이성을 갖는 LbCas12a 뉴클레아제에 관한 것이다. 일부 실시양태에서, 본 발명은 변형된 *라크노스피라세아에* 박테리아 CRISPR (클러스터링된 규칙적 간격의 짧은 회문식 반복부) Cas12a (LbCas12a) 폴리펩티드를 제공하며 여기서 변형된 LbCas12a 폴리펩티드는 **서열식별번호:1** (LbCas12a)의 아미노산 서열과 적어도 80% 동일성 (예를 들어, 약 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100% 동일성; 예를 들어, 약 80% 내지 약 100%, 약 85% 내지 약 100%, 약 90% 내지 약 100% 약 95% 내지 약 100%)을 갖는 아미노산 서열 및 **서열식별번호:1**의 위치 넘버링에 관하여 하기 위치 K116, K120, K121, D122, E125, T148, T149, T152, D156, E159, Q529, G532, D535, K538, D541, Y542, L585, K591, M592, K595, V596, S599, K600, K601, Y616, Y646, 및/또는 W649 중 하나 또는 하나 초과 (예를 들어, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13개 또는 그 초과)에 돌연변이, 임의로 **서열식별번호:1**의 위치 넘버링에 관하여 하기 위치 K116, K120, K121, D122, E125, T152, D156, E159, G532, D535, K538, D541, 및/또는 K595 중 하나 또는 하나 초과에 돌연변이를 포함한다. 일부 실시양태에서, Cas12a (LbCas12a) 폴리펩티드의 돌연변이는 임의의 조합으로 서열식별번호:1의 위치 넘버링에 관하여 하기 위치 K116, K120, K121, D122, E125, T152, D156, E159, G532, D535, K538, D541, 및/또는 K595 중 하나 또는 하나 초과에 돌연변이를 포함하거나, 이로 본질적으로 이루어지거나, 이로 이루어진다. 따라서, 본 발명의 변형된 LbCas12a 폴리펩티드는 **서열식별번호:1**의 위치 넘버링에 관하여 K116, K120, K121, D122, E125, T148, T149, T152, D156, E159, Q529, G532, D535, K538, D541, Y542, L585, K591, M592, K595, V596, S599, K600, K601, Y616, Y646, 및/또는 W649의 위치 중 임의의 하나에 단일 돌연변이를 포함할 수 있거나, 또는 **서열식별번호:1**의 위치 넘버링에 관하여 K116, K120, K121, D122, E125, T148, T149, T152, D156, E159, Q529, G532, D535, K538, D541, Y542, L585, K591, M592, K595, V596, S599, K600, K601, Y616, Y646, 및/또는 W649 중 임의의 2개 이상의 위치에 돌연변이의 조합을 포함할 수 있다.

[0103] 일부 실시양태에서, Cas12a (LbCas12a) 폴리펩티드의 돌연변이는 **서열식별번호:1**의 위치 넘버링에 관하여 하기

돌연변이 K116N, K116R, K120H, K120N, K120Q, K120R, K120T, K121D, K121G, K121H, K121Q, K121R, K121S, K121T, D122H, D122K, D122N, D122R, E125K, E125Q, E125R, E125Y, T148A, T148C, T148H, T148S, T149C, T149F, T149G, T149H, T149N, T149P, T149S, T149V, T152E, T152F, T152H, T152K, T152L, T152Q, T152R, T152W, T152Y, D156E, D156H, D156I, D156K, D156L, D156Q, D156R, D156W, D156Y, E159K, E159Q, E159R, E159Y, Q529A, Q529D, Q529F, Q529G, Q529H, Q529N, Q529P, Q529S, Q529T, Q529W, G532A, G532C, G532D, G532F, G532H, G532K, G532L, G532N, G532Q, G532S, D535A, D535H, D535K, D535N, D535S, D535T, D535V, K538C, K538F, K538G, K538H, K538L, K538M, K538Q, K538R, K538V, K538W, K538Y, D541A, D541E, D541H, D541I, D541N, D541R, D541Y, Y542F, Y542H, Y542K, Y542L, Y542M, Y542N, Y542R, Y542T, Y542V, L585F, L585G, L585H, K591A, K591F, K591G, K591H, K591R, K591S, K591W, K591Y, M592A, M592E, M592Q, K595H, K595L, K595M, K595Q, K595R, K595S, K595W, K595Y, V596H, V596T, S599G, S599H, S599N, K600G, K600H, K600R, K601H, K601Q, K601R, K601T, Y616E, Y616F, Y616H, Y616K, Y616R, Y646E, Y646H, Y646K, Y646N, Y646Q, Y646R, Y646W, W649H, W649K, W649R, W649S 및/또는 W649Y 중 하나 또는 하나 초과를 포함하거나, 이로 본질적으로 이루어지거나, 이로 이루어진다. 이해될 것인 바와 같이, 2개 이상의 돌연변이를 갖는 임의의 단일 Cas12a 폴리펩티드는 임의의 주어진 위치에 단지 단일 돌연변이를 포함할 것이다. 따라서, 예를 들어, 폴리펩티드는 D535A, D535H, D535K, D535N, D535S, D535T, 또는 D535V 중 임의의 하나의 위치 D535에 돌연변이를 가질 수 있지만, 동일한 폴리펩티드는 본원에 기재된 바와 같은 임의의 다른 위치 중 하나 또는 하나 초과에 돌연변이를 추가로 포함할 수 있다. 일부 실시양태에서, Cas12a (LbCas12a) 폴리펩티드의 돌연변이는 임의의 조합으로 **서열식별번호:1**의 잔기 위치 넘버링에 관하여 K116N, K116R, K120H, K120N, K120Q, K120R, K120T, K121D, K121G, K121H, K121Q, K121R, K121S, K121T, D122H, D122K, D122N, D122R, E125K, E125Q, E125R, E125Y, T152E, T152F, T152H, T152K, T152L, T152Q, T152R, T152W, T152Y, D156E, D156H, D156I, D156K, D156L, D156Q, D156R, D156W, D156Y, E159K, E159Q, E159R, E159Y, G532A, G532C, G532D, G532F, G532H, G532K, G532L, G532N, G532Q, G532S, D535A, D535H, D535K, D535N, D535S, D535T, D535V, K538C, K538F, K538G, K538H, K538L, K538M, K538Q, K538R, K538V, K538W, K538Y, D541A, D541E, D541H, D541I, D541N, D541R, D541Y, K595H, K595L, K595M, K595Q, K595R, K595S, K595W, 및/또는 K595Y 중 하나 또는 하나 초과를 포함하거나, 이로 본질적으로 이루어지거나, 이로 이루어진다. 일부 실시양태에서, Cas12a (LbCas12a) 폴리펩티드의 돌연변이는 **서열식별번호:1**의 위치 넘버링에 관하여 K116R, K116N, K120Y, K121S, K121R, D122H, D122N, E125K, T152R, T152K, T152Y, T152Q, T152E, T152F, D156R, D156W, D156Q, D156H, D156I, D156V, D156L, D156E, E159K, E159R, G532N, G532S, G532H, G532K, G532R, G532L, D535N, D535H, D535T, D535S, D535A, D535W, K538R, K538V, K538Q, K538W, K538Y, K538F, K538H, K538L, K538M, K538C, K538G, K538A, D541E, K595R, K595Q, K595Y, K595W, K595H, K595S, 및/또는 K595M 중 하나 또는 하나 초과를 포함하거나, 이로 본질적으로 이루어지거나, 이로 이루어진다. 이해될 것인 바와 같이, 2개 이상의 돌연변이를 갖는 임의의 단일 Cas12a 폴리펩티드는 임의의 주어진 위치에 단지 단일 돌연변이를 포함할 것이다. 따라서, 예를 들어, 폴리펩티드는 D535A, D535H, D535K, D535N, D535S, D535T, 또는 D535V 중 임의의 하나의 위치 D535에 돌연변이를 가질 수 있고, 본원에 기재된 바와 같은 임의의 다른 위치에 하나 또는 하나 초과에 돌연변이를 추가로 포함할 수 있다.

[0104] 일부 실시양태에서, 돌연변이는 **서열식별번호:1**의 위치 넘버링에 관하여 D156R, G532R, K538R, K538V, Y542R 또는 K595R의 돌연변이를 포함하거나, 이로 본질적으로 이루어지거나, 이로 이루어지지 않는다. 일부 실시양태에서, Cas12a (LbCas12a) 폴리펩티드의 돌연변이는 **서열식별번호:1**의 위치 넘버링에 관하여 G532R 및 K595R의, G532R, K538V 및 Y542R의, 또는 D156R, G532R 및 K532R의 돌연변이의 조합을 포함하거나, 이로 본질적으로 이루어지거나, 이로 이루어지지 않는다.

[0105] 일부 실시양태에서, 변형된 LbCas12a 폴리펩티드는 **표 2에 (실시예 2에)** 제시된 바와 같은 **서열식별번호:1**의 1개 이상의 아미노산 돌연변이를 포함할 수 있다.

[0106] 일부 실시양태에서, 변형된 LbCas12a 폴리펩티드는 야생형 LbCas12a (예를 들어, **서열식별번호:1**)와 비교하여 변경된 프로토스페이서 인접 모티프 (PAM) 특이성을 포함할 수 있다. 본 발명의 변형된 LbCas12a 폴리펩티드는 변경된 PAM 특이성을 포함할 수 있고, 여기서 변경된 PAM 특이성은 NNNG, NNNT, NNNA, NNNC, NNG, NNT, NNC, NNA, NG, NT, NC, NA, NN, NNN, NNNN을 포함하나 이에 제한되지 않고, 여기서 각각의 서열의 각각의 N은 T, C, G, 또는 A 중 임의의 것으로부터 독립적으로 선택된다. 일부 실시양태에서, 변경된 PAM 특이성은 TTTA, TTTC, TTTG, TTTT, TTCA, TTCC, TTCG, TTCT, ATTC, CTTA, CTTC, CTG, GTTC, TATA, TATC, CTCC, TCCG, TACA, TCCG, TACA, TCCG, TCCC, TCCA, 및/또는 TATG를 포함할 수 있으나 이에 제한되지는 않는다. 일부 실시양태에서, 변경된 PAM 특이성은 NNNN일 수 있고, 여기서 각각의 서열의 각각의 N은 T, C, G, 또는 A 중 임의의 것으로부터 독

립적으로 선택된다.

- [0107] 변경된 PAM 인식 특이성을 갖는 것 외에도, 변형된 LbCas12a 뉴클레아제는 뉴클레아제 활성 부위 (예를 들어, RuvC 도메인)에 돌연변이를 추가로 포함할 수 있다 (예를 들어, 죽은LbCas12a, dLbCas12a). 이러한 변형은 감소된 뉴클레아제 활성 (예를 들어, 니카제 활성)을 갖거나 뉴클레아제 활성을 갖지 않는 LbCas12a 폴리펩티드를 발생시킬 수 있다.
- [0108] 일부 실시양태에서, (a) (i) 본 발명의 변형된 LbCas12a 뉴클레아제 또는 본 발명의 변형된 LbCas12a 뉴클레아제를 코딩하는 핵산, 및 (ii) 관심의 폴리펩티드 또는 관심의 폴리펩티드를 코딩하는 핵산을 포함하는 융합 단백질; 및 (b) 스페이서 서열 및 반복 서열을 포함하는 가이드 핵산 (CRISPR RNA, CRISPR DNA, crRNA, crDNA)을 포함하며, 여기서 가이드 핵산은 변형된 LbCas12a 뉴클레아제 또는 융합 단백질과 복합체를 형성할 수 있고, 스페이서 서열은 표적 핵산에 혼성화하여, 변형된 LbCas12a 뉴클레아제 및 관심의 폴리펩티드를 표적 핵산에 가이드할 수 있고, 그에 의해 시스템은 표적 핵산을 변형시키거나 (예를 들어, 절단 또는 편집) 또는 조정할 (예를 들어, 전사를 조정함) 수 있는 것인 유형 V 클러스터링된 규칙적 간격의 짧은 회문식 반복부 (CRISPR) - 연관된 (Cas) (CRISPR-Cas) 시스템이 제공된다. 일부 실시양태에서, 시스템은 임의로 펩티드 링커를 통해 변형된 LbCas12a 뉴클레아제 (예를 들어, 융합 단백질)의 C-말단 및/또는 N-말단에 연결된 관심의 폴리펩티드를 포함한다.
- [0109] 추가로, 본 발명의 변형된 Cas12a 뉴클레아제를 포함하는 융합 단백질에 본원에서 제공된다. 일부 실시양태에서, 융합 단백질은 변형된 LbCas12a의 C-말단에 및/또는 N-말단에 연결된 관심의 폴리펩티드를 포함할 수 있다. 일부 실시양태에서, 본 발명은 변형된 LbCas12a 및 관심의 폴리펩티드를 연결하는 임의적 개재 링커를 포함하는 융합 단백질을 제공한다.
- [0110] 융합 단백질의 활성을 방해하지 않는 관련 기술분야에 공지되어 있거나 나중에 확인될 임의의 링커가 사용될 수 있다. 융합 단백질의 활성을 "방해하지" 않는 링커는 융합 단백질 (예를 들어, 뉴클레아제 및/또는 관심의 폴리펩티드)의 폴리펩티드의 활성을 감소시키거나 제거하지 않는 링커이며; 즉, 뉴클레아제 활성, 핵산 결합 활성, 편집 활성, 및/또는 뉴클레아제 또는 관심의 펩티드의 임의의 다른 활성은 뉴클레아제 및 관심의 폴리펩티드가 링커를 통해 서로에 결합된 융합 단백질에서 유지된다. 일부 실시양태에서, 펩티드 링커는 (예를 들어, 그의 N-말단에서) 변형된 LbCas12a의 C-말단에 연결될 수 있고, 임의로 여기서 융합 단백질은 링커의 C-말단에 연결된 관심의 폴리펩티드를 추가로 포함할 수 있다. 일부 실시양태에서, 펩티드 링커는 (예를 들어, 그의 C-말단에서) 변형된 LbCas12a의 N-말단에 연결될 수 있고, 임의로 여기서 융합 단백질은 링커의 N-말단에 연결된 관심의 폴리펩티드를 추가로 포함할 수 있다. 일부 실시양태에서, 본 발명의 변형된 LbCas12a는 (직접적으로 또는 링커를 통해) 그의 C-말단 및 N-말단 둘 다에서 링커 및/또는 관심의 폴리펩티드에 연결될 수 있다.
- [0111] 일부 실시양태에서, 본 발명에 유용한 링커는 아미노산 또는 펩티드일 수 있다. 일부 실시양태에서, 본 발명에 유용한 펩티드 링커는 약 4 내지 약 100개 또는 그 초과 아미노산의 길이, 예를 들어, 약 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100개 또는 그 초과 아미노산의 길이 (예를 들어, 약 4 내지 약 40, 약 4 내지 약 50, 약 4 내지 약 60, 약 5 내지 약 40, 약 5 내지 약 50, 약 5 내지 약 60, 약 9 내지 약 40, 약 9 내지 약 50, 약 9 내지 약 60, 약 10 내지 약 40, 약 10 내지 약 50, 약 10 내지 약 60개, 또는 약 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25개의 아미노산 내지 약 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100개 또는 그 초과 아미노산의 길이)일 수 있다. 일부 실시양태에서, 펩티드 링커는 GS 링커일 수 있다.
- [0112] 본 발명에 유용한 관심의 폴리펩티드는 데아미나제 (탈아미노화) 활성, 니카제 활성, 리코비나제 활성, 트랜스포사제 활성, 메틸라제 활성, 글리코실라제 (DNA 글리코실라제) 활성, 글리코실라제 억제제 활성 (예를 들어, 우라실-DNA 글리코실라제 억제제 (UGI)), 데메틸라제 활성, 전사 활성화 활성, 전사 억제 활성, 전사 방출 인자 활성, 히스톤 변형 활성, 뉴클레아제 활성, 단일-가닥 RNA 절단 활성, 이중-가닥 RNA 절단 활성, 제한 엔도뉴클레아제 활성 (예를 들어, Fok1), 핵산 결합 활성, 메틸트랜스퍼라제 활성, DNA 복구 활성, DNA 손상 활성, 디스

뮤타제 활성, 알킬화 활성, 탈퓨린화 활성, 산화 활성, 피리미딘 이량체 형성 활성, 인테그라제 활성, 트랜스포사제 활성, 폴리머라제 활성, 리가제 활성, 헬리카제 활성, 및/또는 포토리아제 활성을 갖는 폴리펩티드 또는 단백질 도메인을 포함할 수 있으나 이에 제한되지는 않는다.

[0113] 일부 실시양태에서, 관심의 폴리펩티드는 데아미나제 활성을 갖는 적어도 1종의 폴리펩티드 또는 단백질 도메인을 포함할 수 있다. 일부 실시양태에서, 적어도 1종의 폴리펩티드 또는 단백질 도메인은 아데닌 데아미나제 도메인일 수 있다. 본 발명에 유용한 아데닌 데아미나제 (또는 아데노신 데아미나제)는 임의의 유기체로부터의 임의의 공지된 또는 나중에 확인될 아데닌 데아미나제일 수 있다 (예를 들어, 아데닌 데아미나제의 그의 개시내용에 대해 본원에 참조로 포함되는 미국 특허 번호 10,113,163 참조). 아데닌 데아미나제는 아데닌 또는 아데노신의 가수분해적 탈아미노화를 촉매할 수 있다. 일부 실시양태에서, 아데닌 데아미나제는 아데노신 또는 테옥시아데노신의 각각 이노신 또는 테옥시이노신으로의 가수분해적 탈아미노화를 촉매할 수 있다. 일부 실시양태에서, 아데노신 데아미나제는 DNA에서의 아데닌 또는 아데노신의 가수분해적 탈아미노화를 촉매할 수 있다. 일부 실시양태에서, 본 발명의 핵산 구축물에 의해 코딩되는 아데닌 데아미나제는 표적 핵산의 센스 (예를 들어, "+"; 주형) 가닥에서의 A→G 전환 또는 표적 핵산의 안티센스 (예를 들어, "-", 상보적) 가닥에서의 T→C 전환을 생성할 수 있다.

[0114] 일부 실시양태에서, 아데노신 데아미나제는 천연-발생 아데닌 데아미나제의 변이체일 수 있다. 따라서, 일부 실시양태에서, 본 발명에 유용한 아데노신 데아미나제는 야생형 아데닌 데아미나제와 약 70% 내지 100% 동일할 수 있다 (예를 들어, 천연 발생 아데닌 데아미나제와 약 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일함, 및 그 안의 임의의 범위 또는 값). 일부 실시양태에서, 데아미나제 또는 데아미나제는 자연에서 발생하지 않으며, 조작된, 돌연변이된 또는 진화된 아데노신 데아미나제로 지칭될 수 있다. 따라서, 예를 들어, 조작된, 돌연변이된 또는 진화된 아데닌 데아미나제 폴리펩티드 또는 아데닌 데아미나제 도메인은 천연 발생 아데닌 데아미나제 폴리펩티드/도메인과 약 70% 내지 99.9% 동일할 수 있다 (예를 들어, 천연 발생 아데닌 데아미나제 폴리펩티드 또는 아데닌 데아미나제 도메인과 약 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8% 또는 99.9% 동일함, 및 그 안의 임의의 범위 또는 값). 일부 실시양태에서, 아데노신 데아미나제는 박테리아 (예를 들어, *에스케리키아 콜라이*(*Escherichia coli*), *스타필로코쿠스 아우레우스*(*Staphylococcus aureus*), *해모필루스 인플루엔자에*(*Haemophilus influenzae*), *카울로박터 크레스투스*(*Caulobacter crescentus*) 등)로부터의 것일 수 있다. 일부 실시양태에서, 아데닌 데아미나제 폴리펩티드/도메인을 코딩하는 폴리뉴클레오티드는 유기체에서의 발현을 위해 코돈 최적화될 수 있다.

[0115] 일부 실시양태에서, 아데닌 데아미나제 도메인은 야생형 tRNA-특이적 아데노신 데아미나제 도메인, 예를 들어, tRNA-특이적 아데노신 데아미나제 (TadA) 및/또는 돌연변이된/진화된 아데노신 데아미나제 도메인, 예를 들어, 돌연변이된/진화된 tRNA-특이적 아데노신 데아미나제 도메인 (TadA*)일 수 있다. 일부 실시양태에서, TadA 도메인은 *이. 콜라이*로부터의 것일 수 있다. 일부 실시양태에서, TadA는 변형될, 예를 들어, 말단절단될, 전장 TadA에 비해 1개 이상의 N-말단 및/또는 C-말단 아미노산을 소실할 수 있다 (예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, 또는 20개의 N-말단 및/또는 C 말단 아미노산 잔기는 전장 TadA에 비해 소실될 수 있다). 일부 실시양태에서, TadA 폴리펩티드 또는 TadA 도메인은 N-말단 메티오닌을 포함하지 않는다. 일부 실시양태에서, 야생형 *이. 콜라이* TadA는 **서열식별번호:18**의 아미노산 서열을 포함한다. 일부 실시양태에서, 돌연변이된/진화된 *이. 콜라이* TadA*는 **서열식별번호:19 내지 22**의 아미노산 서열을 포함한다. 일부 실시양태에서, TadA/TadA*를 코딩하는 폴리뉴클레오티드는 유기체에서의 발현을 위해 코돈 최적화될 수 있다.

[0116] 본 발명에 유용한 시토신 데아미나제 (또는 시티딘 데아미나제)는 임의의 유기체로부터의 임의의 공지된 또는 나중에 확인될 시토신 데아미나제일 수 있다 (예를 들어, 시토신 데아미나제의 그의 개시내용에 대해 본원에 참조로 포함되는 미국 특허 번호 10,167,457 참조). 일부 실시양태에서, 적어도 1종의 폴리펩티드 또는 단백질 도메인은 시토신 데아미나제 폴리펩티드 또는 도메인일 수 있다. 일부 실시양태에서, 시토신 데아미나제 폴리펩티드/도메인은 아포지질단백질 B mRNA 편집 촉매 폴리펩티드-유사 (APOBEC) 도메인일 수 있다. 일부 실시양태에서, 관심의 폴리펩티드는 글리코실라제 억제제 활성을 갖는 적어도 1종의 폴리펩티드 또는 단백질 도메인을 포함할 수 있다. 일부 실시양태에서, 관심의 폴리펩티드는 우라실-DNA 글리코실라제 억제제 (UGI) 폴리펩티드/도메인일 수 있다. 일부 실시양태에서, 본 발명의 변형된 LbCas12a 뉴클레아제 및 시토신 데아미나제 도메인을

코딩하는 (예를 들어, 변형된 LbCas12a 뉴클레아제 및 시토신 데아미나제 도메인을 포함하는 융합 단백질을 코딩하는) 핵산 구조물은 우라실-DNA 글리코실라제 억제제 (UGI)를 추가로 코딩할 수 있고, 여기서 UGI는 유기체에서의 발현을 위해 코돈 최적화된다. 일부 실시양태에서, 본 발명은 변형된 LbCas12a 뉴클레아제, 시토신 데아미나제 도메인, 및 UGI 및/또는 이를 코딩하는 1종 이상의 폴리뉴클레오티드를 포함하는 융합 단백질을 제공하고, 임의로 여기서 1종 이상의 폴리뉴클레오티드는 유기체에서의 발현을 위해 코돈 최적화될 수 있다.

[0117] 시토신 데아미나제는 시티딘 또는 테옥시시티딘의 각각 우리딘 또는 테옥시우리딘으로의 가수분해적 탈아미노화를 촉매한다. 일부 실시양태에서, 데아미나제 또는 데아미나제 도메인은 시토신의 우라실로의 가수분해적 탈아미노화를 촉매하는 시티딘 데아미나제 도메인일 수 있다. 일부 실시양태에서, 시토신 데아미나제는 영양류 (예를 들어, 인간, 원숭이, 침팬지, 고릴라), 개, 소, 래트 또는 마우스를 포함하나 이에 제한되지는 않는 천연-발생 시토신 데아미나제의 변이체일 수 있다. 따라서, 일부 실시양태에서, 본 발명에 유용한 시토신 데아미나제는 야생형 시토신 데아미나제와 약 70% 내지 약 100% 동일할 수 있다 (예를 들어, 천연 발생 시토신 데아미나제와 약 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일함, 및 그 안의 임의의 범위 또는 값). 일부 실시양태에서, 시토신 데아미나제 폴리펩티드/도메인을 코딩하는 폴리뉴클레오티드는 유기체에서의 발현을 위해 코돈 최적화될 수 있다.

[0118] 일부 실시양태에서, 본 발명에 유용한 시토신 데아미나제는 아포지질단백질 B mRNA-편집 복합체 (APOBEC) 패밀리의 데아미나제일 수 있다. 일부 실시양태에서, 시토신 데아미나제는 APOBEC1 데아미나제, APOBEC2 데아미나제, APOBEC3A 데아미나제, APOBEC3B 데아미나제, APOBEC3C 데아미나제, APOBEC3D 데아미나제, APOBEC3F 데아미나제, APOBEC3G 데아미나제, APOBEC3H 데아미나제, APOBEC4 데아미나제, 인간 활성화 유도된 데아미나제 (hAID), rAPOBEC1, FERNY, 및/또는 CDA1, 임의로 pmCDA1, atCDA1 (예를 들어, At2g19570), 및 이의 진화된 버전일 수 있다. 일부 실시양태에서, 시토신 데아미나제는 **서열식별번호:23**, **서열식별번호:44** 또는 **서열식별번호:46**의 아미노산 서열을 갖는 APOBEC1 데아미나제일 수 있다. 일부 실시양태에서, 시토신 데아미나제는 **서열식별번호:24**의 아미노산 서열을 갖는 APOBEC3A 데아미나제일 수 있다. 일부 실시양태에서, 시토신 데아미나제는 CDA1 데아미나제, 임의로 **서열식별번호:25** 또는 **서열식별번호:43**의 아미노산 서열을 갖는 CDA1일 수 있다. 일부 실시양태에서, 시토신 데아미나제는 FERNY 데아미나제, 임의로 **서열식별번호:42** 또는 **서열식별번호:45**의 아미노산 서열을 갖는 FERNY일 수 있다. 일부 실시양태에서, 시토신 데아미나제는 **서열식별번호:47** 또는 **서열식별번호:48**의 아미노산 서열을 갖는 인간 활성화 유도된 데아미나제 (hAID)일 수 있다. 일부 실시양태에서, 본 발명에 유용한 시토신 데아미나제는 천연 발생 시토신 데아미나제 (예를 들어, 진화된 데아미나제)의 아미노산 서열과 약 70% 내지 약 100% 동일할 (예를 들어, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% 또는 100% 동일할) 수 있다. 일부 실시양태에서, 본 발명에 유용한 시토신 데아미나제는 **서열식별번호:23**, **서열식별번호:24**, **서열식별번호:25** 또는 **서열식별번호:42** 내지 **48**의 아미노산 서열과 약 70% 내지 약 99.5% 동일할 (예를 들어, 약 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 99.5% 동일할) (예를 들어, **서열식별번호:23**, **서열식별번호:24**, **서열식별번호:25** 또는 **서열식별번호:42** 내지 **48**의 아미노산 서열과 적어도 80%, 적어도 85%, 적어도 90%, 적어도 92%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일할) 수 있다. 일부 실시양태에서, 시토신 데아미나제를 코딩하는 폴리뉴클레오티드는 유기체에서의 발현을 위해 코돈 최적화될 수 있고, 코돈 최적화된 폴리펩티드는 참조 폴리뉴클레오티드와 약 70% 내지 99.5% 동일할 수 있다.

[0119] 본 발명에 유용한 "우라실 글리코실라제 억제제" (UGI)는 우라실-DNA 글리코실라제 염기-삭제 복구 효소를 억제할 수 있는 임의의 단백질일 수 있다. 일부 실시양태에서, UGI 도메인은 야생형 UGI 또는 그의 단편을 포함한다. 일부 실시양태에서, 본 발명에 유용한 UGI 도메인은 천연 발생 UGI 도메인의 아미노산 서열과 약 70% 내지 약 100% 동일할 수 있다 (예를 들어, 약 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% 또는 100% 동일함 및 그 안의 임의의 범위 또는 값). 일부 실시양태에서, UGI 도메인은 **서열식별번호:26**의 아미노산 서열 또는 **서열식별번호:26**의 아미노산 서열과 약 70% 내지 약 99.5% 동일성을 갖는 폴리펩티드를 포함할 수 있다 (예를 들어, **서열식별번호:26**의 아미노산 서열과 적어도 80%, 적어도 85%, 적어도 90%, 적어도 92%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 적어도 99.5% 동일함). 예를 들어, 일부 실시양태에서, UGI 도메인은 **서열식별번호:26**의 아미노산 서열의 연속적 뉴클레오티드 (예를 들어, 약 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80개의 연속적 뉴클레오티드; 예를 들어, 약 10, 15, 20, 25,

30, 35, 40, 45, 내지 약 50, 55, 60, 65, 70, 75, 80개의 연속적 뉴클레오타이드)의 부분과 100% 동일한 서열식별번호:26의 아미노산 서열의 단편을 포함할 수 있다. 일부 실시양태에서, UGI 도메인은 공지된 UGI와 70% 내지 약 99.5% 동일성 (예를 들어, 약 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% 동일성, 및 그 안의 임의의 범위 또는 값)을 갖는 공지된 UGI (예를 들어, **서열식별번호:26**)의 변이체일 수 있다. 일부 실시양태에서, UGA를 코딩하는 폴리뉴클레오타이드는 유기체에서의 발현을 위해 코돈 최적화될 수 있고, 코돈 최적화된 폴리펩티드는 참조 폴리뉴클레오타이드와 약 70% 내지 약 99.5% 동일할 수 있다.

[0120] 일부 실시양태에서, 변형된 LbCas12a 뉴클레아제는 그의 뉴클레아제 활성 부위 (예를 들어, RuvC)에 돌연변이를 포함할 수 있다. 그의 뉴클레아제 활성 부위(들)에 돌연변이를 갖고 뉴클레아제 활성을 더 이상 포함하지 않는 변형된 LbCas12a 뉴클레아제는 통상적으로 "죽은", 예를 들어, dLbCas12a로 지칭된다. 일부 실시양태에서, 그의 뉴클레아제 활성 부위(들)에 돌연변이를 갖는 변형된 LbCas12a 도메인 또는 폴리펩티드는 돌연변이를 갖지 않는 동일한 LbCas12a 뉴클레아제와 비교하여 손상된 활성 또는 감소된 활성 (예를 들어, 니카제 활성)을 가질 수 있다.

[0121] 본 발명의 변형된 LbCas12a 뉴클레아제는 변형된 LbCas12a 뉴클레아제와 함께 기능하여 표적 핵산을 변형시키도록 디자인된 가이드 RNA (gRNA, CRISPR 어레이, CRISPR RNA, crRNA)와 조합으로 사용될 수 있다. 본 발명에 유용한 가이드 핵산은 적어도 스페이서 서열 및 반복 서열을 포함한다. 가이드 핵산은 변형된 LbCas12a 뉴클레아제를 코딩하는 본 발명의 폴리뉴클레오타이드/핵산 구축물에 의해 코딩되고 발현되는 LbCas12a 뉴클레아제 도메인과 복합체를 형성할 수 있고, 스페이서 서열은 표적 핵산에 혼성화할 수 있고, 그에 의해 핵산 구축물 (예를 들어, 변형된 LbCas12a 뉴클레아제 (및/또는 관심의 폴리펩티드))을 표적 핵산에 가이드하고, 여기서 표적 핵산은 변형된 LbCas12a 뉴클레아제 (및/또는 코딩된 데아미나제 도메인 및/또는 관심의 폴리펩티드)에 의해 변형될 (예를 들어, 절단될 또는 편집될) 또는 조정될 (예를 들어, 전사를 조정할) 수 있다. 예로서, 시토신 데아미나제 도메인에 연결된 LbCas12a 도메인 (예를 들어, 융합 단백질)을 코딩하는 핵산 구축물은 LbCas12a 가이드 핵산과 조합으로 사용되어 표적 핵산을 변형시킬 수 있고, 여기서 융합 단백질의 시토신 데아미나제 도메인은 표적 핵산에서 시토신 염기를 탈아미노화하고, 그에 의해 표적 핵산을 편집한다. 추가의 예에서, 아데닌 데아미나제 도메인에 연결된 LbCas12a 도메인 (예를 들어, 융합 단백질)을 코딩하는 핵산 구축물은 LbCas12a 가이드 핵산과 조합으로 사용되어 표적 핵산을 변형시킬 수 있고, 여기서 융합 단백질의 아데닌 데아미나제 도메인은 표적 핵산에서 아데닌 염기를 탈아미노화하고, 그에 의해 표적 핵산을 편집한다.

[0122] 본원에 사용된 "가이드 핵산", "가이드 RNA", "gRNA", "CRISPR RNA/DNA", "crRNA" 또는 "crDNA"는 표적 핵산 (예를 들어, 프로토스페이서)에 상보적인 (및 그에 혼성화하는) 적어도 1개의 스페이서 서열, 및 적어도 1개의 반복 서열 (예를 들어, 유형 V Cas12a CRISPR-Cas 시스템의 반복부, 또는 그의 단편 또는 부분)을 포함하는 핵산을 의미하고, 여기서 반복 서열은 스페이서 서열의 5' 단부 및/또는 3' 단부에 연결될 수 있다. 본 발명의 gRNA의 디자인은 유형 V Cas12a 시스템에 기반할 수 있다.

[0123] 일부 실시양태에서, Cas12a gRNA는 5'에서 3'로, 반복 서열 (전장 또는 그의 부분 ("핸들")); 예를 들어, 슈도노트-유사 구조) 및 스페이서 서열을 포함할 수 있다.

[0124] 일부 실시양태에서, 가이드 핵산은 1개 초과 "반복 서열-스페이서" 서열 (예를 들어, 2, 3, 4, 5, 6, 7, 8, 9, 10개, 또는 그 초과 반복부-스페이서 서열) (예를 들어, 반복부-스페이서-반복부, 예를 들어, 반복부-스페이서-반복부-스페이서-반복부-스페이서-반복부-스페이서-반복부-스페이서 등)을 포함할 수 있다. 본 발명의 가이드 핵산은 합성, 인간-제조이며, 자연에서 발견되지 않는다. gRNA는 매우 길 수 있고, 압타머 (MS2 동원 전락에서와 같이) 또는 스페이서를 늘어뜨린 다른 RNA 구조로서 사용될 수 있다.

[0125] 본원에 사용된 "반복 서열"은 예를 들어, 야생형 Cas12a 유전자좌 (예를 들어, LbCas12a 유전자좌)의 임의의 반복 서열 또는 본 발명의 핵산 구축물에 의해 코딩되는 LbCas12a 뉴클레아제와 함께 기능적인 합성 crRNA의 반복 서열을 지칭한다. 본 발명에 유용한 반복 서열은 Cas12a 유전자좌의 임의의 공지된 또는 나중에 확인될 반복 서열일 수 있거나, 또는 이는 Cas12a 유형 V CRISPR-Cas 시스템에서 기능하도록 디자인된 합성 반복부일 수 있다. 반복 서열은 헤어핀 구조 및/또는 줄기 루프 구조를 포함할 수 있다. 일부 실시양태에서, 반복 서열은 그의 5' 단부에서 슈도노트-유사 구조 (즉, "핸들")를 형성할 수 있다. 따라서, 일부 실시양태에서, 반복 서열은 야생형 V CRISPR-Cas 유전자좌 (예를 들어, 야생형 Cas12a 유전자좌)로부터의 반복 서열과 동일하거나 실질적으로 동일할 수 있다. 야생형 Cas12a 유전자좌로부터의 반복 서열은 확립된 알고리즘을 통해, 예컨대 CRISPRdb를 통해 제공되는 CRISPR파인더를 사용하여 결정될 수 있다 (문헌 [Grissa et al. *Nucleic Acids Res.* 35(Web

Server issue):W52-7] 참조). 일부 실시양태에서, 반복 서열 또는 그의 부분은 그의 3' 단부에서 스페이스 서열의 5' 단부에 연결되고, 그에 의해 반복부-스페이스 서열 (예를 들어, 가이드 RNA, crRNA)을 형성한다.

[0126] 일부 실시양태에서, 반복 서열은 특정 반복부 및 반복부를 포함하는 가이드 RNA가 프로세싱되는지 비프로세싱되는지 여부에 따라 적어도 10개의 뉴클레오타이드 (예를 들어, 약 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 내지 100개 또는 그 초과)의 뉴클레오타이드, 또는 그 안의 임의의 범위 또는 값; 예를 들어, 약)를 포함하거나, 이로 본질적으로 이루어지거나, 이로 이루어진다. 일부 실시양태에서, 반복 서열은 약 10 내지 약 20, 약 10 내지 약 30, 약 10 내지 약 45, 약 10 내지 약 50, 약 15 내지 약 30, 약 15 내지 약 40, 약 15 내지 약 45, 약 15 내지 약 50, 약 20 내지 약 30, 약 20 내지 약 40, 약 20 내지 약 50, 약 30 내지 약 40, 약 40 내지 약 80, 약 50 내지 약 100개, 또는 그 초과)의 뉴클레오타이드를 포함하거나, 이로 본질적으로 이루어지거나, 이로 이루어진다.

[0127] 스페이스 서열의 5' 단부에 연결된 반복 서열은 반복 서열의 부분 (예를 들어, 야생형 반복 서열의 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35개 또는 그 초과)의 연속적 뉴클레오타이드)을 포함할 수 있다. 일부 실시양태에서, 스페이스 서열의 5' 단부에 연결된 반복 서열의 부분은 약 5 내지 약 10개의 연속적 뉴클레오타이드의 길이 (예를 들어, 약 5, 6, 7, 8, 9, 10개의 뉴클레오타이드)일 수 있고, 야생형 CRISPR Cas 반복 뉴클레오타이드 서열의 동일한 영역 (예를 들어, 5' 단부)과 적어도 90% 동일성 (예를 들어, 적어도 약 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 그 초과)을 갖는다. 일부 실시양태에서, 반복 서열의 부분은 그의 5' 단부에 슈도노트-유사 구조 (예를 들어, "헨들")를 포함할 수 있다.

[0128] 본원에 사용된 "스페이스 서열"은 표적 핵산 (예를 들어, 표적 DNA) (예를 들어, 프로토스페이스)에 상보적인 뉴클레오타이드 서열이다. 스페이스 서열은 표적 핵산에 완전히 상보적이거나 실질적으로 상보적일 수 있다 (예를 들어, 적어도 약 70% 상보적 (예를 들어, 약 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 그 초과, 및 그 안의 임의의 범위 또는 값)). 따라서, 일부 실시양태에서, 스페이스 서열은 표적 핵산과 비교하여 1, 2, 3, 4, 또는 5개의 미스매치를 가질 수 있으며, 미스매치는 인접하거나 비인접할 수 있다. 일부 실시양태에서, 스페이스 서열은 표적 핵산에 약 70% 상보성을 가질 수 있다. 다른 실시양태에서, 스페이스 뉴클레오타이드 서열은 표적 핵산에 약 80% 상보성을 가질 수 있다. 추가의 다른 실시양태에서, 스페이스 뉴클레오타이드 서열은 표적 핵산 (프로토스페이스)에 약 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 99.5% 상보성 등을 가질 수 있다. 일부 실시양태에서, 스페이스 서열은 표적 핵산에 100% 상보적이다. 스페이스 서열은 약 15개의 뉴클레오타이드 내지 약 30개의 뉴클레오타이드 (예를 들어, 약 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 또는 30개의 뉴클레오타이드, 또는 그 안의 임의의 범위 또는 값)의 길이를 가질 수 있다. 따라서, 일부 실시양태에서, 스페이스 서열은 적어도 약 15개의 뉴클레오타이드 내지 약 30개의 뉴클레오타이드의 길이일 수 있는 표적 핵산 (예를 들어, 프로토스페이스)의 영역에 걸쳐 완전한 상보성 또는 실질적 상보성을 가질 수 있다. 일부 실시양태에서, 스페이스는 약 20, 21, 22, 23, 24, 또는 25개의 뉴클레오타이드의 길이일 수 있다. 일부 실시양태에서, 스페이스는 23개의 뉴클레오타이드의 길이일 수 있다.

[0129] 일부 실시양태에서, 가이드 RNA의 스페이스 서열의 5' 영역은 표적 핵산과 동일할 수 있는 반면, 스페이스의 3' 영역은 표적 핵산에 실질적으로 상보적일 수 있거나 (예를 들어, 유형 V CRISPR-Cas), 또는 가이드 RNA의 스페이스 서열의 3' 영역은 표적 핵산과 동일할 수 있는 반면, 스페이스의 5' 영역은 표적 핵산에 실질적으로 상보적일 수 있고 (예를 들어, 유형 II CRISPR-Cas), 따라서, 표적 핵산에 대한 스페이스 서열의 전체적 상보성은 100% 미만일 수 있다. 따라서, 예를 들어, 유형 V CRISPR-Cas 시스템에 대한 가이드에서, 예를 들어, 20 뉴클레오타이드 스페이스 서열의 5' 영역 (즉, 시드 영역)의 최초 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개의 뉴클레오타이드는 표적 핵산에 100% 상보적일 수 있는 반면, 스페이스 서열의 3' 영역에서 나머지 뉴클레오타이드는 표적 핵산에 실질적으로 상보적 (예를 들어, 적어도 약 70% 상보적)일 수 있다. 일부 실시양태에서, 스페이스 서열의 5' 단부의 최초 1 내지 8개의 뉴클레오타이드 (예를 들어, 최초 1, 2, 3, 4, 5, 6, 7, 8개의 뉴클레오타이드, 및 그 안의 임의의 범위)는 표적 핵산에 100% 상보적일 수 있는 반면, 스페이스 서열의 3' 영역에서의 나머지 뉴클레오타이드는 표적 핵산에 실질적으로 상보적 (예를 들어, 적어도 약 50% 상보적 (예를 들어, 약 50%, 55%, 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 그 초과))일 수 있다.

[0130] 일부 실시양태에서, 스페이스의 시드 영역은 약 8 내지 약 10개의 뉴클레오타이드의 길이, 약 5 내지 약 6개의 뉴

클레오티드의 길이, 또는 약 6개의 뉴클레오티드의 길이일 수 있다.

[0131] 본원에 사용된 "표적 핵산", "표적 DNA", "표적 뉴클레오티드 서열", "표적 영역", 또는 "게놈에서의 표적 영역"은 본 발명의 가이드 RNA에서의 스페이서 서열에 완전히 상보적 (100% 상보적) 또는 실질적으로 상보적 (예를 들어, 적어도 70% 상보적 (예를 들어, 약 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 초과, 및 그 안의 임의의 범위 또는 값))인 유기체의 게놈의 영역을 지칭한다. 일부 실시양태에서, 유형 V CRISPR-Cas 시스템 (예를 들어, LbCas12a)에 유용한 표적 영역은 유기체의 게놈 (예를 들어, 식물 게놈, 동물 게놈, 박테리아 게놈)에서 PAM 서열에 바로 3'에 위치한다. 일부 실시양태에서, 표적 영역은 PAM 서열에 바로 인접하여 위치한 임의의 적어도 15개의 연속적 뉴클레오티드 (예를 들어, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30개 또는 그 초과)의 뉴클레오티드, 및 그 안의 임의의 범위 또는 값; 예를 들어, 약 19 내지 약 25개의 뉴클레오티드, 약 20 내지 약 24개의 뉴클레오티드의 길이 등)로부터 선택될 수 있다.

[0132] "프로토스페이서 서열"은 CRISPR 반복부-스페이서 서열 (예를 들어, 가이드 RNA, CRISPR 어레이, crRNA)의 스페이서 서열에 완전히 또는 실질적으로 상보적인 (및 혼성화하는) 표적 핵산 및 구체적으로 표적 핵산의 부분 (예를 들어, 또는 게놈에서의 표적 영역)을 지칭한다.

[0133] 유형 V CRISPR-Cas Cas12a 시스템의 경우에, 프로토스페이서 서열은 프로토스페이서 인접 모티프 (PAM)에 플랭킹된다 (그에 바로 인접함). PAM은 비-표적 가닥 상의 5' 단부에 및 표적 가닥의 3' 단부에 위치한다 (예로서, 하기 참조).

5'-NNNNNNNNNNNNNNNNNNNNNN-3' RNA 스페이서 (서열식별번호:27)
 |||||
 3'-AAANNNNNNNNNNNNNNNNNNNN-5' 표적 가닥 (서열식별번호:28)
 ||||
 5'-TTNNNNNNNNNNNNNNNNNNNN-3' 비-표적 가닥 (서열식별번호:29)

[0134]

[0135] 정준 Cas12a PAM은 T 풍부성이다. 일부 실시양태에서, 정준 Cas12a PAM 서열은 5'-TTN, 5'-TTTN, 또는 5'-TTTV일 수 있다.

[0136] 본 발명의 폴리펩티드, 융합 단백질 및/또는 시스템은 폴리뉴클레오티드 또는 핵산 구축물에 의해 코딩될 수 있다. 일부 실시양태에서, 본 발명의 폴리펩티드, 융합 단백질 및/또는 시스템을 코딩하는 폴리뉴클레오티드/핵산 구축물은 본원에 기재된 바와 같은 관심의 유기체 및/또는 관심의 유기체의 세포에서의 발현을 위해 조절 요소 (예를 들어, 프로모터, 종결자 등)와 작동가능하게 회합될 수 있다. 일부 실시양태에서, 본 발명의 폴리펩티드, 융합 단백질 및/또는 시스템을 코딩하는 폴리뉴클레오티드/핵산 구축물은 유기체에서의 발현을 위해 코돈 최적화될 수 있다.

[0137] 일부 실시양태에서, 본 발명은 (a) 본 발명의 변형된 LbCas12a 폴리펩티드 또는 본 발명의 융합 단백질 및 (b) 가이드 핵산 (예를 들어, CRISPR RNA, CRISPR DNA, crRNA, crDNA)을 포함하는 복합체를 제공한다.

[0138] 일부 실시양태에서, 본 발명은 (a) 본 발명의 변형된 LbCas12a 폴리펩티드 또는 본 발명의 융합 단백질 및 (b) 가이드 핵산을 포함하는 조성물을 제공한다.

[0139] 일부 실시양태에서, 본 발명은 본 발명의 폴리뉴클레오티드/핵산 구축물을 포함하는 발현 카세트 및/또는 벡터를 제공한다. 일부 실시양태에서, 본 발명의 폴리뉴클레오티드/핵산 구축물 및/또는 1종 이상의 가이드 핵산을 포함하는 발현 카세트 및/또는 벡터가 제공될 수 있다. 일부 실시양태에서, 본 발명의 변형된 CRISPR-Cas 뉴클레아제 및/또는 변형된 CRISPR-Cas 뉴클레아제를 포함하는 융합 단백질을 코딩하는 핵산 구축물은 가이드 핵산을 포함하는 것과 동일한 또는 별개의 발현 카세트 또는 벡터에 포함될 수 있다. 핵산 구축물이 가이드 핵산을 포함하는 것과 별개의 발현 카세트 또는 벡터에 포함되는 경우, 표적 핵산은 가이드 핵산을 포함하는 발현 카세트가 제공되기 (예를 들어, 표적 핵산과 접촉되기) 전에, 그와 공동으로, 또는 그 후에 본 발명의 핵산 구축물을 포함하는 발현 카세트 또는 벡터와 접촉될 (예를 들어, 제공될) 수 있다.

[0140] 일부 실시양태에서, 본 발명은 본 발명의 조성물 및/또는 복합체를 코딩하거나 본 발명의 시스템을 포함하는 발현 카세트 및/또는 벡터를 제공한다.

[0141] 일부 실시양태에서, 유기체에서의 발현을 위해 최적화된 본 발명의 폴리뉴클레오티드, 핵산 구축물, 발현 카세트

트 및/또는 벡터는 본 발명의 동일한 변형된 CRISPR-Cas 뉴클레아제 또는 융합 단백질을 코딩하지만, 유기체에서의 발현을 위해 코돈 최적화되지 않은 폴리뉴클레오타이드, 핵산 구축물, 발현 카세트 및/또는 벡터와 약 70% 내지 약 100% 동일할 (예를 들어, 약 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% 또는 100%, 및 그 안의 임의의 값 또는 범위) 수 있다. 폴리뉴클레오타이드 또는 핵산 구축물이 최적화될 수 있는 유기체는 동물, 식물, 진균, 고세균, 또는 박테리아를 포함할 수 있으나 이에 제한되지는 않는다. 일부 실시양태에서, 본 발명의 폴리뉴클레오타이드 또는 핵산 구축물은 식물에서의 발현을 위해 코돈 최적화된다.

[0142] 일부 실시양태에서, 본 발명은 본 발명의 1종 이상의 폴리뉴클레오타이드, 가이드 핵산, 핵산 구축물, 시스템, 발현 카세트 및/또는 벡터를 포함하는 세포를 제공한다.

[0143] 본 발명의 핵산 구축물 (예를 들어, 본 발명의 변형된 CRISPR-Cas 뉴클레아제 및/또는 본 발명의 변형된 CRISPR-Cas 뉴클레아제를 포함하는 융합 단백질을 코딩함) 및 이를 포함하는 발현 카세트/벡터는 생체내에서 (예를 들어, 유기체 또는 유기체의 세포; 예를 들어, 식물에서) 및 시험관내에서 (예를 들어, 세포 또는 무세포 시스템에서) 표적 핵산 및/또는 그들의 발현을 변형시키는데 사용될 수 있다.

[0144] 본 발명은 Cas12a 폴리펩티드의 PAM 특이성을 변경시키는 방법을 추가로 제공한다. 일부 실시양태에서, 돌연변이를 Cas12a 폴리펩티드 내로 도입하는 것을 포함하는, PAM 특이성을 변경시키는 방법이며, 여기서 돌연변이는 **서열식별번호:1**의 위치 넘버링에 관하여 아미노산 잔기 K116, K120, K121, D122, E125, T148, T149, T152, D156, E159, Q529, D535, K538, D541, Y542, L585, K591, M592, K595, V596, S599, K600, K601, Y616, Y646, W649에 있는 것인 방법이 제공된다. 일부 실시양태에서, Cas12a 폴리펩티드 내로 도입되는 돌연변이는 **서열식별번호:1**의 위치 넘버링에 관하여 K116R, K116N, K120R, K120H, K120N, K120T, K120Y, K120Q, K121S, K121T, K121H, K121R, K121G, K121D, K121Q, D122R, D122K, D122H, D122E, D122N, E125R, E125K, E125Q, E125Y, T148H, T148S, T148A, T148C, T149A, T149C, T149S, T149G, T149H, T149P, T149F, T149N, T149D, T149V, T152R, T152K, T152W, T152Y, T152H, T152Q, T152E, T152L, T152F, D156R, D156K, D156Y, D156W, D156Q, D156H, D156I, D156V, D156L, D156E, E159K, E159R, E159H, E159Y, E159Q, Q529N, Q529T, Q529H, Q529A, Q529F, Q529G, Q529C, Q529S, Q529P, Q529W, Q529D, G532D, G532N, G532S, G532H, G532F, G532K, G532R, G532Q, G532A, G532L, G532C, D535N, D535H, D535V, D535T, D535S, D535A, D535W, D535K, K538R, K538V, K538Q, K538W, K538Y, K538F, K538H, K538L, K538M, K538C, K538G, K538A, K538P, D541N, D541H, D541R, D541K, D541Y, D541I, D541A, D541S, D541E, Y542R, Y542K, Y542H, Y542Q, Y542F, Y542L, Y542M, Y542P, Y542V, Y542N, Y542T, L585G, L585H, L585F, K591W, K591F, K591Y, K591H, K591R, K591S, K591A, K591G, K591P, M592R, M592K, M592Q, M592E, M592A, K595R, K595Q, K595Y, K595L, K595W, K595H, K595E, K595S, K595D, K595M, V596T, V596H, V596G, V596A, S599G, S599H, S599N, S599D, K600R, K600H, K600G, K601R, K601H, K601Q, K601T, Y616K, Y616R, Y616E, Y616F, Y616H, Y646R, Y646E, Y646K, Y646H, Y646Q, Y646W, Y646N, W649H, W649K, W649Y, W649R, W649E, W649S, W649V, 및/또는 W649T이다. 일부 실시양태에서, Cas12a 폴리펩티드 내로 도입되는 돌연변이는 **서열식별번호:1**의 위치 넘버링에 관하여 아미노산 잔기 위치 K116, K120, K121, D122, E125, T152, D156, E159, G532, D535, K538, D541, 및/또는 K595에 있고, 임의로 여기서 돌연변이는 **서열식별번호:1**의 위치 넘버링에 관하여 K116R, K116N, K120Y, K121S, K121R, D122H, D122N, E125K, T152R, T152K, T152Y, T152Q, T152E, T152F, D156R, D156W, D156Q, D156H, D156I, D156V, D156L, D156E, E159K, E159R, G532N, G532S, G532H, G532K, G532R, G532L, D535N, D535H, D535T, D535S, D535A, D535W, K538R, K538V, K538Q, K538W, K538Y, K538F, K538H, K538L, K538M, K538C, K538G, K538A, D541E, K595R, K595Q, K595Y, K595W, K595H, K595S, 및/또는 K595M이다. 도입되는 돌연변이는 단일 돌연변이일 수 있거나, 또는 이는 2개 이상의 돌연변이의 조합일 수 있다. 이해될 것인 바와 같이, 2개 이상의 돌연변이를 갖는 임의의 단일 Cas12a 폴리펩티드는 임의의 주어진 위치에 단지 단일 돌연변이를 포함할 것이다. 일부 실시양태에서, 본 발명의 방법에 의해 PAM 특이성에 있어서 변경되는 Cas12a 폴리펩티드는 LbCas12a 폴리펩티드 (라크노스피라세아에 박테리아)이다.

[0145] 본 발명의 변형된 Cas12a 폴리펩티드 또는 뉴클레아제 (예를 들어, LbCas12a 뉴클레아제)는 세포 또는 무세포 시스템에서 표적 핵산을 변형시키는데 (예를 들어, 표적 핵산을 변경시키는데, 세포/유기체의 계층을 변경시키는데) 사용될 수 있다. 따라서, 일부 실시양태에서, 표적 핵산을 (a) (i) 본 발명의 변형된 LbCas12a 폴리펩티드, 또는 본 발명의 융합 단백질을 (예를 들어, 본 발명의 변형된 LbCas12a 폴리펩티드 및 관심의 폴리펩티드 (예를 들어, 데아미나제)), 및 (ii) 가이드 핵산; (b) (i) 본 발명의 변형된 LbCas12a 폴리펩티드 또는 융합 단백질, 및 (ii) 가이드 핵산을 포함하는 본 발명의 복합체; (c) (i) 본 발명의 변형된 LbCas12a 폴리펩티드, 또는

본 발명의 융합 단백질, 및 (ii) 가이드 핵산을 포함하는 조성물; 및/또는 (d) 본 발명의 시스템과 접촉시키고, 그에 의해 표적 핵산을 변형시키는 것을 포함하는, 표적 핵산을 변형시키는 방법이 제공된다. 일부 실시양태에서, 세포/유기체의 게놈에서의 표적 핵산을 (a) (i) 본 발명의 변형된 LbCas12a 폴리펩티드, 또는 본 발명의 융합 단백질 (예를 들어, 본 발명의 변형된 LbCas12a 폴리펩티드 및 관심의 폴리펩티드 (예를 들어, 데아미나제)), 및 (ii) 가이드 핵산; (b) (i) 본 발명의 변형된 LbCas12a 폴리펩티드 또는 융합 단백질, 및 (ii) 가이드 핵산을 포함하는 본 발명의 복합체; (c) (i) 본 발명의 변형된 CRISPR-Cas 뉴클레아제 (예를 들어, 변형된 LbCas12a 폴리펩티드), 또는 본 발명의 융합 단백질, 및 (ii) 가이드 핵산을 포함하는 조성물; 및/또는 (d) 본 발명의 시스템과 접촉시키고, 그에 의해 세포 또는 유기체의 게놈을 변형시키는/변경시키는 것을 포함하는, 세포 또는 유기체의 게놈을 변형시키는/변경시키는 방법이 제공된다. 일부 실시양태에서, 세포 또는 유기체는 식물 세포 또는 식물이다.

[0146] 일부 실시양태에서, 표적 핵산을 포함하는 세포 또는 무세포 시스템을 (a) (i) 본 발명의 폴리뉴클레오티드 (예를 들어, 본 발명의 변형된 LbCas12a 폴리펩티드를 코딩하거나, 또는 본 발명의 변형된 LbCas12a 폴리펩티드 및 관심의 폴리펩티드 (예를 들어, 데아미나제)를 포함하는 융합 단백질을 코딩함), 또는 이를 포함하는 발현 카세트 또는 벡터, 및 (ii) 가이드 핵산, 또는 이를 포함하는 발현 카세트 및/또는 벡터; 및/또는 (b) 본 발명의 변형된 LbCas12a 폴리펩티드, 또는 본 발명의 변형된 LbCas12a 폴리펩티드 및 관심의 폴리펩티드 (예를 들어, 데아미나제)를 포함하는 융합 단백질을 포함하는 본 발명의 복합체를 코딩하는 핵산 구축물, 또는 이를 포함하는 발현 카세트 및/또는 벡터와 접촉시키고, 여기서 접촉은 폴리뉴클레오티드 및/또는 핵산 구축물이 발현되고, 변형된 LbCas12a 폴리펩티드 및/또는 융합 단백질이 생산되어, 가이드 핵산과의 복합체를 형성하는 조건 하에서 수행되고, 그에 의해 표적 핵산을 변형시키는 것을 포함하는, 표적 핵산을 변형시키는 방법이 제공된다. 일부 실시양태에서, 표적 핵산을 포함하는 세포 및/또는 유기체에서의 세포를 (a) (i) 본 발명의 폴리뉴클레오티드 (예를 들어, 본 발명의 변형된 LbCas12a 폴리펩티드를 코딩하거나, 또는 본 발명의 변형된 LbCas12a 폴리펩티드 및 관심의 폴리펩티드 (예를 들어, 데아미나제)를 포함하는 융합 단백질을 코딩함), 또는 이를 포함하는 발현 카세트 또는 벡터, 및 (ii) 가이드 핵산, 또는 이를 포함하는 발현 카세트 및/또는 벡터; 및/또는 (b) 본 발명의 변형된 LbCas12a 폴리펩티드, 또는 본 발명의 변형된 LbCas12a 폴리펩티드 및 관심의 폴리펩티드 (예를 들어, 데아미나제)를 포함하는 융합 단백질을 포함하는 본 발명의 복합체를 코딩하는 핵산 구축물, 또는 이를 포함하는 발현 카세트 및/또는 벡터와 접촉시키고, 여기서 접촉은 폴리뉴클레오티드 및/또는 핵산 구축물이 발현되고, 변형된 LbCas12a 폴리펩티드 및/또는 융합 단백질이 생산되어, 가이드 핵산과의 복합체를 형성하는 조건 하에서 수행되고, 그에 의해 표적 핵산을 변형시키는 것을 포함하는, 세포 및/또는 유기체의 게놈을 변형시키는/변경시키는 방법이 제공된다.

[0147] 일부 실시양태에서, 본 발명은 표적 핵산을 (a) (i) 본 발명의 융합 단백질 (본 발명의 변형된 LbCas12a 폴리펩티드 및 관심의 폴리펩티드 (예를 들어, 데아미나제)를 포함함, 및 (a) (ii) 가이드 핵산; (b) 본 발명의 융합 단백질, 및 가이드 핵산을 포함하는 복합체; (c) 본 발명의 융합 단백질 및 가이드 핵산을 포함하는 조성물; 및/또는 (d) 본 발명의 시스템과 접촉시키고, 그에 의해 표적 핵산을 편집하는 것을 포함하는, 표적 핵산을 편집하는 방법을 제공한다.

[0148] 일부 실시양태에서, 본 발명은 표적 핵산을 포함하는 세포 또는 무세포 시스템을 (a) (i) 본 발명의 융합 단백질 (예를 들어, 본 발명의 변형된 LbCas12a 폴리펩티드 및 관심의 폴리펩티드 (예를 들어 데아미나제))를 코딩하는 폴리뉴클레오티드 또는 이를 포함하는 발현 카세트 및/또는 벡터, 및 (a) (ii) 가이드 핵산, 또는 이를 포함하는 발현 카세트 및/또는 벡터; (b) 본 발명의 융합 단백질, 및 가이드 핵산을 포함하는 복합체를 코딩하는 핵산 구축물, 또는 이를 포함하는 발현 카세트 및/또는 벡터; 및/또는 (c) 본 발명의 시스템과 접촉시키고, 여기서 접촉은 폴리뉴클레오티드 및/또는 핵산 구축물이 발현되고, 변형된 CRISPR-Cas 뉴클레아제 및/또는 융합 단백질이 생산되어 가이드 핵산과의 복합체를 형성하는 조건 하에서 수행되고, 그에 의해 표적 핵산을 편집하는 것을 포함하는, 표적 핵산을 편집하는 방법을 제공한다.

[0149] 변형된 PAM 인식 특이성을 갖는 CRISPR-Cas 뉴클레아제는 인텔 (NHEJ)을 생성하는 것, 상동성 지정된 복구에서, 뉴클레아제 기능을 갖지 않는 게놈 인식 요소 (즉, Cpf1)로서, 부분적으로 기능적 뉴클레아제를 갖는 게놈 인식 요소 (니카제 Cpf1)로서, 게놈 DNA의 촉매 편집을 위한 융합 단백질 (DNA 염기 편집제)에서, RNA의 촉매 편집을 위한 융합 단백질 (RNA 염기 편집제)에서, 다른 거대분자를 특이적 게놈 영역에 표적화하기 위한 것; 소형 화학물질을 특이적 게놈 영역에 표적화하기 위한 것, 특이적 게놈 영역의 표지화를 위한 것 및/또는 CRISPR-지정된 게놈 재조합 전략을 위한 것을 포함하지만 이에 제한되지 않는 많은 방식으로 이용될 수 있다.

[0150] 상이한 핵산 구축물, 발현 벡터, 및/또는 벡터 상에서 제공되는 경우, 본 발명의 핵산 구축물은 표적 핵산을 가

이드 핵산과 접촉시키기 전에, 그와 공동으로 또는 그 후에 표적 핵산과 접촉될 수 있다.

[0151] 본 발명의 변형된 CRISPR-Cas 뉴클레아제 및 폴리펩티드 및 이를 코딩하는 핵산 구축물은 동물, 식물, 진균, 고세균, 또는 박테리아를 포함하나 이에 제한되지는 않는 임의의 유기체에서 표적 핵산을 변형시키는데 사용될 수 있다. 동물은 포유동물, 곤충, 어류, 조류 등을 포함할 수 있으나 이에 제한되지는 않는다. 본 발명이 유용할 수 있는 예시적인 포유동물은 영장류 (인간 및 비-인간 (예를 들어, 침팬지, 비비, 원숭이, 고릴라 등)), 고양이, 개, 마우스, 래트, 페렛, 거빌, 햄스터, 소, 돼지, 말, 염소, 당나귀, 또는 양을 포함하나 이에 제한되지는 않는다.

[0152] 임의의 식물 또는 식물 부분의 표적 핵산은 본 발명의 핵산 구축물을 사용하여 변형될 및/또는 편집될 (예를 들어, 돌연변이될, 예를 들어, 염기 편집될, 절단될, 니킹될 등) 수 있다. 피자식물, 나자식물, 단자엽식물, 쌍자엽식물, C3, C4, CAM 식물, 선대식물, 양치식물 및/또는 양치식물 동류, 미세조류, 및/또는 거대조류를 포함한 임의의 식물 (또는 예를 들어, 속 또는 보다 높은 목 분류로의 식물의 그룹핑)은 본 발명의 핵산 구축물을 사용하여 변형될 수 있다. 본 발명에 유용한 식물 및/또는 식물 부분은 임의의 식물 종/변종/품종의 식물 및/또는 식물 부분일 수 있다. 본원에 사용된 용어 "식물 부분"은 배, 화분, 배우, 종자, 잎, 줄기, 순, 꽃, 가지, 과실, 인, 이삭, 대, 껍질, 자루, 뿌리, 근단, 꽃밥, 식물 및/또는 식물의 부분에서 무손상인 식물 세포를 포함한 식물 세포, 식물 원형질체, 식물 조직, 식물 세포 조직 배양물, 식물 캘러스, 식물 무리 등을 포함하나 이에 제한되지는 않는다. 본원에 사용된 "순"은 잎 및 줄기를 포함한 땅 위의 부분을 지칭한다. 또한, 본원에 사용된 "식물 세포"는 세포벽을 포함하고 또한 원형질체로 지칭될 수 있는 식물의 구조적 및 생리학적 단위를 지칭한다. 식물 세포는 단리된 단일 세포의 형태일 수 있거나, 배양된 세포일 수 있거나, 또는 고등-유기체화된 단위, 예컨대, 예를 들어, 식물 조직 또는 식물 기관의 일부일 수 있다.

[0153] 본 발명에 유용한 식물의 비-제한적 예는 잔디풀 (예를 들어, 블루그래스, 벤프그래스, 독보리, 김의털), 깃털갈대풀, 쯤새풀, 억새, 물대, 지팡이풀, 채소 작물, 예컨대 아티초크, 콜라비, 루콜라, 리크, 아스파라거스, 상추 (예를 들어, 결구, 잎, 로메인), 토란, 멜론 (예를 들어, 머스크멜론, 수박, 크렌쇼, 감로, 칸탈루프), 콜작물 (예를 들어, 방울 양배추, 양배추, 콜리플라워, 브로콜리, 콜라드, 케일, 중국 양배추, 청경채), 카르도니, 당근, 나파, 오크라, 양파, 셀러리, 파슬리, 병아리콩, 파스닙, 치커리, 피망, 감자, 조롱박 (예를 들어, 매로우, 오이, 주키니, 스쿼시, 호박, 감로 멜론, 수박, 칸탈루프), 무, 건구식 양파, 루타바가, 가지, 서양우엉, 꽃상추, 샐러드, 엔다이브, 마늘, 시금치, 파, 스쿼시, 그린, 비트 (예를 들어, 사탕무 및 사료용 비트), 고구마, 근대, 서양고추냉이, 토마토, 순무, 및 향신료; 과실 작물, 예컨대 사과, 살구, 체리, 천도복숭아, 복숭아, 배, 자두, 프룬, 체리, 마르멜로, 무화과, 너트 (예를 들어, 밤, 피칸, 피스타치오, 헤이즐넛, 피스타치오, 땅콩, 호두, 마카다미아 너트, 아몬드 등), 감귤류 (예를 들어, 귤, 금귤, 오렌지, 자몽, 탄제린, 만다린, 레몬, 라임 등), 블루베리, 블랙 라즈베리, 보이즈넬리, 크랜베리, 커런트, 구스베리, 로건베리, 산딸기, 딸기, 블랙베리, 포도 (예를 들어, 와인 및 테이블), 아보카도, 바나나, 키위, 감, 석류, 파인애플, 열대 과일, 인과, 멜론, 망고, 파파야, 및 리치, 필드 작물 식물, 예컨대 클로버, 알팔파, 티모시, 달맞이꽃, 메도우 폼, 옥수수/메이즈 (예를 들어, 필드, 스위트, 팝콘), 홉, 호호바, 메밀, 홍화, 퀴노아, 밀, 벼, 보리, 호밀, 기장, 수수, 귀리, 라이밀, 수수, 담배, 케이폭, 콩과 식물 (콩 (예를 들어, 생두 및 건조된), 렌틸, 완두콩, 대두), 오일 시드 식물 (예를 들어, 유채, 카놀라, 머스타드, 피피, 올리브, 해바라기, 코코넛, 피마자유 식물, 코코아빈, 땅콩, 기름 야자, 대두, 카멜리나 등), 쯤개구리밥, 아라비도시스, 섬유 식물 (면화, 아마, 대마, 황마), 칸나비스(Cannabis) (예를 들어, 칸나비스 사티바(*Cannabis sativa*), 칸나비스 인디카(*Cannabis indica*), 및 칸나비스 루데랄리스(*Cannabis ruderalis*)), 녹나무과 (예를 들어, 계피, 장뇌), 또는 식물, 예컨대 커피, 사탕수수, 차, 및 천연 고무 식물; 및/또는 화초 식물, 예컨대 화훼 식물, 선인장, 다육식물 및/또는 장식용 식물 (예를 들어, 장미, 튜립, 제비꽃), 뿐만 아니라 나무, 예컨대 삼림수 (활엽수 및 상록수, 예컨대 침엽수; 예를 들어, 느릅나무, 물푸레나무, 떡갈나무, 단풍나무, 진나무, 가문비나무, 삼나무, 소나무, 자작나무, 사이프러스, 유칼립투스, 버드나무), 뿐만 아니라 관목 및 다른 묘목을 포함한다. 일부 실시양태에서, 본 발명의 핵산 구축물 및/또는 이를 코딩하는 발현 카세트 및/또는 벡터는 메이즈, 대두, 밀, 카놀라, 벼, 토마토, 피망, 해바라기, 산딸기, 블랙베리, 블랙 라즈베리 및/또는 체리를 변형시키는데 사용될 수 있다.

[0154] 본 발명은 본 발명의 방법을 수행하는 키트를 추가로 포함한다. 본 발명의 키트는 시약, 완충제, 및/또는 혼합, 측정, 분류, 표지화 등을 위한 장치, 뿐만 아니라 표적 핵산을 변형시키는데 적절한 것인 바와 같은 지시서 등을 포함할 수 있다.

[0155] 일부 실시양태에서, 본 발명은 본 발명의 1종 이상의 폴리뉴클레오티드 및/또는 핵산 구축물, 및/또는 이를 포함하는 발현 카세트 및/또는 벡터를 그의 사용을 위한 임의적 지시서와 함께 포함하는 키트를 제공한다. 일부

실시양태에서, 키트는 관심의 폴리펩티드 및/또는 이를 코딩하는 폴리뉴클레오티드 및 이를 포함하는 발현 카세트 및/또는 벡터를 추가로 포함할 수 있다. 일부 실시양태에서, 가이드 핵산은 본 발명의 핵산 구축물과 동일한 발현 카세트 및/또는 벡터 상에 제공될 수 있다. 일부 실시양태에서, 가이드 핵산은 본 발명의 핵산 구축물을 포함하는 것과는 별개의 발현 카세트 또는 벡터 상에 제공될 수 있다.

[0156] 따라서, 일부 실시양태에서, (a) 본원에서 제공된 바와 같은 변형된 CRISPR-Cas 뉴클레아제를 코딩하는 폴리뉴클레오티드 및 (b) (a)의 폴리뉴클레오티드의 발현을 유도하는 프로모터를 포함하는 핵산 구축물을 포함하는 키트가 제공된다. 일부 실시양태에서, 키트는 가이드 핵산을 코딩하는 핵산 구축물을 추가로 포함할 수 있고, 여기서 구축물은 가이드 핵산의 백본 내로 표적 핵산 서열과 동일하거나 그에 상보적인 핵산 서열을 클로닝하기 위한 클로닝 부위를 포함한다.

[0157] 일부 실시양태에서, 키트는 1개 이상의 핵 국제화 신호를 포함하는/코딩하는 핵산 구축물을 포함할 수 있고, 여기서 핵 국제화 신호는 CRISPR-Cas 뉴클레아제에 융합된다. 일부 실시양태에서, 본 발명의 변형된 CRISPR-Cas 뉴클레아제를 코딩하는 본 발명의 핵산 구축물, 또는, 및/또는 이를 포함하는 발현 카세트 및/또는 벡터를 포함하는 키트가 제공되고, 여기서 핵산 구축물, 발현 카세트 및/또는 벡터는 형질전환체 (예를 들어, 항생제 저항성 유전자, 제초제 저항성 유전자 등을 코딩하는 핵산)를 확인하는데 유용한 1종 이상의 선택가능한 마커를 추가로 코딩할 수 있다. 일부 실시양태에서, 핵산 구축물은 코딩된 CRISPR-Cas 뉴클레아제 내의 1개 이상의 인트론을 코딩하는 mRNA일 수 있다. 일부 실시양태에서, 키트는 본 발명의 폴리펩티드 및 핵산 구축물의 발현에 사용하기 위한 프로모터 및 인트론을 갖는 프로모터를 포함할 수 있다.

[0158] **CRISPR-Cas 뉴클레아제 및 관련된 조성물의 PAM 특이성을 변형시키는 방법**

[0159] CRISPR-Cas 시스템은 2가지 주요한 기준을 사용하여 표적 핵산을 향해 지정된다: 표적화된 DNA 서열과의 가이드 RNA의 상동성, 및 특정 서열의 프로토스페이서 인접 모티프 (PAM)의 존재. 상이한 CRISPR-Cas 뉴클레아제는 상이한 PAM 서열 요건, 예컨대 SpCas9에 대해 NGG 또는 LbCas12a (Cpf1)에 대해 TTTV (여기서 V는 임의의 비-티미딘 뉴클레오티드임)를 갖는다. 새로운 CRISPR 뉴클레아제 또는 그의 돌연변이체를 그들의 PAM 요건에 대해 스크리닝하는 것은 너무 많은 반복이 가능하기 때문에 복잡하고 예측불가능할 수 있다. 시험관내 검정, 특히 PAM 결정 검정 (PAMDA)은 임의의 특정 CRISPR 뉴클레아제 또는 그의 돌연변이체에 대한 PAM 특이성을 스크리닝하는데 사용될 수 있다. 이들 검정은 한정된/공지된 프로토스페이서 서열에 인접한 DNA의 무작위적 부분에 의존한다. 가이드 RNA는 공지된 프로토스페이서 서열을 표적화하도록 디자인될 수 있으며, 무작위적 PAM 영역이 적절한 DNA 서열 (예를 들어, CRISPR 뉴클레아제 또는 그의 돌연변이체에 의해 인식됨)을 함유하는 경우, CRISPR 뉴클레아제는 표적에 결합하고 이를 커팅할 수 있다.

[0160] CRISPR-Cas 뉴클레아제에 대한 PAM 인식 부위는 PAM 부위-고갈 검정 (예를 들어, PAM 고갈 검정) 또는 PAM 결정 검정 (PAMDA)을 사용하여 평가될 수 있다 (Kleinstiver et al. *Nat Biotechnol* **37**: 276-282 (2019)). PAM 고갈 검정을 위해, 프로토스페이서에 인접한 무작위적 뉴클레오티드 (염기 쌍)를 함유하는 플라스미드의 라이브러리는 박테리아 (예를 들어, *이. 콜라이*)에서 CRISPR 뉴클레아제에 의해 절단에 대해 시험된다. 플라스미드는 예를 들어, 무작위적 PAM 서열에 인접한 항생제 저항성을 부여하는 폴리뉴클레오티드를 포함할 수 있다. CRISPR-Cas 뉴클레아제에의 노출시 커팅되지 않는 그러한 서열은 세포가 항생제 저항성 유전자의 존재로 인해 항생제의 존재 하에서 생존하는 것을 가능하게 하는 반면, 표적화가능한 PAM을 함유하는 플라스미드는 절단되고, 세포 사멸로 인해 라이브러리로부터 고갈된다. 플라스미드의 생존하는 (비절단된) 집단의 시퀀싱은 선택 후 PAM 고갈 값의 계산을 가능하게 하며, 이는 CRISPR-Cas 뉴클레아제에 노출되지 않은 라이브러리와 비교된다. 실험 라이브러리에서 서열의 풀로부터 고갈된 그러한 서열은 CRISPR-Cas 뉴클레아제에 의해 인식되는 PAM 서열(들)을 함유한다.

[0161] PAM 서열을 확인하는데 사용될 수 있는 또 다른 방법은 PAM 결정 검정 (PAMDA)이다 (Kleinstiver et al. *Nat Biotechnol* **37**: 276-282 (2019)). 이 경우, 절단은 살아 있는 세포의 외부에서 수행된다. PAMDA에서, 단일 DNA 가닥은 한정된 프로토스페이서 서열 옆의 뉴클레오티드의 무작위적 부분으로 합성된다. 올리고뉴클레오티드는 합성된 DNA 가닥의 3' 단부에 어닐링되고, 엑소뉴클레아제 마이너스 (-엑소) 클레노우(Klenow) 단편을 사용하여 확장되어, 한정된 및 무작위 서열에 걸쳐 중합한다. 이는 두가닥 라이브러리를 생성하고, 이는 이어서 제한 엔도뉴클레아제로 커팅되고, 총 DNA를 증폭시키기 위해 박테리아 내로 클로닝된다. 플라스미드는 추출되고, 또 다른 제한 엔도뉴클레아제로 선형화되어 선형 주형을 제조한다. 주형은 CRISPR-Cas 뉴클레아제-가이드 RNA 복합체와 접촉된다. 단지 CRISPR-Cas 뉴클레아제에 의해 인식되는 PAM을 함유하는 그러한 서열이 절단될 것이다. 이어서 실험 라이브러리 및 대조군 라이브러리 (CRISPR-Cas 뉴클레아제에 노출되지 않음) 둘 다는 PCR

을 통해 증폭된다. 단지 CRISPR-Cas 뉴클레아제에 의해 절단되지 않은 서열이 증폭될 것이다. 대조군 라이브러리 및 실험 라이브러리 (CRISPR-Cas 뉴클레아제로부터 처리됨)로부터의 PCR 증폭된 서열은 시퀀싱되고 비교된다. 대조군 (CRISPR-Cas 뉴클레아제에 노출되지 않음) 라이브러리에 존재하지만 실험 라이브러리에는 존재하지 않는 PAM 서열은 CRISPR-Cas 뉴클레아제에 의해 인식되는 PAM 서열이다 (그에 의해, 프로토스페이서가 절단되는 것을 허용함).

[0162] 시험관내에서 뉴클레아제 요건의 평가를 위해, 무작위적 PAM 라이브러리가 제조된다. 이 방법에 대해 기재된 단계는 평가되는 모든 PAM 서열을 함유하는 비편향된 무작위적 DNA 라이브러리의 제조, 플라스미드 내로의 클로닝, 출발 DNA의 총량을 증가시키는 박테리아 내로의 라이브러리의 도입, 플라스미드의 추출, 슈퍼코일링을 제거하는 제한 효소로의 플라스미드의 선형화, 선형화된 분자를 CRISPR-Cas 뉴클레아제에 노출시키는 것, 단편을 증폭시키는 것 (예를 들어, PCR), 및 마지막으로 시퀀싱 분석 (예를 들어, 차세대 시퀀싱, NGS)을 수반한다. 비편향된 라이브러리를 생성하는 처음 단계 및 제한 소화는 적어도 2종의 제한 효소, 클레노우 연장, 및 벡터 내로 라이게이션하기 전에 생성물의 정화를 요구한다. 2 내지 3종의 제한 효소를 사용하는 것은 전형적으로 일부 PAM 서열을 라이브러리로부터 제거하며, 이는 라이브러리 내로 편향을 도입한다. 또한, 후속 클레노우 연장 및 정화 단계는 또한 PAM 서열을 제거하고, 그에 의해 라이브러리 내로 추가의 편향을 도입할 수 있다. PAM 서열의 소실을 회피하고 보다 완전하고 비편향된 라이브러리를 생성하기 위해, 본 발명은 제한 엔도뉴클레아제 및 클레노우 연장 대신 오버행을 갖는 중첩하는 고체-상태 합성된 올리고뉴클레오티드 (예를 들어, 어닐링된 올리고뉴클레오티드)를 사용하여 무작위적 PAM 라이브러리를 생성하는 새로운 방법을 제공한다 (예를 들어, 도 1b 참조). 이어서 본 발명의 방법을 사용하여 생산된 무작위적 PAM 라이브러리는 종래 기술 방법을 통해 생산된 라이브러리로 이전에 이용가능한 것보다 더 많은 정확성을 갖고 CRISPR-Cas 뉴클레아제의 PAM 특이성을 시험하는데 사용될 수 있다.

[0163] 따라서, 일부 실시양태에서, 본 발명은 하기 단계를 포함하는 2개 이상의 이중 가닥 핵산 분자를 제조하는 것, 그에 의해 이중 가닥 핵산 분자를 포함하는 무작위적 DNA 라이브러리를 구축하는 것을 포함하는, 프로토스페이서의 5' 단부에 프로토스페이서 인접 모티프 (PAM) 인식 부위를 갖는 CRISPR-Cas 뉴클레아제의 PAM 요건/특이성을 결정하기 위한 이중 가닥 핵산 분자를 포함하는 무작위적 DNA 라이브러리를 구축하는 방법을 제공한다: (a) 비-표적 올리고뉴클레오티드 (제1) 가닥 및 표적 올리고뉴클레오티드 (제2) 가닥을 2개 이상의 이중 가닥 핵산 분자의 각각에 대해 합성하는 단계이며, 여기서 비-표적 올리고뉴클레오티드 가닥은 5'에서 3'로 (i) 약 5 내지 약 15개의 뉴클레오티드 (예를 들어, 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 또는 15개의 뉴클레오티드, 및 그 안의 임의의 범위)를 갖는 제1 서열, (ii) 적어도 4개의 무작위적 뉴클레오티드 (예를 들어, 적어도 4, 5, 6, 7, 8, 9, 10개, 또는 그 초과, 및 그 안의 임의의 범위)를 갖는 제2 서열, (iii) 약 16 내지 약 25개의 뉴클레오티드 (예를 들어, 약 16, 17, 18, 19, 20, 21, 22, 23, 24, 또는 25개의 뉴클레오티드, 및 그 안의 임의의 범위)를 포함하는 프로토스페이서 서열, 및 (iv) 약 5 내지 약 20개의 뉴클레오티드 (예를 들어, 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개의 뉴클레오티드, 및 그 안의 임의의 범위)를 갖는 제3 서열을 포함하고, (i)의 약 5 내지 15개의 뉴클레오티드를 갖는 제1 서열은 (ii)의 제2 서열의 5' 단부에 바로 인접하고, (ii)의 제2 서열은 (iii)의 프로토스페이서 서열의 5' 단부에 바로 인접하고, 프로토스페이서 서열은 (iv)의 제3 서열의 5' 단부에 바로 인접하고; 표적 올리고뉴클레오티드 (제2) 가닥은 비-표적 올리고뉴클레오티드 가닥에 상보적인 것인 단계; 및 (b) 비-표적 올리고뉴클레오티드 가닥을 상보적 표적 올리고뉴클레오티드 가닥에 어닐링하여 이중 가닥 핵산 분자를 생산하는 단계이며, 여기서 제1 서열은 제한 부위를 (그의 5' 단부에) 포함하고, 제3 서열은 제한 부위를 (그의 3' 단부에) 포함하고, 2개 이상의 이중 가닥 핵산 분자의 각각의 제1 서열 (i), 프로토스페이서 서열 (iii) 및 제3 서열 (iv)은 동일한 것인 단계. 일부 실시양태에서, 표적 가닥 및/또는 비-표적 가닥은 5' 인산화될 수 있다.

[0164] 일부 실시양태에서, 본 발명은 하기 단계를 포함하는 2개 이상의 이중 가닥 핵산 분자를 제조하는 것, 그에 의해 이중 가닥 핵산 분자를 포함하는 무작위적 DNA 라이브러리를 구축하는 것을 포함하는, 프로토스페이서의 3' 단부에 프로토스페이서 인접 모티프 (PAM) 인식 부위를 갖는 CRISPR-Cas 뉴클레아제의 PAM 요건/특이성을 결정하기 위한 이중 가닥 핵산 분자를 포함하는 무작위적 DNA 라이브러리를 구축하는 방법을 제공한다: (a) 비-표적 올리고뉴클레오티드 (제1) 가닥 및 표적 올리고뉴클레오티드 (제2) 가닥을 2개 이상의 이중 가닥 핵산 분자의 각각에 대해 합성하는 단계이며, 여기서 비-표적 올리고뉴클레오티드 가닥은 5'에서 3'로 (i) 약 5 내지 약 20개의 뉴클레오티드 (예를 들어, 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개의 뉴클레오티드, 및 그 안의 임의의 범위)를 갖는 제1 서열, (ii) 약 16 내지 약 25개의 뉴클레오티드 (예를 들어, 약 16, 17, 18, 19, 20, 21, 22, 23, 24, 또는 25개의 뉴클레오티드, 및 그 안의 임의의 범위)를 포함하는 프로토스페이서 서열, (iii) 적어도 4개의 무작위적 뉴클레오티드 (예를 들어, 적어도 4, 5, 6, 7, 8, 9, 10개, 또는 그

초과, 및 그 안의 임의의 범위)를 갖는 제2 서열, 및 (iv) 약 5 내지 약 15개의 뉴클레오타이드 (예를 들어, 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15개의 뉴클레오타이드, 및 그 안의 임의의 범위)를 갖는 제3 서열을 포함하고, (i)의 약 5 내지 20개의 뉴클레오타이드를 갖는 제1 서열은 (ii)의 프로토스페이서 서열의 5' 단부에 바로 인접하고, (iii)의 제2 서열은 (iii)의 프로토스페이서 서열의 3' 단부에 바로 인접하고, (iv)의 제3 서열은 (iii)의 제2 서열의 3' 단부에 바로 인접하고; 표적 올리고뉴클레오타이드 (제2) 가닥은 비-표적 올리고뉴클레오타이드 가닥에 상보적인 것인 단계; 및 (b) 비-표적 올리고뉴클레오타이드 가닥을 상보적 표적 올리고뉴클레오타이드 가닥에 어닐링하여 이중 가닥 핵산 분자를 생산하는 단계이며, 여기서 제1 서열 (i)은 제한 부위를 (그의 5' 단부에) 포함하고, 제3 서열 (iv)은 제한 부위를 (그의 3' 단부에) 포함하고, 2개 이상의 이중 가닥 핵산 분자의 각각의 제1 서열 (i), 프로토스페이서 서열 (ii) 및 제3 서열 (iv)은 동일한 것인 단계. 일부 실시양태에서, 표적 가닥 및/또는 비-표적 가닥은 5' 인산화될 수 있다.

[0165] 일부 실시양태에서, 이중 가닥 핵산 분자는 벡터 내로 라이게이션되어 무작위적 DNA 라이브러리를 포함하는 벡터를 생산할 수 있다. 일부 실시양태에서, 벡터는 고 카피 수 벡터일 수 있다. 일부 실시양태에서, 무작위적 DNA 라이브러리는 예를 들어, 무작위적 DNA 라이브러리를 포함하는 벡터를 1개 이상의 박테리아 세포 내로 도입하고, 1개 이상의 박테리아 세포를 배양함으로써 증폭될 수 있다. 일부 실시양태에서, 무작위적 DNA 라이브러리를 포함하는 벡터는 배양 후에 1개 이상의 박테리아 세포로부터 분리될 수 있다. 이어서 분리된 벡터는 예를 들어, CRISPR-Cas 뉴클레아제의 PAM 인식 특이성의 분석에 사용하기 위해 선형화될 수 있다 (예를 들어, 벡터를 1종 이상의 제한 효소; 예를 들어, ScaI 또는 PfoI과 접촉시킴으로써). 일부 실시양태에서, PfoI은 분리된 벡터를 선형화하는데 사용될 수 있다.

[0166] 일부 실시양태에서, 프로토스페이서의 5' 단부 상에 프로토스페이서 인접 모티프 (PAM) 인식 부위를 갖는 CRISPR-Cas 뉴클레아제의 PAM 요건/특이성을 결정하기 위한 무작위적 DNA 라이브러리가 제공되며, 여기서 무작위적 DNA 라이브러리는 2개 이상의 이중 가닥 핵산 분자를 포함하고, 그의 각각은 (a) 비-표적 올리고뉴클레오타이드 (제1) 가닥 및 표적 올리고뉴클레오타이드 (제2) 가닥 (여기서 비-표적 올리고뉴클레오타이드 가닥은 5'에서 3'로 (i) 약 5 내지 약 15개의 뉴클레오타이드 (예를 들어, 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 또는 15개의 뉴클레오타이드, 및 그 안의 임의의 범위)를 갖는 제1 서열, (ii) 적어도 4개의 무작위적 뉴클레오타이드 (예를 들어, 적어도 4, 5, 6, 7, 8, 9, 10개, 또는 그 초과, 및 그 안의 임의의 범위)를 갖는 제2 서열, (iii) 약 16 내지 약 25개의 뉴클레오타이드 (예를 들어, 약 16, 17, 18, 19, 20, 21, 22, 23, 24, 또는 25개의 뉴클레오타이드, 및 그 안의 임의의 범위)를 포함하는 프로토스페이서 서열, 및 (iv) 약 5 내지 약 20개의 뉴클레오타이드 (예를 들어, 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개의 뉴클레오타이드, 및 그 안의 임의의 범위)를 갖는 제3 서열을 포함하고, (i)의 약 5 내지 15개의 뉴클레오타이드를 갖는 제1 서열은 (i)의 제2 서열의 5' 단부에 바로 인접하고, (ii)의 제2 서열은 (iii)의 프로토스페이서 서열의 5' 단부에 바로 인접하고, 프로토스페이서 서열은 (iv)의 제3 서열의 5' 단부에 바로 인접하고; 표적 올리고뉴클레오타이드 (제2) 가닥은 비-표적 올리고뉴클레오타이드 가닥에 상보적임); 및 (b) 상보적 표적 올리고뉴클레오타이드 가닥에 어닐링되어 이중 가닥 핵산을 생산하는 비-표적 올리고뉴클레오타이드 가닥 (여기서 제1 서열은 제한 부위를 (그의 5' 단부에) 포함하고, 제3 서열은 제한 부위를 (그의 3' 단부에) 포함하고, 2개 이상의 이중 가닥 핵산 분자의 각각의 제1 서열 (i), 프로토스페이서 서열 (iii) 및 제3 서열 (iv)은 동일함)을 포함한다. 일부 실시양태에서, 표적 가닥 및/또는 비-표적 가닥은 5' 인산화될 수 있다.

[0167] 일부 실시양태에서, 프로토스페이서의 3' 단부 상에 프로토스페이서 인접 모티프 (PAM) 인식 부위를 갖는 CRISPR-Cas 뉴클레아제의 PAM 요건/특이성을 결정하기 위한 무작위적 DNA 라이브러리가 제공될 수 있으며, 여기서 무작위적 DNA 라이브러리는 2개 이상의 이중 가닥 핵산 분자를 포함하고, 그의 각각은 (a) 비-표적 올리고뉴클레오타이드 (제1) 가닥 및 표적 올리고뉴클레오타이드 (제2) 가닥 (여기서 비-표적 올리고뉴클레오타이드 가닥은 5'에서 3'로 (i) 약 5 내지 약 20개의 뉴클레오타이드 (예를 들어, 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개의 뉴클레오타이드, 및 그 안의 임의의 범위)를 갖는 제1 서열, (ii) 약 16 내지 약 25개의 뉴클레오타이드 (예를 들어, 약 16, 17, 18, 19, 20, 21, 22, 23, 24, 또는 25개의 뉴클레오타이드, 및 그 안의 임의의 범위)를 포함하는 프로토스페이서 서열, (iii) 적어도 4개의 무작위적 뉴클레오타이드 (예를 들어, 적어도 4, 5, 6, 7, 8, 9, 10개, 또는 그 초과, 및 그 안의 임의의 범위)를 갖는 제2 서열, 및 (iv) 약 5 내지 약 15개의 뉴클레오타이드 (예를 들어, 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15개의 뉴클레오타이드, 및 그 안의 임의의 범위)를 갖는 제3 서열을 포함하고, (i)의 약 5 내지 20개의 뉴클레오타이드를 갖는 제1 서열은 (ii)의 프로토스페이서 서열의 5' 단부에 바로 인접하고, (iii)의 제2 서열은 (iii)의 프로토스페이서 서열의 3' 단부에 바로 인접하고, (iv)의 제3 서열은 (iii)의 제2 서열의 3' 단부에 바로 인접하고; 표적 올리고뉴클레오타이드 (제2) 가닥은 비-표적 올리고뉴클레오타이드 가닥에 상보적임); 및 (b) 상보적 표적 올리고뉴클레오타이드 가닥에 어닐링

되어 이중 가닥 핵산 분자를 생산하는 비-표적 올리고뉴클레오타이드 가닥 (여기서 제1 서열은 제한 부위를 (그의 5' 단부에) 포함하고, 제3 서열은 제한 부위를 (그의 3' 단부에) 포함하고, 2개 이상의 이중 가닥 핵산 분자의 각각의 제1 서열 (i), 프로토스페이서 서열 (ii) 및 제3 서열 (iv)은 동일함)을 포함한다. 일부 실시양태에서, 표적 가닥 및/또는 비-표적 가닥은 5' 인산화될 수 있다.

[0168] 일부 실시양태에서, 본 발명은 CRISPR-Cas 뉴클레아제를 본 발명의 무작위적 DNA 라이브러리와 접촉시키는 단계; CRISPR-Cas 뉴클레아제와의 접촉 전에 (예를 들어, 대조군) 및 후에 무작위적 DNA 라이브러리의 이중 가닥 핵산 분자를 시퀀싱하는 단계이며, 여기서 CRISPR-Cas 뉴클레아제와의 접촉 전에 무작위적 DNA 라이브러리에 존재하지만, CRISPR-Cas 뉴클레아제와의 접촉 후에 무작위적 DNA 라이브러리에 부재하는 이중 가닥 핵산 분자는 CRISPR-Cas 뉴클레아제의 프로토스페이서 인접 모티프 (PAM) 인식 서열을 확인하고, 그에 의해 CRISPR-Cas 뉴클레아제의 PAM 특이성을 결정하는 것인 단계를 포함하는, CRISPR-Cas 뉴클레아제의 PAM 특이성을 결정하는 방법을 제공한다.

[0169] 일부 실시양태에서, CRISPR-Cas 뉴클레아제의 프로토스페이서 인접 모티프 (PAM) 특이성을 결정하는 방법은 CRISPR-Cas 뉴클레아제를 본 발명의 무작위적 DNA 라이브러리와 접촉시키고; CRISPR-Cas 뉴클레아제와의 접촉 전에 (예를 들어, 대조군) 및 후에 무작위적 DNA 라이브러리의 이중 가닥 핵산 분자를 시퀀싱하고, 뉴클레아제의 PAM 인식 서열을 확인하는 것을 포함하고, 여기서 확인은 CRISPR-Cas 뉴클레아제와의 접촉 전에 라이브러리에 존재하는 이중 가닥 핵산 분자를 CRISPR-Cas 뉴클레아제와의 접촉 후에 라이브러리에 존재하는 이중 가닥 핵산 분자와 비교하는 것을 포함하고, CRISPR-Cas 뉴클레아제와의 접촉 전에 무작위적 DNA 라이브러리에 존재하지만, CRISPR-Cas 뉴클레아제와의 접촉 후에 무작위적 DNA 라이브러리로부터 부재하는 이중 가닥 핵산 분자는 CRISPR-Cas 뉴클레아제의 PAM 특이성을 확인한다.

[0170] 접촉 전의 무작위적 라이브러리의 시퀀싱의 결과는 접촉 후의 시퀀싱의 결과에 대한 대조군으로서 역할을 할 수 있다. 일부 실시양태에서, CRISPR-Cas 뉴클레아제의 PAM 특이성을 결정하는 것은 핵산 시퀀싱을 수행하는 것을 포함할 수 있다. 일부 실시양태에서, 서열은 차세대 시퀀싱 (NGS)을 포함할 수 있다.

[0171] 임의의 CRISPR-Cas 뉴클레아제는 PAM 인식 특이성을 변형시키기 위한 본 발명의 방법에 사용될 수 있다. 따라서, 야생형과 비교하여 상이한 PAM 특이성을 갖도록 변형될 수 있는 CRISPR-Cas 뉴클레아제는 Cas9, C2c1, C2c3, Cas12a (또한 Cpf1로 지칭됨), Cas12b, Cas12c, Cas12d, Cas12e, Cas13a, Cas13b, Cas13c, Cas13d, Cas1, Cas1B, Cas2, Cas3, Cas3', Cas3", Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (또한 Csn1 및 Csx12로 공지됨), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Cse1, Cse2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 (dinG), 및/또는 Csf5 폴리펩티드 또는 도메인을 포함할 수 있으나 이에 제한되지는 않는다.

[0172] Cas12a는 원래 *프레보텔라(Prevotella)* 종 및 *프란시셀라(Francisella)* 종에서 확인된 유형 V 클러스터링된 규칙적 간격의 짧은 회문식 반복부 (CRISPR)-Cas 뉴클레아제이다. Cas12a (이전에 Cpf1로 지칭됨)는 보다 널리 공지된 유형 II CRISPR Cas9 뉴클레아제와는 몇몇 측면에서 상이하다. 예를 들어, Cas9는 그의 가이드 RNA (gRNA, sgRNA) 결합 부위 (프로토스페이서, 표적 핵산, 표적 DNA) (3'-NGG)에 대해 3'인 G-풍부 프로토스페이서 인접 모티프 (PAM)를 인식하는 반면, Cas12a는 결합 부위 (프로토스페이서, 표적 핵산, 표적 DNA)에 대해 5'에 위치한 T-풍부 PAM (5'-TTN, 5'-TTTN)을 인식한다. 사실, Cas9 및 Cas12a가 그들의 가이드 RNA에 결합하는 배향은 그들의 N 및 C 말단에 관하여 아주 거의 반대이다. 더욱이, Cas12a 효소는 천연 Cas9 시스템에서 발견되는 이중 가이드 RNA (sgRNA (예를 들어, crRNA 및 tracrRNA))라기 보다는 단일 가이드 RNA (gRNA, CRISPR 어레이, crRNA)를 사용하고, Cas12a는 그 자신의 gRNA를 프로세싱한다. 추가로, Cas12a 뉴클레아제 활성화는 Cas9 뉴클레아제 활성화에 의해 생성된 블런트 단부 대신 엇갈린 DNA 이중 가닥 파단을 생성하고, Cas12a는 둘 다의 DNA 가닥을 절단하기 위해 단일 RuvC 도메인에 의존하는 반면, Cas9는 절단을 위해 HNH 도메인 및 RuvC 도메인을 이용한다.

[0173] 본 발명에 유용한 CRISPR Cas12a 폴리펩티드 또는 CRISPR Cas12a 도메인은 임의의 공지된 또는 나중에 확인될 Cas12a 뉴클레아제일 수 있다 (예를 들어, Cpf1 (Cas12a) 서열의 그의 개시내용에 대해 참조로 포함되는 미국 특허 번호 9,790,490 참조). 용어 "Cas12a", "Cas12a 폴리펩티드" 또는 "Cas12a 도메인"은 Cas12a의 가이드 핵산 결합 도메인 및/또는 Cas12a의 활성화, 불활성, 또는 부분적으로 활성화 DNA 절단 도메인을 포함하는 Cas12a 폴리펩티드, 또는 그의 단편을 포함하는 RNA-가이드된 뉴클레아제를 지칭한다. 일부 실시양태에서, 본 발명에 유용한 Cas12a는 뉴클레아제 활성화 부위 (예를 들어, Cas12a 도메인의 RuvC 부위)에 돌연변이를 포함할 수 있다.

그의 뉴클레아제 활성 부위에 돌연변이를 갖고, 따라서 더 이상 뉴클레아제 활성을 포함하지 않는 Cas12a 도메인 또는 Cas12a 폴리펩티드는 통상적으로 죽은Cas12a (예를 들어, dCas12a)로 지칭된다. 일부 실시양태에서, 그의 뉴클레아제 활성 부위에 돌연변이를 갖는 Cas12a 도메인 또는 Cas12a 폴리펩티드는 동일한 돌연변이 (예를 들어, 니카제 활성)를 갖지 않는 동일한 Cas12a 폴리펩티드와 비교하여 손상된/감소된 활성을 가질 수 있다.

[0174] 일부 실시양태에서, Cas12a 도메인은 **서열식별번호:1 내지 17** (예를 들어, **서열식별번호:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16** 및/또는 **17**) 중 임의의 하나의 아미노산 서열 또는 이를 코딩하는 폴리뉴클레오티드를 포함할 수 있으나 이에 제한되지는 않는다. 일부 실시양태에서, 본 발명의 융합 단백질은 *라크노스 피라세아에* 박테리아 ND2006 Cas12a (LbCas12a) (예를 들어, **서열식별번호:1**)로부터의 Cas12a 도메인을 포함할 수 있다.

[0175] 본 발명에 유용한 CRISPR Cas9 폴리펩티드 또는 CRISPR Cas9 도메인은 임의의 공지된 또는 나중에 확인될 Cas9 뉴클레아제일 수 있다. 일부 실시양태에서, 본 발명에 유용한 Cas9 폴리펩티드는 임의의 공지된 Cas9의 아미노산 서열과 적어도 70% 동일성 (예를 들어, 약 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 등)을 포함한다. CRISPR-Cas9 시스템은 관련 기술분야에 널리 공지되어 있으며, *레지오넬라 뉴모필라*(*Legionella pneumophila*) 균주 파리, *스트렙토코쿠스 써모필루스*(*Streptococcus thermophilus*) CNRZ1066, *스트렙토코쿠스 피오게네스*(*Streptococcus pyogenes*) MI, 또는 *네이세리아 락타미카*(*Neisseria lactamica*) 020-06 등으로부터의 Cas9 폴리펩티드를 포함하나 이에 제한되지는 않는다.

[0176] 신규한 PAM 인식 서열을 확인하기 위한 본 발명에 유용할 수 있는 다른 뉴클레아제는 C2c1, C2c3, Cas12b, Cas12c, Cas12d, Cas12e, Cas13a, Cas13b, Cas13c, Cas13d, Cas1, Cas1B, Cas2, Cas3, Cas3', Cas3", Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (또한 Csn1 및 Csx12로 공지됨), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Cse1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 (dinG), 및/또는 Csf5를 포함하나 이에 제한되지는 않는다.

[0177] 본 발명은 이제 하기 실시예를 참고로 기재될 것이다. 이들 실시예는 본 발명에 대한 청구항의 범위를 제한하는 것으로 의도되지 않으며, 오히려 특정 실시양태를 예시하는 것으로 의도됨이 인정되어야 한다. 통상의 기술자에게 일어나는 예시화된 방법의 임의의 변동은 본 발명의 범위 내에 해당하는 것으로 의도된다.

[0178] **실시예**

[0179] **실시예 1.**

[0180] **무작위적 라이브러리**

[0181] 시험관내 절단 검정 (PAM 결정 검정 (PAMDA))을 위한 라이브러리의 효율적이고 비용-효과적인 생성을 위한 본 발명의 방법의 실시예가 제공된다. 2개의 라이브러리를 프로토스페이스 1 및 2에 대해 생성하였다 (표 1 참조). 5'-단부에 무작위적 5 뉴클레오티드 서열을 갖는 올리고뉴클레오티드를 합성하고, 각각의 프로토스페이스 서열이 등물 비를 점유하도록 확인하였다 (인테그레이티드 DNA 테크놀로지스(Integrated DNA Technologies)) (표 1). 프로토스페이스 1 (PM0518, PM0519)에 대한 및 프로토스페이스 2 (PM0520, PM0521)에 대한 올리고뉴클레오티드를 혼합물을 95℃에서 5분 동안 열 사이클러에 정치하고, 0.1℃/초 동안 25℃/실온으로 냉각시킴으로써 어닐링하였다.

표 1

프로토스페이스 서 명칭	올리고 번호	서열
PAMDA 라이브러리 1	프로토스페이스 1	GGAATCCCTTCTGCAGCACCTGG (SEQ ID NO:30)
PAMDA 라이브러리 2	프로토스페이스 2	CTGATGGTCCATGTCTGTTACTC (SEQ ID NO:31)
프로토스페이스 서 1 상부 가닥	PM0518	/5Phos/CGATGTNNNNNGAATCCCTTCTGCAGCACCTGGGCG CAGGTCACGAGG (SEQ ID NO:32)
프로토스페이스 서 1 하부 가닥	PM0519	/5Phos/AATTCCTCGTGACCTGCGCCAGGTGCTGCAGAAGGG ATTCCNNNNNACATCGCATG (SEQ ID NO:33)
프로토스페이스 서 2 상부 가닥	PM0520	/5Phos/CGATGTNNNNNCTGATGGTCCATGTCTGTTACTCGCG CAGGTCACGAGG (SEQ ID NO:34)
프로토스페이스 서 2 하부 가닥	PM0521	/5Phos/AATTCCTCGTGACCTGCGCGAGTAACAGACATGGACC ATCAGNNNNNACATCGCATG (SEQ ID NO:35)

어닐링된 이중 가닥 단편을 SphI 및 EcoRI 소화된 pUC19 벡터에 직접적으로 라이게이션하였다. 라이게이션된 프로토스페이스서 구축물을 사용하여 XL1-블루 전기 적격 *E. coli* 콜라이 세포 (애질런트(Agilent))를 형질전환시키고, 1 ml SOC 배지에서 37℃에서 1시간 동안 회수하였다. 카르베니실린 플레이트를 사용하여 *E. coli* 콜라이 세포에서의 라이게이션된 생성물의 존재에 대해 점검하였다. 형질전환된 *E. coli* 콜라이 세포를 카르베니실린 (50 mg/mL)으로 보충된 LB 브로쓰 (200 ml)에서 16시간 동안 성장시켰다. 프로토스페이스서 구축물을 포함하는 플라스미드를 자이모(Zymo) 미디프랩 키트를 사용하여 정제하였다. 플라스미드/벡터를 덤 시퀀싱 분석으로 처리하여 일루미나 마이섹을 사용하여 각각의 PAM 위치에서의 A/T/G/C의 빈도를 계산하였다.

이 방법은 선택의 임의의 프로토스페이스서 올리고뉴클레오타이드(들)를 사용하여 PAM 결정을 위한 라이브러리를 생성하는데 사용될 수 있고, 여기서 어닐링된 올리고뉴클레오타이드는 라이브러리에서 PAM 서열의 전체 상보체를 보유하도록 선택된 임의의 적절한 제한 부위를 포함할 수 있다.

실시예 2

라크노스피라세아에 박테리아 ND2006 Cpf1 (LbCpf1)은 고도로 특이적인 프로토스페이스서 인접 모티프 (PAM)를 요구한다. 'TTTV' 서열은 무작위 뉴클레오타이드에 비교하여 85개의 염기 중 단지 약 1개 발생한다. 이는 무작위 DNA에서 16개의 염기 중 약 1개 발생하는 SpCas9에 대한 NGG, 무작위 DNA에서 16개의 염기 중 약 1개 발생하는 AaC2c1에 대한 TTN, 및 그의 NG PAM 요건이 4개의 염기 중 약 1개 발생하는 xCas9/Cas9-NG의 상대 난잡성과는 대조적이다. Cpf1 PAM은 Cas9 PAM보다 옥수수 및 대두 유전자에서 훨씬 덜 풍부하다 (도 2). 추가로, 아데닌 및 시토신 (염기 편집제에 대한 현재의 표적)은 그의 엄격한 PAM 요건에 기반하여 LbCpf1에 훨씬 덜 접근가능하다 (도 3).

CRISPR-Cas 뉴클레아제에 대한 도 3에 나타내어진 바와 같은 이러한 엄격성은 잠재적 표적 및 새로운 소질의 생성을 크게 감소시킨다. 본 발명은 접근가능한 PAM 서열의 개선된 비를 갖는 CRISPR-Cas 뉴클레아제, 특히, LbCpf1 (Cas12a) 뉴클레아제 (예를 들어, 약 1:4 또는 그보다 양호한 비로 발생하는 PAM 인식 부위를 갖는 뉴클레아제)의 생성에 관한 것이다. 이러한 조작된 Cas12a PAM 돌연변이체는 (NHEJ 또는 HDR 적용을 위한) 뉴클레아제로서 사용될 수 있거나, 또는 불활성화된 버전은 게놈 편집 툴에서 게놈 인식 요소로서 사용될 수 있다.

PAMDA 검증

PAM 결정 검증 (PAMDA)은 미지의 PAM 인식을 갖는 CRISPR 효소에 대한 PAM 요건을 시험하는데 유용하다. 이들은 단지 성공적인 PAM 결합 후에 표적 서열을 절단하는 CRISPR-Cas 뉴클레아제의 능력을 이용하는 시험관내 검증이다. 간략하게, 무작위적 PAM 서열을 갖는 DNA 기체의 라이브러리를 CRISPR 뉴클레아제와 함께 인큐베이션하고, 이어서 DNA를 PCR을 통해 증폭시킨다. 단지 무손상 단편 (예를 들어, 뉴클레아제에 의해 인식되지 않는 것들)이 증폭된다. 절단된 단편 (뉴클레아제에 의해 인식되는 것들)은 증폭되지 않는다. 뉴클레아제에 노출된 라이브러리 및 대조군 라이브러리 (뉴클레아제에 노출되지 않음) 둘 다로부터의 DNA를 시퀀싱한다. 시퀀싱 결과의 2개의 세트를 비교하여 어느 서열이 절단되었고, 따라서 뉴클레아제에의 노출 후 시퀀싱 어셈블리에 존재하지 않는지를 결정한다 (예로서, 도 4 참조). 다중 시점을 사용하는 변형된 PAMDA를 사용하여 PAM 결합 및 후속 절단을 결정하였다.

[0191] *LbCpf1* 돌연변이유발

[0192] 백팔십육 (186)개의 점 돌연변이 (표 2)를 디자인하고, 본원에 기재된 바와 같은 PAMDA 검정에서 개별적으로 시험하였다. 성공적인 조작은 PAM 인식 서열을 변화시켜 신규한 PAM 인식 *LbCpf1*을 생성할 수 있거나, 또는 PAM 엄격성을 완화시켜 보다 난잡한 *LbCpf1*을 발생시킬 수 있다.

[0193] 표 2. 서열식별번호:1에서의 치환을 위한 잔기

K116R	K120R	K121S	D122R	E125R	T148H	T149A
K116N	K120H	K121T	D122K	E125K	T148S	T149C
	K120N	K121H	D122H	E125Q	T148A	T149S
	K120T	K121R	D122E	E125Y	T148C	T149G
	K120Y	K121G	D122N			T149H
	K120Q	K121D				T149P
		K121Q				T149F
						T149N
						T149D
						T149V
T152R	D156R	E159K	Q529N	G532D	D535N	K538R
T152K	D156K	E159R	Q529T	G532N	D535H	K538V
T152W	D156Y	E159H	Q529H	G532S	D535V	K538Q
T152Y	D156W	E159Y	Q529A	G532H	D535T	K538W
T152H	D156Q	E159Q	Q529F	G532F	D535S	K538Y
T152Q	D156H		Q529G	G532K	D535A	K538F
T152E	D156I		Q529S	G532R	D535W	K538H
T152L	D156V		Q529P	G532Q	D535K	K538L
T152F	D156L		Q529W	G532A		K538M
	D156E		Q529D	G532L		K538C
				G532C		K538G
						K538A
						K538P

[0194]

D541N	Y542R	L585G	K591W	M592R	K595R	V596T
D541H	Y542K	L585H	K591F	M592K	K595Q	V596H
D541R	Y542H	L585F	K591Y	M592Q	K595Y	V596G
D541K	Y542Q		K591H	M592E	K595L	V596A
D541Y	Y542F		K591R	M592A	K595W	
D541I	Y542L		K591S		K595H	
D541A	Y542M		K591A		K595E	
D541S	Y542P		K591G		K595S	
D541E	Y542V		K591P		K595D	
	Y542N				K595M	
	Y542T					
S599G	K600R	K601R	Y616K	Y646R	W649H	
S599H	K600H	K601H	Y616R	Y646E	W649K	
S599N	K600G	K601Q	Y616E	Y646K	W649R	
S599D		K601T	Y616F	Y646H	W649Y	
			Y616H	Y646Q	W649E	
				Y646W	W649S	
				Y646N	W649V	
					W649T	

[0195]

[0196] 개별적 돌연변이 외에도, PAM 인식을 변경시키는 돌연변이의 조합을 조합하고, PAMDA를 통해 평가하여 LbCpf1 돌연변이의 제2 세대를 제공한다.

[0197] 실시예 3.

[0198] 3가지 방법을 사용하여 186개의 돌연변이를 시험하였다:

[0199] (1) 정제된 단백질 및 플라스미드 라이브러리를 사용하여 라이브러리에 걸쳐 각각의 점 돌연변이를 시험하는 PAMDA 검정으로 공지된 시험관내 방법 (Kleinstiver et al. *Nat Biotechnol* 37,:276-282 (2019)). 라이브러리 구성원의 고갈을 차세대 시퀀싱 (NGS)을 사용하여 점수화하였다. 고갈을 라이브러리 그 자체에 대해 (특정 PAM에 대한 절대 활성을 결정하기 위해) 또는 야생형 LbCas12a에 의한 절단에 대해 (돌연변이가 야생형과 비교하여 새로운 PAM 인식을 부여하였는지를 결정하기 위해) 계산하였다.

[0200] (2) *에스케리키아 콜라이*에서의 라이브러리를 사용하여 256개의 가능한 PAM NNNN 변이체체의 Cas12a 돌연변이의 결합을 시험하는 PAM-SCANR로 공지된 박테리아 방법 (Leenay et al. *Mol Cell* 62:137-147 (2016)). 이는 절단이 아니라, 단지 결합을 시험한다. 생성된 돌연변이는 촉매 영역 부근의 아무 데도 없기 때문에, 결합은 절단을 또한 반영할 것으로 예상된다 (293T 검정에서 나중에 확인됨). PAM-SCANR에 대한 이점은 점 돌연변이 뿐만 아니라 아미노산 점 돌연변이의 조합을 빠르고 정확한 방식으로 신속하게 시험하는 능력이다. 이 검정은 시험관내 절단 검정보다 더 엄격할 수 있다.

[0201] (3) 인간 HEK293T 세포에서의 인텔 검정. 이 검정은 가치 있는 진핵생물 인텔 데이터를 제공한다. 진핵생물에서 삽입 및 결실을 얻기 위해, 다수의 기준이 충족되어야 한다: CRISPR 효소는 세포에서 발현되고 안정할 필요가 있고, crRNA는 발현되고 정확하게 프로세싱될 필요가 있고, 단백질:RNA 복합체가 형성될 필요가 있고, 복합체는 안정할 필요가 있고, 복합체는 충분한 양으로 핵 내로 전위될 필요가 있고, 표적 DNA는 접근가능할 필요가 있고, DNA는 특정 가이드-RNA 디자인에 의해 잘 표적화되어야 하고, 이중-가닥 파단은 삽입 또는 결실 (인텔)을 통해 가끔의 DNA 복구 실수를 생성하는데 충분하게 높은 비율로 일어날 필요가 있다. 이는 진핵생물 검정을 이 연구에서 가장 엄격한 검정으로 만든다. 실험이 저-처리량임으로 인해 모든 256개라기 보다는, 하기 기재된 3개의 점 돌연변이체의 각각에 대해 수십 개의 PAM을 시험하였다. 종종 특정 가이드는 표적 접근성으로 인해 비효과적이기 때문에 가-음성을 회피하도록 시도하기 위해 3개의 상이한 표적을 각각의 PAM-돌연변이체 조합에 대

해 선택하였다.

[0202] 1. 시험관내에서 PAM 결합 및 절단의 결정

[0203] PAM 플라스미드-기반 라이브러리의 수립

[0204] 23 뉴클레오타이드 스페이스 서열에 대해 직접적으로 5'인 5개의 무작위 뉴클레오타이드로 이루어진 DNA 라이브러리를 제조하였다. LbCas12a는 4 뉴클레오타이드 프로토스페이스 인접 모티프 (PAM)를 갖는 것으로 공지되어 있지만, 본 발명자들은 실험 내에서 복제를 허용하기 위해 4라기 보다는 5개의 무작위 뉴클레오타이드를 사용하는 것을 선택하였다. 사용된 스페이스 서열은 5'-GGAATCCCTTCTGCAGCACCTGG (서열식별번호:30)였다. 라이브러리는 서열 5'-NNNNNGGAATCCCTTCTGCAGCACCTGG (서열식별번호:36)를 함유하였다. 5개의 무작위 뉴클레오타이드를 갖는 것은 이 라이브러리에서 검정된 1024개의 가능한 PAM을 생성한다.

[0205] 본 발명자들은 신규한 방식을 사용하여 이 라이브러리를 생성하였다. PAM-스페이스 융합물의 단일 무작위적 풀을 사용하고, 이전에 기재된 바와 같은 상보적 가닥을 생성하는 폴리머라제를 사용하기 보다는 (Kleinstiver et al. *Nat Biotechnol* 37,:276-282 (2019)), 본 발명자들은 보다 직접적인 방법을 선택하였다. 2개의 5'-인산화된 서열을 합성하였다:

[0206] 5'phos/CGATGTNNNNNGGAATCCCTTCTGCAGCACCTGGGCGCAGGTCACGAGG (서열식별번호:32) 및

[0207] AATTCCTCGTGACCTGCGCCAGGTGCTGCAGAAGGGATTCCNNNNNACATCGCATG/5'phos (서열식별번호:35).

[0208] 가열되고 어닐링된 경우, 2개의 NNNNN 서열 사이의 상보적 서열이 어닐링되고, 생성된 단부는 SphI 및 EcoRI 제한 엔도뉴클레아제에 의해 생성된 오버행에 상응하는 오버행을 갖는다. 2개의 올리고뉴클레오타이드를 95℃에서 5분 동안 열 사이클러에서 등물 비로 어닐링하고, 0.1℃/초 동안 25℃/실온으로 냉각시켰다.

[0209] 어닐링된 이중 가닥 단편을 SphI 및 EcoRI 소화된 pUC19 벡터에 직접적으로 라이게이션하였다. 라이게이션된 스페이스 구축물을 사용하여 XL1-블루 전기 적격 *E. coli* *콜라이* 세포 (에질런트)를 형질전환시키고, 글루코스 (SOC) 배지를 갖는 1 ml 슈퍼 옵티멀(Super Optimal) 브로쓰에서 37℃에서 1시간 동안 회수하였다. 분취물의 부분을 카르베니실린 플레이트 상에 플레이팅하여 라이게이션된 생성물의 존재에 대해 점검하였다. 나머지 형질전환된 세포를 50 mg/mL 카르베니실린으로 보충된 200 ml 루리아(Luria) 브로쓰 (LB)에서 16시간 동안 성장시켰다. 스페이스 플라스미드를 플라스미드 미디프랩 키트 (자이모 리서치(Zymo Research))를 사용하여 정제하였다.

[0210] PAM 라이브러리의 확인

[0211] 스페이스 벡터를 딥 시퀀싱 분석으로 처리하여 일루미나 마이섹을 제조업체의 프로토콜에 따라 사용하여 PAM의 각각의 위치에서의 A/T/G/C의 빈도를 계산하였다. 간략하게, 10 ng의 DNA를 PCR을 위한 주형으로서 사용하였다. 위상조정 유전자 특이적 정방향 및 역방향 PCR 프라이머를 표적 부위에 걸쳐 증폭시키도록 디자인하였다. 앰플리콘 라이브러리를 2-단계 PCR 방법을 사용하여 생성하였으며, 여기서 5' 꼬리를 갖는 1차 PCR은 2차 PCR이 일루미나 i5 및 i7 어댑터 서열 및 다중화 샘플을 분류하기 위한 바코드를 첨가하는 것을 허용한다. PCR 증폭을 하기 파라미터를 사용하여 수행하였다: 30초 동안 98℃; PCR1에 대해 25 사이클 및 PCR2에 대해 8 사이클 (98℃ 10초, 55℃ 20초, 72℃ 30초); 5분 동안 72℃; 12℃에서 보유. PCR 반응을 Q5 고-충실도 DNA 폴리머라제 (뉴 잉글랜드 바이오랩스(New England BioLabs), 미국 매사추세츠주 베버리)로 수행하였다. 2차 PCR 앰플리콘 샘플을 AM퓨어(AMPure) XP 비드를 제조업체의 지시서 (베크만 쿨터(Beckman Coulter), 미국 캘리포니아주 브레아)에 따라 사용하여 개별적으로 정제하고; 모든 정제된 샘플을 플레이트 판독기를 사용하여 정량화하고, 등물 비로 풀링하고, AATI 단편 분석기 (에질런트 테크놀로지스(Agilent Technologies), 미국 캘리포니아주 팔로 알토) 상에서 실행하였다. 풀링된 앰플리콘 라이브러리를 일루미나 마이섹 (2 X 250 쌍형성된 단부) 상에서 마이섹 시약 키트 v2 (일루미나, 미국 캘리포니아주 샌 디에고)를 사용하여 시퀀싱하였다.

[0212] 3개의 별개의 판독물을 라이브러리에 대해 생성하고, 평균하였다. 생성된 1024개의 라이브러리 구성원은 39개의 판독물의 평균 판독물 카운트 및 11.9개의 판독물의 표준 편차를 가졌다. 임의의 PAM 서열에 대한 평균 판독물의 최대 수는 74이고, 최소는 12였다. PAM 카운트는 정규 분포를 따랐다 (도 5).

[0213] LbCas12a 돌연변이의 클로닝

[0214] LbCas12a 서열, 이어서 뉴클레오플라스민 NLS 및 6x 히스티딘 태그로 구성된 DNA 카세트를 합성하고 (진위즈 (GeneWiz)) (서열식별번호:52), NcoI 및 XhoI 사이의 pET28a 벡터 내로 클로닝하여, pWISE450 (서열식별번호:53)을 생성하였다.

호:53)을 생성하였다. 추가의 글리신을 Met-1 및 Ser-2 사이의 서열에 첨가하여 클로닝을 용이하게 하였다. 본 문서 전반에 걸쳐 넘버링은 이 가의 글리신을 배제한다. 이어서, 186개의 상이한 아미노산 점 돌연변이 (**표 2**)를 유사한 전략을 사용하여 생성하여 186개의 상이한 플라스미드 벡터를 생성하였다.

[0215] LbCas12a 돌연변이의 발현 및 정제

[0216] BL21 스타(Star) (DE3) 세포 (써모피셔 사이언티픽(ThermoFisher Scientific))에서의 각각의 돌연변이체의 글리세롤 스톱을 사용하여 1 mL의 배지를 24-웰 블록에서 50 µg/mL의 카나마이신으로 접종하였다. 배양물을 에어포어(AirPore) 테이프 쉬트 (키아젠(Qiagen))로 밀봉하고, 37°C에서 밤새 진탕하면서 성장시켰다. 다음 날 아침, 카나마이신을 갖는 4 mL의 ZYP 자가유도 배지를 100 µL의 밤샘 배양물로 접종하고, 37°C에서 진탕하면서 0.2 내지 0.5의 OD 600nm 범위까지 인큐베이션하였다. 온도를 18°C로 감소시키고, 배양물을 단백질 발현을 위해 밤새 성장시켰다. 세포를 원심분리에 의해 수거하고, 펠릿을 -80°C에서 저장하였다.

[0217] 하기 완충제를 세포 용해 및 정제에 사용하였다. 비-이온성 세정제, 용해제, 환원제, 프로테아제 억제제, 완충제, 및 염을 포함하는 용해 완충제. 이 용액은 박테리아를 용해시키고, 점도를 감소시키고, 간접 뉴클레아제 없이 효소의 하류 정제를 허용할 수 있었다. 완충제 A는 20 mM 헤페스(Hepes)-KOH pH 7.5, 0.5 M NaCl, 10% 글리세롤, 2 mM TCEP 및 10 mM 이미다졸 pH 7.5로 구성되었다. 완충제 B는 완충제 A와 동일하지만 또한 20 mM의 이미다졸을 포함하였다. 완충제 C는 20 mM 헤페스-KOH pH 7.5, 150 mM NaCl, 10% 글리세롤, 0.5 mM TCEP, 및 200 mM 이미다졸 pH 7.5를 함유하였다.

[0218] 정제를 다중-웰 형식을 사용하여 수행하였다. 2개의 스테인레스 스틸 5/32" BB를 세포 펠릿을 함유하는 모든 웰에 첨가하였다. 펠릿을 0.5 mL 차가운 용해 완충제에 재현탁시키고, 실온에서 30분 동안 케도 혼합하면서 인큐베이션하였다. 조 용해물, 0.5 mL를 사전-평형화된 His 멀티트랩(MultiTrap)TM 플레이트 (싸이타바 라이프사이언시스(Cytiva LifeSciences))에 첨가하였다. 플레이트를 실온에서 5분 동안 인큐베이션하여 단백질 결합을 허용하였다. 나머지 단계는 제조업체의 지시서에 따라 수행하였다. 간략하게, 플레이트를 0.5 mL 완충제 A로 2회 세척하고, 이어서 0.5 mL 완충제 B로 1회 세척한 후, 0.2 mL의 완충제 C에서 용리하였다. 단백질 농도를 피어스(Pierce)TM 쿠마시 플러스(Coomassie Plus) (브래드포드(Bradford)) 검정 시약을 사용하여 결정하였다. 단백질 용리액을 4°C에서 저장하였다.

[0219] wtLbCas12a에 의한 PAM 라이브러리의 시험 점단

[0220] 실험의 3가지 측면을 평가하기 위한 사전-시험을 수행하였다: 실험이 비-특이적 뉴클레아제가 없음을 보장하고, crRNA 가이드의 첨가시 라이브러리로부터 NTTV PAM에서 고갈이 있었음을 보장하고, CTTTA의 첨가된 샘플에 대해 15분에서 고갈의 정도를 보기 위해.

[0221] 시험 고갈을 위한 반응 조건은 하기와 같았다: 뉴클레아제-무함유수, 3 µL NEB 완충제 2.1 (뉴 잉글랜드 바이오랩스), 300 nM의 crRNA의 3 µL 스톱 (5'-AAUUCUACUAAGUGUAGAUGGAUCCUUCUGCAGCACCUGG-3' (**서열식별번호:62**), 신테고 코포레이션(Synthego Corporation)), 및 1 µM 스톱의 1 µL의 정제된 wtLbCas12a를 함유하는 27 µL 총 부피를 실온에서 10분 동안 인큐베이션하였다. 3 µL의 10 ng/µL 스톱을 첨가하여 반응을 시작시켰다. 라이브러리를 그대로 첨가하거나, 1 µL의 CTTTA-함유 플라스미드를 먼저 0.75 ng/µL에서 첨가하였다. 총 부피는 모두 30 µL였다. 반응물을 37°C에서 15분 동안 인큐베이션하였다.

[0222] **표 3**은 실험의 결과를 제공한다. NTTV 서열에 대한 라이브러리 카운트는 219 내지 515 카운트였으며 (칼럼 2), crRNA의 부재 하에서 기재된 바와 같이 정제된 야생형 단백질을 첨가하는 것은 라이브러리 구성원의 고갈을 발생시키지 않았고 (칼럼 3), crRNA 및 단백질의 첨가는 모든 NTTV 함유 PAM의 고갈을 발생시켰고 (칼럼 4), 라이브러리에 CTTTA를 첨가한 것은 CTTTA NGS 카운트만큼 많은 대략 35배를 발생시켰고 (칼럼 5), wtLbCas12a 및 crRNA의 첨가는 10,776에서 193 카운트로 감소하는 CTTTA를 포함하여 모든 라이브러리 구성원의 고갈을 발생시켰다. ACGA가 LbCas12a에 의해 인식되는 PAM이 아니기 때문에 예상되는 바와 같이, 시험된 조건 하에서 고갈을 나타내지 않는 NACGA PAM-함유 라이브러리 구성원이 또한 나타내어진다. 따라서, **표 3**에 나타내어진 바와 같이, NTTV PAM 라이브러리 구성원은 wtLbCas12a에 의해 효율적으로 절단되고 고갈되는 반면, wtLbCas12a에 의해 인식되지 않는 PAM (NACGA)은 그렇지 않다.

[0223] 표 3. PAM 라이브러리 절단 및 고갈.

PAM	라이브러리 구성원 NGS 카운트	라이브러리 +wtLbCas12a NO crRNA	라이브러리 +wtLbCas12a+crRNA (15 분)	CTTTA 가 첨가된 라이브러리 (15 분)	CTTTA+wtLbCas12a+crRNA 가 첨가된 라이브러리
(기재)					
ATTTG	314	336	13	294	30
CTTTC	219	245	13	225	15
ATTTT	317	392	22	294	22
ATTTA	515	490	37	396	37
TTTTT	359	382	33	286	28
TTTTA	489	666	46	446	32
GTTTG	330	438	34	297	30
GTTTA	395	492	41	378	31
CTTTG	299	333	35	263	28
TTTTG	348	430	41	343	24
CTTTA	293	379	35	10776	193
GTTTC	299	349	39	269	14
(비-기재)					
AACGA	215	271	237	197	211
GACGA	273	282	274	251	282
CACGA	264	294	253	194	235
TACGA	302	268	255	225	243

[0224]

[0225]

표 3에서의 결과는 (1) wtLbCas12a의 효과적이고 뉴클레아제-무함유 정제가 달성되었고, (2) 라이브러리는 PAM-기재를 함유하는 구성원에 대해 시험된 조건 하에서 고갈될 수 있고, (3) 스페이서-표적화 crRNA의 첨가시 고갈이 발생하고, (4) 효소-crRNA 복합체는 다량의 CTTTA 기재가 기재의 고갈을 변경시키지 않기 때문에 개별적 라이브러리 구성원을 다량 초과함을 나타낸다.

[0226]

LbCas12a 돌연변이체에 의한 PAM 라이브러리의 절단

[0227]

동일한 반응 조건을 wtLbCas12a의 시험 실시예에 나타내어진 바와 같이 186개의 PAM 돌연변이의 각각에 대해 시험하였다. 3개의 시점을 각각의 돌연변이에 대해 선택하였다: 37℃에서 75, 435, 및 900초. 다수의 라이브러리-단독 대조군이 포함되었다. 생성물을 일루미나 하이섹(HiSeq) 분석 (진위즈)으로 처리하였다. 데이터는 표 4에 보고된다.

[0228]

절대 고갈 점수 프로세싱

[0229]

본 발명자들은 임의의 5 뉴클레오타이드 PAM에 대해 4가지 가능성 사이에 고갈의 작은 차이를 관찰하였다. 다시 말해서: 임의의 4 뉴클레오타이드 서열에 대해 ANNNN, CNNNN, GNNNN, 및 TNNNN은 유사한 PAM 고갈을 가졌다. 이는 NTTTT 서열이 N이 A, C, G, 또는 T인 경우와 무관하게 모두 유사한 양으로 고갈되었음을 나타낸 야생형 LbCas12a 실험에서 본 발명자들이 관찰한 것과 일치하였다 (표 3). 두번째로, 본 발명자들은 75, 435, 및 900초의 3개의 시점이 모두 유사한 고갈을 가졌음을 관찰하였다. 이는 반응이 37℃에서 단지 75초 후에 거의 완결되었음을 지시하였다. 따라서, 본 발명자들은 5nt 라이브러리로부터 모든 4개의 4nt PAM을 평균하고, 모든 3개의 시점을 평균하여, 각각의 PAM에 대해 12개의 데이터 점을 효과적으로 발생시킬 수 있었다. 이어서, 본 발명자들은 그 평균을 취하고, 이를 각각의 PAM에 대한 중위 라이브러리 값으로 나누었다. 이는 본 발명자들에게 모든 186개의 돌연변이체에 대한 각각의 4 뉴클레오타이드 PAM에 대한 고갈 점수를 제공하였다. 10의 고갈은 그 4 nt PAM을 갖는 4개의 모 플라스미드 라이브러리 구성원의 90%가 고갈되었음을 지시하는 반면, 20의 점수는 95% 고갈을 지시한다.

[0230]

야생형 LbCas12a에 대한 고갈 점수는 TTTT 서열에 대해 9.2였으며, 따라서 이 9.2 점수에서 또는 그보다 양호하게 PAM을 절단한 임의의 돌연변이체는 기준으로서 야생형을 사용하여 효율적인 것으로 간주되었다. 예를 들어, 표 4는 wtLbCas12a에서 또는 그보다 양호하게 라이브러리로부터 45개의 상이한 PAM 4mer을 시험관내에서 고갈시키는 돌연변이체 LbCas12a-K595Y가 TTTT-함유 서열을 절단함을 나타낸다. 이 분석을 사용하여 186개의 돌연변이의 각각을 점수화하여 돌연변이체의 각각에 의한 시험관내 PAM 인식 및 절단을 결정하였다. 인식 서열을 함유하는 데이터는 표 4에 나타내어지며, 이는 9.2의 TTTT-함유 서열에 대한 wtLbCas12a 점수에서 또는 그보다 양

호한 LbCas12a-K595Y PAMDA 고갈 점수를 나타낸다.

표 4.

PAM 고갈 데이터

돌연변이체	4mer	고갈
D122H	TCCC	18.76
D122H	ATTC	17.15
D122H	GTTG	16.7
D122H	TCCG	15.92
D122H	TTCA	14.95
D122H	TCGC	14.8
D122H	TTCT	14.68
D122H	CCCC	14.61
D122H	CTTG	14.37
D122H	GTTG	14.14
D122H	ATTA	13.66
D122H	TCTA	12.99
D122H	CCTG	12.96
D122H	GTTA	12.87
D122H	CCTA	12.77
D122H	TTCG	12.68
D122H	CCTC	12.5
D122H	ATTG	12.43
D122H	TTGA	12.14
D122H	TTAC	11.87
D122H	GCGC	11.86
D122H	CTTA	11.78
D122H	CTCA	11.77
D122H	TCTC	11.75
D122H	TCCA	11.74
D122H	CCGC	11.64
D122H	TCCG	11.3
D122H	CCCG	11.05
D122H	CTTC	11.01
D122H	TCAC	10.9
D122H	TTTT	10.85
D122H	TTTC	10.27
D122H	TTTA	10.18
D122H	ACTC	10.17
D122H	GCTC	10.01
D122H	TTGC	9.8
D122H	TTCC	9.8
D122H	TTGG	9.66
D122H	CCCA	9.34
D122K	TCCC	13.07
D122K	TCCG	13.02
D122K	TCTA	12.97
D122K	GTTA	11.7
D122K	TCTC	11.54
D122K	CCCC	11.24
D122K	TCTG	10.77
D122K	GTTG	10.77
D122K	TTAC	10.53
D122K	ATTC	10.5
D122K	CCGC	10.33
D122K	TTCT	10.1
D122K	CTCC	10.09
D122K	GTTG	10.04
D122K	ATTG	10.02
D122K	TCGC	9.84
D122K	TCCG	9.84
D122K	ATTA	9.76

D122K	TTCA	9.46
D122K	CTTG	9.44
D122K	CTTC	9.39
D122K	CCTA	9.34
D122K	TTTT	9.23
D122K	TCCA	9.11
D122K	TCAC	8.97
D122K	GCTA	8.88
D122K	TTCC	8.77
D122K	GCGC	8.76
D122K	CCTC	8.6
D122K	TTGC	8.56
D122K	TTCG	8.43
D122K	TTTC	8.41
D122N	TCCC	11.48
D122N	TTTT	11
D122N	TCTA	10.38
D122N	GTTG	10.31
D122N	ATTC	10.27
D122N	CCCC	10.22
D122N	TTCA	10.16
D122N	TCCG	10.02
D122N	TTCT	9.86
D122N	CTTG	9.79
D122N	TCGC	9.72
D122N	TCTG	9.66
D122N	ATTA	9.62
D122N	CTTA	9.61
D122N	CTTC	9.49
D122N	TCCA	9.46
D122N	TTCC	9.44
D122N	TTGC	9.39
D122N	TTAC	9.36
D122N	GTTC	9.22
D122N	CCTA	9.18
D122N	CCTC	9.16
D122N	GTTA	9.14
D122N	TTCG	9.04
D122N	TCTC	9.01
D122N	TTGG	8.97
D122N	ATTG	8.7
D122R	TCCC	10.09
D122R	GTTC	9.87
D122R	CCCC	9.85
D122R	ATTG	9.82
D122R	TTGC	9.8
D122R	GTTA	9.53
D122R	TTAC	8.94
D122R	TTCG	8.87
D122R	CTTA	8.8
D122R	GCGC	8.73
D122R	TTCT	8.6
D122R	CCGC	8.6
D122R	CCTC	8.53
D122R	ATTA	8.25
D122R	TCAC	8.17
D122R	TTTT	8.09
D122R	GTTG	7.81
D156E	TTCG	14.53
D156E	TTTT	14.35
D156E	CTTA	13.86
D156E	TTGC	13.64
D156E	ATTC	13.58
D156E	TTCT	13.51
D156E	TTCA	13.24

D156E	TCTG	13.06
D156E	TCCC	13.05
D156E	CTTG	13
D156E	TTCC	12.52
D156E	ATTG	12.43
D156E	TCTA	12.35
D156E	TTAC	12.06
D156E	TCTC	11.97
D156E	GTTG	11.9
D156E	ATTA	11.88
D156E	CCCG	11.88
D156E	CTTC	11.51
D156E	TTTC	11.46
D156E	CCTG	11.34
D156E	CCTA	11.19
D156E	GTTA	11.16
D156E	GTTG	10.73
D156E	TTTG	10.61
D156E	TTTA	10.25
D156E	CCCC	10.16
D156E	CCTC	9.91
D156E	CTCC	9.83
D156E	CTCA	9.57
D156E	TTGG	8.81
D156E	GCTC	8.57
D156H	CCTG	13.3
D156H	TCTG	12.9
D156H	TTCC	12.22
D156H	TTAC	12.08
D156H	ATTC	12.04
D156H	CCTC	11.12
D156H	CCCA	10.92
D156H	TCGC	10.89
D156H	TTGC	10.87
D156H	GCTA	10.68
D156H	GTTG	10.6
D156H	CCCG	10.37
D156H	CCTA	10.3
D156H	GTTA	10.27
D156H	GCTC	10.15
D156H	CTCC	10.03
D156H	GCGC	10
D156H	ATTA	10
D156H	TCCG	9.99
D156H	TCAC	9.98
D156H	TCTA	9.95
D156H	TCTC	9.74
D156H	TTCT	9.69
D156H	GTTG	9.51
D156H	CCGC	9.45
D156H	TTCA	9.4
D156H	CCCC	9.28
D156H	TCGA	9.14
D156H	TTTT	9.13
D156H	GCCC	9.1
D156H	TCCA	9.05
D156H	ACTC	8.98
D156H	TTCG	8.89
D156H	ATTG	8.88
D156H	ACTA	8.4
D156H	CTCG	8.13
D156I	CCGC	10.23
D156I	TCAG	9.16
D156I	TCGA	8.68
D156I	ACCC	8.52

D156I	GCGC	8.42
D156I	ATTC	8.37
D156I	TCCC	8.33
D156I	ACTC	8.31
D156I	GCCC	8.05
D156I	GTTG	7.84
D156K	TTCC	12.11
D156K	TTCT	12.02
D156K	ATTA	11.58
D156K	CCTG	11.51
D156K	CCTA	11.41
D156K	TCCC	11.39
D156K	TCCG	11.32
D156K	GTTG	11.24
D156K	TTGA	11.23
D156K	ATTC	11.16
D156K	ATTG	10.87
D156K	GCTC	10.8
D156K	GTTC	10.77
D156K	TTGC	10.71
D156K	CTTC	10.65
D156K	TTCA	10.57
D156K	TCTG	10.41
D156K	CCCC	10.25
D156K	TCGC	10.15
D156K	CTCC	10
D156K	TCTA	9.99
D156K	TTCG	9.98
D156K	TTAC	9.83
D156K	CCTC	9.73
D156K	TCCA	9.65
D156K	TTTT	9.64
D156K	ACTA	9.56
D156K	GTTG	9.44
D156K	TCTC	9.37
D156K	CCCG	9.21
D156K	TCAC	9.18
D156K	TTTC	9.13
D156K	CTTG	9.05
D156K	CTCA	9
D156K	TTTG	8.92
D156K	GCTG	8.91
D156K	TTGG	8.53
D156K	ACTC	8.42
D156L	GTTG	15.01
D156L	ATTC	14.48
D156L	CTTC	14.39
D156L	TCCC	14.28
D156L	ATTA	13.69
D156L	TCCG	13.57
D156L	CCCG	13.52
D156L	TTAC	13.32
D156L	CCTG	13.25
D156L	TCGC	13.04
D156L	TTCT	12.9
D156L	CCTC	12.55
D156L	ATTG	12.52
D156L	CCCC	12.48
D156L	TTGC	12.37
D156L	GTTG	12.31
D156L	ACTA	12.29
D156L	TCTC	11.89
D156L	TCTA	11.73
D156L	CCTA	11.49
D156L	TTTT	11.47

D156L	CTTA	11.36
D156L	CTTG	11.36
D156L	TTCC	11.36
D156L	TCTG	11.07
D156L	GTTA	11.02
D156L	CCCA	10.93
D156L	GCTA	10.87
D156L	TTTG	10.8
D156L	TTCC	10.73
D156L	TCCA	10.63
D156L	CTCC	10.14
D156L	TTCA	10.07
D156L	TTTC	9.96
D156L	TTCA	9.82
D156L	TCAC	9.8
D156L	TTTA	9.69
D156L	ACTG	9.67
D156L	GCGC	9.08
D156L	GCGC	8.68
D156L	GCTC	8.48
D156L	TCAG	8.46
D156L	CTCG	8.35
D156Q	GTTT	15.61
D156Q	TCCC	15.02
D156Q	ATTG	14.2
D156Q	TTCA	14.13
D156Q	TCTA	13.69
D156Q	CCTA	13.4
D156Q	CCTG	13.17
D156Q	ATTG	12.94
D156Q	TCTG	12.76
D156Q	TCTC	12.71
D156Q	CTCC	12.58
D156Q	CCCC	12.54
D156Q	CTTG	12.5
D156Q	GTTG	12.46
D156Q	CCTC	11.6
D156Q	CTCA	11.56
D156Q	TTTT	11.44
D156Q	CCCG	11.41
D156Q	ACTG	11.38
D156Q	ATTA	11.35
D156Q	GTTA	11.2
D156Q	CTTA	11.06
D156Q	TTCA	10.96
D156Q	CTTC	10.86
D156Q	TTCC	10.7
D156Q	TTGC	10.65
D156Q	TTAC	10.6
D156Q	ACTA	10.58
D156Q	TGCG	10.58
D156Q	TTCT	10.47
D156Q	TTGA	10.35
D156Q	TCCG	10.31
D156Q	GCTA	10.26
D156Q	GCTG	10.21
D156Q	TTTG	10.17
D156Q	GCTC	10.12
D156Q	TTTC	10.01
D156Q	TCAC	9.87
D156Q	CCCA	9.86
D156Q	CTCG	9.78
D156Q	GCGC	9.58
D156Q	TTCC	9.43
D156Q	TTTA	9.38

D156Q	CTTT	9.36
D156Q	TTAG	9.1
D156R	CCTG	11.46
D156R	GCTA	11.19
D156R	TTCT	11.15
D156R	CCGC	11.01
D156R	CCTC	10.98
D156R	TCCA	10.93
D156R	CCCC	10.57
D156R	TCTG	10.53
D156R	TTGC	10.34
D156R	CTTC	9.92
D156R	TCGC	9.46
D156R	ACTA	9.41
D156R	GTTG	9.21
D156R	TCTA	9.19
D156R	TTTT	9.1
D156R	TCCC	8.96
D156R	CTCA	8.63
D156W	ACTA	13.47
D156W	CCTG	12.62
D156W	CCCA	11.99
D156W	GTTG	11.77
D156W	TCTG	11.72
D156W	CCTA	11.4
D156W	CCCG	11.4
D156W	GCTA	11.28
D156W	TCGC	11.27
D156W	CCTC	11.21
D156W	CCCC	11.16
D156W	TTGC	10.91
D156W	TCCG	10.66
D156W	ATTG	10.65
D156W	GTTT	10.5
D156W	TCCA	10.43
D156W	TCTC	10.34
D156W	TCTA	10.25
D156W	TTCC	10.18
D156W	TTCT	10.1
D156W	GTTA	10.08
D156W	TTCC	9.92
D156W	TTTT	9.65
D156W	CTCA	9.63
D156W	CTTC	9.54
D156W	TCCC	9.49
D156W	ATTA	9.47
D156W	CTTG	9.45
D156Y	TCTG	12.9
D156Y	ATTG	12.41
D156Y	TCTA	12.32
D156Y	TCCC	11.75
D156Y	TCTC	11.68
D156Y	ATTG	10.93
D156Y	CTTA	10.52
D156Y	TTCC	10.34
D156Y	GTTG	10.31
D156Y	CTTC	10.17
D156Y	TTCT	10.02
D156Y	TTGC	9.75
D156Y	TTTA	9.64
D156Y	TTCC	9.59
D535A	TTCC	17.75
D535A	TTCA	16.69
D535A	CTTG	16.57
D535A	CTTA	15.51

D535A	TCTC	15.39
D535A	TCGC	15.38
D535A	TCTA	15.02
D535A	TTGC	14.6
D535A	TCCC	14.6
D535A	TTTA	14.59
D535A	TTAC	14.56
D535A	TTTT	13.92
D535A	TTCC	13.69
D535A	TTTC	12.57
D535A	CTTC	12.57
D535A	TCTG	12.5
D535A	TCCG	12.08
D535A	TTTG	11.96
D535A	ATTA	11.71
D535A	TTAG	11.43
D535A	CCCC	11.28
D535A	ATTG	11.06
D535A	TTGA	10.6
D535A	TTCT	10.56
D535A	GCGC	10.49
D535A	ATTC	10.41
D535A	TCCA	10.31
D535A	GTTG	10.15
D535A	TCAC	9.7
D535A	GTTC	8.85
D535A	CCCG	8.33
D535A	TCAG	8.01
D535H	TCAC	12.82
D535H	TTCA	11.72
D535H	TCCG	11.42
D535H	TTCC	11.3
D535H	TCTA	11.27
D535H	TCTC	10.82
D535H	CCCC	10.68
D535H	TTCT	10.51
D535H	TTTT	10.43
D535H	TCAG	10.24
D535H	ATTA	10.22
D535H	ATTG	10.1
D535H	TTAC	10.04
D535H	TCCC	10.02
D535H	TTCC	9.97
D535H	CTTC	9.97
D535H	TCGC	9.95
D535H	TTGC	9.74
D535H	TTTG	9.52
D535H	CCTA	9.5
D535H	CTTA	9.28
D535H	CCAG	9.27
D535H	TCTG	9.24
D535H	TTTA	9.15
D535H	TTGG	8.73
D535H	CCAC	8.4
D535H	GCAC	8.08
D535K	TTTA	14.65
D535K	TTTC	13.64
D535K	TTTG	13.58
D535K	TTGC	11.64
D535K	TTAC	10.49
D535N	TTGC	10.94
D535N	TTAG	10.83
D535N	TTCC	10.61
D535N	TTTC	10.54
D535N	TTTT	10.52

D535N	TCGC	10.51
D535N	TTAC	10.39
D535N	TTTA	10.34
D535N	TCAC	10.22
D535N	TTTG	10.01
D535N	TTAA	9.04
D535N	TCAG	8.94
D535N	TCCC	8.46
D535N	TTGG	7.82
D535S	TTCC	12.34
D535S	TCTG	12.09
D535S	CTTG	12.07
D535S	CTTA	11.85
D535S	TCTC	11.41
D535S	TTTA	10.73
D535S	TCCC	10.7
D535S	TTAC	10.6
D535S	TTAG	10.49
D535S	TTCC	10.47
D535S	TCAG	10.46
D535S	TCTA	10.4
D535S	TTCA	10.36
D535S	TCGC	10.31
D535S	ATTA	10.14
D535S	TTGA	10.13
D535S	TTTT	10.1
D535S	TCAC	10.07
D535S	TTTC	9.76
D535S	CCCC	9.75
D535S	TTGC	9.72
D535S	CCCG	9.66
D535S	ATTG	9.64
D535S	GTTG	9.6
D535S	TTCT	9.51
D535S	TTTG	9.4
D535S	CCTG	9.34
D535S	CCCG	9.18
D535S	TCCA	9.11
D535S	TTGG	9.02
D535S	TTAA	8.98
D535T	CTTA	14.25
D535T	TTCC	13.61
D535T	TTTT	13.51
D535T	TCCC	12.73
D535T	TCTA	12.41
D535T	CTTG	12.36
D535T	TCTG	11.75
D535T	TTGC	11.65
D535T	TCCG	11.65
D535T	TCTC	11.38
D535T	TCGC	11.38
D535T	TTAC	11.38
D535T	TTTA	11.34
D535T	TTTG	11.25
D535T	CTTC	11.2
D535T	TTGG	10.8
D535T	TTCT	10.42
D535T	TCAC	10.06
D535T	TCAG	9.75
D535T	TTTC	9.07
D535T	GTTA	8.61
D535T	CTCC	8.06
D535V	TTTG	15.43
D535V	TTTA	13.9
D535V	TTTC	12.89

[0234]

D535V	TTGC	9.81
D535V	TTTT	9.77
D541A	ATTG	9.99
D541E	CTTC	15.41
D541E	TTTG	14.24
D541E	TTTT	14.15
D541E	CTTA	13.23
D541E	TTTC	12.78
D541E	CTTG	11.68
D541E	TTCA	11.56
D541E	TTCC	11.02
D541E	TTTA	10.69
D541E	CCGC	10.13
D541E	CTTC	9.86
D541E	TCTC	9.77
D541E	TCCG	9.75
D541E	TCCC	9.37
D541E	TTGC	9.33
D541E	TTCG	9.03
D541E	CCCC	8.91
D541E	TTAC	8.69
D541E	CCTG	8.3
D541E	ATTG	8.18
D541E	CCTA	7.96
D541E	GTTA	7.84
D541H	TTCG	8.67
D541H	CTTC	8.58
D541I	TTTC	11.44
D541I	TTTG	10
D541I	CTTA	9.11
D541I	TTTT	8.7
D541N	TCCC	8.36
D541R	TTTG	8.99
D541Y	CTTA	10.32
D541Y	TTTC	9.36
D541Y	CTTC	8
E125K	TTGC	10.22
E125K	TCTC	9.44
E125K	ATTG	9.32
E125K	TTAC	9.11
E125K	CCCC	8.94
E125K	TTCG	8.93
E125K	CTTC	8.83
E125K	TCCC	8.81
E125K	TTCT	8.62
E125K	TTTA	8.56
E125K	TTTT	8.43
E125K	TTGA	8.25
E125K	TCGC	8.15
E125K	ATTG	8.04
E125Q	TTCA	12.74
E125Q	CTTC	12.55
E125Q	CTTA	12.14
E125Q	TCCC	11.7
E125Q	TTCG	11.66
E125Q	TCTA	11.06
E125Q	TTGC	10.88
E125Q	ATTG	10.84
E125Q	ATTA	10.53
E125Q	TTTG	10.51
E125Q	TTTT	10.08
E125Q	CCCC	10.04
E125Q	TCTC	10.03
E125Q	CCCG	9.96
E125Q	TCCA	9.94

E125Q	TTGG	9.92
E125Q	TTCC	9.88
E125Q	TCCG	9.84
E125Q	TTCT	9.61
E125Q	CTTG	9.55
E125Q	ATTG	9.45
E125Q	GTTG	9.29
E125Q	TCTG	9.17
E125Q	TTTA	9.1
E125Q	TTAC	8.56
E125Q	TCGC	8.55
E125Q	TCAC	8.5
E125Q	CCTG	8.42
E125Q	CTCG	8.25
E125R	CCCG	11.21
E125R	TCCC	8.35
E125R	TCCA	8.03
E125Y	TTGC	11.77
E125Y	TTGC	10.69
E125Y	CTTC	9.94
E125Y	CTTC	9.82
E125Y	TTTT	9.39
E159K	TTCT	12.66
E159K	CCCC	12.33
E159K	ATTG	11.53
E159K	CTTG	11.38
E159K	CTTA	11.24
E159K	TCTC	11.21
E159K	ATTG	11.05
E159K	TCCC	11.04
E159K	TTCG	10.86
E159K	TTCA	10.67
E159K	TCTA	10.64
E159K	TTGC	10.56
E159K	CTTC	10.36
E159K	TCTG	10.17
E159K	TTAC	10.12
E159K	TTCC	9.92
E159K	TCGC	9.78
E159K	GTTA	9.69
E159K	TCCG	9.67
E159K	CCTG	9.31
E159K	GTTG	9.04
E159K	GCCC	8.85
E159K	CCTC	8.85
E159K	GCGC	8.65
E159K	ACTC	8.23
E159Q	TTTT	11.37
E159Q	TTCG	10.89
E159Q	CTTA	10.51
E159Q	TTGC	10.5
E159Q	TTTC	10.15
E159Q	TTCC	9.95
E159Q	CTTG	9.91
E159Q	TTCA	9.8
E159Q	TTTA	9.44
E159Q	TTTG	9.27
E159R	GTTG	14.77
E159R	GTTG	13.13
E159R	TCCG	12.2
E159R	TTGC	11.82
E159R	TTCG	11.71
E159R	CCTC	11.67
E159R	ATTG	11.35
E159R	CCCC	11.09

E159R	TTAC	11.05
E159R	CTTC	10.88
E159R	TCTG	10.85
E159R	ATTG	10.72
E159R	TTCT	10.71
E159R	TCTC	10.63
E159R	GTTA	10.62
E159R	TTCA	10.59
E159R	CTTG	10.51
E159R	TCGC	10.29
E159R	TCTA	10.16
E159R	ATTA	10.06
E159R	ACCC	9.76
E159R	TTTG	9.55
E159R	CCTA	9.52
E159R	CTTA	9.47
E159R	GCGC	9.32
E159R	ACTC	8.95
E159R	CCGC	8.91
E159Y	TTTT	11.72
E159Y	CTTC	11.71
E159Y	TTTA	11.42
E159Y	CTTA	10.79
E159Y	TTTG	10.39
E159Y	TTGC	10.37
E159Y	TTCC	9.94
E159Y	CTTG	9.51
E159Y	TCCC	9.43
G532A	TTTT	13.95
G532A	CTTG	12.83
G532A	ATTG	12.56
G532A	CTTC	11.75
G532A	TCTC	11.56
G532A	TTCC	11.55
G532A	TTCG	11.5
G532A	TCTA	11.01
G532A	TTCA	10.91
G532A	CCCC	10.36
G532A	CTTA	10.35
G532A	TTGC	10.19
G532A	GTTG	10.18
G532A	TCCC	10.1
G532A	TCTG	9.94
G532A	ATTG	9.6
G532A	TTTA	9.47
G532A	GTTA	9.4
G532C	TTTT	15.09
G532C	TTCC	12.43
G532C	TTTC	12.3
G532C	TTGC	11.56
G532C	CTTA	11.27
G532C	TTAC	11.02
G532C	TTTG	10.94
G532C	TTCA	10.93
G532C	CTTG	10.62
G532C	TTTA	10.31
G532C	TCCC	10.29
G532C	TCTA	10.26
G532C	TCTG	10.19
G532C	GTTA	9.93
G532C	TCTC	9.91
G532C	CTTC	9.78
G532C	GTTG	8.88
G532D	CTTG	21.38
G532D	CTTA	15.81

G532D	TTCC	15.28
G532D	TTTT	14.57
G532D	CTTC	13.93
G532D	TTTC	13.65
G532D	TTTG	13.59
G532D	TTCG	13.54
G532D	TTCT	13.13
G532D	TTCA	12.84
G532D	CCCC	12.66
G532D	TTTA	12.54
G532D	TTGC	12.49
G532D	GTTG	11.87
G532D	TCCG	11.41
G532D	ATTG	11.27
G532D	ATTA	11
G532D	TCCC	10.91
G532D	ATTG	10.56
G532D	TCTA	10.36
G532D	CTCC	9.98
G532D	TTAC	9.81
G532D	TCGC	9.2
G532F	CTTA	13.56
G532F	ATTG	11.28
G532F	CTTC	11.1
G532F	TTCA	10.82
G532F	TTGC	10.74
G532F	TTTT	10.6
G532F	TCTA	10.52
G532F	TTAC	10.48
G532F	TTTA	10.31
G532F	TTTG	10.25
G532F	ATTA	10.2
G532F	GTTA	9.82
G532F	TTTC	9.64
G532F	ATTG	9.47
G532F	TCCC	9.44
G532F	CTTG	9.28
G532F	TCTC	9.23
G532F	TCCG	8.95
G532F	TCTG	8.91
G532F	CCCC	8.33
G532F	GTTG	8.17
G532F	CCTA	8.08
G532H	TTCA	16.2
G532H	CTTC	15.09
G532H	GTTG	14.43
G532H	TTTT	13.93
G532H	TTCG	13.87
G532H	CTTG	13.72
G532H	TCTG	13.22
G532H	GTTG	12.71
G532H	TTCC	12.66
G532H	CTTA	12.54
G532H	TTGC	12.31
G532H	TCTA	12.2
G532H	TTCT	12.15
G532H	ATTG	11.94
G532H	CCCC	11.91
G532H	TTTC	11.89
G532H	TCCG	11.71
G532H	TTTA	11.67
G532H	TCCC	11.6
G532H	TTTG	11.23
G532H	CTCA	11.23
G532H	ATTA	11.16

G532H	TCCA	11.11
G532H	ATTG	11.02
G532H	TTAC	10.74
G532H	TCTC	10.55
G532H	TCCG	10.42
G532H	CCTA	10.3
G532H	CTCC	9.99
G532H	GTTA	9.77
G532H	CCCA	9.61
G532H	CCTG	9.15
G532H	CCTC	9.03
G532H	CCCG	8.84
G532H	TCAC	7.76
G532K	CTCA	13.19
G532K	GCCC	12.17
G532K	CTTA	11.31
G532K	CTTC	11.17
G532K	TTGC	10.91
G532K	TCTC	10.78
G532K	ATTA	10.64
G532K	GCCG	10.53
G532K	TTAC	10.41
G532K	TTTT	9.96
G532K	TTCA	9.94
G532K	ACCG	9.8
G532K	CCCC	9.69
G532L	CTTC	15.2
G532L	TTTT	14.33
G532L	TTTA	12.17
G532L	CTTG	12.04
G532L	CTTA	11.73
G532L	TTGC	11.67
G532L	TTCA	11.66
G532L	TTAC	11.57
G532L	TTTG	11.5
G532L	ATTA	11.36
G532L	TTTC	11.15
G532L	TTCC	10.89
G532L	ATTG	10.02
G532L	GTTG	9.71
G532L	TCTA	9.46
G532L	CCTA	9.29
G532L	ATTC	8.98
G532L	TCCG	8.77
G532L	TTGA	8.33
G532N	CTTG	15.59
G532N	TCCC	15.09
G532N	TTCG	14.11
G532N	ATTG	13.73
G532N	CTTC	12.97
G532N	TTGC	12.7
G532N	TTAC	12.61
G532N	TTCT	12.13
G532N	ATTA	12.03
G532N	TCTA	12.02
G532N	TTCC	12
G532N	TTTT	11.95
G532N	TCTC	11.91
G532N	ATTG	11.89
G532N	CTTA	11.8
G532N	TTTC	11.52
G532N	TTCA	11.33
G532N	TCTG	11.3
G532N	TTTA	11.23
G532N	TTTG	11.18

G532N	TCCG	11.17
G532N	GTTG	11.14
G532N	GTTA	10.53
G532N	CCCG	10.51
G532N	GTTC	9.62
G532N	CTCC	9.48
G532N	CCCC	9.46
G532N	TCCA	9.45
G532N	CCTA	9.35
G532N	CCTG	9.28
G532Q	TTTG	16.46
G532Q	TTGC	14.95
G532Q	TTCA	14.64
G532Q	CTTC	13
G532Q	TTTT	12.79
G532Q	ATTA	12.41
G532Q	TTCC	12.4
G532Q	TTAC	11.52
G532Q	TTCT	11.34
G532Q	CTTG	11.2
G532Q	TCTA	11.19
G532Q	TTTA	10.72
G532Q	CTTA	10.58
G532Q	ATTG	10.43
G532Q	TTTC	10.32
G532Q	TTCC	10.23
G532Q	TCCC	10.05
G532Q	GTTG	9.8
G532Q	CCCC	9.74
G532Q	TCTC	9.47
G532S	TTAC	13.85
G532S	TTTT	13.71
G532S	GTTA	13.01
G532S	TCTC	12.51
G532S	TTCA	12.5
G532S	CTTA	12.33
G532S	TCCC	12.3
G532S	ATTG	12.3
G532S	CTTC	12.28
G532S	ATTC	12.08
G532S	TTCC	11.95
G532S	TCTA	11.79
G532S	TCTG	11.78
G532S	TCCG	11.58
G532S	ATTA	11.51
G532S	TTTC	11.49
G532S	CTTG	11.43
G532S	GTTC	11.1
G532S	TTCG	11.03
G532S	GTTG	10.72
G532S	TTTG	10.63
G532S	TTTA	10.56
G532S	TTCT	10.36
G532S	TTGC	10.15
G532S	TCGC	9.46
G532S	TTAG	9.38
G532S	CCCC	9.11
G532S	CTCC	8.12
K116N	CTTA	12.52
K116N	CTTC	11.57
K116N	TCCC	11.05
K116N	TTAC	10.99
K116N	TCTA	10.97
K116N	TTCT	10.9
K116N	TCCA	10.84

K116N	TCTG	10.77
K116N	TCCG	10.73
K116N	TTCC	10.7
K116N	TTCA	10.66
K116N	ATTA	10.47
K116N	TTTT	10.21
K116N	TTGC	10.17
K116N	TTTC	9.92
K116N	CTTG	9.87
K116N	TTCC	9.85
K116N	TTCT	9.64
K116N	CCTA	9.42
K116N	GTTA	9.31
K116N	CCTG	9.21
K116N	CCCC	9.13
K116N	TCGC	9.04
K116N	GTTC	8.91
K116N	CTCA	8.61
K116N	CCGC	8.39
K116N	TTGA	7.69
K116R	TCCG	13.47
K116R	ATTC	13.02
K116R	TTAC	12.37
K116R	GTTC	11.83
K116R	CCTA	11.44
K116R	TTCT	11.35
K116R	ATTG	11.11
K116R	TCTA	11.03
K116R	CCCC	11.03
K116R	CTCC	10.97
K116R	TCTC	10.84
K116R	TCGC	10.29
K116R	GTTG	10.27
K116R	GCTC	10.2
K116R	CTTA	10.15
K116R	TTGC	10.12
K116R	CCTG	10.04
K116R	ATTA	10.02
K116R	TTTT	9.98
K116R	CCTC	9.78
K116R	TCCC	9.72
K116R	GTTA	9.65
K116R	CCCG	9.62
K116R	ACTA	9.61
K116R	TTCC	9.54
K116R	TCCA	9.36
K116R	GCGC	9.27
K116R	TTCC	9.22
K116R	TTCA	9.14
K116R	GCTA	9.14
K116R	CCCA	8.96
K116R	CTTC	8.95
K116R	TCTG	8.93
K116R	TTGG	8.82
K116R	CCGC	8.81
K116R	GCTG	8.67
K116R	ACGC	8.32
K116R	ACTC	8.29
K116R	CTCA	8.06
K120H	TCTC	10.27
K120H	TTCC	10.07
K120H	TTTT	9.97
K120H	CTTG	9.77
K120H	CTTA	9.69
K120H	TTCC	9.3

K120H	TTCA	9.02
K120H	TTTG	8.5
K120H	TCGC	8.43
K120H	TTTC	8.26
K120H	TTTA	8.02
K120H	TTGC	7.89
K120H	TCCC	7.56
K120H	CCCC	7.55
K120N	TTTA	12.31
K120N	TTTT	11.86
K120N	CCCC	11.41
K120N	TCTC	11.4
K120N	TTCC	11.21
K120N	TCCC	11.18
K120N	CTTA	10.93
K120N	TTTC	10.91
K120N	TTGC	10.55
K120N	TCTG	10.41
K120N	ATTA	10.4
K120N	CTTC	10.32
K120N	ATTC	10.31
K120N	TTTG	9.99
K120N	TTCA	9.95
K120N	TCTA	9.86
K120N	CCCG	9.64
K120N	CTTG	9.6
K120N	TCCG	9.49
K120N	TTCC	9.48
K120N	TTCT	9.32
K120N	GCGC	9.07
K120N	CCTA	8.43
K120N	CTCC	8.18
K120Q	TTTC	12.01
K120Q	TTCC	11.56
K120Q	TTCA	11.1
K120Q	CTTG	10.57
K120Q	TTGC	10.37
K120Q	TTTG	10.25
K120Q	TTCC	10.25
K120Q	TTTT	9.91
K120Q	CTTC	9.83
K120Q	TTTA	9.72
K120Q	ATTG	9.25
K120Q	TTAC	9.25
K120Q	TCTA	9.18
K120Q	GTTC	9.03
K120Q	CTTA	8.96
K120Q	ATTC	8.71
K120R	TTCA	10.74
K120R	CTTA	9.7
K120T	TTTT	7.73
K120T	TTCC	7.73
K120T	TTGA	7.34
K121D	TTTA	12.69
K121D	TTTG	11.07
K121D	TTTC	10.18
K121G	TTTT	16.86
K121G	TTCA	15.35
K121G	TTCC	15.22
K121G	CTTG	14.67
K121G	TTGC	14.52
K121G	CTTA	14.02
K121G	TTTC	13.96
K121G	TTTA	13.42
K121G	CTTC	13.37

K121G	TTCG	13.33
K121G	TTTG	12.87
K121G	TTCT	12.72
K121G	ATTG	12.35
K121G	ATTC	11.96
K121G	ATTA	11.89
K121G	TTAC	11.65
K121G	TCCC	11.12
K121G	TCTA	10.53
K121G	TCGC	9.78
K121G	GTGA	9.62
K121G	GTTG	9.32
K121G	TCTC	9.08
K121G	CCCC	8.06
K121G	GTTT	7.89
K121H	TTCA	10.11
K121H	TTGC	9.93
K121H	TCTA	9.6
K121H	TTCC	9.53
K121H	CTTG	9.11
K121H	TTTT	8.99
K121H	TTCG	8.82
K121H	TTTC	8.53
K121H	TTTA	8.28
K121H	CTTC	7.95
K121H	TCCG	7.77
K121H	CCTG	7.66
K121H	TCTC	7.51
K121Q	TTTA	8.76
K121Q	TTTG	8.76
K121Q	TTTT	8.28
K121Q	TTCT	7.93
K121Q	TTCC	7.89
K121Q	TTCT	14.44
K121R	CTTC	13.38
K121R	TTCA	13.05
K121R	TTTT	13.03
K121R	CCTG	12.96
K121R	ATTG	12.3
K121R	TTAC	12.28
K121R	CCCC	12.28
K121R	TCGC	12.01
K121R	ATTA	12
K121R	TCCG	11.99
K121R	TCTA	11.97
K121R	TTCG	11.79
K121R	TCTC	11.72
K121R	TTTG	11.55
K121R	CCGC	11.31
K121R	GTGA	11.23
K121R	GTTG	11.05
K121R	TTTG	10.86
K121R	ATTC	10.84
K121R	CTTA	10.83
K121R	CCTA	10.71
K121R	CCTC	10.52
K121R	TTCC	10.4
K121R	CCCG	10.37
K121R	CTCC	10.21
K121R	TCCA	10.18
K121R	TTTA	10.07
K121R	TCCC	9.9
K121R	CTTG	9.72
K121R	GCGC	9.67
K121R	TCAC	9.59

K121R	CCCA	9.45
K121R	GTTC	9.23
K121R	CTCA	9.04
K121R	TTTC	8.93
K121R	TCTG	8.87
K121S	TTGC	11.77
K121S	CCTG	11.74
K121S	TCCA	11.66
K121S	TTTT	10.88
K121S	TTCA	10.85
K121S	CCTC	10.72
K121S	TCTC	10.62
K121S	CTTA	10.33
K121S	TTCT	10.1
K121S	TCTA	10.06
K121S	CCCC	10.02
K121S	TTCC	9.9
K121S	CTTG	9.88
K121S	TCCC	9.81
K121S	TTCG	9.71
K121S	CCTC	9.65
K121S	TCCG	9.59
K121S	CTTC	9.59
K121S	ATTC	9.53
K121S	TCTG	9.47
K121S	TTAC	9.46
K121S	TTTA	9.31
K121S	GCGC	9.26
K121S	GTTC	9.25
K121S	ATTA	9.2
K121S	CCGC	9.14
K121S	ACTA	8.94
K121S	GCCT	8.9
K121S	TTTC	8.89
K121S	CTTT	8.79
K121S	TTGA	8.76
K121S	TCGC	8.63
K121S	CCCG	8.61
K121S	GTGA	8.5
K121S	GTTG	8.35
K121S	ACTG	7.89
K121S	CTCG	7.8
K121S	GCTA	7.57
K121T	TTTC	13.85
K121T	TTTT	13.26
K121T	TTTA	13.02
K121T	TTTG	12.09
K121T	CTTA	11.65
K121T	CTTC	10.99
K121T	CTTG	10.52
K121T	TTCA	10.44
K121T	TTCC	10.36
K121T	TTGC	9.85
K538C	CTTC	9.67
K538F	TTTA	17.03
K538F	TTTC	15.83
K538F	CTTG	11.01
K538F	CTTC	10.62
K538F	CTTA	9.14
K538G	TTTC	11.42
K538G	TTTA	10.22
K538G	TTTG	9.79
K538G	CTTC	9.18
K538H	TTTG	11.31
K538H	CTTC	10.74

K538L	TTTC	10.95
K538L	TTTA	10
K538L	CTTA	9.31
K538M	TTTG	13.69
K538M	TTTA	13.09
K538M	TTTC	11.36
K538M	CTTC	10
K538M	CTTA	9.97
K538Q	CTTC	13.23
K538Q	TTTC	12.93
K538Q	TTTA	12.91
K538Q	CTTA	11.77
K538Q	TTTG	11.17
K538Q	CTTG	11.03
K538R	TTTG	12.33
K538R	TTTC	12.28
K538R	TTTA	12.26
K538R	CTTA	11.95
K538R	TTTT	9.32
K538R	CTTG	9.16
K538V	CTTC	18.4
K538V	CTTG	18.05
K538V	TTTA	17.53
K538V	TTTG	17.49
K538V	TTTC	15.98
K538V	CTTA	14.13
K538W	TTTC	11.69
K538W	TTTG	11.18
K538W	TTTA	11.03
K538W	TATA	10.38
K538W	CTTC	10
K538W	TATC	9.66
K538W	CTTA	9.26
K538Y	CTTG	9.14
K538Y	CTTA	8.3
K538Y	CTTC	8.08
K591A	TTTG	8.62
K591F	TTTA	13.44
K591F	TTTG	12.35
K591F	TTTC	11.17
K591F	TTTT	10.9
K591G	TTTA	11.71
K591G	TTTC	11.25
K591G	TTTT	10.92
K591G	TTTG	10.78
K591H	TTTC	16.42
K591H	TTTA	14.07
K591H	TTTG	14.02
K591H	TTTT	12.41
K591R	TTTG	10.31
K591R	TTCC	9.81
K591R	TTTC	9.07
K591R	CTTG	8.61
K591S	TTTT	11.24
K591S	TTCC	9.81
K591S	TTTA	9.37
K591S	TTTC	9.23
K591S	CTTG	8.98
K591S	TTTG	8.89
K591S	TTCC	8.04
K591W	TTTA	13.93
K591W	TTTC	11.17
K591W	TTTG	11.14
K591Y	TTTC	12.29
K591Y	TTTA	11.4

K591Y	TTTG	10.79
K595H	TTGC	12.45
K595H	TTTT	11.42
K595H	TTTG	10.91
K595H	TTCA	10.39
K595H	TTTC	10.21
K595H	TTCG	9.36
K595L	TTTA	16.84
K595L	TTCC	15.92
K595L	TTTG	14.92
K595L	CCTG	13.79
K595L	TTCG	13.45
K595L	TTTC	13.12
K595L	TCTA	12.6
K595L	TCTC	12.15
K595L	TCCA	12.14
K595L	TCTG	11.7
K595L	ACTC	11.46
K595L	TCCG	11.37
K595L	CCTC	10.6
K595L	TTCA	10.5
K595L	TCCC	10.24
K595L	ACTG	9.69
K595M	TTGC	13.18
K595M	TTTA	12.88
K595M	CCTG	12.36
K595M	TTTG	12.07
K595M	TCCC	11.4
K595M	TTTT	11.26
K595M	CCTC	10.96
K595M	TCTG	10.95
K595M	TTTC	10.84
K595M	TTCA	10.81
K595M	ACTG	10.74
K595M	TCTA	10.34
K595M	TCTT	10.1
K595M	TCCG	9.83
K595M	ACGC	9.6
K595M	TTCC	9.51
K595M	CCCC	9.4
K595M	TCGG	9.21
K595M	GCTG	9.12
K595M	CTTC	8.62
K595M	CTTA	8.14
K595Q	TCCG	15.13
K595Q	TTCC	14.56
K595Q	TTCA	13.57
K595Q	TCTG	13.57
K595Q	TCTC	12.93
K595Q	TTTG	12.45
K595Q	TTTA	12.22
K595Q	TCTA	12.06
K595Q	TTCG	11.85
K595Q	TTTC	11.69
K595Q	TCCC	10.62
K595Q	ACCC	10.09
K595Q	CCCC	9.21
K595R	TCTT	8.5
K595R	TCCC	8
K595S	TCCC	14.55
K595S	TTGC	14.34
K595S	TCGC	14.23
K595S	TTTA	13.22
K595S	TCTC	12.48
K595S	TCTA	12.47

K595S	TCCG	12.3
K595S	CCGC	12.29
K595S	ACGC	11.62
K595S	TTCC	11.53
K595S	TTTC	10.96
K595S	TCCG	10.45
K595S	CCCC	10.13
K595S	TTCA	9.48
K595S	TTCG	8.92
K595W	CCCC	11.92
K595W	TCCA	11.62
K595W	ACGG	11.48
K595W	ACCC	11.35
K595W	TTTG	11.35
K595W	TCTC	11.28
K595W	TTTT	10.53
K595W	GCCG	10.42
K595W	TTTA	10.21
K595W	ACTA	10.18
K595W	ACGA	9.87
K595W	CCGC	9.81
K595W	ATGC	9.77
K595W	TCGT	9.71
K595W	ATTG	9.63
K595W	TTCT	9.25
K595W	CCGG	9.19
K595W	TTGT	9.08
K595W	TCCG	9.03
K595W	TTCT	8.95
K595W	ACTG	8.88
K595Y	ACCC	18.7
K595Y	TCCC	18.62
K595Y	ACCG	18.58
K595Y	CCCG	18.42
K595Y	TTCG	17.75
K595Y	GCGC	17.47
K595Y	TCCA	16.89
K595Y	CCGC	16.66
K595Y	CCGG	16.59
K595Y	GCCG	16.46
K595Y	TCCG	16.23
K595Y	CCCC	16.05
K595Y	TTTC	15.82
K595Y	TCCA	14.88
K595Y	TCCG	14.74
K595Y	TTTA	14.31
K595Y	CCGA	14.29
K595Y	TTTG	14.12
K595Y	CTGC	14.03
K595Y	TCGT	14.01
K595Y	ACGC	13.83
K595Y	ACGA	13.62
K595Y	ACGG	13.55
K595Y	TCTG	13.36
K595Y	TCTC	13.31
K595Y	TCCG	13.26
K595Y	GCCG	13.19
K595Y	GCCC	13.15
K595Y	TTGA	12.32
K595Y	TTCT	12.28
K595Y	TTCA	12.22
K595Y	TTGT	12.12
K595Y	TTGG	12.01
K595Y	TCTA	11.81
K595Y	GCGA	11.57

K595Y	CCCA	11.52
K595Y	CTCG	11.51
K595Y	TTGC	11.46
K595Y	TTCC	11.3
K595Y	ACCA	11.3
K595Y	ACTG	11.29
K595Y	CCTG	10.54
K595Y	CTCC	10.3
K595Y	ACTC	9.54
K595Y	ATGC	9.28
K600G	ATTG	14.1
K600G	TTTT	13.53
K600G	TCTC	13.48
K600G	CTTG	13.02
K600G	TTTA	12.84
K600G	CTTC	12.82
K600G	TCTG	12.69
K600G	ATTA	12.52
K600G	TCTA	12.41
K600G	TTCA	12.15
K600G	TTTC	12.11
K600G	CTTA	12.01
K600G	TTGC	11.67
K600G	TTCC	11.66
K600G	TTTG	11.58
K600G	TTAC	11.52
K600G	TCCC	11.4
K600G	CCTG	11.38
K600G	CCCC	10
K600G	GTTA	9.75
K600G	GTTG	9.72
K600G	TCCG	9.69
K600G	TTCT	9.48
K600G	CCTA	9.4
K600G	TCCA	9.38
K600H	CTTC	16.65
K600H	TTTG	14.45
K600H	TTCC	14.11
K600H	TTTT	13.78
K600H	TTCA	13.27
K600H	CTTA	12.42
K600H	TCTA	12.26
K600H	TCTC	12.2
K600H	TCCC	12.16
K600H	ATTG	12.02
K600H	ATTA	11.77
K600H	CTTG	11.23
K600H	TTCG	11.13
K600H	TTTC	10.82
K600H	TTAC	10.54
K600H	GTTC	9.1
K600H	GTTG	8.65
K600H	CCGC	8.41
K600R	ATTG	13.41
K600R	TTGC	12.52
K600R	TTCT	11.83
K600R	TTCC	11.61
K600R	TTCA	11.08
K600R	TTTT	11.07
K600R	CCCC	11.05
K600R	TCGC	11.02
K600R	TCTA	10.96
K600R	GTTC	10.91
K600R	TCCA	10.84
K600R	TCCC	10.63

K600R	CCTC	10.42
K600R	TTCG	10.1
K600R	CTCA	8.96
K601H	CCTG	14.33
K601H	CTTG	12.41
K601H	TCTA	12.13
K601H	TCTC	11.97
K601H	ATTC	11.89
K601H	TCCG	11.84
K601H	TTTG	11.78
K601H	CCCG	11.7
K601H	TTTT	11.35
K601H	CCTA	11.15
K601H	TTGC	10.89
K601H	CTTC	10.79
K601H	TTTC	10.65
K601H	CTTA	10.45
K601H	CCCC	10.42
K601H	TTCG	10.39
K601H	ATTG	10.26
K601H	TTTA	10.21
K601H	ATTA	10.2
K601H	TTCA	10.12
K601H	TCTG	10.09
K601H	CCTC	9.82
K601H	TCGC	9.4
K601H	GCTC	8.68
K601Q	CTTA	13.97
K601Q	TTTG	11.17
K601Q	TTTC	10.17
K601Q	CTTG	9.73
K601Q	TTTT	9.44
K601R	TCTA	12.21
K601R	TTGC	12.09
K601R	ATTC	11.08
K601R	TTTT	10.98
K601R	TTCC	10.95
K601R	TCTG	10.8
K601R	TTCA	10.77
K601R	CCTA	10.4
K601R	ATTG	10.33
K601R	CTTG	10.31
K601R	CTTC	10.11
K601R	TCTC	9.85
K601R	TCCG	9.65
K601R	ATTA	9.5
K601R	TCCA	8.91
K601R	CCTG	8.54
K601R	TTTC	8.41
K601T	CTTA	12.42
K601T	TTTC	11.55
K601T	TTCC	11.36
K601T	TTGC	11.26
K601T	TTTT	11.13
K601T	CTTC	11.06
K601T	TTTA	10.99
K601T	TTCA	10.85
K601T	TTTG	10.2
K601T	TCCC	10.17
K601T	TCCG	9.65
K601T	CCTG	9.53
K601T	TCTC	9.29
K601T	CCCC	8.18
L585F	TTTG	13.66
L585F	TTTT	11.79

L585F	TTTA	11.72
L585F	TTTC	10.51
L585G	TTCC	13.33
L585G	TTAC	11.98
L585G	CTTG	11.26
L585G	TTTT	11.21
L585G	TCCC	11.09
L585G	TTCA	10.97
L585G	CTTC	10.32
L585G	TTTA	9.92
L585G	TCTC	9.8
L585G	TTGC	9.57
L585G	CTTA	9.48
L585G	TCAC	9.48
L585G	TTTG	9.02
L585G	TTTC	8.8
L585H	TTCC	9.49
L585H	TTTT	8.29
L585H	TCTC	8.06
M592A	TTTG	15.81
M592A	TTTA	11.41
M592A	TTTC	10.98
M592A	CTTA	10.8
M592A	TTCA	10.53
M592A	TCTA	10.38
M592A	TTTT	9.48
M592E	TTTA	12.63
M592E	TTTG	11.85
M592E	TTTC	11.64
M592Q	TTCA	12.54
M592Q	TTTA	10.68
M592Q	TCCG	10.66
M592Q	CTTG	10.19
M592Q	TTCG	9.93
M592Q	CTTA	9.45
M592Q	CCCG	9.41
M592Q	TCTG	9.39
M592Q	TTTG	9.37
M592Q	CCCA	9.22
M592Q	TCCC	9.17
M592Q	TTCC	9.16
M592Q	CCCC	8.93
M592Q	TTTC	8.89
M592Q	CTCA	8.41
Q529A	TCTG	13.89
Q529A	CCCC	13.38
Q529A	TTCC	12.5
Q529A	ATTG	12.31
Q529A	TCCA	12.26
Q529A	TCTC	12.21
Q529A	TTGC	11.88
Q529A	CCCG	11.81
Q529A	TTTT	11.69
Q529A	TTCG	11.43
Q529A	TCTA	11.31
Q529A	CTTC	11.2
Q529A	ATTG	10.94
Q529A	CCTA	10.85
Q529A	ATTA	10.79
Q529A	TTTA	10.73
Q529A	CTTA	10.63
Q529A	CCTG	10.54
Q529A	CTTG	10.47
Q529A	TCCG	10.4
Q529A	CCTC	10.38

Q529A	TCCC	10.18
Q529A	TTCA	10.16
Q529A	TTCT	9.89
Q529A	CCGC	9.68
Q529A	TTTG	9.62
Q529A	TTTC	9.51
Q529A	GTTA	9.4
Q529D	TTTC	13.64
Q529D	TTTT	13.03
Q529D	TTTG	12.45
Q529D	TTCC	12.36
Q529D	CTTC	12.27
Q529D	TTTA	12.24
Q529D	TTGC	10.98
Q529D	CTTG	10.85
Q529D	CTTA	10.81
Q529D	TTCA	10.36
Q529D	TCCC	8.52
Q529F	CTTA	14.01
Q529F	CTTG	13.75
Q529F	TTTG	12.84
Q529F	TTTA	11.62
Q529F	TTTC	10.91
Q529G	CTTC	13.9
Q529G	TTCA	11.91
Q529G	TCTA	11.71
Q529G	CCCC	11.18
Q529G	TTTA	10.86
Q529G	TCTG	10.79
Q529G	TTTC	10.63
Q529G	CTTA	10.61
Q529G	TTTG	10.19
Q529G	CTTG	9.88
Q529G	CCTA	9.67
Q529G	TCCC	9.5
Q529G	TTTG	9.15
Q529G	TCTC	9.11
Q529G	ATTG	8.98
Q529G	CTCG	8.96
Q529G	ATTA	8.33
Q529G	TCGC	8.2
Q529G	TCCG	8.11
Q529H	TTTC	11.09
Q529H	CTTG	10.43
Q529H	TTTA	10.39
Q529H	TTCC	10.29
Q529H	TTCG	10.29
Q529H	CTTC	9.59
Q529H	ATTG	9.52
Q529H	TTTG	9.12
Q529H	TTCA	9.07
Q529H	ATTA	8.47
Q529N	TCGC	13.73
Q529N	TCTC	13.48
Q529N	TCCG	12.62
Q529N	TTTT	12.48
Q529N	CTTG	12.38
Q529N	TTGC	12.3
Q529N	TTCC	12.2
Q529N	TCCA	11.98
Q529N	TCCC	11.38
Q529N	TCTA	11.04
Q529N	TTCA	10.89
Q529N	TTCG	10.81
Q529N	ATTG	10.79

Q529N	CCCC	10.6
Q529N	TTTA	10.57
Q529N	CTTG	10.24
Q529N	CTTA	10.18
Q529N	TTCT	9.91
Q529N	CTTC	9.87
Q529N	ATTA	9.85
Q529N	TTTC	9.83
Q529N	CCTA	9.49
Q529N	GTTA	9.32
Q529N	GTTC	9.18
Q529N	TCTT	8.71
Q529N	TCAC	8.3
Q529N	CCCG	8.23
Q529P	TTTA	14.13
Q529P	TTTG	12.06
Q529P	TTTC	10.87
Q529S	TTCA	11.88
Q529S	CTTA	11.14
Q529S	TTTG	11.09
Q529S	TTTA	11.06
Q529S	TCTA	10.17
Q529S	TTTC	10.1
Q529S	CCCC	10.1
Q529S	TTTT	10.03
Q529S	TTCC	10.02
Q529S	CTTG	9.9
Q529S	TCCC	9.77
Q529S	TCTG	9.68
Q529S	ATTA	8.81
Q529T	CTTG	15.39
Q529T	TTCA	14.62
Q529T	TCTG	13.74
Q529T	TCTC	13.63
Q529T	TCTA	13.2
Q529T	TTCC	13.17
Q529T	TTTT	12.6
Q529T	TCCG	12.44
Q529T	TCCA	12.34
Q529T	ATTA	12.15
Q529T	CTTC	12.13
Q529T	TTCG	11.89
Q529T	TCCC	11.35
Q529T	TTGC	11.17
Q529T	CCTA	11.07
Q529T	GTTG	10.96
Q529T	ATTG	10.9
Q529T	TTTC	10.88
Q529T	TTTA	10.73
Q529T	TTTG	10.51
Q529T	ATTC	10.41
Q529T	CTTA	10.37
Q529T	CCTC	10.12
Q529T	TCGC	10.06
Q529T	GTTA	9.83
Q529T	CCCC	9.78
Q529T	GTTC	9.72
Q529T	CCTG	9.3
Q529T	TTCT	8.86
Q529T	TTAC	7.99
Q529W	TTTG	15.75
Q529W	CTTA	14.5
Q529W	TTTA	12.69
Q529W	TTTC	11.31
Q529W	CTTC	9.28

S599G	CTTG	12.09
S599G	TTTT	11.66
S599G	TTTA	10.2
S599G	TTCC	9.87
S599G	CCCC	9.47
S599G	TTCA	9.36
S599H	TTTA	10.82
S599H	TTTG	9.61
S599N	CTTA	13.93
S599N	TTTG	13.04
S599N	TTTA	11.64
S599N	TTTC	11.3
S599N	CTTC	11.15
S599N	CTTG	10.54
S599N	TTCA	8.73
T148A	TTTC	13.06
T148A	TTTA	11.98
T148A	TTTT	11.53
T148A	TTGC	11.48
T148A	TTTG	11.33
T148A	TTCC	10.97
T148A	CTTG	9.83
T148A	TTCA	9.26
T148C	TTCA	15.44
T148C	TTTG	15.17
T148C	TTTC	15.09
T148C	CTTG	14.75
T148C	CTTA	13.98
T148C	CTTC	13.67
T148C	TTTT	13.43
T148C	TTTA	13.16
T148C	TTCC	12.16
T148C	TTGC	11.1
T148C	TCTA	10.77
T148C	GTTG	10.01
T148C	TCTC	9.93
T148H	TTTC	14.35
T148H	TTTG	13.68
T148H	TTTA	13.64
T148S	TTTT	13.35
T148S	CTTA	12.93
T148S	TTTA	12.48
T148S	TTCG	12.31
T148S	TTCA	11.85
T148S	TTGC	11.56
T148S	TTTG	11.55
T148S	TTCC	11.39
T148S	TTTC	11.01
T148S	CTTC	11.01
T148S	CTTG	10.8
T148S	ATTA	9.58
T148S	TCTA	8.54
T148S	ATTG	8.31
T149C	TTTC	11.35
T149C	TTTA	11.08
T149C	TTTG	10.49
T149C	CTTA	10.35
T149C	TTTT	9.98
T149F	TTTC	13.56
T149F	TTTG	13.11
T149F	TTTA	11.72
T149G	TTTG	12.06
T149G	TTTA	10.59
T149G	CTTG	10.14
T149G	TCCC	10.11

T149G	GTTG	10.08
T149G	TTTC	9.96
T149G	TTCA	9.78
T149G	TTTT	9.77
T149G	CCCC	9.29
T149G	ATTG	9.26
T149G	TTCC	9.2
T149G	CTTC	8.83
T149G	TCTA	8.83
T149G	ATTG	8.6
T149G	TCCG	8.58
T149G	CCCG	8.57
T149G	TCTC	8.45
T149H	TTTA	15.77
T149H	TTTC	11.31
T149N	TTTA	13.97
T149N	TTTC	13.27
T149N	TTTG	10.86
T149P	TTTC	10.27
T149P	TTTA	10.14
T149P	TTTG	9.22
T149S	TTCC	14.11
T149S	CTTC	14.01
T149S	ATTG	13.88
T149S	TCTG	13.24
T149S	TTCA	12.74
T149S	ATTG	12.36
T149S	CTTA	12.32
T149S	TTTT	12.24
T149S	CTTG	11.87
T149S	CTCG	11.55
T149S	ATTA	11.51
T149S	TCTC	11.38
T149S	TTCG	11.34
T149S	TCTA	11.17
T149S	GTTA	10.66
T149S	GTTG	10.45
T149S	GTTG	10.44
T149S	TCGC	10.43
T149S	CCCC	10.4
T149S	TCCG	10.35
T149S	TTGC	10.2
T149S	TCCA	9.82
T149S	CCTA	9.7
T149S	TTTC	9.68
T149S	TCCC	9.33
T149S	TTCT	9.18
T149S	TTTA	9.08
T149S	CCGC	8.88
T149S	CCCG	8.63
T149S	CCTC	8.45
T149V	TTCA	13.47
T149V	TTTA	13.34
T149V	TTCC	12.51
T149V	TTTG	12.3
T149V	TTTT	11.86
T149V	TTCG	11.21
T149V	TTTC	10.58
T149V	TCCC	9.41
T149V	CTTG	8.81
T149V	TCTC	8.75
T152E	CCTC	18.07
T152E	CTTC	16.54
T152E	TTGC	15.69
T152E	TTCA	14.71

T152E	TCTG	14.5
T152E	TCCC	14.43
T152E	TCGC	14.25
T152E	TCTC	14.21
T152E	GTTC	14.11
T152E	TTTC	14.1
T152E	TTTT	14.1
T152E	ATTC	13.83
T152E	CTTG	13.72
T152E	TTTA	13.71
T152E	TCCG	13.52
T152E	CCCC	12.98
T152E	TTCC	12.87
T152E	TTCC	12.63
T152E	ACTC	12.45
T152E	CCGC	12.05
T152E	GCGC	12.01
T152E	CTTA	11.75
T152E	TTTG	11.67
T152E	TTCT	11.58
T152E	TCAC	10.76
T152E	CCAC	10.72
T152E	GTTC	10.4
T152F	TTCC	16.18
T152F	CTTC	15.86
T152F	CTTA	15.51
T152F	TTTC	14.6
T152F	TTGC	14.54
T152F	TTCA	14.36
T152F	TTTT	13.43
T152F	TTTG	13.17
T152F	CTTG	12.61
T152F	ATTA	12.5
T152F	TCTC	11.97
T152F	TTCC	11.93
T152F	TCCC	11.76
T152F	TTTA	11.14
T152F	TCGC	10.44
T152F	CCCC	9.98
T152H	TCCC	12.04
T152H	TTGC	11.41
T152H	CCCC	11.3
T152H	TTTT	11.27
T152H	CTTA	11.1
T152H	TTCC	11.06
T152H	TCTC	10.51
T152H	CTTG	10.27
T152H	TTTC	10.23
T152H	TTTA	10.06
T152H	TCGC	10.02
T152H	ATTC	10.01
T152H	TTTG	9.94
T152H	TCTG	9.88
T152H	CCTC	9.85
T152H	CTTC	9.81
T152H	TCCG	9.76
T152H	TTCA	9.71
T152H	ATTA	9.64
T152H	TCTA	9.63
T152H	TTCT	9.34
T152H	TCCA	9.05
T152H	TTAC	8.95
T152H	ATTG	8.74
T152H	TTCC	8.46
T152K	CTTG	14.08

T152K	CCGC	13.88
T152K	CCTA	13.58
T152K	ATTA	13.28
T152K	ACTC	13.28
T152K	TCCC	13.26
T152K	GTTC	13.22
T152K	GCTC	13.21
T152K	CCCC	12.99
T152K	TTCT	12.84
T152K	TCAC	12.63
T152K	GCTA	12.51
T152K	CCTG	12.5
T152K	CTTA	12.46
T152K	GTTC	12.46
T152K	ATTG	12.29
T152K	ATTC	12.24
T152K	TCTC	12.23
T152K	TCTG	12.2
T152K	CTTC	12.12
T152K	TTCA	11.99
T152K	TTCC	11.88
T152K	CCTC	11.81
T152K	CCAC	11.76
T152K	ACTG	11.5
T152K	TCTA	11.44
T152K	GCGC	11.41
T152K	TCGC	11.15
T152K	TCCA	11.08
T152K	GTTC	10.94
T152K	TTAC	10.94
T152K	CTCC	10.78
T152K	TTTA	10.44
T152K	TCGA	10.16
T152K	ACTA	10.14
T152K	TTCC	10.09
T152K	GCCC	9.97
T152K	TTTT	9.85
T152K	ACGC	9.71
T152K	TTGC	9.68
T152K	GCTG	9.59
T152K	TCCG	9.4
T152K	TTGA	9.39
T152K	CCCC	9.34
T152K	ACCC	9.3
T152K	TTTG	9.21
T152K	CTTT	8.75
T152L	CTTA	15.37
T152L	ATTC	14.54
T152L	TCGC	13.95
T152L	TCTA	13.76
T152L	CTTG	13.75
T152L	CTTC	13.54
T152L	TCTC	13.15
T152L	CCCC	12.44
T152L	TTGC	12.23
T152L	TTCC	12.04
T152L	TTCC	11.39
T152L	ATTG	10.85
T152L	TTTT	10.5
T152L	CCTA	10.43
T152L	TTAC	10.32
T152L	GTTC	9.55
T152L	CCTC	9.38
T152L	TCTG	9.22
T152L	TTTA	8.96

T152L	CCTG	8.96
T152Q	TCGC	11.48
T152Q	CCTG	11.36
T152Q	CCTA	11.27
T152Q	CCCA	10.85
T152Q	TCAC	9.99
T152Q	CCGC	9.9
T152Q	ATTC	9.85
T152Q	TCTG	9.82
T152Q	CTTC	9.78
T152Q	TTGC	9.7
T152Q	TCTC	9.7
T152Q	GTTC	9.64
T152Q	TTGA	9.56
T152Q	TCTA	9.37
T152Q	ATTG	9.12
T152Q	TTAG	9.09
T152Q	TTCT	9.04
T152Q	ATTA	8.96
T152Q	TIGG	8.65
T152Q	TCCA	8.57
T152Q	TTTG	8.56
T152Q	ACTC	8.52
T152Q	GCCC	8.37
T152Q	TTTT	8.32
T152Q	CTTA	7.83
T152Q	TTCC	7.74
T152Q	TTCC	7.35
T152R	CTTC	16.05
T152R	CCTG	14.82
T152R	CCTC	13.67
T152R	GCCC	12.97
T152R	CCCC	12.95
T152R	GTTC	12.67
T152R	GCCG	12.64
T152R	ACTC	12.49
T152R	CCGC	12.38
T152R	CCGG	12.22
T152R	TCGA	12.16
T152R	CTGC	11.86
T152R	CCCG	11.78
T152R	TCGG	11.59
T152R	ATTA	11.55
T152R	CCCA	11.27
T152R	TCAG	11.2
T152R	TTCA	11.03
T152R	CTTA	10.87
T152R	CCTA	10.82
T152R	ATTG	10.8
T152R	TTCC	10.71
T152R	TCCG	10.7
T152R	ACGA	10.69
T152R	TCCA	10.66
T152R	ATTC	10.62
T152R	TCGT	10.58
T152R	GCAC	10.54
T152R	TTAC	10.46
T152R	TCTA	10.44
T152R	ACCC	10.44
T152R	GCTA	10.41
T152R	TTTT	10.4
T152R	GCCA	10.38
T152R	GTTC	10.31
T152R	ACTG	10.25
T152R	GCTG	10.24

T152R	GCGA	10.23
T152R	ACTA	10.21
T152R	TTGT	10.12
T152R	TTCT	10.07
T152R	GCTC	10.05
T152R	ACCG	10.01
T152R	CCGA	9.99
T152R	TCGC	9.92
T152R	GTTC	9.85
T152R	CTCG	9.82
T152R	GCCG	9.74
T152R	TTAG	9.7
T152R	TCCC	9.66
T152R	TCTC	9.58
T152R	ACCA	9.58
T152R	CTTG	9.53
T152R	TTGG	9.53
T152R	CTCA	9.17
T152R	TTGC	9.03
T152R	ACGG	8.22
T152W	TTGC	16.19
T152W	TTCA	14.29
T152W	TTCC	13.39
T152W	TTCC	13.36
T152W	TTTT	13.32
T152W	TCTA	12.94
T152W	TTTG	12.53
T152W	CTTC	12.39
T152W	CTTG	12.21
T152W	ATTA	12.11
T152W	TCCG	12.04
T152W	CCTG	11.72
T152W	CTTA	11.67
T152W	TTTC	11.21
T152W	TTTA	11.1
T152W	ATTG	11.02
T152W	CCTA	10.57
T152W	TCTC	10.48
T152W	TTAC	10.29
T152W	ATTC	9.95
T152W	GTTC	9.86
T152W	GTTC	9.45
T152Y	TCCA	16.58
T152Y	TCTG	15.9
T152Y	TCTC	15.89
T152Y	CCTC	15.47
T152Y	TCTA	15.43
T152Y	TTAC	15.31
T152Y	CCTG	15.3
T152Y	CTTG	15.16
T152Y	ACTG	14.71
T152Y	GCTA	14.52
T152Y	ATTA	14.36
T152Y	ACTC	14.29
T152Y	ATTG	14.24
T152Y	TTCC	14.05
T152Y	CCGC	13.93
T152Y	GCCC	13.92
T152Y	TCCG	13.65
T152Y	CTTC	13.65
T152Y	GCTC	13.61
T152Y	GTTC	13.55
T152Y	TCAC	13.42
T152Y	ATTC	13.38
T152Y	CTTA	13.2

T152Y	TTGC	13.03
T152Y	TCGC	12.87
T152Y	GTTA	12.47
T152Y	TCCC	12.23
T152Y	ACTA	12.07
T152Y	CCTA	12.04
T152Y	GCTG	11.97
T152Y	ACGC	11.96
T152Y	TTCA	11.9
T152Y	TTCC	11.9
T152Y	GTTC	11.89
T152Y	TCGG	11.8
T152Y	TTCT	11.68
T152Y	CCCC	11.49
T152Y	CCAC	11.49
T152Y	GCGC	11.42
T152Y	CCCA	11.39
T152Y	GCAC	11.15
T152Y	TTGA	10.7
T152Y	CTCC	10.58
T152Y	CCCG	10.42
T152Y	CTTT	10.32
T152Y	TTTT	10.31
T152Y	ACCC	10.09
T152Y	TTGG	9.8
T152Y	TTTC	9.37
T152Y	ACAC	8.65
V596H	TTTG	11.75
V596H	TTTC	11.08
V596T	CTTA	20.31
V596T	TTTG	16.49
V596T	TTCA	15.21
V596T	CTTG	14.59
V596T	TTTA	14.55
V596T	CTTC	14.43
V596T	TCTA	13.11
V596T	TTCC	13
V596T	TTGC	12.93
V596T	TCCC	12.65
V596T	TTTT	12.27
V596T	TCTC	12.27
V596T	TCTG	12.07
V596T	CCCC	12.03
V596T	TTTC	12.02
V596T	TTCG	10.87
V596T	CCCG	9.96
V596T	TCCG	9.6
V596T	GTTC	9.3
V596T	GCGC	9.17
W649H	TTTC	11.36
W649H	TTTA	11.18
W649H	TTTG	10.47
W649K	TTTA	13.34
W649R	TTTA	8.84
W649S	TTTA	9.77
W649Y	TTTC	11.78
W649Y	TTTG	11.48
W649Y	TTTA	11.05
Y542F	TTTA	11.24
Y542F	TTTC	10.25
Y542F	TTTG	9.9
Y542H	CTTA	11.46
Y542H	TTTT	11.44
Y542H	CTTC	10.96
Y542H	TTTC	10.6

Y542H	TCCC	10.37
Y542H	GTTA	10.23
Y542H	TTCG	10.16
Y542H	TTCA	10.06
Y542H	TTTG	10.04
Y542H	CCAC	8.42
Y542H	GCTC	8.27
Y542K	TTTA	10.62
Y542K	TTTG	9.4
Y542K	TTCA	8.45
Y542L	CTTA	12.93
Y542L	TTTA	10.46
Y542L	TTTC	10.2
Y542L	TTTG	10.04
Y542M	TTTG	11.54
Y542M	TTTA	11.15
Y542M	TTTC	10.32
Y542M	CTTG	10.17
Y542M	CTTA	10.05
Y542M	CTTC	8.67
Y542N	TTGC	11.39
Y542N	TTTA	11.38
Y542N	TTTG	11.27
Y542N	TTCC	11.06
Y542N	TTTT	10.31
Y542N	GTTC	10.16
Y542N	CCCC	9.87
Y542N	ATTA	9.39
Y542N	TCCC	9.37
Y542N	TTTC	9.21
Y542R	TTTT	10.21
Y542R	CTTG	8.75
Y542R	TTCA	8.73
Y542R	TCCC	8.68
Y542R	GTTA	7.7
Y542T	TTTA	11.46
Y542T	TTTC	9.68
Y542T	TTTG	9.59
Y542V	TTTG	12.03
Y542V	TTTA	10.76
Y616E	TTTG	12.92
Y616E	TTTA	12.58
Y616E	TTTC	12.02
Y616F	TTTT	12.56
Y616F	TTTA	12.36
Y616F	CTTA	12.33
Y616F	CTTG	11.97
Y616F	TTTG	11.74
Y616F	TTCA	11.5
Y616F	ATTG	11.46
Y616F	CTTC	11.34
Y616F	TTCC	11.06
Y616F	TCTA	10.99
Y616F	ATTA	10.97
Y616F	TTTC	10.79
Y616F	TTCT	10.63
Y616F	GTTG	10.46
Y616F	TTGC	9.95
Y616F	TTCG	9.74
Y616F	TCTG	9.65
Y616F	ATTC	9.45
Y616F	TTAC	8.77
Y616H	CTTA	12.23
Y616H	CTTG	11.73
Y616H	TTTC	11.26

Y616H	CTTC	11.18
Y616H	TTTT	11.11
Y616H	TTTA	11
Y616H	TTTG	11
Y616H	TTCA	10.45
Y616H	TTCC	10.28
Y616H	TTGC	10.16
Y616K	TTTG	9.23
Y616K	TTCC	8.66
Y616R	TTTG	10.22
Y616R	TTCC	9.77
Y646E	TTTA	13.43
Y646E	CTTA	12.79
Y646E	TTTT	11.99
Y646E	TTTG	11.63
Y646E	TTTC	11.28
Y646E	CTTC	11.26
Y646E	CTTG	10.71
Y646E	TTCC	9.96
Y646H	TTGC	13.8
Y646H	GTTC	12.39
Y646H	GTTA	12.24
Y646H	TCTA	11.57
Y646H	TCTC	11.45
Y646H	ATTA	10.92
Y646H	TTCT	10.69
Y646H	TCTG	10.57
Y646H	TCCC	10.3
Y646H	CTTC	10.25
Y646H	ATTG	10.23
Y646H	CTTG	10.13
Y646H	TTCC	9.81
Y646H	CCTG	9.69
Y646H	ATTC	9.61
Y646H	GTTG	9.13
Y646K	TTTT	10.71
Y646K	CTTA	9.82
Y646K	TTTC	9.2
Y646N	TTCA	10.45
Y646N	ATTA	10.38
Y646N	TTTA	8.91
Y646N	TCTA	8.61
Y646Q	ATTA	11.69
Y646Q	ATTG	11.67
Y646Q	TCTA	11.39
Y646Q	TTCC	11.24
Y646Q	CTTC	10.99
Y646Q	TTTT	10.98
Y646Q	TTCA	10.58
Y646Q	CTTA	9.82
Y646Q	TCTC	9.52
Y646Q	TTTG	9.27
Y646R	TTTC	10.17
Y646R	TCTA	9.17
Y646R	TTTA	8.09
Y646W	TTTT	10.69
Y646W	TTTC	9.78
Y646W	CTTG	9.6

[0241]

[0242]

본 발명자들은 2개의 LbCas12a 대조군을 사용하여 시험관내에서 절단된 36개의 고유한 PAM 서열을 관찰하였다. 이는 시험관내 wtLbCas12a가 단지 TTTV보다 더 많은 서열을 인식하고 절단할 수 있다는 관찰과 일치한다. TTCN, CTTN, TCTN, 및 다른 것들은 시험관내에서 LbCas12a에 의해 인식되고 절단되는 것으로 나타났으며, 여기서 AsCas12a는 단지 TTTN을 절단하는 것으로 나타났다 (Zetsche et al. *Cell* 163:759-771 (2015)).

[0243]

돌연변이체의 일부는 wtLbCas12a와 비교하여 시험관내에서 PAM 서열 인식 및 절단의 총 수를 증가시켰다 (표 5). 이는 절대 PAM 인식 서열에 대해서가 아니라, 개별적 돌연변이에 의해 부여된 전체적 난잡성을 말한다. 일부 개별적 점 돌연변이체는 야생형보다 더 난잡하였다. 예를 들어, T152R은 57개의 상이한 PAM을 인식하였고, K959Y는 45개를 인식하였다.

[0244]

표 5: 시험관내에서 인식되고 절단된 PAM의 수 (nPAM)에 의해 판단된 바와 같은 보다 난잡한 LbCas12a 돌연변이체의 일부

돌연변이체	nPAM
T152R	57
T152Y	50
T152K	47
D156Q	45
K595Y	45
D156L	43
D122H	39
K116R	39
D156K	38
K121S	38
K121R	37
야생형	36

[0245]

[0246]

야생형 LbCas12a에 대한 고갈의 비교

[0247]

시험관내 wtLbCas12a는 단지 TTTV보다 더 많은 서열을 인식하고 절단할 수 있다 (Zetsche et al. *Cell* 163:759-771 (2015)). TTCN, CTN, TCTN, 및 다른 것들은 시험관내에서 LbCas12a에 의해 인식되고 절단되는 것으로 나타났으며, 여기서 AsCas12a는 단지 TTTN을 절단하는 것으로 나타났다 (Zetsche et al. *Cell* 163:759-771 (2015)). 이 연구의 목표는 LbCas12a의 PAM 인식을 그의 야생형 능력 너머로 확장시키는 것이었다. 이를 위해, 본 발명자들은 라이브러리 고갈 점수 단독과는 상이한 분석을 채용하였다. 그들 점수는 시험관내에서 절대 PAM 인식 및 절단을 결정하는데 중요하지만, 도입된 점 돌연변이로 인해 효소에 의한 PAM 인식에 대한 변화를 용이하게 조명하지 않는다.

[0248]

먼저, 5 뉴클레오티드 고갈 결과를 전과 같이 4 뉴클레오티드 PAM으로 붕괴시켰다. 각각의 시점은 개별적으로 유지되었다. 각각의 돌연변이체-시점 NGS 총 카운트를 NGS 칩에 대한 로딩 차이를 처리하기 위해 PAM 당 100 카운트로 정규화하였다. 이어서 각각의 4 nt PAM에 대한 전체적 중위값을 각각의 돌연변이체-시점과 비교하였다. 이는 총 라이브러리와 비교한 고갈이라기 보다는 야생형과 비교한 고갈을 제공하였다. 결과는 어느 돌연변이가 PAM 인식 프로파일을 변화시켰는지를 조명한다. 본 발명자들은 보존적 접근법을 취하였으며, 돌연변이체에 의한 새로운 PAM 인식의 지표로서 4 이상의 고갈 점수를 선택하였다. 4의 고갈 점수는 야생형에 대한 중위값과 비교하여 그 특정 PAM-함유 라이브러리 구성원의 4배가 절단되었음을 지시하였다. 예를 들어 100 NGS 카운트가 야생형에 대해 GCGC를 갖는 PAM에 대해 남아 있고, 25 카운트가 특정 돌연변이체-시점에 대해 남아 있는 경우, 4의 점수가 계산되었다.

[0249]

186개의 돌연변이의 각각의 요약은 하기 표 6에 나타내어진다. 볼드체 글자로 제공된 돌연변이는 돌연변이가 4 초과점수를 갖는 야생형과 비교하여 3개 초과점수의 새로운 PAM 서열을 인식하고 절단하였음을 지시한다. 이탤릭체로 제공된 돌연변이는 돌연변이체가 4 초과점수를 갖는 야생형과 비교하여 1 내지 3개 또는 새로운 PAM 서열을 획득하였음을 지시한다. 정자체 (볼드체 또는 이탤릭체가 아님)의 돌연변이는 점 돌연변이가 4 초과점수를 갖는 야생형과 비교하여 새로운 PAM 서열을 절단하지 않았음을 지시한다. 특정 아미노산, 예컨대 T149는 단백질의 PAM-인식 도메인 부근에 있고 10개의 새로운 아미노산을 시험함에도 불구하고 새로운 PAM 인식을 획득하지 않았다. 다른 아미노산, 예컨대 D156은 새로운 PAM 인식 모티프를 조작하기 위한 핫-스팟인 것으로 나타났다. 아스파르트레이트 156은 10개의 상이한 아미노산으로 변화하는 경우 다수의 새로운 PAM을 인식하는 7개의 돌연변이를 가졌으며, 1개는 소수의 새로운 PAM을 나타내고, 2개는 wtLbCas12a와 비교하여 새로운 PAM을 획득하지 않았다. 일반적으로, 야생형과 비교하여 PAM 인식 및 절단의 차이를 나타낸 임의의 위치는 이중, 삼중, 또는 다중 돌연변이로 조합되어 PAM 인식을 추가로 변경시킬 수 있었다. 총 130개/186개의 점 돌연변이는 4의 점수 초과점의 wtLbCas12a (정자체/볼드체 또는 이탤릭체가 아님)에 비해 새로운 PAM을 획득하지 않았고, 40개/186개는 많은 새로운 PAM을 획득하였으며 (볼드체), 16개/186개는 1 내지 3개의 새로운 PAM을 획득하였다 (이탤릭체). 전체 30% 성공률 (56개/186개)은 효율적인 방법을 사용하여 LbCas12a에 대한 점 돌연변이를 생성함으로써 신규한 PAM 인식 모티프를 디자인하였음을 지시한다.

[0250] 표 6. 186개의 LbCas12a 점 돌연변이에 대한 요약 표 (참조 서열 서열식별번호:1).

K116R	K120R	K121S	D122R	E125R	T148H	T149A	T152R	D156R	E159K	Q529N	G532D	D535N
K176N	K120H	K121T	D122K	E125K	T148S	T149C	T152K	D156K	E159R	Q529T	G532N	D535H
	K120N	K121H	D122H	E125Q	T148A	T149S	T152W	D156Y	E159H	Q529H	G532S	D535V
	K120T	K121R	D122E	E125Y	T148C	T149G	T152Y	D156W	E159Y	Q529A	G532H	D535T
	K120Y	K121G	D122N			T149H	T152H	D156Q	E159Q	Q529F	G532F	D535S
	K120Q	K121D				T149P	T152Q	D156H		Q529G	G532K	D535A
						T149F	T152E	D156I		Q529S	G532R	D535W
						T149N	T152L	D156V		Q529P	G532Q	D535K
						T149D	T152F	D156L		Q529W	G532A	
						T149V		D156E		Q529D	G532C	
K538R	D541N	Y542R	L585G	K591W	M592R	K595R	V596T	S599G	K600R	K601R	Y616K	Y646R
K538V	D541H	Y542K	L585H	K591F	M592K	K595Q	V596H	S599H	K600H	K601H	Y616R	Y646E
K538Q	D541R	Y542H	L585F	K591Y	M592Q	K595Y	V596G	S599N	K600G	K601Q	Y616E	Y646K
K538W	D541K	Y542Q		K591H	M592E	K595L	V596A	S599D		K601T	Y616F	Y646H
K538Y	D541Y	Y542F		K591R	M592A	K595W					Y616H	Y646Q
K538F	D541I	Y542L		K591S		K595H						Y646W
K538H	D541A	Y542M		K591A		K595E					Y646N	W649S
K538L	D541S	Y542P		K591G		K595S						W649V
K538M	D541E	Y542V		K591P		K595D						W649T
K538C		Y542N				K595M						
K538G		Y542T										
K538A												
K538P												

[0251]

[0252]

많은 점 돌연변이는 신규한 PAM 인식을 LbCas12a에 부여하여, 그것이 시험관내에서 이들 서열이 선행하는 DNA를 절단하는 것을 허용하였다. 일부 돌연변이는 전체적 난잡성의 증가를 유발한 반면, 디자인되고 시험된 다른 것들은 wtLbCas12a 인식 및 절단을 변화시키는 것으로 나타나지 않았다. 전체 130개/186개의 점 돌연변이는 4의 점수 초과인 wtLbCas12a에 비해 새로운 PAM을 획득하지 않았고 (표 6, 정자체/이탤릭체 또는 볼드체가 아님), 40개/186개는 많은 새로운 PAM을 획득하였고 (표 6, 볼드체), 16개/186개는 1 내지 3개의 새로운 PAM을 획득하였다 (표 6, 이탤릭체). 전체 30% 성공률 (56개/186개)은 효율적인 방법을 사용하여 LbCas12a에 대한 점 돌연변이를 생성함으로써 신규한 PAM 인식 모티프를 디자인하였음을 지시한다.

[0253]

2. 원핵생물에서의 점 돌연변이 및 조합의 결합의 결정

[0254]

개별적 돌연변이의 조합은 단일 돌연변이보다 훨씬 더 PAM 인식을 변경시킬 수 있다. 그러나, 이러한 실험은 시험하기 위해 다수의 조합으로 신속하게 규모화된다. LbCas12a가 3개 이상의 새로운 PAM을 인식하는 것을 유발한 단지 40개의 돌연변이를 취하고, 이중 돌연변이체의 라이브러리를 생성하여, 총 40^2 또는 1,600개의 효소가 시험될 수 있었다. 삼중-돌연변이체를 생성하여 라이브러리는 반복실험에서 정제하고 검정하기 위한 40^3 또는 64,000개의 효소를 발생시킬 것이며, 이는 실체적이지 않다. 따라서, 본 발명자들은 PAM-SCANR로 공지된 박테리아 방법을 적응시켜 (Leenay et al. *Mol Cell* 62, 137-147 (2016)) 조합적 돌연변이를 평가하였다. 본 발

명자들은 *에스케리키아 콜라이*에서의 라이브러리를 사용하여 256개의 가능한 PAM NNNN 변이체체의 Cas12a 돌연변이의 결합을 시험하였다. 이 검정은 절단을 시험하지 않으며, 오히려 이는 생체내에서 결합을 시험한다. 생성된 돌연변이는 촉매 영역 부근의 어디에도 없었기 때문에, 결합은 절단을 또한 반영할 것으로 예상된다 (이는 293T 검정에서 나중에 확인되었음). PAM-SCANR에 대한 이점은 점 돌연변이 뿐만 아니라 아미노산 점 돌연변이의 조합을 빠르고 정확한 방식으로 신속하게 시험하는 능력이다. 이는 또한 시험관내 절단 검정보다 더 엄격한 경향이 있다.

[0255] 리포터 플라스미드

[0256] 플라스미드 pWISE1963을 256개의 PAM의 각각을 갖는 리포터를 생성하기 위한 기본 벡터로서 사용하였다. 플라스미드는 *lac* 프로모터의 제어 하에서 스펙티노마이신 저항성, ColE1 복제 원점, LacI, 및 eGFP를 함유한다. NotI 및 SmaI 제한 부위 (*lacI* 유전자 내로 *lacI* 프로모터의 단지 5'로부터) 사이에 단편을 함유하는 256개의 유전자 블록을 트위스트 바이오사이언스(Twist Bioscience)에 의해 합성하였다. 각각의 단편은 *lacI* 프로모터의 직접적으로 5'에 상이한 4-mer PAM을 함유하였다. 각각의 유전자 블록을 제한 및 라이게이션을 통해 pWISE1963 내로 클로닝하였다. 클론을 각각의 변이체에 대해 선택하고, PAM의 신원을 생거(Sanger) 시퀀싱에 의해 확인하였다.

[0257] CRISPR-Cas 플라스미드

[0258] 플라스미드 pWISE2031을 모든 CRISPR-Cas 플라스미드를 생성하기 위한 기본 벡터로서 사용하였다. 플라스미드는 클로람페니콜 저항성, CloDF13 복제 원점, 프로모터 BbaJ23108에 의해 유도된 dLbCas12a, 및 BbaJ23119 프로모터에 의해 유도된 *lacI* 프로모터를 표적화하는 crRNA를 갖는 LbCas12a를 함유한다. 음성 대조군 플라스미드, pWISE1961은 비-표적화 crRNA가 사용된 것을 제외하고는 pWISE2031과 동일한 성분을 함유한다. 각각의 점 및 조합적 돌연변이체 (pWISE2984 내지 pWISE3007)를 진위즈에서의 pWISE2031의 부위-지정된 돌연변이유발을 통해 구축하였다.

[0259] 세포주

[0260] *lacI* 유전자, JW0336의 염색체 결실을 함유하는 *이. 콜라이* 세포주를 다마콘 호리즌 디스커버리(Dharmacon Horizon Discovery)로부터 얻었다. 전기적격 세포를 기재된 프로토콜에 따라 제조하였다 (Sambrook, J., and Russell, D.W. (2006). *Transformation of E. coli by Electroporation*. Cold Spring Harb Protoc 2006, pdb.prot3933). *이. 콜라이* JW0336은 모든 라이브러리 형질전환 및 세포 분류 실험에 사용되었다.

[0261] 리포터 라이브러리의 제조

[0262] 상기 섹션에 기재된 10 ng의 각각의 리포터 플라스미드를 단일 튜브 내로 풀링하여 형질전환 및 증폭을 위한 라이브러리를 생성하였다. 풀링된 플라스미드 라이브러리의 1/20 (대략 0.5 ng의 각각의 리포터)을 초적격 XL1-블루 내로 제조업체의 지시서에 따라 형질전환시켰다. 225 rpm에서 진탕하면서 37℃에서 1-시간의 회수 후, 전체 형질전환물을 1-L의 LB 스펙티노마이신에 옮기고, 225 rpm에서 진탕하면서 37℃에서 밤새 성장시켰다. 다음 날, 플라스미드 DNA를 자이모퓨어(ZymoPURE) 플라스미드 기가프렙 키트를 제조업체의 지시서에 따라 사용하여 밤샘 배양물로부터 추출하였다. DNA를 나노드롭에 의해 정량화하고, 모든 후속 라이브러리 형질전환에 사용하였다.

[0263] 라이브러리 형질전환 및 세포 분류

[0264] 100 ng의 리포터 플라스미드 라이브러리 및 100 ng의 Crispr/Cas 플라스미드를 전기천공에 의해 40 uL의 JW0336 내로 공동-형질전환시켰다. 형질전환물을 225 rpm에서 1-시간 동안 진탕하면서 37℃에서 회수하였다. 회수의 종료시, 10 uL의 형질전환물을 제거하고, 90 uL LB와 혼합하고, 클로람페니콜 및 스펙티노마이신을 갖는 LB 한천 플레이트 상으로 플레이팅하여 형질전환 효율을 결정하였다. 나머지 양의 회수물 (990 uL)을 스펙티노마이신 및 클로람페니콜을 갖는 29 mL의 LB를 함유하는 밤샘 배양물에 옮겼다. 배양물을 225 rpm에서 진탕하면서 37℃에서 밤새 성장시켰다. 다음 날 아침, 콜로니를 형질전환 플레이트 상에서 카운팅하여 형질전환 효율을 결정하였으며; 그들 중 2개를 제외한 모든 형질전환물은 >2,000개의 형질전환체를 나타내었으며, 이는 리포터 라이브러리의 10X 이상의 커버리지와 동등하였다. 10X 이상의 커버리지를 나타내지 않은 2개의 샘플을 반복하였다. 이어서 밤샘 배양물의 글리세롤 스톱을 제조하고, -80℃에서 저장하고, 6-mL의 각각의 배양물을 쿼젠 미니프렙 키트를 제조업체의 지시서에 따라 사용하여 미니프렙하였다. 이들 미니프렙물을 "분류 전"으로 표지하고, 4℃에서 저장하였다.

[0265] 1 광학 밀도 (OD)의 각각의 밤샘 라이브러리 배양물을 8,000 rpm, 4℃에서 5분 동안 테이블탑 미세원심분리기에 서 스핀 다운시켰다. 상청액을 피펫팅하고, 1 mL의 필터 멸균된 1X PBS 완충제를 각각의 튜브에 첨가하였다. 펠릿을 피펫팅에 의해 주의 깊게 재-현탁시켰다. 1X PBS로의 세척을 2회 더 반복하고, 최종 재-현탁 후에 세포 (1X PBS 중 mL 당 약 10^8 개의 세포)를 얼음 상에 정치하였다. 각각의 샘플을 베크만-쿨터 모플로(MoFlo) XDP 세포 분류기 상에서 분류하였다. 음성 대조군 (WT-dLbCas12a + 비-표적화 crRNA +리포터 라이브러리) 및 양성 대조군 (WT-dLbCas12a + 표적화 crRNA + 리포터 라이브러리)을 사용하여 세포 분류를 위한 게이팅 파라미터를 설정하였다. 샘플을 전압 425, ssc 전압 535, 4.0의 fsc 전압 (게인)으로 단일-세포 순도 모드 상에서 분류하였다. 전형적인 분류 속도는 약 4000개의 사건/초였다. 각각의 샘플은 최소 1.0×10^6 개의 사건을 가졌으며; 세포 분류를 50,000개의 GFP-양성 사건이 수집되거나 샘플이 고갈될 때까지 수행하였다. 샘플이 고갈된 경우에, 최소 200개의 GFP 양성 사건을 수집하였다. GFP-양성 사건을 스펙티노마이신 및 클로람페니콜을 갖는 2-mL의 LB를 함유하는 튜브 내로 수집하였다. 분류 후, 샘플을 스펙티노마이신 및 클로람페니콜을 갖는 추가의 LB로 6-mL로 희석하고, 이어서 225 rpm에서 진탕하면서 37℃에서 밤새 성장시켰다.

[0266] 예시 분류는 도 6 내지 11에 제공된다. 도 6은 wtLbCas12a 및 플라스미드 스페이서를 표적화하지 않은 crRNA를 함유하는 음성 대조군의 세포 분류 결과를 나타낸다. GFP 높음 샘플로부터의 분류된 세포는 분류된 분획에서 세포를 나타내지 않고 (좌측 패널), 단일 피크에 의해 지시된 GFP 신호의 단일 집단을 나타낸다 (우측 패널). 도 7은 wtLbCas12a 및 플라스미드 스페이서를 표적화하는 crRNA의 세포 분류 결과를 나타낸다. GFP 높음 샘플로부터의 분류된 세포는 분류된 분획에서 세포를 나타내고 (좌측 패널, GFP hi), 2개의 우세한 피크에 의해 지시된 GFP 신호의 2개의 집단을 나타낸다 (우측 패널, GFP neg 및 GFP 높음). GFP 높음 분류된 세포 형광화가 또한 나타내어진다 (하부 우측 패널). 도 8은 LbCas12a-K595Y 및 플라스미드 스페이서를 표적화하는 crRNA의 세포 분류 결과를 나타낸다. GFP 높음 샘플로부터의 분류된 세포는 분류된 분획에서 세포를 나타내고 (좌측 패널, GFP hi), 2개의 우세한 피크에 의해 지시된 집단을 나타낸다 (우측 패널, GFP neg 및 GFP 높음). 도 9는 LbCas12a-G532R-K595R 이중 돌연변이 대조군 (Gao et al. Nat Biotechnol 35, nbt.3900 (2017)) 및 플라스미드 스페이서를 표적화하는 crRNA의 세포 분류 결과를 나타낸다. GFP 높음 샘플로부터의 분류된 세포는 분류된 분획에서 세포를 나타내고 (좌측 패널, GFP hi), 2개의 우세한 피크에 의해 지시된 집단을 나타낸다 (우측 패널). 도 10은 플라스미드 스페이서를 표적화하는 crRNA와 함께 사용된 점 돌연변이 중 2개의 조합인 LbCas12a-T152R-K595Y 이중 돌연변이의 세포 분류 결과를 나타낸다. GFP 높음 샘플로부터의 분류된 세포는 분류된 분획에서 세포를 나타내고 (좌측 패널, GFP hi), 2개의 우세한 피크에 의해 지시된 집단을 나타낸다 (우측 패널). 도 11은 플라스미드 스페이서를 표적화하는 crRNA와 함께 사용된 점 돌연변이 중 3개의 조합인 LbCas12a-T152R-K538W-K595Y 삼중 돌연변이의 세포 분류 결과를 나타낸다. GFP 높음 샘플로부터의 분류된 세포는 분류된 분획에서 세포를 나타내고 (좌측, GFP hi), 지시된 집단을 나타낸다 (우측, 녹색 선).

[0267] 차세대 시퀀싱

[0268] 분류 후 다음 날 아침, 각각의 밤샘 배양물의 글리세롤 스톱을 제조하고, -80℃에서 저장하였다. 나머지 양의 각각의 6-mL 배양물을 쿼아젠 미니프랩 키트로 제조업체의 지시서에 따라 미니프랩하였다. 이들 미니프랩물을 "분류 후"로 표지하고, 4℃에서 저장하였다. 분류 전 및 분류 후 미니프랩물을 나노드랍에 의해 정량화하고, 10-배 희석하고, 일루미나 마이-섹 상에서의 시퀀싱을 위해 핸드 오프하였다.

[0269] 스페이서 벡터를 딥 시퀀싱 분석으로 처리하여 일루미나 마이섹을 제조업체의 프로토콜에 따라 사용하여 PAM의 각각의 위치에서의 A/T/G/C의 빈도를 계산하였다. 간략하게, 10 ng의 DNA를 PCR을 위한 주형으로서 사용하였다. 위상조정 유전자 특이적 정방향 및 역방향 PCR 프라이머를 표적 부위에 걸쳐 증폭시키도록 디자인 하였다. 앰플리콘 라이브러리를 2-단계 PCR 방법을 사용하여 생성하였으며, 여기서 5' 꼬리를 갖는 1차 PCR은 2차 PCR이 일루미나 i5 및 i7 어댑터 서열 및 다중화 샘플을 분류하기 위한 바코드를 첨가하는 것을 허용한다. PCR 증폭을 하기 파라미터를 사용하여 수행하였다: 30초 동안 98℃; PCR1에 대해 25 사이클 및 PCR2에 대해 8 사이클 (98℃ 10초, 55℃ 20초, 72℃ 30초); 5분 동안 72℃; 12℃에서 보유. PCR 반응을 Q5 고-충실도 DNA 폴리머라제 (뉴 잉글랜드 바이오랩스, 미국 매사추세츠주 베버리)로 수행하였다. 2차 PCR 앰플리콘 샘플을 AM퓨어 XP 비드를 제조업체의 지시서 (베크만 쿨터, 미국 캘리포니아주 브레아)에 따라 사용하여 개별적으로 정제하고; 모든 정제된 샘플을 플레이트 판독기를 사용하여 정량화하고, 등물 비로 풀링하고, AATI 단편 분석기 (애질런트 테크놀로지스, 미국 캘리포니아주 팔로 알토) 상에서 실행하였다. 풀링된 앰플리콘 라이브러리를 일루미나 마이섹 (2 X 250 쌍형성된 단부) 상에서 마이섹 시약 키트 v2 (일루미나, 미국 캘리포니아주 샌 디에고)를 사용하여 시퀀싱하였다.

[0270] 고도로 형광 분류된 세포의 시퀀싱 결과

[0271] wtdLbCas12a 및 비-표적화 crRNA를 함유하는 2개의 음성 대조군 샘플을 256-구성원 리포터 라이브러리의 존재 하에서 실행하였다. 1.0으로의 정규화 후, 라이브러리의 각각의 구성원에 대한 값을 히스토그램으로서 플롯팅 하였다 (도 12). 도 12는 2개의 별개의 crRNA 없는 대조군 및 야생형 dLbCas12a 및 리포터 라이브러리에 대한 총 정규화된 NGS 카운트를 나타낸다. 2개의 별개의 샘플을 분석하고 조합하였다 (256개의 PAM x 2를 나타내는 총 512개의 점). 본 발명자들은 그 초과를 본 발명자들이 PAM 결합으로서 점수화한, 이들 실험에 대한 컷오프로서, 가장 높은 카운트인 보존적 1.67 값을 선택하였다. 표준 편차는 0.16이었다. 컷오프로서 표준 편차의 일부 승수를 선택하기 보다는, 본 발명자들은 2개의 음성 대조군 중 어느 하나에서 발견된 절대 가장 큰 값, 1.67을 선택하였다. 이는 데이터의 표준 편차의 10배 초과와 고도로 엄격한 컷오프를 제공하였다. 사실, 단지 3개의 PAM 서열이 1.5 초과에서 발견되었다.

[0272] 분류 전의 풀을 분류 전에 모두 시퀀싱하였다. PAM 당 평균 판독물은 샘플에 따라 약 250 내지 500 NGS 판독물 이었다. 고도로 형광 분류 후의 풀을 시퀀싱하였으며, PAM 당 약 250 내지 500개의 판독물에서 PAM 당 유사한 판독물 카운트를 가졌다. 이어서 둘 다의 샘플을 각각의 NGS 실험에서 작은 로딩 차이에 대해 대조군에 대해 정규화하였다. 이어서 2개의 값을 차감하고, 1.0으로 정규화하였다. 많은 PAM 서열은 1.67 컷오프 초과와 점 돌연변이 라이브러리에 의해 결합되었다 (도 13, 표 7).

[0273] 표 7. 정규화된 점수에 의해 분류된 1.67 역치 초과와 PAM에 결합하는 점 돌연변이의 능력. 야생형에 대한 TTTV 서열은 **볼드체**이다.

PAM 야생형		T152R PAM		K538W PAM		K595Y PAM		D122H PAM	
야생형 점수		T152R 점수		K538W 점수		K595Y 점수		D122H 점수	
TTTC	2.96	TTTG	2.94	TTTA	4.71	TCGC	3.23	TTTC	3.03
CTTG	2.56	TTTA	2.9	CTTA	2.80	TCCC	3.03	TTTG	2.87
TTTA	2.54	CTTA	2.72	AGCT	2.70	TTGC	2.98	TTTA	2.56
TTTG	2.47	TTTC	2.66	AATG	2.19	GCGC	2.93	TTTT	2.46
CCGC	2.23	GTTA	2.48	CTTC	2.03	ACGC	2.40	CTTA	2.20
CTTA	2.19	CTTG	2.19	TTTC	1.82	TCCA	2.35	TTCC	2.19
CTTC	1.97	GTTG	2.1			TTCC	2.32	CTTG	2.10
GTTT	1.82	CTTC	2.02			TCCG	2.17	GTTA	1.97
TTCC	1.82	AGAG	1.91			TTCCG	2.01	ATTG	1.96
GTTG	1.80	ATTG	1.85			TCGG	1.93	CTTC	1.89
GTTA	1.75	TCCA	1.79			CCGC	1.84	TCTG	1.82
		GCCA	1.72			CCCA	1.79	CCCA	1.79
		TCTA	1.72			ACCC	1.74	GTTC	1.72
		TTTT	1.72						
		GTTC	1.71						
D156V PAM		D156Q PAM		D535N PAM		E159K PAM		G532H PAM	
D156V 점수		D156Q 점수		D535N 점수		E159K 점수		G532H 점수	
AATG	3.13	TGTG	3.96	TTTG	4.66	TTTG	2.37	AATG	2.30
GTTG	2.16	CCTC	2.42	TTTC	4.23	TTTC	2.36	CCTA	1.89
TTTT	2.15	CCCA	2.14	TTTA	3.91	TTTA	2.25	GTTG	1.83
TTTA	2.02	GTTG	2.09	TTTT	2.04	CTTA	2.07	GTTA	1.79
CTTA	1.99	GTTT	2.07	CACG	2.03	CTTG	1.94	TTTA	1.77
GTTA	1.98	CTTC	1.97			CTTC	1.86	TCTA	1.73
ATTA	1.97	ATTC	1.96			CCCA	1.68	CTTA	1.68
CTTG	1.93	TTCC	1.90			GTTA	1.67		
GTTT	1.85	TTGC	1.90						
CTTC	1.77	TTCA	1.88						
TCTC	1.71	ACTC	1.86						
TTCA	1.70	ACAC	1.85						
		CTTG	1.82						
		AGCT	1.80						
		AATG	1.80						
		AAGC	1.72						
		GTAC	1.70						
		TCGC	1.69						
		CCTG	1.69						
		GTTA	1.69						
		TCCC	1.67						
G532K PAM		K121R PAM		K595W PAM		K538W PAM		D535N PAM	
G532K 점수		K121R 점수		K595W 점수		K538W 점수		D535N 점수	
TTTA	2.76	TTTG	3.28	TTTG	2.35	TTTA	11.18	GCCG	18.70
TTTC	2.66	TTTA	3.15	TTCC	2.27	CTTA	9.57	TTTG	16.34

TTTG	2.56	TTTC	2.95	TTTA	2.23	TTTC	8.45	TTTA	13.15
CCGC	2.51	CTTG	2.18	GCGC	2.03	TTTG	7.27	TTTC	3.87
TTCC	2.37	CTTA	2.17	TTCCG	2.02	CTTC	6.70	TTGG	3.83
TTCA	2.16	TTTT	2.14	TTTC	2.01	CTTG	6.33	TTAC	2.94
TCCA	2.13	CTTC	1.85	TTCA	1.99	TATC	5.35	TCAC	2.78
TCCA	1.76	GTTG	1.80	ACCG	1.97	TATA	4.90	TTGC	2.38
GTTA	1.75	GTTA	1.78	TCCG	1.87	TATG	4.39	TTCC	1.87
ATTG	1.72	TTGG	1.74	TCGC	1.85	AATG	2.75	CCAC	1.87
CCCC	1.69	TTCC	1.71	TCCC	1.84	CCGC	2.19	TCGC	1.84
CTTA	1.68			AACG	1.84	AATC	2.13	CTTG	1.67
				TTGC	1.83	GTTA	1.89	TTTT	1.67
				TCCA	1.76	AATA	1.68		

[0274] 본 발명자들은 3개의 점 돌연변이 T152R, K538W, 및 K595Y를 다양한 조합으로 조합하여 이중 및 삼중 dLbCas12a 돌연변이체 (T152R+K538W, K538W+K595Y, 및 T152R+K538W+K595Y)를 제조하였다. 이를 그의 LbCas12a 돌연변이가 G532R+K595R 대조군에 상응하는 'RR'로 공지된 AsCas12a에서 개발된 이전에 기재된 대조군과 비교하였다 (Gao et al. *Nat Biotechnol* 35(8):789-792 (2017)). 'RR'은 AsCas12a 및 이어서 LbCas12a에서 TYCV+CCCC 서열에서 인텔을 유발할 수 있는 것으로 기재되었다.

[0275] 동일한 방법론을 적용하여 조합적 돌연변이를 점수화하였다. PAM 라이브러리 구성원 당 각각 평균 약 250 내지 500개의 마이섹 NGS 판독물을 갖는 분류 전 및 후의 풀을 시퀀싱하였다. 분류 전 및 후의 풀을 정규화하고, 차

감하여, 차이를 1.0로 정규화하였다. 많은 PAM 서열은 1.67 컷오프 초과에 의해 결합되었다 (도 14, 표 8).

표 8. 정규화된 점수에 의해 분류된, 1.67 역치 초과에 PAM에 결합하는 조합적 돌연변이의 능력. 야생형에 대한 TTV 서열은 볼드체이다. G532R+K595R 대조군에 대한 TYCV+CCCC 서열은 밑줄친다.

야생형 PAM	점수	G532R+K595R PAM	점수	T152R+K538W PAM	점수	K538W+K595Y PAM	점수	T152R+K538W+K595Y PAM	점수
TTTC	2.96	TTTT	5.07	TTTC	11.96	TTTC	11.96	AGCC	13.65
CTTG	2.56	TTTC	4.81	TTTA	11.62	TTTA	11.62	TGGC	10.28
TTTA	2.54	ACCC	4.52	TTTG	11.19	TTTG	11.19	TACC	9.35
TTTG	2.47	TCCG	4.51	GTTG	7.75	GTTG	7.75	TGCC	7.68
CCGC	2.23	TTCA	4.46	GTTA	6.68	GTTA	6.68	CTGC	6.69
CTTA	2.19	TTTA	4.30	CTTG	5.53	CTTG	5.53	GGGC	6.09
CTTC	1.97	TTTG	4.28	CTTA	4.93	CTTA	4.93	GGCC	5.71
GTTC	1.82	GCCG	4.15	GTTC	4.34	GTTC	4.34	AGGC	5.26
TTCC	1.82	TTCT	4.07	TTCC	4.11	TTCC	4.11	TAGC	4.96
GTTG	1.80	CTCC	4.03	TCCC	3.82	TCCC	3.82	CGGC	4.89
GTTA	1.75	GCCA	4.01	TTCA	3.78	TTCA	3.78	ACCC	4.15
		CCCG	4.00	ATTG	3.78	ATTG	3.78	GACC	4.08
		TCCA	3.92	TCCG	3.48	TCCG	3.48	AAGC	3.81
		ACTA	3.90	TTGG	3.43	TTGG	3.43	AGCA	3.69
		GTCC	3.87	GCCC	3.32	GCCC	3.32	GCGC	3.49
		CCCC	3.87	CTTC	3.30	CTTC	3.30	TACA	3.26
		TTTC	3.84	CCCC	2.95	CCCC	2.95	TCCC	3.17
		CCCA	3.78	TCTA	2.88	TCTA	2.88	CCCA	3.15
		TCTA	3.75	GCCA	2.86	GCCA	2.86	TGCA	2.87
		TCCT	3.68	TCCA	2.86	TCCA	2.86	GCCC	2.66
		GCTA	3.63	ATTA	2.44	ATTA	2.44	CCCC	2.49
		GCCC	3.59	CTCC	2.29	CTCC	2.29	TTGC	2.42
		GTTA	3.58	CCCG	2.19	CCCG	2.19	ACGC	2.21
		ACCA	3.43	ACCC	2.16	ACCC	2.16	GGCA	2.21
		TCCC	3.38	TTCC	2.13	TTCC	2.13	ACCA	2.18
		TTCC	3.36	GTCC	2.13	GTCC	2.13	AACC	2.09
		ACCG	3.36	CCCA	2.06	CCCA	2.06	GACA	1.83
		TCTG	3.31	GCCG	1.99	GCCG	1.99	CAGC	1.81
		TCTT	3.05	GCTA	1.97	GCTA	1.97	ACCG	1.80
		CTTA	2.92	TCTG	1.96	TCTG	1.96	GGGG	1.80
		TCTC	2.89	TCTC	1.84	TCTC	1.84	TACG	1.72
		CCCT	2.89	TTTT	1.81	TTTT	1.81	TTCC	1.69
		CTCA	2.86	ATTC	1.79	ATTC	1.79		
		GTCA	2.63	GTCA	1.67				
		GTCC	2.54						
		CCTA	2.54						
		ATCC	2.34						
		GTTC	2.11						
		TTAC	2.10						
		CTCG	2.03						
		ATTA	2.01						
		GCTC	1.94						
		GCCT	1.89						
		CCTC	1.82						
		GCTT	1.71						

PAM-SCANR 데이터의 전체적 분석

야생형 LbCas12a는 강한 TTTV 결합을 나타내고, LbCas12a-G532R-K595R 대조군은 강한 TYCV 및 CCCC 결합을 나타내었다. 'RR'로 지칭되는 이 돌연변이는 AsCas12a에서 개발되었고, TYCV 및 CCCC에 결합하는 것으로 나타났다 (Gao et al. *Nat Biotechnol* 35(8):789-792 (2017)). 그러나, 시험관내 wtLbCas12a는 TTCN, CTNN, TCTN, 및 다른 것들을 인식하고 절단할 수 있으며, 여기서 AsCas12a는 단지 TTTN을 절단하는 것으로 나타났다 (Zetsche et al., *Cell* 163:759-771 (2015)). 본 발명자들은 LbCas12a 맥락에서 정치된 'RR' 돌연변이가 그것이 AsCas12a 맥락에서 정치되는 경우보다 더 난잡할 것으로 추론하였고, 이는 이 대조군에 대해 관찰된 것이었으며, LbCas12a-RR은 45개의 서열을 인식한다. 야생형 및 LbCas12a-RR 결과 둘 다는 선택 및 분류 파라미터의 타당성을 입증한다.

시험된 돌연변이는 신규한 PAM이 시험관내에서 확인된 개별적 점 돌연변이에 의해 생체내에서 인식됨을 명백하게 지시하였다. 예를 들어, K595Y는 1.67 역치 초과에 13개의 PAM에 결합하였으며, 그 중 11개는 wtLbCas12a에

의해 인식되지 않았고, 그 중 어느 것도 Cas12a에 의해 결합되는 것으로 공지된 TTTV 서열을 함유하지 않았다. 유사하게, T152R은 15개의 별개의 PAM을 인식하였지만, 이 경우에 이는 야생형의 TTTV를 보유하였다. 전체적으로 시험된 12개의 점 돌연변이 중에서, 각각은 정준 TTTV 모티프의 외부의 및 dLbCas12a 대조군과는 별개의 신규한 PAM 결합 서열을 가졌다.

[0283] 조합의 효과

[0284] 본 발명자들은 다수의 점 돌연변이를 조합하는 것이 점 돌연변이의 PAM 서열의 선형 첨가를 초래하지 않았음을 밝혀내었다 (도 15). 예를 들어, LbCas12a 점 돌연변이체 K538W 및 K595Y를 조합하는 것은 일부 경우에 K538W (수직 해칭됨) 또는 K595Y (수평 해칭됨)와 PAM 인식 모티프를 공유하는 효소 LbCas12a-K538W-K595Y를 발생시키지만, 보다 종종 완전히 신규한 PAM 인식 서열 (태칭됨)을 발생시킨다. 동일한 예를 사용하여, K538W는 AGCT를 인식하지만, K538W+K595Y는 그렇지 않다. K595Y는 ACGC를 인식하지만, K538W+K595Y는 그렇지 않다. CCCC는 K538W에 의해서도 K595Y에 의해서도 인식되지 않지만, 이중 돌연변이체는 고 친화도로 이에 결합한다.

[0285] 전체적으로, 돌연변이의 조합은 PAM 인식의 선형 확장 초과를 발생시킨다. 예를 들어 K538W는 6개의 PAM 서열을 인식하고, K595Y는 13개를 인식하지만, 함께 이들은 32개의 서열을 인식한다 (도 15). 단순한 상가 효과는 본 발명자들이 관찰한 32개라기 보다는, 이중 돌연변이체에 대해 19개의 PAM을 발생시킬 것이다. 더욱이, 이중 돌연변이체에 의해 인식되는 32개의 서열 중 단지 11개는 2개의 단일 돌연변이 중 어느 하나에 의해 인식된다. 유사한 패턴이 3개의 돌연변이를 조합한 경우 관찰된다 (도 16). T152R, K538W, 및 K595Y의 조합은 3개의 개별적 돌연변이 단독 중 임의의 것과는 상이한 PAM 인식을 갖는 삼중 돌연변이를 발생시킨다. 예를 들어: GGCA, GGCC, GGGC, 및 GGGG 단독은 모든 3개의 돌연변이가 LbCas12a 상에 생성되는 경우 인식된다. 단지 T152R, K538W, 및 K595Y가 함께 모두 돌연변이된 경우를 제외하고는, 이들 PAM 중 어느 것도 임의의 단일 또는 이중 돌연변이에 의해 결합되지 않는다.

[0286] PAM-SCANR 대 시험관내 PAMDA에서의 점 돌연변이 PAM 인식의 비교

[0287] 전체로서 시험된 12개의 점 돌연변이에 대해 취하여, 1.67 초과의 PAM-SCANR 히트는 시험관내 PAMDA 고갈 검정에서 잘 나타내어졌다. K595Y 및 T152R 예가 나타내어진다 (도 17). 도 17은 PAM-SCANR로부터의 1.67 점수 초과를 나타낸 모든 비-TTTV PAM (회색 박스)를 K595Y (좌측 패널) 및 T152R (우측 패널)과 비교한다. 1.67 컷오프 초과의 PAM-SCANR 양성 PAM 중 하나를 제외한 전부는 시험관내에서 9.2 컷오프 초과의 PAM 고갈 점수를 가졌다. 그러나, PAM-SCANR 방법 및 분석은 시험관내 검정 및 분석보다 더 엄격하였다. 예를 들어, PAM-SCANR에서 13개의 상이한 PAM이 분류되고, 시퀀싱되고, 1.67 초과의 값을 갖도록 정규화되었다. 이는 시험관내에서 용이하게 절단된 바와 같은 45개의 서열을 확인한 PAMDA 검정과 대조적이다. 이는 세포의 내부 대 시험 튜브 중의 상대 농도의 함수일 가능성이 있지만, 이는 또한 PAM-SCANR에 대해 너무 엄격한 컷오프의 또는 PAMDA 검정에 대해 너무 허용적인 컷오프의 설정의 함수일 수 있다.

[0288] 촉매 부위로부터 먼 잔기의 본 발명자들의 조작이 PAM 인식 및 결합에 영향을 미치고, 촉매작용에 영향을 미치지 않음을 지시하는 데이터세트 사이에 상관관계가 있다. 이들 잔기에서의 돌연변이가 PAM 결합과 함께 뉴클레아제 활성에 영향을 미친 경우, PAMDA 검정에서 절단을 나타내지 않은 PAM-SCANR 검정 (절단이 아니라 결합을 측정함)에서 많은 히트가 있을 것이다. 본 발명자들은 그 패턴을 보지 않는다. 본 발명자들은 결합의 변화에 영향을 미치고 (PAM-SCANR) 또한 초래한 돌연변이가 시험관내에서 절단되었음을 (PAMDA) 관찰한다.

[0289] 3. 진핵생물에서의 결합, 절단, 및 인텔 형성의 결정

[0290] 본 발명자들은 진핵생물 HEK293T 세포에서 삽입 또는 결실 (인텔)을 유발하는 그들의 능력을 시험하기 위해 3개의 돌연변이 T152R, K538W, 및 K595Y를 선택하였다. 이 검정은 가치 있는 진핵생물 인텔 데이터를 제공한다. 진핵생물에서 삽입 및 결실을 얻기 위해, 다수의 기준 모두가 충족되어야 한다: CRISPR 효소는 세포에서 발현되고 안정할 필요가 있고, crRNA는 발현되고 정확하게 프로세싱될 필요가 있고, 단백질:RNA 복합체가 형성될 필요가 있고, 복합체는 안정할 필요가 있고, 복합체는 충분한 양으로 핵 내로 전위될 필요가 있고, 표적 DNA는 접근 가능할 필요가 있고, DNA는 특정 가이드-RNA 디자인에 의해 잘 표적화되어야 하고, 이중-가닥 파단은 삽입 또는 결실 (인텔)을 통해 가끔의 DNA 복구 실수를 생성하는데 충분하게 높은 비율로 일어날 필요가 있다. 이는 진핵생물 검정을 이 연구에서 가장 엄격한 검정으로 만든다. 실험이 저-처리량임으로 인해 모든 256개라기 보다는, 하기 기재된 3개의 점 돌연변이체의 각각에 대해 수십 개의 PAM을 시험하였다. 종종 특정 가이드는 표적 접근성으로 인해 비효율적이기 때문에 가-음성을 회피하도록 시도하기 위해 3개의 상이한 표적을 각각의 PAM-돌연변이체 조합에 대해 선택하였다.

[0291] HEK293T 세포 시험

[0292] 진핵생물 HEK293T (ATCC CRL-3216) 세포를 10% (v/v) FBS (FBS)로 보충된 돌베코 변형 이글 배지(Dulbecco's Modified Eagle's Medium) 플러스 글루타맥스(GlutaMax) (썬모피셔)에서, 37 °C에서 5% CO2로 배양하였다. 야생형 및 돌연변이체 LbCas12를 고체-상태 합성을 사용하여 합성하고, 이어서 CMV 프로모터 뒤의 플라스미드 내로 클로닝하였다. CRISPR RNA (crRNA)를 인간 U6 프로모터 뒤로 클로닝하였다 (표 9). HEK293T 세포를 48-웰 콜라겐-코팅된 바이오코트(BioCoat) 플레이트 (코닝(Corning)) 상에 시딩하였다. 세포를 약 70% 전면생장물로 형질감염시켰다. 750 ng의 단백질 플라스미드 및 250 ng의 crRNA 발현 플라스미드를 제조업체의 프로토콜에 따라 웰 당 1.5 µl의 리포펙타민 3000 (썬모피셔 사이언티픽)을 사용하여 형질감염시켰다. 형질감염된 세포로부터의 게놈 DNA를 3일 후에 얻고, 인텔을 검출하고, 고-처리량 일루미나 앰플리콘 시퀀싱을 사용하여 정량화하였다.

[0293] 스페이서 벡터를 딥 시퀀싱 분석으로 처리하여 일루미나 마이섹을 제조업체의 프로토콜에 따라 사용하여 PAM의 각각의 위치에서의 A/T/G/C의 빈도를 계산하였다. 간략하게, 10 ng의 DNA를 PCR을 위한 주형으로서 사용하였다. 위상조정 유전자 특이적 정방향 및 역방향 PCR 프라이머를 표적 부위에 걸쳐 증폭시키도록 디자인하였다. 앰플리콘 라이브러리를 2-단계 PCR 방법을 사용하여 생성하였으며, 여기서 5' 꼬리를 갖는 1차 PCR은 2차 PCR이 일루미나 i5 및 i7 어댑터 서열 및 다중화 샘플을 분류하기 위한 바코드를 첨가하는 것을 허용한다. PCR 증폭을 하기 파라미터를 사용하여 수행하였다: 30초 동안 98°C; PCR1에 대해 25 사이클 및 PCR2에 대해 8 사이클 (98°C 10초, 55°C 20초, 72°C 30초); 5분 동안 72°C; 12°C에서 보유. PCR 반응을 Q5 고-충실도 DNA 폴리머라제 (뉴 잉글랜드 바이오랩스, 미국 매사추세츠주 베버리)로 수행하였다. 2차 PCR 앰플리콘 샘플을 AM퓨어 XP 비드를 제조업체의 지시서 (베크만 쿨터, 미국 캘리포니아주 브레아)에 따라 사용하여 개별적으로 정제하고; 모든 정제된 샘플을 플레이트 판독기를 사용하여 정량화하고, 등몰 비로 풀링하고, AATI 단편 분석기 (애질런트 테크놀로지스, 미국 캘리포니아주 팔로 알토) 상에서 실행하였다. 풀링된 앰플리콘 라이브러리를 일루미나 마이섹 (2 X 250 쌍형성된 단부) 상에서 마이섹 시약 키트 v2 (일루미나, 미국 캘리포니아주 샌 디에고)를 사용하여 시퀀싱하였다.

[0294] 야생형 대조군

[0295] 야생형 LbCas12a (wtLbCas12a)는 TTTV (TTTA, TTTC, 및 TTTG)를 인식한다. 본 발명자들은 crRNA 스페이서 (표 9)를 사용하여 TTTV 함유 23 뉴클레오티드 스페이서 표적에 대해 야생형 단백질을 시험하였다 (도 18).

[0296] 표 9. 스페이서 및 표적

스페이서 명칭	HEK 표적	PAM	스페이서 서열	SEQ ID NO
PWsp125	EMX1	TTTC	TCATCTGTGCCCTCCCTCCCTG	54
PWsp449	FANCF1	TTTC	GCGGATGTTCCAATCAGTACGCA	55
PWsp450	HEK2	TTTC	CAGCCCCTGGCCCTGTAAAGGA	56
PWsp453	RNF2	TTTA	TATGAGTTACAACGAACACCTCA	57
PWsp454	RNF2	TTTA	CACGTCTCATATGCCCTTTGGCA	58
PWsp455	RNF2	TTTA	GAACATGAAAACCTTAAATAGAAC	59
PWsp759	DNMT1	TTTG	GTCAGGTTGGCTGCTGGGCTGGC	60
PWsp452	HEK2	TTTG	ACAGATGGGGCTGGACAATTTTT	61

[0297]

[0298] 단백질 및 표적의 선택

[0299] PAMDA 시험관내 검정에서 증가된 PAM 접근성을 나타낸 많은 점 돌연변이가 있었다. 내인성 293T 세포 표적에 대한 모든 효율적인 PAM 돌연변이체를 시험하는 것은 실험의 복잡성, 비용, 및 시간으로 인해 본 발명자들의 많은 점 돌연변이 및 256개의 가능한 4 뉴클레오티드 (nt) PAM을 고려하면 효과적으로 불가능하다. 따라서, 본 발명자들은 3개의 점 돌연변이를 선택하여 PAM의 하위세트에 대해 시험하였다. 시험된 3개의 점 돌연변이는 T152R, K538W, 및 K595Y였다.

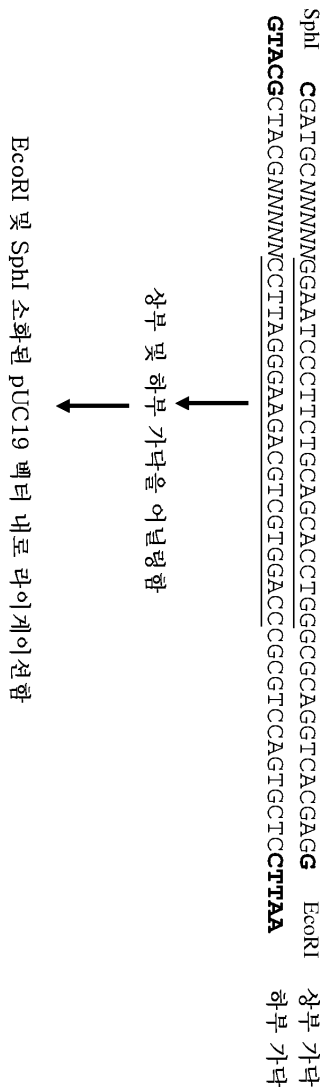
[0300] 293T 세포에서의 게놈 표적을 그들의 PAM 서열에 기반하여 선택하였다. 3개의 점 돌연변이에 대한 게놈 표적을 적절한 4 nt PAM을 갖고 그 PAM으로부터 하류의 23개의 뉴클레오티드를 선택하는 것 이외의 특정 규칙 없이 사용하여 무작위로 선택하였다. 3개의 상이한 스페이서를 선택하여 각각의 PAM을 검정하였다. 이는 CRISPR 효소의 활성이 표적-특이적이고, 종종 예측될 수 없다는 관찰을 설명하기 위한 것이다.

- [0301] 평균적으로, 본 발명자들은 시험된 23개의 뉴클레오티드 wtLbCas12a 스페이스의 대략 절반이 정확한 PAM TTTV 서열을 갖는 것에도 불구하고 비효과적임을 관찰하였다. 각각의 PAM에 대해 단지 3개의 테이터 점 및 표적의 약 50%가 인텔을 생성하지 않을 것이라는 관찰로, 무작위로 디자인된 스페이스에 대한 평균이라기 보다는 PAM 당 최대 인텔 백분율을 가시화함으로써 PAM 인식을 평가하는 것이 보다 유용한 정보를 제공한다 (도 19 내지 21). 각각의 PAM에 대해 보다 많은 수의 스페이스를 검정한 경우, 통계적 시험을 사용하여 그들의 평균 편집 효율을 평가할 수 있다.
- [0302] 본 발명자들의 전체적 형질감염 및 검정 조건은 TTTC에 대해 대략 11 내지 26%, 및 TTTA에 대해 10%, 및 TTTG에 대해 4 내지 10%에서 HEK293T 세포에서 인텔을 유발하는 야생형 LbCas12a를 발생시킨다 (도 18). 이들은 문헌으로부터 사전-정의된 HEK293T 표적 부위 및 가이드이며, 따라서, 임의의 무작위로 선택된 가이드보다 더 효과적일 것으로 예상된다. 돌연변이체에 대한 crRNA 가이드에 대한 무작위 디자인에도 불구하고, 많은 새로운 Cas12a PAM 인식 부위는 TTTV 서열에서 야생형과 유사한 비율로 인텔을 유발하였다 (도 19 내지 21). 0.1% 초과 임의의 인텔은 10,000 NGS 관독물-깊이에서 관독된 시퀀싱 검정의 잡음 초과이다.
- [0303] K595Y는 ACCG에서 25.5%로, CCGC에서 10.9%로, TCGC에서 10.1%로, CCGC에서 9.5%로, GCGC에서 8.3%로, CTGG에서 7.8%로, ACGG에서 6.3%로, CCGC에서 6.0%로, TGGC에서 5.3%로, 및 기타로 인텔을 유발할 수 있었다 (도 19). 이들 수는 무작위로 디자인됨에도 불구하고 모두 wtLbCas12a에 대한 TTTV 대조군의 범위 내에 있다. Cas12a 단백질의 한 주요한 특징은 이들이 T-풍부 PAM을 인식한다는 것이다 (Zetsche et al. *Cell* 163:759-771 (2015)). 이는 게놈 편집 기술에 있어서 그들의 유용성을 제한한다. K595Y는 명백하게 C 및 G-풍부 PAM을 선호하며, 이는 G-풍부 PAM을 이용하는 Cas9 CRISPR 효소의 표적을 이전에 우세하게 표적화하는 Cas2a 유용성을 확장시킬 것이다 (Jinek et al. *Science* 337, 816-821 (2012)). 총 256개의 가능한 4 뉴클레오티드 PAM 중 단지 31개를 K595Y에 대해 293T 세포 (또는 12%)에서 시험하였다. K595Y에 의해 인식되고 진핵생물 세포에서 인텔을 유발할 수 있는 많은 다른 PAM이 있을 가능성이 있다.
- [0304] T152R은 CCTC에서 11.5%로, CCTG에서 10.0%로, CCGA에서 9.6%로, GCCA에서 8.4%로, GCCC에서 7.2%로, CTGC에서 5.1%로, 및 기타로 인텔을 유발할 수 있었다 (도 20). 흥미롭게도, T152R은 TTTC에서 34.9%로, TTTA에서 10.2%로, 및 TTTG에서 6.2%로 인텔을 유발함으로써 wtLbCas12a의 TTTV 인식을 보유하였다. 이는 또한 8.3%로 인텔을 유발하는 TTTT 인식을 선택하였다. 총 256개의 가능한 4 뉴클레오티드 PAM 중 단지 22개를 T152R에 대해 293T 세포 (또는 9%)에서 시험하였다. T152R에 의해 인식되고 진핵생물 세포에서 인텔을 유발할 수 있는 많은 다른 PAM이 있을 가능성이 있다.
- [0305] 도 21에 나타내어진 바와 같이, 28개의 PAM 표적 중 22개는 0.1%의 배경 초과 활성을 가졌으며, 이는 시험된 PAM의 79%가 이 효소에 의해 인식되고 절단되었지만, 가끔은 보다 적은 것이 일부 적용을 위해 바람직할 수 있음을 지시한다. 시험된 6개의 PAM은 선택된 3개의 표적에 대한 배경 초과 임의의 편집을 갖지 않았다. 3개의 TTTV 표적은 모두 여전히 TTTC, TTTG, 및 TTTA에 대해 각각 15.6%, 6.2%, 및 5.8%로 양호한 활성을 가졌다. 1% 초과 인텔 형성을 갖는 다른 PAM 서열은 ATTA (3.5%), TTTT (3.2%), TGTC (1.8%), AGCG (1.8%), AGTC (1.6%), AGCA (1.4%), 및 GGTC (1.1%)를 포함하였다. 이 점 돌연변이를 PAM-SCANR 실험에서 T152R 및/또는 K595Y와 조합으로 사용하여 폭넓게 다양한 PAM 인식을 생성하였지만, 그 자신 상에서 이는 그 검정을 사용하여 상대적으로 소수의 PAM에 결합하였다. 단독이라기 보다는 장래의 이중 돌연변이에서 사용하여 HEK293T 세포에서 인텔을 생성하는 것이 우수한 선택일 수 있다. 다른 2개의 점 돌연변이와 유사하게, 가능한 256개의 4 뉴클레오티드 PAM 중 단지 28개를 시험하였고 (11%), 이 돌연변이체는 여기서 시험되지 않은 PAM 또는 표적을 인식할 수 있음이 가능하다.
- [0306] HEK293T 인텔 및 PAM-SCANR 결합 사이의 상관관계
- [0307] T152R 및 K595Y에 대해 관찰된 최대 인텔 백분율 및 PAM-SCANR 점수 사이에 상관관계가 관찰되었다 (도 22a 내지 22b). 도 22a 내지 22b는 LbCas12a-T152R (도 22a) 및 LbCas12a-K595Y (도 22b)에 대한 % 인텔 (최대) 및 정규화된 박테리아 PAM-SCANR 점수 사이에 선형 상관관계를 나타낸다.
- [0308] 특히, 시험된 1.5 초과 PAM-SCANR 점수를 갖는 임의의 점 돌연변이는 293T 세포에서 5% 초과 비율로 인텔을 생성하였다. 이는 1.5 초과 정규화된 점수 (본 발명자들의 엄격한 1.67 컷오프라기 보다는)를 갖는 PAM-SCANR 실험에서 시험된 임의의 돌연변이가 대부분의 진핵생물 적용에 유용한 비율로 인텔을 생성할 수 있는 가능성이 있음을 시사한다.
- [0309] 상기는 본 발명을 예시하며, 그의 제한인 것으로 해석되지 않아야 한다. 본 발명은 청구항의 등가물이 그 안에

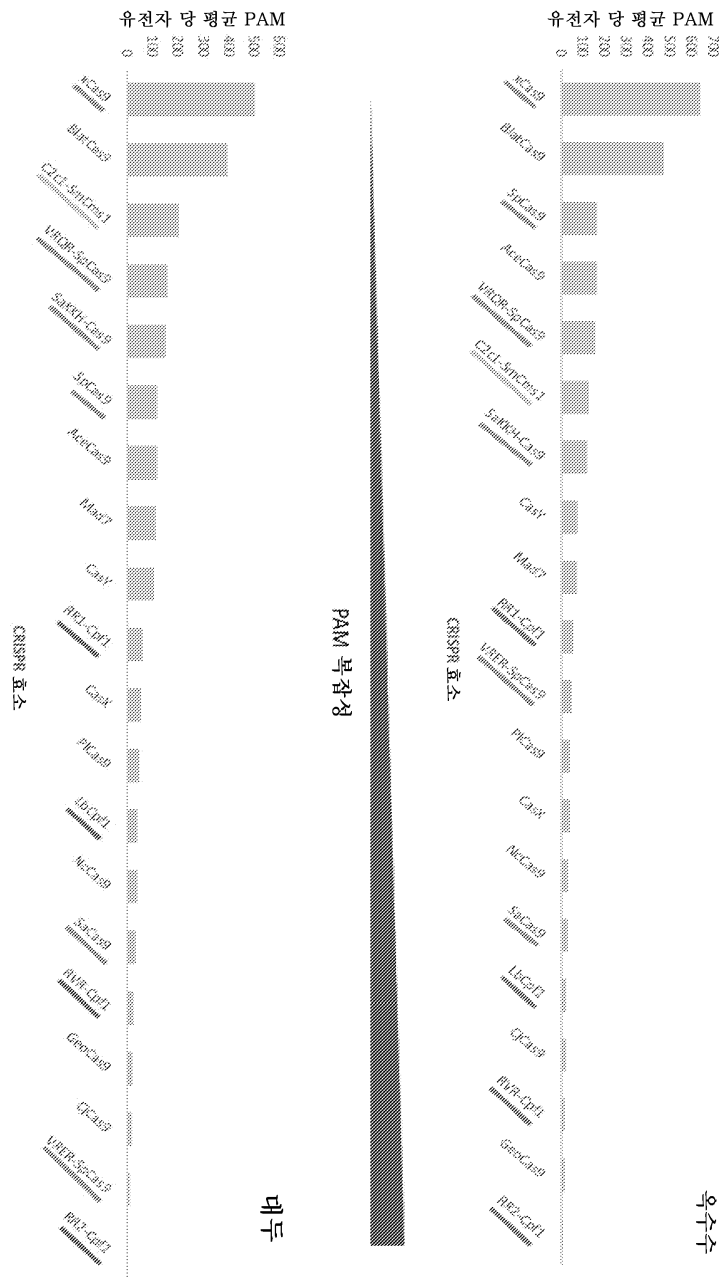
포함되는 하기 청구항에 의해 한정된다.

도면

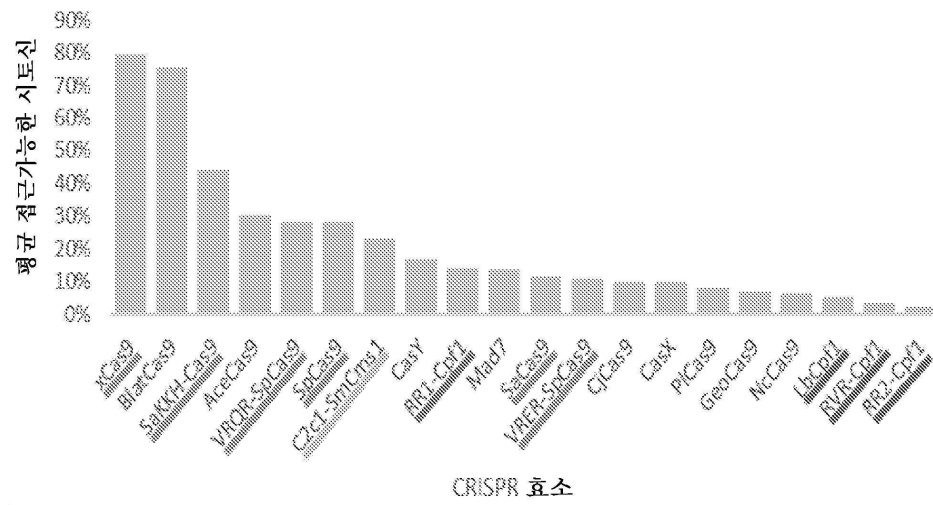
도면1



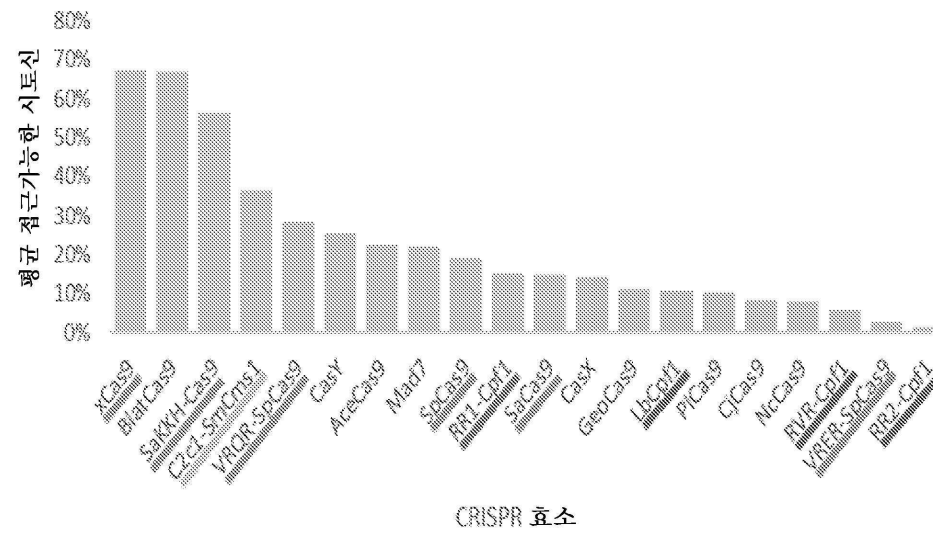
도면2



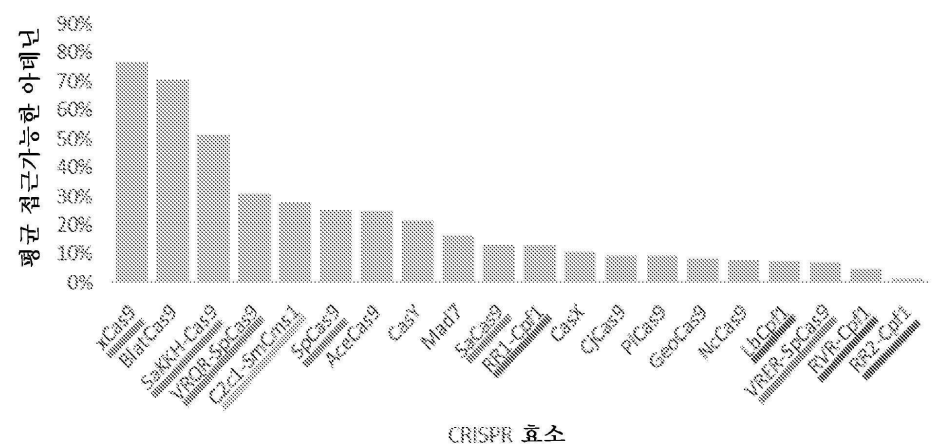
도면3a



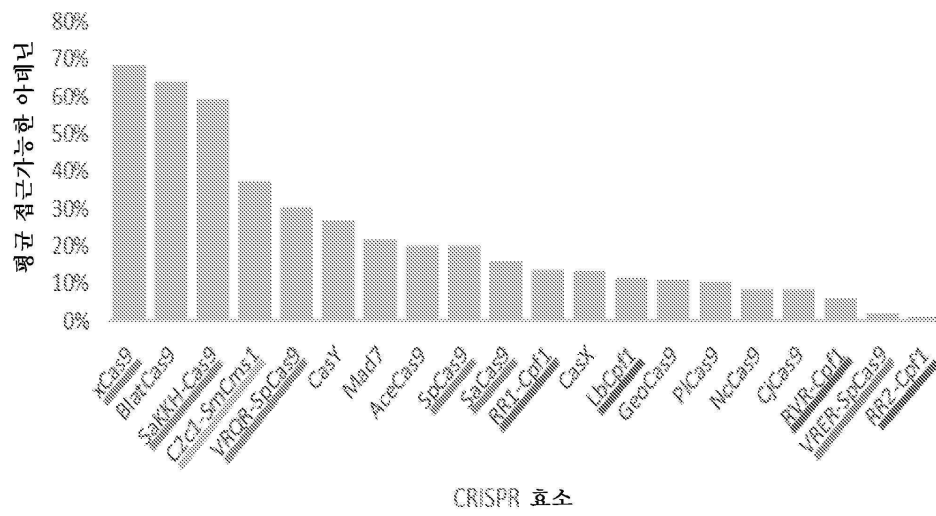
도면3b



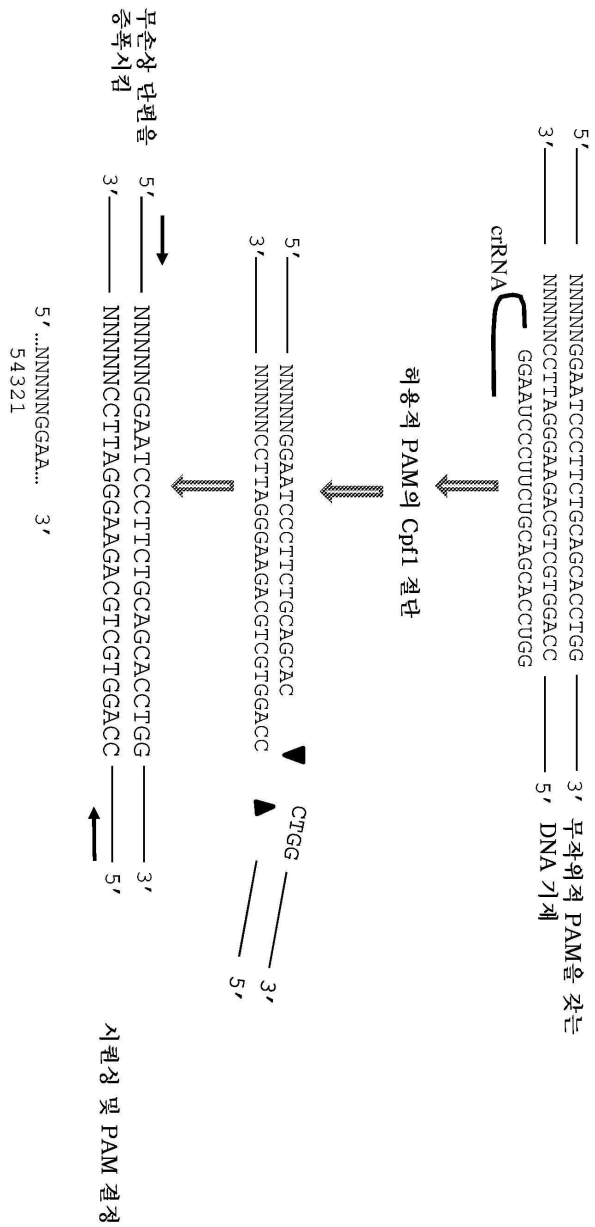
도면3c



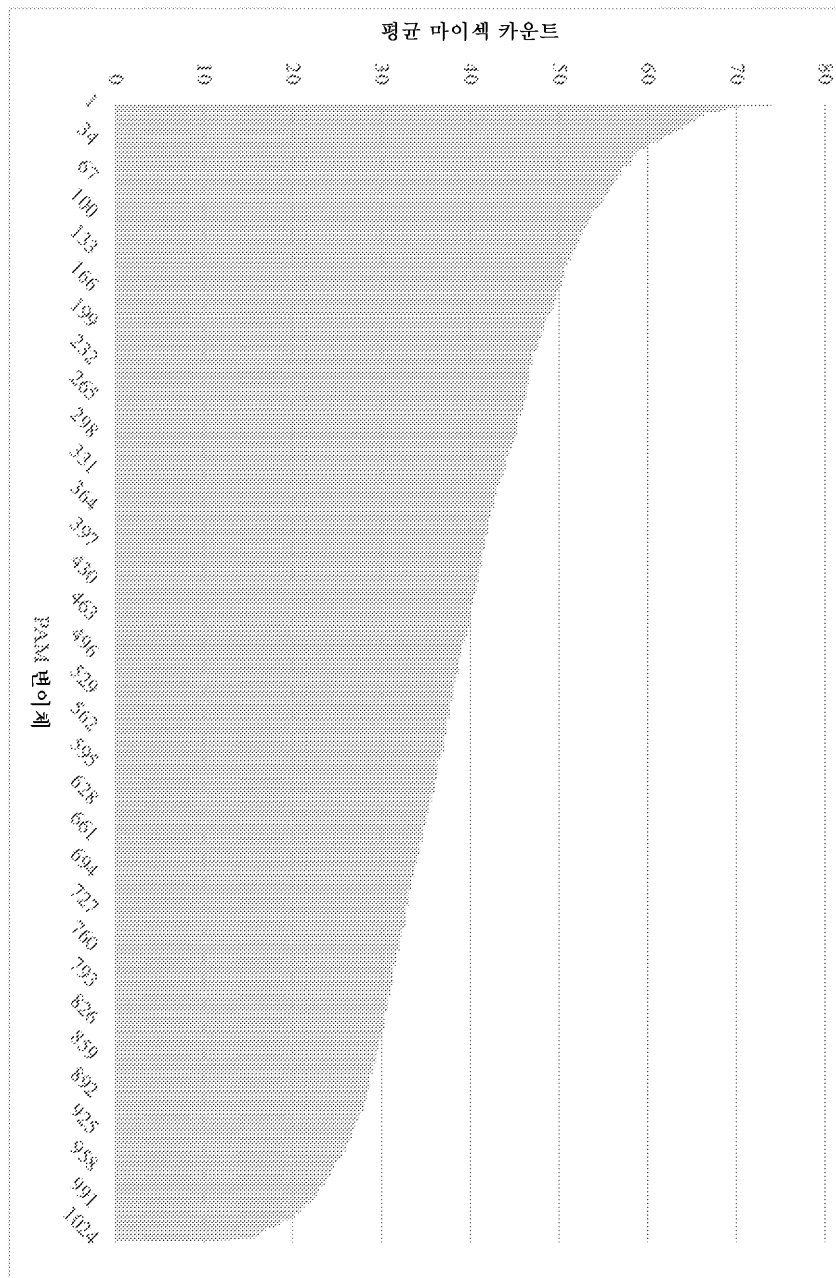
도면3d



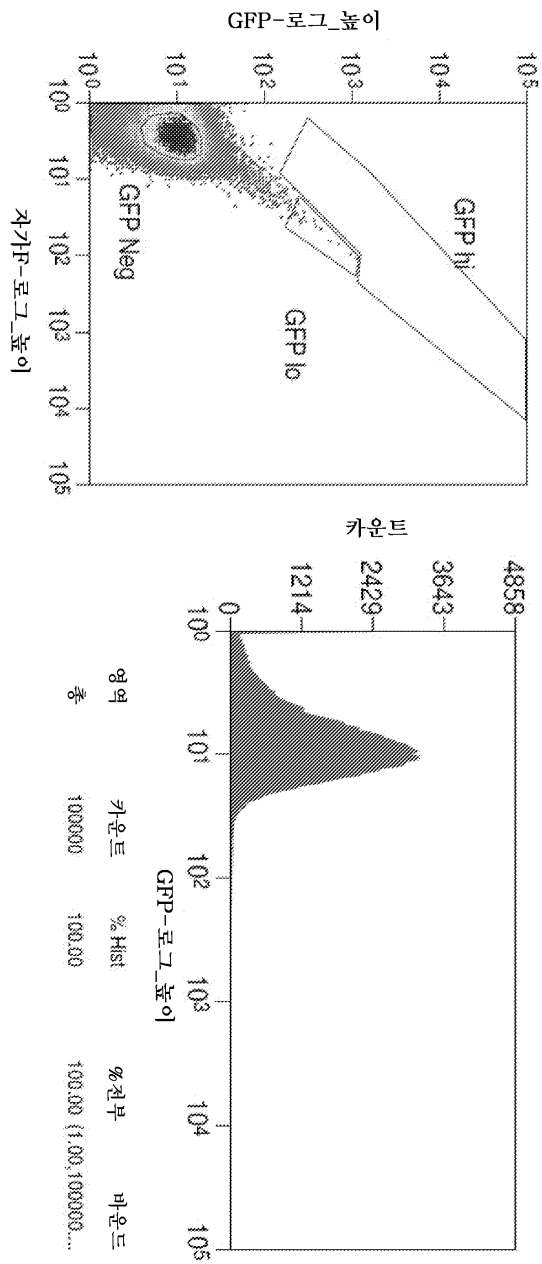
도면4



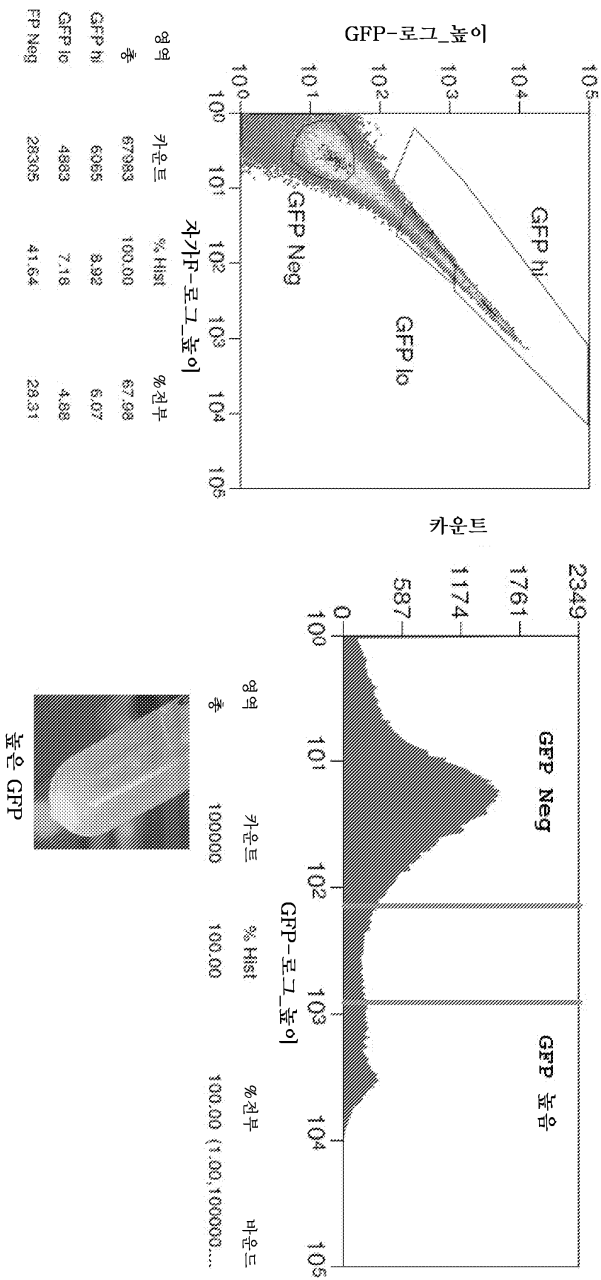
도면5



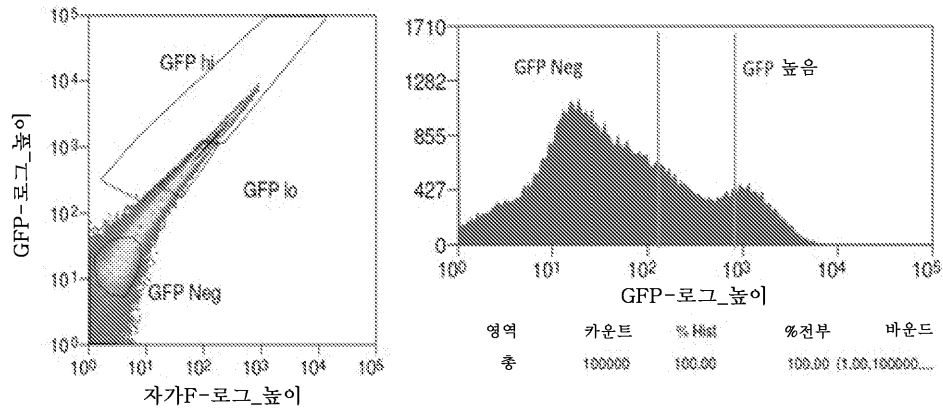
도면6



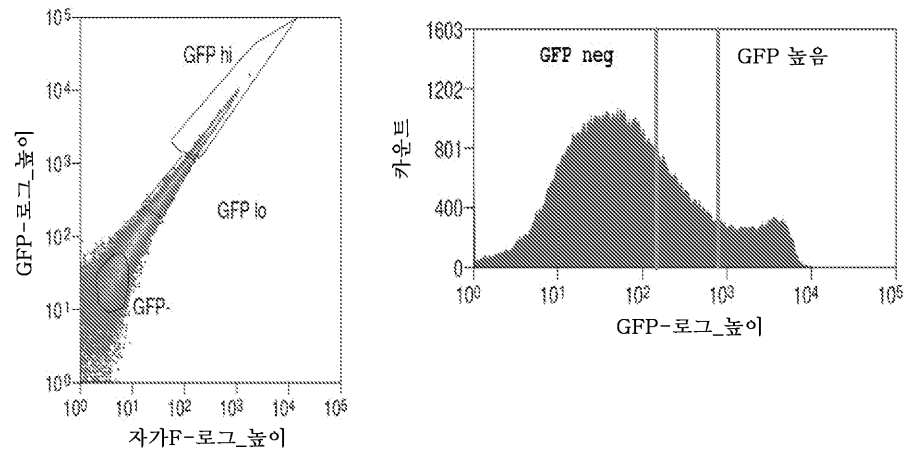
도면7



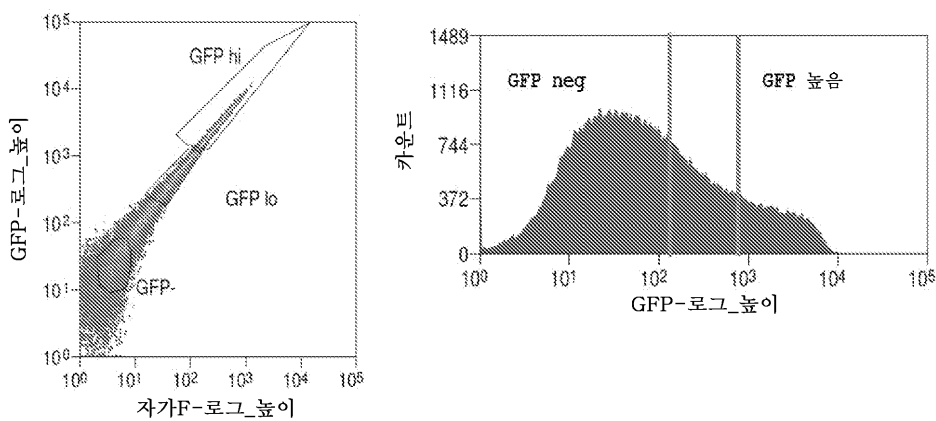
도면8



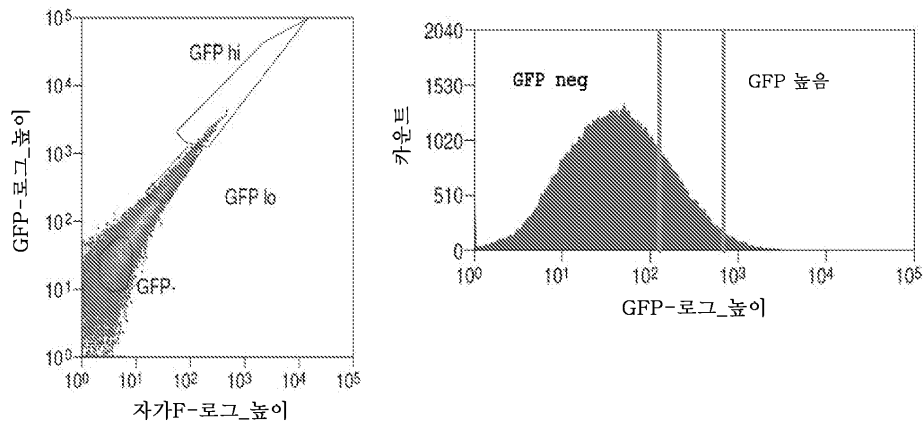
도면9



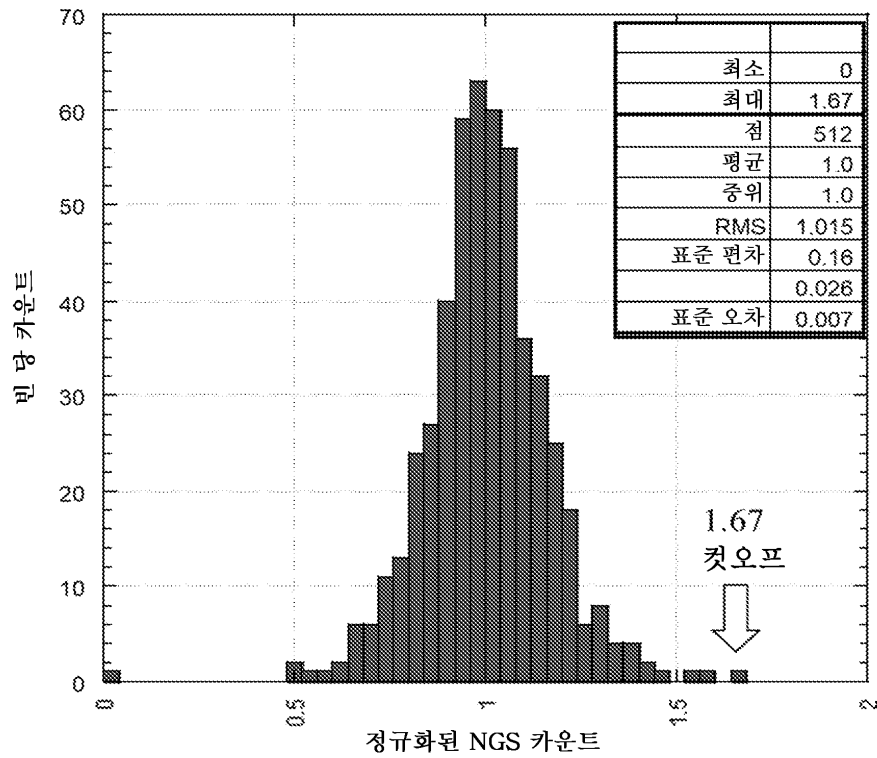
도면10



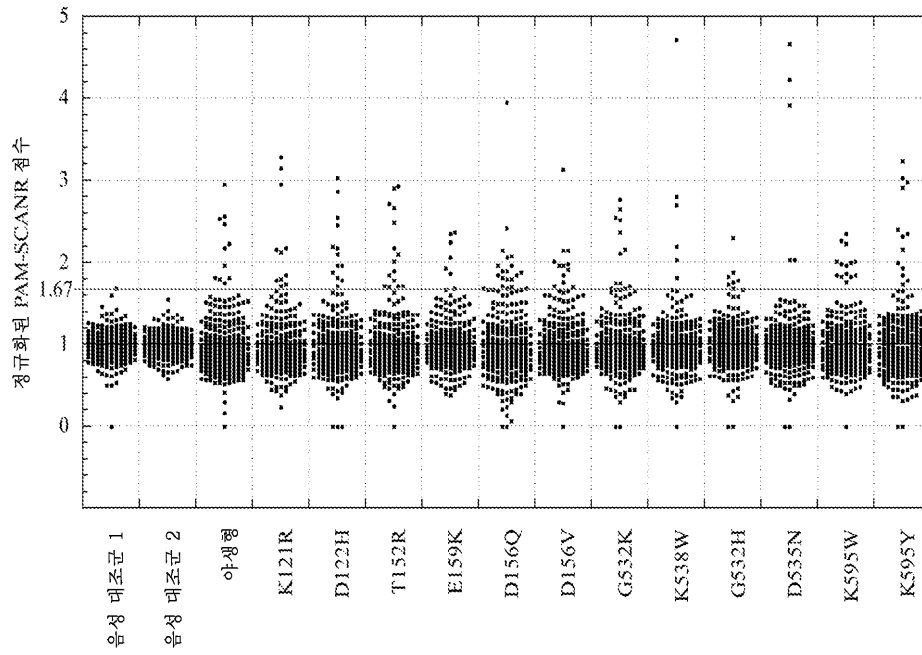
도면11



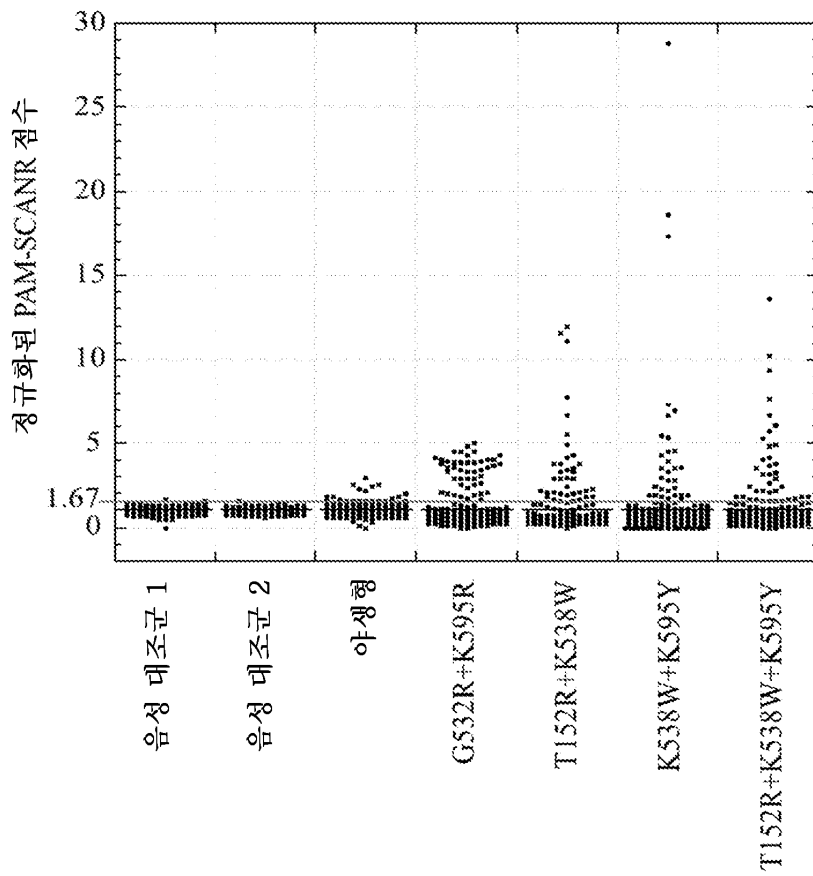
도면12



도면13



도면14



도면15

PAM	K538W	K595Y	K538W+K595Y
AATG			
ACCC			
ACGC			
AGCT			
ATTA			
ATTG			
CCCA			
CCCC			
CCCG			
CCGC			
CTCC			
CTTA			
CTTC			
CTTG			
GCCA			
GCCC			
GCCG			
GCGC			
GCTA			
GTCC			
GTTA			
GTTC			
GTTG			
TCCA			
TCCC			
TCCG			
TCGC			
TCGG			
TCTA			
TCTC			
TCTG			
TTCA			
TTCC			
TTCG			
TTGC			
TTGG			
TTTA			
TTTC			
TTTG			

도면16

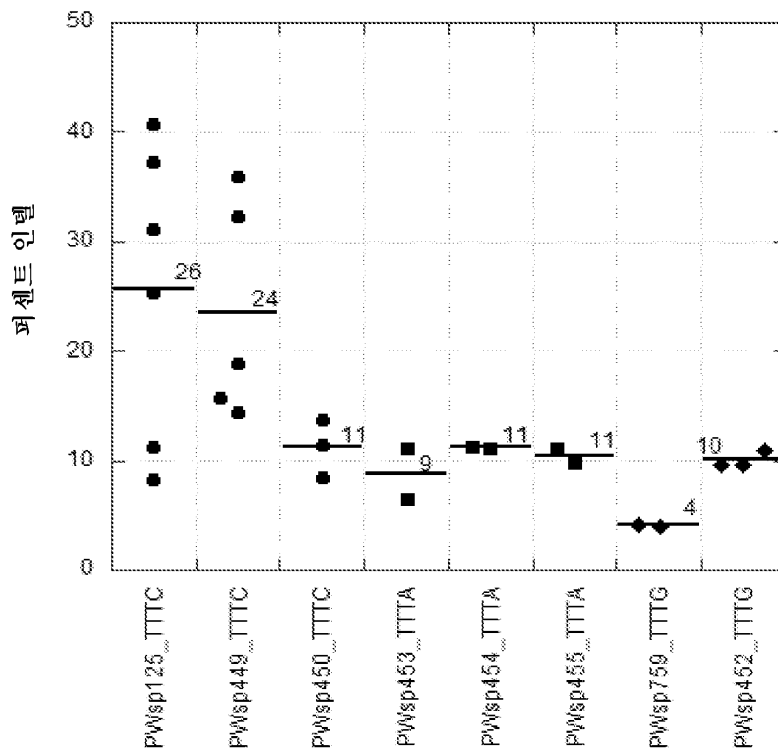
PAM	T152R	K538W	K595Y	T152R+K538W	K538W+K595Y	T152R+K538W+K595Y
AACC						
AAGC						
AATG						
ACCA						
ACCC						
ACCG						
ACGC						
AGAG						
AGCA						
AGCC						
AGCT						
AGGC						
ATTA						
ATTC						
ATTG						
CAGC						
CCCA						
CCCC						
CCCG						
CCGC						
CGGC						
CTCC						
CTGC						
CTTA						
CTTC						
CTTG						
GACA						
GACC						
GCCA						
GCCC						
GCCG						
GCGC						
GCTA						
GGCA						
GGCC						
GGGC						
GGGG						
GTCA						
GTCC						
GTTA						
GTTC						
GTTG						
TACA						
TACC						
TACG						
TAGC						
TCCA						
TCCC						
TCCG						
TCGC						
TCGG						
TCTA						
TCTC						
TCTG						
TGCA						
TGCC						
TGGC						
TTCA						
TTCC						
TTCG						
TTGC						
TTGG						
TTTA						
TTTC						
TTTG						
TTTT						

도면17

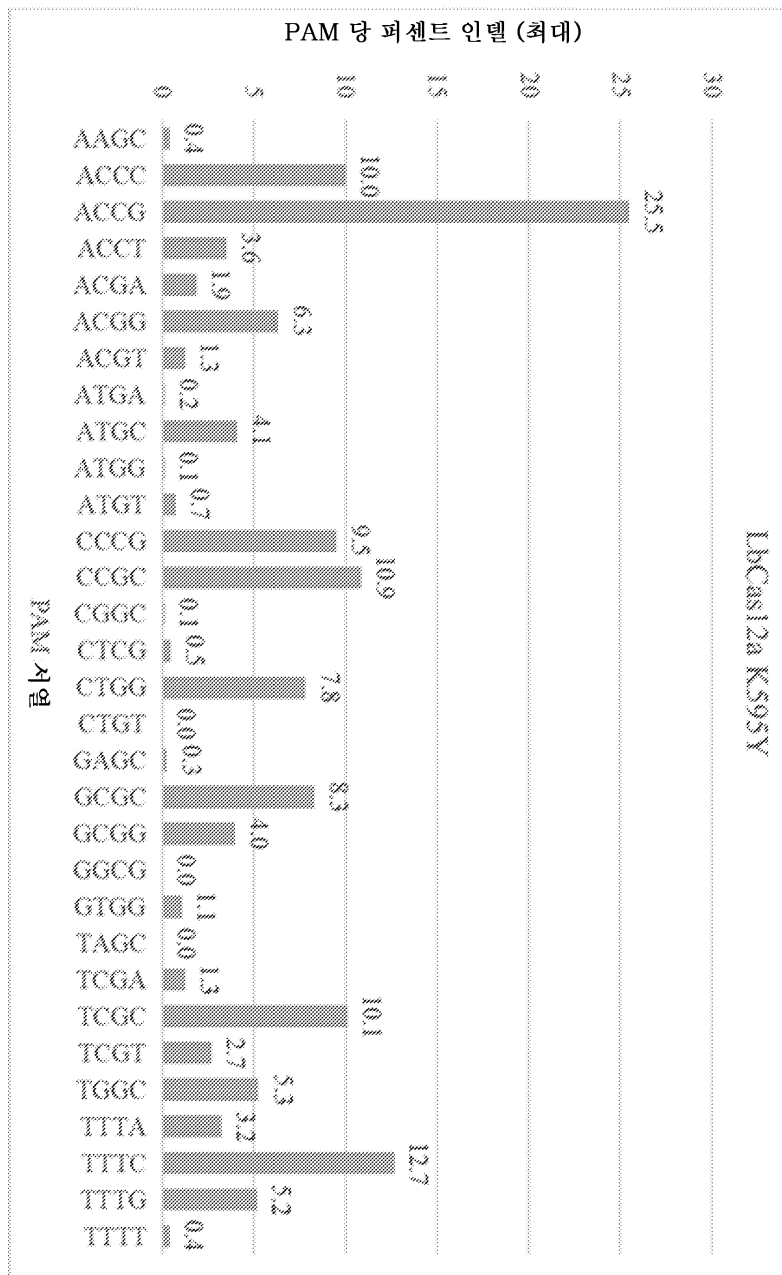
PAM	K595Y	PAM 고갈 점수
ACCC		18.7
ACGC		13.8
CCCA		11.5
CCGC		16.7
GCGC		17.5
TCCA		16.9
TCCC		18.6
TCCG		16.2
TCGC		13.3
TCGG		14.7
TTCC		11.3
TTCG		17.8
TTGC		11.5

PAM	T152R	PAM 고갈 점수
AGAG		
ATTG		10.8
CTTA		10.9
CTTC		16.1
CTTG		9.5
GCCA		10.4
GTTA		10.3
GTTC		9.9
GTTG		12.7
TCCA		10.7
TCTA		10.4
TTTT		10.4

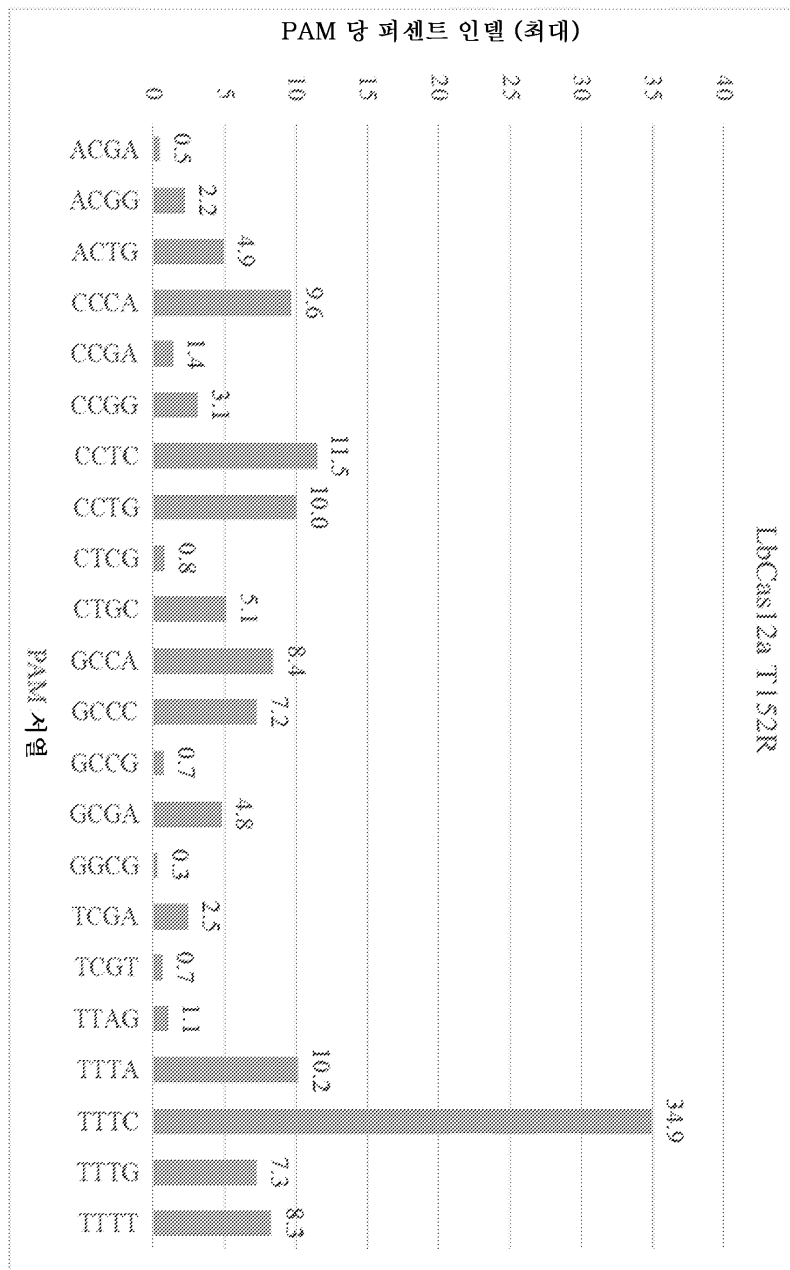
도면18



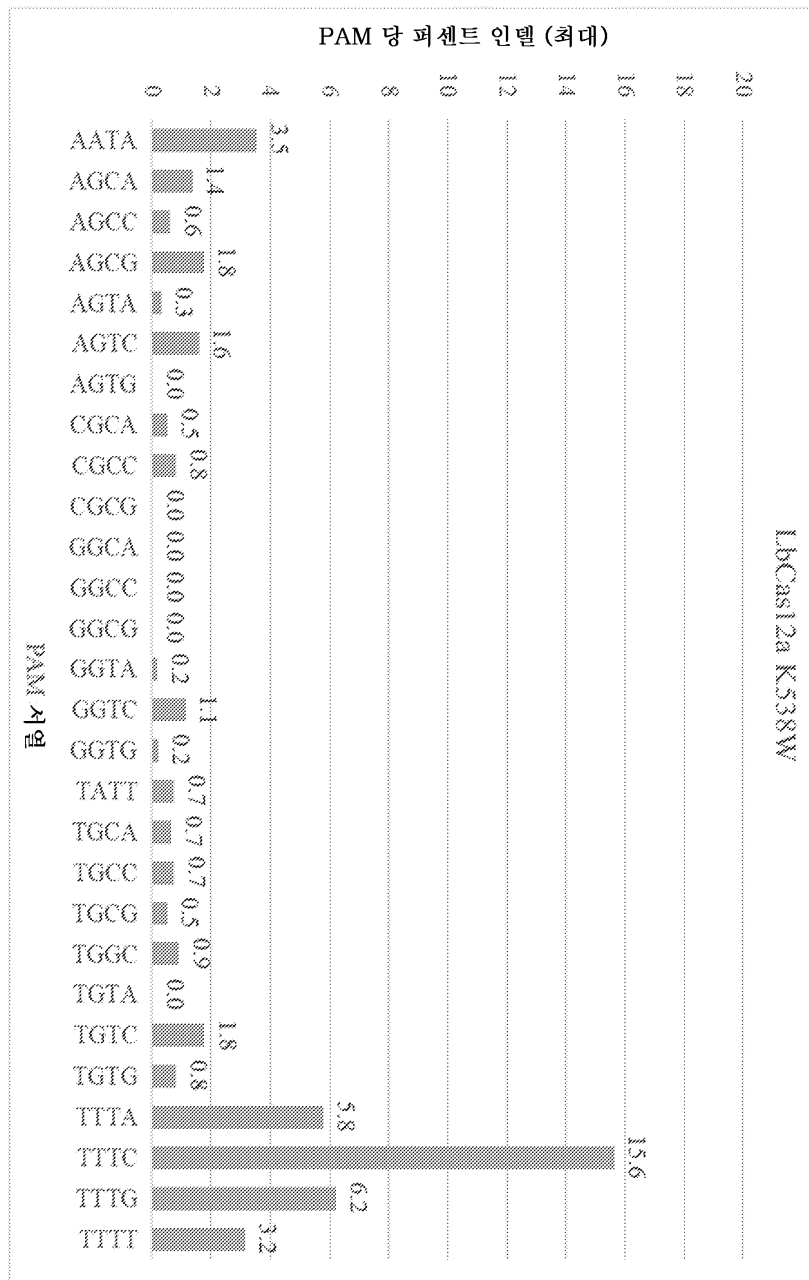
도면19



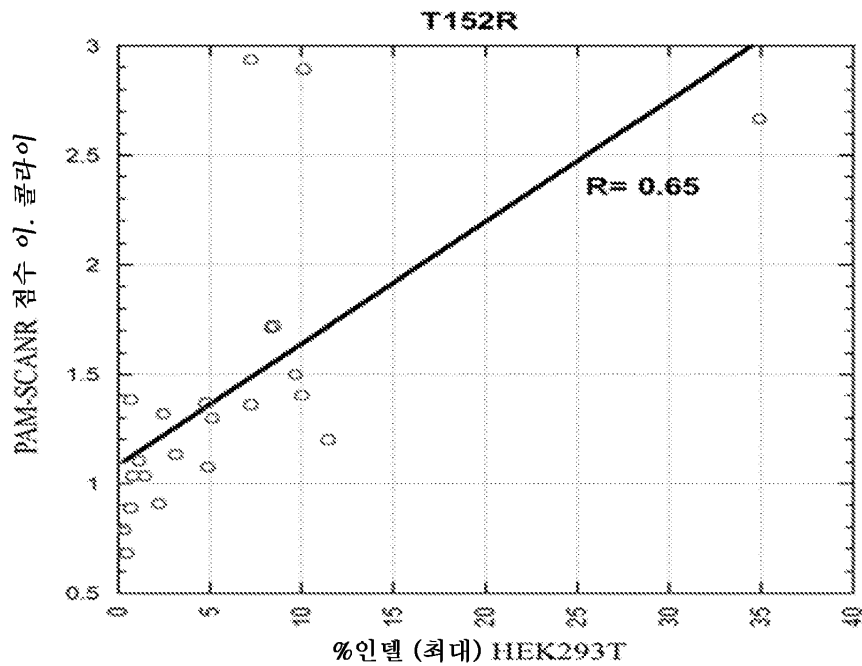
도면20



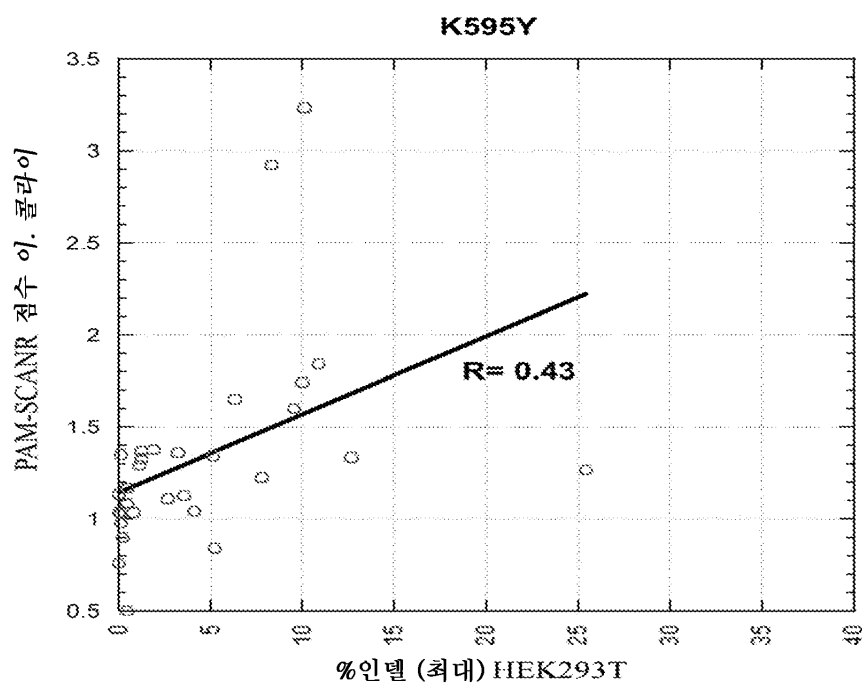
도면21



도면22a



도면22b



서열 목록

SEQUENCE LISTING

<110> Pairwise Plants Services, Inc.

Watts, Joseph Matthew

Jali, Sathya Sheela

<120> VARIANTS OF CAS12A NUCLEASES AND METHODS OF MAKING AND USE
THEREOF

<130> 1499.7.WO

<150> US 62/916,392

<151> 2019-10-17

<160> 62

<170> PatentIn version 3.5

<210> 1

<211> 1228

<212> PRT

<213> Unknown

<220><223> Lachnospiraceae bacterium

<400> 1

Met Ser Lys Leu Glu Lys Phe Thr Asn Cys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Lys Ala Ile Pro Val Gly Lys Thr Gln Glu Asn Ile Asp
20 25 30

Asn Lys Arg Leu Leu Val Glu Asp Glu Lys Arg Ala Glu Asp Tyr Lys
35 40 45

Gly Val Lys Lys Leu Leu Asp Arg Tyr Tyr Leu Ser Phe Ile Asn Asp
50 55 60

Val Leu His Ser Ile Lys Leu Lys Asn Leu Asn Asn Tyr Ile Ser Leu
65 70 75 80

Phe Arg Lys Lys Thr Arg Thr Glu Lys Glu Asn Lys Glu Leu Glu Asn
85 90 95

Leu Glu Ile Asn Leu Arg Lys Glu Ile Ala Lys Ala Phe Lys Gly Asn
100 105 110

Glu Gly Tyr Lys Ser Leu Phe Lys Lys Asp Ile Ile Glu Thr Ile Leu
115 120 125

Pro Glu Phe Leu Asp Asp Lys Asp Glu Ile Ala Leu Val Asn Ser Phe
130 135 140

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

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

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

 Phe Val Lys Asn Gly Pro Ala Ile Ser Thr Ile Ser Lys Asp Ile Phe
 340 345 350
 Gly Glu Trp Asn Val Ile Arg Asp Lys Trp Asn Ala Glu Tyr Asp Asp
 355 360 365
 Ile His Leu Lys Lys Lys Ala Val Val Thr Glu Lys Tyr Glu Asp Asp
 370 375 380
 Arg Arg Lys Ser Phe Lys Lys Ile Gly Ser Phe Ser Leu Glu Gln Leu

385			390				395				400				
Gln	Glu	Tyr	Ala	Asp	Ala	Asp	Leu	Ser	Val	Val	Glu	Lys	Leu	Lys	Glu
405				410				415							
Ile	Ile	Ile	Gln	Lys	Val	Asp	Glu	Ile	Tyr	Lys	Val	Tyr	Gly	Ser	Ser
420				425				430							
Glu	Lys	Leu	Phe	Asp	Ala	Asp	Phe	Val	Leu	Glu	Lys	Ser	Leu	Lys	Lys
435				440				445							
Asn	Asp	Ala	Val	Val	Ala	Ile	Met	Lys	Asp	Leu	Leu	Asp	Ser	Val	Lys
450				455				460							
Ser	Phe	Glu	Asn	Tyr	Ile	Lys	Ala	Phe	Phe	Gly	Glu	Gly	Lys	Glu	Thr
465				470				475				480			
Asn	Arg	Asp	Glu	Ser	Phe	Tyr	Gly	Asp	Phe	Val	Leu	Ala	Tyr	Asp	Ile
485				490				495							
Leu	Leu	Lys	Val	Asp	His	Ile	Tyr	Asp	Ala	Ile	Arg	Asn	Tyr	Val	Thr
500				505				510							
Gln	Lys	Pro	Tyr	Ser	Lys	Asp	Lys	Phe	Lys	Leu	Tyr	Phe	Gln	Asn	Pro
515				520				525							
Gln	Phe	Met	Gly	Gly	Trp	Asp	Lys	Asp	Lys	Glu	Thr	Asp	Tyr	Arg	Ala
530				535				540							
Thr	Ile	Leu	Arg	Tyr	Gly	Ser	Lys	Tyr	Tyr	Leu	Ala	Ile	Met	Asp	Lys
545				550				555				560			
Lys	Tyr	Ala	Lys	Cys	Leu	Gln	Lys	Ile	Asp	Lys	Asp	Asp	Val	Asn	Gly
565				570				575							
Asn	Tyr	Glu	Lys	Ile	Asn	Tyr	Lys	Leu	Leu	Pro	Gly	Pro	Asn	Lys	Met
580				585				590							
Leu	Pro	Lys	Val	Phe	Phe	Ser	Lys	Lys	Trp	Met	Ala	Tyr	Tyr	Asn	Pro
595				600				605							
Ser	Glu	Asp	Ile	Gln	Lys	Ile	Tyr	Lys	Asn	Gly	Thr	Phe	Lys	Lys	Gly
610				615				620							
Asp	Met	Phe	Asn	Leu	Asn	Asp	Cys	His	Lys	Leu	Ile	Asp	Phe	Phe	Lys
625				630				635				640			

Asp Ser Ile Ser Arg Tyr Pro Lys Trp Ser Asn Ala Tyr Asp Phe Asn
645 650 655

Phe Ser Glu Thr Glu Lys Tyr Lys Asp Ile Ala Gly Phe Tyr Arg Glu
660 665 670

Val Glu Glu Gln Gly Tyr Lys Val Ser Phe Glu Ser Ala Ser Lys Lys
675 680 685

Glu Val Asp Lys Leu Val Glu Glu Gly Lys Leu Tyr Met Phe Gln Ile
690 695 700

Tyr Asn Lys Asp Phe Ser Asp Lys Ser His Gly Thr Pro Asn Leu His
705 710 715 720

Thr Met Tyr Phe Lys Leu Leu Phe Asp Glu Asn Asn His Gly Gln Ile
725 730 735

Arg Leu Ser Gly Gly Ala Glu Leu Phe Met Arg Arg Ala Ser Leu Lys
740 745 750

Lys Glu Glu Leu Val Val His Pro Ala Asn Ser Pro Ile Ala Asn Lys
755 760 765

Asn Pro Asp Asn Pro Lys Lys Thr Thr Thr Leu Ser Tyr Asp Val Tyr
770 775 780

Lys Asp Lys Arg Phe Ser Glu Asp Gln Tyr Glu Leu His Ile Pro Ile
785 790 795 800

Ala Ile Asn Lys Cys Pro Lys Asn Ile Phe Lys Ile Asn Thr Glu Val
805 810 815

Arg Val Leu Leu Lys His Asp Asp Asn Pro Tyr Val Ile Gly Ile Asp
820 825 830

Arg Gly Glu Arg Asn Leu Leu Tyr Ile Val Val Val Asp Gly Lys Gly
835 840 845

Asn Ile Val Glu Gln Tyr Ser Leu Asn Glu Ile Ile Asn Asn Phe Asn
850 855 860

Gly Ile Arg Ile Lys Thr Asp Tyr His Ser Leu Leu Asp Lys Lys Glu
865 870 875 880

Lys Glu Arg Phe Glu Ala Arg Gln Asn Trp Thr Ser Ile Glu Asn Ile

885 890 895
 Lys Glu Leu Lys Ala Gly Tyr Ile Ser Gln Val Val His Lys Ile Cys
 900 905 910

 Glu Leu Val Glu Lys Tyr Asp Ala Val Ile Ala Leu Glu Asp Leu Asn
 915 920 925
 Ser Gly Phe Lys Asn Ser Arg Val Lys Val Glu Lys Gln Val Tyr Gln
 930 935 940
 Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Met Val Asp Lys
 945 950 955 960
 Lys Ser Asn Pro Cys Ala Thr Gly Gly Ala Leu Lys Gly Tyr Gln Ile
 965 970 975

 Thr Asn Lys Phe Glu Ser Phe Lys Ser Met Ser Thr Gln Asn Gly Phe
 980 985 990
 Ile Phe Tyr Ile Pro Ala Trp Leu Thr Ser Lys Ile Asp Pro Ser Thr
 995 1000 1005
 Gly Phe Val Asn Leu Leu Lys Thr Lys Tyr Thr Ser Ile Ala Asp
 1010 1015 1020
 Ser Lys Lys Phe Ile Ser Ser Phe Asp Arg Ile Met Tyr Val Pro
 1025 1030 1035

 Glu Glu Asp Leu Phe Glu Phe Ala Leu Asp Tyr Lys Asn Phe Ser
 1040 1045 1050
 Arg Thr Asp Ala Asp Tyr Ile Lys Lys Trp Lys Leu Tyr Ser Tyr
 1055 1060 1065
 Gly Asn Arg Ile Arg Ile Phe Arg Asn Pro Lys Lys Asn Asn Val
 1070 1075 1080
 Phe Asp Trp Glu Glu Val Cys Leu Thr Ser Ala Tyr Lys Glu Leu
 1085 1090 1095

 Phe Asn Lys Tyr Gly Ile Asn Tyr Gln Gln Gly Asp Ile Arg Ala
 1100 1105 1110
 Leu Leu Cys Glu Gln Ser Asp Lys Ala Phe Tyr Ser Ser Phe Met
 1115 1120 1125

Ala Leu Met Ser Leu Met Leu Gln Met Arg Asn Ser Ile Thr Gly
1130 1135 1140

Arg Thr Asp Val Asp Phe Leu Ile Ser Pro Val Lys Asn Ser Asp
1145 1150 1155

Gly Ile Phe Tyr Asp Ser Arg Asn Tyr Glu Ala Gln Glu Asn Ala
1160 1165 1170

Ile Leu Pro Lys Asn Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala
1175 1180 1185

Arg Lys Val Leu Trp Ala Ile Gly Gln Phe Lys Lys Ala Glu Asp
1190 1195 1200

Glu Lys Leu Asp Lys Val Lys Ile Ala Ile Ser Asn Lys Glu Trp
1205 1210 1215

Leu Glu Tyr Ala Gln Thr Ser Val Lys His
1220 1225

<210> 2

<211> 1307

<212> PRT

<213> Acidaminococcus sp.

<400> 2

Met Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Lys His Ile Gln
20 25 30

Glu Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys
35 40 45

Glu Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln
50 55 60

Cys Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile
65 70 75 80

Asp Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile
85 90 95

Glu Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly

100	105	110	
Arg Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile			
115	120	125	
Tyr Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys			
130	135	140	
Gln Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg			
145	150	155	160
Ser Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg			
165	170	175	
Lys Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg			
180	185	190	
Ile Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe			
195	200	205	
Thr Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn			
210	215	220	
Val Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val			
225	230	235	240
Phe Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp			
245	250	255	
Leu Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu			
260	265	270	
Lys Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn			
275	280	285	
Asp Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro			
290	295	300	
Leu Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu			
305	310	315	320
Glu Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr			
325	330	335	
Lys Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu			
340	345	350	

Phe Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His
 355 360 365

 Lys Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr
 370 375 380
 Leu Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys
 385 390 395 400
 Ile Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu
 405 410 415
 Asp Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser
 420 425 430

 Glu Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala
 435 440 445
 Ala Leu Asp Gln Pro Leu Pro Thr Thr Leu Lys Lys Gln Glu Glu Lys
 450 455 460
 Glu Ile Leu Lys Ser Gln Leu Asp Ser Leu Leu Gly Leu Tyr His Leu
 465 470 475 480
 Leu Asp Trp Phe Ala Val Asp Glu Ser Asn Glu Val Asp Pro Glu Phe
 485 490 495

 Ser Ala Arg Leu Thr Gly Ile Lys Leu Glu Met Glu Pro Ser Leu Ser
 500 505 510
 Phe Tyr Asn Lys Ala Arg Asn Tyr Ala Thr Lys Lys Pro Tyr Ser Val
 515 520 525
 Glu Lys Phe Lys Leu Asn Phe Gln Met Pro Thr Leu Ala Ser Gly Trp
 530 535 540
 Asp Val Asn Lys Glu Lys Asn Asn Gly Ala Ile Leu Phe Val Lys Asn
 545 550 555 560

 Gly Leu Tyr Tyr Leu Gly Ile Met Pro Lys Gln Lys Gly Arg Tyr Lys
 565 570 575
 Ala Leu Ser Phe Glu Pro Thr Glu Lys Thr Ser Glu Gly Phe Asp Lys
 580 585 590
 Met Tyr Tyr Asp Tyr Phe Pro Asp Ala Ala Lys Met Ile Pro Lys Cys

595 600 605
 Ser Thr Gln Leu Lys Ala Val Thr Ala His Phe Gln Thr His Thr Thr
 610 615 620

 Pro Ile Leu Leu Ser Asn Asn Phe Ile Glu Pro Leu Glu Ile Thr Lys
 625 630 635 640
 Glu Ile Tyr Asp Leu Asn Asn Pro Glu Lys Glu Pro Lys Lys Phe Gln
 645 650 655
 Thr Ala Tyr Ala Lys Lys Thr Gly Asp Gln Lys Gly Tyr Arg Glu Ala
 660 665 670
 Leu Cys Lys Trp Ile Asp Phe Thr Arg Asp Phe Leu Ser Lys Tyr Thr
 675 680 685

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

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

 Pro Ile Pro Asp Thr Leu Tyr Gln Glu Leu Tyr Asp Tyr Val Asn His
 820 825 830
 Arg Leu Ser His Asp Leu Ser Asp Glu Ala Arg Ala Leu Leu Pro Asn
 835 840 845

Val Ile Thr Lys Glu Val Ser His Glu Ile Ile Lys Asp Arg Arg Phe
 850 855 860
 Thr Ser Asp Lys Phe Phe Phe His Val Pro Ile Thr Leu Asn Tyr Gln
 865 870 875 880

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

Asp Asn Arg Glu Lys Glu Arg Val Ala Ala Arg Gln Ala Trp Ser Val
 945 950 955 960
 Val Gly Thr Ile Lys Asp Leu Lys Gln Gly Tyr Leu Ser Gln Val Ile
 965 970 975
 His Glu Ile Val Asp Leu Met Ile His Tyr Gln Ala Val Val Val Leu
 980 985 990
 Glu Asn Leu Asn Phe Gly Phe Lys Ser Lys Arg Thr Gly Ile Ala Glu
 995 1000 1005

Lys Ala Val Tyr Gln Gln Phe Glu Lys Met Leu Ile Asp Lys Leu
 1010 1015 1020
 Asn Cys Leu Val Leu Lys Asp Tyr Pro Ala Glu Lys Val Gly Gly
 1025 1030 1035
 Val Leu Asn Pro Tyr Gln Leu Thr Asp Gln Phe Thr Ser Phe Ala
 1040 1045 1050
 Lys Met Gly Thr Gln Ser Gly Phe Leu Phe Tyr Val Pro Ala Pro
 1055 1060 1065

Tyr Thr Ser Lys Ile Asp Pro Leu Thr Gly Phe Val Asp Pro Phe
 1070 1075 1080
 Val Trp Lys Thr Ile Lys Asn His Glu Ser Arg Lys His Phe Leu

1085	1090	1095
Glu Gly Phe Asp Phe Leu His Tyr Asp Val Lys Thr Gly Asp Phe		
1100	1105	1110
Ile Leu His Phe Lys Met Asn Arg Asn Leu Ser Phe Gln Arg Gly		
1115	1120	1125
Leu Pro Gly Phe Met Pro Ala Trp Asp Ile Val Phe Glu Lys Asn		
1130	1135	1140
Glu Thr Gln Phe Asp Ala Lys Gly Thr Pro Phe Ile Ala Gly Lys		
1145	1150	1155
Arg Ile Val Pro Val Ile Glu Asn His Arg Phe Thr Gly Arg Tyr		
1160	1165	1170
Arg Asp Leu Tyr Pro Ala Asn Glu Leu Ile Ala Leu Leu Glu Glu		
1175	1180	1185
Lys Gly Ile Val Phe Arg Asp Gly Ser Asn Ile Leu Pro Lys Leu		
1190	1195	1200
Leu Glu Asn Asp Asp Ser His Ala Ile Asp Thr Met Val Ala Leu		
1205	1210	1215
Ile Arg Ser Val Leu Gln Met Arg Asn Ser Asn Ala Ala Thr Gly		
1220	1225	1230
Glu Asp Tyr Ile Asn Ser Pro Val Arg Asp Leu Asn Gly Val Cys		
1235	1240	1245
Phe Asp Ser Arg Phe Gln Asn Pro Glu Trp Pro Met Asp Ala Asp		
1250	1255	1260
Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Gln Leu Leu Leu		
1265	1270	1275
Asn His Leu Lys Glu Ser Lys Asp Leu Lys Leu Gln Asn Gly Ile		
1280	1285	1290
Ser Asn Gln Asp Trp Leu Ala Tyr Ile Gln Glu Leu Arg Asn		
1295	1300	1305

<210> 3

<211> 1241

<212> PRT

<213> *utyrvibrio proteoclasticus*

<400> 3

Met Leu Leu Tyr Glu Asn Tyr Thr Lys Arg Asn Gln Ile Thr Lys Ser

1 5 10 15

Leu Arg Leu Glu Leu Arg Pro Gln Gly Lys Thr Leu Arg Asn Ile Lys

20 25 30

Glu Leu Asn Leu Leu Glu Gln Asp Lys Ala Ile Tyr Ala Leu Leu Glu

35 40 45

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

50 55 60

Thr Leu Lys Asn Cys Glu Leu Ser Phe Glu Lys Leu Tyr Glu His Phe

65 70 75 80

Leu Ser Gly Asp Lys Lys Ala Tyr Ala Lys Glu Ser Glu Arg Leu Lys

85 90 95

Lys Glu Ile Val Lys Thr Leu Ile Lys Asn Leu Pro Glu Gly Ile Gly

100 105 110

Lys Ile Ser Glu Ile Asn Ser Ala Lys Tyr Leu Asn Gly Val Leu Tyr

115 120 125

Asp Phe Ile Asp Lys Thr His Lys Asp Ser Glu Glu Lys Gln Asn Ile

130 135 140

Leu Ser Asp Ile Leu Glu Thr Lys Gly Tyr Leu Ala Leu Phe Ser Lys

145 150 155 160

Phe Leu Thr Ser Arg Ile Thr Thr Leu Glu Gln Ser Met Pro Lys Arg

165 170 175

Val Ile Glu Asn Phe Glu Ile Tyr Ala Ala Asn Ile Pro Lys Met Gln

180 185 190

Asp Ala Leu Glu Arg Gly Ala Val Ser Phe Ala Ile Glu Tyr Glu Ser

195 200 205

Ile Cys Ser Val Asp Tyr Tyr Asn Gln Ile Leu Ser Gln Glu Asp Ile

210 215 220

Asp Ser Tyr Asn Arg Leu Ile Ser Gly Ile Met Asp Glu Asp Gly Ala

225 230 235 240
 Lys Glu Lys Gly Ile Asn Gln Thr Ile Ser Glu Lys Asn Ile Lys Ile

 245 250 255
 Lys Ser Glu His Leu Glu Glu Lys Pro Phe Arg Ile Leu Lys Gln Leu
 260 265 270
 His Lys Gln Ile Leu Glu Glu Arg Glu Lys Ala Phe Thr Ile Asp His
 275 280 285
 Ile Asp Ser Asp Glu Glu Val Val Gln Val Thr Lys Glu Ala Phe Glu
 290 295 300
 Gln Thr Lys Glu Gln Trp Glu Asn Ile Lys Lys Ile Asn Gly Phe Tyr

 305 310 315 320
 Ala Lys Asp Pro Gly Asp Ile Thr Leu Phe Ile Val Val Gly Pro Asn
 325 330 335
 Gln Thr His Val Leu Ser Gln Leu Ile Tyr Gly Glu His Asp Arg Ile
 340 345 350
 Arg Leu Leu Leu Glu Glu Tyr Glu Lys Asn Thr Leu Glu Val Leu Pro
 355 360 365
 Arg Arg Thr Lys Ser Glu Asp Ala Arg Tyr Asp Lys Phe Val Asn Ala

 370 375 380
 Val Pro Lys Lys Val Ala Lys Glu Ser His Thr Phe Asp Gly Leu Gln
 385 390 395 400
 Lys Met Thr Gly Asp Asp Arg Leu Phe Ile Leu Tyr Arg Asp Glu Leu
 405 410 415
 Ala Arg Asn Tyr Met Arg Ile Lys Glu Ala Tyr Gly Thr Phe Glu Arg
 420 425 430
 Asp Ile Leu Lys Ser Arg Arg Gly Ile Lys Gly Asn Arg Asp Val Gln

 435 440 445
 Glu Ser Leu Val Ser Phe Tyr Asp Glu Leu Thr Lys Phe Arg Ser Ala
 450 455 460
 Leu Arg Ile Ile Asn Ser Gly Asn Asp Glu Lys Ala Asp Pro Ile Phe
 465 470 475 480

Tyr Asn Thr Phe Asp Gly Ile Phe Glu Lys Ala Asn Arg Thr Tyr Lys
 485 490 495
 Ala Glu Asn Leu Cys Arg Asn Tyr Val Thr Lys Ser Pro Ala Asp Asp
 500 505 510
 Ala Arg Ile Met Ala Ser Cys Leu Gly Thr Pro Ala Arg Leu Arg Thr
 515 520 525
 His Trp Trp Asn Gly Glu Glu Asn Phe Ala Ile Asn Asp Val Ala Met
 530 535 540
 Ile Arg Arg Gly Asp Glu Tyr Tyr Tyr Phe Val Leu Thr Pro Asp Val
 545 550 555 560
 Lys Pro Val Asp Leu Lys Thr Lys Asp Glu Thr Asp Ala Gln Ile Phe
 565 570 575
 Val Gln Arg Lys Gly Ala Lys Ser Phe Leu Gly Leu Pro Lys Ala Leu
 580 585 590
 Phe Lys Cys Ile Leu Glu Pro Tyr Phe Glu Ser Pro Glu His Lys Asn
 595 600 605
 Asp Lys Asn Cys Val Ile Glu Glu Tyr Val Ser Lys Pro Leu Thr Ile
 610 615 620
 Asp Arg Arg Ala Tyr Asp Ile Phe Lys Asn Gly Thr Phe Lys Lys Thr
 625 630 635 640
 Asn Ile Gly Ile Asp Gly Leu Thr Glu Glu Lys Phe Lys Asp Asp Cys
 645 650 655
 Arg Tyr Leu Ile Asp Val Tyr Lys Glu Phe Ile Ala Val Tyr Thr Arg
 660 665 670
 Tyr Ser Cys Phe Asn Met Ser Gly Leu Lys Arg Ala Asp Glu Tyr Asn
 675 680 685
 Asp Ile Gly Glu Phe Phe Ser Asp Val Asp Thr Arg Leu Cys Thr Met
 690 695 700
 Glu Trp Ile Pro Val Ser Phe Glu Arg Ile Asn Asp Met Val Asp Lys
 705 710 715 720
 Lys Glu Gly Leu Leu Phe Leu Val Arg Ser Met Phe Leu Tyr Asn Arg

725 730 735
Pro Arg Lys Pro Tyr Glu Arg Thr Phe Ile Gln Leu Phe Ser Asp Ser
740 745 750
Asn Met Glu His Thr Ser Met Leu Leu Asn Ser Arg Ala Met Ile Gln

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

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

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

945 950 955 960
Ile Ser Lys Leu Ala Thr Lys Tyr Asn Ala Val Val Val Val Glu Ser
965 970 975

Met Ser Ser Thr Phe Lys Asp Lys Phe Ser Phe Leu Asp Glu Gln Ile
980 985 990

Phe Lys Ala Phe Glu Ala Arg Leu Cys Ala Arg Met Ser Asp Leu Ser
995 1000 1005

Phe Asn Thr Ile Lys Glu Gly Glu Ala Gly Ser Ile Ser Asn Pro
1010 1015 1020

Ile Gln Val Ser Asn Asn Asn Gly Asn Ser Tyr Gln Asp Gly Val
1025 1030 1035

Ile Tyr Phe Leu Asn Asn Ala Tyr Thr Arg Thr Leu Cys Pro Asp
1040 1045 1050

Thr Gly Phe Val Asp Val Phe Asp Lys Thr Arg Leu Ile Thr Met
1055 1060 1065

Gln Ser Lys Arg Gln Phe Phe Ala Lys Met Lys Asp Ile Arg Ile
1070 1075 1080

Asp Asp Gly Glu Met Leu Phe Thr Phe Asn Leu Glu Glu Tyr Pro
1085 1090 1095

Thr Lys Arg Leu Leu Asp Arg Lys Glu Trp Thr Val Lys Ile Ala
1100 1105 1110

Gly Asp Gly Ser Tyr Phe Asp Lys Asp Lys Gly Glu Tyr Val Tyr
1115 1120 1125

Val Asn Asp Ile Val Arg Glu Gln Ile Ile Pro Ala Leu Leu Glu
1130 1135 1140

Asp Lys Ala Val Phe Asp Gly Asn Met Ala Glu Lys Phe Leu Asp
1145 1150 1155

Lys Thr Ala Ile Ser Gly Lys Ser Val Glu Leu Ile Tyr Lys Trp
1160 1165 1170

Phe Ala Asn Ala Leu Tyr Gly Ile Ile Thr Lys Lys Asp Gly Glu
1175 1180 1185

Lys Ile Tyr Arg Ser Pro Ile Thr Gly Thr Glu Ile Asp Val Ser
1190 1195 1200

Lys Asn Thr Thr Tyr Asn Phe Gly Lys Lys Phe Met Phe Lys Gln

1205 1210 1215
 Glu Tyr Arg Gly Asp Gly Asp Phe Leu Asp Ala Phe Leu Asn Tyr
 1220 1225 1230
 Met Gln Ala Gln Asp Ile Ala Val
 1235 1240
 <210> 4
 <211> 1238
 <212> PRT
 <213> Candidatus Methanoplasma termitum
 <400> 4
 Met Asn Asn Tyr Asp Glu Phe Thr Lys Leu Tyr Pro Ile Gln Lys Thr

 1 5 10 15
 Ile Arg Phe Glu Leu Lys Pro Gln Gly Arg Thr Met Glu His Leu Glu
 20 25 30
 Thr Phe Asn Phe Phe Glu Glu Asp Arg Asp Arg Ala Glu Lys Tyr Lys
 35 40 45
 Ile Leu Lys Glu Ala Ile Asp Glu Tyr His Lys Lys Phe Ile Asp Glu
 50 55 60
 His Leu Thr Asn Met Ser Leu Asp Trp Asn Ser Leu Lys Gln Ile Ser

 65 70 75 80
 Glu Lys Tyr Tyr Lys Ser Arg Glu Glu Lys Asp Lys Lys Val Phe Leu
 85 90 95
 Ser Glu Gln Lys Arg Met Arg Gln Glu Ile Val Ser Glu Phe Lys Lys
 100 105 110
 Asp Asp Arg Phe Lys Asp Leu Phe Ser Lys Lys Leu Phe Ser Glu Leu
 115 120 125
 Leu Lys Glu Glu Ile Tyr Lys Lys Gly Asn His Gln Glu Ile Asp Ala

 130 135 140
 Leu Lys Ser Phe Asp Lys Phe Ser Gly Tyr Phe Ile Gly Leu His Glu
 145 150 155 160
 Asn Arg Lys Asn Met Tyr Ser Asp Gly Asp Glu Ile Thr Ala Ile Ser
 165 170 175

Asn Arg Ile Val Asn Glu Asn Phe Pro Lys Phe Leu Asp Asn Leu Gln
 180 185 190
 Lys Tyr Gln Glu Ala Arg Lys Lys Tyr Pro Glu Trp Ile Ile Lys Ala
 195 200 205
 Glu Ser Ala Leu Val Ala His Asn Ile Lys Met Asp Ile Val Phe Ser
 210 215 220
 Leu Glu Tyr Phe Asn Lys Val Leu Asn Gln Glu Gly Ile Gln Arg Tyr
 225 230 235 240
 Asn Leu Ala Leu Gly Gly Tyr Val Thr Lys Ser Gly Glu Lys Met Met
 245 250 255
 Gly Leu Asn Asp Ala Leu Asn Leu Ala His Gln Ser Glu Lys Ser Ser
 260 265 270
 Lys Gly Arg Ile His Met Thr Pro Leu Phe Lys Gln Ile Leu Ser Glu
 275 280 285
 Lys Glu Ser Phe Ser Tyr Ile Pro Asp Val Phe Thr Glu Asp Ser Gln
 290 295 300
 Leu Leu Pro Ser Ile Gly Gly Phe Phe Ala Gln Ile Glu Asn Asp Lys
 305 310 315 320
 Asp Gly Asn Ile Phe Asp Arg Ala Leu Glu Leu Ile Ser Ser Tyr Ala
 325 330 335
 Glu Tyr Asp Thr Glu Arg Ile Tyr Ile Arg Gln Ala Asp Ile Asn Arg
 340 345 350
 Val Ser Asn Val Ile Phe Gly Glu Trp Gly Thr Leu Gly Gly Leu Met
 355 360 365
 Arg Glu Tyr Lys Ala Asp Ser Ile Asn Asp Ile Asn Leu Glu Arg Thr
 370 375 380
 Cys Lys Lys Val Asp Lys Trp Leu Asp Ser Lys Glu Phe Ala Leu Ser
 385 390 395 400
 Asp Val Leu Glu Ala Ile Asp Arg Thr Gly Asn Asn Asp Ala Phe Asn
 405 410 415
 Glu Tyr Ile Ser Lys Met Arg Thr Ala Arg Glu Lys Ile Asp Ala Ala

420 425 430
Arg Lys Glu Met Lys Phe Ile Ser Glu Lys Ile Ser Gly Asp Glu Glu
435 440 445
Ser Ile His Ile Ile Lys Thr Leu Leu Asp Ser Val Gln Gln Phe Leu

450 455 460
His Phe Phe Asn Leu Phe Lys Ala Arg Gln Asp Ile Pro Leu Asp Gly
465 470 475 480
Ala Phe Tyr Ala Glu Phe Asp Glu Val His Ser Lys Leu Phe Ala Ile
485 490 495
Val Pro Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Lys Asn Asn Leu
500 505 510
Asn Thr Lys Lys Ile Lys Leu Asn Phe Lys Asn Pro Thr Leu Ala Asn

515 520 525
Gly Trp Asp Gln Asn Lys Val Tyr Asp Tyr Ala Ser Leu Ile Phe Leu
530 535 540
Arg Asp Gly Asn Tyr Tyr Leu Gly Ile Ile Asn Pro Lys Arg Lys Lys
545 550 555 560
Asn Ile Lys Phe Glu Gln Gly Ser Gly Asn Gly Pro Phe Tyr Arg Lys
565 570 575
Met Val Tyr Lys Gln Ile Pro Gly Pro Asn Lys Asn Leu Arg Pro Val

580 585 590
Phe Leu Thr Ser Thr Lys Gly Lys Lys Glu Tyr Lys Pro Ser Lys Glu
595 600 605
Ile Ile Glu Gly Tyr Glu Ala Asp Lys His Ile Arg Gly Asp Lys Phe
610 615 620
Asp Leu Asp Phe Cys His Lys Leu Ile Asp Phe Phe Lys Glu Ser Ile
625 630 635 640
Glu Lys His Lys Asp Trp Ser Lys Phe Asn Phe Tyr Phe Ser Pro Thr

645 650 655
Glu Ser Tyr Gly Asp Ile Ser Glu Phe Tyr Leu Asp Val Glu Lys Gln
660 665 670

Gly Tyr Arg Met His Phe Glu Asn Ile Ser Ala Glu Thr Ile Asp Glu
 675 680 685
 Tyr Val Glu Lys Gly Asp Leu Phe Leu Phe Gln Ile Tyr Asn Lys Asp
 690 695 700
 Phe Val Lys Ala Ala Thr Gly Lys Lys Asp Met His Thr Ile Tyr Trp

 705 710 715 720
 Asn Ala Ala Phe Ser Pro Glu Asn Leu Gln Asp Val Val Val Lys Leu
 725 730 735
 Asn Gly Glu Ala Glu Leu Phe Tyr Arg Asp Lys Ser Asp Ile Lys Glu
 740 745 750
 Ile Val His Arg Glu Gly Glu Ile Leu Val Asn Arg Thr Tyr Asn Gly
 755 760 765
 Arg Thr Pro Val Pro Asp Lys Ile His Lys Lys Leu Thr Asp Tyr His

 770 775 780
 Asn Gly Arg Thr Lys Asp Leu Gly Glu Ala Lys Glu Tyr Leu Asp Lys
 785 790 795 800
 Val Arg Tyr Phe Lys Ala His Tyr Asp Ile Thr Lys Asp Arg Arg Tyr
 805 810 815
 Leu Asn Asp Lys Ile Tyr Phe His Val Pro Leu Thr Leu Asn Phe Lys
 820 825 830
 Ala Asn Gly Lys Lys Asn Leu Asn Lys Met Val Ile Glu Lys Phe Leu

 835 840 845
 Ser Asp Glu Lys Ala His Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn
 850 855 860
 Leu Leu Tyr Tyr Ser Ile Ile Asp Arg Ser Gly Lys Ile Ile Asp Gln
 865 870 875 880
 Gln Ser Leu Asn Val Ile Asp Gly Phe Asp Tyr Arg Glu Lys Leu Asn
 885 890 895
 Gln Arg Glu Ile Glu Met Lys Asp Ala Arg Gln Ser Trp Asn Ala Ile

 900 905 910
 Gly Lys Ile Lys Asp Leu Lys Glu Gly Tyr Leu Ser Lys Ala Val His

915 920 925
 Glu Ile Thr Lys Met Ala Ile Gln Tyr Asn Ala Ile Val Val Met Glu
 930 935 940
 Glu Leu Asn Tyr Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln
 945 950 955 960
 Ile Tyr Gln Lys Phe Glu Asn Met Leu Ile Asp Lys Met Asn Tyr Leu

 965 970 975
 Val Phe Lys Asp Ala Pro Asp Glu Ser Pro Gly Gly Val Leu Asn Ala
 980 985 990
 Tyr Gln Leu Thr Asn Pro Leu Glu Ser Phe Ala Lys Leu Gly Lys Gln
 995 1000 1005
 Thr Gly Ile Leu Phe Tyr Val Pro Ala Ala Tyr Thr Ser Lys Ile
 1010 1015 1020
 Asp Pro Thr Thr Gly Phe Val Asn Leu Phe Asn Thr Ser Ser Lys

 1025 1030 1035
 Thr Asn Ala Gln Glu Arg Lys Glu Phe Leu Gln Lys Phe Glu Ser
 1040 1045 1050
 Ile Ser Tyr Ser Ala Lys Asp Gly Gly Ile Phe Ala Phe Ala Phe
 1055 1060 1065
 Asp Tyr Arg Lys Phe Gly Thr Ser Lys Thr Asp His Lys Asn Val
 1070 1075 1080
 Trp Thr Ala Tyr Thr Asn Gly Glu Arg Met Arg Tyr Ile Lys Glu

 1085 1090 1095
 Lys Lys Arg Asn Glu Leu Phe Asp Pro Ser Lys Glu Ile Lys Glu
 1100 1105 1110
 Ala Leu Thr Ser Ser Gly Ile Lys Tyr Asp Gly Gly Gln Asn Ile
 1115 1120 1125
 Leu Pro Asp Ile Leu Arg Ser Asn Asn Asn Gly Leu Ile Tyr Thr
 1130 1135 1140
 Met Tyr Ser Ser Phe Ile Ala Ala Ile Gln Met Arg Val Tyr Asp

 1145 1150 1155

Gly Lys Glu Asp Tyr Ile Ile Ser Pro Ile Lys Asn Ser Lys Gly
1160 1165 1170
Glu Phe Phe Arg Thr Asp Pro Lys Arg Arg Glu Leu Pro Ile Asp
1175 1180 1185
Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Leu Arg Gly Glu Leu
1190 1195 1200
Thr Met Arg Ala Ile Ala Glu Lys Phe Asp Pro Asp Ser Glu Lys
1205 1210 1215
Met Ala Lys Leu Glu Leu Lys His Lys Asp Trp Phe Glu Phe Met
1220 1225 1230
Gln Thr Arg Gly Asp
1235
<210> 5
<211> 1281
<212> PRT
<213> Eubacterium eligens
<400> 5
Met Asn Gly Asn Arg Ser Ile Val Tyr Arg Glu Phe Val Gly Val Ile
1 5 10 15
Pro Val Ala Lys Thr Leu Arg Asn Glu Leu Arg Pro Val Gly His Thr
20 25 30
Gln Glu His Ile Ile Gln Asn Gly Leu Ile Gln Glu Asp Glu Leu Arg
35 40 45
Gln Glu Lys Ser Thr Glu Leu Lys Asn Ile Met Asp Asp Tyr Tyr Arg
50 55 60
Glu Tyr Ile Asp Lys Ser Leu Ser Gly Val Thr Asp Leu Asp Phe Thr
65 70 75 80
Leu Leu Phe Glu Leu Met Asn Leu Val Gln Ser Ser Pro Ser Lys Asp
85 90 95
Asn Lys Lys Ala Leu Glu Lys Glu Gln Ser Lys Met Arg Glu Gln Ile
100 105 110
Cys Thr His Leu Gln Ser Asp Ser Asn Tyr Lys Asn Ile Phe Asn Ala

115 120 125
 Lys Leu Leu Lys Glu Ile Leu Pro Asp Phe Ile Lys Asn Tyr Asn Gln
 130 135 140
 Tyr Asp Val Lys Asp Lys Ala Gly Lys Leu Glu Thr Leu Ala Leu Phe

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

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

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

 340 345 350
 Leu Ser Cys Phe Met Ser Gly Asn Trp Asn Leu Ile Thr Gly Cys Val
 355 360 365

Glu Asn Phe Tyr Asp Glu Asn Ile His Ala Lys Gly Lys Ser Lys Glu
 370 375 380
 Glu Lys Val Lys Lys Ala Val Lys Glu Asp Lys Tyr Lys Ser Ile Asn
 385 390 395 400
 Asp Val Asn Asp Leu Val Glu Lys Tyr Ile Asp Glu Lys Glu Arg Asn
 405 410 415
 Glu Phe Lys Asn Ser Asn Ala Lys Gln Tyr Ile Arg Glu Ile Ser Asn
 420 425 430
 Ile Ile Thr Asp Thr Glu Thr Ala His Leu Glu Tyr Asp Asp His Ile
 435 440 445
 Ser Leu Ile Glu Ser Glu Glu Lys Ala Asp Glu Met Lys Lys Arg Leu
 450 455 460
 Asp Met Tyr Met Asn Met Tyr His Trp Ala Lys Ala Phe Ile Val Asp
 465 470 475 480
 Glu Val Leu Asp Arg Asp Glu Met Phe Tyr Ser Asp Ile Asp Asp Ile
 485 490 495
 Tyr Asn Ile Leu Glu Asn Ile Val Pro Leu Tyr Asn Arg Val Arg Asn
 500 505 510
 Tyr Val Thr Gln Lys Pro Tyr Asn Ser Lys Lys Ile Lys Leu Asn Phe
 515 520 525
 Gln Ser Pro Thr Leu Ala Asn Gly Trp Ser Gln Ser Lys Glu Phe Asp
 530 535 540
 Asn Asn Ala Ile Ile Leu Ile Arg Asp Asn Lys Tyr Tyr Leu Ala Ile
 545 550 555 560
 Phe Asn Ala Lys Asn Lys Pro Asp Lys Lys Ile Ile Gln Gly Asn Ser
 565 570 575
 Asp Lys Lys Asn Asp Asn Asp Tyr Lys Lys Met Val Tyr Asn Leu Leu
 580 585 590
 Pro Gly Ala Asn Lys Met Leu Pro Lys Val Phe Leu Ser Lys Lys Gly
 595 600 605
 Ile Glu Thr Phe Lys Pro Ser Asp Tyr Ile Ile Ser Gly Tyr Asn Ala

610 615 620
 His Lys His Ile Lys Thr Ser Glu Asn Phe Asp Ile Ser Phe Cys Arg
 625 630 635 640
 Asp Leu Ile Asp Tyr Phe Lys Asn Ser Ile Glu Lys His Ala Glu Trp
 645 650 655
 Arg Lys Tyr Glu Phe Lys Phe Ser Ala Thr Asp Ser Tyr Ser Asp Ile

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

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

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

 850 855 860

Lys Tyr Ile Ala Gln Asn Asp Asp Ile His Val Ile Gly Ile Asp Arg
 865 870 875 880
 Gly Glu Arg Asn Leu Ile Tyr Ile Ser Val Ile Asp Ser His Gly Asn
 885 890 895
 Ile Val Lys Gln Lys Ser Tyr Asn Ile Leu Asn Asn Tyr Asp Tyr Lys
 900 905 910
 Lys Lys Leu Val Glu Lys Glu Lys Thr Arg Glu Tyr Ala Arg Lys Asn
 915 920 925
 Trp Lys Ser Ile Gly Asn Ile Lys Glu Leu Lys Glu Gly Tyr Ile Ser
 930 935 940
 Gly Val Val His Glu Ile Ala Met Leu Ile Val Glu Tyr Asn Ala Ile
 945 950 955 960
 Ile Ala Met Glu Asp Leu Asn Tyr Gly Phe Lys Arg Gly Arg Phe Lys
 965 970 975
 Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Ser Met Leu Ile Asn Lys
 980 985 990
 Leu Asn Tyr Phe Ala Ser Lys Glu Lys Ser Val Asp Glu Pro Gly Gly
 995 1000 1005
 Leu Leu Lys Gly Tyr Gln Leu Thr Tyr Val Pro Asp Asn Ile Lys
 1010 1015 1020
 Asn Leu Gly Lys Gln Cys Gly Val Ile Phe Tyr Val Pro Ala Ala
 1025 1030 1035
 Phe Thr Ser Lys Ile Asp Pro Ser Thr Gly Phe Ile Ser Ala Phe
 1040 1045 1050
 Asn Phe Lys Ser Ile Ser Thr Asn Ala Ser Arg Lys Gln Phe Phe
 1055 1060 1065
 Met Gln Phe Asp Glu Ile Arg Tyr Cys Ala Glu Lys Asp Met Phe
 1070 1075 1080
 Ser Phe Gly Phe Asp Tyr Asn Asn Phe Asp Thr Tyr Asn Ile Thr
 1085 1090 1095
 Met Gly Lys Thr Gln Trp Thr Val Tyr Thr Asn Gly Glu Arg Leu

1100 1105 1110
 Gln Ser Glu Phe Asn Asn Ala Arg Arg Thr Gly Lys Thr Lys Ser
 1115 1120 1125
 Ile Asn Leu Thr Glu Thr Ile Lys Leu Leu Leu Glu Asp Asn Glu
 1130 1135 1140
 Ile Asn Tyr Ala Asp Gly His Asp Ile Arg Ile Asp Met Glu Lys
 1145 1150 1155
 Met Asp Glu Asp Lys Lys Ser Glu Phe Phe Ala Gln Leu Leu Ser

1160 1165 1170
 Leu Tyr Lys Leu Thr Val Gln Met Arg Asn Ser Tyr Thr Glu Ala
 1175 1180 1185
 Glu Glu Gln Glu Asn Gly Ile Ser Tyr Asp Lys Ile Ile Ser Pro
 1190 1195 1200
 Val Ile Asn Asp Glu Gly Glu Phe Phe Asp Ser Asp Asn Tyr Lys
 1205 1210 1215
 Glu Ser Asp Asp Lys Glu Cys Lys Met Pro Lys Asp Ala Asp Ala

1220 1225 1230
 Asn Gly Ala Tyr Cys Ile Ala Leu Lys Gly Leu Tyr Glu Val Leu
 1235 1240 1245
 Lys Ile Lys Ser Glu Trp Thr Glu Asp Gly Phe Asp Arg Asn Cys
 1250 1255 1260
 Leu Lys Leu Pro His Ala Glu Trp Leu Asp Phe Ile Gln Asn Lys
 1265 1270 1275

Arg Tyr Glu

1280

<210> 6

<211> 1300

<212> PRT

<213> Francisella novicida

<400> 6

Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr

1

5

10

15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
 20 25 30
 Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
 35 40 45
 Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
 50 55 60
 Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
 65 70 75 80
 Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
 85 90 95
 Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
 100 105 110
 Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
 115 120 125
 Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
 130 135 140
 Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
 145 150 155 160
 Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
 165 170 175
 Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Val Asn Tyr Ser Ser
 180 185 190
 Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
 195 200 205
 Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
 210 215 220
 Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
 225 230 235 240
 Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
 245 250 255
 Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr

260 265 270
 Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
 275 280 285
 Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
 290 295 300
 Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys

 305 310 315 320
 Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
 325 330 335
 Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
 340 345 350
 Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
 355 360 365
 Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln

 370 375 380
 Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
 385 390 395 400
 Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
 405 410 415
 Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
 420 425 430
 Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala

 435 440 445
 Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
 450 455 460
 Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
 465 470 475 480
 Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
 485 490 495
 Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys

 500 505 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
 515 520 525
 Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
 530 535 540
 Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
 545 550 555 560
 Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
 565 570 575
 Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
 580 585 590
 Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
 595 600 605
 Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
 610 615 620
 Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
 625 630 635 640
 Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
 645 650 655
 Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
 660 665 670
 Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
 675 680 685
 Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
 690 695 700
 Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
 705 710 715 720
 Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
 725 730 735
 Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
 740 745 750
 Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn

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

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

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

 945 950 955 960
 Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
 965 970 975
 Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
 980 985 990
 Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
 995 1000 1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val

1010 1015 1020

Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu

1025 1030 1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg

1040 1045 1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly

1055 1060 1065

Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser

1070 1075 1080

Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys

1085 1090 1095

Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp

1100 1105 1110

Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe

1115 1120 1125

Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr

1130 1135 1140

Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp

1145 1150 1155

Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu

1160 1165 1170

Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly

1175 1180 1185

Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe

1190 1195 1200

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg

1205 1210 1215

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val

1220 1225 1230

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys

1235 1240 1245
Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly

1250 1255 1260
Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu

1265 1270 1275
Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu

1280 1285 1290
Phe Val Gln Asn Arg Asn Asn

1295 1300

<210> 7

<211> 1206

<212> PRT

<213> Unknown

<220><223> Lachnospiraceae bacterium

<400> 7

Met Tyr Tyr Glu Ser Leu Thr Lys Gln Tyr Pro Val Ser Lys Thr Ile

1 5 10 15
Arg Asn Glu Leu Ile Pro Ile Gly Lys Thr Leu Asp Asn Ile Arg Gln

20 25 30
Asn Asn Ile Leu Glu Ser Asp Val Lys Arg Lys Gln Asn Tyr Glu His

35 40 45
Val Lys Gly Ile Leu Asp Glu Tyr His Lys Gln Leu Ile Asn Glu Ala

50 55 60
Leu Asp Asn Cys Thr Leu Pro Ser Leu Lys Ile Ala Ala Glu Ile Tyr

65 70 75 80
Leu Lys Asn Gln Lys Glu Val Ser Asp Arg Glu Asp Phe Asn Lys Thr

85 90 95
Gln Asp Leu Leu Arg Lys Glu Val Val Glu Lys Leu Lys Ala His Glu

100 105 110
Asn Phe Thr Lys Ile Gly Lys Lys Asp Ile Leu Asp Leu Leu Glu Lys

115 120 125

Leu Pro Ser Ile Ser Glu Asp Asp Tyr Asn Ala Leu Glu Ser Phe Arg

130 135 140

Asn Phe Tyr Thr Tyr Phe Thr Ser Tyr Asn Lys Val Arg Glu Asn Leu

145 150 155 160

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

165 170 175

Glu Asn Phe Pro Lys Phe Leu Asp Asn Val Lys Ser Tyr Arg Phe Val

180 185 190

Lys Thr Ala Gly Ile Leu Ala Asp Gly Leu Gly Glu Glu Glu Gln Asp

195 200 205

Ser Leu Phe Ile Val Glu Thr Phe Asn Lys Thr Leu Thr Gln Asp Gly

210 215 220

Ile Asp Thr Tyr Asn Ser Gln Val Gly Lys Ile Asn Ser Ser Ile Asn

225 230 235 240

Leu Tyr Asn Gln Lys Asn Gln Lys Ala Asn Gly Phe Arg Lys Ile Pro

245 250 255

Lys Met Lys Met Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Glu Ser

260 265 270

Phe Ile Asp Glu Phe Gln Ser Asp Glu Val Leu Ile Asp Asn Val Glu

275 280 285

Ser Tyr Gly Ser Val Leu Ile Glu Ser Leu Lys Ser Ser Lys Val Ser

290 295 300

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

305 310 315 320

Lys Asn Asp Leu Ala Lys Thr Ala Met Ser Val Ile Val Phe Glu Asn

325 330 335

Trp Arg Thr Phe Asp Asp Leu Leu Asn Gln Glu Tyr Asp Leu Ala Asn

340 345 350

Glu Asn Lys Lys Lys Asp Asp Lys Tyr Phe Glu Lys Arg Gln Lys Glu

355 360 365

Leu Lys Lys Asn Lys Ser Tyr Ser Leu Glu His Leu Cys Asn Leu Ser

370 375 380
 Glu Asp Ser Cys Asn Leu Ile Glu Asn Tyr Ile His Gln Ile Ser Asp

 385 390 395 400
 Asp Ile Glu Asn Ile Ile Ile Asn Asn Glu Thr Phe Leu Arg Ile Val
 405 410 415
 Ile Asn Glu His Asp Arg Ser Arg Lys Leu Ala Lys Asn Arg Lys Ala
 420 425 430
 Val Lys Ala Ile Lys Asp Phe Leu Asp Ser Ile Lys Val Leu Glu Arg
 435 440 445
 Glu Leu Lys Leu Ile Asn Ser Ser Gly Gln Glu Leu Glu Lys Asp Leu

 450 455 460
 Ile Val Tyr Ser Ala His Glu Glu Leu Leu Val Glu Leu Lys Gln Val
 465 470 475 480
 Asp Ser Leu Tyr Asn Met Thr Arg Asn Tyr Leu Thr Lys Lys Pro Phe
 485 490 495
 Ser Thr Glu Lys Val Lys Leu Asn Phe Asn Arg Ser Thr Leu Leu Asn
 500 505 510
 Gly Trp Asp Arg Asn Lys Glu Thr Asp Asn Leu Gly Val Leu Leu Leu

 515 520 525
 Lys Asp Gly Lys Tyr Tyr Leu Gly Ile Met Asn Thr Ser Ala Asn Lys
 530 535 540
 Ala Phe Val Asn Pro Pro Val Ala Lys Thr Glu Lys Val Phe Lys Lys
 545 550 555 560
 Val Asp Tyr Lys Leu Leu Pro Val Pro Asn Gln Met Leu Pro Lys Val
 565 570 575
 Phe Phe Ala Lys Ser Asn Ile Asp Phe Tyr Asn Pro Ser Ser Glu Ile

 580 585 590
 Tyr Ser Asn Tyr Lys Lys Gly Thr His Lys Lys Gly Asn Met Phe Ser
 595 600 605
 Leu Glu Asp Cys His Asn Leu Ile Asp Phe Phe Lys Glu Ser Ile Ser
 610 615 620

Lys His Glu Asp Trp Ser Lys Phe Gly Phe Lys Phe Asp Thr Gln Ala
 625 630 635 640
 Ser Tyr Asn Asp Ile Ser Glu Phe Tyr Arg Glu Val Glu Lys Gln Gly

 645 650 655
 Tyr Lys Leu Thr Tyr Thr Asp Ile Asp Glu Thr Tyr Ile Asn Asp Leu
 660 665 670
 Ile Glu Arg Asn Glu Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe
 675 680 685
 Ser Met Tyr Ser Lys Gly Lys Leu Asn Leu His Thr Leu Tyr Phe Met
 690 695 700
 Met Leu Phe Asp Gln Arg Asn Ile Asp Asp Val Val Tyr Lys Leu Asn

 705 710 715 720
 Gly Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Ser Glu Asp Glu
 725 730 735
 Leu Ile Ile His Lys Ala Gly Glu Glu Ile Lys Asn Lys Asn Pro Asn
 740 745 750
 Arg Ala Arg Thr Lys Glu Thr Ser Thr Phe Ser Tyr Asp Ile Val Lys
 755 760 765
 Asp Lys Arg Tyr Ser Lys Asp Lys Phe Thr Leu His Ile Pro Ile Thr

 770 775 780
 Met Asn Phe Gly Val Asp Glu Val Lys Arg Phe Asn Asp Ala Val Asn
 785 790 795 800
 Ser Ala Ile Arg Ile Asp Glu Asn Val Asn Val Ile Gly Ile Asp Arg
 805 810 815
 Gly Glu Arg Asn Leu Leu Tyr Val Val Val Ile Asp Ser Lys Gly Asn
 820 825 830
 Ile Leu Glu Gln Ile Ser Leu Asn Ser Ile Ile Asn Lys Glu Tyr Asp

 835 840 845
 Ile Glu Thr Asp Tyr His Ala Leu Leu Asp Glu Arg Glu Gly Gly Arg
 850 855 860
 Asp Lys Ala Arg Lys Asp Trp Asn Thr Val Glu Asn Ile Arg Asp Leu

865 870 875 880
 Lys Ala Gly Leu Tyr Leu Gln Val Val Asn Val Val Ala Lys Leu Val
 885 890 895
 Leu Lys Tyr Asn Ala Ile Ile Cys Leu Glu Asp Leu Asn Phe Gly Phe

 900 905 910
 Lys Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu
 915 920 925
 Lys Met Leu Ile Asp Lys Leu Asn Tyr Leu Val Ile Asp Lys Ser Arg
 930 935 940
 Glu Gln Thr Ser Pro Lys Glu Leu Gly Gly Ala Leu Asn Ala Leu Gln
 945 950 955 960
 Leu Thr Ser Lys Phe Lys Ser Phe Lys Glu Leu Gly Lys Gln Ser Gly

 965 970 975
 Val Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile Asp Pro Thr
 980 985 990
 Thr Gly Phe Ala Asn Leu Phe Tyr Met Lys Cys Glu Asn Val Glu Lys
 995 1000 1005
 Ser Lys Arg Phe Phe Asp Gly Phe Asp Phe Ile Arg Phe Asn Ala
 1010 1015 1020
 Leu Glu Asn Val Phe Glu Phe Gly Phe Asp Tyr Arg Ser Phe Thr

 1025 1030 1035
 Gln Arg Ala Cys Gly Ile Asn Ser Lys Trp Thr Val Cys Thr Asn
 1040 1045 1050
 Gly Glu Arg Ile Ile Lys Tyr Arg Asn Pro Asp Lys Asn Asn Met
 1055 1060 1065
 Phe Asp Glu Lys Val Val Val Val Thr Asp Glu Met Lys Asn Leu
 1070 1075 1080
 Phe Glu Gln Tyr Lys Ile Pro Tyr Glu Asp Gly Arg Asn Val Lys

 1085 1090 1095
 Asp Met Ile Ile Ser Asn Glu Glu Ala Glu Phe Tyr Arg Arg Leu
 1100 1105 1110

Tyr Arg Leu Leu Gln Gln Thr Leu Gln Met Arg Asn Ser Thr Ser
1115 1120 1125

Asp Gly Thr Arg Asp Tyr Ile Ile Ser Pro Val Lys Asn Lys Arg
1130 1135 1140

Glu Ala Tyr Phe Asn Ser Glu Leu Ser Asp Gly Ser Val Pro Lys

1145 1150 1155

Asp Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Arg Lys Gly Leu
1160 1165 1170

Trp Val Leu Glu Gln Ile Arg Gln Lys Ser Glu Gly Glu Lys Ile
1175 1180 1185

Asn Leu Ala Met Thr Asn Ala Glu Trp Leu Glu Tyr Ala Gln Thr
1190 1195 1200

His Leu Leu

1205

<210> 8

<211> 1233

<212> PRT

<213> Unknown

<220><223> Lachnospiraceae bacterium

<400> 8

Met Asp Tyr Gly Asn Gly Gln Phe Glu Arg Arg Ala Pro Leu Thr Lys

1 5 10 15

Thr Ile Thr Leu Arg Leu Lys Pro Ile Gly Glu Thr Arg Glu Thr Ile

20 25 30

Arg Glu Gln Lys Leu Leu Glu Gln Asp Ala Ala Phe Arg Lys Leu Val

35 40 45

Glu Thr Val Thr Pro Ile Val Asp Asp Cys Ile Arg Lys Ile Ala Asp

50 55 60

Asn Ala Leu Cys His Phe Gly Thr Glu Tyr Asp Phe Ser Cys Leu Gly

65 70 75 80

Asn Ala Ile Ser Lys Asn Asp Ser Lys Ala Ile Lys Lys Glu Thr Glu

85 90 95

Lys Val Glu Lys Leu Leu Ala Lys Val Leu Thr Glu Asn Leu Pro Asp
 100 105 110
 Gly Leu Arg Lys Val Asn Asp Ile Asn Ser Ala Ala Phe Ile Gln Asp
 115 120 125
 Thr Leu Thr Ser Phe Val Gln Asp Asp Ala Asp Lys Arg Val Leu Ile
 130 135 140
 Gln Glu Leu Lys Gly Lys Thr Val Leu Met Gln Arg Phe Leu Thr Thr
 145 150 155 160
 Arg Ile Thr Ala Leu Thr Val Trp Leu Pro Asp Arg Val Phe Glu Asn
 165 170 175
 Phe Asn Ile Phe Ile Glu Asn Ala Glu Lys Met Arg Ile Leu Leu Asp
 180 185 190
 Ser Pro Leu Asn Glu Lys Ile Met Lys Phe Asp Pro Asp Ala Glu Gln
 195 200 205
 Tyr Ala Ser Leu Glu Phe Tyr Gly Gln Cys Leu Ser Gln Lys Asp Ile
 210 215 220
 Asp Ser Tyr Asn Leu Ile Ile Ser Gly Ile Tyr Ala Asp Asp Glu Val
 225 230 235 240
 Lys Asn Pro Gly Ile Asn Glu Ile Val Lys Glu Tyr Asn Gln Gln Ile
 245 250 255
 Arg Gly Asp Lys Asp Glu Ser Pro Leu Pro Lys Leu Lys Lys Leu His
 260 265 270
 Lys Gln Ile Leu Met Pro Val Glu Lys Ala Phe Phe Val Arg Val Leu
 275 280 285
 Ser Asn Asp Ser Asp Ala Arg Ser Ile Leu Glu Lys Ile Leu Lys Asp
 290 295 300
 Thr Glu Met Leu Pro Ser Lys Ile Ile Glu Ala Met Lys Glu Ala Asp
 305 310 315 320
 Ala Gly Asp Ile Ala Val Tyr Gly Ser Arg Leu His Glu Leu Ser His
 325 330 335
 Val Ile Tyr Gly Asp His Gly Lys Leu Ser Gln Ile Ile Tyr Asp Lys

340 345 350
 Glu Ser Lys Arg Ile Ser Glu Leu Met Glu Thr Leu Ser Pro Lys Glu
 355 360 365
 Arg Lys Glu Ser Lys Lys Arg Leu Glu Gly Leu Glu Glu His Ile Arg

370 375 380
 Lys Ser Thr Tyr Thr Phe Asp Glu Leu Asn Arg Tyr Ala Glu Lys Asn
 385 390 395 400
 Val Met Ala Ala Tyr Ile Ala Ala Val Glu Glu Ser Cys Ala Glu Ile
 405 410 415
 Met Arg Lys Glu Lys Asp Leu Arg Thr Leu Leu Ser Lys Glu Asp Val
 420 425 430
 Lys Ile Arg Gly Asn Arg His Asn Thr Leu Ile Val Lys Asn Tyr Phe

435 440 445
 Asn Ala Trp Thr Val Phe Arg Asn Leu Ile Arg Ile Leu Arg Arg Lys
 450 455 460
 Ser Glu Ala Glu Ile Asp Ser Asp Phe Tyr Asp Val Leu Asp Asp Ser
 465 470 475 480
 Val Glu Val Leu Ser Leu Thr Tyr Lys Gly Glu Asn Leu Cys Arg Ser
 485 490 495
 Tyr Ile Thr Lys Lys Ile Gly Ser Asp Leu Lys Pro Glu Ile Ala Thr

500 505 510
 Tyr Gly Ser Ala Leu Arg Pro Asn Ser Arg Trp Trp Ser Pro Gly Glu
 515 520 525
 Lys Phe Asn Val Lys Phe His Thr Ile Val Arg Arg Asp Gly Arg Leu
 530 535 540
 Tyr Tyr Phe Ile Leu Pro Lys Gly Ala Lys Pro Val Glu Leu Glu Asp
 545 550 555 560
 Met Asp Gly Asp Ile Glu Cys Leu Gln Met Arg Lys Ile Pro Asn Pro

565 570 575
 Thr Ile Phe Leu Pro Lys Leu Val Phe Lys Asp Pro Glu Ala Phe Phe
 580 585 590

Arg Asp Asn Pro Glu Ala Asp Glu Phe Val Phe Leu Ser Gly Met Lys
595 600 605

Ala Pro Val Thr Ile Thr Arg Glu Thr Tyr Glu Ala Tyr Arg Tyr Lys
610 615 620

Leu Tyr Thr Val Gly Lys Leu Arg Asp Gly Glu Val Ser Glu Glu Glu

625 630 635 640

Tyr Lys Arg Ala Leu Leu Gln Val Leu Thr Ala Tyr Lys Glu Phe Leu
645 650 655

Glu Asn Arg Met Ile Tyr Ala Asp Leu Asn Phe Gly Phe Lys Asp Leu
660 665 670

Glu Glu Tyr Lys Asp Ser Ser Glu Phe Ile Lys Gln Val Glu Thr His
675 680 685

Asn Thr Phe Met Cys Trp Ala Lys Val Ser Ser Ser Gln Leu Asp Asp

690 695 700

Leu Val Lys Ser Gly Asn Gly Leu Leu Phe Glu Ile Trp Ser Glu Arg
705 710 715 720

Leu Glu Ser Tyr Tyr Lys Tyr Gly Asn Glu Lys Val Leu Arg Gly Tyr
725 730 735

Glu Gly Val Leu Leu Ser Ile Leu Lys Asp Glu Asn Leu Val Ser Met
740 745 750

Arg Thr Leu Leu Asn Ser Arg Pro Met Leu Val Tyr Arg Pro Lys Glu

755 760 765

Ser Ser Lys Pro Met Val Val His Arg Asp Gly Ser Arg Val Val Asp
770 775 780

Arg Phe Asp Lys Asp Gly Lys Tyr Ile Pro Pro Glu Val His Asp Glu
785 790 795 800

Leu Tyr Arg Phe Phe Asn Asn Leu Leu Ile Lys Glu Lys Leu Gly Glu
805 810 815

Lys Ala Arg Lys Ile Leu Asp Asn Lys Lys Val Lys Val Lys Val Leu

820 825 830

Glu Ser Glu Arg Val Lys Trp Ser Lys Phe Tyr Asp Glu Gln Phe Ala

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

Glu Gly Cys Leu Ile Met Lys Tyr Ser Gly Asp Ser Leu Pro Thr
1085 1090 1095

His Asn Thr Asp Asn Arg Val Trp Asn Cys Cys Cys Asn His Pro
1100 1105 1110

Ile Thr Asn Tyr Asp Arg Glu Thr Lys Lys Val Glu Phe Ile Glu
1115 1120 1125

Glu Pro Val Glu Glu Leu Ser Arg Val Leu Glu Glu Asn Gly Ile
1130 1135 1140

Glu Thr Asp Thr Glu Leu Asn Lys Leu Asn Glu Arg Glu Asn Val
1145 1150 1155

Pro Gly Lys Val Val Asp Ala Ile Tyr Ser Leu Val Leu Asn Tyr
1160 1165 1170

Leu Arg Gly Thr Val Ser Gly Val Ala Gly Gln Arg Ala Val Tyr
1175 1180 1185

Tyr Ser Pro Val Thr Gly Lys Lys Tyr Asp Ile Ser Phe Ile Gln
1190 1195 1200

Ala Met Asn Leu Asn Arg Lys Cys Asp Tyr Tyr Arg Ile Gly Ser
1205 1210 1215

Lys Glu Arg Gly Glu Trp Thr Asp Phe Val Ala Gln Leu Ile Asn
1220 1225 1230

<210> 9

<211> 1227

<212> PRT

<213> Unknown

<220><223> Lachnospiraceae bacterium

<400> 9

Met Ser Lys Leu Glu Lys Phe Thr Asn Cys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Lys Ala Ile Pro Val Gly Lys Thr Gln Glu Asn Ile Asp
20 25 30

Asn Lys Arg Leu Leu Val Glu Asp Glu Lys Arg Ala Glu Asp Tyr Lys
35 40 45

Gly Val Lys Lys Leu Leu Asp Arg Tyr Tyr Leu Ser Phe Ile Asn Asp
50 55 60

Val Leu His Ser Ile Lys Leu Lys Asn Leu Asn Asn Tyr Ile Ser Leu
65 70 75 80

Phe Arg Lys Lys Thr Arg Thr Glu Lys Glu Asn Lys Glu Leu Glu Asn
85 90 95

Leu Glu Ile Asn Leu Arg Lys Glu Ile Ala Lys Ala Phe Lys Gly Asn
100 105 110

Glu Gly Tyr Lys Ser Leu Phe Lys Lys Asp Ile Ile Glu Thr Ile Leu
115 120 125

Pro Glu Phe Leu Asp Asp Lys Asp Glu Ile Ala Leu Val Asn Ser Phe
130 135 140

Asn Gly Phe Thr Thr Ala Phe Thr Gly Phe Phe Asp Asn Arg Glu Asn
145 150 155 160

Met Phe Ser Glu Glu Ala Lys Ser Thr Ser Ile Ala Phe Arg Cys Ile
165 170 175

Asn Glu Asn Leu Thr Arg Tyr Ile Ser Asn Met Asp Ile Phe Glu Lys
180 185 190

Val Asp Ala Ile Phe Asp Lys His Glu Val Gln Glu Ile Lys Glu Lys
195 200 205

Ile Leu Asn Ser Asp Tyr Asp Val Glu Asp Phe Phe Glu Gly Glu Phe
210 215 220

Phe Asn Phe Val Leu Thr Gln Glu Gly Ile Asp Val Tyr Asn Ala Ile
225 230 235 240

Ile Gly Gly Phe Val Thr Glu Ser Gly Glu Lys Ile Lys Gly Leu Asn
245 250 255

Glu Tyr Ile Asn Leu Tyr Asn Gln Lys Thr Lys Gln Lys Leu Pro Lys
260 265 270

Phe Lys Pro Leu Tyr Lys Gln Val Leu Ser Asp Arg Glu Ser Leu Ser
275 280 285

Phe Tyr Gly Glu Gly Tyr Thr Ser Asp Glu Glu Val Leu Glu Val Phe

290 295 300
 Arg Asn Thr Leu Asn Lys Asn Ser Glu Ile Phe Ser Ser Ile Lys Lys
 305 310 315 320
 Leu Glu Lys Leu Phe Lys Asn Phe Asp Glu Tyr Ser Ser Ala Gly Ile
 325 330 335

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

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

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

 Gln Phe Met Gly Gly Trp Asp Lys Asp Lys Glu Thr Asp Tyr Arg Ala
 530 535 540

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

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

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

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

 Lys Asp Lys Arg Phe Ser Glu Asp Gln Tyr Glu Leu His Ile Pro Ile

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

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

 Cys Glu Leu Val Glu Lys Tyr Asp Ala Val Ile Ala Leu Glu Asp Leu
 915 920 925
 Asn Ser Gly Phe Lys Asn Ser Arg Val Lys Val Glu Lys Gln Val Tyr
 930 935 940
 Gln Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Met Val Asp
 945 950 955 960
 Lys Lys Ser Asn Pro Cys Ala Thr Gly Gly Ala Leu Lys Gly Tyr Gln
 965 970 975

 Ile Thr Asn Lys Phe Glu Ser Phe Lys Ser Met Ser Thr Gln Asn Gly
 980 985 990
 Phe Ile Phe Tyr Ile Pro Ala Trp Leu Thr Ser Lys Ile Asp Pro Ser
 995 1000 1005
 Thr Gly Phe Val Asn Leu Leu Lys Thr Lys Tyr Thr Ser Ile Ala
 1010 1015 1020
 Asp Lys Lys Phe Ile Ser Ser Phe Asp Arg Ile Met Tyr Val Pro
 1025 1030 1035

Glu Glu Asp Leu Phe Glu Phe Ala Leu Asp Tyr Lys Asn Phe Ser
 1040 1045 1050
 Arg Thr Asp Ala Asp Tyr Ile Lys Lys Trp Lys Leu Tyr Ser Tyr
 1055 1060 1065
 Gly Asn Arg Ile Arg Ile Phe Arg Asn Pro Lys Lys Asn Asn Val
 1070 1075 1080
 Phe Asp Trp Glu Glu Val Cys Leu Thr Ser Ala Tyr Lys Glu Leu
 1085 1090 1095

Phe Asn Lys Tyr Gly Ile Asn Tyr Gln Gln Gly Asp Ile Arg Ala
 1100 1105 1110
 Leu Leu Cys Glu Gln Ser Asp Lys Ala Phe Tyr Ser Ser Phe Met
 1115 1120 1125
 Ala Leu Met Ser Leu Met Leu Gln Met Arg Asn Ser Ile Thr Gly
 1130 1135 1140
 Arg Thr Asp Val Asp Phe Leu Ile Ser Pro Val Lys Asn Ser Asp
 1145 1150 1155

Gly Ile Phe Tyr Asp Ser Arg Asn Tyr Glu Ala Gln Glu Asn Ala
 1160 1165 1170
 Ile Leu Pro Lys Asn Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala
 1175 1180 1185
 Arg Lys Val Leu Trp Ala Ile Gly Gln Phe Lys Lys Ala Glu Asp
 1190 1195 1200
 Glu Lys Leu Asp Lys Val Lys Ile Ala Ser Asn Lys Glu Trp Leu
 1205 1210 1215

Glu Tyr Ala Gln Thr Ser Val Lys His
 1220 1225

<210> 10

<211> 1264

<212> PRT

<213> Leptospira inadai

<400> 10

Met Glu Asp Tyr Ser Gly Phe Val Asn Ile Tyr Ser Ile Gln Lys Thr

1 5 10 15
 Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu His Ile Glu
 20 25 30
 Lys Lys Gly Phe Leu Lys Lys Asp Lys Ile Arg Ala Glu Asp Tyr Lys
 35 40 45

 Ala Val Lys Lys Ile Ile Asp Lys Tyr His Arg Ala Tyr Ile Glu Glu
 50 55 60
 Val Phe Asp Ser Val Leu His Gln Lys Lys Lys Lys Asp Lys Thr Arg
 65 70 75 80
 Phe Ser Thr Gln Phe Ile Lys Glu Ile Lys Glu Phe Ser Glu Leu Tyr
 85 90 95
 Tyr Lys Thr Glu Lys Asn Ile Pro Asp Lys Glu Arg Leu Glu Ala Leu
 100 105 110

 Ser Glu Lys Leu Arg Lys Met Leu Val Gly Ala Phe Lys Gly Glu Phe
 115 120 125
 Ser Glu Glu Val Ala Glu Lys Tyr Asn Lys Asn Leu Phe Ser Lys Glu
 130 135 140
 Leu Ile Arg Asn Glu Ile Glu Lys Phe Cys Glu Thr Asp Glu Glu Arg
 145 150 155 160
 Lys Gln Val Ser Asn Phe Lys Ser Phe Thr Thr Tyr Phe Thr Gly Phe
 165 170 175

 His Ser Asn Arg Gln Asn Ile Tyr Ser Asp Glu Lys Lys Ser Thr Ala
 180 185 190
 Ile Gly Tyr Arg Ile Ile His Gln Asn Leu Pro Lys Phe Leu Asp Asn
 195 200 205
 Leu Lys Ile Ile Glu Ser Ile Gln Arg Arg Phe Lys Asp Phe Pro Trp
 210 215 220
 Ser Asp Leu Lys Lys Asn Leu Lys Lys Ile Asp Lys Asn Ile Lys Leu
 225 230 235 240

 Thr Glu Tyr Phe Ser Ile Asp Gly Phe Val Asn Val Leu Asn Gln Lys
 245 250 255

Gly Ile Asp Ala Tyr Asn Thr Ile Leu Gly Gly Lys Ser Glu Glu Ser
260 265 270

Gly Glu Lys Ile Gln Gly Leu Asn Glu Tyr Ile Asn Leu Tyr Arg Gln
275 280 285

Lys Asn Asn Ile Asp Arg Lys Asn Pro Leu Asn Val Lys Ile Leu Phe
290 295 300

Lys Gln Ile Leu Gly Asp Arg Glu Thr Lys Ser Phe Ile Pro Glu Ala
305 310 315 320

Phe Pro Asp Asp Gln Ser Val Leu Asn Ser Ile Thr Glu Phe Ala Lys
325 330 335

Tyr Leu Lys Leu Asp Lys Lys Lys Lys Ser Ile Ile Ala Glu Leu Lys
340 345 350

Lys Phe Leu Ser Ser Phe Asn Arg Tyr Glu Leu Asp Gly Ile Tyr Leu
355 360 365

Ala Asn Asp Asn Ser Leu Ala Ser Ile Ser Thr Phe Leu Phe Asp Asp
370 375 380

Trp Ser Phe Ile Lys Lys Ser Val Ser Phe Lys Tyr Asp Glu Ser Val
385 390 395 400

Gly Asp Pro Lys Lys Lys Ile Lys Ser Pro Leu Lys Tyr Glu Lys Glu
405 410 415

Lys Glu Lys Trp Leu Lys Gln Lys Tyr Tyr Thr Ile Ser Phe Leu Asn
420 425 430

Asp Ala Ile Glu Ser Tyr Ser Lys Ser Gln Asp Glu Lys Arg Val Lys
435 440 445

Ile Arg Leu Glu Ala Tyr Phe Ala Glu Phe Lys Ser Lys Asp Asp Ala
450 455 460

Lys Lys Gln Phe Asp Leu Leu Glu Arg Ile Glu Glu Ala Tyr Ala Ile
465 470 475 480

Val Glu Pro Leu Leu Gly Ala Glu Tyr Pro Arg Asp Arg Asn Leu Lys
485 490 495

Ala Asp Lys Lys Glu Val Gly Lys Ile Lys Asp Phe Leu Asp Ser Ile

500	505	510	
Lys Ser Leu Gln Phe Phe Leu Lys Pro Leu Leu Ser Ala Glu Ile Phe			
515	520	525	
Asp Glu Lys Asp Leu Gly Phe Tyr Asn Gln Leu Glu Gly Tyr Tyr Glu			
530	535	540	
Glu Ile Asp Ile Ser Gly His Leu Tyr Asn Lys Val Arg Asn Tyr Leu			
545	550	555	560
Thr Gly Lys Ile Tyr Ser Lys Glu Lys Phe Lys Leu Asn Phe Glu Asn			
565	570	575	
Ser Thr Leu Leu Lys Gly Trp Asp Glu Asn Arg Glu Val Ala Asn Leu			
580	585	590	
Cys Val Ile Phe Arg Glu Asp Gln Lys Tyr Tyr Leu Gly Val Met Asp			
595	600	605	
Lys Glu Asn Asn Thr Ile Leu Ser Asp Ile Pro Lys Val Lys Pro Asn			
610	615	620	
Glu Leu Phe Tyr Glu Lys Met Val Tyr Lys Leu Ile Pro Thr Pro His			
625	630	635	640
Met Gln Leu Pro Arg Ile Ile Phe Ser Ser Asp Asn Leu Ser Ile Tyr			
645	650	655	
Asn Pro Ser Lys Ser Ile Leu Lys Ile Arg Glu Ala Lys Ser Phe Lys			
660	665	670	
Glu Gly Lys Asn Phe Lys Leu Lys Asp Cys His Lys Phe Ile Asp Phe			
675	680	685	
Tyr Lys Glu Ser Ile Ser Lys Asn Glu Asp Trp Ser Arg Phe Asp Phe			
690	695	700	
Lys Phe Ser Lys Thr Ser Ser Tyr Glu Asn Ile Ser Glu Phe Tyr Arg			
705	710	715	720
Glu Val Glu Arg Gln Gly Tyr Asn Leu Asp Phe Lys Lys Val Ser Lys			
725	730	735	
Phe Tyr Ile Asp Ser Leu Val Glu Asp Gly Lys Leu Tyr Leu Phe Gln			
740	745	750	

Ile Tyr Asn Lys Asp Phe Ser Ile Phe Ser Lys Gly Lys Pro Asn Leu

755 760 765

His Thr Ile Tyr Phe Arg Ser Leu Phe Ser Lys Glu Asn Leu Lys Asp

770 775 780

Val Cys Leu Lys Leu Asn Gly Glu Ala Glu Met Phe Phe Arg Lys Lys

785 790 795 800

Ser Ile Asn Tyr Asp Glu Lys Lys Lys Arg Glu Gly His His Pro Glu

805 810 815

Leu Phe Glu Lys Leu Lys Tyr Pro Ile Leu Lys Asp Lys Arg Tyr Ser

820 825 830

Glu Asp Lys Phe Gln Phe His Leu Pro Ile Ser Leu Asn Phe Lys Ser

835 840 845

Lys Glu Arg Leu Asn Phe Asn Leu Lys Val Asn Glu Phe Leu Lys Arg

850 855 860

Asn Lys Asp Ile Asn Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu

865 870 875 880

Leu Tyr Leu Val Met Ile Asn Gln Lys Gly Glu Ile Leu Lys Gln Thr

885 890 895

Leu Leu Asp Ser Met Gln Ser Gly Lys Gly Arg Pro Glu Ile Asn Tyr

900 905 910

Lys Glu Lys Leu Gln Glu Lys Glu Ile Glu Arg Asp Lys Ala Arg Lys

915 920 925

Ser Trp Gly Thr Val Glu Asn Ile Lys Glu Leu Lys Glu Gly Tyr Leu

930 935 940

Ser Ile Val Ile His Gln Ile Ser Lys Leu Met Val Glu Asn Asn Ala

945 950 955 960

Ile Val Val Leu Glu Asp Leu Asn Ile Gly Phe Lys Arg Gly Arg Gln

965 970 975

Lys Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Lys Met Leu Ile Asp

980 985 990

Lys Leu Asn Phe Leu Val Phe Lys Glu Asn Lys Pro Thr Glu Pro Gly

995	1000	1005
Gly Val Leu Lys Ala Tyr Gln	Leu Thr Asp Glu Phe	Gln Ser Phe
1010	1015	1020
Glu Lys Leu Ser Lys Gln Thr	Gly Phe Leu Phe Tyr	Val Pro Ser
1025	1030	1035
Trp Asn Thr Ser Lys Ile Asp	Pro Arg Thr Gly Phe	Ile Asp Phe
1040	1045	1050
Leu His Pro Ala Tyr Glu Asn	Ile Glu Lys Ala Lys	Gln Trp Ile
1055	1060	1065
Asn Lys Phe Asp Ser Ile Arg	Phe Asn Ser Lys Met	Asp Trp Phe
1070	1075	1080
Glu Phe Thr Ala Asp Thr Arg	Lys Phe Ser Glu Asn	Leu Met Leu
1085	1090	1095
Gly Lys Asn Arg Val Trp Val	Ile Cys Thr Thr Asn	Val Glu Arg
1100	1105	1110
Tyr Phe Thr Ser Lys Thr Ala	Asn Ser Ser Ile Gln	Tyr Asn Ser
1115	1120	1125
Ile Gln Ile Thr Glu Lys Leu	Lys Glu Leu Phe Val	Asp Ile Pro
1130	1135	1140
Phe Ser Asn Gly Gln Asp Leu	Lys Pro Glu Ile Leu	Arg Lys Asn
1145	1150	1155
Asp Ala Val Phe Phe Lys Ser	Leu Leu Phe Tyr Ile	Lys Thr Thr
1160	1165	1170
Leu Ser Leu Arg Gln Asn Asn	Gly Lys Lys Gly Glu	Glu Glu Lys
1175	1180	1185
Asp Phe Ile Leu Ser Pro Val	Val Asp Ser Lys Gly	Arg Phe Phe
1190	1195	1200
Asn Ser Leu Glu Ala Ser Asp	Asp Glu Pro Lys Asp	Ala Asp Ala
1205	1210	1215
Asn Gly Ala Tyr His Ile Ala	Leu Lys Gly Leu Met	Asn Leu Leu
1220	1225	1230

Val Leu Asn Glu Thr Lys Glu Glu Asn Leu Ser Arg Pro Lys Trp
1235 1240 1245

Lys Ile Lys Asn Lys Asp Trp Leu Glu Phe Val Trp Glu Arg Asn
1250 1255 1260

Arg

<210> 11

<211> 1373

<212> PRT

<213> Moraxella bovoculi

<400> 11

Met Leu Phe Gln Asp Phe Thr His Leu Tyr Pro Leu Ser Lys Thr Val
1 5 10 15

Arg Phe Glu Leu Phe Ile Asp Arg Thr Leu Glu His Ile His Ala Lys
20 25 30

Asn Phe Leu Ser Gln Asp Glu Thr Met Ala Asp Met His Gln Lys Val
35 40 45

Lys Val Ile Leu Asp Asp Tyr His Arg Asp Phe Ile Ala Asp Met Met
50 55 60

Gly Glu Val Lys Leu Thr Lys Leu Ala Glu Phe Tyr Asp Val Tyr Leu
65 70 75 80

Lys Phe Arg Lys Asn Pro Lys Asp Asp Glu Leu Gln Lys Ala Gln Leu
85 90 95

Lys Asp Leu Gln Ala Val Leu Arg Lys Glu Ile Val Lys Pro Ile Gly
100 105 110

Asn Gly Gly Lys Tyr Lys Ala Gly Tyr Asp Arg Leu Phe Gly Ala Lys
115 120 125

Leu Phe Lys Asp Gly Lys Glu Leu Gly Asp Leu Ala Lys Phe Val Ile
130 135 140

Ala Gln Glu Gly Glu Ser Ser Pro Lys Leu Ala His Leu Ala His Phe
145 150 155 160

Glu Lys Phe Ser Thr Tyr Phe Thr Gly Phe His Asp Asn Arg Lys Asn

165 170 175

Met Tyr Ser Asp Glu Asp Lys His Thr Ala Ile Ala Tyr Arg Leu Ile

180 185 190

His Glu Asn Leu Pro Arg Phe Ile Asp Asn Leu Gln Ile Leu Thr Thr

195 200 205

Ile Lys Gln Lys His Ser Ala Leu Tyr Asp Gln Ile Ile Asn Glu Leu

210 215 220

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

225 230 235 240

His Lys Leu Leu Thr Gln Glu Gly Ile Thr Ala Tyr Asn Thr Leu Leu

245 250 255

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

260 265 270

Glu Leu Ile Asn Ser His His Asn Gln His Cys His Lys Ser Glu Arg

275 280 285

Ile Ala Lys Leu Arg Pro Leu His Lys Gln Ile Leu Ser Asp Gly Met

290 295 300

Ser Val Ser Phe Leu Pro Ser Lys Phe Ala Asp Asp Ser Glu Met Cys

305 310 315 320

Gln Ala Val Asn Glu Phe Tyr Arg His Tyr Ala Asp Val Phe Ala Lys

325 330 335

Val Gln Ser Leu Phe Asp Gly Phe Asp Asp His Gln Lys Asp Gly Ile

340 345 350

Tyr Val Glu His Lys Asn Leu Asn Glu Leu Ser Lys Gln Ala Phe Gly

355 360 365

Asp Phe Ala Leu Leu Gly Arg Val Leu Asp Gly Tyr Tyr Val Asp Val

370 375 380

Val Asn Pro Glu Phe Asn Glu Arg Phe Ala Lys Ala Lys Thr Asp Asn

385 390 395 400

Ala Lys Ala Lys Leu Thr Lys Glu Lys Asp Lys Phe Ile Lys Gly Val

405 410 415
His Ser Leu Ala Ser Leu Glu Gln Ala Ile Glu His Tyr Thr Ala Arg

420 425 430
His Asp Asp Glu Ser Val Gln Ala Gly Lys Leu Gly Gln Tyr Phe Lys

435 440 445
His Gly Leu Ala Gly Val Asp Asn Pro Ile Gln Lys Ile His Asn Asn

450 455 460
His Ser Thr Ile Lys Gly Phe Leu Glu Arg Glu Arg Pro Ala Gly Glu

465 470 475 480
Arg Ala Leu Pro Lys Ile Lys Ser Gly Lys Asn Pro Glu Met Thr Gln

485 490 495
Leu Arg Gln Leu Lys Glu Leu Leu Asp Asn Ala Leu Asn Val Ala His

500 505 510
Phe Ala Lys Leu Leu Thr Thr Lys Thr Thr Leu Asp Asn Gln Asp Gly

515 520 525
Asn Phe Tyr Gly Glu Phe Gly Val Leu Tyr Asp Glu Leu Ala Lys Ile

530 535 540
Pro Thr Leu Tyr Asn Lys Val Arg Asp Tyr Leu Ser Gln Lys Pro Phe

545 550 555 560
Ser Thr Glu Lys Tyr Lys Leu Asn Phe Gly Asn Pro Thr Leu Leu Asn

565 570 575
Gly Trp Asp Leu Asn Lys Glu Lys Asp Asn Phe Gly Val Ile Leu Gln

580 585 590
Lys Asp Gly Cys Tyr Tyr Leu Ala Leu Leu Asp Lys Ala His Lys Lys

595 600 605
Val Phe Asp Asn Ala Pro Asn Thr Gly Lys Ser Ile Tyr Gln Lys Met

610 615 620
Ile Tyr Lys Tyr Leu Glu Val Arg Lys Gln Phe Pro Lys Val Phe Phe

625 630 635 640
Ser Lys Glu Ala Ile Ala Ile Asn Tyr His Pro Ser Lys Glu Leu Val

645 650 655

Glu Ile Lys Asp Lys Gly Arg Gln Arg Ser Asp Asp Glu Arg Leu Lys
 660 665 670
 Leu Tyr Arg Phe Ile Leu Glu Cys Leu Lys Ile His Pro Lys Tyr Asp
 675 680 685
 Lys Lys Phe Glu Gly Ala Ile Gly Asp Ile Gln Leu Phe Lys Lys Asp
 690 695 700
 Lys Lys Gly Arg Glu Val Pro Ile Ser Glu Lys Asp Leu Phe Lys Asp
 705 710 715 720
 Ile Asn Gly Ile Phe Ser Ser Lys Pro Lys Leu Glu Met Glu Asp Phe
 725 730 735
 Phe Ile Gly Glu Phe Lys Arg Tyr Asn Pro Ser Gln Asp Leu Val Asp
 740 745 750
 Gln Tyr Asn Ile Tyr Lys Lys Ile Asp Ser Asn Asp Asn Arg Lys Lys
 755 760 765
 Glu Asn Phe Tyr Asn Asn His Pro Lys Phe Lys Lys Asp Leu Val Arg
 770 775 780
 Tyr Tyr Tyr Glu Ser Met Cys Lys His Glu Glu Trp Glu Glu Ser Phe
 785 790 795 800
 Glu Phe Ser Lys Lys Leu Gln Asp Ile Gly Cys Tyr Val Asp Val Asn
 805 810 815
 Glu Leu Phe Thr Glu Ile Glu Thr Arg Arg Leu Asn Tyr Lys Ile Ser
 820 825 830
 Phe Cys Asn Ile Asn Ala Asp Tyr Ile Asp Glu Leu Val Glu Gln Gly
 835 840 845
 Gln Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Pro Lys Ala
 850 855 860
 His Gly Lys Pro Asn Leu His Thr Leu Tyr Phe Lys Ala Leu Phe Ser
 865 870 875 880
 Glu Asp Asn Leu Ala Asp Pro Ile Tyr Lys Leu Asn Gly Glu Ala Gln
 885 890 895
 Ile Phe Tyr Arg Lys Ala Ser Leu Asp Met Asn Glu Thr Thr Ile His

900	905	910	
Arg Ala Gly Glu Val Leu Glu Asn Lys Asn Pro Asp Asn Pro Lys Lys			
915	920	925	
Arg Gln Phe Val Tyr Asp Ile Ile Lys Asp Lys Arg Tyr Thr Gln Lys			
930	935	940	
Asp Phe Met Leu His Val Pro Ile Thr Met Asn Phe Gly Val Gln Gly			
945	950	955	960
Met Thr Ile Lys Glu Phe Asn Lys Lys Val Asn Gln Ser Ile Gln Gln			
965	970	975	
Tyr Asp Glu Val Asn Val Ile Gly Ile Asp Arg Gly Glu Arg His Leu			
980	985	990	
Leu Tyr Leu Thr Val Ile Asn Ser Lys Gly Glu Ile Leu Glu Gln Cys			
995	1000	1005	
Ser Leu Asn Asp Ile Thr Thr Ala Ser Ala Asn Gly Thr Gln Met			
1010	1015	1020	
Thr Thr Pro Tyr His Lys Ile Leu Asp Lys Arg Glu Ile Glu Arg			
1025	1030	1035	
Leu Asn Ala Arg Val Gly Trp Gly Glu Ile Glu Thr Ile Lys Glu			
1040	1045	1050	
Leu Lys Ser Gly Tyr Leu Ser His Val Val His Gln Ile Ser Gln			
1055	1060	1065	
Leu Met Leu Lys Tyr Asn Ala Ile Val Val Leu Glu Asp Leu Asn			
1070	1075	1080	
Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Ile Tyr			
1085	1090	1095	
Gln Asn Phe Glu Asn Ala Leu Ile Lys Lys Leu Asn His Leu Val			
1100	1105	1110	
Leu Lys Asp Lys Ala Asp Asp Glu Ile Gly Ser Tyr Lys Asn Ala			
1115	1120	1125	
Leu Gln Leu Thr Asn Asn Phe Thr Asp Leu Lys Ser Ile Gly Lys			
1130	1135	1140	

Gln Thr Gly Phe Leu Phe Tyr Val Pro Ala Trp Asn Thr Ser Lys		
1145	1150	1155
Ile Asp Pro Glu Thr Gly Phe Val Asp Leu Leu Lys Pro Arg Tyr		
1160	1165	1170
Glu Asn Ile Gln Ala Ser Gln Ala Phe Phe Gly Lys Phe Asp Lys		
1175	1180	1185
Ile Cys Tyr Asn Ala Asp Lys Asp Tyr Phe Glu Phe His Ile Asp		
1190	1195	1200
Tyr Ala Lys Phe Thr Asp Lys Ala Lys Asn Ser Arg Gln Ile Trp		
1205	1210	1215
Thr Ile Cys Ser His Gly Asp Lys Arg Tyr Val Tyr Asp Lys Thr		
1220	1225	1230
Ala Asn Gln Asn Lys Gly Ala Ala Lys Gly Ile Asn Val Asn Asp		
1235	1240	1245
Ile Leu Lys Ser Leu Phe Ala Arg His His Ile Asn Glu Lys Gln		
1250	1255	1260
Pro Asn Leu Val Met Asp Ile Cys Gln Asn Asn Asp Lys Glu Phe		
1265	1270	1275
His Lys Ser Leu Met Tyr Leu Leu Lys Thr Leu Leu Ala Leu Arg		
1280	1285	1290
Tyr Ser Asn Ala Ser Ser Asp Glu Asp Phe Ile Leu Ser Pro Val		
1295	1300	1305
Ala Asn Asp Glu Gly Val Phe Phe Asn Ser Ala Leu Ala Asp Asp		
1310	1315	1320
Thr Gln Pro Gln Asn Ala Asp Ala Asn Gly Ala Tyr His Ile Ala		
1325	1330	1335
Leu Lys Gly Leu Trp Leu Leu Asn Glu Leu Lys Asn Ser Asp Asp		
1340	1345	1350
Leu Asn Lys Val Lys Leu Ala Ile Asp Asn Gln Thr Trp Leu Asn		
1355	1360	1365
Phe Ala Gln Asn Arg		

1370

<210> 12

<211> 1352

<212> PRT

<213> Unknown

<220><223> Parcubacteria bacterium

<400> 12

Met Glu Asn Ile Phe Asp Gln Phe Ile Gly Lys Tyr Ser Leu Ser Lys

1 5 10 15

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

20 25 30

Lys Ile Asn Lys Val Phe Glu Lys Asp Gln Thr Ile Asp Asp Ser Tyr

35 40 45

Asn Gln Ala Lys Phe Tyr Phe Asp Ser Leu His Gln Lys Phe Ile Asp

50 55 60

Ala Ala Leu Ala Ser Asp Lys Thr Ser Glu Leu Ser Phe Gln Asn Phe

65 70 75 80

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

85 90 95

Glu Met Gly Ala Leu Arg Lys Arg Asp Lys Asn Ala Val Gly Ile Asp

100 105 110

Arg Leu Gln Lys Glu Ile Asn Asp Ala Glu Asp Ile Ile Gln Lys Glu

115 120 125

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

130 135 140

Glu Ser Trp Lys Thr Tyr Tyr Gln Glu Arg Glu Val Asp Gly Lys Lys

145 150 155 160

Ile Thr Glu Ser Lys Ala Asp Leu Lys Gln Lys Gly Ala Asp Phe Leu

165 170 175

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

180 185 190

Lys Glu Lys Glu Phe Gln Ala Lys Asn Gln Pro Ser Leu Phe Val Glu

195 200 205
 Glu Lys Glu Asn Pro Gly Gln Lys Arg Tyr Ile Phe Asp Ser Phe Asp
 210 215 220
 Lys Phe Ala Gly Tyr Leu Thr Lys Phe Gln Gln Thr Lys Lys Asn Leu

 225 230 235 240
 Tyr Ala Ala Asp Gly Thr Ser Thr Ala Val Ala Thr Arg Ile Ala Asp
 245 250 255
 Asn Phe Ile Ile Phe His Gln Asn Thr Lys Val Phe Arg Asp Lys Tyr
 260 265 270
 Lys Asn Asn His Thr Asp Leu Gly Phe Asp Glu Glu Asn Ile Phe Glu
 275 280 285
 Ile Glu Arg Tyr Lys Asn Cys Leu Leu Gln Arg Glu Ile Glu His Ile

 290 295 300
 Lys Asn Glu Asn Ser Tyr Asn Lys Ile Ile Gly Arg Ile Asn Lys Lys
 305 310 315 320
 Ile Lys Glu Tyr Arg Asp Gln Lys Ala Lys Asp Thr Lys Leu Thr Lys
 325 330 335
 Ser Asp Phe Pro Phe Phe Lys Asn Leu Asp Lys Gln Ile Leu Gly Glu
 340 345 350
 Val Glu Lys Glu Lys Gln Leu Ile Glu Lys Thr Arg Glu Lys Thr Glu

 355 360 365
 Glu Asp Val Leu Ile Glu Arg Phe Lys Glu Phe Ile Glu Asn Asn Glu
 370 375 380
 Glu Arg Phe Thr Ala Ala Lys Lys Leu Met Asn Ala Phe Cys Asn Gly
 385 390 395 400
 Glu Phe Glu Ser Glu Tyr Glu Gly Ile Tyr Leu Lys Asn Lys Ala Ile
 405 410 415
 Asn Thr Ile Ser Arg Arg Trp Phe Val Ser Asp Arg Asp Phe Glu Leu

 420 425 430
 Lys Leu Pro Gln Gln Lys Ser Lys Asn Lys Ser Glu Lys Asn Glu Pro
 435 440 445

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

690 695 700
 Gln Glu Ile Leu Arg Ile Lys Lys Glu Lys Thr Phe Lys Arg Glu Ser
 705 710 715 720
 Lys Asn Phe Ser Leu Arg Asp Leu His Ala Leu Ile Glu Tyr Tyr Arg
 725 730 735
 Asn Cys Ile Pro Gln Tyr Ser Asn Trp Ser Phe Tyr Asp Phe Gln Phe

 740 745 750
 Gln Asp Thr Gly Lys Tyr Gln Asn Ile Lys Glu Phe Thr Asp Asp Val
 755 760 765
 Gln Lys Tyr Gly Tyr Lys Ile Ser Phe Arg Asp Ile Asp Asp Glu Tyr
 770 775 780
 Ile Asn Gln Ala Leu Asn Glu Gly Lys Met Tyr Leu Phe Glu Val Val
 785 790 795 800
 Asn Lys Asp Ile Tyr Asn Thr Lys Asn Gly Ser Lys Asn Leu His Thr

 805 810 815
 Leu Tyr Phe Glu His Ile Leu Ser Ala Glu Asn Leu Asn Asp Pro Val
 820 825 830
 Phe Lys Leu Ser Gly Met Ala Glu Ile Phe Gln Arg Gln Pro Ser Val
 835 840 845
 Asn Glu Arg Glu Lys Ile Thr Thr Gln Lys Asn Gln Cys Ile Leu Asp
 850 855 860
 Lys Gly Asp Arg Ala Tyr Lys Tyr Arg Arg Tyr Thr Glu Lys Lys Ile

 865 870 875 880
 Met Phe His Met Ser Leu Val Leu Asn Thr Gly Lys Gly Glu Ile Lys
 885 890 895
 Gln Val Gln Phe Asn Lys Ile Ile Asn Gln Arg Ile Ser Ser Ser Asp
 900 905 910
 Asn Glu Met Arg Val Asn Val Ile Gly Ile Asp Arg Gly Glu Lys Asn
 915 920 925
 Leu Leu Tyr Tyr Ser Val Val Lys Gln Asn Gly Glu Ile Ile Glu Gln

 930 935 940

Ala Ser Leu Asn Glu Ile Asn Gly Val Asn Tyr Arg Asp Lys Leu Ile
945 950 955 960
Glu Arg Glu Lys Glu Arg Leu Lys Asn Arg Gln Ser Trp Lys Pro Val
965 970 975
Val Lys Ile Lys Asp Leu Lys Lys Gly Tyr Ile Ser His Val Ile His
980 985 990
Lys Ile Cys Gln Leu Ile Glu Lys Tyr Ser Ala Ile Val Val Leu Glu
995 1000 1005
Asp Leu Asn Met Arg Phe Lys Gln Ile Arg Gly Gly Ile Glu Arg
1010 1015 1020
Ser Val Tyr Gln Gln Phe Glu Lys Ala Leu Ile Asp Lys Leu Gly
1025 1030 1035
Tyr Leu Val Phe Lys Asp Asn Arg Asp Leu Arg Ala Pro Gly Gly
1040 1045 1050
Val Leu Asn Gly Tyr Gln Leu Ser Ala Pro Phe Val Ser Phe Glu
1055 1060 1065
Lys Met Arg Lys Gln Thr Gly Ile Leu Phe Tyr Thr Gln Ala Glu
1070 1075 1080
Tyr Thr Ser Lys Thr Asp Pro Ile Thr Gly Phe Arg Lys Asn Val
1085 1090 1095
Tyr Ile Ser Asn Ser Ala Ser Leu Asp Lys Ile Lys Glu Ala Val
1100 1105 1110
Lys Lys Phe Asp Ala Ile Gly Trp Asp Gly Lys Glu Gln Ser Tyr
1115 1120 1125
Phe Phe Lys Tyr Asn Pro Tyr Asn Leu Ala Asp Glu Lys Tyr Lys
1130 1135 1140
Asn Ser Thr Val Ser Lys Glu Trp Ala Ile Phe Ala Ser Ala Pro
1145 1150 1155
Arg Ile Arg Arg Gln Lys Gly Glu Asp Gly Tyr Trp Lys Tyr Asp
1160 1165 1170
Arg Val Lys Val Asn Glu Glu Phe Glu Lys Leu Leu Lys Val Trp

1175 1180 1185
 Asn Phe Val Asn Pro Lys Ala Thr Asp Ile Lys Gln Glu Ile Ile
 1190 1195 1200
 Lys Lys Ile Lys Ala Gly Asp Leu Gln Gly Glu Lys Glu Leu Asp
 1205 1210 1215
 Gly Arg Leu Arg Asn Phe Trp His Ser Phe Ile Tyr Leu Phe Asn
 1220 1225 1230
 Leu Val Leu Glu Leu Arg Asn Ser Phe Ser Leu Gln Ile Lys Ile

1235 1240 1245
 Lys Ala Gly Glu Val Ile Ala Val Asp Glu Gly Val Asp Phe Ile
 1250 1255 1260
 Ala Ser Pro Val Lys Pro Phe Phe Thr Thr Pro Asn Pro Tyr Ile
 1265 1270 1275
 Pro Ser Asn Leu Cys Trp Leu Ala Val Glu Asn Ala Asp Ala Asn
 1280 1285 1290
 Gly Ala Tyr Asn Ile Ala Arg Lys Gly Val Met Ile Leu Lys Lys

1295 1300 1305
 Ile Arg Glu His Ala Lys Lys Asp Pro Glu Phe Lys Lys Leu Pro
 1310 1315 1320
 Asn Leu Phe Ile Ser Asn Ala Glu Trp Asp Glu Ala Ala Arg Asp
 1325 1330 1335
 Trp Gly Lys Tyr Ala Gly Thr Thr Ala Leu Asn Leu Asp His
 1340 1345 1350

<210> 13

<211> 1260

<212> PRT

<213> Porphyromonas crevioricanis

<400> 13

Met Asp Ser Leu Lys Asp Phe Thr Asn Leu Tyr Pro Val Ser Lys Thr

1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu Asn Ile Glu

20 25 30

Lys Ala Gly Ile Leu Lys Glu Asp Glu His Arg Ala Glu Ser Tyr Arg
 35 40 45
 Arg Val Lys Lys Ile Ile Asp Thr Tyr His Lys Val Phe Ile Asp Ser
 50 55 60

 Ser Leu Glu Asn Met Ala Lys Met Gly Ile Glu Asn Glu Ile Lys Ala
 65 70 75 80
 Met Leu Gln Ser Phe Cys Glu Leu Tyr Lys Lys Asp His Arg Thr Glu
 85 90 95
 Gly Glu Asp Lys Ala Leu Asp Lys Ile Arg Ala Val Leu Arg Gly Leu
 100 105 110
 Ile Val Gly Ala Phe Thr Gly Val Cys Gly Arg Arg Glu Asn Thr Val
 115 120 125

 Gln Asn Glu Lys Tyr Glu Ser Leu Phe Lys Glu Lys Leu Ile Lys Glu
 130 135 140
 Ile Leu Pro Asp Phe Val Leu Ser Thr Glu Ala Glu Ser Leu Pro Phe
 145 150 155 160
 Ser Val Glu Glu Ala Thr Arg Ser Leu Lys Glu Phe Asp Ser Phe Thr
 165 170 175
 Ser Tyr Phe Ala Gly Phe Tyr Glu Asn Arg Lys Asn Ile Tyr Ser Thr
 180 185 190

 Lys Pro Gln Ser Thr Ala Ile Ala Tyr Arg Leu Ile His Glu Asn Leu
 195 200 205
 Pro Lys Phe Ile Asp Asn Ile Leu Val Phe Gln Lys Ile Lys Glu Pro
 210 215 220
 Ile Ala Lys Glu Leu Glu His Ile Arg Ala Asp Phe Ser Ala Gly Gly
 225 230 235 240
 Tyr Ile Lys Lys Asp Glu Arg Leu Glu Asp Ile Phe Ser Leu Asn Tyr
 245 250 255

 Tyr Ile His Val Leu Ser Gln Ala Gly Ile Glu Lys Tyr Asn Ala Leu
 260 265 270
 Ile Gly Lys Ile Val Thr Glu Gly Asp Gly Glu Met Lys Gly Leu Asn

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

 Leu Ser Tyr Leu Pro Glu Ser Phe Glu Lys Asp Glu Glu Leu Leu Arg
 325 330 335
 Ala Leu Lys Glu Phe Tyr Asp His Ile Ala Glu Asp Ile Leu Gly Arg
 340 345 350
 Thr Gln Gln Leu Met Thr Ser Ile Ser Glu Tyr Asp Leu Ser Arg Ile
 355 360 365
 Tyr Val Arg Asn Asp Ser Gln Leu Thr Asp Ile Ser Lys Lys Met Leu
 370 375 380

 Gly Asp Trp Asn Ala Ile Tyr Met Ala Arg Glu Arg Ala Tyr Asp His
 385 390 395 400
 Glu Gln Ala Pro Lys Arg Ile Thr Ala Lys Tyr Glu Arg Asp Arg Ile
 405 410 415
 Lys Ala Leu Lys Gly Glu Glu Ser Ile Ser Leu Ala Asn Leu Asn Ser
 420 425 430
 Cys Ile Ala Phe Leu Asp Asn Val Arg Asp Cys Arg Val Asp Thr Tyr
 435 440 445

 Leu Ser Thr Leu Gly Gln Lys Glu Gly Pro His Gly Leu Ser Asn Leu
 450 455 460
 Val Glu Asn Val Phe Ala Ser Tyr His Glu Ala Glu Gln Leu Leu Ser
 465 470 475 480
 Phe Pro Tyr Pro Glu Glu Asn Asn Leu Ile Gln Asp Lys Asp Asn Val
 485 490 495
 Val Leu Ile Lys Asn Leu Leu Asp Asn Ile Ser Asp Leu Gln Arg Phe
 500 505 510

 Leu Lys Pro Leu Trp Gly Met Gly Asp Glu Pro Asp Lys Asp Glu Arg
 515 520 525

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

 Trp Asp Arg Asn Lys Glu Lys Asp Asn Ser Cys Val Ile Leu Arg Lys
 580 585 590
 Gly Gln Asn Phe Tyr Leu Ala Ile Met Asn Asn Arg His Lys Arg Ser
 595 600 605
 Phe Glu Asn Lys Met Leu Pro Glu Tyr Lys Glu Gly Glu Pro Tyr Phe
 610 615 620
 Glu Lys Met Asp Tyr Lys Phe Leu Pro Asp Pro Asn Lys Met Leu Pro
 625 630 635 640

 Lys Val Phe Leu Ser Lys Lys Gly Ile Glu Ile Tyr Lys Pro Ser Pro
 645 650 655
 Lys Leu Leu Glu Gln Tyr Gly His Gly Thr His Lys Lys Gly Asp Thr
 660 665 670
 Phe Ser Met Asp Asp Leu His Glu Leu Ile Asp Phe Phe Lys His Ser
 675 680 685
 Ile Glu Ala His Glu Asp Trp Lys Gln Phe Gly Phe Lys Phe Ser Asp
 690 695 700

 Thr Ala Thr Tyr Glu Asn Val Ser Ser Phe Tyr Arg Glu Val Glu Asp
 705 710 715 720
 Gln Gly Tyr Lys Leu Ser Phe Arg Lys Val Ser Glu Ser Tyr Val Tyr
 725 730 735
 Ser Leu Ile Asp Gln Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys
 740 745 750
 Asp Phe Ser Pro Cys Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr
 755 760 765

 Trp Arg Met Leu Phe Asp Glu Arg Asn Leu Ala Asp Val Ile Tyr Lys

770	775	780	
Leu Asp Gly Lys Ala Glu Ile Phe Phe Arg Glu Lys Ser Leu Lys Asn			
785	790	795	800
Asp His Pro Thr His Pro Ala Gly Lys Pro Ile Lys Lys Lys Ser Arg			
	805	810	815
Gln Lys Lys Gly Glu Glu Ser Leu Phe Glu Tyr Asp Leu Val Lys Asp			
	820	825	830
Arg Arg Tyr Thr Met Asp Lys Phe Gln Phe His Val Pro Ile Thr Met			
	835	840	845
Asn Phe Lys Cys Ser Ala Gly Ser Lys Val Asn Asp Met Val Asn Ala			
	850	855	860
His Ile Arg Glu Ala Lys Asp Met His Val Ile Gly Ile Asp Arg Gly			
865	870	875	880
Glu Arg Asn Leu Leu Tyr Ile Cys Val Ile Asp Ser Arg Gly Thr Ile			
	885	890	895
Leu Asp Gln Ile Ser Leu Asn Thr Ile Asn Asp Ile Asp Tyr His Asp			
	900	905	910
Leu Leu Glu Ser Arg Asp Lys Asp Arg Gln Gln Glu His Arg Asn Trp			
	915	920	925
Gln Thr Ile Glu Gly Ile Lys Glu Leu Lys Gln Gly Tyr Leu Ser Gln			
	930	935	940
Ala Val His Arg Ile Ala Glu Leu Met Val Ala Tyr Lys Ala Val Val			
945	950	955	960
Ala Leu Glu Asp Leu Asn Met Gly Phe Lys Arg Gly Arg Gln Lys Val			
	965	970	975
Glu Ser Ser Val Tyr Gln Gln Phe Glu Lys Gln Leu Ile Asp Lys Leu			
	980	985	990
Asn Tyr Leu Val Asp Lys Lys Lys Arg Pro Glu Asp Ile Gly Gly Leu			
	995	1000	1005
Leu Arg Ala Tyr Gln Phe Thr Ala Pro Phe Lys Ser Phe Lys Glu			
1010	1015	1020	

Met Gly	Lys Gln Asn Gly Phe	Leu Phe Tyr Ile Pro	Ala Trp Asn
1025	1030	1035	
Thr Ser	Asn Ile Asp Pro Thr	Thr Gly Phe Val Asn	Leu Phe His
1040	1045	1050	
Val Gln	Tyr Glu Asn Val Asp	Lys Ala Lys Ser Phe	Phe Gln Lys
1055	1060	1065	
Phe Asp	Ser Ile Ser Tyr Asn	Pro Lys Lys Asp Trp	Phe Glu Phe
1070	1075	1080	
Ala Phe	Asp Tyr Lys Asn Phe	Thr Lys Lys Ala Glu	Gly Ser Arg
1085	1090	1095	
Ser Met	Trp Ile Leu Cys Thr	His Gly Ser Arg Ile	Lys Asn Phe
1100	1105	1110	
Arg Asn	Ser Gln Lys Asn Gly	Gln Trp Asp Ser Glu	Glu Phe Ala
1115	1120	1125	
Leu Thr	Glu Ala Phe Lys Ser	Leu Phe Val Arg Tyr	Glu Ile Asp
1130	1135	1140	
Tyr Thr	Ala Asp Leu Lys Thr	Ala Ile Val Asp Glu	Lys Gln Lys
1145	1150	1155	
Asp Phe	Phe Val Asp Leu Leu	Lys Leu Phe Lys Leu	Thr Val Gln
1160	1165	1170	
Met Arg	Asn Ser Trp Lys Glu	Lys Asp Leu Asp Tyr	Leu Ile Ser
1175	1180	1185	
Pro Val	Ala Gly Ala Asp Gly	Arg Phe Phe Asp Thr	Arg Glu Gly
1190	1195	1200	
Asn Lys	Ser Leu Pro Lys Asp	Ala Asp Ala Asn Gly	Ala Tyr Asn
1205	1210	1215	
Ile Ala	Leu Lys Gly Leu Trp	Ala Leu Arg Gln Ile	Arg Gln Thr
1220	1225	1230	
Ser Glu	Gly Gly Lys Leu Lys	Leu Ala Ile Ser Asn	Lys Glu Trp
1235	1240	1245	
Leu Gln	Phe Val Gln Glu Arg	Ser Tyr Glu Lys Asp	
1250	1255	1260	

<210> 14

<

211> 1324

<212> PRT

<213> Prevotella disiens

<400> 14

Met Glu Asn Tyr Gln Glu Phe Thr Asn Leu Phe Gln Leu Asn Lys Thr

1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Cys Glu Leu Leu Glu

20 25 30

Glu Gly Lys Ile Phe Ala Ser Gly Ser Phe Leu Glu Lys Asp Lys Val

35 40 45

Arg Ala Asp Asn Val Ser Tyr Val Lys Lys Glu Ile Asp Lys Lys His

50 55 60

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

65 70 75 80

Leu Leu Lys Gln Tyr Phe Asp Cys Tyr Asn Glu Leu Lys Ala Phe Lys

85 90 95

Lys Asp Cys Lys Ser Asp Glu Glu Glu Val Lys Lys Thr Ala Leu Arg

100 105 110

Asn Lys Cys Thr Ser Ile Gln Arg Ala Met Arg Glu Ala Ile Ser Gln

115 120 125

Ala Phe Leu Lys Ser Pro Gln Lys Lys Leu Leu Ala Ile Lys Asn Leu

130 135 140

Ile Glu Asn Val Phe Lys Ala Asp Glu Asn Val Gln His Phe Ser Glu

145 150 155 160

Phe Thr Ser Tyr Phe Ser Gly Phe Glu Thr Asn Arg Glu Asn Phe Tyr

165 170 175

Ser Asp Glu Glu Lys Ser Thr Ser Ile Ala Tyr Arg Leu Val His Asp

180 185 190

Asn Leu Pro Ile Phe Ile Lys Asn Ile Tyr Ile Phe Glu Lys Leu Lys

195 200 205

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

450 455 460
 Arg Lys Gly Asp Phe Gly Ser Asn Asp Leu Ile Glu Asn Ile Lys Thr
 465 470 475 480
 Lys Leu Ser Ala Ala Glu Lys Leu Leu Gly Thr Lys Tyr Gln Glu Thr
 485 490 495
 Ala Lys Asp Ile Phe Lys Lys Asp Glu Asn Ser Lys Leu Ile Lys Glu

 500 505 510
 Leu Leu Asp Ala Thr Lys Gln Phe Gln His Phe Ile Lys Pro Leu Leu
 515 520 525
 Gly Thr Gly Glu Glu Ala Asp Arg Asp Leu Val Phe Tyr Gly Asp Phe
 530 535 540
 Leu Pro Leu Tyr Glu Lys Phe Glu Glu Leu Thr Leu Leu Tyr Asn Lys
 545 550 555 560
 Val Arg Asn Arg Leu Thr Gln Lys Pro Tyr Ser Lys Asp Lys Ile Arg

 565 570 575
 Leu Cys Phe Asn Lys Pro Lys Leu Met Thr Gly Trp Val Asp Ser Lys
 580 585 590
 Thr Glu Lys Ser Asp Asn Gly Thr Gln Tyr Gly Gly Tyr Leu Phe Arg
 595 600 605
 Lys Lys Asn Glu Ile Gly Glu Tyr Asp Tyr Phe Leu Gly Ile Ser Ser
 610 615 620
 Lys Ala Gln Leu Phe Arg Lys Asn Glu Ala Val Ile Gly Asp Tyr Glu

 625 630 635 640
 Arg Leu Asp Tyr Tyr Gln Pro Lys Ala Asn Thr Ile Tyr Gly Ser Ala
 645 650 655
 Tyr Glu Gly Glu Asn Ser Tyr Lys Glu Asp Lys Lys Arg Leu Asn Lys
 660 665 670
 Val Ile Ile Ala Tyr Ile Glu Gln Ile Lys Gln Thr Asn Ile Lys Lys
 675 680 685
 Ser Ile Ile Glu Ser Ile Ser Lys Tyr Pro Asn Ile Ser Asp Asp Asp

 690 695 700

Lys Val Thr Pro Ser Ser Leu Leu Glu Lys Ile Lys Lys Val Ser Ile
 705 710 715 720
 Asp Ser Tyr Asn Gly Ile Leu Ser Phe Lys Ser Phe Gln Ser Val Asn
 725 730 735
 Lys Glu Val Ile Asp Asn Leu Leu Lys Thr Ile Ser Pro Leu Lys Asn
 740 745 750
 Lys Ala Glu Phe Leu Asp Leu Ile Asn Lys Asp Tyr Gln Ile Phe Thr
 755 760 765
 Glu Val Gln Ala Val Ile Asp Glu Ile Cys Lys Gln Lys Thr Phe Ile
 770 775 780
 Tyr Phe Pro Ile Ser Asn Val Glu Leu Glu Lys Glu Met Gly Asp Lys
 785 790 795 800
 Asp Lys Pro Leu Cys Leu Phe Gln Ile Ser Asn Lys Asp Leu Ser Phe
 805 810 815
 Ala Lys Thr Phe Ser Ala Asn Leu Arg Lys Lys Arg Gly Ala Glu Asn
 820 825 830
 Leu His Thr Met Leu Phe Lys Ala Leu Met Glu Gly Asn Gln Asp Asn
 835 840 845
 Leu Asp Leu Gly Ser Gly Ala Ile Phe Tyr Arg Ala Lys Ser Leu Asp
 850 855 860
 Gly Asn Lys Pro Thr His Pro Ala Asn Glu Ala Ile Lys Cys Arg Asn
 865 870 875 880
 Val Ala Asn Lys Asp Lys Val Ser Leu Phe Thr Tyr Asp Ile Tyr Lys
 885 890 895
 Asn Arg Arg Tyr Met Glu Asn Lys Phe Leu Phe His Leu Ser Ile Val
 900 905 910
 Gln Asn Tyr Lys Ala Ala Asn Asp Ser Ala Gln Leu Asn Ser Ser Ala
 915 920 925
 Thr Glu Tyr Ile Arg Lys Ala Asp Asp Leu His Ile Ile Gly Ile Asp
 930 935 940
 Arg Gly Glu Arg Asn Leu Leu Tyr Tyr Ser Val Ile Asp Met Lys Gly

945 950 955 960
 Asn Ile Val Glu Gln Asp Ser Leu Asn Ile Ile Arg Asn Asn Asp Leu
 965 970 975
 Glu Thr Asp Tyr His Asp Leu Leu Asp Lys Arg Glu Lys Glu Arg Lys
 980 985 990
 Ala Asn Arg Gln Asn Trp Glu Ala Val Glu Gly Ile Lys Asp Leu Lys
 995 1000 1005
 Lys Gly Tyr Leu Ser Gln Ala Val His Gln Ile Ala Gln Leu Met

 1010 1015 1020
 Leu Lys Tyr Asn Ala Ile Ile Ala Leu Glu Asp Leu Gly Gln Met
 1025 1030 1035
 Phe Val Thr Arg Gly Gln Lys Ile Glu Lys Ala Val Tyr Gln Gln
 1040 1045 1050
 Phe Glu Lys Ser Leu Val Asp Lys Leu Ser Tyr Leu Val Asp Lys
 1055 1060 1065
 Lys Arg Pro Tyr Asn Glu Leu Gly Gly Ile Leu Lys Ala Tyr Gln

 1070 1075 1080
 Leu Ala Ser Ser Ile Thr Lys Asn Asn Ser Asp Lys Gln Asn Gly
 1085 1090 1095
 Phe Leu Phe Tyr Val Pro Ala Trp Asn Thr Ser Lys Ile Asp Pro
 1100 1105 1110
 Val Thr Gly Phe Thr Asp Leu Leu Arg Pro Lys Ala Met Thr Ile
 1115 1120 1125
 Lys Glu Ala Gln Asp Phe Phe Gly Ala Phe Asp Asn Ile Ser Tyr

 1130 1135 1140
 Asn Asp Lys Gly Tyr Phe Glu Phe Glu Thr Asn Tyr Asp Lys Phe
 1145 1150 1155
 Lys Ile Arg Met Lys Ser Ala Gln Thr Arg Trp Thr Ile Cys Thr
 1160 1165 1170
 Phe Gly Asn Arg Ile Lys Arg Lys Lys Asp Lys Asn Tyr Trp Asn
 1175 1180 1185

Tyr Glu Glu Val Glu Leu Thr Glu Glu Phe Lys Lys Leu Phe Lys

1190 1195 1200

Asp Ser Asn Ile Asp Tyr Glu Asn Cys Asn Leu Lys Glu Glu Ile

1205 1210 1215

Gln Asn Lys Asp Asn Arg Lys Phe Phe Asp Asp Leu Ile Lys Leu

1220 1225 1230

Leu Gln Leu Thr Leu Gln Met Arg Asn Ser Asp Asp Lys Gly Asn

1235 1240 1245

Asp Tyr Ile Ile Ser Pro Val Ala Asn Ala Glu Gly Gln Phe Phe

1250 1255 1260

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

1265 1270 1275

Asn Gly Ala Tyr Asn Ile Ala Arg Lys Gly Leu Trp Asn Ile Arg

1280 1285 1290

Gln Ile Lys Gln Thr Lys Asn Lys Asp Asp Leu Asn Leu Ser Ile

1295 1300 1305

Ser Ser Thr Glu Trp Leu Asp Phe Val Arg Glu Lys Pro Tyr Leu

1310 1315 1320

Lys

<210> 15

<211> 1484

<212> PRT

<213> Peregrinibacteria bacterium

<220><221> misc_feature

<222> (1073)..(1073)

<223> Xaa can be any naturally occurring amino acid

<400> 15

Met Ser Asn Phe Phe Lys Asn Phe Thr Asn Leu Tyr Glu Leu Ser Lys

1 5 10 15

Thr Leu Arg Phe Glu Leu Lys Pro Val Gly Asp Thr Leu Thr Asn Met

20								25								30							
Lys	Asp	His	Leu	Glu	Tyr	Asp	Glu	Lys	Leu	Gln	Thr	Phe	Leu	Lys	Asp								
35				40				45															
Gln	Asn	Ile	Asp	Asp	Ala	Tyr	Gln	Ala	Leu	Lys	Pro	Gln	Phe	Asp	Glu								
50				55				60															
Ile	His	Glu	Glu	Phe	Ile	Thr	Asp	Ser	Leu	Glu	Ser	Lys	Lys	Ala	Lys								
65				70				75				80											
Glu	Ile	Asp	Phe	Ser	Glu	Tyr	Leu	Asp	Leu	Phe	Gln	Glu	Lys	Lys	Glu								
85				90				95															
Leu	Asn	Asp	Ser	Glu	Lys	Lys	Leu	Arg	Asn	Lys	Ile	Gly	Glu	Thr	Phe								
100				105				110															
Asn	Lys	Ala	Gly	Glu	Lys	Trp	Lys	Lys	Glu	Lys	Tyr	Pro	Gln	Tyr	Glu								
115				120				125															
Trp	Lys	Lys	Gly	Ser	Lys	Ile	Ala	Asn	Gly	Ala	Asp	Ile	Leu	Ser	Cys								
130				135				140															
Gln	Asp	Met	Leu	Gln	Phe	Ile	Lys	Tyr	Lys	Asn	Pro	Glu	Asp	Glu	Lys								
145				150				155				160											
Ile	Lys	Asn	Tyr	Ile	Asp	Asp	Thr	Leu	Lys	Gly	Phe	Phe	Thr	Tyr	Phe								
165				170				175															
Gly	Gly	Phe	Asn	Gln	Asn	Arg	Ala	Asn	Tyr	Tyr	Glu	Thr	Lys	Lys	Glu								
180				185				190															
Ala	Ser	Thr	Ala	Val	Ala	Thr	Arg	Ile	Val	His	Glu	Asn	Leu	Pro	Lys								
195				200				205															
Phe	Cys	Asp	Asn	Val	Ile	Gln	Phe	Lys	His	Ile	Ile	Lys	Arg	Lys	Lys								
210				215				220															
Asp	Gly	Thr	Val	Glu	Lys	Thr	Glu	Arg	Lys	Thr	Glu	Tyr	Leu	Asn	Ala								
225				230				235				240											
Tyr	Gln	Tyr	Leu	Lys	Asn	Asn	Asn	Lys	Ile	Thr	Gln	Ile	Lys	Asp	Ala								
245				250				255															
Glu	Thr	Glu	Lys	Met	Ile	Glu	Ser	Thr	Pro	Ile	Ala	Glu	Lys	Ile	Phe								
260				265				270															

Asp Val Tyr Tyr Phe Ser Ser Cys Leu Ser Gln Lys Gln Ile Glu Glu
275 280 285

Tyr Asn Arg Ile Ile Gly His Tyr Asn Leu Leu Ile Asn Leu Tyr Asn
290 295 300

Gln Ala Lys Arg Ser Glu Gly Lys His Leu Ser Ala Asn Glu Lys Lys
305 310 315 320

Tyr Lys Asp Leu Pro Lys Phe Lys Thr Leu Tyr Lys Gln Ile Gly Cys
325 330 335

Gly Lys Lys Lys Asp Leu Phe Tyr Thr Ile Lys Cys Asp Thr Glu Glu
340 345 350

Glu Ala Asn Lys Ser Arg Asn Glu Gly Lys Glu Ser His Ser Val Glu
355 360 365

Glu Ile Ile Asn Lys Ala Gln Glu Ala Ile Asn Lys Tyr Phe Lys Ser
370 375 380

Asn Asn Asp Cys Glu Asn Ile Asn Thr Val Pro Asp Phe Ile Asn Tyr
385 390 395 400

Ile Leu Thr Lys Glu Asn Tyr Glu Gly Val Tyr Trp Ser Lys Ala Ala
405 410 415

Met Asn Thr Ile Ser Asp Lys Tyr Phe Ala Asn Tyr His Asp Leu Gln
420 425 430

Asp Arg Leu Lys Glu Ala Lys Val Phe Gln Lys Ala Asp Lys Lys Ser
435 440 445

Glu Asp Asp Ile Lys Ile Pro Glu Ala Ile Glu Leu Ser Gly Leu Phe
450 455 460

Gly Val Leu Asp Ser Leu Ala Asp Trp Gln Thr Thr Leu Phe Lys Ser
465 470 475 480

Ser Ile Leu Ser Asn Glu Lys Leu Lys Ile Ile Thr Asp Ser Gln Thr
485 490 495

Pro Ser Glu Ala Leu Leu Lys Met Ile Phe Asn Asp Ile Glu Lys Asn
500 505 510

Met Glu Ser Phe Leu Lys Glu Thr Asn Asp Ile Ile Thr Leu Lys Lys

515 520 525
 Tyr Lys Gly Asn Lys Glu Gly Thr Glu Lys Ile Lys Gln Trp Phe Asp
 530 535 540

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

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

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

 Glu Glu Cys Lys Ile Ser Lys Gln Asp Phe Glu Leu Asn Asn Lys Val
 740 745 750
 Phe Asn Lys Asn Glu Leu Ser Val Thr Ala Met Arg Tyr Asp Leu Ser
 755 760 765

Ser Thr Gln Glu Lys Gln Tyr Ile Lys Ala Phe Gln Lys Glu Tyr Trp
770 775 780
Glu Leu Leu Phe Lys Gln Glu Lys Arg Asp Thr Lys Leu Thr Asn Asn
785 790 795 800

Glu Ile Phe Asn Glu Trp Ile Asn Phe Cys Asn Lys Lys Tyr Ser Glu
805 810 815
Leu Leu Ser Trp Glu Arg Lys Tyr Lys Asp Ala Leu Thr Asn Trp Ile
820 825 830
Asn Phe Cys Lys Tyr Phe Leu Ser Lys Tyr Pro Lys Thr Thr Leu Phe
835 840 845
Asn Tyr Ser Phe Lys Glu Ser Glu Asn Tyr Asn Ser Leu Asp Glu Phe
850 855 860

Tyr Arg Asp Val Asp Ile Cys Ser Tyr Lys Leu Asn Ile Asn Thr Thr
865 870 875 880
Ile Asn Lys Ser Ile Leu Asp Arg Leu Val Glu Glu Gly Lys Leu Tyr
885 890 895
Leu Phe Glu Ile Lys Asn Gln Asp Ser Asn Asp Gly Lys Ser Ile Gly
900 905 910
His Lys Asn Asn Leu His Thr Ile Tyr Trp Asn Ala Ile Phe Glu Asn
915 920 925

Phe Asp Asn Arg Pro Lys Leu Asn Gly Glu Ala Glu Ile Phe Tyr Arg
930 935 940
Lys Ala Ile Ser Lys Asp Lys Leu Gly Ile Val Lys Gly Lys Lys Thr
945 950 955 960
Lys Asn Gly Thr Trp Ile Ile Lys Asn Tyr Arg Phe Ser Lys Glu Lys
965 970 975
Phe Ile Leu His Val Pro Ile Thr Leu Asn Phe Cys Ser Asn Asn Glu
980 985 990

Tyr Val Asn Asp Ile Val Asn Thr Lys Phe Tyr Asn Phe Ser Asn Leu
995 1000 1005
His Phe Leu Gly Ile Asp Arg Gly Glu Lys His Leu Ala Tyr Tyr

1010	1015	1020
Ser Leu Val Asn Lys Asn Gly Glu Ile Val Asp Gln Gly Thr Leu		
1025	1030	1035
Asn Leu Pro Phe Thr Asp Lys Asp Gly Asn Gln Arg Ser Ile Lys		
1040	1045	1050
Lys Glu Lys Tyr Phe Tyr Asn Lys Gln Glu Asp Lys Trp Glu Ala		
1055	1060	1065
Lys Glu Val Asp Xaa Trp Asn Tyr Asn Asp Leu Leu Asp Ala Met		
1070	1075	1080
Ala Ser Asn Arg Asp Met Ala Arg Lys Asn Trp Gln Arg Ile Gly		
1085	1090	1095
Thr Ile Lys Glu Ala Lys Asn Gly Tyr Val Ser Leu Val Ile Arg		
1100	1105	1110
Lys Ile Ala Asp Leu Ala Val Asn Asn Glu Arg Pro Ala Phe Ile		
1115	1120	1125
Val Leu Glu Asp Leu Asn Thr Gly Phe Lys Arg Ser Arg Gln Lys		
1130	1135	1140
Ile Asp Lys Ser Val Tyr Gln Lys Phe Glu Leu Ala Leu Ala Lys		
1145	1150	1155
Lys Leu Asn Phe Leu Val Asp Lys Asn Ala Lys Arg Asp Glu Ile		
1160	1165	1170
Gly Ser Pro Thr Lys Ala Leu Gln Leu Thr Pro Pro Val Asn Asn		
1175	1180	1185
Tyr Gly Asp Ile Glu Asn Lys Lys Gln Ala Gly Ile Met Leu Tyr		
1190	1195	1200
Thr Arg Ala Asn Tyr Thr Ser Gln Thr Asp Pro Ala Thr Gly Trp		
1205	1210	1215
Arg Lys Thr Ile Tyr Leu Lys Ala Gly Pro Glu Glu Thr Thr Tyr		
1220	1225	1230
Lys Lys Asp Gly Lys Ile Lys Asn Lys Ser Val Lys Asp Gln Ile		
1235	1240	1245

Ile Glu Thr Phe Thr Asp Ile Gly Phe Asp Gly Lys Asp Tyr Tyr		
1250	1255	1260
Phe Glu Tyr Asp Lys Gly Glu Phe Val Asp Glu Lys Thr Gly Glu		
1265	1270	1275
Ile Lys Pro Lys Lys Trp Arg Leu Tyr Ser Gly Glu Asn Gly Lys		
1280	1285	1290
Ser Leu Asp Arg Phe Arg Gly Glu Arg Glu Lys Asp Lys Tyr Glu		
1295	1300	1305
Trp Lys Ile Asp Lys Ile Asp Ile Val Lys Ile Leu Asp Asp Leu		
1310	1315	1320
Phe Val Asn Phe Asp Lys Asn Ile Ser Leu Leu Lys Gln Leu Lys		
1325	1330	1335
Glu Gly Val Glu Leu Thr Arg Asn Asn Glu His Gly Thr Gly Glu		
1340	1345	1350
Ser Leu Arg Phe Ala Ile Asn Leu Ile Gln Gln Ile Arg Asn Thr		
1355	1360	1365
Gly Asn Asn Glu Arg Asp Asn Asp Phe Ile Leu Ser Pro Val Arg		
1370	1375	1380
Asp Glu Asn Gly Lys His Phe Asp Ser Arg Glu Tyr Trp Asp Lys		
1385	1390	1395
Glu Thr Lys Gly Glu Lys Ile Ser Met Pro Ser Ser Gly Asp Ala		
1400	1405	1410
Asn Gly Ala Phe Asn Ile Ala Arg Lys Gly Ile Ile Met Asn Ala		
1415	1420	1425
His Ile Leu Ala Asn Ser Asp Ser Lys Asp Leu Ser Leu Phe Val		
1430	1435	1440
Ser Asp Glu Glu Trp Asp Leu His Leu Asn Asn Lys Thr Glu Trp		
1445	1450	1455
Lys Lys Gln Leu Asn Ile Phe Ser Ser Arg Lys Ala Met Ala Lys		
1460	1465	1470
Arg Lys Lys Lys Arg Pro Ala Ala Thr Lys Lys		

1475 1480

<210> 16

<211> 1245

<212> PRT

<213> *Porphyromonas macacae*

<400> 16

Met Lys Thr Gln His Phe Phe Glu Asp Phe Thr Ser Leu Tyr Ser Leu

1 5 10 15

Ser Lys Thr Ile Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Leu Glu

20 25 30

Asn Ile Lys Lys Asn Gly Leu Ile Arg Arg Asp Glu Gln Arg Leu Asp

35 40 45

Asp Tyr Glu Lys Leu Lys Lys Val Ile Asp Glu Tyr His Glu Asp Phe

50 55 60

Ile Ala Asn Ile Leu Ser Ser Phe Ser Phe Ser Glu Glu Ile Leu Gln

65 70 75 80

Ser Tyr Ile Gln Asn Leu Ser Ile Ser Glu Ala Arg Ala Lys Ile Glu

85 90 95

Lys Thr Met Arg Asp Thr Leu Ala Lys Ala Phe Ser Glu Asp Glu Arg

100 105 110

Tyr Lys Ser Ile Phe Lys Lys Glu Leu Val Lys Lys Asp Ile Pro Val

115 120 125

Trp Cys Pro Ala Tyr Lys Ser Leu Cys Lys Lys Phe Asp Asn Phe Thr

130 135 140

Thr Ser Leu Val Pro Phe His Glu Asn Arg Lys Asn Leu Tyr Thr Ser

145 150 155 160

Asn Glu Ile Thr Ala Ser Ile Pro Tyr Arg Ile Val His Val Asn Leu

165 170 175

Pro Lys Phe Ile Gln Asn Ile Glu Ala Leu Cys Glu Leu Gln Lys Lys

180 185 190

Met Gly Ala Asp Leu Tyr Leu Glu Met Met Glu Asn Leu Arg Asn Val

195 200 205

Trp Pro Ser Phe Val Lys Thr Pro Asp Asp Leu Cys Asn Leu Lys Thr
 210 215 220
 Tyr Asn His Leu Met Val Gln Ser Ser Ile Ser Glu Tyr Asn Arg Phe
 225 230 235 240

 Val Gly Gly Tyr Ser Thr Glu Asp Gly Thr Lys His Gln Gly Ile Asn
 245 250 255
 Glu Trp Ile Asn Ile Tyr Arg Gln Arg Asn Lys Glu Met Arg Leu Pro
 260 265 270
 Gly Leu Val Phe Leu His Lys Gln Ile Leu Ala Lys Val Asp Ser Ser
 275 280 285
 Ser Phe Ile Ser Asp Thr Leu Glu Asn Asp Asp Gln Val Phe Cys Val
 290 295 300

 Leu Arg Gln Phe Arg Lys Leu Phe Trp Asn Thr Val Ser Ser Lys Glu
 305 310 315 320
 Asp Asp Ala Ala Ser Leu Lys Asp Leu Phe Cys Gly Leu Ser Gly Tyr
 325 330 335
 Asp Pro Glu Ala Ile Tyr Val Ser Asp Ala His Leu Ala Thr Ile Ser
 340 345 350
 Lys Asn Ile Phe Asp Arg Trp Asn Tyr Ile Ser Asp Ala Ile Arg Arg
 355 360 365

 Lys Thr Glu Val Leu Met Pro Arg Lys Lys Glu Ser Val Glu Arg Tyr
 370 375 380
 Ala Glu Lys Ile Ser Lys Gln Ile Lys Lys Arg Gln Ser Tyr Ser Leu
 385 390 395 400
 Ala Glu Leu Asp Asp Leu Leu Ala His Tyr Ser Glu Glu Ser Leu Pro
 405 410 415
 Ala Gly Phe Ser Leu Leu Ser Tyr Phe Thr Ser Leu Gly Gly Gln Lys
 420 425 430

 Tyr Leu Val Ser Asp Gly Glu Val Ile Leu Tyr Glu Glu Gly Ser Asn
 435 440 445
 Ile Trp Asp Glu Val Leu Ile Ala Phe Arg Asp Leu Gln Val Ile Leu

450 455 460
 Asp Lys Asp Phe Thr Glu Lys Lys Leu Gly Lys Asp Glu Glu Ala Val
 465 470 475 480
 Ser Val Ile Lys Lys Ala Leu Asp Ser Ala Leu Arg Leu Arg Lys Phe
 485 490 495

 Phe Asp Leu Leu Ser Gly Thr Gly Ala Glu Ile Arg Arg Asp Ser Ser
 500 505 510
 Phe Tyr Ala Leu Tyr Thr Asp Arg Met Asp Lys Leu Lys Gly Leu Leu
 515 520 525
 Lys Met Tyr Asp Lys Val Arg Asn Tyr Leu Thr Lys Lys Pro Tyr Ser
 530 535 540
 Ile Glu Lys Phe Lys Leu His Phe Asp Asn Pro Ser Leu Leu Ser Gly
 545 550 555 560

 Trp Asp Lys Asn Lys Glu Leu Asn Asn Leu Ser Val Ile Phe Arg Gln
 565 570 575
 Asn Gly Tyr Tyr Tyr Leu Gly Ile Met Thr Pro Lys Gly Lys Asn Leu
 580 585 590
 Phe Lys Thr Leu Pro Lys Leu Gly Ala Glu Glu Met Phe Tyr Glu Lys
 595 600 605
 Met Glu Tyr Lys Gln Ile Ala Glu Pro Met Leu Met Leu Pro Lys Val
 610 615 620

 Phe Phe Pro Lys Lys Thr Lys Pro Ala Phe Ala Pro Asp Gln Ser Val
 625 630 635 640
 Val Asp Ile Tyr Asn Lys Lys Thr Phe Lys Thr Gly Gln Lys Gly Phe
 645 650 655
 Asn Lys Lys Asp Leu Tyr Arg Leu Ile Asp Phe Tyr Lys Glu Ala Leu
 660 665 670
 Thr Val His Glu Trp Lys Leu Phe Asn Phe Ser Phe Ser Pro Thr Glu
 675 680 685

 Gln Tyr Arg Asn Ile Gly Glu Phe Phe Asp Glu Val Arg Glu Gln Ala
 690 695 700

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

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

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

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

 Ala Ile Val Val Leu Glu Asn Leu Asn Leu Ser Phe Met Lys Gly Arg

945	950	955	960
Lys Lys Val Glu Lys Ser Val Tyr Glu Lys Phe Glu Arg Met Leu Val			
	965	970	975
Asp Lys Leu Asn Tyr Leu Val Val Asp Lys Lys Asn Leu Ser Asn Glu			
	980	985	990
Pro Gly Gly Leu Tyr Ala Ala Tyr Gln Leu Thr Asn Pro Leu Phe Ser			
	995	1000	1005
Phe Glu Glu Leu His Arg Tyr Pro Gln Ser Gly Ile Leu Phe Phe			
	1010	1015	1020
Val Asp Pro Trp Asn Thr Ser Leu Thr Asp Pro Ser Thr Gly Phe			
	1025	1030	1035
Val Asn Leu Leu Gly Arg Ile Asn Tyr Thr Asn Val Gly Asp Ala			
	1040	1045	1050
Arg Lys Phe Phe Asp Arg Phe Asn Ala Ile Arg Tyr Asp Gly Lys			
	1055	1060	1065
Gly Asn Ile Leu Phe Asp Leu Asp Leu Ser Arg Phe Asp Val Arg			
	1070	1075	1080
Val Glu Thr Gln Arg Lys Leu Trp Thr Leu Thr Thr Phe Gly Ser			
	1085	1090	1095
Arg Ile Ala Lys Ser Lys Lys Ser Gly Lys Trp Met Val Glu Arg			
	1100	1105	1110
Ile Glu Asn Leu Ser Leu Cys Phe Leu Glu Leu Phe Glu Gln Phe			
	1115	1120	1125
Asn Ile Gly Tyr Arg Val Glu Lys Asp Leu Lys Lys Ala Ile Leu			
	1130	1135	1140
Ser Gln Asp Arg Lys Glu Phe Tyr Val Arg Leu Ile Tyr Leu Phe			
	1145	1150	1155
Asn Leu Met Met Gln Ile Arg Asn Ser Asp Gly Glu Glu Asp Tyr			
	1160	1165	1170
Ile Leu Ser Pro Ala Leu Asn Glu Lys Asn Leu Gln Phe Asp Ser			
	1175	1180	1185

Arg Leu Ile Glu Ala Lys Asp Leu Pro Val Asp Ala Asp Ala Asn
1190 1195 1200

Gly Ala Tyr Asn Val Ala Arg Lys Gly Leu Met Val Val Gln Arg
1205 1210 1215

Ile Lys Arg Gly Asp His Glu Ser Ile His Arg Ile Gly Arg Ala
1220 1225 1230

Gln Trp Leu Arg Tyr Val Gln Glu Gly Ile Val Glu
1235 1240 1245

<210> 17

<

211> 1250

<212> PRT

<213> Smithella sp.

<400> 17

Met Gln Thr Leu Phe Glu Asn Phe Thr Asn Gln Tyr Pro Val Ser Lys
1 5 10 15

Thr Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Lys Asp Phe Ile
20 25 30

Glu Gln Lys Gly Leu Leu Lys Lys Asp Glu Asp Arg Ala Glu Lys Tyr
35 40 45

Lys Lys Val Lys Asn Ile Ile Asp Glu Tyr His Lys Asp Phe Ile Glu
50 55 60

Lys Ser Leu Asn Gly Leu Lys Leu Asp Gly Leu Glu Lys Tyr Lys Thr
65 70 75 80

Leu Tyr Leu Lys Gln Glu Lys Asp Asp Lys Asp Lys Lys Ala Phe Asp
85 90 95

Lys Glu Lys Glu Asn Leu Arg Lys Gln Ile Ala Asn Ala Phe Arg Asn
100 105 110

Asn Glu Lys Phe Lys Thr Leu Phe Ala Lys Glu Leu Ile Lys Asn Asp
115 120 125

Leu Met Ser Phe Ala Cys Glu Glu Asp Lys Lys Asn Val Lys Glu Phe
130 135 140

Glu Ala Phe Thr Thr Tyr Phe Thr Gly Phe His Gln Asn Arg Ala Asn

145 150 155 160
 Met Tyr Val Ala Asp Glu Lys Arg Thr Ala Ile Ala Ser Arg Leu Ile
 165 170 175
 His Glu Asn Leu Pro Lys Phe Ile Asp Asn Ile Lys Ile Phe Glu Lys

 180 185 190
 Met Lys Lys Glu Ala Pro Glu Leu Leu Ser Pro Phe Asn Gln Thr Leu
 195 200 205
 Lys Asp Met Lys Asp Val Ile Lys Gly Thr Thr Leu Glu Glu Ile Phe
 210 215 220
 Ser Leu Asp Tyr Phe Asn Lys Thr Leu Thr Gln Ser Gly Ile Asp Ile
 225 230 235 240
 Tyr Asn Ser Val Ile Gly Gly Arg Thr Pro Glu Glu Gly Lys Thr Lys

 245 250 255
 Ile Lys Gly Leu Asn Glu Tyr Ile Asn Thr Asp Phe Asn Gln Lys Gln
 260 265 270
 Thr Asp Lys Lys Lys Arg Gln Pro Lys Phe Lys Gln Leu Tyr Lys Gln
 275 280 285
 Ile Leu Ser Asp Arg Gln Ser Leu Ser Phe Ile Ala Glu Ala Phe Lys
 290 295 300
 Asn Asp Thr Glu Ile Leu Glu Ala Ile Glu Lys Phe Tyr Val Asn Glu

 305 310 315 320
 Leu Leu His Phe Ser Asn Glu Gly Lys Ser Thr Asn Val Leu Asp Ala
 325 330 335
 Ile Lys Asn Ala Val Ser Asn Leu Glu Ser Phe Asn Leu Thr Lys Met
 340 345 350
 Tyr Phe Arg Ser Gly Ala Ser Leu Thr Asp Val Ser Arg Lys Val Phe
 355 360 365
 Gly Glu Trp Ser Ile Ile Asn Arg Ala Leu Asp Asn Tyr Tyr Ala Thr

 370 375 380
 Thr Tyr Pro Ile Lys Pro Arg Glu Lys Ser Glu Lys Tyr Glu Glu Arg
 385 390 395 400

Lys Glu Lys Trp Leu Lys Gln Asp Phe Asn Val Ser Leu Ile Gln Thr
 405 410 415
 Ala Ile Asp Glu Tyr Asp Asn Glu Thr Val Lys Gly Lys Asn Ser Gly
 420 425 430
 Lys Val Ile Ala Asp Tyr Phe Ala Lys Phe Cys Asp Asp Lys Glu Thr
 435 440 445
 Asp Leu Ile Gln Lys Val Asn Glu Gly Tyr Ile Ala Val Lys Asp Leu
 450 455 460
 Leu Asn Thr Pro Cys Pro Glu Asn Glu Lys Leu Gly Ser Asn Lys Asp
 465 470 475 480
 Gln Val Lys Gln Ile Lys Ala Phe Met Asp Ser Ile Met Asp Ile Met
 485 490 495
 His Phe Val Arg Pro Leu Ser Leu Lys Asp Thr Asp Lys Glu Lys Asp
 500 505 510
 Glu Thr Phe Tyr Ser Leu Phe Thr Pro Leu Tyr Asp His Leu Thr Gln
 515 520 525
 Thr Ile Ala Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Gln Lys Pro
 530 535 540
 Tyr Ser Thr Glu Lys Ile Lys Leu Asn Phe Glu Asn Ser Thr Leu Leu
 545 550 555 560
 Gly Gly Trp Asp Leu Asn Lys Glu Thr Asp Asn Thr Ala Ile Ile Leu
 565 570 575
 Arg Lys Asp Asn Leu Tyr Tyr Leu Gly Ile Met Asp Lys Arg His Asn
 580 585 590
 Arg Ile Phe Arg Asn Val Pro Lys Ala Asp Lys Lys Asp Phe Cys Tyr
 595 600 605
 Glu Lys Met Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro
 610 615 620
 Lys Val Phe Phe Ser Gln Ser Arg Ile Gln Glu Phe Thr Pro Ser Ala
 625 630 635 640
 Lys Leu Leu Glu Asn Tyr Ala Asn Glu Thr His Lys Lys Gly Asp Asn

645 650 655
 Phe Asn Leu Asn His Cys His Lys Leu Ile Asp Phe Phe Lys Asp Ser
 660 665 670
 Ile Asn Lys His Glu Asp Trp Lys Asn Phe Asp Phe Arg Phe Ser Ala
 675 680 685
 Thr Ser Thr Tyr Ala Asp Leu Ser Gly Phe Tyr His Glu Val Glu His

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

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

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

 885 890 895

Val Asp Tyr His Asn Leu Leu Asp Lys Lys Glu Gly Asp Arg Ala Thr
900 905 910

Ala Arg Gln Glu Trp Gly Val Ile Glu Thr Ile Lys Glu Leu Lys Glu
915 920 925

Gly Tyr Leu Ser Gln Val Ile His Lys Leu Thr Asp Leu Met Ile Glu
930 935 940

Asn Asn Ala Ile Ile Val Met Glu Asp Leu Asn Phe Gly Phe Lys Arg

945 950 955 960

Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu Lys Met
965 970 975

Leu Ile Asp Lys Leu Asn Tyr Leu Val Asp Lys Asn Lys Lys Ala Asn
980 985 990

Glu Leu Gly Gly Leu Leu Asn Ala Phe Gln Leu Ala Asn Lys Phe Glu
995 1000 1005

Ser Phe Gln Lys Met Gly Lys Gln Asn Gly Phe Ile Phe Tyr Val

1010 1015 1020

Pro Ala Trp Asn Thr Ser Lys Thr Asp Pro Ala Thr Gly Phe Ile
1025 1030 1035

Asp Phe Leu Lys Pro Arg Tyr Glu Asn Leu Asn Gln Ala Lys Asp
1040 1045 1050

Phe Phe Glu Lys Phe Asp Ser Ile Arg Leu Asn Ser Lys Ala Asp
1055 1060 1065

Tyr Phe Glu Phe Ala Phe Asp Phe Lys Asn Phe Thr Glu Lys Ala

1070 1075 1080

Asp Gly Gly Arg Thr Lys Trp Thr Val Cys Thr Thr Asn Glu Asp
1085 1090 1095

Arg Tyr Gln Trp Asn Arg Ala Leu Asn Asn Asn Arg Gly Ser Gln
1100 1105 1110

Glu Lys Tyr Asp Ile Thr Ala Glu Leu Lys Ser Leu Phe Asp Gly
1115 1120 1125

Lys Val Asp Tyr Lys Ser Gly Lys Asp Leu Lys Gln Gln Ile Ala

1130 1135 1140
 Ser Gln Glu Ser Ala Asp Phe Phe Lys Ala Leu Met Lys Asn Leu
 1145 1150 1155
 Ser Ile Thr Leu Ser Leu Arg His Asn Asn Gly Glu Lys Gly Asp
 1160 1165 1170
 Asn Glu Gln Asp Tyr Ile Leu Ser Pro Val Ala Asp Ser Lys Gly
 1175 1180 1185
 Arg Phe Phe Asp Ser Arg Lys Ala Asp Asp Asp Met Pro Lys Asn

1190 1195 1200
 Ala Asp Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Leu Trp
 1205 1210 1215
 Cys Leu Glu Gln Ile Ser Lys Thr Asp Asp Leu Lys Lys Val Lys
 1220 1225 1230
 Leu Ala Ile Ser Asn Lys Glu Trp Leu Glu Phe Val Gln Thr Leu
 1235 1240 1245

Lys Gly

1250

<210> 18

<211> 166

<212> PRT

<213> Escherichia coli

<400> 18

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr

1 5 10 15

Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala Val

20 25 30

Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro Ile

35 40 45

Gly Arg His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln

50 55 60

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr

65 70 75 80

Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser
85 90 95
Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly Ala
100 105 110
Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His Arg
115 120 125

Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu
130 135 140
Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys Lys
145 150 155 160
Ala Gln Ser Ser Thr Asp
165

<210> 19

<211> 166

<212> PRT

<213> Artificial

<220><223> adenosine deaminase

<400> 19

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr
1 5 10 15

Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val
20 25 30

Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile
35 40 45

Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln
50 55 60

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr
65 70 75 80

Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser
85 90 95
Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala
100 105 110

Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg
 115 120 125
 Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu
 130 135 140

Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys
 145 150 155 160
 Ala Gln Ser Ser Thr Asp
 165

<210> 20

<211> 166

<212> PRT

<213> Artificial

<220><223> adenosine deaminase

<400> 20

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr
 1 5 10 15
 Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala Val
 20 25 30

Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ser Ile
 35 40 45
 Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln
 50 55 60
 Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr
 65 70 75 80
 Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser
 85 90 95

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

Cys Tyr Phe Phe Arg Met Arg Arg Gln Val Phe Asn Ala Gln Lys Lys
145 150 155 160

Ala Gln Ser Ser Thr Asp
165

<210> 21

<211> 166

<212> PRT

<213> Artificial

<220><223> adenosine deaminase

<400> 21

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr

1 5 10 15

Leu Ala Lys Arg Ala Leu Asp Glu Arg Glu Val Pro Val Gly Ala Val

20 25 30

Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile

35 40 45

Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln

50 55 60

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr

65 70 75 80

Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser

85 90 95

Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala

100 105 110

Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg

115 120 125

Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Asn Ala Leu Leu

130 135 140

Cys Tyr Phe Phe Arg Met Arg Arg Gln Val Phe Asn Ala Gln Lys Lys

145 150 155 160

Ala Gln Ser Ser Thr Asp

165

<210> 22

<211> 166

<212> PRT

<213> Artificial

<220

><223> adenosine deaminase

<400> 22

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr

1 5 10 15

Leu Ala Lys Arg Ala Leu Asp Glu Arg Glu Val Pro Val Gly Ala Val

20 25 30

Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile

35 40 45

Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln

50 55 60

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr

65 70 75 80

Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser

85 90 95

Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala

100 105 110

Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg

115 120 125

Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Asn Ala Leu Leu

130 135 140

Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys

145 150 155 160

Ala Gln Ser Ser Thr Asp

165

<210> 23

<211> 228

<212> PRT

<213> Rattus norvegicus

<400> 23

Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg Arg
1 5 10 15

Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu Arg
20 25 30

Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His Ser
35 40 45

Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val Asn
50 55 60

Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr Arg
65 70 75 80

Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys Ser
85 90 95

Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu Phe
100 105 110

Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg Gln
115 120 125

Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met Thr
130 135 140

Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser Pro
145 150 155 160

Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg Leu
165 170 175

Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys Leu
180 185 190

Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile Ala
195 200 205

Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp Ala
210 215 220

Thr Gly Leu Lys

225

<210> 24

<211> 199

<212> PRT

<213> Homo sapiens

<400> 24

Met Glu Ala Ser Pro Ala Ser Gly Pro Arg His Leu Met Asp Pro His

1 5 10 15

Ile Phe Thr Ser Asn Phe Asn Asn Gly Ile Gly Arg His Lys Thr Tyr

20 25 30

Leu Cys Tyr Glu Val Glu Arg Leu Asp Asn Gly Thr Ser Val Lys Met

35 40 45

Asp Gln His Arg Gly Phe Leu His Asn Gln Ala Lys Asn Leu Leu Cys

50 55 60

Gly Phe Tyr Gly Arg His Ala Glu Leu Arg Phe Leu Asp Leu Val Pro

65 70 75 80

Ser Leu Gln Leu Asp Pro Ala Gln Ile Tyr Arg Val Thr Trp Phe Ile

85 90 95

Ser Trp Ser Pro Cys Phe Ser Trp Gly Cys Ala Gly Glu Val Arg Ala

100 105 110

Phe Leu Gln Glu Asn Thr His Val Arg Leu Arg Ile Phe Ala Ala Arg

115 120 125

Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys Glu Ala Leu Gln Met Leu Arg

130 135 140

Asp Ala Gly Ala Gln Val Ser Ile Met Thr Tyr Asp Glu Phe Lys His

145 150 155 160

Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys Pro Phe Gln Pro Trp

165 170 175

Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg Leu Arg Ala

180 185 190

Ile Leu Gln Asn Gln Gly Asn

195

<210> 25

<211> 621

<212> DNA

<213> *Petromyzon marinus*

<400> 25

acagatgcag agtatgtgag aattcacgaa aagctggaca tctatacctt caagaagcag 60
ttctttaaca ataagaagtc tgtgagccat aggtgctacg tgctgttcga gctgaagaga 120
aggggtgaaa gaagggcatg tttttggggg tatgctgtga acaagcccca gtctggaact 180

gagagaggca ttcacgccga aattttcagc atcagaaagg tggaggaata cctgagggat 240
aaccttgac agtttacaat taattggtat tctagctggt ctccatgcgc tgactgtgcc 300
gagaagatcc tggaatggta caaccaggag ctgagaggaa atggccatac cctgaagatt 360
tgggcctgca agctgtacta tgaaaagaac gcaagaaatc agatcggact gtggaacctg 420
agggataatg gtgtggggct gaacgtgatg gtgtccgagc actatcagtg ctgtagaaag 480
attttcattc agtcctcaca taatcagctg aacgagaata gatggctgga aaagactctg 540
aagagggctg agaagagaag gtccgaactg tcaattatga tccaggtgaa gatcctgcac 600

accactaagt cacctgccgt g 621

<210> 26

<211> 83

<212> PRT

<213> *Bacillus phage*

<400> 26

Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu Val

1 5 10 15

Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile

20 25 30

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

35 40 45

Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr

50 55 60

Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile

65 70 75 80

Lys Met Leu

<210> 27
 <211> 19
 <212> RNA
 <213> Artificial
 <220><223> polynucleotide
 <220><221> misc_feature
 <222> (1)..(19)
 <223> wherein n is A, C, T or G
 <400> 27
 nnnnnnnnnn nnnnnnnnnn 19

<210> 28
 <211> 22
 <212> DNA
 <213> Artificial
 <220><223> polynucleotide
 <220><221> misc_feature
 <222> (4)..(22)
 <223> wherein n is A, C, T or G
 <400> 28
 aaannnnnnn nnnnnnnnnn nn 22

<210> 29
 <211> 22
 <212> DNA
 <213> Artificial
 <220><223> polynucleotide
 <220><221> misc_feature
 <222> (4)..(22)
 <223> wherein n is A, C, T or G
 <400> 29
 tttnnnnnnn nnnnnnnnnn nn 22

<

210> 30
 <211> 23
 <212> DNA

<213> Artificial
 <220><223> polynucleotide
 <400> 30
 ggaatccctt ctgcagcacc tgg 23
 <210> 31
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> polynucleotide
 <400> 31
 ctgatgggtcc atgtctgtta ctc 23
 <210> 32
 <211> 49
 <212> DNA
 <213> Artificial
 <220><223> polynucleotide
 <220><221> misc_feature
 <222> (7)..(11)
 <223> wherein n is A, T, C or G

 <400> 32
 cgatgtnnnn nggaatccct tctgcagcac ctgggcgcag gtcacgagg 49
 <210> 33
 <211> 57
 <212> DNA
 <213> Artificial
 <220><223> polynucleotide
 <220><221> misc_feature
 <222> (43)..(46)
 <223> n is a, c, g, or t
 <220><221> misc_feature
 <222> (47)..(51)
 <223> wherein n is A, T, C or G
 <400> 33
 aattcctcgt gacctgcgcc caggtgctgc agaagggatt ccnnnnnaca tcgcatg 57

<210> 34
 <211> 49
 <212> DNA
 <213> Artificial
 <220><223> polynucleotide
 <220><
 221> misc_feature
 <222> (7)..(11)
 <223> wherein n is A, T, C or G
 <400> 34
 cgatgtnnnn nctgatggtc catgtctgtt actcgcgag gtcacgagg 49
 <210> 35
 <211> 57
 <212> DNA
 <213> Artificial
 <220><223> polynucleotide
 <220><221> misc_feature
 <222> (43)..(46)
 <223> n is a, c, g, or t
 <220><221> misc_feature
 <222> (47)..(51)
 <223> wherein n is A, T, C or G
 <400> 35
 aattcctcgt gacctgcgag agtaacagac atggaccatc agnnnnnaca tcgcatg 57
 <210> 36
 <211>
 28
 <212> DNA
 <213> Artificial
 <220><223> polynucleotide
 <220><221> misc_feature
 <222> (1)..(5)
 <223> wherein n is A, C, T or G
 <400> 36
 nnnnnngaatt cccttctgca gcacctgg 28

<210> 37
 <211> 28
 <212> DNA
 <213> Artificial
 <220><223> polynucleotide
 <220><221> misc_feature
 <222> (1)..(5)
 <223> wherein n is A, C, T or G
 <400> 37
 nnnnncctta gggaagacgt cgtggacc 28
 <210> 38
 <211> 23
 <212> RNA
 <213> Artificial
 <220><223> polynucleotide
 <400> 38
 ggaaucuccuu cugcagcacc ugg 23
 <210> 39
 <211> 24
 <212> DNA
 <213> Artificial
 <220><223> polynucleotide
 <220><221> misc_feature
 <222> (1)..(5)
 <223> wherein n is A, C, T or G
 <400> 39
 nnnnnngaatt cccttctgca gcac 24
 <210> 40
 <211> 1592
 <212> DNA
 <213> Medicago truncatula
 <400> 40
 actgttaata atttttaaac gtcagcgac taaaaaacg aaaagacgga cacgtgaaaa 60

taaaaaacac acactagttt atgacgcaat actattttac ttatgatttg ggtacattag 120
 acaaaaccgt gaaagagatg tatcagctat gaaacctgta tacttcaata cagagactta 180
 ctcatatcgg atacgtacgc acgaagtatc atattaatta ttttaatttt taataaatat 240
 tttatcggat acttatgtga tactctacat atacacaagg atatttctaa gatactttat 300
 agatacgtat cctagaaaaa catgaagagt aaaaaagtga gacaatgttg taaaaattca 360
 ttataaatgt atatgattca attttagata tgcacagta taattgattc tcgatgaaac 420
 acttaaaatt atatttcttg tggaagaacg tagcgagaga ggtgattcag ttagacaaca 480

ttaaataaaa ttaatgttaa gttcttttaa tgatgtttct ctcaatatca catcatatga 540
 aaatgtaata tgatttataa gaaaattttt aaaaaattta ttttaataat cacatgtact 600
 attttttaaa aattgtatct tttataataa tacaataata aagagtaatc agtgtaatt 660
 tttcttcaaa tataagtttt attataaatc attgttaacg tatcataagt cattaccgta 720
 tcgtatctta attttttttt aaaaaccgct aattcacgta cccgtattgt attgtaccg 780
 cacctgtatc acaatcgatc ttagttagaa gaattgtctc gaggcggtgc aagacagcat 840
 ataatagacg tggactctct tataccaaac gttgtcgtat cacaagggt taggtaacaa 900

gtcacagttt gtccacgtgt cacgttttaa ttggaagagg tgccgttggc gtaataaac 960
 agccaatcga tttttgctat aaaagcaaact caggtaaact aaacttcttc attcttttct 1020
 tccccatcgc tacaaaaccg gttcttttgg aaaagagatt cattcaaacc tagcacccaa 1080
 ttccgtttca aggtataatc tactttctat tcttcgatta ttttattatt attagctact 1140
 atcgtttaat cgatcttttc ttttgatccg tcaaatttaa attcaattag ggttttggtc 1200
 ttttctttca tctgattgaa atccttctga attgaaccgt ttacttgatt ttactgttta 1260
 ttgtatgatt taatcctttg tttttcaaag acagtcttta gattgtgatt aggggttcat 1320

ataaattttt agatttggat ttttgatttg tatgattcaa aaaatacgtc cttaatttag 1380
 attagtacat ggatattttt taccgatttt attgattgtc agggagaatt tgatgagcaa 1440
 gtttttttga tgtctgttgt aaattgaatt gattataatt gctgatctgc tgcttccagt 1500
 ttccataacc catattcttt taaccttgtt gtacacacaa tgaaaaattg gtgattgatt 1560
 catttgtttt tctttgtttt ggattataca gg 1592

<210> 41

<211> 2000

<212> DNA

<213> Zea mays

<400> 41

gtcgtgcccc tccttagaga taaagagcat tgcattgtcta aagtataaaa aattaccaca 60

tatTTTTTtg tcacacttat ttgaagtgtg gtttatctat ctctatacat atatttaaac 120

ttcactctac aaataatata gtctataata ctaaaataat attagtgttt tagaggatca 180

tataaataaa ctgctagaca tggctctaaag gataattgaa tatTTtgaca atctacagtt 240

ttatctTTTT agtgtgcatg tgatctctct gTTTTTTTTg caaatagctt gacctatata 300

atacttcate catTTtatta gtacatccat ttaggattta gggttgatgg tttctataga 360

ctaattTTta gtacatccat tttattcttt ttagtctcta aattTTTtaa aactaaaact 420

ctatttttagt tttttattta ataatttaga tataaaatga aataaaataa attgactaca 480

aataaaacaa atacccttta agaaataaaa aaactaagca aacatttttc ttgtttcgag 540

tagataatga caggctgttc aacgccgtcg acgagtctaa cggacaccaa ccagcgaacc 600

agcagcgtcg cgtcgggcca agcgaagcag acggcacggc atctctgtag ctgcctctgg 660

accctctctg agagtccgc tccaccgttg gacttgctcc gctgtcggca tccagaaatt 720

gcgtggcggg gcggcagacg tgaggcggca cggcaggcgg cctcttctc ctctcacggc 780

accggcagct acgggggatt ctttccccc cgtctcttcg ctttcccttc ctgcccgcgc 840

glaataaata gacacccctt ccacaccctt tttccccaac ctctgttctg ttcggagcgc 900

acacacacgc aaccagatct cccccaaatc cagccgtcgg caccctcgct tcaaggtacg 960

ccgctcatcc tccccccccc cctctctcta ccttctctag atcggcgatc cggttccatgg 1020

ttagggcccc gtagttctac ttctgttcat gtttgtgtta gagcaaacat gttcatgttc 1080

atgtttgtga tgatgtggtc tggttgggcg gtcgttctag atcggagtag gatactgttt 1140

caagctacct ggiggattta ttaattttgt atctgtatgt gtgtgccata catcttcata 1200

gttacgagtt taagatgatg gatggaaata tcgatctagg ataggtatac atgttgatgc 1260

gggttttact gatgcatata cagagatgct tttttctcgc cttggttgtg atgatatggt 1320

ctggttgggc ggtcgttcta gatcggagta gaatactgtt tcaaaactacc tgggtggattt 1380

attaaaggat aaagggtcgt tctagatcgg agtagaatac tgtttcaaac tacctggtgg 1440

atttattaaa ggatctgtat gtatgtgcct acatcttcac agttacgagt ttaagatgat 1500

ggatggaaat atcgatctag gataggtata catgttgatg cgggttttac tgatgcatat 1560

acagagatgc ttttttccg ttggttgtga tgatgtggtc tggttgggcg gtcgttctag 1620

atcggagtag aaactgttt caaactacct ggtggattta ttaattttgt atctttatgt 1680

gltgtccata catcttcata gttacgagtt taagatgatg gatggaaata ttgatctagg 1740

ataggtatac atgttgatgt gggttttact gatgcatata catgatggca tatgcggcat 1800
ctattcatat gctctaacct tgagtaccta tctattataa taaacaagta tgttttataa 1860
ttattttgat cttgatatac ttggatgatg gcatatgcag cagctatatg tggatttttt 1920
agccctgcct tcataccta tttatttgc tggctactgtt tcttttgtcc gatgctcacc 1980
ctgttgtttg gtgatacttc 2000

<210> 42

<211> 160

<212> PRT

<213> Unknown

<220><223> FERNY deaminase

<400> 42

Phe Glu Arg Asn Tyr Asp Pro Arg Glu Leu Arg Lys Glu Thr Tyr Leu

1 5 10 15
Leu Tyr Glu Ile Lys Trp Gly Lys Ser Gly Lys Leu Trp Arg His Trp
20 25 30
Cys Gln Asn Asn Arg Thr Gln His Ala Glu Val Tyr Phe Leu Glu Asn
35 40 45
Ile Phe Asn Ala Arg Arg Phe Asn Pro Ser Thr His Cys Ser Ile Thr
50 55 60
Trp Tyr Leu Ser Trp Ser Pro Cys Ala Glu Cys Ser Gln Lys Ile Val

65 70 75 80
Asp Phe Leu Lys Glu His Pro Asn Val Leu Glu Ile Tyr Val Ala Arg
85 90 95
Leu Tyr Tyr His Glu Asp Glu Arg Asn Arg Gln Gly Leu Arg Asp Leu
100 105 110
Val Asn Ser Gly Val Thr Ile Arg Ile Met Asp Leu Pro Asp Tyr Asn
115 120 125
Tyr Cys Trp Lys Thr Phe Val Ser Asp Gln Gly Gly Asp Glu Asp Tyr

130 135 140
Trp Pro Gly His Phe Ala Pro Trp Ile Lys Gln Tyr Ser Leu Lys Leu
145 150 155 160
<210> 43

<211> 207

<212> PRT

<213> Artificial

<220><223> evolved cytidine deaminase

<400> 43

Thr Asp Ala Glu Tyr Val Arg Ile His Glu Lys Leu Asp Ile Tyr Thr

1 5 10 15

Phe Lys Lys Gln Phe Ser Asn Asn Lys Lys Ser Val Ser His Arg Cys

20 25 30

Tyr Val Leu Phe Glu Leu Lys Arg Arg Gly Glu Arg Arg Ala Cys Phe

35 40 45

Trp Gly Tyr Ala Val Asn Lys Pro Gln Ser Gly Thr Glu Arg Gly Ile

50 55 60

His Ala Glu Ile Phe Ser Ile Arg Lys Val Glu Glu Tyr Leu Arg Asp

65 70 75 80

Asn Pro Gly Gln Phe Thr Ile Asn Trp Tyr Ser Ser Trp Ser Pro Cys

85 90 95

Ala Asp Cys Ala Glu Lys Ile Leu Glu Trp Tyr Asn Gln Glu Leu Arg

100 105 110

Gly Asn Gly His Thr Leu Lys Ile Trp Val Cys Lys Leu Tyr Tyr Glu

115 120 125

Lys Asn Ala Arg Asn Gln Ile Gly Leu Trp Asn Leu Arg Asp Asn Gly

130 135 140

Val Gly Leu Asn Val Met Val Ser Glu His Tyr Gln Cys Cys Arg Lys

145 150 155 160

Ile Phe Ile Gln Ser Ser His Asn Gln Leu Asn Glu Asn Arg Trp Leu

165 170 175

Glu Lys Thr Leu Lys Arg Ala Glu Lys Arg Arg Ser Glu Leu Ser Ile

180 185 190

Met Phe Gln Val Lys Ile Leu His Thr Thr Lys Ser Pro Ala Val

195 200 205

<210> 44

<211> 228

<212> PRT

<213> Artificial

<220><223> evolved cytidine deaminase

<400> 44

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

1 5 10 15
Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu Arg

20 25 30
Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His Ser

35 40 45
Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val Asn

50 55 60
Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr Arg

65 70 75 80
Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys Ser

85 90 95
Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro Asn Val Thr Leu Phe

100 105 110
Ile Tyr Ile Ala Arg Leu Tyr His Leu Ala Asn Pro Arg Asn Arg Gln

115 120 125
Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met Thr

130 135 140
Glu Gln Glu Ser Gly Tyr Cys Trp His Asn Phe Val Asn Tyr Ser Pro

145 150 155 160
Ser Asn Glu Ser His Trp Pro Arg Tyr Pro His Leu Trp Val Arg Leu

165 170 175
Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys Leu

180 185 190
Asn Ile Leu Arg Arg Lys Gln Ser Gln Leu Thr Ser Phe Thr Ile Ala

195 200 205

Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp Ala
210 215 220
Thr Gly Leu Lys
225
<210> 45
<211> 162
<212> PRT
<213> Artificial
<220><223> evolved cytidine deaminase
<400> 45
Ser Phe Glu Arg Asn Tyr Asp Pro Arg Glu Leu Arg Lys Glu Thr Tyr
1 5 10 15
Leu Leu Tyr Glu Ile Lys Trp Gly Lys Ser Gly Lys Leu Trp Arg His
20 25 30
Trp Cys Gln Asn Asn Arg Thr Gln His Ala Glu Val Tyr Phe Leu Glu
35 40 45
Asn Ile Phe Asn Ala Arg Arg Phe Asn Pro Ser Thr His Cys Ser Ile
50 55 60
Thr Trp Tyr Leu Ser Trp Ser Pro Cys Ala Glu Cys Ser Gln Lys Ile
65 70 75 80
Val Asp Phe Leu Lys Glu His Pro Asn Val Asn Leu Glu Ile Tyr Val
85 90 95
Ala Arg Leu Tyr Tyr Pro Glu Asn Glu Arg Asn Arg Gln Gly Leu Arg
100 105 110
Asp Leu Val Asn Ser Gly Val Thr Ile Arg Ile Met Asp Leu Pro Asp
115 120 125
Tyr Asn Tyr Cys Trp Lys Thr Phe Val Ser Asp Gln Gly Gly Asp Glu
130 135 140
Asp Tyr Trp Pro Gly His Phe Ala Pro Trp Ile Lys Gln Tyr Ser Leu
145 150 155 160
Lys Leu

<210> 46

<211> 229

<212> PRT

<213> Rattus norvegicus

<400> 46

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His

35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val

50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr

65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys

85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu

100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser

145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg

165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys

180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp

210 215 220

Ala Thr Gly Leu Lys

225

<210> 47

<211> 198

<212> PRT

<213> Homo sapiens

<400> 47

Met Asp Ser Leu Leu Met Asn Arg Arg Lys Phe Leu Tyr Gln Phe Lys

1 5 10 15

Asn Val Arg Trp Ala Lys Gly Arg Arg Glu Thr Tyr Leu Cys Tyr Val

20 25 30

Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly Tyr

35 40 45

Leu Arg Asn Lys Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr

50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp

65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp

85 90 95

Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg

100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg

115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr

130 135 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe Lys

145 150 155 160

Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu

165 170 175

Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala

180 185 190

Phe Arg Thr Leu Gly Leu

195

<210> 48

<211> 197

<212> PRT

<213> Artificial

<220><223> cytosine deaminase

<400> 48

Met Asp Ser Leu Leu Met Asn Arg Arg Glu Phe Leu Tyr Gln Phe Lys

1 5 10 15

Asn Val Arg Trp Ala Lys Gly Arg Arg Glu Thr Tyr Leu Cys Tyr Val

20 25 30

Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly Tyr

35 40 45

Leu Arg Asn Lys Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr

50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp

65 70 75 80

Phe Ile Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp

85 90 95

Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg

100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg

115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr

130 135 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Gly Arg Thr Phe Lys

145 150 155 160

Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu

165 170 175

Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala

180

185

190

Phe Arg Thr Cys Thr

195

<210> 49

<211> 3987

<212> DNA

<213> Artificial

<220><223> polynucleotide

<400> 49

atggccggga gcaagaagcg ccggataaag caggacacgc agttcgaggg cttcaccaac 60

ctgtaccaag tctccaagac gctccgggtc gagcttatcc cgcaaggga gaccctgaaa 120

cacatccagg aacaaggttt catcgaggag gacaaggccc gcaacgacca ctacaaggag 180

ctcaagccca taatcgatcg gatctacaag acgtacgccg accagtgcct ccaactggtg 240

cagctcgact gggagaacct gagcgccgcc attgacagct accgcaagga aaagacggag 300

gagacgcga acgcccttat tgaggagcaa gccacctacc gcaacgccat ccacgactac 360

ttcatcgggc gcaccgacaa cctgacggac gcgatcaaca agcgccacgc ggaaatctac 420

aagggccttt tcaaggccga gctcttcaac gggaagggtc taaaacagct cgggactgtc 480

acgacaaccg agcatgagaa cgccctcctt cgcagcttcg acaagttcac cacatacttc 540

tcgggcttct accggaaccg caagaacgtt ttcagcgccg aggacatctc caccgccatc 600

ccgcacagga tcgtccagga caacttcccc aagttcaagg agaactgcca catcttcacg 660

cgctgatta cagccgtacc ttactttcgt gagcacttcg agaacgtcaa aaaggccatc 720

gggatcttcg tctccacgtc catcgaggag gtattctctt tcccgttcta taaccagctc 780

ctgaccaga cgcagatcga cctctacaac cagctactgg gcggcatcag ccgggaggcc 840

gggaccgaga aaataaaggc cctcaacgaa gttctcaacc tggccatcca gaagaacgac 900

gagaccgcgc atatcatcgc atccctgccg catcgcttca ttcctttgtt caagcagata 960

ttgagcgacc ggaacaccct ctcgttcac ctcgaagaat tcaagagcga cgaggaggtc 1020

attcagtctt tctgcaagta caagacgtc ctacggaatg agaatgtgct ggagaccgcg 1080

gaggcactct tcaatgagct gaactccatt gacctgacct acatcttcat tagccacaag 1140

aaactggaga cgatctccag gcgcctgtgc gacctggg acactctccg caacgcctc 1200

tacgaacgcc ggatctccga acttaccggc aagataacta agtcggctaa ggagaagggtg 1260

caacggagcc tcaagcacga ggacatcaac cttcaggaaa tcatttcagc gcggggcaag 1320

gagctgagcg aggcgtttta gcagaaaaca tcggagatac tgagccacgc gcacgcggcc	1380
ctggatcaac cgctgccgac gactctcaag aagcaagagg agaaggaaat ccttaagtcc	1440
cagctcgact cgctgctcgg cctctatcac ttgctcgact ggttcgcggt tgatgagtcc	1500
aacgaggtgg acccggagtt ctccgcgcgc ctacagggtta ttaagctgga gatggagcca	1560
agcttaagct tctacaacaa ggcccgaac tacgcgacca aaaaaccgta ctcagtcgag	1620
aaattcaagc tgaatttcca gatgcctaca ttggcgaggg ggtgggacgt gaaccgcgag	1680
aagaacaatg gagccatcct gtctgtcaaa aatgggttgt actacctggg catcatgccc	1740
aagcagaagg gccgttacaa ggccctgtca ttcgagccta ccgagaagac ctcgaggggc	1800
ttcgacaaga tgiactacga ctatttcccg gacgccgcca agatgatccc gaagtgtccc	1860
acgcagctca aagccgtcac ggcccacttc cagacgcata ccacgccgat acttctgagc	1920
aacaacttca ttgagccgct agagatcacg aaggagatat acgacctaaa caaccccgaa	1980
aaggagccca agaagttcca gacagcctac gctaagaaga caggtgatca gaagggatat	2040
agggaggcac tctgcaagtg gatcgacttc acgcgcgact tcctgtcgaa atatacaaag	2100
acgaccagca ttgacctaaq ttctctccgc ccatacctccc agtacaagga tctgggcgag	2160
tattatgcgg agctgaaccc attgtgttac cacatcagct tccagaggat cgccgagaag	2220
gagattatgg acgcggtgga gacggggaaa ctatacctgt tccaaatata taacaaggac	2280
ttcgctaaag ggcaccacgg gaagcccaac ctgcacacac tctactggac gggcttgttt	2340
tcgccagaaa atttggccaa gacttcgac aagctcaacg gccaggcgga gttgttttac	2400
cgteccaagt ctgcatgaa gcgcatggcg catcgctcg gagagaaaat gcttaacaag	2460
aagctcaagg atcagaagac gccatacct gatacgttgt accaggaatt gtacgactac	2520
gtgaaccacc gcctatcgca cgaccttca gacgagggcc gcgccctcct cccaaacgtg	2580
attactaagg aggtttccca tgaaataatc aaggaccgac ggttcaccag cgacaaattt	2640
ttttccacg tgctatcac gctcaattac caggcgcca actccccatc gaagttcaac	2700
cagcgctga acgcctacct taaggagcac ccggagaccc caatcatcgg gatcgaccgt	2760
ggcgagcgga acctgatcta tattacggtg atcgatagca ccgggaagat cctggagcag	2820
cgctccctga acacaatcca gcagtttgac taccagaaga aactcgacaa ccgggagaag	2880
gagcgcgtcg cagcccgga agcatggagt gtggtcggca ccataaagga cctgaaacag	2940
ggttacctaa gtcaagttat ccacgagatc gttgacctga tgatacacta tcaagccgta	3000
gtcgtgctgg agaacctcaa cttcgggttt aagtccaagc gcaccggcat cgcggagaag	3060
gcggtgtacc agcagttcga gaagatgctg atcgacaagc tgaactgcct ggtgctcaag	3120

gactaccctg cggagaaggt cggcgggggtc ttgaaccgt accagctaac cgaccagttc 3180

acgagcttcg ccaaaatggg cacgcagtcc ggattcttgt tttatgtccc ggctccatat 3240

acaagtaaga tcgaccgcgt gacagggttt gttgacccat tcgtgtggaa gaccatcaag 3300

aaccacgaga gcaggaaaca cttcttagag ggcttcgact tcctgcatta cgacgttaag 3360

acaggcgact tcactctgca cttcaagatg aaccgcaacc tgtcgttcca gaggggcctg 3420

cccggcttca tgccgcctg ggatatcgtc ttgagaaga atgagacgca gttcgacgcg 3480

aaggggacgc cgttcacgcg tggaaagcgg atcgtgccgg tcacgagaa ccaccgcttc 3540

acgggtcgct accgagattt ataccccgcc aacgaactaa ttgcgtgct ggaggagaag 3600

gggatcgtgt tccgagatgg cagcaacatt ctcccgaagc tgctggagaa cgacgactcg 3660

cacgctattg acacgatggc cgccctcata cggagcgtgc ttcagatgcg gaacagtaac 3720

gctgccacgg gcgaggacta cattaactcc ccgctccgag acctcaacgg ggtctgcttc 3780

gatagccgct tccagaacct ggagtggcct atggatgcgg acgcgaacgg ggcctaccac 3840

atcgccctca agggccaact cctgctcaac cacttgaagg aaagcaaaga cctcaaattg 3900

cagaatggca tcagtaacca ggactggctc gcgtacatcc aggaactgag aaacgggtcc 3960

aagaagcggc gtatcaagca agattga 3987

<210> 50

<211> 3987

<212> DNA

<213> Artificial

<220><223> polynucleotide

<400> 50

atggcgggaa gcaaaaagcg ccggattaag caagacacgc agttcgaggg cttcacgaac 60

ctctaccaag tcagcaagac cctccggttc gagctgatac cacagggaac gacgtcaag 120

cacatccagg aacagggttt catcgaggag gacaaggcgc gcaacgacca ctacaaggag 180

ttgaaaccga tcacgaccg catctacaag acgtacgccg accagtgcct ccagctcgtg 240

cagctcgact gggagaacct ctccgccgcc attgactcgt accggaagga gaagactgag 300

gagaccgcga acgccctgat cgaggagcaa gcaacctacc ggaacgccat ccacgactac 360

ttcatcggcc gcaccgacaa cctcaccgac gcgatcaaca agcggcacgc ggagatatac 420

aaagggtgtg tcaaggcgga gctgttcaac ggcaaggtgc tcaagcagct agggacggtg 480

accacgaccg agcacgagaa cgcgtcctc cgcagcttcg acaagttcac cacctacttc 540

agcggcttct accggaaccg caagaatgtg ttcagcgcgg aggacatcag cacggccatc 600

ccgcaccgca tcgtccagga caacttcccg aagttcaagg agaactgcc aatcttcacc	660
cgctgataa ccgccgtccc ctccctgcgg gagcacttcg agaacgtcaa aaaggcaatt	720
gggatcttcg tctcgaccag cattgaggag gtgttcagct tccccttcta caaccagctc	780
ctcaccaga cgcatcgga cctgtacaat cagttgctcg gcgggataag ccgcgaggcg	840
ggaaccgaaa aaatcaaggg gctgaacgaa gtgttgaaacc tcgccatcca gaagaacgac	900
gagaccgccc acatcatcg ctccttgcgc caccggttca tcccgtgtt caagcagatc	960
ctctctgacc ggaacacctt gtccttcatt ctgaggagt tcaagtcgga cgaggaggtc	1020
atccagagct tctgcaagta caagacgtg ctacggaacg agaacgtgct ggagacggcg	1080
gaggcactgt tcaacgagct aaacagcatc gacctcacgc acatcttcat cagtcacaag	1140
aaactggaga ccatctctc cgcgctgtgc gacctggg acacgtcag gaacgcgtc	1200
tacgagcgcc gaatcagtga gctgacgggc aagatcacga agtccgcgaa ggagaagtg	1260
cagcggtccc tcaagcacga ggacatcaac ctccaggaga tcatctcagc ggctgggaaa	1320
gagctgtccg aggcgttcaa gcagaaaacg agcgaaatcc tgtcccacgc gcacgcggcc	1380
ctggatcagc ctctgccgac gacctcaag aaacaagaag aaaaggaaat cctcaagtcg	1440
cagctcgact cgctgctggg cctgtacat ctctcgact ggttcgccgt ggacgagagc	1500
aacgaggtgg accccgagtt ctccgcgcgg ctacgggga tcaagctgga gatggagccc	1560
agcctgtcct tctacaaca ggcgcgcaac tacgccacca agaagcccta cagcgtggag	1620
aagttcaagc tcaacttcca gatgccact ctgcacgtg ggtgggacgt caaccgcgaa	1680
aaaaataatg ggccgatcct gttcgtcaag aacggcctgt actacttggg catcatgccg	1740
aaacagaagg gccgctacaa ggccctgagc ttgaaccga ccgagaaaac gagcgagggg	1800
ttcgacaaga tgtactagca ctacttccc gagccgcga agatgattcc aaagtgtcc	1860
acgcagctta agccgtgac ggcccacttc cagacgcaca cgaccccgat cctcctcagc	1920
aacaacttca tcgagccctt ggagatcacg aaggagatat acgacctgaa caaccggag	1980
aaggagccca agaaattcca gaccgcctac gccagaaga caggcgacca aaagggttac	2040
agggaggccc tctgcaagtg gatcgacttc actagggact tcctgtccaa gtacaccaag	2100
actacctta tcgacctgtc cagcctccgc ccgtcgtccc agtacaaga ttgggagag	2160
tattacgagg agctgaacc actgtctac cacatcagct tccagcgcat cgcggagaag	2220
gagatcatgg acgcagtga gacgggcaag ctatacctat ttcatatata caacaagac	2280
ttcgctaagg gacaccacgg caagcctaac ctgcacacc tctactggac ggggtcttc	2340
agcccgagga acctgcca gacctgatc aagctcaac gccaggccga gctgttctac	2400

cggcccaagt cccgcatgaa gcggatggcc caccggctcg gggagaaaat gctcaacaag 2460

aaattgaagg accaaaaaac gccgataccc gacaccctat accaggagct gtacgactat 2520

gtgaaccacc gcctgagcca cgacctcagc gacgaggcgc gggccctcct gccgaacgtc 2580

atcacaaagg aggtcagcca cgagatcatc aaggaccggc gcttcacctc cgacaagttt 2640

ttcttttcac tgcccatcac gctcaactac caggccgcca actcgccgtc caagttcaac 2700

cagcgctga acgcctacct caaggagcac cccgagaccc cgatcatcgg gattgaccga 2760

ggggagcgga acctcatcta catcacctc atcgacagca ccgggaagat ccttgaacag 2820

cggtcgctca acaccatcca gcagttcgac taccagaaga aactcgacaa ccgggagaag 2880

gagagagtgg cgcccccca ggcttgggtcc gtcgtcggga cgattaagga cttgaacaa 2940

ggttacctgt cgcaagtgat ccacgagatc gttgacctga tgatccacta ccaagccgtc 3000

gtggtcctgg agaacctcaa cttcggcttc aagagcaaac gaaccggcat cgcggagaag 3060

gccgtgtacc agcagttcga aaaaatgctg atcgacaagc tgaactgcct cgtgctcaag 3120

gactaccccg ctgagaaggt cggcgggggtg ctgaaccctg accagctcac tgaccagttc 3180

accagcttcg caaagatggg caccagctcc ggcttctgt tctacgtgcc tgcgccatac 3240

acctgaaga tcgaccctc caccgggttc gtggaccct tcgtctggaa gaccatcaag 3300

aaccacgaga gccgcaagca ctctctggag ggcttcgact tctccacta cgacgtcaag 3360

accggggact tcatctgca cttaagatg aaccgcaacc tcagtttcca gcgcggcctg 3420

ccgggggttca tgcccgctt ggatatagtc ttcgagaaga atgagacgca gttcgacgcg 3480

aagggcaccc cgttcacgc cggaagcgc atcgtgccgg tcatcgagaa ccaccggttc 3540

accgggcgtt accgcgacct ataccggcg aacgagttga tcgccctcct ggaggagaag 3600

ggcatcgtgt tccgcgacgg ctccaacatc ctcccgaagc tgctcgaaaa cgacgactcc 3660

cacgccatcg acacgatggt cgcgtgatc cggtcggtgc tccagatgcg gaactccaac 3720

gccgcgacgg gcgaggacta catcaacagt ccggtccgcg atctgaacgg cgtctgcttc 3780

gactcccggt tcagaaccc cgagtggcgg atggacgcgg acgcaacgg cgcataccac 3840

atcgccctaa aagggaatt gctgtcaac cacctcaagg aatccaaaga cctaaagctc 3900

cagaacggca tctccaacca ggactggctg gcgtacatcc aggaactgcg gaacgggagc 3960

aaaaaacgtc ggatcaagca agattga 3987

<210> 51

<211> 3987

<212> DNA

<213> Artificial

<220><223> polynucleotide

<400> 51

atggcgggct ccaagaaacg ccgattaag caagataccc agttcgaggg gttcacgaac 60

ctctaccaag tgagcaagac cctccgattc gaactgattc ctgaggggaa gaccttcaag 120

cacatccagg agcaagggtt catcgaggag gacaaggcgc ggaacgacca ctacaaggaa 180

ctcaaaccca tcatcgaccg catctacaag acctacgccg atcagtcctt ccagctcgtg 240

cagttggact gggagaacct cagcgcggcc attgactcct accggaagga gaaaacggag 300

gagacgcgca acgcgtcat cgaggaacag gcaacctatc gcaacgccat ccacgactac 360

ttcatcgga ggactgaaa cctcactgac gcgattaaca agcggcacgc ggagatatac 420

aagggactct tcaaagcgga gctgtttaac ggcaagggtt tcaagcaact cggcactgtg 480

accacgaccg agcatgagaa cgccctgctc cgtccttcg acaagttcac cacctacttc 540

tccgggttct accgcaaccg caagaatgtc ttgagcgcgg aggacatcag cacggccatt 600

ccacatcgaa tcgtccaaga taacttcccg aagttcaagg agaactgcca catcttcacc 660

cgactcatta ctgctgtacc gtcgttacgc gaacacttcg agaactgcaa gaaggcaatt 720

ggaatcttcg tctctacgtc aatagaggag gtgttcagct tccctttcta caaccagctc 780

cttacgcaga cccagataga cctgtacaat cagctcctcg gtgggacag ccgggaggcg 840

gggactgaga agattaaagg gctcaacgag gtcttgaacc tggccatcca aaaaaacgat 900

gagacggcgc acatcatcgc ctgctgccc caccgggtca tcccgtgtt caagcagatc 960

ctcagtgaca ggaacacctt gagctttatc ctgaggagt tcaagagcga cgaggaggtg 1020

atccagagct tcigcaagta caaaaccttg ctgaggaaac agaactcctt ggagacggcg 1080

gaggcgtgt tcaacgagct gaactctatc gacttaactc acatattcat ctgcacaag 1140

aagctggaga ctattagctc tgactctgc gaccactggg acacctcctg caacgcgtc 1200

tacgagcgcc gcatctcgga gctgaccggg aagatcacca aatccgcgaa ggaaaaggctc 1260

cagcgttccc tcaaacacga ggatattaac ttacaggaga ttatctcagc ggctgggaag 1320

gagttgtcag aggcgttcaa gcagaaaact tccgagatcc tgagccacgc gcacgcagcg 1380

ctcgaccagc ctctgccac cacctcaaa aagcaggaag aaaaagagat cctcaagagc 1440

cagttggact cctgctggg gctctatcac ctctcagct ggttcgccgt cgatgagtcg 1500

aacgaggtgg accccgagtt ctccggcggt ctgaccggca tcaagctaga gatggagccg 1560

tccctcagct tctacaataa ggcccgcaac tacgcgacca aaaaacccta cagcgtggag 1620

aagttcaagc tgaacttcca gatgccgacc ttagcacgcg gttgggacgt aaacaggagag 1680

aagaacaatg gagccatcct gttcgtcaag aacgggcttt actacctcgg gataatgccc 1740

aagcagaagg gccgctacaa ggccctttcc ttcgagccga cggagaaaac ctccgagggg 1800

ttcgacaaga tgtactacga ctacttcccc gacgccgcca agatgatccc gaagtgtca 1860

acgcagctaa aagccgtgac cgcccacttc cagaccaca cgacgccgat cctgctgagc 1920

aacaacttca tcgagccctt tgagatcact aaggagatat acgacctgaa caaccccgag 1980

aaggagccca agaagtttca aaccgcctac gccaaaaaaa ctggcgacca aaagggtac 2040

agggaggcgc tgttaagtg gatcgacttc acacgcgact tcctttcgaa gtatacgaag 2100

acaacctcta ttgacctgag cagcctgcgt cctagctccc agtacaaaga ttgggagcag 2160

tactacgcgg agcttaatcc actactctac cacatctcat tccagcgcat cgctgagaag 2220

gaaatcatgg acgcggtgga gacaggcaaa ctgtacctct tccagatata caacaaagac 2280

ttcgctaagg ggcaccacgg gaagcccaac cttcatacgc tctactggac gggcctattc 2340

agccccgaaa atctggccaa gacctccatc aagctgaacg gccaaagcga gctgttctac 2400

agaccaaga gccggatgaa gcggatggcc cacaggctcg gcgagaaaat gcttaacaaa 2460

aagttgaagg accagaaaaa ccctatcccc gacacctct accaggaact gtacgactac 2520

gtgaaccaca ggctctcgca cgacctttcc gacgaggccc gtgcctact cccgaacgtc 2580

attaccaaag aggtttcgca cgagatcatc aaggaccggc ggttcacgag cgacaagttt 2640

ttctttcacg tccccatcac ccttaactac caggcggcca actccccatc caagttcaac 2700

cagcgtgtga atgcctacct caaggagcac ccagagacct cgatcattgg gatcgaccgg 2760

ggcgagcggg acctgatcta catcaccgtc atcgactcga cgggcaagat tcttgagcag 2820

agatcgttga ataccatata gcagttcgac taccagaaga aactcgacaa ccgcgagaag 2880

gagcgcgtgg cgcccccca ggcgtggctc gtggttggga cgattaagga cttgaacaaa 2940

ggttatctgt cccaagtcac ccacgagatc gttgatctga tgatccacta tcaggcagtg 3000

glggtgctgg agaattctaa cttcggcttc aagagtaagc ggacgggaat cgccgagaag 3060

gccgtgtacc agcagttcga gaagatgctg atcgacaagc tcaactgcct tgtgtgaaa 3120

gactaccggg ccgagaaggt cggcggcgtc ctcaaccgt accaacttac cgaccagttc 3180

acctccttcg ccaagatggg cactcagtcc gggttcttgt tctacgtccc cgcaccttac 3240

acctctaaga tcgacctct gactggcttc gtagatccat tcgtgtggaa gaccattaag 3300

aaccacgaga gccgcaagca cttcctggag ggcttcgact tcctgacta cgacgtgaag 3360

accggggact tcaccttca cttcaagatg aaccggaacc tcagcttcca gcggggcctg 3420

ccgggggttca tgccccgctg ggacatcgtg ttcgagaaga acgagacca gttcgacgcg 3480

aagggcacgc ctttcacgc cgggaagcgt atcgtgccgg tgatcgagaa ccatcgtttc 3540

acgggtcgct accgtgacct ctaccggcg aacgagctta tcgcactcct ggaggagaag 3600

ggcatcgtct tccgggacgg ctccaacatc ctcccgaac tgctggaaaa cgacgactct 3660

cacgccatcg acacgatggt ggccctcatc cggtcctgac tccaaatgcg gaacagcaac 3720

gccgccaccg gtgaggacta catcaacagc ccggtccggg atctgaacgg ggtgtgcttc 3780

gattcgcggt tccagaatcc tgagtggccg atggacgcgg atgcaaacgg ggcgtaccac 3840

atcgcgctca agggccagtt acttctgaac caccttaagg agtctaaaga ttgaaactc 3900

cagaacggga tctcgaaaca ggactggctg gcctacatcc aagagttgcg gaacggcagc 3960

aagaagcggc ggattaagca agattag 3987

<210> 52

<211> 3790

<212> DNA

<213> Artificial

<220><223> expression cassette

<400> 52

ccatgggcag caaactggaa aaatttacga attgttatag cctgtccaag accctgcgtt 60

tcaaagccat ccccgttggc aaaaccagg agaatatga taataaacgt ctgctggttg 120

aggatgaaaa aagagcagaa gactataagg gagtcaaaaa actgctggat cgggtactacc 180

tgagctttat aaatgacgtg ctgcatagca ttaaactgaa aaatctgaat aactatatta 240

gtctgttccg caagaaaacc cgaacagaga aagaaaataa agagctggaa aacctggaga 300

tcaatctgcg taaagagatc gcaaaagctt ttaaaggaaa tgaaggttat aaaagcctgt 360

tcaaaaaaga cattattgaa accatcctgc cggaatttct ggatgataaa gacgagatag 420

cgctcgtgaa cagcttcaac gggttcacga ccgcttcac gggctttttc gataacaggg 480

aaaatatggt tttagaggaa gccaaaagca cctcgatagc gttccgttgc attaatgaaa 540

atttgacaag atatatcagc aacatggata ttttcgagaa agttgatgcg atctttgaca 600

aacatgaagt gcaggagatt aaggaaaaaa ttctgaacag cgattatgat gttgaggatt 660

ttttcgaggg ggaatttttt aactttgtac tgacacagga aggtatagat gtgtataatg 720

ctattatcgg cgggttcgtt accgaatccg gcgagaaaat taagggtctg aatgagtaca 780

tcaatctgta taacccaaaag accaaacaga aactgccaaa attcaaaccg ctgtacaagc 840

aagtcctgag cgatcgggaa agcttgagct tttacggtga aggttatacc agcgacaggg 900

aggtactgga ggtctttcgc aataccctga acaagaacag cgaaattttc agctccatta 960

aaaagctgga gaaactgttt aagaattttg acgagtacag cagcgcaggt atttttgtga 1020

 agaacggacc tgccataagc accattagca aggatatattt tggagagtgg aatgttatcc 1080
 gtgataaatg gaacgcggaa tatgatgaca tacacctgaa aaagaaggct gtggtaactg 1140
 agaaatatga agacgatcgc cgcaaaagct ttaaaaaaat cggcagcttt agcctggagc 1200
 agctgcagga atatgaggac gccgacctga gcgtggctga gaaactgaag gaaattatta 1260
 tccaaaaagt ggatgagatt tacaaggat atggttagcag cgaaaaactg ttgatgcgg 1320
 acttcgttct ggaaaaagc ctgaaaaaaa atgatgctgt tgttgcgac atgaaagacc 1380
 tgctcgatag cgtaagagc ttgaaaaatt acattaaagc attctttggc gagggcaaag 1440

 aaacaaacag agacgaaagc ttttatggcg acttcgtcct ggcttatgac atcctgttga 1500
 aggtagatca tatatatgat gcaattcgta attacgtaac ccaaaagccg tacagcaaag 1560
 ataagttcaa actgtatttc cagaacccgc agtttatggg tggtctgggac aaagacaagg 1620
 agacagacta tcgcgccact attctgcgtt acggcagcaa gtactatctc gccatcatgg 1680
 acaaaaaata tgcaaagtgt ctgcagaaaa tcgataaaga cgacgtgaac ggaaattacg 1740
 aaaagattaa ttataagctg ctgccagggc ccaacaagat gttaccgaaa gtattttttt 1800
 ccaaaaaatg gatggcatat tataacccga gcgaggatat acagaagatt tacaaaaatg 1860

 ggaccttcaa aaagggggat atgttcaatc tgaatgactg ccacaaactg atcgattttt 1920
 ttaaagatag catcagccgt tatcctaaat ggtcaaacgc gtatgatttt aatttctccg 1980
 aaacggagaa atataaagac attgctggtt tctatcgcga agtcgaagaa cagggttata 2040
 aagttagctt tgaatcgcc agcaagaaag aggttgataa actggtggag gagggtaagc 2100
 tgtatatgtt tcagatttat aacaagact tiagcgacaa aagccacggt actcctaate 2160
 tgcatacgtat gtacttttaa ctgctgtttg atgagaataa ccacggccaa atccgtctct 2220
 ccggtggagc agaacttttt atgcggcgctg cgagcctaaa aaaggaagaa ctggtggtgc 2280

 atcccgccaa cagcccgtt gctaacaaaa atccagataa tcctaagaag accaccacac 2340
 tgtcgtacga tgtctataag gataaacgtt tctcggaaga ccagtatgaa ttgcatatac 2400
 cgatagcaat taataaatgc ccaaaaaaca ttttcaaat caacactgaa gttcgtgtgc 2460
 tgctgaacaa tgatgataat ccgtatgtga tcggaattga ccgtggggag agaaatctgc 2520
 tgtatattgt agtcgttgat ggcaagggca acatcgttga gcagtatagc ctgaatgaaa 2580
 taattaataa ttttaacggt atacgtatta aaaccgacta tcatagcctg ctggataaaa 2640
 aggagaaaga gcgttttgag gcacgcaaaa attggacgag catcgaaaac atcaaggaac 2700

tgaaggcagg atatatcagc caagtagtcc ataaaatctg tgaactgggtg gagaagtacg 2760
acgctgtcat tgccttgaa gacctcaata gcggctttaa aaacagccgg gtgaaggtgg 2820
agaaacaggt ataccaaaag ttgtaaaaga tgctcattga taagctgaac tatatggttg 2880
ataaaaagag caacccgtgc gccactggcg gtgcactgaa agggtaacaa attaccaata 2940
aatttgaaag ctttaaaagc atgagcacgc agaattgggtt tattttttat ataccagcat 3000
ggctgacgag caagattgac cccagcactg gttttgtcaa tctgtgaaa accaaataca 3060
caagcattgc ggatagcaaa aaatttattt cgagcttcga ccgtattatg tatgttccgg 3120

aggaagatct gtttgaattt gccctggatt ataaaaactt cagccgcacc gatgcagatt 3180
atatcaaaaa atggaagctg tacagttatg gtaatcgtat acgtatcttc cgtaatccga 3240
agaaaaacaa tgtgttcgat tgggaagagg tctgtctgac cagcgcgtat aaagaactgt 3300
tcaacaagta cggaataaat tatcagcaag gtgacattcg cgcactgctg tgtgaacagt 3360
cagataaagc attttatagc agctttatgg cgtgatgag cctgatgctc cagatgcgca 3420
acagcataac cggtcgcaca gatgttgact ttctgatcag ccctgtgaag aatagcgacg 3480
gcattctcta cgattccagg aactatgaag cacaggaaaa cgctattctg cctaaaaatg 3540

ccgatccaa cggcgcctat aatattgcac ggaaggttct gtgggcgatt ggacagttca 3600
agaaagcgga agatgagaag ctggataagg taaaaattgc tattagcaat aaggaatggc 3660
tggagtacgc acagacatcg gttaaacacg gtagtaaaag gccggcggcc acgaaaaagg 3720
ccggccaggc aaaaaagaaa aagggagcgg ccgcactcga gcaccaccac caccaccact 3780
gagcggccgc 3790

<210> 53
<211> 8986
<212> DNA
<213> Artificial
<220><223> vector
<400> 53

tggcgaatgg gacgcgcctt gtagcggcgc attaaagcgc gcgggtgtgg tggttacgcg 60

cagcgtgacc gctacacttg ccagcgcctt agcgcgccgt cctttcgctt tcttcccttc 120
ctttctcgcc acgttcgccg gctttccccc tcaagctcta aatcgggggc tcccttagg 180
gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg gtgatggttc 240
acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg agtcacgtt 300
ctttaatagt ggactcttgt tccaaactgg aacaacactc aacctatct cggtctattc 360
ttttgattta taagggattt tgccgatttc ggccatttgg ttaaaaaatg agctgattta 420

acaaaaaattt aacgcgaatt ttaacaaaat attaacgttt acaatttcag gtggcacttt 480

 tcggggaaat gtgcgcggaa ccctattttg tttatttttc taaatacatt caaatatgta 540
 tccgctcatg aattaattct tagaaaaact catcgagcat caaatgaaac tgcaatttat 600
 tcatacagg attatcaata ccatattttt gaaaaagccg tttctgtaat gaaggagaaa 660
 actcaccgag gcagttccat aggatggcaa gatcctggta tcggtctgcg attccgactc 720
 gtccaacatc aatacaacct attaatitcc cctcgtcaaa aataaggta tcaagtgaga 780
 aatcaccatg agtgacgact gaatccggtg agaatggcaa aagtttatgc atttctttcc 840
 agacttgttc aacaggccag ccattacgct cgtcatcaaa atcactcgca tcaaccaaac 900

 cgttattcat tcgtgattgc gcctgagcga gacgaaatac gcgatcgctg ttaaaaggac 960
 aattacaaac aggaatcgaa tgcaaccggc gcaggaacac tgccagcgca tcaacaatat 1020
 tttcacctga atcaggatat tcttctaata cctggaatgc tgttttcccg gggatcgag 1080
 tggtagtaaa ccatgcatca tcaggagtac ggataaaatg cttgatggtc ggaagaggca 1140
 taaattccgt cagccagttt agtctgacca tctcatctgt aacatcattg gcaacgctac 1200
 ctttgccatg tttcagaaac aactctggcg catcgggctt cccatacaat cgatagattg 1260
 tcgcacctga ttgcccgaca ttatcgcgag cccatttata cccatataaa tcagcatcca 1320

 tgttgaatt taatcgcggc ctagagcaag acgtttcccg ttgaatatgg ctcataacac 1380
 cccttgattt acigtattatg taagcagaca gttttattgt tcatgaccaa aatcccttaa 1440
 cgtgagtttt cgttccactg agcgtcagac cccgtagaaa agatcaaagg atcttcttga 1500
 gatccttttt tctgcgctg aatctgctgc ttgcaaacia aaaaaccacc gctaccagcg 1560
 gtggtttgtt tgccgatca agagctacca actctttttc cgaaggtaac tggcttcagc 1620
 agagcgaga taccaaatc tgctcttcta gtgtagccgt agttaggcca ccacttaag 1680
 aactctgtag caccgctac atacctcgt ctgctaatec tgttaccagt ggctgctgcc 1740

 agtggcgata agtcgtgtct taccgggttg gactcaagac gatagttacc ggataaggcg 1800
 cagcggtcgg gctgaacggg gggttcgtgc acacagccca gcttgagcg aacgacctac 1860
 accgaactga gatacctaca gcgtgagcta tgagaaagcg ccacgcttcc cgaagggaga 1920
 aaggcggaca ggtatccgt aagcggcagg gtcggaacag gagagcgac gagggagctt 1980
 ccaggggaa acgcctggta tctttatagt cctgtcgggt ttcgccacct ctgacttgag 2040
 cgtcgatttt tigtatgctc gtcagggggg cggagcctat ggaaaaacgc cagcaacgcg 2100
 gcctttttac ggttcttggc cttttgctgg cttttgctc acatgttctt tctgctgta 2160

tccccgtgatt ctgtggataa ccgtattacc gcctttgagt gagctgatac cgctcgccgc	2220
agccgaacga ccgagcgagc cgagtcagt agcgaggaag cggagagagc cctgatgcgg	2280
tatctctcc ttacgcatt gtgcgggtatt tcacaccgca tataatggtgc actctcagta	2340
caatctgctc tgatgccgca tagttaagcc agtatacact ccgctatcgc tacgtgactg	2400
ggatcatggct gcgccccgac acccgccaac acccgctgac gcgcctgac gggtttgtct	2460
gtccccgga tccgcttaca gacaagctgt gaccgtctcc gggagctgca tgtgtcagag	2520
gttttcaccg tcatcaccga aacgcgcgag gcagctgcgg taaagctcat cagcgtggtc	2580
gtgaagcgat tcacagatgt ctgcctgttc atccgcgtcc agctcgttga gtttctccag	2640
aagcgtaat gtctggcttc tgataaagcg ggccatgtta agggcggttt tttctgttt	2700
ggctactgat gcctccgtgt aagggggatt tctgttcatt ggggtaata taccgatgaa	2760
acgagagagg atgtcacga tacgggttac tgatgatgaa catgcccgtt tactggaacg	2820
ttgtgagggt aaacaactgg cggtatggat gcggcgggac cagagaaaaa tcactcaggg	2880
tcaatgccag cgcttcgtta atacagatgt aggtgttcca cagggtagcc agcagcatcc	2940
tgcgatgcag atccggaaca taatgggtgca gggcgtgac ttccgcgttt ccagacttta	3000
cgaaacacgg aaaccgaaga ccattcatgt tgttgctcag gtcgcagacg ttttgagca	3060
gcagtcgctt cagcttcgtt cgcgtatcgg tgattcattc tgctaaccag taaggcaacc	3120
ccgccagcct agccgggttc tcaacgacag gagcacgac atgcgcaccc gtggggccgc	3180
catgcccggc ataattggct gcttctgcc gaaacgtttg gtggcgggac cagtgcgaa	3240
ggcttgagcg agggcgtgca agattccgaa taccgcaagc gacaggccga tcatcgtgc	3300
gtccagcga aagcggctct cgccgaaaat gaccagagc gctgccgga cctgtctac	3360
gagttgcatg ataaagaaga cagtcataag tgcggcgacg atagtcattc cccgcgcca	3420
ccggaaggag ctgactgggt tgaaggctct caagggcac ggtcgagatc ccggtgccta	3480
atgagtgcgc taacttacat taattgcgtt gcgctcactg cccgctttcc agtcgggaaa	3540
cctgtcgtgc cagctgcatt aatgaatcgg ccaacgcgcg gggagaggcg gtttgcgtat	3600
tgggcgccag ggtgggtttt cttttcacca gtgagacggg caacagctga ttgcccttca	3660
ccgcctggcc ctgagagagt tgcagcaagc ggtccacgt ggtttgcccc agcaggcgaa	3720
aatctgttt gatgggtggt aacggcggga tataacatga gctgtcttcg gtatcgtct	3780
atcccactac cgagatatcc gcaccaacgc gcagcccga ctcggtaatg gcgcgattg	3840
cgccagcgc catctgatcg ttggcaacca gcatcgcagt gggaacgat ccctcattca	3900
gcatttgcat ggtttgttga aaaccggaca tggcactcca gtcgccttcc cgttcgcta	3960
tcggctgaat ttgattgcga gtgagatat tatgccagcc agccagacgc agacgcgccg	4020

agacagaact taatgggccc gctaacagcg cgatttgctg gtgaccaat gcgaccagat 4080
 gctccacgcc cagtcgcgta ccgtcttcat gggagaaaat aatactgttg atgggtgtct 4140
 ggtcagagac atcaagaaat aacgccgga cattaagtga ggcagcttcc acagcaatgg 4200
 catcctggtc atccagcgga tagttaatga tcagccact gacgcgttgc gcgagaagat 4260

 tgtgcaccgc cgctttacag gcttcgacgc cgcttcgttc taccatcgac accaccacgc 4320
 tggcaccag ttgatcgcg cgagatttaa tcgcccgac aatttgacg ggcgcgtgca 4380
 gggccagact ggaggtggca acgccaatca gcaacactg tttgccgcc agttgttgtg 4440
 ccacgcggtt gggaatgtaa ttcagctccg ccacgcgcg ttcactttt tcccgcgttt 4500
 tcgcagaaac gtggctggcc tggttacca cgcgggaaac ggtctgata gagacaccgg 4560
 catactctgc gacatcgat aacgttactg gtttcacatt caccaccctg aattgactct 4620
 ctccggcg ctatcatgcc ataccgcgaa aggttttgcg ccattcgatg gtgtccggga 4680

 tctcgacgct ctcccttatg cgactcctgc attaggaagc agcccagtag taggttagg 4740
 ccgttgagca ccgcccgcc aaggaatggt gcatgcaagg agatggcgcc caacagtc 4800
 ccggccagg ggctgccac cataccacg ccgaaacaag cgctcatgag cccgaagtgg 4860
 cgagcccgat ctccccatc ggtgatgtcg gcgatatagg cgcagcaac gcacctgtg 4920
 gcgccggtga tgccggccac gatgcgtccg gcgtagagga tcgagatctc gatccgcga 4980
 aattaatacg actcactata ggggaattgt gagcggataa caattccct ctagaaataa 5040
 tttgtttaa cttaagaag gagatatac atgggcagca aactggaaa atttacaat 5100

 tgttatagcc tgtccaagac cctgcgtttc aaagccatcc ccgttgcaa aaccaggag 5160
 aatattgata ataaacgtct gctggttgag gatgaaaaa gagcagaaga ctataaggga 5220
 gtcaaaaaac tgctggatcg gtactacctg agctttataa atgacgtgct gcatagcatt 5280
 aaactgaaa atctgaataa ctatattagt ctgttcgca agaaaaccg aacagagaaa 5340
 gaaaataaag agctggaaa cctggagatc aatctgcgt aagagatcg aaaagctttt 5400
 aaaggaaatg aaggttataa aagcctgttc aaaaaagaca ttattgaaac catcctgccg 5460
 gaatttctgg atgataaaga cgagatagcg ctctgaaca gcttcaacgg gttcacgacc 5520

 gccttcacgg gctttttcga taacaggga aatatgttt cagaggaagc caaaagcacc 5580
 tcgatagcgt tccgttgcat taatgaaaat ttgacaagat atatcagca catggatatt 5640
 ttcgagaaag ttgatcgat ctttgacaaa catgaagtgc aggagattaa ggaaaaatt 5700
 ctgaacacg attatgatg tgaggatttt ttcgagggg aattttttaa ctttgtactg 5760
 acacaggaag gtatagatg gtataatgct attatcggcg ggttcgttac cgaatccggc 5820
 gagaaaatta aggtctgaa tgagtcacatc aatctgtata accaaaagac caaacagaaa 5880

ctgccaaaat tcaaaccgct gtacaagcaa gtcctgagcg atcgggaaag cttgagcttt 5940

tacgggtgaag gttataccag cgacgaggag gtactggagg tctttcgcaa taccctgaac 6000

aagaacagcg aaattttcag ctccattaaa aagctggaga aactgtttaa gaattttgac 6060

gagtacagca ggcgaggtat ttttgtgaag aacggacctg ccataagcac cattagcaag 6120

gatatTTTTT gagagtggaa tgttatccgt gataaatgga acgcggaata tgatgacata 6180

cacctgaaaa agaaggctgt ggtaactgag aaatatgaag acgatcgccg caaaagcttt 6240

aaaaaaatcg gcagctttag cctggagcag ctgcaggaat atgcggacgc cgacctgagc 6300

gtggtcgaga aactgaagga aattattatc caaaaagtgg atgagattta caaggtatat 6360

ggtagcagcg aaaaactgtt tgatgcggac ttcgttctgg aaaaaagcct gaaaaaaaat 6420

gatgctgttg ttgcgatcat gaaagacctg ctcgatagcg ttaagagctt tgaaaattac 6480

attaaagcat tctttggcga gggcaaagaa acaaacagag acgaaagctt ttatggcgac 6540

ttcgtcctgg cttatgacat cctgttgaag gtagatcata tataatgatgc aattcgtaat 6600

tacgtaaccc aaaagccgta cagcaaagat aagtccaac tgtatttcca gaaccgcag 6660

tttatgggtg gctgggacaa agacaaggag acagactatc gcgccactat tctgcgttac 6720

ggcagcaagt actatctcgc catcatggac aaaaaatag caaagtgtct gcagaaaatc 6780

gataaagacg acgtgaacgg aaattacgaa aagattaatt ataagctgct gccagggccc 6840

aacaagatgt taccgaaagt atttttttcc aaaaaatgga tggcatacta taaccgcagc 6900

gaggatatac agaagattta caaaaatggg accttcaaaa agggggatata gttcaatctg 6960

aatgactgcc aaaaactgat cgatTTTTTT aaagatagca tcagccgtta tcctaaatgg 7020

tcaaacgcgt atgattttaa tttctccgaa acggagaaat ataagacat tgctggtttc 7080

tatcggaag tcgaagaaca gggttataaa gttagctttg aatcgccag caagaaagag 7140

gttgataaac tggtaggagga gggtaagctg tatatgtttc agatttataa caaagacttt 7200

agcgacaaaa gccacggtac tcctaactctg catacgatgt actttaaact gctgtttgat 7260

gagaataacc acggccaaat ccgtctctcc ggtggagcag aactttttat gcggcgtgcg 7320

agcctaaaaa aggaagaact ggtggtgcat cccgccaaca gcccattgc taacaaaaat 7380

ccagataatc ctaagaagac caccacactg tcgtacgatg tctataagga taaacgtttc 7440

tcggaagacc agtatgaatt gcatataccg atagcaatta ataaatgccc aaaaaacatt 7500

ttcaaaatca aactgaagt tcgtgtgctg ctgaaacatg atgataatcc gtatgtgatc 7560

ggaattgacc gtggggagag aaatctgctg tatattgtag tcgttgatgg caagggaac 7620

atcggttgagc agtatagcct gaatgaaata attaataatt ttaacggtat acgtattaaa 7680

accgactatc atagcctgct ggataaaaag gagaaagagc gttttgaggc acgccaaaat 7740

tggacgagca tcgaaaacat caaggaactg aaggcaggat atatcagcca agtagtccat 7800

aaaatctgtg aactggtgga gaagtiacgac gctgtcattg ccctggaaga cctcaatagc 7860

ggctttaaaa acagccgggt gaaggtggag aaacaggtat accaaaagtt tgaaaagatg 7920

ctcattgata agctgaacta tatggttgat aaaaagagca acccgtgcgc cactggcgggt 7980

gcactgaaag ggiaccaaat taccaataaa ttgaaagct ttaaaagcat gagcacgcag 8040

aatgggttta tttttatat accagcatgg ctgacgagca agattgaccc cagcactggt 8100

tttgtcaatc tgctgaaaac caaatcacaca agcattgcgg atagcaaaaa atttatttcg 8160

agcttcgacc gtattatgta tgttccggag gaagatctgt ttgaatttgc cctggattat 8220

aaaaacttca gccgcaccga tgcagattat atcaaaaaat ggaagctgta cagttatggt 8280

aatcgatac gtatcttccg taatccgaag aaaaacaatg tgttcgattg ggaagaggtc 8340

tgtctgacca gcgcgtataa agaactgttc aacaagtacg gaataaatta tcagcaaggt 8400

gacattcgcg cactgctgtg tgaacagtca gataaagcat tttatagcag ctttatggcg 8460

ctgatgagcc tgatgctcca gatgcgcaac agcataaccg gtcgcacaga tgttgacttt 8520

ctgatcagcc ctgtgaagaa tagcgacggc atctttctacg attccaggaa ctatgaagca 8580

caggaaaacg ctattctgcc taaaatgcc gatgccaacg gcgcctataa tattgcacgg 8640

aaggttctgt gggcgattgg acagttcaag aaagcggaag atgagaagct ggataaggta 8700

aaaattgcta ttagcaataa ggaatggctg gagtacgcac agacatcggt taaacacggt 8760

agtaaaaggc cggcggccac gaaaaaggcc ggccaggcaa aaaagaaaaa gggagcggcc 8820

gcactcgagc accaccacca ccaccactga gatccggctg ctaacaaagc ccgaaaggaa 8880

gctgagttgg ctgctgccac cgctgagcaa taactagcat aacccttgg ggcctctaaa 8940

cgggtcttga ggggtttttt gctgaaagga ggaactatat ccggat 8986

<210> 54

<211> 23

<212> DNA

<213> Artificial

<220><223> spacer sequence

<400> 54

tcactctgtgc cctccctcc ctg 23

<210> 55

<211> 23

<212> DNA
 <213> Artificial
 <220><223> spacer sequence
 <400> 55
 gcggatgttc caatcagtac gca 23
 <210> 56

 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> spacer sequence
 <400> 56
 cagcccgctg gccctgtaaa gga 23
 <210> 57
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> spacer sequence
 <400> 57
 tatgagttac aacgaacacc tca 23
 <210> 58
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> spacer sequence
 <400> 58
 cacgtctcat atgcccttg gca 23

 <210> 59
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> spacer sequence
 <400> 59
 gaacatgaaa acttaaataag aac 23

<210> 60
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> spacer sequence
 <400> 60
 gtcaggttgg ctgctgggct ggc 23
 <210> 61
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> spacer sequence
 <400> 61
 acagatgggg ctggacaatt ttt 23
 <210> 62
 <211> 43
 <212> RNA
 <213> Artificial
 <220><223> CRISPR RNA
 <400> 62
 aaauucuacu aaguguagau ggaaucuuu cugcagcacc ugg 43