



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0116282

(43) 공개일자 2019년10월14일

- (51) 국제특허분류(Int. Cl.)
C12N 15/63 (2006.01) C12N 15/64 (2006.01)
C12N 15/66 (2006.01)
- (52) CPC특허분류
C12N 15/63 (2013.01)
C12N 15/09 (2013.01)
- (21) 출원번호 10-2019-7022014
(22) 출원일자(국제) 2018년02월09일
심사청구일자 없음
- (85) 번역문제출일자 2019년07월25일
(86) 국제출원번호 PCT/US2018/017573
(87) 국제공개번호 WO 2018/148511
국제공개일자 2018년08월16일
- (30) 우선권주장
62/457,493 2017년02월10일 미국(US)
- (71) 출원인
지머젠 인코포레이티드
미국 94608 캘리포니아 에머리빌 스위트 105 호턴 스트리트 5980
- (72) 발명자
제이켄드 브라이언
미국 94611 캘리포니아 오클랜드 몬테 크레스타 애비뉴 91비
반 로섬 헨드릭 마리누스
미국 94608 캘리포니아 오클랜드 발레호 스트리트 6420
- (74) 대리인
석혜선, 김용인

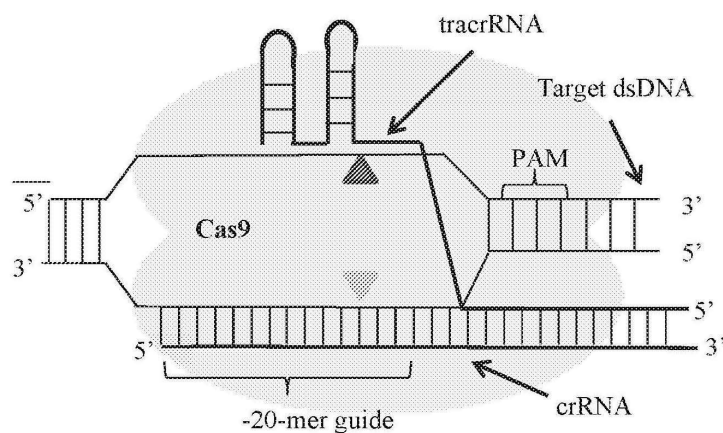
전체 청구항 수 : 총 49 항

(54) 발명의 명칭 복수의 숙주를 위한 다중 DNA 구조체의 조립 및 편집을 위한 모듈식 범용 플라스미드 디자인 전략

(57) 요약

본 발명은 생체외에서 고 처리량 DNA 조립 반응을 위한 방법, 조성물 및 키트를 기술한다. 본 발명은 추가로 미리 확증된 CRISPR 프로토스페이서/프로토스페이서 인접 모티프 서열 조합을 포함하는 복제 태그 분절이 측면에 위치하는 모듈식 삽입체 DNA 부분을 포함하는 모듈식 CRISPR DNA 구조체를 더 기술한다.

대표도 - 도1a



CRISPR- Cas9
WT crRNA/tracrRNA

(52) CPC특허분류

C12N 15/10 (2013.01)

C12N 15/102 (2013.01)

C12N 15/64 (2013.01)

C12N 15/66 (2013.01)

C12N 2310/20 (2017.05)

명세서

청구범위

청구항 1

CRISPR 다중 복제 부위를 포함하는 재조합 모듈식 CRISPR DNA 구조체로서, 상기 다중 복제 부위는 다음을 포함하는 재조합 모듈식 CRISPR DNA 구조체:

- a) 적어도 2개의 별개의 복제 태그(cTAG), 각 cTAG는
 - i) 각각이 프로토스페이서 인접 모티프(PAM)에 작동 가능하게 연결된 프로토스페이서 서열을 포함하는 하나 이상의 확증된 CRISPR 랜딩 부위를 포함하며; 상기 확증된 CRISPR 랜딩 부위 중 적어도 하나는 모듈식 CRISPR DNA 구조체 내에서 고유하다; 및
- b) 하나 이상의 DNA 삽입체 부분(들);
 - i) 상기 별개의 cTAG의 각각은 하나 이상의 DNA 삽입체 부분(들)의 각각의 주위의 플랭킹 위치에 분포된다.

청구항 2

제 1 항에 있어서,
상기 모듈식 CRISPR DNA 구조체는 원형인 모듈식 CRISPR DNA 구조체.

청구항 3

제 1 항에 있어서,
상기 모듈식 CRISPR DNA 구조체는 선형인 모듈식 CRISPR DNA 구조체.

청구항 4

제 1 항에 있어서,
상기 모듈식 CRISPR DNA 구조체는 유기체의 게놈에 통합되는 모듈식 CRISPR DNA 구조체.

청구항 5

제 1 항 내지 제 4 항 중 어느 한 항에 있어서,
상기 별개의 cTAG의 적어도 하나는 적어도 2개의 확증된 CRISPR 랜딩 부위를 포함하는 모듈식 CRISPR DNA 구조체.

청구항 6

제 5 항에 있어서,
CRISPR 랜딩 부위의 적어도 하나는 Cas9 엔도뉴클레아제를 위한 것인 모듈식 CRISPR DNA 구조체.

청구항 7

제 6 항에 있어서,
CRISPR 랜딩 부위의 적어도 하나는 Cpf1 엔도뉴클레아제를 위한 것인 모듈식 CRISPR DNA 구조체.

청구항 8

제 1 항에 있어서,
상기 별개의 cTAG의 적어도 하나는 회귀(≥ 8 염기 길이) 제한 엔도뉴클레아제 부위를 포함하는 모듈식 CRISPR DNA 구조체.

청구항 9

다음 단계를 포함하는 재조합 핵산 분자를 제조하는 방법:

a) 분해된 DNA 말단을 생성하기 위해 복수의 DNA 삽입체 부분의 적어도 2개에서 표적화된 cTAG(s)의 분해를 허용하는 조건하에서 다음을 포함하는 혼합물을 배양하는 단계:

i) 복수의 DNA 삽입체 부분, 여기서 각각의 DNA 삽입체 부분은 2개의 복제 태그(cTAGs)에 인접하고, 각각의 cTAG는

1) 각각이 프로토스페이서 인접 모티프(PAM)에 작동 가능하게 연결된 프로토스페이서 서열을 포함하는 하나 이상의 확증된 CRISPR 랜딩 부위를 포함한다; 및

ii) 복수의 DNA 삽입체 부분의 적어도 2개에 존재하는 상기 cTAGs의 적어도 하나를 표적으로 하는 하나 이상의 CRISPR 복합체, 각각의 CRISPR 복합체는

1) CRISPR 엔도뉴클레아제, 및

2) 상기 표적화된 cTAG의 하나에 상기 CRISPR 엔도뉴클레아제를 모집할 수 있는 가이드 RNA 또는 가이드 RNAs를 포함한다, 및

b) 분해된 DNA 말단의 공유 결합을 허용하는 조건하에서 (a)에서 생성된 분해된 DNA 말단으로 DNA 삽입체 부분들을 배양하는 단계; 여기서 생성된 재조합 핵산 분자는 상기 방법에서 결합된 최초 삽입체 부분의 완전한 cTAG 서열을 포함한다.

청구항 10

제 9 항에 있어서,

분해된 DNA 말단은 분해된 말단의 공유 결합 전에 서로에 대해 혼성화될 수 있는 오버행 서열을 갖는 점착성 말단인 방법.

청구항 11

제 10 항에 있어서,

CRISPR 엔도뉴클레아제는 Cpf1인 방법.

청구항 12

제 9 항에 있어서,

분해된 DNA 말단은 평활 말단인 방법.

청구항 13

제 12 항에 있어서,

평활 말단은 ssDNA 엑소뉴클레아제로 추가로 분해되어 오버행 서열을 갖는 점착성 말단을 생성하며, 여기서 단계 b)는 분해된 말단의 공유 결합 전에 오버행 서열에 혼성화될 수 있는 가교 DNA 서열을 첨가하는 단계를 더 포함하는 방법.

청구항 14

제 9 항, 제 12 항 및 제 13 항 중 어느 한 항에 있어서,

CRISPR 엔도뉴클레아제는 Cas9인 방법.

청구항 15

제 9 항 내지 제 13 항 중 어느 한 항에 있어서,

i) 단계(b) 전에 CRISPR 복합체로부터 분해된 cTAG 서열을 분리하는 단계, 또는 ii) 단계(b) 전에 CRISPR 복합

체를 불활성화시키는 단계를 더 포함하는 방법.

청구항 16

제 15 항에 있어서,

분리 단계는 DNA 정제 단계를 포함하는 방법.

청구항 17

제 15 항에 있어서,

불활성화 단계는 상기 CRISPR 복합체의 열 또는 화학적 불활성화를 포함하는 방법.

청구항 18

제 9 항 내지 제 13 항 중 어느 한 항에 있어서,

복수의 DNA 삽입체 부분의 각각에 대한 2개의 cTAG는 cTAG 쌍을 형성하고, 상기 cTAG 쌍은 상기 방법에서 절찰된 DNA 삽입체 부분의 모든 다른 cTAG 쌍으로부터 독특한 방법.

청구항 19

제 18 항에 있어서,

각 cTAG 쌍에서 cTAGs의 적어도 하나는 상이한 cTAG 쌍에서 적어도 하나의 다른 cTAG와 동일한 방법.

청구항 20

다음 단계를 포함하는 DNA 서열 편집 방법:

a) 다음을 제공하는 단계:

i) 제 1 항의 모듈식 CRISPR DNA 구조체:

ii) 대체 DNA 삽입체 부분, 상기 대체 DNA 삽입체 부분은 제 1 및 제 2 삽입체 cTAG가 측면에 위치된다;

1) 제 1 삽입체 cTAG는 모듈식 CRISPR DNA 구조체의 별개의 cTAG의 하나의 확증된 CRISPR 랜딩 부위를 포함하고, 제 2 삽입체 cTAG는 모듈식 CRISPR DNA 구조체의 별개의 cTAG의 하나의 확증된 CRISPR 랜딩 부위를 포함한다; 및

iii) 제 1 및 제 2 삽입체 cTAG를 표적으로 하는 제 1 및 제 2 CRISPR 복합체, 각각의 CRISPR 복합체는 다음을 포함한다:

1) CRISPR 엔도뉴클레아제, 및

2) 상기 표적화된 삽입체 cTAG의 하나에 상기 CRISPR 엔도뉴클레아제를 모집할 수 있는 가이드 RNA;

여기서 부분 (i) 및 (ii)는 각각 (iii) 부분과 함께 단일 또는 별도의 반응으로 배양되며; 제 1 및 제 2 CRISPR 복합체는 제 1 및 제 2 삽입체 cTAG 및 이들의 상응하는 별개의 cTAG를 절단하여 분해된 DNA 말단을 생성시킨다, 및

b) 분해된 DNA 말단의 공유 결합을 허용하는 조건하에서 단계(a)에서 생성된 분해된 DNA 말단으로 대체 DNA 삽입체 부분 및 모듈식 CRISPR DNA 구조체를 배양하는 단계; 여기서 생성된 편집된 모듈식 CRISPR DNA 구조체는 상기 방법에서 공유 결합된 최초의 삽입체 부분의 완전한 cTAG 서열을 포함한다.

청구항 21

제 20 항에 있어서,

단계(b)의 반응은 기능성 리가아제를 포함하는 방법.

청구항 22

제 20 항에 있어서,

i) 단계(b) 전에 CRISPR 복합체로부터 절단된 제 1 및 제 2 삽입체 cTAG 및 이들의 상응하는 구별된 cTAG를 분리하는 단계, 또는

ii) 단계(b) 전에 CRISPR 복합체를 불활성화시키는 단계를 더 포함하는 방법.

청구항 23

제 22 항에 있어서,

분리 단계는 DNA 정제 단계를 포함하는 방법.

청구항 24

제 22 항에 있어서,

불활성화 단계는 상기 CRISPR 복합체의 열 또는 화학적 불활성화를 포함하는 방법.

청구항 25

제 20 항에 있어서,

분해된 DNA 말단은 분해된 말단의 공유 결합 전에 서로에 대해 혼성화될 수 있는 오버행 서열을 갖는 점착성 말단인 방법.

청구항 26

제 20 항에 있어서,

CRISPR 엔도뉴클레아제는 Cpf1인 방법.

청구항 27

제 20 항에 있어서,

분해된 DNA 말단은 평활 말단인 방법.

청구항 28

제 27 항에 있어서,

평활 말단은 ssDNA 엑소뉴클레아제로 추가로 분해되어 오버행 서열을 갖는 점착성 말단을 생성하며, 여기서 단계 b)는 분해된 말단의 공유 결합 전에 오버행 서열에 혼성화될 수 있는 가교 DNA 서열을 첨가하는 단계를 더 포함하는 방법.

청구항 29

제 20 항, 제 27 항 및 제 28 항 중 어느 한 항에 있어서,

CRISPR 엔도뉴클레아제는 Cas9인 방법.

청구항 30

다음에 포함하는 DNA 서열 편집 방법:

a) 다음을 제 1 반응에 제공하는 단계:

i) 제 1 항의 모듈식 CRISPR DNA 구조체;

ii) 모듈식 CRISPR DNA 구조체에서 적어도 하나의 별개의 cTAG를 각각 표적으로 하는 적어도 2개의 CRISPR 복합체로서, 각각의 CRISPR 복합체는 다음을 포함한다:

1) CRISPR 엔도뉴클레아제, 및

2) 상기 표적화된 별개의 cTAGs의 하나에 상기 CRISPR 엔도뉴클레아제를 모집할 수 있는 가이드 RNA;

여기서 제 1 및 제 2 CRISPR 복합체는 모듈식 CRISPR DNA 구조체에서 표적화된 별개의 cTAG를 절단하여 분해된

cTAG를 생성한다, 및

b) 부분(i)의 모듈식 CRISPR DNA 구조체 및 부분(ii)의 대체 DNA 삽입체 부분의 결합을 허용하여 편집된 모듈식 CRISPR DNA 구조체를 형성하는 조건하에서 다음을 제 2 반응에 제공하는 단계:

i) 단계(a)에서 생성된 분해된 cTAGs를 갖는 모듈식 CRISPR DNA 구조체; 및

ii) 대체 DNA 삽입체 부분, 상기 대체 DNA 삽입체 부분은 제 1 및 제 2 삽입체 cTAG가 측면에 위치되며;

1) 제 1 삽입체 cTAG는 단계(a)에서 절단된 분해되지 않은 별개의 cTAGs의 하나와 동일하거나 실질적으로 동일한 서열을 포함하며, 제 2 삽입체 cTAG는 단계(a)에서 절단된 분해되지 않은 별개의 cTAG의 다른 하나와 동일하거나 실질적으로 동일한 서열을 포함한다; 여기서 생성된 편집된 모듈식 CRISPR DNA 구조체는 단계(a)에서 표적화된 분해되지 않은 별개의 cTAGs와 동일하거나 실질적으로 동일한 재조합 cTAG를 포함한다.

청구항 31

제 30 항에 있어서,

단계(b)의 결합은 결찰이고 대체 DNA 삽입체 부분은 또한 결찰 전에 CRISPR 복합체에 의해 분해되는 방법.

청구항 32

제 30 항에 있어서,

CRISPR 엔도뉴클레아제는 Cpf1인 방법.

청구항 33

제 30 항에 있어서,

단계(b)의 결합은 상동성 재조합인 방법.

청구항 34

제 30 항 또는 제 33 항에 있어서,

CRISPR 엔도뉴클레아제는 Cas9인 방법.

청구항 35

제 30 항에 있어서,

제 1 및 제 2 반응은 분리되지 않는 방법.

청구항 36

제 33 항 또는 제 35 항에 있어서,

DNA 삽입체 부분은 CRISPR 엔도뉴클레아제에 의해 분해되지 않는 방법.

청구항 37

제 33 항 또는 제 35 항에 있어서,

분해된 cTAG를 갖는 모듈식 CRISPR DNA 구조체 및 분해되지 않은 대체 DNA 삽입체 부분은 모두 단일 가닥 DNA(ssDNA) 엑소뉴클레아제로 분해되어, 서로에 대해 혼성화될 수 있는 양립 가능한 오버행 DNA 말단을 생성하는 방법.

청구항 38

제 37 항에 있어서,

제 2 반응은 양립 가능한 오버행 DNA 말단의 공유 결합 전에 임의의 서열 간격에 채워질 수 있는 폴리머라제 또는 리가아제를 더 포함하는 방법.

청구항 39

제 36 항에 있어서,

삽입체 부분은 CRISPR 분해를 방지하도록 화학적으로 변형되는 방법.

청구항 40

제 36 항에 있어서,

제 1 및 제 2 삽입체 cTAG는 확증된 CRISPR 랜딩 부위에 돌연변이 PAM 또는 돌연변이 프로토스페이서를 포함하는 방법.

청구항 41

CRISPR 다중 복제 부위를 포함하는 재조합 모듈식 CRISPR DNA 구조체를 포함하는 숙주 세포 계놈으로서, 상기 다중 복제 부위는 다음을 포함하는 숙주 세포 계놈:

- a) 적어도 2개의 별개의 복제 태그(cTAG), 각 cTAG는
 - i) 각각이 프로토스페이서 인접 모티프(PAM)에 작동 가능하게 연결된 프로토스페이서 서열을 포함하는 하나 이상의 확증된 CRISPR 랜딩 부위를 포함하며; 상기 확증된 CRISPR 랜딩 부위 중 적어도 하나는 모듈식 CRISPR DNA 구조체 내에서 고유하다; 및
- b) 하나 이상의 DNA 삽입체 부분(들);
 - i) 상기 별개의 cTAG의 각각은 하나 이상의 DNA 삽입체 부분(들)의 각각의 주위의 플랭킹 위치에 분포된다.

청구항 42

다음을 포함하여 재조합 핵산 분자를 제조하는 방법:

분해된 DNA 말단을 갖는 삽입체 부분을 생성하기 위해 적어도 2개의 복수의 DNA 삽입체 부분에서 하나 이상의 제한 효소에 의해 표적화된 cTAG의 분해를 허용하는 조건하에서 a) 다음을 포함하는 혼합물을 배양하는 단계:

- i) 2개의 복제 태그(cTAGs)가 측면에 위치되는 복수의 DNA 삽입체 부분, 각 cTAG는
 - 1) 각각이 프로토스페이서 인접 모티프(PAM)에 작동 가능하게 연결된 프로토스페이서 서열을 포함하는 하나 이상의 확증된 CRISPR 랜딩 부위; 및
 - 2) 회귀(≥ 8 개 염기) 제한 효소 인식 부위; 적어도 2개의 삽입체 부분의 cTAGs의 적어도 하나는 동일한 제한 효소 부위를 포함한다;
- ii) 복수의 DNA 삽입체 부분의 적어도 2개에서 회귀 제한 효소 부위를 표적으로 하는 하나 이상의 제한 효소; 및
- b) 분해된 DNA 말단의 공유 결합을 허용하는 조건하에서 단계(a)에서 생성된 분해된 DNA 말단으로 DNA 삽입체 부분들을 배양하는 단계; 여기서 생성된 재조합 핵산 분자는 상기 방법에서 공유 결합된 최초의 삽입체 부분의 완전한 cTAG 서열을 포함한다.

청구항 43

다음을 포함하는 DNA 서열 편집 방법:

- a) 다음을 제공하는 단계:
 - i) 제 1 항의 모듈식 CRISPR DNA 구조체, 적어도 2개의 별개의 cTAGs는 회귀(≥ 8 개 염기) 제한 효소 인식 부위를 포함한다;
 - ii) 대체 DNA 삽입체 부분, 상기 대체 DNA 삽입체 부분은 제 1 및 제 2 삽입체 cTAG가 측면에 위치된다;
 - 1) 제 1 삽입체 cTAG는 모듈식 CRISPR DNA 구조체의 별개의 cTAG의 하나의 회귀 제한 효소 인식 부위를 포함하고, 제 2 삽입체 cTAG는 모듈식 CRISPR DNA 구조체의 다른 별개의 cTAG의 회귀 제한 효소 인식 부위를

포함한다; 및

iii) 제 1 및 제 2 삽입체 cTAG에서 회귀 제한 효소 부위를 표적으로 하는 하나 이상의 제한 효소;

여기서 부분 (i) 및 (ii)는 각각 (iii) 부분과 함께 단일 또는 별도의 반응으로 배양되며; 하나 이상의 제한 효소는 제 1 및 제 2 삽입체 cTAG의 회귀 제한 효소 인식 부위 및 이들의 상응하는 별개의 cTAG를 절단하여 분해된 DNA 말단을 생성시킨다, 및

b) 분해된 DNA 말단의 공유 결합을 허용하는 조건하에서 단계(a)에서 생성된 분해된 DNA 말단으로 대체 DNA 삽입체 부분 및 모듈식 CRISPR DNA 구조체를 배양하는 단계; 여기서 생성된 편집된 모듈식 CRISPR DNA 구조체는 상기 방법에서 공유 결합된 최초의 삽입체 부분의 완전한 cTAG 서열을 포함한다.

청구항 44

제 43 항에 있어서,

제 1 및 제 2 삽입체 cTAG는 동일한 회귀 제한 효소 부위를 포함하는 방법.

청구항 45

다음을 포함하여 재조합 핵산 분자를 제조하는 방법:

적어도 2개의 DNA 삽입체 부분에서 공유된 cTAG의 분해를 허용하여, 적어도 2개의 DNA 삽입체 부분에서 양립 가능한 오버행 DNA 말단을 생성하는 조건하에서 a) 다음을 포함하는 혼합물을 배양하는 단계:

i) 각각이 2개의 복제 태그(cTAGs)가 측면에 위치되는 복수의 DNA 삽입체 부분, 각 cTAG는

1) 각각이 프로토스페이서 인접 모티프(PAM)에 작동 가능하게 연결된 프로토스페이서 서열을 포함하는 하나 이상의 확증된 CRISPR 랜딩 부위; DNA 삽입체 부분의 적어도 2개는 동일한 cTAG를 공유한다;

ii) 단일 가닥 DNA(ssDNA) 엑소뉴클레아제, 및

b) 적어도 2개의 DNA 삽입체 부분의 양립 가능한 오버행 DNA 말단의 혼성화 및 공유 결합을 허용하는 조건하에서 (a)에서 생성된 분해된 cTAG로 DNA 삽입체 부분을 배양하는 단계; 여기서 생성된 재조합 핵산 분자는 분해 전 공유된 cTAG의 완전한 cTAG 서열을 포함한다.

청구항 46

제 45 항에 있어서,

단계(a)의 혼합물은 양립 가능한 오버행 DNA 말단의 공유 결합 전에 임의의 서열 간격에 채워질 수 있는 폴리머라제를 더 포함하는 방법.

청구항 47

제 45 항에 있어서,

복수의 DNA 삽입체 부분은 단계(a) 전에 이중 가닥 DNA(dsDNA) 엑소뉴클레아제로 분해되는 방법.

청구항 48

제 45 항에 있어서,

하나 이상의 복수의 DNA 삽입체 부분은 단계(a)에서 배양 전에 CRISPR 엔도뉴클레아제에 의해 절단되는 방법.

청구항 49

하나 이상의 cTAG가 SEQ ID NO: 65-74, 78-81 및 이들의 조합으로 이루어진 그룹으로부터 선택되는 제 1 항 내지 제 48 항 중 어느 한 항의 삽입체 부분.

발명의 설명

기술 분야

[0001] 본 발명은 생체외에서 유도된 유전자 서열 편집을 위해 사용되는 시스템, 방법 및 조성물에 관한 것이다. 본 발명은, 특히, 개선된 DNA 복제, 올리고뉴클레오타이드의 조립 및 미생물의 개선을 위해 유도된 서열 편집 복합체를 사용하는 방법을 기술한다.

[0002] 본 발명은 DARPA에 의해 수여된 계약 번호 HR0011-15-9-0014에 따라 정부 지원으로 이루어졌다. 정부는 발명에 대해 특정 권리를 갖는다.

배경 기술

[0003] 생물학에서 중요한 관심 영역은 유전자 서열의 생체 외 및 생체 내 표적화된 변형이다. 사실, 학술적 및 상업적 유전자 연구의 가장 중요한 병목 중 하나는 새로운 유전체 구조체가 테스트 전에 생성되거나 나중에 변형될 수 있는 속도이었다.

[0004] 제한 효소 부위 인식 또는 DNA 혼성화 및 증폭에 의존하는 현재 이용 가능한 복제 기술은 느리고, 신뢰성이 낮으며, 이후의 변형에 대해 다루기가 어려운 것으로 입증되었다. 덩어리진 규칙적인 간격을 갖는 짧은 회문구조 반복체(CRISPR) 유전자 편집 시스템의 발견은 연구자들에게 유전자 변형을 위한 추가 경로를 제공했다. 그러나, 이러한 새로운 접근법은 높은 처리량의 모듈식 복제 응용분야에서는 비실용적이다.

[0005] 예를 들어, CRISPR 편집 위치는 종종 프로토스페이서 인접 모티프(PAMs)의 위치에 의해 제한된다. 새로운 CRISPR 가이드 RNA 디자인과 유전자 표적화는 시간과 비용이 많이들뿐만 아니라 효율성이 떨어지며 표적 외의 돌연변이 가능성이 있다.

발명의 내용

해결하려는 과제

[0006] 따라서, 유전자 서열의 표적화된 변경을 위한 개선된 조성물 및 방법에 대한 필요성이 존재한다.

과제의 해결 수단

[0007] 일부 실시태양에서, 본 발명은 모듈식 CRISPR DNA 구조체를 이용하여 생체 내 및 생체외 고 처리량 DNA 조립 반응을 위한 방법, 조성물 및 키트를 교시한다.

[0008] 따라서, 일부 실시태양에서, 본 발명은 사전-검증된 CRISPR 프로토스페이스/ 프로토스페이스 인접 모티프(PAM) 서열 조합을 포함하는 tag 분절을 복제함으로써 측면으로 위치하는 모듈식 삽입 DNA 부분을 포함하는 CRISPR DNA 구조체를 교시한다. 일부 실시태양에서, 본 발명은 CRISPR 엔도뉴클레아제로 DNA를 분해하는 것을 교시한다. 일부 실시태양에서, 본 발명은 타입 II - 클래스 2 CRISPR 엔도뉴클레아제(예를 들어, Cas9)로 DNA를 분해하는 것을 교시한다. 일부 실시태양에서, 본 발명은 타입 V - 클래스 2 CRISPR 엔도뉴클레아제로 DNA를 분해하는 것을 교시한다. 일부 실시태양에서, 본 발명은 Cpf1 엔도뉴클레아제로 DNA를 분해하는 것을 교시한다.

[0009] 일부 실시태양에서, 본 발명은 CRISPR 다중 복제 부위를 포함하는 재조합 모듈식 CRISPR DNA 구조체를 교시하고, 상기 다중 복제 부위는 a) 적어도 2개의 별개의 복제 태그(cTAG)를 포함하며, 각 cTAG는 i) 각각이 프로토스페이서 인접 모티프(PAM)에 작동 가능하게 연결된 프로토스페이서 서열을 포함하는 하나 이상의 확증된 CRISPR 랜딩 부위를 포함하며; 상기 확증된 CRISPR 랜딩 부위 중 적어도 하나는 모듈식 CRISPR DNA 구조체 내에서 고유하다; 및 b) 하나 이상의 DNA 삽입체 서열을 포함한다; i) 상기 cTAG의 각각은 하나 이상의 DNA 삽입체 서열의 각각의 주위의 플랭킹 위치(flanking positions)에 분포되며; 및 ii) 상기 DNA 삽입체 서열의 적어도 하나는 선택 마커를 포함한다.

[0010] 일부 실시태양에서, 본 발명은 재조합 모듈식 CRISPR DNA 구조체를 교시하며, 상기 모듈식 CRISPR DNA 구조체는 원형이다.

[0011] 일부 실시태양에서, 본 발명은 재조합 모듈식 CRISPR DNA 구조체를 교시하며, 상기 모듈식 CRISPR DNA 구조체는 선형이다.

[0012] 일부 실시태양에서, 본 발명은 재조합 모듈식 CRISPR DNA 구조체를 교시하며, 상기 모듈식 CRISPR DNA 구조체는 유기체의 게놈에 통합된다.

[0013] 일부 실시태양에서, 본 발명은 재조합 모듈식 CRISPR DNA 구조체를 교시하며, 상기 별개의 cTAG의 적어도 하나

는 적어도 2개의 확증된 CRISPR 랜딩 부위를 포함한다.

- [0014] 일부 실시태양에서, 본 발명은 재조합 모듈식 CRISPR DNA 구조체를 교시하며, CRISPR 랜딩 부위의 적어도 하나는 Cas9 엔도뉴클레아제를 위한 것이다.
- [0015] 일부 실시태양에서, 본 발명은 재조합 모듈식 CRISPR DNA 구조체를 교시하며, CRISPR 랜딩 부위의 적어도 하나는 Cpf1 엔도뉴클레아제를 위한 것이다.
- [0016] 일부 실시태양에서, 본 발명은 재조합 모듈식 CRISPR DNA 구조체를 교시하며, 상기 별개의 cTAG의 적어도 하나는 희귀(≥ 8 염기 길이) 제한 엔도뉴클레아제 부위를 포함한다.
- [0017] 일부 양태에서, 본 발명은 재조합 모듈식 CRISPR DNA 구조체를 "메가모듈식(MegaModular)" 구조체로 지칭한다.
- [0018] 일부 실시태양에서, 본 발명은 재조합 핵산 분자를 제조하는 방법을 교시하며, 상기 방법은 a) 다음을 포함하는 혼합물을 형성하는 단계: i) 복수의 DNA 삽입체 부분, 여기서 각각의 DNA 삽입체 부분은 2개의 복제 태그(cTAGs)에 인접하고, 각각의 cTAG는 1) 각각이 프로토스페이서 인접 모티프(PAM)에 작동 가능하게 연결된 프로토스페이서 서열을 포함하는 하나 이상의 확증된 CRISPR 랜딩 부위를 포함한다; ii) 복수의 DNA 삽입체 부분의 적어도 2개에 존재하는 상기 cTAGs의 적어도 하나를 표적으로 하는 하나 이상의 CRISPR 복합체, 각각의 CRISPR 복합체는 1) CRISPR 엔도뉴클레아제, 및 2) 상기 표적화된 cTAG의 하나에 상기 CRISPR 엔도뉴클레아제를 모집할 수 있는 가이드 RNA 또는 가이드 RNAs를 포함하며; 여기서 혼합물은 오버행 말단을 생성하기 위해 복수의 DNA 삽입체 부분의 적어도 2개에서 표적화된 cTAG(s)의 분해를 허용하는 조건하에서 배양된다, 및 b) 양립 가능한 오버행 말단의 혼성화 및 혼성화된 말단의 공유 결합을 허용하는 조건에서 (a)에서 생성된 분해 생성물을 배양하는 단계를 포함하며; 생성된 재조합 핵산 분자는 상기 방법에서 결찰된 최초 삽입체 부분의 완전한 cTAG 서열을 포함한다.
- [0019] 일부 실시태양에서, 본 발명은 재조합 핵산 분자의 제조 방법을 교시하며, CRISPR 엔도뉴클레아제는 Cpf1이다.
- [0020] 일부 실시태양에서, 본 발명은 재조합 핵산 분자를 제조하는 방법을 교시하며, CRISPR 엔도뉴클레아제는 Cas9이다.
- [0021] 일부 실시태양에서, 본 발명은 재조합 핵산 분자를 제조하는 방법을 교시하며, 상기 방법은 i) 결찰 전에 CRISPR 복합체로부터 분해된 cTAG 서열을 분리하는 단계 또는 ii) 결찰 전에 CRISPR을 불활성화시키는 단계를 포함한다.
- [0022] 일부 실시태양에서, 본 발명은 재조합 핵산 분자를 제조하는 방법을 교시하며, 분리 단계는 DNA 정제 단계를 포함한다.
- [0023] 일부 실시태양에서, 본 발명은 재조합 핵산 분자의 제조 방법을 교시하며, 불활성화 단계는 상기 CRISPR 복합체의 열 또는 화학적 불활성화를 포함한다.
- [0024] 일부 실시태양에서, 본 발명은 재조합 핵산 분자를 제조하는 방법을 교시하며, 복수의 DNA 삽입체 부분의 각각에 대한 2개의 cTAG는 cTAG 쌍을 형성하고, 상기 cTAG 쌍은 상기 방법에서 결찰된 DNA 삽입체 부분의 모든 다른 cTAG 쌍으로부터 독특하다.
- [0025] 일부 실시태양에서, 본 발명은 재조합 핵산 분자를 제조하는 방법을 교시하며, 각 cTAG 쌍에서 cTAGs의 적어도 하나는 상이한 cTAG 쌍에서 적어도 하나의 다른 cTAG와 동일하다.
- [0026] 일부 실시태양에서, 본 발명은 DNA 서열 편집 방법을 교시하며 상기 방법은 a) 다음을 반응에 도입하는 단계: i) 본 발명의 모듈식 CRISPR DNA 구조체; ii) 대체 DNA 삽입체 부분, 상기 대체 DNA 삽입체 부분은 제 1 및 제 2 삽입체 cTAG가 측면에 위치되며; 1) 제 1 삽입체 cTAG는 모듈식 CRISPR DNA 구조체의 별개의 cTAGs의 하나의 확증된 CRISPR 랜딩 부위(들)를 포함하고, 제 2 삽입체 cTAG는 모듈식 CRISPR DNA 구조체의 다른 별개의 cTAG의 확증된 CRISPR 랜딩 부위(들)를 포함한다; 및 iii) 각각 제 1 및 제 2 삽입체 cTAG를 표적으로하는 제 1 및 제 2 CRISPR 복합체, 각각의 CRISPR 복합체는 1) CRISPR 엔도뉴클레아제 및 2) 상기 표적화된 삽입체 cTAGs의 하나에 상기 CRISPR 엔도뉴클레아제를 모집할 수 있는 가이드 RNA를 포함하며; 상기 제 1 및 제 2 CRISPR 복합체는 상기 제 1 및 제 2 삽입체 cTAG 및 상응하는 별개의 cTAG를 절단하여 오버행 말단을 생성한다, 및 b) 양립 가능한 오버행 말단의 혼성화 및 혼성화된 말단의 공유 결합을 허용하는 조건하에서 (a)에서 생성된 분해된 cTAGs로 대체 DNA 삽입체 부분 및 모듈식 CRISPR DNA 구조체를 배양하는 단계를 포함하며; 여기서 생성된 편집된 모듈식 CRISPR DNA 구조체는 상기 방법에서 결찰된 최초의 삽입체 부분의 완전한 cTAG 서열을 포함한다.

- [0027] 일부 실시태양에서, 본 발명은 DNA 서열 편집 방법을 교시하며, 단계(b)의 반응은 기능성 리가아제를 포함한다.
- [0028] 일부 실시태양에서, 본 발명은 DNA 서열 편집 방법을 교시하며, 상기 방법은 a) 다음을 반응에 도입하는 단계: i) 본 발명의 모듈식 CRISPR DNA 구조체; ii) 모듈식 CRISPR DNA 구조체에서 2개의 별개의 cTAG를 표적으로 하는 적어도 2개의 CRISPR 복합체로서, 각각의 CRISPR 복합체는 1) CRISPR 엔도뉴클레아제, 및 2) 상기 표적화된 별개의 cTAGs의 하나에 상기 CRISPR 엔도뉴클레아제를 모집할 수 있는 가이드 RNA를 포함하며; 제 1 및 제 2 CRISPR 복합체는 모듈식 CRISPR DNA 구조체에서 2개의 별개의 cTAG를 절단하고, 최종 별개의 cTAGs는 오버행 말단을 포함한다, 및 b) 다음을 제 2 반응에 도입하는 단계: i) (a)에서 생성된 분해된 cTAGs를 갖는 모듈식 CRISPR DNA 구조체; 및 ii) 대체 DNA 삽입체 부분, 상기 대체 DNA 삽입체 부분은 제 1 및 제 2 삽입체 cTAG가 측면에 위치되며; 1) 제 1 삽입체 cTAG는 (a)에서 절단된 분해되지 않은 별개의 cTAGs의 하나의 폴리뉴클레오타이드 서열을 포함하고, 제 2 삽입체 cTAG는 (a)에서 절단된 다른 분해되지 않은 별개의 cTAG의 폴리뉴클레오타이드 서열을 포함하고; 및 2) 상기 제 1 및 제 2 삽입체 cTAGs는 양립 가능한 오버행 말단의 혼성화 및 혼성화된 말단의 공유 결합을 허용하는 조건(a)로부터 별개의 cTAGs의 오버행 말단과 양립 가능한 오버행 말단을 포함하며; 여기서 생성된 편집된 모듈식 CRISPR DNA 구조체는 (a)에서 표적화된 최초의 분해되지 않은 별개의 cTAG의 완전한 서열을 포함한다.
- [0029] 일부 실시태양에서, 본 발명은 DNA 서열 편집 방법을 교시하며, 단계(b)의 반응은 기능적 리가아제를 포함한다.
- [0030] 일부 실시태양에서, 본 발명은 DNA 서열 편집 방법을 교시하며, CRISPR 엔도뉴클레아제가 Cpf1이다.
- [0031] 일부 실시태양에서, 본 발명은 DNA 서열 편집 방법을 교시하며, 여기서 단계(a)는 2개의 절단된 별개의 cTAG를 단일 가닥의 엑소뉴클레아제로 분해시켜 오버행 말단을 갖는 별개의 cTAG를 생성하는 단계를 추가로 포함한다. 일부 양태에서, 엑소뉴클레아제 단계 이후에 폴리머라제 및 리가아제로 접합부를 수리하기 위해 리가아제 및 폴리머라제를 첨가할 수 있다. 일부 양태에서, 이 반응은 또한 Cas9 분해되고, 평할 말단 절단으로 수행될 수 있다.
- [0032] 일부 실시태양에서, 본 발명은 DNA 서열 편집 방법을 교시하며, CRISPR 엔도 뉴클레아제는 Cas9이다.
- [0033] 일부 실시태양에서, 본 발명은 CRISPR 다중 복제 부위를 포함하는 재조합 모듈식 CRISPR DNA 구조체를 포함하는 숙주 세포 계통을 제공하며, 상기 다중 복제 부위는 a) 적어도 2개의 별개의 복제 태그(cTAG)를 포함하며, 각 cTAG는 다음을 포함한다: i) 각각이 프로토스페이서 인접 모티프(PAM)에 작동 가능하게 연결된 프로토스페이서 서열을 포함하는 하나 이상의 확증된 CRISPR 랜딩 부위; 상기 확증된 CRISPR 랜딩 부위의 적어도 하나는 모듈식 CRISPR DNA 구조체 내에서 독특한 것이다; 및 b) 하나 이상의 DNA 삽입체 부분(들); i) 상기 별개의 cTAGs의 각각은 하나 이상의 DNA 삽입체 부분(들)의 각각의 주위의 플랭킹 위치에 분포되어 있다.
- [0034] 일부 실시태양에서, 본 발명은 재조합 핵산 분자를 제조하는 방법을 제공하며, 상기 방법은 분해된 DNA 말단을 갖는 삽입체 부분을 생성하기 위해 적어도 2개의 복수의 DNA 삽입체 부분에서 하나 이상의 제한 효소에 의해 표적화된 cTAG의 분해를 허용하는 조건하에서 a) 다음을 포함하는 혼합물을 배양하는 단계: i) 2개의 복제 태그(cTAGs)가 측면에 위치되는 복수의 DNA 삽입체 부분, 각 cTAG는 1) 각각이 프로토스페이서 인접 모티프(PAM)에 작동 가능하게 연결된 프로토스페이서 서열을 포함하는 하나 이상의 확증된 CRISPR 랜딩 부위; 및 2) 회귀(≥ 8 개 염기) 제한 효소 인식 부위; 적어도 2개의 삽입체 부분의 cTAGs의 적어도 하나는 동일한 제한 효소 부위를 포함한다; ii) 복수의 DNA 삽입체 부분의 적어도 2개에서 회귀 제한 효소 부위를 표적으로 하는 하나 이상의 제한 효소; 및 b) 분해된 DNA 말단의 공유 결합을 허용하는 조건하에서 단계(a)에서 생성된 분해된 DNA 말단으로 DNA 삽입체 부분(들)을 배양하는 단계를 포함하며; 생성된 재조합 핵산 분자는 상기 방법에서 공유 결합된 최초의 삽입체 부분의 완전한 cTAG 서열을 포함한다.
- [0035] 일부 실시태양에서, 본 발명은 DNA 서열 편집 방법을 제공하며, 상기 방법은 a) 다음을 제공하는 단계: i) 본 발명의 모듈식 CRISPR DNA 구조체, 적어도 2개의 별개의 cTAGs는 회귀(≥ 8 개 염기) 제한 효소 인식 부위를 포함한다; ii) 대체 DNA 삽입체 부분, 상기 대체 DNA 삽입체 부분은 제 1 및 제 2 삽입체 cTAG가 측면에 위치된다; 1) 제 1 삽입체 cTAG는 모듈식 CRISPR DNA 구조체의 별개의 cTAG의 하나의 회귀 제한 효소 인식 부위를 포함하고, 제 2 삽입체 cTAG는 모듈식 CRISPR DNA 구조체의 다른 별개의 cTAG의 회귀 제한 효소 인식 부위를 포함한다; 및 iii) 제 1 및 제 2 삽입체 cTAG에서 회귀 제한 효소 부위를 표적으로 하는 하나 이상의 제한 효소; 부분 (i) 및 (ii)는 각각 (iii) 부분과 함께 단일 또는 별도의 반응으로 배양되며; 하나 이상의 제한 효소는 제 1 및 제 2 삽입체 cTAG의 회귀 제한 효소 인식 부위 및 이들의 상응하는 별개의 cTAG를 절단하여 분해된 DNA 말단을 생성시킨다, 및 b) 분해된 DNA 말단의 공유 결합을 허용하는 조건하에서 단계(a)에서 생성된 분

해된 DNA 말단으로 대체 DNA 삽입체 부분 및 모듈식 CRISPR DNA 구조체를 배양하는 단계를 포함하며; 생성된 편집된 모듈식 CRISPR DNA 구조체는 상기 방법에서 공유 결합된 최초의 삽입체 부분의 완전한 cTAG 서열을 포함한다.

[0036] 일부 실시태양에서, 본 발명은 재조합 핵산 분자를 제조하는 방법을 제공하며, 상기 방법은 적어도 2개의 DNA 삽입체 부분에서 공유된 cTAG의 분해를 허용하여, 적어도 2개의 DNA 삽입체 부분에서 양립 가능한 오버행 DNA 말단을 생성하는 조건하에서 a) 다음을 포함하는 혼합물을 배양하는 단계: i) 각각이 2개의 복제 태그(cTAGs)가 측면에 위치되는 복수의 DNA 삽입체 부분, 각 cTAG는 1) 각각이 프로토스페이서 인접 모티프(PAM)에 작동 가능하게 연결된 프로토스페이서 서열을 포함하는 하나 이상의 확증된 CRISPR 랜딩 부위; DNA 삽입체 부분의 적어도 2개는 동일한 cTAG를 공유한다; ii) 단일 가닥 DNA(ssDNA) 엑소뉴클레아제, 및 b) 적어도 2개의 DNA 삽입체 부분의 양립 가능한 오버행 DNA 말단의 혼성화 및 공유 결합을 허용하는 조건하에서 (a)에서 생성된 분해된 cTAG로 DNA 삽입체 부분을 배양하는 단계를 포함하며; 생성된 재조합 핵산 분자는 분해 전 공유된 cTAG의 완전한 cTAG 서열을 포함한다. 이 반응은 또한 접합을 고정하는데 사용되는 폴리머라제 및 또는 리가아제로 수행될 수 있다. 또한, 이것은 미리 분해된 벡터로 수행될 수 있다.

발명의 효과

[0037] 본 발명의 내용 중에 포함되어 있다.

도면의 간단한 설명

[0038] 도 1a-c는 본 발명의 CRISPR/Cas9 및 CRISPR/Cpf1 시스템의 비교를 도시한다. 도 1a - Cas9 엔도뉴클레아제는 tracrRNA 및 crRNA 복합체에 의해 표적 dsDNA에 모집된다. 도 1b - Cas9 엔도뉴클레아제는 인위적으로 융합된 tracrRNA 및 단일 가이드 RNA(sgRNA)로 알려진 crRNA 서열에 의해 표적 dsDNA에 모집될 수 있다. Cas9 엔도뉴클레아제는 평활 말단을 생성한다. 도 1c - Cpf1 엔도뉴클레아제는 crRNA 가이드 폴리리보뉴클레오티드만을 필요로 한다. Cpf1 엔도뉴클레아제 절단은 5'오버행을 갖는 이중 가닥 절단을 생산한다.

도 2a-2c는 본 발명의 모듈식 CRISPR 구조체를 이용하는 본 발명의 복제 방법의 한 실시태양을 도시한다. 도 2a는 본 발명에 따라 Cas9 또는 Cpf1 뉴클레아제로 쉽게 변형될 수 있는 모듈식 CRISPR 플라스미드를 나타낸다. 전술한 바와 같이, 본 발명의 모듈식 CRISPR 구조체는 "MegaModular" 구조체라 할 수 있다. 숫자로 표시되는 상호 교환 가능한 부분은 문자로 표시되는 불변 cTAG 서열이 측면에 위치된다. 부분은 사전 조립되거나 cTAG 서열 동일성을 기반으로 생체외에서 조립될 수 있다. 삽입체 부분의 예는 도 2a의 오른쪽에 도시된다. 도 2b - cTAG에서의 Cas9, Cpf1 또는 제한 엔도뉴클레아제 절단과 같은 여러 전략을 사용하여 전체 플라스미드를 재구성하지 않고도 개별 부분을 대체할 수 있다. cTAG 서열은 Cas9, Cpf1, 제한 및/또는 재조합 부위를 포함하나 이에 제한되지 않는 하나 이상의 복제 부위를 포함할 수 있다. 도 2c - 일단 유기체의 게놈에 통합되면, cTAG는 사전 확증된 Cas9 또는 Cpf1 랜딩 부위로 계속 사용될 수 있어, 사전 검증된 직교 gRNA 염기 서열을 가진 게놈적으로 통합된 DNA의 대체, 삽입 또는 제거를 가능하게 한다.

도 3a-d는 본 발명의 모듈식 CRISPR 구조체를 이용하는 본 발명의 복제 방법의 한 실시태양을 도시한다. 도 3a는 본 발명에 따라, Cas9 또는 Cpf1 뉴클레아제로 쉽게 변형될 수 있는 모듈식 CRISPR 플라스미드를 나타낸다. 숫자로 표시되는 상호 교환 가능한 부분은 문자로 표시되는 불변 cTAG 서열에 의해 인접된다. 부분은 사전 조립되거나 cTAG 서열 동일성을 기반으로 생체내 또는 생체외에서 조립될 수 있다. 예시적 삽입체 부분은 도 3a의 오른쪽에 도시된다. 도 3b - cTAG에서의 Cas9, Cpf1 또는 제한 엔도뉴클레아제 절단과 같은 여러 전략을 사용하여 전체 플라스미드를 재구성하지 않고도 개별 부분을 대체할 수 있다. cTAG 서열은 Cas9, Cpf1, 제한 및/또는 재조합 부위를 포함하나 이에 제한되지 않는 하나 이상의 복제 부위를 포함할 수 있다. 도 3c는 삽입cp 부분을 제거하거나 기존의 모듈식 플라스미드로부터 스테퍼 서열을 추가하기 위한 본 발명의 방법을 예시한다. 도 3d - 본 발명의 모듈식 플라스미드의 삽입체 부분은 숙주 세포의 게놈 속으로 모듈식 CRISPR 벡터의 일부 또는 전부의 게놈 통합을 위한 서열로서 작용할 수 있다.

도 4는 실시예 1의 원-포트(one-pot) 생체의 모듈식 CRISPR 복제를 도시한다. 구체적으로, 원-포트 반응에서 하나의 플라스미드로부터 다른 플라스미드로 인서트를 옮김으로써 플라스미드 13001009086의 생성이 도시된다. 이 반응의 세부 사항은 실시예 1에 기재되어 있다.

도 5는 실시예 2의 생체외 모듈식 CRISPR 복제 방법의 실시태양을 예시한다. 각 패널은 실시예 2에 기술된 실험 디자인의 설명을 제공한다. 클로람페니콜 내성 유전자는 카나마이신 내성 백본 플라스미드에 복제시켜 이중 내

성 플라스미드를 생성하였다. 그런 후에 이중 내성 플라스미드는 박테리아로 형질전환된 이후, 카나마이신 및 클로람페니콜 항생제로 강화된 배지에서 배양하였다. 내성 콜로니는 성공적인 Cpf1 복제 어셈블리를 나타내었다.

도 6은 실시예 2의 생체의 모듈식 CRISPR 복제 방법의 결과를 도시한다. y-축은 카나마이신 및 클로람페니콜로 강화된 배지에서 성장하는 회수된 콜로니 수를 나타낸다. 내성 콜로니는 성공적인 Cpf1 복제 어셈블리를 나타낸다. 결과는 이중 내성 플라스미드의 리가아제 의존성 어셈블리가 나타났다.

도 7은 pJDI427에 대한 벡터 지도를 묘사한다. Cpf1 어셈블리에 사용된 CRISPR 랜딩 부위는 cTAG M 및 cTAG N 으로 레이블된다. 관련 서열 정보는 SEQ ID NO: 102에서 찾을 수 있다.

도 8은 pJDI429에 대한 벡터 지도를 묘사한다. Cpf1 어셈블리에 사용된 CRISPR 랜딩 부위는 cTAG N 및 cTAG O 로 레이블된다. 관련 서열 정보는 SEQ ID NO: 103에서 찾을 수 있다.

도 9는 pJDI430에 대한 벡터 지도를 묘사한다. Cpf1 어셈블리에 사용된 CRISPR 랜딩 부위는 cTAG P 및 cTAG N 으로 레이블된다. 관련 서열 정보는 SEQ ID NO: 104에서 찾을 수 있다.

도 10은 pJDI431에 대한 벡터 지도를 묘사한다. Cpf1 어셈블리에 사용된 CRISPR 랜딩 부위는 cTAG P 및 cTAG O 로 레이블된다. 관련 서열 정보는 SEQ ID NO: 105에서 찾을 수 있다.

도 11은 pJDI432에 대한 벡터 지도를 묘사한다. Cpf1 어셈블리에 사용된 CRISPR 랜딩 부위는 cTAG M 및 cTAG N 으로 레이블된다. 관련 서열 정보는 SEQ ID NO: 106에서 찾을 수 있다.

도 12는 pJDI434에 대한 벡터 지도를 묘사한다. Cpf1C 어셈블리에 사용된 CRISPR 랜딩 부위는 cTAG N 및 cTAG O 로 레이블된다. 관련 서열 정보는 SEQ ID NO: 107에서 찾을 수 있다.

도 13은 pJDI435에 대한 벡터 지도를 묘사한다. Cpf1 어셈블리에 사용된 CRISPR 랜딩 부위는 cTAG P 및 cTAG N 으로 레이블된다. 관련 서열 정보는 SEQ ID NO: 108에서 찾을 수 있다.

도 14는 pJDI436에 대한 벡터 지도를 묘사한다. Cpf1 어셈블리에 사용된 CRISPR 랜딩 부위는 cTAG P 및 cTAG O 로 레이블된다. 관련 서열 정보는 SEQ ID NO: 109에서 찾을 수 있다.

도 15는 본 발명의 방법에 따라 모듈식 CRISPR 구조체의 예시적인 유전자 편집을 도시한다. 구체적으로, 도 15는 실시예 3의 메가모듈식 디자인을 사용하여 제한 효소 분해 및 결합에 의한 플라스미드 어셈블리를 도시한다. 도 15는 모듈식 CRISPR 플라스미드 백본 p1300283391 및 양립 가능한 GFP-함유 삽입체 DNA 부분이 각각 ApaI 및 PvuI 제한 효소로 분해 되어 양립 가능한 복제 태그 말단을 생성하는 것을 나타낸다. 분해된 백본과 인서트는 생체외에서 결합되어 새로운 모듈식 CRISPR 구조체를 생성한다.

발명을 실시하기 위한 구체적인 내용

정의

다음의 용어는 당해 기술 분야에서 통상의 지식을 가진 자에 의해 잘 이해되는 것으로 생각되지만, 다음 정의는 본 발명에 개시된 주제의 설명을 용이하게 하기 위해 제시된다.

용어 하나("a" 또는 "an")는 그 실체 중 하나 이상을 의미하며, 즉 복수의 지시대상을 의미할 수 있다. 이와 같이, 용어 "하나", "하나 이상" 및 "적어도 하나"라는 본 발명에서 상호 교환적으로 사용된다. 또한, 부정관사에 의한 "한 요소"에 대한 언급은, 내용이 분명하게는 요소의 하나 및 단지 하나가 존재하는 것을 요구하지 않는 한, 하나 이상의 요소가 존재하는 가능성을 배제하지 않는다.

용어 "원핵 생물"은 당업계에서 공지되어 있으며 핵을 함유하지 않는 세포를 의미한다. 원핵생물은 일반적으로 두 영역, 박테리아와 고세균의 하나로 분류된다. 고세균과 박테리아 영역의 유기체 사이의 명확한 차이는 16S 리보솜 RNA에서 뉴클레오타이드 염기 서열의 근본적인 차이에 기초한다.

"진핵 생물"은 세포가 막 내에 둘러싸인 핵 및 다른 소기관을 함유하는 임의의 유기체이다. 진핵 생물은 분류군(Eukarya 또는 Eukaryota)에 속한다. 진핵 생물 세포를 원핵 생물(앞에서 언급한 박테리아와 고세균)과 구분시키는 정의하는 특징은 막으로 둘러싸인 유전자 물질, 특히 유전자 물질을 함유하고 핵막에 의해 둘러싸인 핵을 가진다는 것이다.

용어 "고세균"은 전형적으로 특이한 환경에서 발견되고 세포벽에서 리보솜 단백질의 수 및 뮤라민산의 부족을 포

함하는 몇몇 기준에 의해 나머지 원핵 생물과 구별되는 멘도시쿠테스(Mendosicutes) 문의 유기체의 범주를 의미한다. ssrRNA 분석에 기초하여, 고세균은 두 계통발생학적으로 다른 그룹으로 구성된다: 크렌고세균(Crenarchaeota) 및 유리고세균(Euryarchaeota). 생리학을 기초로, 고세균은 세 가지 유형으로 구성될 수 있다: 메테인 생성균(methanogens)(메테인을 생성하는 원핵 생물); 고염성 세균(extreme halophiles)(매우 높은 농도의 염(NaCl)에서 사는 원핵 생물; 및 고온성(초고온성) 세균(extreme(hyper) thermophilus)(초고온에서 사는 원핵 생물). 박테리아와 구별되는 통일된 고세균의 특징(즉, 세포벽, 에스터-연결 막 지질 등에 류테인 없음)이 외에, 이런 원핵 생물은 이들의 특정한 서식 환경에 적응시키는 특이한 구조 또는 생화학적 특성을 나타낸다. 크렌고세균은 주로 초고온성 황 의존성 원핵 생물로 이루어지며 유리고세균은 메테인 생성균과 고염성 세균을 함유한다.

[0045] "박테리아" 또는 "진정세균(eubacteria)"는 원핵 생물의 영역을 의미한다. 박테리아는 다음과 같이 적어도 11개의 구별된 그룹을 포함한다: (1) 그람 양성 (그람+) 박테리아, 2개의 주요 세부구분이 존재한다: (1) 높은 G+C 그룹(액티노마이세테스, 마이코박테리아, 마이크로코코커스, 기타) (2) 낮은 G+C 그룹(바실러스, 클로스트리디아, 락토바실러스, 스태필로코키, 스트렙토코키, 마이코플라스마스); (2) 프로테오박테리아, 예를 들어, 보라색 광합성 + 비 광합성 그람 음성 박테리아 (대부분의 "일반적인" 그람 음성 박테리아 포함); (3) 사이아노박테리아, 예를 들면, 산소성 광영양생물; (4) 스피로체테스 및 관련 종; (5) 플랑크토마이세스; (6) 박테로이데스, 플라보박테리아; (7) 클라미디아; (8) 녹색 황 박테리아; (9) 녹색 비 황 박테리아(또한 혐기성 광영양생물); (10) 방사성 저항성 마이크로코키 및 동족; (11) 써모토가(Thermotoga) 및 써모시포 써모필레스(Thermosiphon thermophiles).

[0046] 용어 "유전자 변형된 숙주 세포", "재조합 숙주 세포" 및 "재조합 균주"는 본 발명에서 상호 교환적으로 사용되고 본 발명의 클로닝 및 형질전환 방법에 의해 유전자 변형된 숙주 세포를 의미한다. 따라서, 이 용어는 유전자 변형, 변형 또는 조작되어, 숙주 세포가 유도된 자연 발생 유기체와 비교하여 변형, 변형 또는 상이한 유전자형 및/또는 표현형을 나타내는(예를 들어, 유전자 변형이 미생물의 핵산 서열 암호화에 영향을 미칠 때) 숙주 세포(예를 들어, 박테리아, 효모 세포, 균류 세포, CHO, 인간 세포 등)를 포함한다. 이 용어는 문제의 특정 재조합 미생물뿐만 아니라 이런 미생물의 자손 또는 잠재적 자손을 의미하는 것으로 이해된다.

[0047] 용어 "유전자 조작된"은 (예를 들어, 핵산의 삽입 또는 결실에 의한) 숙주 세포의 게놈의 임의적 조작을 의미할 수 있다.

[0048] 본 발명에서 사용된 "선택 가능한 마커"는 흔히 특정 조건하에서 분자(예를 들어, 레플리콘) 또는 이를 함유하는 세포를 선택할 수 있게 하는 핵산 분절이다. 이들 마커는 RNA, 펩타이드 또는 단백질의 생산과 같으나 이에 제한되지 않는 활성을 암호화할 수 있거나 RNA, 펩타이드, 단백질, 무기 및 유기 화합물 또는 조성물 등을 위한 결합 부위를 제공할 수 있다. 선택 가능한 마커의 예는 (1) 독성 화합물에 대한 내성을 제공하는 생산물(예를 들어, 항생제)을 암호화하는 핵산 분절; (2) 수용체 세포에서 결핍된 생산물(예를 들어, tRNA 유전자, 영양요구성 마커)을 암호화하는 핵산 분절; (3) 유전자 생산물의 활성을 억제하는 생산물을 암호화하는 핵산 분절; (4) 쉽게 식별될 수 있는 생산물(예를 들어, β -갈락토시다제, 녹색 형광 단백질(GFP), 황색 형광 단백질(YFP), 시안 형광 단백질(CFP) 및 세포 표면 단백질과 같은 표현형 마커)을 암호화하는 핵산 분절; (5) 세포 생존 및/또는 기능에 해를 끼치는 다른 생산물과 결합하는 생산물을 암호화하는 핵산 분절; (6) 가시적인 또는 선택 가능한 표현형을 생산하는 임의의 핵산 단편의 활성을 억제하는 핵산(예를 들어, 안티센스 올리고뉴클레오타이드)을 암호화하는 핵산 분절; (7) 기질을 변형시키는 다른 생성물과 결합하는 생성물(예를 들어, 제한 엔도뉴클레아제)을 암호화하는 핵산 분절; (8) 원하는 분자(예를 들어, 특정 단백질 결합 부위)를 분리 또는 동정하는 데 사용될 수 있는 핵산 분절; (9) 비 기능적일 수 있는 특정 뉴클레오타이드 서열(예를 들어, 분자의 아집단의 PCR 증폭용)을 암호화하는 핵산 분절; (10) 부재시에는 특정 화합물에 대해 직접 또는 간접적으로 내성 또는 민감성을 부여하는 핵산 분절을 포함하나 이에 제한되지 않는다.

[0049] 본 발명에서 사용된 "역선택 가능한 마커" 또는 "역선택 마커"는 선택시 숙주 유기체의 성장을 제거 또는 억제하는 핵산 분절이다. 일부 실시태양에서, 본 발명의 역선택 가능한 마커는 세포를 하나 이상의 화학 물질/성장 조건/유전적 배경에 민감하게 만든다. 일부 실시태양에서, 본 발명의 역선택 가능한 마커는 독성 유전자이다. 일부 실시태양에서, 역선택 가능한 마커는 유도성 프로모터에 의해 발현된다.

[0050] 본 발명에 사용된 용어 "핵산"은 임의의 길이의 중합체 형태의 리보뉴클레오타이드 또는 데옥시리보뉴클레오타이드 또는 이의 유사체를 의미한다. 이 용어는 분자의 1 차 구조를 의미하며, 따라서 이중 나선 및 단일 가닥의 DNA뿐 아니라 이중 및 단일 가닥의 RNA를 포함한다. 또한, 메틸화 및/또는 캡핑된 핵산과 같은 변형된 핵산, 변

형된 염기를 함유하는 핵산, 골격 변형 등과 같은 변형 핵산을 포함한다. 용어 "핵산" 및 "뉴클레오타이드 서열"은 상호 교환적으로 사용된다.

[0051] 본 발명에 사용된 용어 "유전자"는 생물학적 기능과 관련된 DNA의 임의의 분절을 의미한다. 따라서, 유전자는 암호화 서열 및/또는 그의 발현에 요구되는 조절 서열을 포함하나 이에 제한되지 않는다. 유전자는 또한, 예를 들어, 다른 단백질에 대한 인식 서열을 형성하는 비 발현 DNA 분절을 포함할 수 있다. 유전자는 관심 공급원으로부터의 클로닝 또는 공지되거나 예측된 서열 정보로부터의 합성을 포함하는 다양한 공급원으로부터 얻을 수 있고, 원하는 파라미터를 갖도록 디자인된 서열을 포함할 수 있다.

[0052] 본 발명에 사용된 용어 "상동성" 또는 "동족체" 또는 "오르쏘로그(ortholog)"는 당업계에 공지되어 있고 공통 조상 또는 가족 구성원을 공유하고 서열 동일성의 정도에 기초하여 결정되는 관련 서열을 의미한다. 용어 "상동 관계", "상동성", "실질적으로 유사" 및 "상응하게 실질적으로"는 본 발명에서 상호 교환적으로 사용된다. 이들은 하나 이상의 뉴클레오타이드 염기의 변화가 유전자 발현을 중재하거나 특정 표현형을 생성시키는 핵산 단편의 능력에 영향을 미치지 않는 핵산 단편을 의미한다. 이런 용어는 또한 초기의 변형되지 않은 단편에 비해 생성된 핵산 단편의 기능적 특성을 실질적으로 변화시키지 않는 하나 이상의 뉴클레오타이드의 결실 또는 삽입과 같은 본 발명의 핵산 단편의 변형을 의미한다. 따라서, 당업자가 알 수 있는 바와 같이, 본 발명은 특정 예시적인 서열 이상을 포함하는 것으로 이해된다. 이런 용어는 한 종, 아종, 품종, 품종 또는 군주에서 발견된 유전자 및 다른 종, 아종, 품종, 품종 또는 군주에서 상응하는 또는 동등한 유전자 사이의 관계를 기술한다. 본 발명을 위해서, 상동성 서열이 비교된다. "상동 서열" 또는 "상동체" 또는 "오르쏘로그"는 기능적으로 관련이 있다고 생각되거나 믿거나 알려진다. 기능적 관계는 (a) 서열 동일성 및/또는 (b) 동일하거나 유사한 생물학적 기능을 포함하나 이에 제한되지 않는 다수의 방식 중 임의의 하나로 표시될 수 있다. 바람직하게는, (a) 및 (b) 모두가 표시된다. 상동성은 Current Protocols in Molecular Biology(F.M. Ausubel et al., eds., 1987) Supplement 30, 섹션 7.718, 표 7.71에서 논의된 바와 같은 당해분야에서 용이하게 이용 가능한 소프트웨어 프로그램을 사용하여 결정될 수 있다. 일부 정렬 프로그램은 맥벡터(MacVector)(Oxford Molecular Ltd, Oxford, U.K.), ALIGN 플러스(Plus)(Scientific and Educational Software, Pennsylvania) 및 AlignX(Vector NTI, Invitrogen, Carlsbad, CA)이다. 다른 정렬 프로그램은 기본 매개변수를 사용하여 시퀀처(Sequencher)(Gene Codes, Ann Arbor, Michigan)이다.

[0053] 본 발명에 사용된 용어 "뉴클레오타이드 변화"는 당업계에서 잘 알려진 바와 같이, 예를 들어 뉴클레오타이드 치환, 결실 및/또는 삽입을 의미한다. 예를 들어 돌연변이는 침묵 치환, 추가 또는 결실을 생성하나 암호화된 단백질의 특성 또는 활성 또는 단백질이 어떻게 만들어지는지를 변형하지 않는 변형을 함유한다.

[0054] 본 발명에 사용된 용어 "단백질 변형"은, 예를 들어, 당업계에서 잘 이해되는 바와 같이, 아미노산 치환, 아미노산 변형, 결실 및/또는 삽입을 의미한다.

[0055] 본 발명에 사용된 용어 핵산 또는 폴리펩타이드의 "적어도 일부" 또는 "단편"은 전장 분자를 포함하는 전장 분자의 이런 서열 또는 임의의 더 큰 단편의 최소 크기 특성을 갖는 부분을 의미한다. 본 발명의 폴리뉴클레오타이드 단편은 유전자 조절 요소의 생물학적 활성 부분을 암호화할 수 있다. 유전자 조절 요소의 생물학적 활성 부분은 유전자 조절 요소를 포함하는 본 발명의 폴리뉴클레오타이드 중 하나의 일부를 분리하고 본 발명에 기재된 바와 같은 활성을 평가함으로써 제조될 수 있다. 유사하게, 폴리펩타이드의 일부는 전장 폴리펩타이드까지 이르는 4개 아미노산, 5개 아미노산, 6개 아미노산, 7개 아미노산 등일 수 있다. 사용될 부분의 길이는 특정 용도에 따라 다를 것이다. 하이브리드화 프로브로서 유용한 핵산의 일부는 12개 뉴클레오타이드 정도로 짧을 수 있으며; 일부 실시태양에서, 이것은 20개 뉴클레오타이드이다. 에피토프로서 유용한 폴리펩타이드의 일부는 4개의 아미노산 정도로 짧을 수 있다. 전장 폴리펩타이드의 기능을 수행하는 폴리펩타이드의 일부는 일반적으로 4개 이상의 아미노산보다 길 수 있다.

[0056] 본 발명에 개시된 폴리뉴클레오타이드의 PCR 증폭을 위해, 임의의 관심 유기체로부터 추출된 cDNA 또는 게놈 DNA로부터의 상응하는 DNA 서열을 증폭시키기 위한 PCR 반응에 사용하기 위해 올리고뉴클레오타이드 프라이머가 디자인될 수 있다. PCR 프라이머 및 PCR 클로닝을 디자인하기 위한 방법은 당업계에 일반적으로 공지되어 있으며, Sambrook *et al.*(2001) Molecular Cloning: A Laboratory Manual (3rd ed., Cold Spring Harbor Laboratory Press, Plainview, New York). See also Innis *et al.*, eds. (1990) PCR Protocols: A Guide to Methods and Applications (Academic Press, New York); Innis and Gelfand, eds. (1995) PCR Strategies (Academic Press, New York); and Innis and Gelfand, eds. (1999) PCR Methods Manual (Academic Press, New York)에 개시된다. PCR의 공지된 방법은 쌍을 이룬 프라이머, 네스티드 프라이머, 단일 특이적 프라이머, 축퇴

성 프라이머, 유전자 특이적 프라이머, 벡터 특이적 프라이머, 부분적 불일치 프라이머 등을 사용하는 방법을 포함하나 이에 제한되지 않는다.

[0057] 본 발명에 사용된 용어 "프라이머"는 DNA 중합효소를 부착시키는 증폭 표적에 대한 어닐링을 행할 수 있어, 프라이머 연장 생성물의 합성이 유도되는 조건하에 놓일 때, 즉, 뉴클레오타이드 및 DNA 중합효소와 같은 중합화제의 존재하에서 및 적절한 온도 및 pH에서 DNA 합성의 개시점으로서 작용하는 올리고뉴클레오타이드를 의미한다. (증폭) 프라이머는 증폭 효율을 최대화하기 위해 바람직하게는 단일 가닥이다. 바람직하게는, 프라이머는 올리고데옥시리보뉴클레오타이드이다. 프라이머는 중합화제의 존재 하에서 증량 생성물의 합성을 시작하기에 충분히 길어야 한다. 프라이머의 정확한 길이는 프라이머의 온도 및 조성(A/T 대 G/C 함량)을 포함하는 많은 요소에 따라 달라질 것이다. 한 쌍의 양방향성 프라이머는 PCR 증폭과 같은 DNA 증폭 기술 분야에서 일반적으로 사용되는 것과 같은 하나의 순방향 및 역방향 프라이머로 구성된다.

[0058] 용어 "엄격성" 또는 "엄격한 혼성화 조건"은 하이브리드의 안정성에 영향을 주는 혼성화 조건, 예를 들어 온도, 염 농도, pH, 포름아마이드 농도 등을 의미할 수 있다. 이러한 조건은 특이적 결합을 최대화하고 표적 핵산 서열에 대한 프라이머 또는 프로브의 비특이적 결합을 최소화하기 위해 경험적으로 최적화된다. 사용된 용어는 프로브 또는 프라이머가 다른 서열보다 검출 가능하게 더 큰 정도로 (예를 들어, 배경보다 적어도 2배) 이의 표적 서열에 혼성화될 조건에 대한 언급을 포함한다. 엄격한 조건은 서열 의존성이며 다른 상황에서 상이할 것이다. 긴 서열은 고온에서 특이적으로 혼성화된다. 일반적으로, 엄격한 조건은 정의된 이온 강도 및 pH에서 특정 서열에 대한 열 융점(T_m)보다 약 5°C 낮은 것으로 선택된다. T_m 은 상보적 표적 서열의 50%가 완벽하게 매치된 프로브 또는 프라이머에 하이브리드화되는 온도(정의된 이온 강도 및 pH 하에서)이다. 전형적으로, 엄격한 조건은 염 농도가 pH 7.0 내지 8.3에서 약 1.0 M Na^+ 이온 미만, 전형적으로는 약 0.01 내지 1.0 M Na^+ 이온 농도(또는 다른 염)이고 온도는 짧은 프로브 또는 프라이머(예를 들어, 10 내지 50개 뉴클레오타이드)의 경우 적어도 약 30°C 이고 긴 프로브 또는 프라이머(예를 들어, 50개 초과 뉴클레오타이드)의 경우 적어도 약 60°C 이다. 엄격한 조건은 또한 포름아마이드와 같은 불안정화제의 첨가로 성취될 수 있다. 예시적인 낮은 엄격한 조건 또는 "감소된 엄격한 조건"은 37°C 에서 30% 포름아마이드, 1M NaCl , 1% SDS의 완충액에 의한 하이브리드화 및 40°C 에서 2 x SSC의 세척을 포함한다. 예시적인 높은 엄격 조건은 37°C 에서 50% 포름아마이드, 1M NaCl , 1% SDS에서 하이브리드화 및 60°C 에서 0.1 x SSC 세척을 포함한다. 하이브리드화 절차는 당업계에 주지되어 있으며 예를 들어, Ausubel *et al.*, 1998 and Sambrook *et al.*, 2001에 의해 기술된다. 일부 실시태양에서, 엄격한 조건은 $1\text{mM Na}_2\text{EDTA}$, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19% 또는 20%와 같은 45°C 의 0.5-20% 황산 도데실 나트륨을 함유하는 $0.25\text{ M Na}_2\text{HPO}_4$ 완충액(pH 7.2)에서의 하이브리드화, 및 뒤이은 55°C 내지 65°C 에서 0.1%(w/v) 황산 도데실 나트륨을 함유하는 5 x SSC 세척이다.

[0059] 본 발명에 사용된 용어 "실질적으로 동일한"은 1개, 2개, 3개, 4개, 5개, 6개 또는 7개 이하의 뉴클레오타이드가 변하는 2개의 폴리뉴클레오타이드 서열을 의미한다. cTAGs의 문맥에서 사용되는 경우, 실질적으로 동일한 용어는 적어도 하나의 CRISPR 랜딩 부위에서 CRISPR 절단을 폐기하기 위해 디자인된 cTAGs의 하나의 PAM 또는 프로토스페이스 영역의 돌연변이를 제외하고는 동일할 2개의 cTAG를 나타낸다. 용어 "부분적" 서열 또는 cTAG와 함께 용어 실질적으로 동일한이 사용되는 경우, 상기 조합은 상기한 바와 같이 2개의 실질적으로 동일한 cTAG 사이의 비교를 의미하며, cTAG의 하나가 CRISPR 엔도뉴클레아제로 분해되었다. 따라서 이 용어는 기술된 cTAG가 PAM 또는 프로토스페이스 영역의 돌연변이를 제외하고 (분해되지 않은 형태의) 제 2 cTAG와 동일하다는 것을 나타내는 데 사용될 수 있다.

[0060] 본 발명에 사용된 용어 "프로모터"는 암호화 서열 또는 기능성 RNA의 발현을 제어할 수 있는 DNA 서열을 의미한다. 프로모터 서열은 근위 및 더 원위 상류 요소로 구성될 수 있고, 후자의 요소는 종종 인핸서로 의미된다. 따라서, "인핸서"는 프로모터 활성을 자극할 수 있는 DNA 서열이며, 프로모터의 선천적인 요소 또는 프로모터의 수준 또는 조직 특이성을 향상시키기 위해 삽입된 이종성 요소일 수 있다.

[0061] 본 발명에 사용된 용어 "이종성"은 특정 유기체에서 자연적으로 발견되지 않는 핵산 서열을 의미한다.

[0062] 본 발명에 사용된 용어 "내인성", "내인성 유전자"는 유전자의 자연 발생 복사물을 의미한다.

[0063] 본 발명에 사용된 용어 "자연적으로 발생하는"은 자연적으로 발생하는 원료로부터 유래된 유전자 또는 서열을 의미한다. 일부 실시태양에서, 자연적으로 발생하는 유전자는, 상이한 유기체에 도입될 때, 원료 유기체 내의 내인성 설정에 위치되거나 "이종성" 설정에 놓인 경우, 야생형(비-형질전환 유전자) 유전자의 유전자를 의미한다. 따라서, 본 발명의 목적 상, "자연적으로 발생하지 않은" 서열은 공지된 천연 서열과 상이한 서열을 갖도록 합성되고, 돌연변이되고 또는 변형된 서열이다. 일부 실시태양에서, 변형은 단백질 수준일 수 있다(예를 들어,

아미노산 치환). 다른 실시태양에서, 변형은 단백질 서열에 영향을 주지 않으면서 DNA 수준일 수 있다(예를 들어, 코돈 최적화). 일부 실시태양에서, 자연적으로 발생하지 않는 서열은 구조체일 수 있다.

[0064] 본 발명에 사용된 용어 "외인성"은 용어 "이종성(heterologous)"과 상호 교환적으로 사용되고, 자연 공급원 이외의 일부 공급원으로부터 유도하는 물질을 의미한다. 예를 들어, 용어 "외인성 단백질" 또는 "외인성 유전자"는 비 자연 공급원 또는 위치의 단백질 또는 유전자 및 생물학적 시스템에 공급된 단백질 또는 유전자를 의미한다. 내인성 유전자의 인공적으로 돌연변이된 변이체는 본 발명의 목적 상 "외인성"으로 고려된다.

[0065] 본 발명에서 사용된 바와 같이, 용어 "재조합 구조체", "발현 구조체", "키메라 구조체", "구조체" 및 "재조합 DNA 구조체"는 본 발명에서 상호 교환적으로 사용된다. 재조합 구조체는 천연에서 함께 발견되지 않는 조절 및 암호화 서열과 같은 핵산 단편의 인위적인 조합을 포함한다. 예를 들어, 키메라 구조체는 상이한 공급원으로부터 유도된 조절 서열 및 암호화 서열, 또는 동일한 공급원으로부터 유도되지만 자연계에서 발견되는 것과 상이한 방식으로 배열된 조절 서열 및 암호화 서열을 포함할 수 있다. 이러한 구조체는 그 자체로 사용되거나 벡터와 함께 사용될 수 있다. 벡터가 사용되는 경우 벡터의 선택은 당업자에게 주지된 바와 같이 숙주 세포를 형질 전환하는 데 사용될 방법에 의존한다. 예를 들어, 플라스미드 벡터가 사용될 수 있다. 당업자는 본 발명의 분리된 핵산 단편을 포함하는 숙주 세포를 성공적으로 형질전환, 선별 및 증식시키기 위해 벡터 상에 존재해야 하는 유전자 요소를 잘 알고 있다. 당업자는 또한 상이한 독립적인 형질전환 사건이 발현의 상이한 수준 및 패턴을 유도한다는 것을 인식할 것이며(Jones *et al.*, (1985) EMBO J. 4 : 2411-2418; De Almeida *et al.*, (1989) Mol. Gen. Genetics 218 : 78-86), 따라서 다수의 사건은 원하는 발현 수준 및 패턴을 나타내는 라인을 수득하기 위해 선별되어야 한다. 이러한 선별은 다른 것들 중에서, DNA의 서던 분석, mRNA 발현의 노던 분석, 단백질 발현의 면역 블로팅 분석 또는 표현형 분석에 의해 실행될 수 있다. 벡터는 자율적으로 복제하거나 숙주 세포의 염색체에 통합될 수 있는 플라스미드, 바이러스, 박테리오파지, 프로-바이러스, 파지미드, 트랜스포존, 인공 염색체 등일 수 있다. 벡터는 또한 자율적으로 복제하지 않는 네이키드 RNA 폴리뉴클레오타이드, 네이키드 DNA 폴리뉴클레오타이드, 동일한 가닥 내의 DNA 및 RNA 모두로 구성된 폴리뉴클레오타이드, 폴리-라이신-컨쥬게이트된 DNA 또는 RNA, 펩타이드-컨쥬게이트된 DNA 또는 RNA, 리포솜-컨쥬게이트된 DNA 등일 수 있다. 본 발명에 사용된 용어 "발현"은 기능적 최종 산물, 예를 들어 mRNA 또는 단백질(전구체 또는 성숙)의 생산을 의미한다.

[0066] 용어 "작동 가능하게 연결된"은 추가 폴리뉴클레오타이드의 전사를 초래하는 추가의 올리고뉴클레오타이드 또는 폴리뉴클레오타이드에 의한 본 발명에 따른 프로모터 폴리뉴클레오타이드의 순차적 배열을 의미한다. 일부 실시태양에서, 본 발명의 프로모터 서열은 유전자의 5'UTR 또는 오픈 리딩 프레임 직전에 삽입된다. 다른 실시태양에서, 본 발명의 작동 가능하게 연결된 프로모터 서열 및 유전자 서열은 하나 이상의 링커 뉴클레오타이드에 의해 분리된다. CRISPR 프로토스페이서 및 프로토스페이서 인접 모티프(PAMs)의 문맥에서 용어 "작동 가능하게 연결된"은 CRISPR 엔도뉴클레아제 복합체에 의해 고효율로 절단될 수 있는 인접 배치된 프로토스페이서/PAM 조합 서열을 의미한다.

[0067] 용어 "CRISPR RNA" 또는 "crRNA"는 표적 DNA 서열과의 혼성화 및 CRISPR 엔도뉴클레아제의 모집을 담당하는 RNA 가닥을 의미한다. crRNA는 자연 발생적이거나 RNA를 생산하는 임의의 공지된 방법에 따라 합성될 수 있다.

[0068] 용어 "가이드 서열" 또는 "스페이서"는 표적 DNA와의 혼성화를 담당하는 crRNA 또는 가이드 RNA(gRNA)의 일부분을 의미한다.

[0069] 용어 "프로토스페이서"는 crRNA 또는 가이드 가닥에 의해 표적화된 DNA 서열을 의미한다. 일부 실시태양에서, 프로토스페이서 서열은 CRISPR 복합체의 crRNA 가이드 서열과 혼성화한다.

[0070] 용어 "씨드 영역"은 DNA 서열 CRISPR 리보뉴클레오타이드 단백질 복합체 간의 초기 착물화를 담당하는 리보핵산 서열을 의미한다. 씨드 영역과 표적 DNA 서열 사이의 불일치는 crRNA/sgRNA 서열의 나머지 부분보다 표적 부위 인식 및 절단에 더 강한 영향을 미친다. 일부 실시태양에서, crRNA/gRNA의 씨드 영역에서의 단일 불일치는 CRISPR 복합체를 그 결합 부위에서 불활성으로 만들 수 있다. 일부 실시태양에서, Cas9 엔도뉴클레아제의 씨드 영역은 PAM에 인접한 프로토스페이서 표적 서열의 부분에 상응(혼성화)하는 가이드 서열의 3' 부분의 마지막 ~ 12nt를 따라 위치된다. 일부 실시태양에 있어서, Cpf1 엔도뉴클레아제의 씨드 영역은 PAM에 인접한 프로토스페이서 표적 서열의 부분에 상응(혼성화)하는 가이드 서열의 5' 부분의 첫 번째 ~ 5nt를 따라 위치된다.

[0071] 용어 "tracrRNA"는 작은 트랜스-암호화 RNA를 의미한다. TracrRNA는 crRNA와 상보적이고 염기 쌍을 이뤄 CRISPR 엔도뉴클레아제를 표적 서열에 모집할 수 있는 crRNA/tracrRNA 하이브리드를 형성한다.

[0072] 본 발명에 사용된 용어 "가이드 RNA" 또는 "gRNA"는 CRISPR 엔도뉴클레아제를 표적 서열에 모집할 수 있는 RNA

서열 또는 서열 조합을 의미한다. 따라서, 본 발명에 사용된 바와 같이, 가이드 RNA는 천연 또는 합성 crRNA(예를 들어, Cpf1의 경우), 천연 또는 합성 crRNA/tracrRNA 하이브리드(예를 들어, Cas9의 경우) 또는 단일 가이드 RNA(sgRNA)일 수 있다.

[0073] 본 발명에 사용된 용어 "CRISPR 랜딩 부위"는 CRISPR 복합체에 의해 표적화될 수 있는 DNA 서열을 의미한다. 따라서, 일부 실시태양에서, CRISPR 랜딩 부위는 CRISPR 엔도뉴클레아제 복합체가 절단될 수 있는 인접 배치된 프로토스페이서/프로토스페이서 인접 모티프 조합 서열을 포함한다. 용어 "확증된 CRISPR 랜딩 부위"는 상기 서열의 고효율 절단을 유도할 수 있는 가이드 RNA가 존재하는 CRISPR 랜딩 부위를 의미한다. 따라서, 용어 "확증된"은 서열이 이전에 CRISPR 복합체에 의해 절단 가능한 것으로 입증되었다는 의미로 해석되어야 한다. 각각의 "확증된 CRISPR 랜딩 부위"는 정의에 의해 확증과 관련된 테스트된 가이드 RNA의 존재를 확인할 것이다.

[0074] 용어 "점착성 말단(들)"은 서열 오버행을 포함하는 이중 가닥 폴리뉴클레오타이드 분자 말단을 의미한다. 일부 실시태양에서, 점착성 말단은 5' 또는 3' 서열 오버행을 갖는 dsDNA 분자 말단일 수 있다. 일부 실시태양에서, 본 발명의 점착성 말단은 동일 또는 다른 분자의 양립 가능한 점착성 말단과 혼성화할 수 있다. 따라서, 한 실시태양에서, 제 1 DNA 단편의 3' 상의 점착성 말단은 제 2 DNA 단편 상의 양립 가능한 점착성 말단과 혼성화될 수 있다. 일부 실시태양에서, 이들 혼성화 점착성 말단은 리가아제에 의해 함께 봉합될 수 있다. 다른 실시태양에서, 점착성 말단은 절찰 전에 dsDNA 분자를 완성하기 위해 오버행의 연장을 필요로 할 수 있다. 용어 "유전적 흔적(들)"은 DNA 조작 방법에 의해 핵산 서열에 도입된 임의의 바람직하지 않은 서열을 의미한다. 예를 들어, 일부 실시태양에서, 본 발명은 제한 효소 결합 부위, 복제를 수용하기 위한 서열 어택터 또는 스페이서, TA 부위, NHEJ로부터 남겨진 흔적 등과 같은 유전적 흔적을 교시한다. 일부 실시태양에서, 본 발명은 흔적없는 복제 및 유전자 편집을 교시한다.

[0075] 본 발명에 사용된 용어 "표적화된"은 하나의 품목 또는 분자가 비 표적 품목 또는 분자를 배제하기 위해 특이성의 정도로 다른 품목 또는 분자와 상호 작용할 것이라는 기대를 의미한다. 예를 들어, 본 발명에 따른, 제 2 폴리뉴클레오타이드에 대해 표적화된 제 1 폴리뉴클레오타이드는 서열 특이적 방식으로 (예를 들어, 왓슨-크릭 염기쌍을 통해) 제 2 폴리뉴클레오타이드와 혼성화하도록 디자인되었다. 일부 실시태양에서, 혼성화의 선택된 영역은 혼성화를 하나 이상의 표적화된 영역에 특유하게 하도록 디자인된다. 제 2 폴리뉴클레오타이드는 그의 표적화 서열(혼성화 영역)이 돌연변이되거나 그렇지 않으면 제 2 폴리뉴클레오타이드로부터 제거/분리된다면, 제 1 표적화 폴리뉴클레오타이드의 표적이 되는 것을 중지시킬 수 있다.

[0076] 본 발명은 "메가모듈식" 구조체 또는 디자인으로서 교시되고 기술된 범용 모듈식 CRISPR DNA 구조체 또는 디자인을 의미한다.

[0077] DNA 뉴클레아제

[0078] 일부 실시태양에서, 본 발명은 DNA 뉴클레아제를 이용한 유전자 편집/복제를 위한 방법 및 조성물을 교시한다. CRISPR 복합체, 전사 활성화제 유사 작용제 뉴클레아제(TALENs), 아연 핑거 뉴클레아제(ZFNs) 및 FokI 제한 효소는 유전자 편집 도구로 사용된 서열 특이성 뉴클레아제의 일부이다. 이들 효소는 관심 서열을 인식하도록 조작된 가이드 영역과의 상호 작용을 통해 원하는 표적 유전자좌에 그들의 뉴클레아제 활성을 표적화할 수 있다. 일부 실시태양에서, 본 발명은 CRISPR-기반 유전자 편집 방법을 교시한다.

[0079] 생체내 CRISPR-기반 편집의 원리는 주로 천연 세포 DNA 복구 시스템에 의존한다. 뉴클레아제에 의해 도입된 이중 가닥 dsDNA 절단은 비 상동성 말단 결합(NHEJ) 또는 상동성 직접 수선(HDR) 또는 단일 가닥 어닐링(SSA) 또는 미세상동 말단 결합(MMEJ)에 의해 복구된다.

[0080] HDR은 DNA 절단의 표적화된 부위를 둘러싸는 영역에 상동성인 서열을 함유하는 주형 DNA에 의존한다. 세포 수리 단백질은 외인성으로 공급된 또는 내인성 DNA 서열과 dsDNA 절단을 복구하기 위한 DNA 절단 주위 부위 사이의 상동성을 사용하여 절단을 주형 DNA 상의 서열로 대체한다. 그러나 주형 DNA를 통합하는데 실패는 NHEJ, MMEJ 또는 SSA를 발생시킬 수 있다. NHEJ, MMEJ 및 SSA는 종종 표적 부위에서 뉴클레오타이드(indels)의 삽입 또는 결실을 수반하는 오류가 발생하기 쉬운 과정으로, 조기 정지 코돈의 프레임시프트 돌연변이 또는 삽입에 의한 게놈의 표적화된 영역의 유전자 녹아웃(silencing)을 초래한다. Cpf1 매개 편집은 또한 엔도뉴클레아제에 의해 생성된 오버행의 전통적인 혼성화와 이어지는 절찰을 통해 기능할 수 있다.

[0081] CRISPR 엔도뉴클레아제는 또한 본 발명의 이후 섹션에서 논의되는 바와 같이 생체의 DNA 조작에 유용하다.

[0082] CRISPR 시스템

- [0083] CRISPR(Clustered Regularly Interspaced Short Palindromic Repeats) 및 CRISPR-관련(CRISPR-associated(cas)) 엔도뉴클레아제는 원래 바이러스 및 플라스미드 침입으로부터 보호하기 위해 박테리아 및 고세균에 의해 진화된 적응 면역 시스템으로서 발견되었다. 박테리아에서 자연적으로 발생하는 CRISPR/Cas 시스템은 하나 이상의 Cas 유전자와 이전에 만났던 바이러스 및 플라스미드에서 얻은 게놈 표적화 서열에 의해 분리된 염기 서열의 짧은 회문 반복체(스페이서라고 함)로 이루어진 하나 이상의 CRISPR 어레이로 구성된다. (Wiedenheft, B., *et al.* Nature. 2012; 482:331; Bhaya, D., *et al.*, Annu. Rev. Genet. 2011; 45:231; and Terms, M.P. *et al.*, Curr. Opin. Microbiol. 2011; 14:321). 하나 이상의 CRISPR 유전자좌를 가진 박테리아와 고세균은 CRISPR 어레이의 근위 말단에서 외래 서열(프로토스페이서)의 짧은 단편을 숙주 염색체에 통합 시킴으로써 바이러스 또는 플라스미드 공격에 대응한다. CRISPR 유전자좌의 전사는 이전에 만났던 침입하는 핵산과 상보적인 서열을 함유하는 CRISPR 유래 RNA(crRNA)의 라이브러리를 생성한다(Haurwitz, R.E., *et al.*, Science. 2012:329;1355; Gesner, E.M., *et al.*, Nat. Struct. Mol. Biol. 2001:18;688; Jinek, M., *et al.*, Science. 2012:337; 816-21). crRNA에 의한 표적 인지는 Cas 단백질에 의한 외래 서열의 절단을 지시하는 표적 DNA와의 상보적 염기쌍을 통해 일어난다. (Jinek *et al.* 2012 "A Programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Science. 2012:337; 816-821).
- [0084] 적어도 5개의 주요 CRISPR 시스템 유형(유형 I, II, III, IV 및 V) 및 적어도 16개의 별개의 아형이 존재한다 (Makarova, K.S., *et al.*, Nat Rev Microbiol. 2015. Nat. Rev. Microbiol. 13, 722-736). CRISPR 시스템은 또한 이펙터 단백질에 따라 분류된다. 클래스 1 시스템은 다중 서브 유닛 crRNA-이펙터 복합체를 보유하는 반면, 클래스 2 시스템에서 이펙터 복합체의 모든 기능이 단일 단백질(예를 들어, Cas9 또는 Cpf1)에 의해 수행된다. 일부 실시태양에서, 본 발명은 유형 II 및/또는 유형 V 단일-서브유닛 이펙터 시스템을 사용하는 것을 교시한다. 따라서, 일부 실시태양에서, 본 발명은 클래스 2 CRISPR 시스템들을 사용하여 교시한다.
- [0085] **CRISPR/Cas9**
- [0086] 일부 실시태양에서, 본 발명은 유형 II CRISPR 시스템을 사용하여 유전자 편집 방법을 교시한다. 일부 실시태양에서, 유형 II CRISPR 시스템은 Cas9 효소를 사용한다. 유형 II 시스템은 i) 단일 엔도뉴클레아제 단백질, ii) 트랜스활성화 crRNA(tracrRNA), 및 iii) crRNA의 5' 말단의 ~ 20 뉴클레오타이드(nt) 부분이 표적 핵산에 상보적인 crRNA에 의존한다. 표적 DNA 프로토스페이서에 상보적인 CRISPR crRNA 가닥의 영역은 본 발명에서 "가이드 서열"이라고 불린다.
- [0087] 일부 실시태양에서, 유형 II 시스템의 tracrRNA 및 crRNA 성분은 단일 가이드 RNA(sgRNA)로 대체될 수 있다. sgRNA는, 예를 들어, 표적 DNA 서열(가이드 서열)에 상보적인 적어도 12-20개 뉴클레오타이드 서열을 포함하는 뉴클레오타이드 서열을 포함할 수 있고 이의 3' 말단에 공통 스캐폴드 RNA 서열을 포함할 수 있다. 본 발명에 사용된 "공통 스캐폴드 RNA"는 tracrRNA 서열을 모방하는 임의의 RNA 서열 또는 tracrRNA로서 기능하는 임의의 RNA 서열을 의미한다.
- [0088] Cas9 엔도뉴클레아제는 평활 말단 DNA 절단을 생산하고 crRNA 및 tracrRNA 올리고머의 조합에 의해 표적 DNA에 모집되며, 이것이 RNA CRISPR 복합체의 상보적 혼성화를 통해 엔도뉴클레아제를 연결한다. (도 1a의 검은 색 삼각형 화살표 참조).
- [0089] 일부 실시태양에서, crRNA/엔도뉴클레아제 복합체에 의한 DNA 인식은 표적 프로토스페이서로부터의 하류인 표적 DNA의 3' 부분에 위치한 프로토스페이서 인접 모티프(protospacer adjacent motif(PAM))(예를 들어, 5'-NGG-3')와의 추가적인 상보적인 염기쌍을 필요로 한다. (Jinek, M., *et al.*, Science. 2012:337;816-821). 일부 실시태양에서, Cas9에 의해 인식되는 PAM 모티프는 상이한 Cas9 단백질에 대해 다양하다.
- [0090] 일부 실시태양에서, 당업자는 본 발명에 개시된 Cas9가 임의의 공급원으로부터 유도되거나 분리된 임의의 변이체일 수 있음을 인식할 수 있다. 예를 들어, 일부 실시태양에서, 본 발명의 Cas9 펩타이드는 SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, 및 SEQ ID NO: 6로부터 선택된 하나 이상의 SEQ ID NOs를 포함할 수 있다. 다른 실시태양에서, 본 발명의 Cas9 펩타이드는 Fonfara *et al.* Nucleic Acids Res. 2014 Feb;42(4):2577-90; Nishimasu H. *et al.* Cell. 2014 Feb 27;156(5):935-49; Jinek M. *et al.* Science. 2012 337:816-21; 및 Jinek M. *et al.* Science. 2014 Mar 14;343(6176)에 기술된 기능성 돌연변이를 포함하나 이에 제한되지 않는 문헌에 기술된 하나 이상의 돌연변이를 포함할 수 있고; 또한 본 발명에 참조로 포함되는 2013년 3월 15일에 출원된 미국 특허 출원 제 13/842,859호 참조; 또한, 본 발명에 참조로 포함되는 미국 특허 제8,697,359호; 제8,771,945호; 제8,795,965호; 제8,865,406호; 제8,871,445호; 제8,889,356호; 제8,895,308호; 제8,906,616호; 제8,932,814호; 제8,945,839호; 제8,993,233호; 및 제8,999,641호에 참조. 따

라서, 일부 실시태양에서, 본 발명에 개시된 시스템 및 방법은 이중 가닥 뉴클레아제 활성을 갖는 야생형 Cas9 단백질, 단일 가닥 뉴클레아제로서 작용하는 Cas9 돌연변이체 또는 변형된 뉴클레아제 활성을 갖는 다른 돌연변이체와 함께 사용될 수 있다.

[0091] **CRISPR/Cpf1**

[0092] 다른 실시태양에서, 본 발명은 유형 V CRISPR 시스템을 사용하여 유전자 편집 방법을 교시한다. 일부 실시태양에서, 본 발명은 프레보텔라 및 프란시셀라(*Prevotella* and *Francisella*) 1(Cpf1)의 CRISPR을 사용하는 방법을 교시한다.

[0093] 본 발명의 Cpf1 CRISPR 시스템은 i) 단일 엔도뉴클레아제 단백질, 및 ii) crRNA의 3' 말단의 일부가 표적 핵산에 상보적인 가이드 서열을 함유하는 crRNA를 포함한다. 이 시스템에서, Cpf1 뉴클레아제는 crRNA에 의해 표적 DNA에 직접 모집된다(도 1b의 검은 색 삼각형 화살표 참조). 일부 실시태양에서, Cpf1에 대한 가이드 서열은 검출 가능한 DNA 절단을 달성하기 위해 적어도 12nt, 13nt, 14nt, 15nt 또는 16nt이어야 하고, 효율적인 DNA 절단을 달성하기 위해 최소 14nt, 15nt, 16nt, 17nt 또는 18nt이어야 한다.

[0094] 본 발명의 Cpf1 시스템은 다양한 방식으로 Cas9와 상이하다. 첫 번째, Cas9와는 달리, Cpf1은 분열을 위한 별도의 tracrRNA를 필요로 하지 않는다. 일부 실시태양에서, Cpf1 crRNA는 길이가 약 42-44 염기 정도로 짧을 수 있으며, 그 중 23-25nt는 가이드 서열이고 19nt는 구조적 직접 반복체 서열이다. 대조적으로, Cas9 tracrRNA 및 crRNA 합성 서열의 결합은 약 100 염기 길이일 수 있다. 일부 실시태양에서, 본 발명은 "가이드 RNA"로서 Cpf1에 대한 crRNA를 언급할 것이다.

[0095] 두 번째, Cpf1은 이의 표적의 5' 상류에 위치한 "TTN" PAM 모티프를 선호한다. 이것은 Cas9 시스템의 표적 DNA의 3' 상에 위치한 "NGG" PAM 모티프와는 대조적이다. 일부 실시태양에서, 가이드 서열 직전의 우라실 염기는 치환될 수 없다(모든 목적을 위해 그 전체가 참조 문헌으로 본 발명에 포함되는 Zetsche, B. *et al.* 2015. "Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System" Cell 163, 759-771).

[0096] 세 번째로, Cpf1 절단 부위는 약 3-5개 염기에 의해 엇갈리게 배치되어 "끈끈한 말단"을 만든다(2016년 6월 온라인 발행된 Kim *et al.*, 2016. "Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells"). ~3-5nt 오버행을 갖는 이 점착성 말단은 NHEJ 매개 결찰을 촉진하고, 말단이 일치하는 DNA 단편의 유전자 편집을 향상시키는 것으로 생각된다. 절단 부위는 PAM이 있는 5' 말단에 대해 멀리 있는 표적 DNA의 3' 말단에 있다. 절단 위치는 대개 비 혼성화 가닥 상의 18번째 염기와 crRNA에 혼성화된 상보적 가닥 상의 상응하는 23번째 염기를 따른다(도 1b).

[0097] 넷째, Cpf1 복합체에서, "씨드" 영역은 가이드 서열의 처음 5nt 내에 위치한다. Cpf1 crRNA 씨드 영역은 돌연변이에 매우 민감하며 이 영역에서의 단일 염기 치환조차도 절단 활성을 크게 감소시킬 수 있다(Zetsche B. *et al.* 2015 "Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System" Cell 163, 759-771 참조). 비판적으로, Cas9 CRISPR 표적과 달리, Cpf1 시스템의 절단 부위 및 씨드 영역은 중첩되지 않는다. 올리고스를 표적으로 하는 Cpf1 crRNA를 디자인하는 것에 대한 추가 지침은 (Zetsche B. *et al.* 2015. "Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System" Cell 163, 759-771)에서 이용 가능하다.

[0098] 당업자는 본 발명에 개시된 Cpf1이 임의의 공급원으로부터 유도되거나 분리된 임의의 변이체일 수 있음을 인식할 것이다. 예를 들어, 일부 실시태양에서, 본 발명의 Cpf1 펩타이드는 SEQ ID NO: 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 또는 이의 임의의 변이체로부터 선택된 하나 이상의 SEQ ID Nos를 포함할 수 있다.

[0099] **리가아제**

[0100] 일부 실시태양에서, 본 발명은 표적화된 Cpf1 복합체를 통해 표적 DNA를 절단한 다음, 생성된 점착성 말단을 DNA 삽입체와 결찰시키는 방법을 교시한다. 일부 실시태양에서, 본 발명은 표적 DNA를 절단하는 Cpf1 복합체 및 DNA를 다시 "봉합"하는 리가아제를 제공하는 방법을 교시한다. 다른 실시태양에서, 본 발명은 연결된 리가아제 효소를 포함하는 변형된 Cpf1 복합체를 교시한다.

[0101] 본 발명에 사용된 용어 "리가아제"는 임의의 수의 효소 또는 비 효소 시약을 포함할 수 있다. 예를 들어, 리가아제는 적절한 조건하에서 DNA 분자, RNA 분자 또는 하이드리드의 인접한 뉴클레오타이드의 3'-OH와 5'-포스페이트 사이에 포스포다이에스터 결합을 형성하는 효소적 결찰 시약 또는 촉매이다.

- [0102] 일부 실시태양에서, 본 발명은 효소적 리가아제의 사용을 교시한다. 양립 가능한 온도 민감성 효소적 리가아제는 박테리오파지 T4 리가아제 및 대장균 리가아제를 포함하나 이에 제한되지 않는다. 열 안정성 리가아제는 Afu 리가아제, Taq 리가아제, Tfl 리가아제, Tth 리가아제, Tth HB8 리가아제, 썬머스(Thermus) 종 AK16D 리가아제 및 Pfu 리가아제를 포함하나 이에 제한되지 않는다(예를 들어 공개된 PCT 출원 WO/2000/026381, Wu *et al.*, Gene, 76(2):245-254, (1989), 및 Luo *et al.*, Nucleic Acids Research, 24(15): 3071-3078 (1996)). 당업자는 임의의 수의 열 안정성 리가아제는 호열성 또는 과호열성 유기체, 예를 들어 진균 및 고세균의 특정 종으로부터 얻을 수 있음과 이런 리가아제가 개시된 방법 및 키트에 사용될 수 있음을 인식할 것이다. 일부 실시태양에서, 가역적으로 불활성화된 효소(예를 들어, 미국 특허 제5,773,258호 참조)가 본 발명의 일부 실시태양에서 사용될 수 있다.
- [0103] 다른 실시태양에서, 본 발명은 화학적 결찰제의 사용을 교시한다. 화학적 결합제는 카보다이미드, 시아노겐 브로마이드(BrCN), N-사이아노이미다졸, 이미다졸, 1-메틸이미다졸/카보다이미드/시스타민, 다이티오프레이톨(DTT) 및 자외선과 같은 활성제, 촉합제 및 환원제를 제한없이 포함한다. 자동 결찰, 즉 결찰제 부재 하에서의 자발적 결찰도 또한 본 발명의 교시의 범위 내에 있다. 화학적 결찰 방법에 대한 상세한 프로토콜 및 적절한 반응성 그룹에 대한 설명은, 다른 곳 중에서, Xu *et al.*, Nucleic Acid Res., 27:875-81 (1999); Gryaznov and Letsinger, Nucleic Acid Res. 21:1403-08 (1993); Gryaznov *et al.*, Nucleic Acid Res. 22:2366-69 (1994); Kanaya and Yanagawa, Biochemistry 25:7423-30 (1986); Luebke and Dervan, Nucleic Acids Res. 20:3005-09 (1992); Sievers and von Kiedrowski, Nature 369:221-24 (1994); Liu and Taylor, Nucleic Acids Res. 26:3300-04 (1999); Wang and Kool, Nucleic Acids Res. 22:2326-33 (1994); Purmal *et al.*, Nucleic Acids Res. 20:3713-19 (1992); Ashley and Kushlan, Biochemistry 30:2927-33 (1991); Chu and Orgel, Nucleic Acids Res. 16:3671-91 (1988); Sokolova *et al.*, FEBS Letters 232:153-55 (1988); Naylor and Gilham, Biochemistry 5:2722-28 (1966); 및 U.S. Pat. No. 5,476,930에서 발견할 수 있다.
- [0104] 일부 실시태양에서, 본 발명의 방법, 키트 및 조성물은 또한 광결찰 반응과 양립 가능하다. 적절한 파장의 광을 결찰제로 사용하는 광결찰도 이 교시의 범위 내에 있다. 일부 실시태양에서, 광결찰은 4-티오티미딘, 5-바이닐 우라실 및 이의 유도체 또는 이들의 조합을 포함하나 이에 제한되지 않는 뉴클레오타이드 유사체를 포함하는 프로브를 포함한다. 일부 실시태양에서, 결합제는 (a) UV-A 범위(약 320nm 내지 약 400nm), UV-B 범위(약 290nm 내지 약 320nm), 또는 이들의 조합 (b) 약 300nm 내지 약 375nm의 파장을 갖는 광, (c) 약 360nm 내지 약 370nm의 파장을 갖는 광; (d) 약 364nm 내지 약 368nm의 파장을 갖는 광, 또는 (e) 약 366nm의 파장을 갖는 광을 포함한다. 일부 실시태양에서, 광결찰은 가역적이다. 광결찰의 설명은, 다른 곳 중에서, Fujimoto *et al.*, Nucl. Acid Symp. Ser. 42:39-40 (1999); Fujimoto *et al.*, Nucl. Acid Res. Suppl. 1:185-86 (2001); Fujimoto *et al.*, Nucl. Acid Suppl., 2:155-56 (2002); Liu and Taylor, Nucl. Acid Res. 26:3300-04 (1998) and on the world wide web at: sbchem.kyoto-u.ac.jp/saito-lab에서 발견할 수 있다.
- [0105] **범용 모듈식 CRISPR DNA 구조체 및 이의 용도**
- [0106] 일부 실시태양에서, 본 발명은 DNA 구조체의 모듈식 조립을 위한 전략을 기술한다. 일부 실시태양에서, 본 발명의 DNA 조립 방법은 플라스미드, 작은 선형 DNA 및 형질전환된 염색체 유전자좌를 포함하는 임의의 구조체에 적용 가능하다.
- [0107] 양태들에서, 발명자들은 "범용 모듈식 CRISPR DNA 구조체"를 "메가모듈식 (MegaModular)" 디자인으로 언급한다.
- [0108] **전통적인 DNA 편집 및 조립 기술의 단점**
- [0109] 전통적인 다성분 DNA 복제 전략은 복잡한 서열을 갖는 다중 성분 DNA 구조체를 효과적으로 조립 및 변형시키는 능력이 제한적이다. 예를 들어, 제한 효소 복제는 각 DNA 삽입체의 복제 접합부 및 최종 벡터 내에서의 목적 부위에 적절하게 위치한 특유의 제한 효소 인식 부위의 이용 가능성에 의해 제한된다. 게이트웨이 복제 기술은 다성분 조립에 이용될 수 있는 비교적 적은 수의 특유의 재조합 부위에 의해 유사하게 제한된다.
- [0110] 전통적인 DNA 조립 기술에 대한 또 다른 단점은 서열을 편집할 수 있는 능력이 종종 제작 시간에 제한된다는 것이다. 예를 들어, 리가아제 순환 반응과 같은 효율적인 조립 전략의 생산물(LCR은 초기 조립이 완료되면 쉽게 변형되지 않는다(Kok, S, *et al.*, 2014 "Rapid and Reliable DNA assembly via Ligase Cycling Reaction" ACS Synth. Biol., 3 (2): 97-106). 일반적인 제한 인식 부위가 제한 부위를 함유하는 폴리뉴클레오타이드가 조립되는 구조체 속에 삽입되거나 구조체가 상기 부위의 염색체 속에 통합될 때 독특한 복제 지점으로서 기능하는 것을 중지하는 전통적인 제한 효소 복제에 의해 유사한 우려가 나타난다. 따라서 연속적인 제한 복제를 통해 생산

된 벡터는 복제 과정이 잘 진행된다면 서열을 고정하거나 업데이트하는데 매우 적은 옵션을 제공한다.

[0111] 종래의 CRISPR DNA 조립 기술과 같은 더 새로운 기술조차도 유사한 복잡성, 이전에 조립된 구조체/벡터 디자인에 대한 반복의 용이성 및 속도 제한으로 어려움을 겪고 있다(Wang, JW. *et al.*, 2015 "CRISPR/Cas9 nuclease combined with Gibson assembly for seamless cloning" *BioTechniques*, Vol 58, No. 4:161-170). CRISPR 복제는 양립 가능한 프로토스페이서 인접 모티프(PAM) 옆에 표적화된 기능성 가이드 RNA의 디자인이 필요하다. 표적 부위 내 적합한 PAM 서열의 이용 가능성은 게놈 또는 구조체 내의 가능한 DNA 삽입 위치의 수에 대한 중요한 디자인 제한을 초래한다.

[0112] 더욱이, 가이드 RNA 서열의 디자인 및 테스트는 다성분 조립에 대한 중요한 기술적 도전을 부과한다. 당업자는 예를 들어, 모든 gRNA 서열이 기능적이지는 않으며, CRISPR DNA 조립의 효과적인 구현은 종종 다중 gRNA 서열 변이체의 디자인 및 검증을 필요로 한다는 것을 인식할 것이다. 이러한 제한은 원하는 변형을 성공적으로 일으키는 단일 gRNA 서열의 고장이 원래 복제 계획에 더 이상 포함되지 않는 후속 조립 성분을 재디자인할 필요성을 야기할 수 있는 다성분 조립에서 특히 부담이 된다. 따라서 다성분 조립의 모든 접합부에 대해 다수의 주문형 가이드 RNA를 필요로 하는 기술을 적용하는 것은 또한 매우 고가이고, 부담이 되고, 비실용적일 수 있다.

[0113] 모듈식 CRISPR 태그 조립 벡터 및 이의 사용 방법

[0114] 일부 실시태양에서, 본 발명은 상기 기술한 전통적인 기술과 관련된 많은 제한을 극복하는 DNA 조립 방법을 교시한다. 일부 실시태양에서, 본 발명은 또한 본 발명의 방법과 함께 사용하기 위한 모듈식 CRISPR 조립 구조체, 조성물 및 키트를 교시한다.

[0115] 일부 실시태양에서, 본 발명은 하나 이상의 CRISPR 다중 복제 부위(cMCS)를 포함하는 DNA 구조체를 교시한다. 일부 실시태양에서, 본 발명의 cMCS는 기술된 DNA 구조체의 단지 일부를 나타낸다(즉, 구조체의 일부만이 본 발명의 방법에 따라 쉽게 편집 가능하다). 다른 실시태양에서, 본 발명의 cMCS는 전체 구조체 내의 주요 위치 상에 위치되어, 전체 DNA 구조체가 쉽게 편집 가능하다. 따라서, 일부 실시태양에서, 모듈식 cTAG 벡터의 모든 기능적 부분(예를 들어, 모든 기원, 마커, 화물, 조립에 필요한 요소)은 삽입체 DNA 부분 내에 포함되며 본 발명의 유전자 편집 방법을 통해 쉽게 교환될 수 있다.

[0116] 일부 실시태양에서, 본 발명의 cMCS는 적어도 하나의 확증된 CRISPR 표적화 부위를 각각 포함하는 하나 이상의 복제 태그(cTAG)를 포함한다. 일부 실시태양에서, 본 발명의 cMCS는 각각 한 쌍의 cTAG가 측면에 위치되는 DNA 삽입체 부분을 추가로 포함하여, 하나 이상의 cTAG를 표적으로 하는 하나 이상의 CRISPR 엔도뉴클레아제를 갖는 cMCS의 분해가 상기 측면에 위치한 삽입체 부분을 방출하여, 양립 가능한 도너 DNA 부분의 삽입을 허용할 것이다.

[0117] 본 명세서의 도 2 및 3은 본 발명의 방법에 따른 모듈식 CRISPR 조립 플라스미드 구조체의 한 실시태양을 도시한다. 개시된 예시적 플라스미드는 각각 도 2a의 한 쌍의 cTAG(태그 A-H)에 의해 측면에 위치되는 일련의 DNA 삽입(도 2a의 파트 1-8)을 함유한다. 적절한 CRISPR/가이드 서열 복합체를 갖는 본 실시예의 cTAG A 및 B의 분해는 플라스미드의 파트 2를 방출하여 원하는 특성을 갖는 대체 파트 2 삽입체의 삽입을 가능하게 한다.

[0118] 당업자는 생체내 및 생체외에서 벡터의 서열 특이적 모듈식 복제/편집을 가능하게 하는 현재 기술된 벡터 시스템의 이점을 즉각 인식할 것이다. 아래의 섹션은 개시된 모듈식 복제 벡터의 다양한 양태뿐만 아니라 분자 생물학, 유전자 치료 및 유전자 편집에 대한 이들의 다양한 응용을 약술할 것이다.

[0119] 모듈식 CRISPR 벡터 삽입체 부분

[0120] 일부 실시태양에서, 본 발명의 삽입체 부분은 CRISPR 분해 후 상동성 재조합 삽입을 위한 도너 DNA 서열이다. 따라서, 일부 실시태양에서, 본 발명의 삽입체 부분 서열은 서열의 상동성 재조합, 혼성화 및 삽입을 유발하기 위해, 분해된 모듈식 CRISPR 구조체의 말단에 충분한 상동성을 갖는 서열이 측면에 위치하는 관심 삽입체 서열을 포함한다.

[0121] 다른 실시태양에서, 본 발명의 삽입체 부분은 점착성 말단(예를 들어, Cpf1 분해, 제한 효소 분해, 김슨 조립, 또는 LCR을 포함하는 기타 혼성화 기반 조립 이후)을 통한 혼성화 및 결찰이 가능한 도너 DNA 서열이다. 따라서, 일부 실시태양에서, 본 발명의 삽입체 부분 서열은 점착성 말단의 혼성화를 허용하기 위해 분해된 모듈식 CRISPR 구조체의 말단과 충분한 상동성을 갖는 서열이 측면에 위치하는 관심 삽입체 서열을 포함한다.

[0122] 또 다른 실시태양에서, 본 발명의 삽입체 부분은 평활 말단 결찰을 위한 도너 DNA 서열이다.

- [0123] 일부 실시태양에서, 본 발명의 모듈식 CRISPR DNA 구조체는 임의의 삽입체 부분 서열과 양립 가능하다. 따라서, 본 벡터의 부분은 선택 가능한 마커, 복제 기원, 프로모터, 터미네이터 서열; 다른 조절 서열, 바코드, 재조합 부위 또는 사용자에게 관심있는 다른 서열을 포함하나 이에 제한되지 않는다. 일부 실시태양에서, 본 발명의 삽입체 부분은 하나 이상의 유전자좌로의 상동성 재조합 및 삽입을 유발하기 위한 상동성 서열을 포함할 수 있다. 일부 실시태양에서, 상기 상동성 재조합 삽입체 부분은 재조합 사건을 통해 게놈으로 삽입될 다른 삽입체 부분에 선행하고 후행할 것이다.
- [0124] 일부 실시태양에서, 본 발명은 각 삽입체 부분이 단일 서열(예를 들어, 단지 프로모터 또는 단지 관심 유전자, 도 2a, 파트 8 참조)을 포함한다는 것을 교시한다. 다른 실시태양에서, 본 발명은 하나 이상의 삽입체 부분이 관심 프로모터-유전자(GOI) 조합, 다중-아단위 키메라 단백질 융합체, 또는 심지어 전체 구조체와 같은 다중 요소(도 2a, 파트 5 참조, 프로모터-GOI-터미네이터 조합을 포함)를 함유할 수 있다.
- [0125] 일부 실시태양에서, 본 발명은 결합되지 않은 개별 삽입체 부분을 교시한다. 즉, 일부 실시태양에서, 본 발명은 하나 또는 복수의 연결되지 않은 삽입체 부분을 교시한다(도 2a 참조, 우측은 결합되지 않은 삽입체 부분의 목록을 나타낸다). 일부 실시태양에서, 본 발명은 상기 복수의 부분을 하나 이상의 모듈식 CRISPR 구조체로 조립하는 방법을 교시한다. 일부 양태에서, 본 발명은 메가모듈식 구조체를 조립하기 위한 키트를 교시한다.
- [0126] 다른 실시태양에서, 본 발명은 부분적으로 또는 완전히 조립된 모듈식 CRISPR DNA 구조체를 교시한다. 예를 들어, 일부 실시태양에서, 본 발명은 1, 2, 3, 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, 100 또는 그 이상의 조립된 삽입체 부분 및 이들 사이의 임의의 범위를 포함하는 모듈식 CRISPR DNA 구조체를 교시한다. 본 발명은 또한 상기 삽입체 부분을 포함하는 키트를 교시한다.
- [0127] 일부 실시태양에서, 상기 조립되거나 부분 조립된 모듈식 CRISPR DNA 구조체는 선형이다. 일부 실시태양에서, 상기 조립되거나 부분 조립된 모듈식 CRISPR DNA 구조체는 원형(예를 들어, 플라스미드)이다. 일부 실시태양에서, 상기 조립되거나 부분 조립된 모듈식 CRISPR DNA 구조체는 게놈 DNA에 통합된다.
- [0128] 일부 실시태양에서, 본 발명의 구조체는 초기에 추가 복제를 위한 플레이스 홀더로서 단지 짧은 스페이서 서열만을 함유할 것이다(도 3c의 "스터퍼(stuffer)"서열 참조). 일부 실시태양에서, 삽입체 부분 플레이스홀더는 작고 무작위화된 서열이다. 다른 실시태양에서, 본 발명의 벡터는 초기에 하나 이상의 미리 선택된 삽입체 DNA 부분을 포함할 것이다. 예를 들어, 일부 실시태양에서, 모듈식 CRISPR 구조체는 초기에 적어도 하나의 선택 가능한 마커 및/또는 적어도 하나의 복제 기점을 포함할 것이다.
- [0129] 적합한 선택 가능한 마커는 항생제 내성을 부여하는 유전자, 형광 단백질을 암호화하는 유전자, tRNA 유전자, 영양요구성 마커, 독성 유전자, 표현형 마커, 안티센스 올리고뉴클레오타이드, 제한 엔도뉴클레아제, 제한 엔도뉴클레아제 절단 부위, 효소 절단 부위, 단백질 결합 부위 및 PCR 프라이머 서열에 상보적인 서열을 포함하나 이에 제한되지 않는다.
- [0130] 적절한 항생제 내성 유전자는 클로람페니콜 내성 유전자, 암피실린 내성 유전자, 테트라사이클린 내성 유전자, 제오신 저항성 유전자, 스펙티노마이신 내성 유전자 및 카나마이신 내성 유전자를 포함하나 이에 제한되지 않는다.
- [0131] 본 발명의 특정 실시태양에서, 역선택 가능한 마커는 독성 유전자이다. 적절한 독성 유전자는 ccdB 유전자, 하나 이상의 ter 부위에 결합하는 tus 단백질을 암호화하는 유전자, kicB 유전자, sacB 유전자, ASK1 유전자, ΦX174 E 유전자 및 DpnI 유전자를 포함하나 이에 제한되지 않는다. 일부 실시태양에서, 독성 선택 가능한 마커의 존재는 삽입이 수행되지 않았거나 성공적이지 못한 지표로서 작용한다. 독성 선택 가능한 마커는 또한 여전히 독성 유전자가 있는 변형되지 않은 벡터를 보유하고 있는 세포에 사멸을 유발함으로써, 양성 세포의 비 변형 부모 벡터의 배경을 감소시키는 역할을 한다.
- [0132] 본 발명의 방법의 추가 실시태양에서, 모듈식 CRISPR 구조체는 하나 이상의 독성 유전자 및 하나 이상의 항생제 내성 유전자 모두를 포함할 수 있다.
- [0133] 일부 실시태양에서, 모듈식 CRISPR 구조체는 초기에 적어도 하나의 조절 서열을 포함할 것이다. 일부 실시태양에서, 본 발명은 제한 없이, 매트릭스 부착 영역, 발현 인슐레이터 서열, 발현 인핸서 서열, 프로모터, 5'UTR,

3'UTR, 터미네이터 서열, 종지 코돈, 개시 코돈 등을 포함하는 벡터를 교시한다. 일부 실시태양에서, 모듈식 CRISPR 구조체는 초기에 상기 구조체의 염색체 삽입을 용이하게 하기 위한 서열(예를 들어, t-DNA 경계, Cre/Lox 또는 염색체 서열에 대한 상동성 말단)을 포함할 것이다. 일부 실시태양에서, 염색체 삽입을 위한 서열은 유기체의 게놈에 전체 모듈식 CRISPR 구조체를 삽입하도록 위치된다. 다른 실시태양에서, 염색체 삽입을 위한 서열은 모듈식 CRISPR 구조체의 일부만 삽입하도록 위치된다(도 3d 참조).

[0134] 일부 실시태양에서, 본 발명의 삽입체 부분은 추가의 cTAG를 포함할 수 있다. 삽입체 부분을 통한 cTAG의 추가는 이용 가능한 복제 체계의 복잡성을 증가시킬 수 있으며 또한 대체될 수 있는 이용 가능한 삽입체 부분의 수를 확장하여 구조체의 크기를 확장할 수 있다.

[0135] 일부 실시태양에서, 본 발명의 삽입체 부분은 전통적인 복제 부위를 포함할 수 있다. 예를 들어, 일부 실시태양에서, 본 발명은 게이트웨이 재조합 부위, 제한 부위, Cre/Lox 부위, 또는 다른 전통적인 복제 부위를 포함하는 삽입체 부분을 교시한다.

[0136] 일부 실시태양에서, 본 발명은 전통적인 DNA 구조체로부터 삽입체 부분을 생산하는 방법을 교시한다. 즉, 일부 실시태양에서, 본 발명은 전통적인 DNA 구조체(예를 들어, 올리고, PCR 단편, 플라스미드 또는 다른 이용 가능한 DNA 단편)에 cTAG를 첨가하는 방법을 교시한다. 일부 실시태양에서, 본 발명은 관심 유전자(GOI), 프로모터와 같은 단일 성분에 cTAG를 첨가하는 방법을 교시한다. 다른 실시태양에서, 본 발명은 다중 요소 구조체에 cTAG를 첨가하는 방법을 교시한다.

[0137] 당업자는 삽입체 부분을 제조하는 방법을 인식할 것이다. 예를 들어, 일부 실시태양에서, cTAG는 상기 cTAG를 포함하는 프라이머로 PCR 증폭을 통해 DNA 분자 내로 혼입될 수 있다. 다른 실시태양에서, cTAG는 전통적인 복제 기술(예를 들어, 제한 효소, 겔순 또는 다른 조립 방법)을 통해 혼입될 수 있다. 또 다른 실시태양에서, cTAG는 평활 말단 연결을 통해 혼입될 수 있다.

[0138] 일부 실시태양에서, 본 발명의 삽입체 부분은 넓은 종의 상용성 스펙트럼을 가질 수 있다(예를 들어, 마커가 원핵 생물 및 진핵 생물 발현 서열 둘 다를 함유하여 다수의 생물체에서 효과적이게 할 수 있다). 다른 실시태양에서, 본 발명의 삽입체 부분은 단일 종/속/과/목/강/문/계 또는 영역 내의 유기체에 대한 제한된 응용 가능성을 갖도록 디자인된다. 일부 실시태양에서, 예를 들어, 복제 기원 부분은 단지 단일 종 또는 종 그룹에서 플라스미드를 유지할 수 있다. 다른 실시태양에서, 형광 마커는 원핵 생물 및 진핵 생물 영역 모두에서 기능하도록 코돈 최적화될 수 있다.

[0139] 일부 실시태양에서, Cas9 엔도뉴클레아제는 표적 서열의 PAM으로부터 3-4개 뉴클레오타이드 상류를 절단한다. 따라서 Cas9 복합체에 의한 cTAG 분해는 PAM 서열 또는 표적의 프로토스페이스 서열의 손실을 통한 cTAG 기능의 상실을 초래할 수 있다. 일부 실시태양에서, 본 발명은 도너 삽입체 서열을 디자인함으로써 상기 cTAG 서열의 기능성을 유지시켜 이들이 삽입시(예를 들어, 이전에 분실된 PAM 또는 프로토스페이스 서열의 삽입을 통해) cTAG 서열을 재구성하는 방법을 교시한다. 유사한 대비가 Cpf1 엔도뉴클레아제를 통해 절단된 서열에 대해 계획된다.

[0140] 도 2b는 cTAG 수리의 현재 개시된 개념을 도시한다. Cas9 엔도뉴클레아제에 의한 삽입체 부분 2의 절단은 또한 cTAG A 및 B의 일부의 손실을 초래한다. 상동성 재조합을 통한 삽입체 부분 2a-2d 중 임의의 하나의 후속 삽입은 전체 cTAG 서열의 복원을 초래한다.

[0141] 당업자는 삽입체 부분에 대한 거의 무한한 옵션을 인식할 것이다. 상기 삽입체 목록은 예시적인 것으로 의도되었으며, 현재 개시된 방법, 키트 및 구조체의 응용 가능성을 제한하는 것으로 결코 해석되어서는 안 된다.

[0142] **모듈식 CRISPR 복제 태그**

[0143] 일부 실시태양에서, 본 발명의 모듈식 CRISPR 구조체는 하나 이상의 복제 태그(cTAGs)를 포함한다. 일부 실시태양에서, 본 발명의 모듈식 CRISPR 구조체는 1, 2, 3, 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, 100 또는 그 이상의 cTAG를 포함한다.

[0144] 일부 실시태양에서, 본 발명은 각각의 cTAG가 적어도 하나의 확증된 CRISPR 프로토스페이스/PAM 조합 서열("CRISPR 랜딩 부위")를 포함함을 교시한다. 즉, 일부 실시태양에서, cTAG는 적어도 하나의 실험적으로 확증된

고효율 CRISPR 랜딩 부위를 포함한다. 일부 실시태양에서, 본 발명의 cTAG는 습식 벤치 실험(wet bench experimentation)(예를 들어, 상기 CRISPR 랜딩 부위를 표적으로 하는 CRISPR 복합체에 의한 cTAG 서열의 생체 외 절단)에 의해 확증될 수 있다. 다른 실시태양에서, cTAG 확증은 피어-리뷰 저널에서의 절단 리포트로부터 가정될 수 있다.

[0145] 일부 실시태양에서, 본 발명의 cTAG는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 또는 그 이상의 CRISPR 랜딩 부위를 포함한다. 일부 실시태양에서, CRISPR 랜딩 부위는 서로 중첩된다. 다른 실시태양에서, CRISPR 랜딩 부위는 cTAG 내의 뚜렷한 중첩되지 않는 영역을 차지한다. 일부 실시태양에서, CRISPR 랜딩 부위는 Cas9 또는 Cpf1 엔도뉴클레아제 절단에 특이적일 수 있다. 일부 실시태양에서, CRISPR 랜딩 부위는 임의의 다른 현재 또는 앞으로 발견될 CRISPR 엔도뉴클레아제에 특이적일 수 있다.

[0146] 다른 실시태양에서, 본 발명은 단일 cTAG 내의 다중 복제 부위가 상이한 유기체를 통해 기능하도록 디자인될 수 있음을 교시한다. 따라서, 일부 실시태양에서, cTAG Cpf1 랜딩 부위는 HR 기계장치가 결여되거나 하향조절되는 유기체에서 바람직할 수 있다. 다른 실시태양에서, cTAG의 제한 부위는 초기의 생체의 복제에 바람직할 수 있는 반면, Cas9 또는 Cpf1 랜딩 부위는 선택된 진행 생물에서 생체내에서 발생하는 보다 복잡한 편집에 바람직할 수 있다.

[0147] 일부 실시태양에서, 본 발명은 cTAG가 하나 이상의 비-CRISPR 복제 서열을 포함할 수 있음을 교시한다. 예를 들어, 일부 실시태양에서, 본 발명의 cTAG는 제한 효소 부위, 재조합 부위, 토포아이스머라제 부위, 스플라이싱 부위 및 Cre-Lox 부위로 이루어진 그룹으로부터 선택된 하나 이상의 요소를 포함할 수 있다.

[0148] 일부 실시태양에서, 적합한 제한 효소 부위는 제한 없이 AaII, AarI, AasI, AatII, Acc65I, AccB7I, AccI, AccIII, AciI, AclI, AcuI, AdeI, AfeI, AflIII, AflIII, AgeI, AhdI, AleI, AloI, AluI, Alw21I, Alw26I, Alw44I, AlwI, AlwNI, ApaI, ApaLI, ApeKI, ApoI, AscI, AseI, AsiSI, AvaI, AvaII, AvrII, BaeI, BalI, BamHI, BanI, BanII, BbsI, BbuI, BbvCI, BbvI, BccI, BceAI, BcgI, BciVI, BclI, BcnI, BcuI, BfaI, BfiI, BfmI, BfrBI, BfuAI, BfuCI, BfuI, BglI, BglII, BlpI, Bme1390I, Bme1580I, BmgBI, BmrI, BmtI, BoxI, BpiI, BplI, BpmI, Bpu10I, Bpu1102I, BpuEI, BsaAI, BsaBI, BsaHI, BsaI, BsaJI, BsaMI, BsaWI, BsaXI, BseDI, BseGI, BseJI, BseLI, BseMI, BseMII, BseNI, BseRI, BseSI, BseXI, BseYI, BsgI, Bsh1236I, Bsh1285I, BshNI, BshTI, BsiEI, BsiHKAI, BsiWI, BslI, BsmAI, BsmBI, BsmFI, BsmI, BsoBI, Bsp19I, Bsp120I, Bsp1286I, Bsp1407I, Bsp143I, Bsp143II, Bsp68I, BspCNI, BspDI, BspEI, BspHI, BspLI, BspMI, BspPI, BspQI, BspTI, BsrBI, BsrDI, BsrFI, BsrGI, BsrI, BsrSI, BssHII, BssKI, BssSI, Bst1107I, Bst98I, BstAPI, BstBI, BstEII, BstF5I, BstNI, BstOI, BstUI, BstXI, BstYI, BstZI, BstZ17I, Bsu15I, Bsu36I, BsuRI, BtgI, BtgZI, BtsCI, BtsI, BveI, Cac8I, CaiI, CfoI, Cfr10I, Cfr13I, Cfr42I, Cfr9I, CfrI, ClaI, CpoI, Csp45I, Csp6I, CspI, CspCI, CvialI, CviKI-1, CviQI, DdeI, DpnI, DpnII, DraI, DraIII, DrdI, EaeI, EagI, Eam1104I, Eam1105I, EarI, EciI, Ecl136II, EclHKI, Eco105I, Eco130I, Eco147I, Eco24I, Eco31I, Eco32I, Eco47I, Eco47III, Eco52I, Eco57I, Eco57MI, Eco72I, Eco81I, Eco88I, Eco91I, EcoICRI, EcoNI, EcoO109I, EcoP15I, EcoRI, EcoRV, EheI, Esp3I, FatI, FauI, Fnu4HI, FokI, FseI, FspI, FspAI, GsuI, HaeII, HaeIII, HgaI, HhaI, HinII, Hin4I, Hin6I, HincII, HindIII, HinfI, HinPII, HpaI, HpaII, HphI, Hpy166II, Hpy188I, Hpy188III, Hpy8I, Hpy99I, HpyAV, HpyCH4III, HpyCH4IV, HpyCH4V, HpyF10VI, Hsp92I, Hsp92II, I-PpoI, KasI, Kpn2I, KpnI, KspAI, LweI, MbiI, MboI, MboII, MfeI, MisI, MluI, MlyI, MmeI, MnlI, Mph1103I, MscI, MseI, MslI, MspAI, MspI, MssI, MunI, Mva1269I, MvaI, MwoI, NaeI, NarI, NciI, NcoI, NdeI, NdeII, NgoMIV, NheI, NheI-HF, NlaIII, NlaIV, NmeAIII, NmuCI, NotI, NruI, NsbI, NsiI, NspI, OliI, PacI, PaeI, Paer7I, PagI, PauI, PciI, PdiI, PdmI, Pfl23II, PflFI, PflMI, PfoI, PhoI, PleI, PmeI, PmlI, PpiI, PpuMI, PshAI, PsiI, Psp1406I, Psp5II, PspGI, PspOMI, PspXI, PstI, PsuI, PsyI, PvuI, PvuII, PvuII-HF, RsaI, RsrII, SacI, SacII, SalI, SalI-HF, SapI, SatI, Sau3AI, Sau96I, SbfI, ScaI, ScaI-HF, SchI, ScrFI, SdaI, SduI, SexAI, SfaNI, SfcI, SfiI, SfoI, SgfI, SgrAI, SinI, SmaI, SmiI, SmlI, SmuI, SnaBI, SpeI, SphI, SphI-HF, SspI, StuI, StyD4I, StyI, SwaI, TaaI, TaiI, Taq aI, TaqI, TasI, TatI, TauI, TfiI, TliI, TruII, Tru9I, TseI, Tsp45I, Tsp509I, TspMI, TspRI, Tth111I, TurboNaeI, TurboNarI, Van9II, VspI, XagI, XapI, XbaI, XceI, XcmI, XhoI, XhoII, XmaI, XmaJI, XmiI, XmnI, 및 ZraI로 이루어진 그룹으로부터 선택된 제한 효소에 의해 인식되는 부위를 포함한다. 양태는 또한 I-SceI, I-CeuI 및 PI-PspI와 같은 호밍 엔도뉴클레아제(homing endonuclease)를 포함한다. 이런 효소에 대한 상응하는 절단 부위는 당업계에 공지되어 있다.

[0149] 일부 실시태양에서, 본 발명은 길이가 8개 이상의 뉴클레오타이드(≥ 8 제한 효소)를 인식하는 회귀 제한 효소의 사용을 교시한다. 일부 실시태양에서, 본 발명은 각각의 cTAG에서 단일 회귀 제한 부위의 사용을 교시한다. 다

른 실시태양에서, 본 발명의 cTAG는 둘 이상의 제한 부위를 포함할 수 있다. 하기 표 1은 본 발명에 따른 cTAG의 목록을 제공하며, 각각 회귀 제한 효소 부위가 굵게 표시되어 있다.

표 1

[0150] 실시예 cTAG 서열, CRISPR 랜딩 부위 및 회귀한 제한 효소 부위(굵은 서열 부분은 제한 효소 부위이다)

TagA	ACTGGGTGGAATCCCTTCTGCAGCACCTGGATT ACCCTGTTATCCCTAGT ACTGGGTGGAATCCCTTCTGCAGCACCTGGATT ACCCTGTTATCCCTAGT	I-SceI	SEQ ID NO: 65
TagB	TAATGAGTAGTCCCTCATCTCCCTCAAGCAGG GCGCGGCGGTACTGCCATC TAATGAGTAGTCCCTCATCTCCCTCAAGCAGG GCGCGGCGGTACTGCCATC	MreI	SEQ ID NO: 66
TagC	CATATAATCTCCCTCAAGCAGGCCCGCTGG GCGCGCGAATGTTAGGAA	MauBI	SEQ ID NO: 67
TagD	GCCTATAATGTGAAGAGCTTCACTGAGTAGG GCCCGGGCTGTAAACGGTT	SrfI	SEQ ID NO: 68
TagE	ATTCGCTAGCAGATGTAGTGTTCACAGGG GCGATCGTGATATGGGTC	AsiSI	SEQ ID NO: 69
TagF	ACTACCTAGCTGCATTTTCAGGAGGAAGCGATGG GCGGCGCACACCTTC	NotI	SEQ ID NO: 70
TagG	TGATAATGGGTGAGTGAGTGTGTGCGTGTGG GCGCGCCAGATGGGAACA	AscI	SEQ ID NO: 71
TagH	ACTCCAGTCTTTCTAGAAGAT TGGCAAACAGCTATTATGGGTATTATGGGT	PI-PspI	SEQ ID NO: 72
TagI	TAGTGGACGGGGCCACTAGGGACAGGATTGG CCTGCAGGATCCCGTCAA	SbfI	SEQ ID NO: 73
TagJ	TGAACCTAAGGCGGCTGCACAACCAAGTGGAGGCCTAAATGATC	없음	SEQ ID NO: 74

[0151] 일부 실시태양에서, 본 발명에서 사용하기 위한 적합한 재조합 부위는 attB 부위, attP 부위, attL 부위, attR 부위, lox 부위, psi 부위, tnpI 부위, dif 부위, cer 부위, frt 부위 및 이들의 돌연변이체, 변이체 및 유도체를 포함하나 이에 제한되지 않는다. 본 발명의 특정 실시태양에서, 토포아이스머라제 인식 부위는, 존재한다면, 유형 IB 토포아이스머라제 일 수 있는 유형 I 토포아이스머라제에 의해 인식되고 결합된다. 유형 IB 토포아이스머라제의 적합한 유형은 진핵 생물 핵 유형 I 토포아이스머라제 및 폭스 바이러스 토포아이스머라제를 포함하나 이에 제한되지 않는다. 일부 실시태양에서, 적절한 유형의 폭스 바이러스 토포아이스머라제는 우두 바이러스, 쇼프 섬유휘종 바이러스, ORF 바이러스, 계두 바이러스, 물사마귀 바이러스 및 암삭타 모레리엔토모폭스바이러스(Amsacta morreientomopoxvirus)와 같은 바이러스에 의해 생성되거나 이로부터 분리된 폭스바이러스 토포아이스머라제를 포함하나 이에 제한되지 않는다.

[0152] 일부 실시태양에서, CRISPR 및 비-CRISPR 복제 부위의 cTAG 배열은 사용자 선호에 따라 주문할 수 있다. 일부 실시태양에서, 본 발명은 CRISPR 결합 부위가 삽입체 부분으로부터 가장 멀리 떨어져 있도록 주문되어야 한다고 교시한다. 하나의 예시적인 실시태양에서, cTAG는 5'-3':(부분 I)-[R1-A1-C-A2-R2]-(부분 II)로부터 다음과 같이 배열될 수 있으며, 여기서 R = 제한 부위, A = 재조합효소 부위 및 C = CRISPR 랜딩 부위. 일부 실시태양에서, C는 다수의 중첩, 또는 순차적인 CRISPR 및/또는 제한 랜딩 부위를 포함할 수 있다. 일부 실시태양에서, 본 발명의 cTAG 상의 복제 부위의 배열은 대칭(즉, 복제 부위의 유형의 대칭적인 순서를 제공한다)일 것이다.

[0153] 다른 실시태양에서, 본 발명의 cTAG 상의 복제 부위의 배열은 비대칭일 수 있다. 예를 들어, 다른 예시적인 실시태양에서, cTAG는 5'-3':(부분 I)-[R1-A1-C1-C2]-(부분 II)로부터 다음과 같이 배열될 수 있으며, 여기서 R = 제한 부위, A = 재조합효소 부위 및 C1-2 = CRISPR 랜딩 부위(들). 또 다른 실시태양에서, cTAG는 5'-3': i) (부분 I)-[R1-C1-C2]-(부분 II), ii) (부분 I)-[R1-C1]-(부분 II), iii) (부분 I)-[C1-C2]-(부분 II)로부터 다음과 같이 배열될 수 있거나 이들의 역순서로 배열될 수 있으며, 여기서 R = 제한 부위, A = 재조합효소 부위 및 C1-2 = CRISPR 랜딩 부위(들).

[0154] 당업자는 다양한 cTAG 배열의 이점 및 응용을 인식할 것이다. 예를 들어, 단일 태그 실시태양에서, 모듈식 구조체는 단일 CRISPR 엔도뉴클레아제를 소화시키면서 삽입할 수 있지만(추가 cTAG의 추가 분해 없이는) 삽입의 제거 또는 교체를 허용하지 않을 것인데, 이는 제 2 플랭킹 cTAG 부위의 부족 때문이다. 일부 실시태양에서, 본 발명은 삽입된 부분들이 cMCS 내의 가능한 삽입체 부분 위치의 수를 확장하기 위해 추가 cTAG를 자체적으로 포함할 수 있음을 교시한다.

[0155] 다른 실시태양에서, 본 발명은 모듈식 CRISPR 구조체로부터 하나 이상의 삽입체 부분을 제거하는 방법을 교시한다. 일부 실시태양에서, 모듈식 CRISPR 구조체의 둘 이상의 cTAG는 양립 가능한 말단을 생성할 수 있는 제한 효소 결합 부위를 포함한다. 일부 실시태양에서, 제한 효소 부위는 동일하다. 다른 실시태양에서, 제한 효소 부위는 별개이지만, 상기 부위의 최종 분해는 혼성화 및 결합을 위한 양립 가능한 말단을 생성한다. 일부 실시태양에서, 모듈식 CRISPR 구조체의 일부의 결실을 위한 제한 부위는 둘 이상의 cTAGS의 다른 말단에 위치하여, 최종

결찰 구조체는 cTAGs에 대한 삽입체 부분의 동일한 비율을 여전히 유지할 것이다.

- [0156] 일부 실시태양에서, 본 발명은 본 발명의 모듈식 CRISPR 구조체 내의 결실에 사용되는 제한 효소 부위가 양립 가능한 말단을 생성하는 임의의 제한 효소일 수 있음을 교시한다. 다른 실시태양에서, 본 발명은 본 발명의 모듈식 CRISPR 구조체 내의 결실에 사용되는 제한 효소 부위가 양립 가능한 말단을 생성하는 임의의 회귀 8≥ 염기 제한 효소일 수 있음을 교시한다. 선택된 실시태양에서, 본 발명의 모듈식 CRISPR 구조체 내의 결실에 사용되는 제한 효소 부위가 I-SceI 및 PI-PspI일 수 있음을 교시한다.
- [0157] 일부 실시태양에서, 본 발명은 cTAG 쌍을 생성하기 위해 각각의 삽입체 부분에 인접한 2개의 cTAG를 갖는 모듈식 CRISPR 구조체를 교시한다. 일부 실시태양에서, 상기 cTAG 쌍은 삽입체 부분의 선택적 절단/교체를 허용한다. 예를 들어, 도 2b에 예시된 바와 같이, cTAG A 및 B를 표적으로 하는 엔도뉴클레아제에 의한 모듈식 CRISPR 플라스미드의 분해가 삽입체 부분 2의 특이적 제거를 초래할 수 있다.
- [0158] 전술한 바와 같이, 본 발명의 선택된 실시태양은 엔도뉴클레아제 절단 후 cTAG 기능을 회복시키는 대체 삽입체 부분을 제공한다. 따라서, 도 2b에 도시된 바와 같이, 대체 삽입체 부분 2a-2d는 모듈식 CRISPR 플라스미드 내로 삽입시 cTAG A 및 B 기능을 회복시킬 서열을 포함한다.
- [0159] 일부 실시태양에서, 본 발명은 cTAG가 또한 삽입체 부분 방향성을 제어할 수 있음을 교시한다. 삽입체 부분에서 cTAG 말단 및 모듈식 CRISPR 구조체에서 절단된 cTAG 사이의 서열 상동성은, 상동성 재조합 또는 혼성화(예를 들어, 김슨 접근법에서)를 통한 Cas9 절단된 서열에 대한 삽입 방향성을 결정할 것이다. Cpf1 서열의 삽입 방향성은 cTAG 상의 Cpf1 점착성 말단의 왓슨 크릭 혼성화를 통해 또한 조절될 수 있다.
- [0160] 일부 실시태양에서, 본 발명은 또한 대안적인 cTAG 배열을 제공한다. 예를 들어, 일부 실시태양에서, 본 발명의 모듈식 CRISPR 구조체는 중첩된 cTAG의 사용을 위한 기능성을 제공하도록 디자인될 수 있다.
- [0161] 일부 실시태양에서, 본 발명은 생체의 및 생체내 다성분 조립을 가능하게 하는 공유된 중첩 "태그" 영역에 기초한 성분-기초 CRISPR 조립을 교시한다. 일부 실시태양에서, 본 발명의 태그는 DNA 구조체로부터 미래의 복제 또는 생체의 DNA 조립을 용이하게 하기 위한 CRISPR 랜딩 부위를 포함한다. DNA 구조체가 숙주 유기체의 게놈에 통합되면, 미리 선택된 Cas9 또는 Cpf1 랜딩 부위는 쉽게 유전자 변형을 용이하게 할 수 있다. 한 세트의 실험에서, 조립 전략은 많은 수와 유형의 DNA 성분을 함유하는 여러 유기체에서 사용될 수 있는 DNA 플라스미드의 제작을 가능하게 한다.
- [0162] 일부 실시태양에서, 이 조립 전략은 대사 경로를 비롯한 임의의 원하는 세트의 DNA 성분을 암호화하는 플라스미드를 조립하고 빠르게 재조립하는 데 사용될 수 있다. 다른 실시태양에서, 통합 플라스미드로 cTAG를 디자인하는 것은 또한 숙주 유기체의 게놈 내외로 DNA 성분을 직접 교체하여 미래 플라스미드를 복제할 필요성을 회피하는 데 사용될 수 있다.
- [0163] **cTAG 서열 디자인 알고리즘**
- [0164] 일부 실시태양에서, 본 발명은 cTAG 내의 CRISPR 랜딩 부위를 용이하게 하도록 디자인된 알고리즘을 교시한다. 일부 실시태양에서, CRISPR 랜딩 부위는 기존 서열로부터 동정된 서열이다. 따라서, 일부 실시태양에서, 본 발명은 소프트웨어 프로그램의 사용이 원하는 가이드 서열 길이 및 특정 CRISPR 효소에 대한 CRISPR 모티프 서열(PAM, 프로토스페이서 인접 모티프)에 기초하여 입력 DNA 서열의 양 가닥 상의 후보 CRISPR 표적 서열을 동정하도록 디자인된다. 예를 들어, PAM 서열 TTN을 갖는 프란시셀라 노비시다(*Francisella novicida*) U112로부터의 Cpf1에 대한 표적 부위는 입력 서열 및 입력 서열의 역보체 모두에서 5'-TTN-3'을 검색함으로써 동정될 수 있다. PAM 서열을 갖는 라치노스피라세(*Lachnospiraceae*) 박테리아와 아시다미노콕쿠스 에스피(*Acidaminococcus sp.*)로부터의 Cpf1에 대한 표적 부위는 입력 서열 및 입력 서열의 역보체 모두에서 5'-TTN-3'을 탐색함으로써 동정될 수 있다. 마찬가지로, PAM 서열 NNAGAAW를 갖는 에스 썬모필루스(*S. thermophilus*) CRISPR1의 Cas9에 대한 표적 부위는 입력 서열 및 입력 서열의 역보체 모두에서 5'-Nx-NNAGAAW-3'을 검색함으로써 동정될 수 있다. 에스 피로젠스(*S. pyogenes*)의 Cas9에 대한 PAM 서열은 5'-NGG-3'이다.
- [0165] 마찬가지로, PAM 서열 NGGNG를 갖는 에스 썬모필루스(*S. thermophilus*) CRISPR의 Cas9에 대한 표적 부위는 입력 서열 및 입력 서열의 역보체 모두에서 5'-N, -NGGNG-3'을 검색함으로써 동정될 수 있다.
- [0166] 다른 실시태양에서, 본 발명은 처음부터 CRISPR 랜딩 부위를 디자인하는 방법을 교시한다. 당업자는 본 발명의 가이드 RNA와 함께 CRISPR 랜딩 부위를 쉽게 디자인할 수 있으며, 여기서 최종 프로토스페이서 서열은, 전술한 바와 같이, 원하는 CRISPR 엔도뉴클레아제에 적합한 PAM 모티프와 결합된다.

- [0167] 일부 실시태양에서, 본 발명은 SEQ ID NO. 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 78, 79, 80, 81 및 이의 조합으로 이루어진 그룹으로부터 선택된 서열을 포함하는 cTAG를 교시한다.
- [0168] DNA 표적 부위의 게놈에서의 다수의 발생은, 모든 잠재적 부위를 동정한 후에, 비특이적 게놈 편집을 유도할 수 있기 때문에, 본 발명은, 일부 실시태양에서, 관련 참조 게놈 또는 모듈식 CRISPR 구조체에서 나타나는 횃수에 기초하여 서열을 필터링하는 것을 교시한다. 서열 특이성이 '씨드'서열(Cpf1 매개 분할에 대한 가이드 서열의 처음 5nt와 같은)에 의해 결정되는 CRISPR 효소의 경우, 필터링 단계는 또한 동일한 씨드를 갖는 상이한 서열을 필터 아웃할 수 있다.
- [0169] 일부 실시태양에서, 알고리즘 틀은 또한 특정 가이드 서열에 대한 표적 부위로부터의 잠재적 가능성을 확인할 수 있다. 예를 들어, 일부 실시태양에서, Cas-Offinder는 Cpf1에 대한 표적 부위로부터의 잠재적 가능성을 확인하는데 사용될 수 있다(Kim *et al.*, 2016. "Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells" published online June 06, 2016 참조). Zhang 연구실의 crispr.mit.edu 틀을 포함하여 임의의 다른 공개적으로 이용 가능한 CRISPR 디자인/동정 틀이 사용될 수 있다(Hsu, *et al.* 2013 "DNA targeting specificity of RNA-guided Cas9 nucleases" Nature Biotech 31, 827-832 참조).
- [0170] 일부 실시태양에서, 사용자는 씨드 서열의 길이를 선택할 수 있다. 또한 사용자는 필터를 통과하기 위해 게놈에 있는 씨드:PAM 서열의 발생 횃수를 지정할 수 있다. 기본값은 독특한 서열을 선별하는 것이다. 여과 수준은 씨드 서열의 길이와 게놈에서 서열의 발생 횃수를 변화시킴으로써 변경된다. 프로그램은, 부가적으로 또는 선택적으로, 동정된 표적 서열(들)의 역보체를 제공함으로써 보고된 표적 서열(들)에 상보적인 가이드 서열의 서열을 제공할 수 있다.
- [0171] **모듈식 CRISPR DNA 구조체 복제**
- [0172] 일부 실시태양에서, 본 발명은 본 발명의 모듈식 CRISPR DNA 구조체를 사용하여 신규 재조합 핵산 분자를 제조하는 방법을 교시한다. 일부 실시태양에서, 본 발명은 DNA 부분 조립 방법을 교시한다. 각 방법에 대한 설명은 아래에 제공된다.
- [0173] **DNA 조립 방법**
- [0174] 일부 실시태양에서, 본 발명은 DNA 부분의 모듈 조립 방법을 교시한다. 일부 실시태양에서, 본 발명의 DNA 조립 방법은 생체외에서 수행된다. 따라서, 일부 실시태양에서, 본 발명은 i) 적어도 하나의 CRISPR 복합체와 함께 적어도 2개의 삽입체 부분 DNA를 포함하는 혼합물을 형성하는 단계 및 ii) 상기 혼합물을 삽입체 DNA의 CRISPR 분해를 위한 조건에서 배양하는 단계 iii) 2개의 삽입체 부분 DNA의 각각의 분해로부터 양립 가능한 점착성 말단을 혼성화하는 단계, 및 iv) 상기 혼성화된 말단을 서로 결합시켜 새로운 재조합 핵산을 생성시키는 단계를 교시한다. 따라서, 일부 실시태양에서, 본 발명의 삽입체 부분 DNA는 함께 분해된다. 다른 실시태양에서, 본 발명은 동일하거나 상이한 CRISPR 복합체로 각 삽입체 부분 DNA를 개별적으로 분해하는 방법을 교시한다. 일부 실시태양에서, 적어도 하나의 삽입체 부분은 CRISPR 복합체에 의해 분해되지 않는다. 일부 실시태양에서, 본 발명은 (후술되는 절에서 기술된 바와 같이 이중 CRISPR 분해를 위한) 엑소뉴클레아제 처리가 단계 iii)의 혼성화 전에 수행됨을 교시한다.
- [0175] 또 다른 실시태양에서, 본 발명은 삽입체 부분 말단을 ssDNA 엑소뉴클레아제에 노출시키고 최종 점착성 말단을 혼성화하고 폴리머라제에 의한 선택적 충전과 결합에 의해 삽입체 부분의 김슨형 결합을 교시한다. 일부 실시태양에서, 하나 이상의 삽입체 부분은 ssDNA 엑소뉴클레아제 처리 전에 dsDNA 엑소뉴클레아제에 노출된다. 일부 실시태양에서, 본 발명은 하나 이상의 CRISPR 엔도뉴클레아제(예를 들어, 후술되는 절에서 기술된 바와 같은 이중 CRISPR 분해)에 의해 분해된 삽입체 부분 또는 모듈식 CRISPR 벡터의 김슨형 결합을 교시한다.
- [0176] 아래의 절은 본 발명의 삽입체 부분 및 모듈식 CRISPR 구조체가 조립되고 편집될 수 있는 다양한 방식을 설명하는 일련의 예시적인 실시예를 제공한다. 아래 기술된 기술의 목록은 본 발명의 서열의 유용성을 강조하는 예시적인 일련의 실시예를 제공하나, 이에 제한되는 것은 아니다. 당업자는 본 발명에 따른 삽입체 부분의 조립 및 편집을 허용하는 다른 기술을 인식할 것이다.
- [0177] 일부 실시태양에서, 본 발명은 Cpf1 및/또는 Cas9 CRISPR 엔도뉴클레아제를 포함하는 방법을 기술한다. 이러한 특정 CRISPR 엔도뉴클레아제를 언급하는 것은 예시적인 것이고 청구항에 명시된 경우를 제외하고는 제한적이지 않다. 당업자는 다른 기존의 또는 지금까지 발견되지 않은 CRISPR 엔도뉴클레아제의 본 발명의 구조체 및 방법의 응용가능성을 즉시 인식할 것이다. Cpf1에 대한 언급은 점착성 DNA 말단을 생성하기 위해 지그재그형 DNA

절단을 촉매할 수 있는 임의의 현재 공지된 또는 알려지지 않은 CRISPR 엔도뉴클레아제의 사용을 포괄하는 것으로 해석될 수 있다. Cas9에 대한 언급은 유사하게 dsDNA의 평활 말단 절단을 촉매할 수 있는 임의의 현재 공지된 또는 알려지지 않은 CRISPR 엔도뉴클레아제의 사용을 포괄하는 것으로 해석될 수 있다.

[0178] **생체의 Cpf1**

[0179] 일부 실시태양에서, 본 발명의 생체의 DNA 조립은 아래 기술된 바와 같이 Cpf1 CRISPR 복합체로 수행된다. 첫 번째, 2개 이상의 삽입체 부분은 적어도 2개의 삽입체 부분 사이에서 공통적인 cTAG를 표적으로 하는 Cpf1 CRISPR 복합체와 함께 배양된다. 일부 실시태양에서, 삽입체 부분은 단일 혼합물로 함께 배양된다. 다른 실시태양에서, 삽입체 부분은 상이한 혼합물에서 배양된다.

[0180] 두 번째, 일부 실시태양에서, 분해된 생성물은 정제되어 활성 CRISPR 뉴클레아 제를 제거한다. 일부 실시태양에서, 정제는 분해된 삽입체 부분으로부터 활성 Cpf1 복합체를 분리하는 것을 포함한다. 일부 실시태양에서, 정제는 겔 또는 컬럼 정제와 같은 DNA 정제를 통해 달성될 수 있다. 다른 실시태양에서, 정제는 열 또는 화학적 불활성화와 같은 Cpf1 불활성화에 의해 달성될 수 있다.

[0181] 세 번째, 분해된 삽입체 부분은 Cpf1 복합체에 의해 생성된 양립 가능한 점착성 말단의 혼성화에 적합한 조건에서 배양된다. 그런 후에 혼성화된 말단은 본 발명의 초기 부분에서 기술된 것을 포함하는 임의의 공지된 결합 방법에 따라 결합된다.

[0182] **생체의 Cas9**

[0183] 다른 실시태양에서, 본 발명의 생체의 DNA 조립은 아래 기술된 바와 같은 Cas9 CRISPR 복합체로 수행된다. 첫 번째, 2개 이상의 삽입체 부분은 적어도 2개의 삽입체 부분 사이에 공통적인 cTAG를 표적으로 하는 Cas9 CRISPR 복합체와 함께 배양된다. 일부 실시태양에서, 삽입체 부분은 단일 혼합물로 함께 배양된다. 다른 실시태양에서, 삽입체 부분은 상이한 혼합물로 배양된다.

[0184] 두 번째, 일부 실시태양에서, 분해된 생성물을 정제하여 활성 CRISPR 뉴클레아제를 제거한다. 일부 실시태양에서, 정제는 분해된 삽입체 부분으로부터 활성 Cas9 복합체를 분리하는 것을 포함한다. 일부 실시태양에서, 이는 겔 또는 컬럼 정제와 같은 DNA 정제를 통해 달성될 수 있다. 다른 실시태양에서, 정제는 열 또는 화학적 불활성화와 같은 Cas9 불활성화에 의해 달성될 수 있다.

[0185] 일부 실시태양에서, Cas9 분해 생성물에 대한 세 번째 단계는 평활 말단 결합에 적합한 조건으로 삽입체 부분을 배양하는 것이다.

[0186] **이중 CRISPR 조립**

[0187] 다른 실시태양에서, 본 발명은 또한 적어도 하나의 공유된 cTAG 서열을 갖는 CRISPR-분해된 삽입체의 조각을 조립하기 위한 김슨-조립 유형 방법(예를 들어, 상이한 CRISPR 랜딩 부위에서 분해된 양립 가능한 cTAG의 조립)을 교시한다. 따라서, 일부 실시태양에서, 본 발명은 아래 기술된 바와 같은 이중 CRISPR 분해 조립을 교시한다.

[0188] 먼저, 2개 이상의 삽입체 부분은 적어도 2개의 삽입체 부분 사이에서 공통적인 상기 cTAG 내의 각 부분에 인접한 2개의 상이한 CRISPR 랜딩 부위를 표적으로 하는 2개의 CRISPR 복합체와 함께 배양된다.

[0189] 일부 실시태양에서, 2개의 상이한 CRISPR 랜딩 부위는 함께 분해된다. 다른 실시태양에서, 하나의 삽입체 부분 DNA는 하나의 CRISPR 랜딩 부위를 표적으로 하는 하나의 CRISPR 복합체로 분해되고 다른 삽입체 부분 DNA는 별도의 용기에서 제 2 CRISPR 랜딩 표적 부위를 표적으로 하는 상이한 CRISPR 복합체로 분해된다. 각각의 경우에, 이러한 분해의 결과는 2개의 삽입체 DNA cTAG의 각각에 공유된 cTAG가 서로 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20bp의 서열 중복을 포함하는 것이 될 것이다.

[0190] 예를 들어, 예시적인 실시태양에서, 2개의 삽입체 DNA 부분 사이의 공유 cTAG는 5'-3': (부분 I)-[R1-C1-C2]-(부분 II)로부터 다음과 같이 배열될 것이며, R = 제한 부위, C1 = 제 1 CRISPR 랜딩 부위 및 C2 = 제 2 CRISPR 랜딩 부위. 이 예시적인 실시태양에서, 3' 공유된 cTAG를 갖는 제 1 삽입체 DNA 부분은 C2를 표적으로 하는 CRISPR 복합체로 분해되고 5' 공유된 cTAG를 갖는 제 2 삽입체 DNA 부분은 C1을 표적으로 하는 CRISPR 복합체로 분해될 것이다. 이것은 C1-C2에 걸친 중복 서열을 갖는 2개의 DNA체 삽입 부분을 생성할 것이다.

[0191] 두 번째, 일부 실시태양에서, 분해된 생성물은 정제되어 활성 CRISPR 뉴클레아제를 제거한다. 일부 실시태양에서, 정제는 분해된 삽입체 부분으로부터 활성 CRISPR 복합체의 분리를 수반한다. 일부 실시태양에서, 이는 겔 또는 컬럼 정제와 같은 DNA 정제를 통해 달성될 수 있다. 다른 실시태양에서, 정제는 열 또는 화학적 불활성화

와 같은 CRISPR 불활성화에 의해 달성될 수 있다.

[0192] 세 번째로, 일부 실시태양에서, CRISPR-분해된 삽입체 부분은 ssDNA 엑소뉴클레아제와 함께 배양하여 2개의 삽입체 DNA 부분 사이에 중첩하는 점착성 말단을 생성한다.

[0193] 넷째, 분해된 삽입체 부분은 CRISPR 복합체/엑소뉴클레아제 분해에 의해 생성된 양립 가능한 점착성 말단의 혼성화에 적합한 조건하에 배양한다. 그런 후에 혼성화된 말단은 본 발명의 초기 부분에서 기술된 것을 포함하는 임의의 공지된 결찰 방법에 따라 결찰된다. 일부 실시태양에 있어서, 혼성화된 부분은 중합 효소와 함께 배양되어 결찰 전에 임의의 누락된 서열 간격을 채운다.

[0194] **가교 조립**

[0195] 다른 실시태양에서, 본 발명은 삽입체 부분의 분해된 cTAG 서열과 중첩하는 가교 서열을 포함하는 제 3 DNA 서열의 첨가를 통해 Cas9 분해된 부분의 겹손-조립을 교시한다.

[0196] 이 예시적인 예에서, 두 삽입체 부분은 동일한 CRISPR 랜딩 부위를 표적으로 하는 동일한 Cas9 CRISPR 복합체로 분해된다. 이 실시태양에서, 최종 분해된 cTAG는 서열 중첩을 갖지 않을 것이다. 따라서, 일부 실시태양에서, 제 3 단계는 Cas9 분해된 삽입체 부분이 ssDNA 엑소뉴클레아제로 추가로 분해되어 3' 또는 5' 오버행을 생성하는 것이다. 그런 후에 엑소뉴클레아제 분해된 삽입체 부분은 CRISPR 복합체 및 엑소뉴클레아제 분해와 가교 서열의 조합에 의해 생성된 양립 가능한 점착성 말단의 혼성화에 적합한 조건에서 배양된다. 그런 후에 혼성화된 말단은 본 발명의 초기 부분에서 기술된 것을 포함하는 임의의 공지된 결찰 방법에 따라 결찰된다. 일부 실시태양에서, 본 발명의 엑소뉴클레아제 분해는 제 2 단계 전에 수행된다.

[0197] **생체의 HDR**

[0198] 다른 실시태양에서, 본 발명은 Cas9 또는 Cpf1 엔도뉴클레아제로 분해된 삽입체 부분 DNA의 말단을 HDR 복합체와 조립하여, 상기 분해된 삽입체 부분의 재조합을 개시하는 생체의 방법을 교시한다.

[0199] **생체내 상동성 재조합**

[0200] 일부 실시태양에서, 본 발명의 생체내 DNA 조립은 아래 기술된 바와 같이 Cpf1 또는 Cas9 CRISPR 복합체로 수행된다. 한 실시태양에서, 적어도 하나의 공유된 cTAG를 갖는 둘 이상의 삽입체 부분이 숙주 세포 내로 도입된다. 일부 실시태양에서, 상동성 공유된 cTAG 서열을 갖는 DNA 삽입체 부분의 존재는 상동성 재조합 조립(예를 들어, 효모 상동성 재조합)을 유발하기에 충분할 것이다.

[0201] 예를 들어, 일부 실시태양에서, 2개의 삽입체 DNA 부분 사이의 적어도 하나의 공유된 cTAG 서열은 선형 구조체를 생성하도록 조립될 수 있다. 이러한 예시적인 실시태양에서, 2개의 나머지 외측 cTAG는 또한 세포 내 다른 벡터의 cTAG와 재조합되도록 디자인될 수 있다(예를 들어, 기존 플라스미드 또는 염색체 내로의 삽입). 다른 실시태양에서, 2개의 부분은 2개의 삽입체 DNA 부분 사이에 제 2 공유된 cTAG의 재조합을 통해 원형 구조체로 추가로 조립될 수 있다. 조립된 구조체는 조립에 사용된 유기체에서 사용될 수 있거나, 일부 실시태양에서, 제 2 유기체(예를 들어, 효모의 조립 및 박테리아로의 후속 변형)로 정제되고 변형될 수 있다.

[0202] 다른 실시태양에서, 공유된 cTAG를 갖는 하나 이상의 삽입체 부분은 숙주 세포 내로 도입되기 전에 분해될 수 있다. 따라서, 일부 실시태양에서, 본 발명은 방출된 부분의 생체내 조립 전에 큰 벡터로부터 삽입체 부분을 방출시키는 CRISPR 분해를 교시한다. 일부 실시태양에서, 분해는 Cas9로 수행된다. 다른 실시태양에서, 분해는 Cpf1로 수행된다. 다른 실시태양에서, 분해는 제한 엔도뉴클레아제로 수행된다. 일부 실시태양에서, 삽입체 부분의 CRISPR 분해는 생체외에서 수행된다. 일부 실시태양에서, 분해된 생성물은 정제되어 조립 숙주 세포로의 삽입체 부분의 형질전환 전에 활성 CRISPR 엔도뉴클레아제를 제거한다.

[0203] 일부 실시태양에서, 정제 단계는 겔 또는 컬럼 정제와 같은 DNA 정제를 통해 달성될 수 있다. 다른 실시태양에서, 정제는 열 또는 화학적 불활성화와 같은 CRISPR 불활성화에 의해 달성될 수 있다.

[0204] **생체내 결찰**

[0205] 일부 실시태양에서, 본 발명은 CRISPR 엔도뉴클레아제에 의한 재절단으로부터 삽입체 부분을 보호하는 방법을 교시한다. 일부 실시태양에서, 본 발명의 삽입체 부분은 DNA 서열의 화학적 변형을 통해 엔도뉴클레아제 절단으로부터 보호될 수 있다. 예를 들어, 일부 실시태양에서, 본 발명은 포스포로티오에이트 올리고뉴클레오타이드를 교시한다.

- [0206] 일부 실시태양에서, 본 발명의 방법은 다중 부분 DNA 조립에 특히 유용하다.
- [0207] 본 명세서의 도 2a는 본 발명의 방법에 따른 다중 부분 DNA 조립의 예시적인 예를 제공한다. 이 예에서, 2개의 cTAG(태그 A-H)가 있는 일련의 8개의 DNA 부분(부분 1-8)은 생체외에서 결합된 다음 자체 조립될 수 있다(상기한 바와 같이 생체내 상동성 재조합 또는 결합).
- [0208] **DNA 편집 방법**
- [0209] 일부 실시태양에서, 본 발명은 모듈식 CRISPR DNA 구조체의 편집 방법을 교시한다. 일부 실시태양에서, 본 발명의 DNA 편집 방법은 전술한 DNA 조립 방법의 동일한 원리를 적용하지만, 하나 이상의 미리 존재하는 모듈식 CRISPR DNA 구조체를 편집하는 목적으로 수행한다.
- [0210] 일부 실시태양에서, 본 발명의 DNA 편집 방법은 생체외에서 수행된다. 따라서, 일부 실시태양에서, 본 발명은 i) 모듈식 CRISPR DNA 구조체 및 적어도 하나의 삽입체 DNA 부분을 적어도 하나의 CRISPR 복합체와 함께 포함하는 혼합물을 형성하는 단계, ii) 상기 혼합물을 삽입체 DNA의 cTAG 및 이의 상응하는 모듈식 CRISPR DNA 구조체 cTAG의 CRISPR 분해를 위한 조건에서 배양하는 단계, iii) 상기 cTAG 각각의 분해에 의해 생성된 양립 가능한 점착성 말단(Cpf1인 경우)을 혼성화하는 단계 및 iv) 상기 혼성화 말단(또는 Cas9가 사용되는 경우, 평활 말단)을 서로 결합하여 새로운 재조합 핵산을 생성하는 단계를 교시한다. 일부 실시태양에서, 엑소 뉴클레아제 처리 단계 iii)의 혼성화(아래의 절에서 기술된 바와 같은 이중 CRISPR 분해를 위한) 전에 수행된다. 일부 실시태양에서, 본 발명의 분해는 삽입체 부분 DNA 및 모듈식 CRISPR DNA 구조체에 대해 개별적으로 수행된다. 일부 실시태양에서, 모듈식 CRISPR DNA 구조체만이 CRISPR 복합체로 분해된다.
- [0211] **생체외 Cpf1**
- [0212] 일부 실시태양에서, 본 발명의 생체외 DNA 편집 방법은 아래 기술된 바와 같은 Cpf1 CRISPR 복합체를 사용하여 수행한다. 첫 번째, 모듈식 CRISPR DNA 구조체와 적어도 하나의 삽입체 DNA 부분은 삽입체 부분의 cTAG 및 모듈식 CRISPR DNA 구조체 내의 상응하는 태그를 표적으로 하는 Cpf1 CRISPR 복합체로 배양된다. 일부 실시태양에서, 모듈식 CRISPR DNA 및 삽입체 부분 DNA의 분해는 개별 반응으로 수행된다.
- [0213] 두 번째, 일부 실시태양에서, 분해된 생성물은 정제되어 활성 CRISPR 뉴클레아제를 제거한다. 일부 실시태양에서, 정제는 분해된 뉴클레오타이드로부터 활성 Cpf1 복합체를 분리하는 것을 포함한다. 일부 실시태양에서, 정제는 겔 또는 컬럼 정제와 같은 DNA 정제를 통해 달성될 수 있다. 다른 실시태양에서, 정제는 열 또는 화학적 불활성화와 같은 Cpf1 불활성화에 의해 달성될 수 있다.
- [0214] 세 번째, 분해된 모듈식 CRISPR DNA 구조체 및 삽입체 부분은 Cpf1 복합체에 의해 생성된 양립 가능한 점착성 말단의 혼성화에 적합한 조건에서 배양된다. 그런 후에 혼성화된 말단은 본 발명의 초기 부분에서 기술된 것을 포함하는 임의의 공지된 결합 방법에 따라 결합된다.
- [0215] **생체외 Cas9**
- [0216] 다른 실시태양에서, 본 발명의 생체외 DNA 편집 방법은 아래 기술된 바와 같은 Cas9 CRISPR 복합체를 사용하여 수행된다. 첫 번째, 모듈식 CRISPR DNA 구조체와 적어도 하나의 삽입체 DNA 부분은 삽입체 부분의 cTAG 및 모듈식 CRISPR DNA 구조체 내의 상응하는 태그를 표적으로 하는 Cas9 CRISPR 복합체로 배양된다. 일부 실시태양에서, 모듈식 CRISPR DNA 및 삽입체 부분 DNA의 분해는 개별 반응으로 수행된다.
- [0217] 두 번째, 일부 실시태양에서, 분해된 생성물은 정제되어 활성 CRISPR 뉴클레아제를 제거한다. 일부 실시태양에서, 정제는 분해된 뉴클레오타이드로부터 활성 Cas9 복합체를 분리하는 것을 포함한다. 일부 실시태양에서, 정제는 겔 또는 컬럼 정제와 같은 DNA 정제를 통해 달성될 수 있다. 다른 실시태양에서, 정제는 열 또는 화학적 불활성화와 같은 Cas9 불활성화에 의해 달성될 수 있다.
- [0218] 일부 실시태양에서, Cas9 분해 생성물에 대한 세 번째 단계는 평활 말단 결합에 적합한 조건에서 삽입체 부분을 배양하는 것이다.
- [0219] **깁슨 편집**
- [0220] 다른 실시태양에서, 본 발명은 또한 손상되지 않은 중첩 cTAG 서열을 함유하는 CRISPR-분해된 구조체 및/또는 분해되지 않은 삽입체 부분의 서열을 편집하기 위한 깁슨-조립 유형 방법을 교시한다. 따라서, 일부 실시태양에서, 제 3 단계는 Cas9 분해된 모듈식 CRISPR DNA 구조체 및 삽입체 부분(들)이 ssDNA 엑소뉴클레아제로 추가로 분해되어 3' 또는 5' 오버행을 생성하는 것이다. 일부 실시태양에서, 본 발명은 ssDNA 분해 전에 비-CRISPR 분

해진 삽입체 부분을 단축시키는 dsDNA 엑소뉴클레아제 분해를 교시한다.

[0221] 그런 후에 엑소뉴클레아제 분해된 DNA 절편은 CRISPR 복합체와 엑소뉴클레아제 분해의 조합에 의해 생성된 양립 가능한 점착성 말단의 혼성화에 적합한 조건하에서 배양된다. 그런 후에 혼성화된 말단은 본 발명의 초기 부분에서 기술된 것을 포함하는 임의의 공지된 결찰 방법에 따라 결찰된다. 일부 실시태양에서, 혼성화된 DNA는 중합 효소와 함께 배양되어 결찰 전에 누락된 DNA 절편을 채운다. 일부 실시태양에서, 본 발명의 엑소뉴클레아제 분해는 CRISPR 불활성화 단계 전에 수행된다.

[0222] 일부 실시태양에서, 분해된 서열의 결찰은 생체외에서 발생할 수 있다.

[0223] 다른 실시태양에서, 본 발명은 Cas9 또는 Cpf1 엔도뉴클레아제로 분해된 모듈식 CRISPR DNA 구조체 및 적어도 하나의 분해되지 않은 삽입체를 HDR 복합체로 조립하여 상기 분해된 모듈식 CRISPR DNA 구조체 및 적어도 하나의 삽입체 DNA 부분의 재조합을 개시하는 생체외 방법을 교시한다 .

[0224] 본 발명의 DNA 편집 방법의 일부 실시태양에서, DNA 삽입체 부분은 제 2 모듈식 CRISPR DNA 구조체 내에 포함된다. 따라서, 일부 실시태양에서, 본 발명의 DNA 편집 방법은 하나의 모듈식 CRISPR DNA 구조체로부터 다른 구조체로 DNA 삽입체 부분의 전달을 포함한다.

[0225] 발현, 정제 및 배달

[0226] 일부 실시태양에서, 본 발명은 CRISPR 복합체를 암호화하는 벡터, 구조체 및 핵산 서열의 방법 및 조성물을 교시한다. 일부 실시태양에서, 본 발명은 Cas9 또는 Cpf1 단백질의 형질전환 또는 일시적인 발현을 위한 플라스미드를 교시한다. 일부 실시태양에서, 본 발명은 리가아제, 링커 및 NLS를 포함하나 이에 제한되지 않는, 본 발명에 기술된 하나 이상의 다른 폴리펩타이드의 단백질 융합을 위한 인 프레임 서열을 포함하는 키메라 Cas9 또는 Cpf1 단백질을 암호화하는 플라스미드를 교시한다.

[0227] 일부 실시태양에서, 본 발명의 플라스미드 및 벡터는 Cas9/Cpf1 단백질을 암호화하고 또한 본 발명의 crRNA/tracrRNA/sgRNA 및/또는 도너 삽입체 서열을 암호화할 것이다. 다른 실시태양에서, 조작된 복합체의 상이한 성분은 하나 이상의 별개의 플라스미드에서 암호화될 수 있다.

[0228] 일부 실시태양에서, 본 발명의 플라스미드는 여러 종에 걸쳐 사용될 수 있다. 다른 실시태양에서, 본 발명의 플라스미드는 형질전환되는 유기체에 맞추어진다. 일부 실시태양에서, 본 발명의 서열은 이의 유전자가 편집되고 있는 유기체에서 발현하도록 코돈-최적화될 것이다. 당업자는 유전자 편집을 위한 적절한 발현을 제공하는 프로모터의 사용의 중요성을 인식할 것이다. 일부 실시태양에서, 상이한 종에 대한 플라스미드는 상이한 프로모터를 필요로 할 것이다.

[0229] 일부 실시태양에서, 본 발명의 플라스미드 및 벡터는 관심 세포에서 선택적으로 발현된다. 따라서, 일부 실시태양에서, 본 발명은 이소성 프로모터, 조직-특이적 프로모터, 발달 조절된 프로모터, 또는 유도성 프로모터의 사용을 교시한다. 일부 실시태양에서, 본 발명은 또한 터미네이터 서열의 사용을 교시한다.

[0230] 일부 실시태양에서, 본 발명은 또한 Cpf1 및/또는 Cas9 엔도뉴클레아제 단백질을 발현 및 정제하는 방법을 교시한다. 일부 실시태양에서, 본 발명은 본 발명의 단백질이 상업적으로 입수 가능한 단백질 생산 및 정제 키트 또는 서비스 중 임의의 것에 의해 생산될 수 있음을 교시한다. 예를 들어, 일부 실시태양에서, 본 발명은 Cas9 및/또는 Cpf1을 폴리히스티딘(His), 글루타티온 s-트랜스퍼라제(GST) 또는 다른 정제 태그 키메라 융합으로 벡터 속에 복제하는 방법을 교시한다. 일부 실시태양에서, 본 발명은 다양한 원핵 생물 및 진핵 생물, 및 무세포 단백질 생산 시스템을 교시한다. 예를 들어, 일부 실시태양에서, 본 발명은 대장균 BL21에서 단백질 발현 플라스미드의 발현을 교시한다. 일부 실시태양에서, 단백질 생산 시스템은 단백질 독성의 영향을 감소시키도록 유도성 일 것이다. 예를 들어, 일부 실시태양에서, 본 발명은 IPTG 또는 아라비노스 유도 시스템을 사용하는 방법을 교시한다.

[0231] 일부 실시태양에서, 본 발명은 또한 친화성 태그(His-Nickel, GST-Glutathione 등)를 포함하는 다양한 단백질 정제 방법을 교시한다. 일부 실시태양에서, 본 발명은 단백질 정제를 위한 천연 조건 및 변성 조건 모두를 교시한다.

[0232] 다른 실시태양에서, 본 발명은 GenScript®, ThermoFisher® 및 NovoProtein®을 포함하나 이에 제한되지 않는 하나 이상의 단백질 생산 서비스를 통한 Cas9 및/또는 Cpf1의 생산을 교시한다.

[0233] 형질전환

- [0234] 일부 실시태양에서, 본 발명은 본 발명에 개시된 플라스미드 및 벡터의 형질전환의 사용을 교시한다. 당업자는 본 발명의 플라스미드가 본 명세서의 다른 부분에서 기술된 바와 같이 임의의 공지된 시스템을 통해 세포로 형질전환될 수 있음을 인식 할 것이다. 예를 들어, 일부 실시태양에서, 본 발명은 입자 폭격, 화학적 형질전환, 아그로박테리움 형질전환, 나노-스파이크 형질전환, 전기천공 및 바이러스 형질전환에 의한 형질전환을 교시한다.
- [0235] 일부 실시태양에서, 본 발명의 벡터는 형질전환, 형질감염, 형질도입, 바이러스 감염, 유전자 총 또는 Ti 매개 유전자 전달을 포함하는 다양한 기술 중 임의의 것을 사용하여 숙주 세포에 도입될 수 있다. 특정 방법은 칼슘 포스페이트 형질감염, DEAE-텍스트란 매개 형질감염, 리포펙션 또는 전기천공을 포함한다(Davis, L., Dibner, M., Battey, I. 1986 "Basic Methods in Molecular Biology"). 다른 형질전환 방법은, 예를 들어, 아세트산 리튬 형질전환 및 전기천공을 포함한다. 예를 들어, Gietz *et al.*, *Nucleic Acids Res.* 27:69-74 (1992); Ito *et al.*, *J. Bacteriol.* 153:163-168 (1983); 및 Becker and Guarente, *Methods in Enzymology* 194:182-187 (1991)를 참조한다. 일부 실시태양에서, 형질전환된 숙주 세포는 재조합 숙주 균주로 불린다.
- [0236] 일부 실시태양에서, 본 발명은 본 발명의 96-웰 플레이트 로봇틱스 플랫폼 및 액체 취급 기계를 사용하여 세포의 고효율 형질전환을 교시한다.
- [0237] 일부 실시태양에서, 본 발명은 외인성 단백질(Cpf1/Cas9 및 DNA 리가아제), RNA(crRNA/tracRNA/가이드 RNA) 및 DNA(삽입체 DNA 부분 또는 모듈식 CRISPR 구조체)를 세포 내로 얻는 방법을 교시한다. 단백질/RNA/DNA의 직접 형질감염 또는 DNA 형질전환과 RNA 및 단백질의 세포내 발현을 포함하는 이를 달성하기 위한 다양한 방법은 이전에 기술되어있다(Dicarlo, J. E. *et al.* "Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems." *Nucleic Acids Res* (2013). doi:10.1093/nar/gkt135; Ren, Z. J., Baumann, R. G. & Black, L. W. "Cloning of linear DNAs in vivo by overexpressed T4 DNA ligase: construction of a T4 phage hoc gene display vector." *Gene* 195, 303-311 (1997); Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. "Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery." *Elife* 3, e04766 (2014)).
- [0238] 일부 실시태양에서, 본 발명은 상기한 바와 같은 하나 이상의 선택 마커로 형질전환된 세포를 선별하는 것을 교시한다. 이런 일 실시태양에서, 카나마이신 내성 마커(KanR)를 포함하는 벡터로 형질전환된 세포는 유효량의 카나마이신 항생제를 함유하는 배지 상에 플레이트된다. 카나마이신-레이스(kanamycin-laced) 배지에서 볼 수 있는 콜로니 형성 단위는 벡터 카세트를 계놈에 통합시킨 것으로 추정된다. 원하는 서열의 삽입은 PCR, 제한 효소 분석 및/또는 관련 삽입 부위의 서열 분석을 통해 확인될 수 있다.
- [0239] 다른 실시태양에서, 본 발명의 일부 또는 전체 복합체는 세포에 직접 전달될 수 있다. 따라서, 일부 실시태양에서, 본 발명은 본 발명의 폴리펩타이드 및 핵산의 발현 및 정제를 교시한다. 당업자는 단백질 및 핵산을 정제하는 많은 방법을 인식할 것이다. 일부 실시태양에서, 폴리펩타이드는 박테리아 시스템, 효모 시스템, 식물 세포 시스템 또는 동물 세포 시스템과 같은 유도성 또는 구조성 단백질 생산 시스템을 통해 발현될 수 있다. 일부 실시태양에서, 본 발명은 또한 친화성 태그 또는 맞춤형 항체 정제를 통한 단백질 및/또는 폴리펩타이드의 정제를 교시한다. 다른 실시태양에서, 본 발명은 또한 폴리뉴클레오타이드에 대한 화학적 합성 방법을 교시한다.
- [0240] 일부 실시태양에서, 당업자는 유전자 발현을 위한 바이러스 벡터 또는 플라스미드가 본 발명에 개시된 복합체를 전달하는 데 사용될 수 있다는 것을 인식할 것이다. 바이러스-유사 입자(VLP)는 리보뉴클레오타이드 복합체를 캡슐화하는데 사용될 수 있고, 본 발명에 개시된 정제된 리보뉴클레오타이드 복합체는 정제되고 전기 천공 또는 주사를 통해 세포로 전달될 수 있다.
- [0241] **키트**
- [0242] 일부 실시태양에서, 본 발명은 상기 방법 및 조성물에 개시된 임의의 하나 이상의 요소를 함유하는 키트를 제공한다. 일부 실시태양에서, 키트는 모듈식 CRISPR DNA 구조체 및 키트 및 임의의 필요한 시약 또는 반응물의 사용 지침을 포함한다. 일부 실시태양에서, 벡터 시스템은, 본 출원에서 상기한 바와 같이, (a) 모듈식 CRISPR DNA 구조체, (b) CRISPR 엔도뉴클레아제 단백질 및 필요한 표적 가이드 RNA(들)(또는 상기 항목을 암호화하는 서열)을 포함하는 CRISPR 복합체 및 선택적으로 (c) 삽입체 DNA 부분을 포함한다.
- [0243] 요소는 개별적으로 또는 조합으로 제공될 수 있으며, 바이알, 병 또는 튜브, 또는 숙주 세포 또는 플라스미드와 같은 임의의 적합한 용기에 제공될 수 있다. 일부 실시태양에서, 키트는 하나 이상의 언어, 예를 들어 하나 초과 언어의 지침을 포함한다.

- [0244] 일부 실시태양에서, 키트는 본 발명에 기술된 하나 이상의 요소(예를 들어, 정제된 Cpf1 엔도뉴클레아제)를 이용하는 공정에서 사용하기 위한 하나 이상의 시약을 포함한다. 시약은 임의의 적절한 용기에 제공될 수 있다. 예를 들어, 키트는 하나 이상의 반응 또는 저장 버퍼를 제공할 수 있다. 시약은 특정 분석에서 가능하거나 사용 전에 하나 이상의 다른 성분의 첨가를 필요로 하는 형태(예를 들어, 농축액 또는 동결건조 형태)로 제공될 수 있다. 완충액은 탄산나트륨 완충액, 중탄산나트륨 완충액, 붕산염 완충액, 트리스 완충액, MOPS 완충액, HEPES 완충액 및 이들의 조합을 포함하나 이에 제한되지 않는 임의의 완충액일 수 있다. 일부 실시태양에서, 완충액은 알칼리성이다. 일부 실시태양에서, 완충액은 약 7 내지 약 10의 pH를 갖는다. 일부 실시태양에서, 키트는 crRNA 서열 및 조절 요소를 작동 가능하게 연결시키기 위해 벡터에 삽입하기 위한 crRNA 서열에 상응하는 하나 이상의 올리고뉴클레오타이드를 포함한다.
- [0245] **실시예**
- [0246] 다음의 실시예는 본 발명의 다양한 실시태양을 설명하기 위해 제공되며 어떠한 방식으로든 본 발명을 제한하려는 것은 아니다. 청구항의 범위에 의해 정의된 바와 같이, 본 발명의 취지 내에 포함되는 본 발명의 변화 및 다른 용도는 당업자에게 자명할 것이다.
- [0247] **실시예 1: 원-포트 생체의 모듈식 CRISPR 복제**
- [0248] 이 실시예는 원 포트 반응에서 하나의 플라스미드로부터 다른 플라스미드로 삽입체를 전달함으로써 플라스미드 13001009086(SEQ ID NO: 82)의 생성을 기술한다. 도 4를 참조.
- [0249] 두 플라스미드는 관심 부위에 인접하는 복제 태그를 운반한다(cTAG K [SEQ ID NO: 78] /cTAG L [SEQ ID NO: 79] 및 cTAG K' [SEQ ID NO: 80] /cTAG L' [SEQ ID NO: 81]). 편집된 플라스미드 대한 복제 반응을 유도하기 위해, Cpf1 스페이서는 수용자 및 도너 플라스미드(각각 K/K' 및 L/L') 상에서 반대 방향에 있다. 이 인사이드-아웃(inside-out)/아웃사이드-인(outside-in) 분해는 최종 생성물의 Cpf1 스페이서를 제거하여, 원하는 생성물의 재결단을 제거한다('485 플라스미드의 인사이드-아웃 및 '784 플라스미드의 아웃사이드-인 분해를 묘사하는 도 4의 곡선 화살표 참조). Cas9 스페이서는 유지되어, 이 부위에서 반복 편집을 가능하게 한다. 따라서, 메가 모듈식 구조체는 반복 편집을 가능하게 하는 신속한 단일-포트 반응 방식을 허용한다.
- [0250] Cpf1 단백질은 젠스트립(Genscript)에 의해 합성하였고, crRNA는 쉐인트코(Synthego)에 의해 합성하였다. crRNA(crRNA 1 및 crRNA 3)로 복합체화된 Cpf1 단백질을 ATP를 함유하는 완충액 속 플라스미드(13000789485 - SEQ ID NO: 83 and 13000823784 - SEQ ID NO: 84) 및 DNA 리가아제에 첨가하였다. 이런 성분을 절단과 결합에 최적화된 온도에서 회수하였다.
- [0251] 반응물을 대장균으로 형질전환시키고 양성 클론을 서열 분석하여 새로운 삽입체의 삽입 및 Cpf1 스페이서의 손실을 확인하였다.
- [0252] 결실을 위해, 사용된 복제 태그 내의 Cpf1 부위는 플라스미드 밀봉을 허용하기 위해 양립 가능한 오버행을 생성해야 한다. cTAG L'은 2개의 Cpf1 스페이서를 포함하도록 디자인되었는데, 하나는 오버행이 cTAG K'와 양립 가능하지 않은 경우에 삽입을 위한 것이고 두 번째는 오버행이 cTAG K'와 양립 가능한 경우에 결실을 위한 것이다.
- [0253] **실시예 2: 생체의 모듈식 CRISPR 복제**
- [0254] 이 실시예는 CRISPR 클로닝의 유연성을 입증하기 위해 디자인되었다. 초기 단계로서, 키나마이신 또는 클로람페니콜 내성 유전자를 암호화하는 몇 가지 내성 플라스미드를 각각 소스 벡터 pzHR039(SEQ ID No: 100) 및 13000223370(SEQ ID No: 101)으로부터 생성하였다. 카나마이신 내성 플라스미드는 각각 GFP 유전자의 측면에 위치하는 다양한 Cpf1 랜딩 부위를 포함하도록 디자인하였다(분해시, 이들 플라스미드는 "카나마이신 내성 플라스미드 백본"을 생성한다). 클로람페니콜 내성 플라스미드는 각각 클로람페니콜 내성 유전자의 측면에 위치하는 다양한 Cpf1 랜딩 부위를 포함하도록 디자인하였다(분해시, 이들 플라스미드는 "클로람페니콜 내성 삽입체"를 생성한다). 이 실시예에서 사용된 각각의 플라스미드에 대한 서열 및 벡터 지도가 표 2에 개시되어 있다.
- [0255] 각각의 카나마이신 및 클로람페니콜 내성 플라스미드는 초기에 타입 II 제한 효소 KpnI-HF 및 PvuI-HF(모두 NEB로부터 상업적으로 입수 가능함)로 선형화시켰다. 각 플라스미드 상의 KpnI 및 PvuI 제한 부위의 위치는 도 7 내지 도 14에 제공된 벡터 지도에 기재되어 있다. 선형화 후 저항성 플라스미드는 더 이상 박테리아 숙주 시스템에서 자체 복제가 불가능하였다.
- [0256] 그런 후에 선형화된 내성 플라스미드를 60 µL의 반응물에서 15 µg(최종 농도 1.58 µM)의 Cpf1 효소와 2 µL의 5

μM 의 각각의 가이드 RNA(최종 농도 $0.167\ \mu\text{M}$)와 혼합하여 활성 CRISPR 복합체를 형성하였다.

[0257] 본 실시예에서 사용된 Cpf1 효소는 IDT로부터 구입하였다. Cpf1은 아시다미노코쿠스 에스피(*Acidaminococcus sp.*) Cpf1(AsCpf1)로부터 얻었다. 효소는 1개의 N-말단 핵 편재 서열(NLS) 및 1개의 C-말단 NLS뿐만 아니라 3개의 N-말단 FLAG 태그 및 C-말단 6-His 태그를 포함하도록 추가로 변형시켰다.

[0258] 이 실시예에서 사용된 가이드 RNA는 IDT로부터 맞춤 주문하였다. 각 가이드 RNA는 선형화된 내성 플라스미드 내에 위치한 상이한 CRISPR 랜딩 부위를 표적으로 하도록 디자인하였다. 이 실시예에서, 백본 플라스미드의 Cpf1 랜딩 부위는 절제되었지만 삽입체의 결찰시 복원되었다. 표 2는 사용된 각 가이드 RNA의 가이드 서열 부분을 제공한다. 혼합물 내의 CRISPR 복합체는 카나마이신 내성 플라스미드 백본을 생성하기 위해 각각의 카나마이신 내성 플라스미드로부터 GFP 유전자를 절단하도록 디자인하였다(도 5, 제 2 패널 참조). 혼합물 내의 CRISPR 복합체는 또한 클로람페니콜 내성 플라스미드로부터 클로람페니콜 내성 유전자를 절단하여 클로람페니콜 내성 삽입체를 생성하도록 디자인하였다(도 5, 제 2 패널 참조). 각 반응의 카나마이신 내성 플라스미드 백본 및 클로람페니콜 내성 삽입체는 말단의 혼성화를 초래하여 "이중 내성" 카나마이신 및 클로람페니콜 플라스미드를 생성하는 양립 가능한 오버행을 생성하도록 유사하게 디자인하였다.

[0259] Cpf1 및 가이드 RNA를 포함하는 선형화된 내성 플라스미드 혼합물을 제조자의 권장된 Cpf1 완충액에서 섭씨 37도에서 3시간 동안 배양시켰다. 선택된 반응을 아가로오스 겔 상에서 수행하고 생성된 단편을 표준 DNA 추출 키트(제조사의 지침에 따라 사용된 Zymo Research 키트)를 사용하여 정제 하였다. 정제(대조군) 및 비 정제(테스트).

[0260] 카나마이신 내성 플라스미드 백본 및 클로람페니콜 내성 삽입체를 포함하는 DNA 단편은 T4 DNA 리가아제(NEB로부터 구입 가능)를 갖는 또는 갖지 않는 새로운 반응에서 결합되어 NEB10-B 세포(NEB로부터 구입 가능)로 형질 전환되었다. 형질전환된 세포를 기능적 내성 플라스미드를 함유하지 않은 임의의 세포의 성장을 방지하도록 디자인된 카나마이신 및 클로람페니콜로 증강된 배지 상에 플레이트팅하였다.

[0261] 개별 콜로니를 서열 분석하도록 보내서 Cpf1 복제 접합부를 확인하였다. 회수된 콜로니는 또한 표 2에 기술된 프라이머를 사용하여 PCR을 통해 확증하였다. 도 5는 상기한 바와 같이 Cpf1 분해 전에 플라스미드를 선형화시킨 것을 제외하고는 상기한 일반적인 실험 디자인을 예시한다.

표 2

[0262] 이 실시예 2에 사용된 서열 목록

성분	설명	SEQ ID NO
GFP Cpf1 cTAG M fwd	5' CAGCACCTGGATTACCCCTGTTATCCCTAGTTTGGGTTAAAGATGGTTAAA TGATTCGAAAATAATAAAGGGAAAATCA 3'	SEQ ID No: 86
GFP Cpf1 cTAG N fwd	5' CAGCACCTGGATTACCCCTGTTATCCCTAGTTTGGGATGTTAAGAGTCCT ATCTTCGAAAATAATAAAGGGAAAATCA 3'	SEQ ID No: 87
GFP Cpf1 cTAG P fwd	5' CAGCACCTGGATTACCCCTGTTATCCCTAGTTTGGAGAGTGTTCAGTCTCC GTGAACTCGAAAATAATAAAGGGAAAATCA 3'	SEQ ID No: 88
GFP Cpf1 cTAG O rvs	5' CGCTTCCTCCTGAAAATGCAGCTAGGTAGTTTGTACCGCCCCCCCCATACC CCAATCGACATGCCGAATCAGAAGTGA 3'	SEQ ID No: 89
GFP Cpf1 cTAG N rvs	5' CGCTTCCTCCTGAAAATGCAGCTAGGTAGTTTGGGATGTTAAGAGTCCT ATCTTCGACATGCCGAATCAGAAGTGA 3'	SEQ ID No: 90
CAT01 Cpf1 cTAG M fwd	5' TTTGGGTTAAAGATGGTTAAATGATTCGACATACACATAAAGTAGCTTGCG 3'	SEQ ID No: 91
CAT01 Cpf1 cTAG N fwd	5' TTTGGGATGTTAAGAGTCCCTATCTTCGACATACACATAAAGTAGCTTGCG 3'	SEQ ID No: 92
CAT01 Cpf1 cTAG P fwd	5' TTTGAGGAGTGTTCAGTCTCCGTGAATCGACATACACATAAAGTAGCTTG CG 3'	SEQ ID No: 93

CAT01 Cpf1 cTAG N rvs	5' TTTGGGATGTTAAGAGTCCCTATCTTCGACTGGAAGGACAAGGGGGACC 3'	SEQ ID No: 94
CAT01 Cpf1 cTAG O rvs	5' TTTGACCGCCCCCCCCATACCCCAATCGACTGGAAGGACAAGGGGGACC 3'	SEQ ID No: 95
Cpf1 cTAG M	5' TTTGGGTAAAGATGGTTAAATGAT 3'	SEQ ID No: 96
RNA targeting cTAG M	5' UAAUUUCUACUCUUGUAGAUUGGUUAAAGAUGGUAAAUGAU 3'	SEQ ID NO: 110
Cpf1 cTAG N	5' TTTGGGATGTTAAGAGTCCCTATCT 3'	SEQ ID No: 97
RNA targeting cTAG N	5' UAAUUUCUACUCUUGUAGAUUGGUUAAAGAGUCCCUAUCU 3'	SEQ ID NO: 111
Cpf1 cTAG O	5' TTTGACCGCCCCCCCCATACCCCAA 3'	SEQ ID No: 98
RNA targeting cTAG O	5' UAAUUUCUACUCUUGUAGAUACCGCCCCCCCAUACCCCAA 3'	SEQ ID NO: 112
Cpf1 cTAG P	5' TTTGACCGCCCCCCCCATACCCCAA 3'	SEQ ID No: 99
RNA targeting cTAG P	5' UAAUUUCUACUCUUGUAGAUAGGAGUGUUCAGUCUCCGUGAAC 3'	SEQ ID NO: 113
pzHR039	카나마이신 내성 및 GFP에 대한 원료. 전체 서열 목록 참조.	SEQ ID NO: 100
13000223370	클로람페니콜 내성 및 GFP에 대한 원료. 전체 서열 목록 참조.	SEQ ID NO: 101
pJDI427	GFP Cpf1 cTAGs M 및 N KanR CENARS TRP1 전체 서열 목록 참조.	SEQ ID NO: 102 도 7
pJDI429	GFP Cpf1 cTAGs N 및 O KanR CENARS TRP1 전체 서열 목록 참조.	SEQ ID NO: 103 도 8
pJDI430	GFP Cpf1 cTAGs N 및 P KanR CENARS TRP1 전체 서열 목록 참조.	SEQ ID NO: 104 도 9
pJDI431	GFP Cpf1 cTAGs O 및 P KanR CENARS TRP1 전체 서열 목록 참조.	SEQ ID NO: 105 도 10
pJDI432	pJET AmpR CmR Cpf1 cTAGs M 및 N 전체 서열 목록 참조.	SEQ ID NO: 106 도 11
pJDI434	pJET AmpR CmR blunt Cpf1 cTAGs N 및 O 전체 서열 목록 참조.	SEQ ID NO: 107 도 12
pJDI435	pJET AmpR CmR blunt Cpf1 cTAGs N 및 P 전체 서열 목록 참조.	SEQ ID NO: 108 도 13
pJDI436	pJET AmpR CmR blunt Cpf1 cTAGs O 및 P 전체 서열 목록 참조.	SEQ ID NO: 109 도 14

[0263] *** SEQ ID NOs: 110-113에 대한 가이드 RNA의 밑줄이 없는 부분은 IDT의 화학적으로 변형된 Alt-R RNA이다. 각각의 cTAG(즉, M-P)에 대한 서열의 상동성 영역은 밑줄이 그어져있다.

[0264] 이 실험의 결과를 표 3 및 도 6에 도시한다. 각각의 형질전환에 대한 반응 수를 상단 행을 따라 도시하고, 사용된 가이드 RNA를 표 3의 좌측 열을 따라 나타내었다. 리가아제가 있는 및 없는 동일한 Cpf1 반응의 비교를 리가아제 효소의 존재하에서 9.9배 형질전환체 증가를 나타내었고, 이는 콜로니 성장은 Cpf1 분해 후 이중 카나마이신 및 클로람페니콜 내성 플라스미드의 형성 때문이었다는 것을 나타내었다. 비 리가아제 반응은 반응이 특이적이라는 것을 입증하기 위해 디자인된 일치된 대조군이며 단순히 분해되지 않은 내성 플라스미드의 오염 수준의 존재에 의한 것이 아니었다.

[0265] 16개의 개별 콜로니를 상류 및 하류 클로닝 접합부 모두를 확인하기 위해 생거(Sanger) 서열 분석하였다. 7개의 상류 서열 분석된 접합부 중 7개 및 9개의 하류 접합부 중 8개에서, T4 DNA 리가아제에 의한 반응으로부터의 Cpf1 매개 클론이 충실한 분해와 결합을 나타내었다.

[0266] 반응 71 및 72를 DNA 겔 정제 단계를 거치지 않은 Cpf1 분해 플라스미드로 형질전환시켰다. 그러나 Cpf1 효소는 T4 DNA 리가아제의 첨가 전에 공급자의 지침에 따라 열 비활성화시켰다(반응 72). 반응 71 및 72는 동일한 리가아제 의존성을 나타냈다.

표 3

[0267]

Cpf1 편집 벡터를 포함하는 내성 형질전환체 콜로니

	55	56	59	60	67	68	71*	72*
가이드 M + N	네	네						
가이드 N + O			네	네				
가이드 P + O					네	네	네	네
T4 리가아제	아니오	네	아니오	네	아니오	네	아니오	네
형질전환체의 # (Kan 내성 콜로니)	1	20	0	141	0	12	2	95

[0268]

* 플레이트 71 및 72를 Cpf1 분해 후 DNA 겔 정제를 거치지 않은 분해 #된 DNA로 형질전환하였다.

[0269]

PCT/US2017/042245(WO 2018/013990 A1, 미국 가출원 제62/362,909호 우선권 주장)의 발명은 그 전체가 본 발명에 포함된다.

[0270]

실시예 3: 메가모듈식 디자인을 사용하는 제한 효소 분해 및 결찰에 의한 플라스미드 조립

[0271]

이 실시예는 본 발명의 방법에 따른 모듈식 CRISPR 벡터의 유전자 편집을 기술한다. 도 15는 이 실시예에서 기술된 모듈식 CRISPR 플라스미드 13000444591의 유전자 편집을 예시한다. 플라스미드 백본은 먼저 제작된 플라스미드로부터 "스터 퍼" 삽입체 DNA 부분을 제거함으로써 제조하였다. 스테퍼 삽입체 DNA 부분은 스테퍼 부분의 인접 복제 태그(cTAGs) D (SEQ ID NO: 68) 및 E (SEQ ID NO: 69)를 제한 효소 ApaI 및 PvuI로 분해시킴으로써 제거하였다. 생성된 단편을 겔 전기 영동을 통해 분리하고, 플라스미드 백본에 상응하는 원하는 8.3kb 단편을 겔로부터 절제하고 표준 실리카 막 컬럼을 사용하여 추출하였다.

[0272]

모듈식 CRISPR 벡터에 대한 새로운 삽입체를 생성하기 위하여, cTAG D 및 cTAG E가 측면에 위치하는 원하는 삽입체 DNA 부분을 범용 cTAG oligos tagD_FWD (SEQ ID NO: 75) 및 tagE_REV (SEQ ID NO: 76)를 사용하여 PCR 증폭시켰다. 생성된 삽입체는 cTAG D 및 cTAG E가 측면에 위치한 GFP 마커 유전자를 함유하였다. 생성된 PCR 단편을 cTAG D 내에서 절단하는 ApaI 효소 및 cTAG E 서열 내에서 절단하는 PvuI 효소로 분해하였다. 분해된 삽입체 DNA 부분을 표준 실리카 막 컬럼을 사용하여 정제하였다.

[0273]

정제된 모듈식 CRISPR 벡터 백본 및 삽입체 DNA 부분을 리가아제와의 단일 반응으로 조합하여 원형 플라스미드를 생성하였다. 생성된 편집 GFP-함유 플라스미드 13000444591에 대한 서열은 (SEQ ID NO: 77)에 제공된다.

[0274]

실시예 4: 메가모듈식 디자인을 사용하는 효모 상동성 재조합에 의한 플라스미드 조립

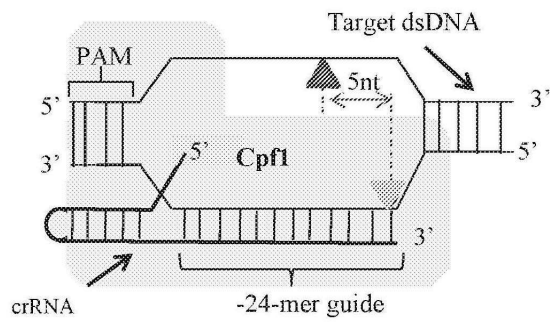
[0275]

플라스미드 13000283399(SEQ ID NO: 85)는 메가모듈식 태그가 측면에 위치하는 PCR 단편의 효모 상동성 재조합에 의해 조립하였다. 조립을 위한 원하는 구조체는 특정 메가모듈식 태그가 측면에 위치하는 방식으로 PCR에 의해 증폭하였다. 이러한 태그는 사카로마이세스 세레비시애(*Saccharomyces cerevisiae*)에서 단편의 방향성 조립을 허용하였는데, 이는 태그 자체가 상동성 재조합을 위한 중첩하는 상동성 영역으로 사용되었기 때문이다. 구체적으로, 5개의 단편은 다음과 같이 메가모듈식 태그가 측면에 위치하게 PCR을 통해 증폭되었다: 태그 A - 단편 1 - 태그 B; 태그 B - 단편 2 - 태그 C; 태그 C - 단편 3 - 태그 D; 태그 D - 단편 4 - 태그 E; 태그 E - 단편 5 - 태그 F. 효모의 복제 기점과 TRP 영양요구성 선별 마커뿐만 아니라 한 쪽 말단에는 태그 A와 다른 쪽에는 태그 F를 포함하는 선형화된 조립 벡터와 함께 이런 단편들을 에스 세레비시애로 형질전환하였다. 원형화되고 조립된 플라스미드는 트립토판이 결여된 배지에서 에스 세레비시애의 성장에 의해 선택하였다. 이들 플라스미드를 회수하여 대장균에서 증폭시켰으며, 서열 분석을 통해 정확한 구조를 확인하였다.

[0276]

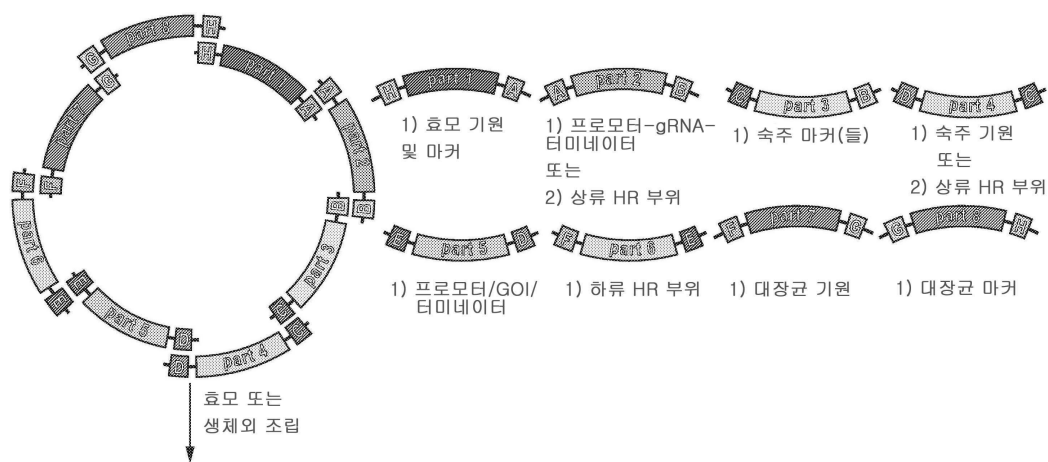
본 발명에 인용된 모든 참고문헌, 기사, 간행물, 특허, 특허 간행물 및 특허 출원은 모든 목적을 위해 그 전문이 참조로 포함된다. 그러나 본 발명에 언급된 모든 참고문헌, 기사, 출판물, 특허, 특허 간행물 및 특허 출원은 세계 어느 나라에서든 유효한 선행 기술을 구성하거나 보통의 일반 지식의 일부를 구성한다는 인정 또는 어떠한 형태의 제안으로 간주되어서는 안 된다.

도면1c

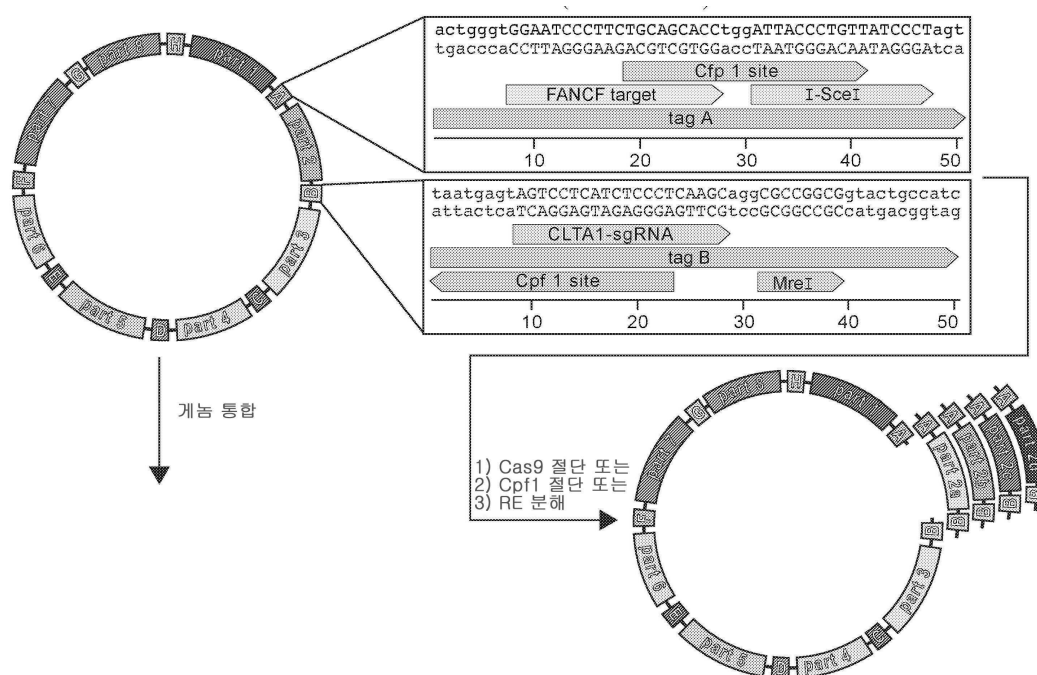


CRISPR- Cpf1

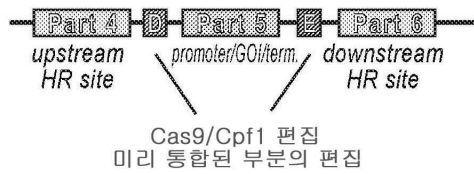
도면2a



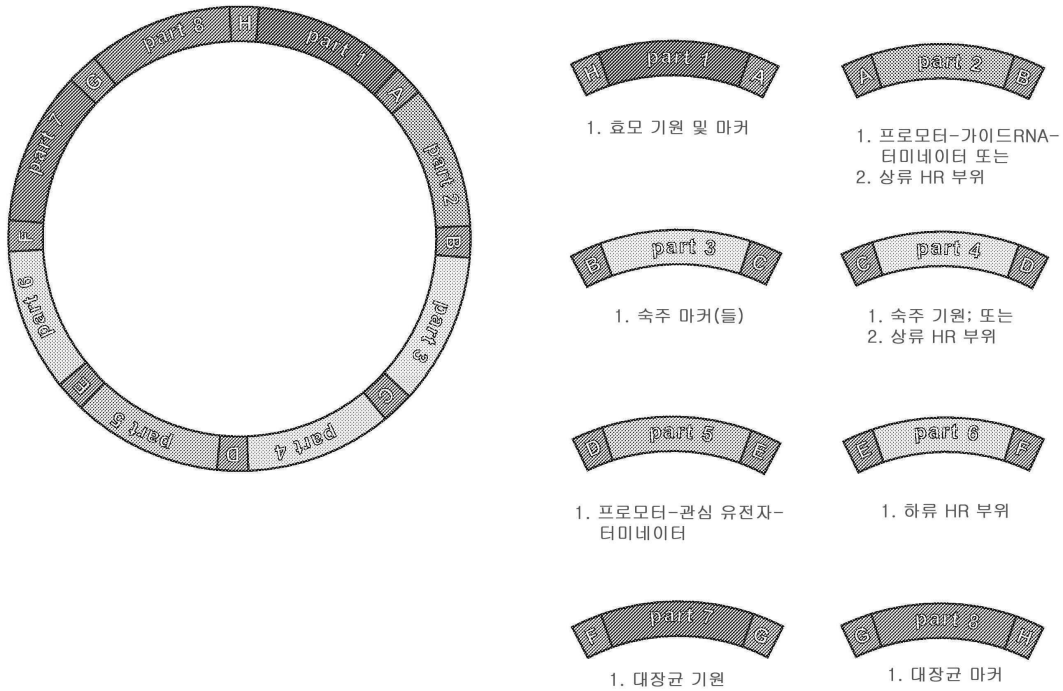
도면2b



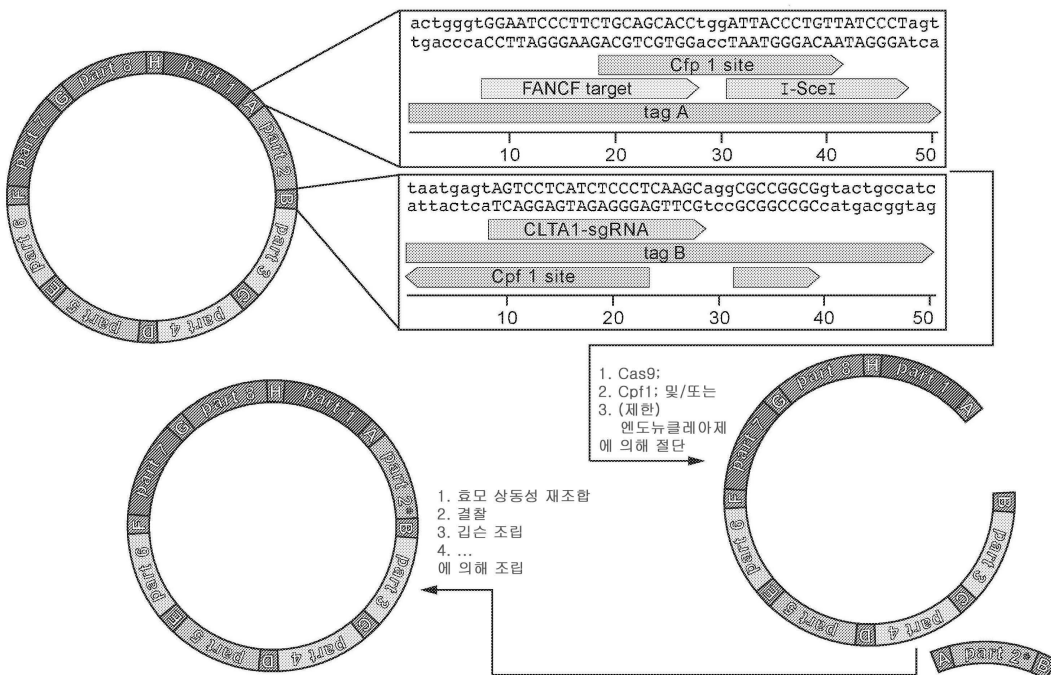
도면2c



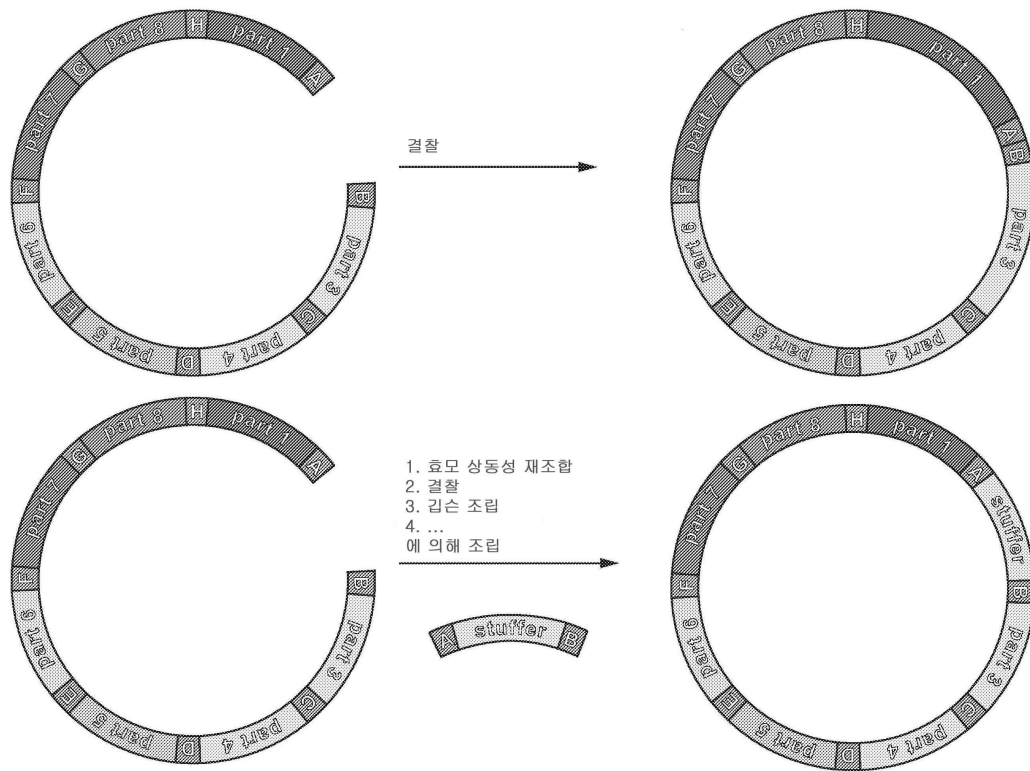
도면3a



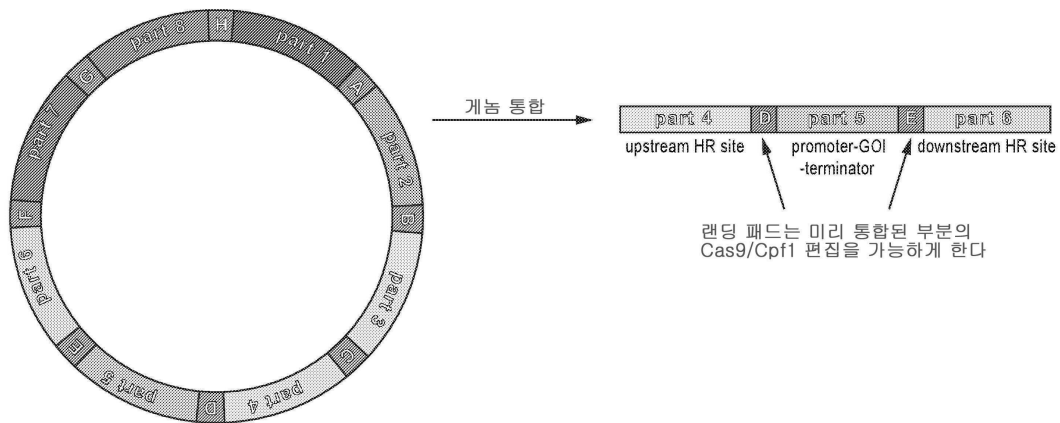
도면3b



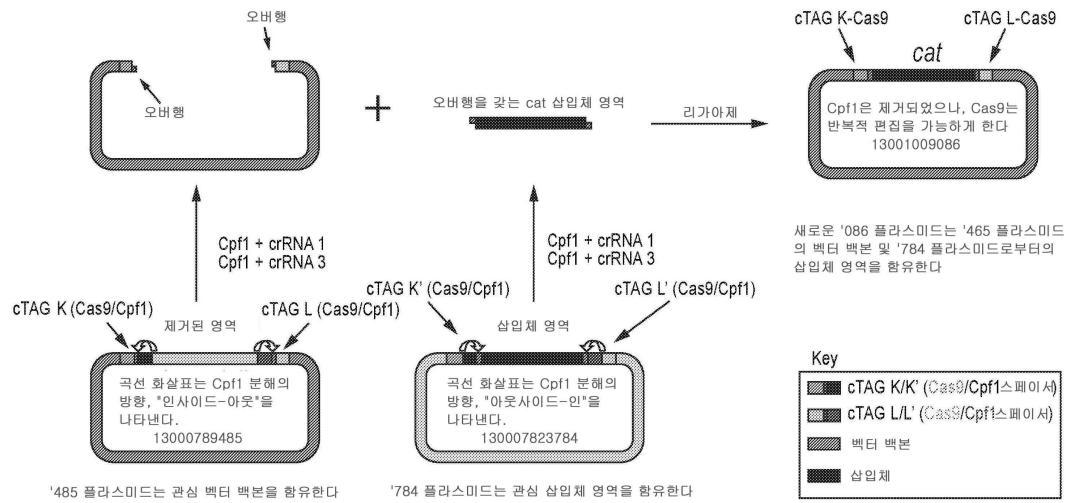
도면3c



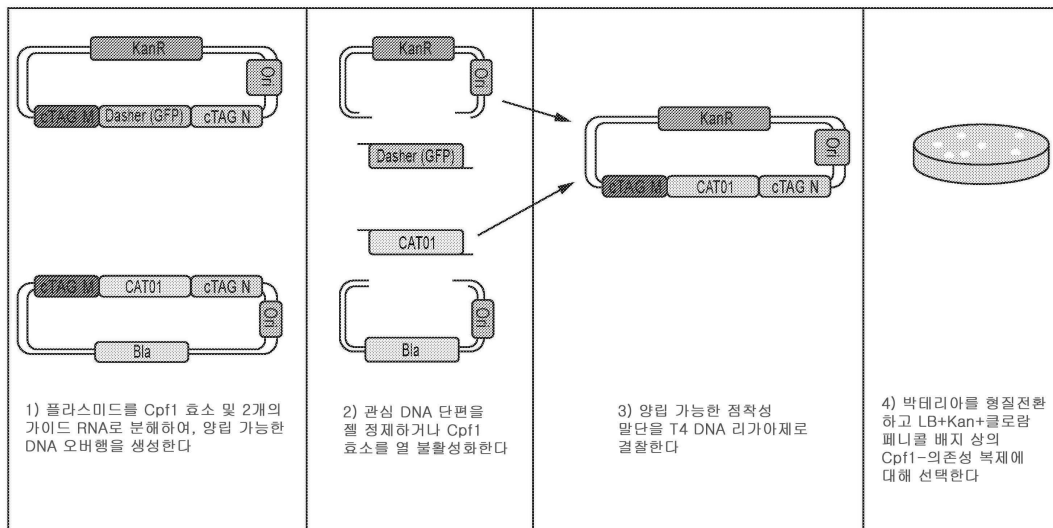
도면3d



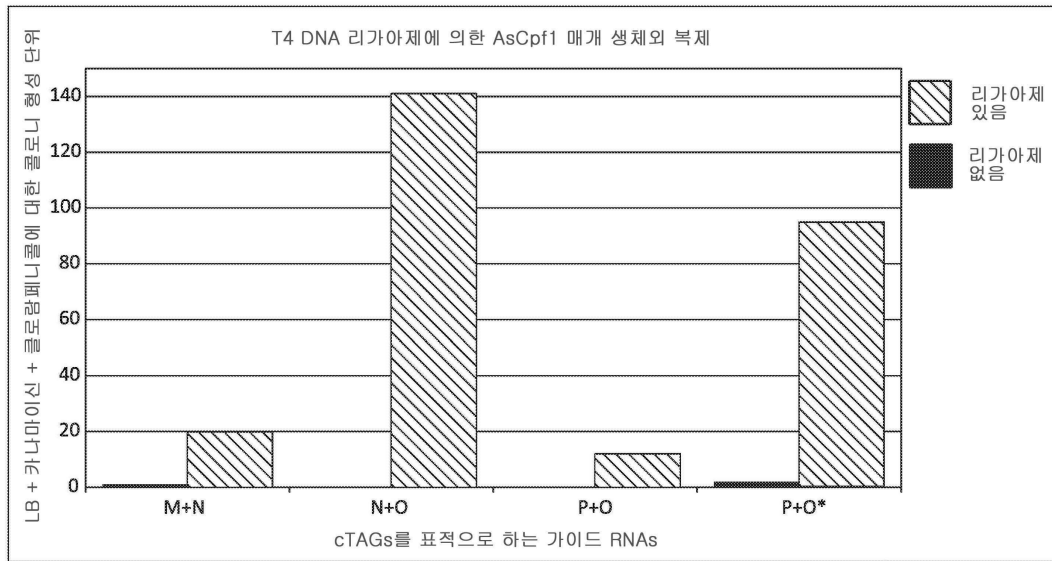
도면4



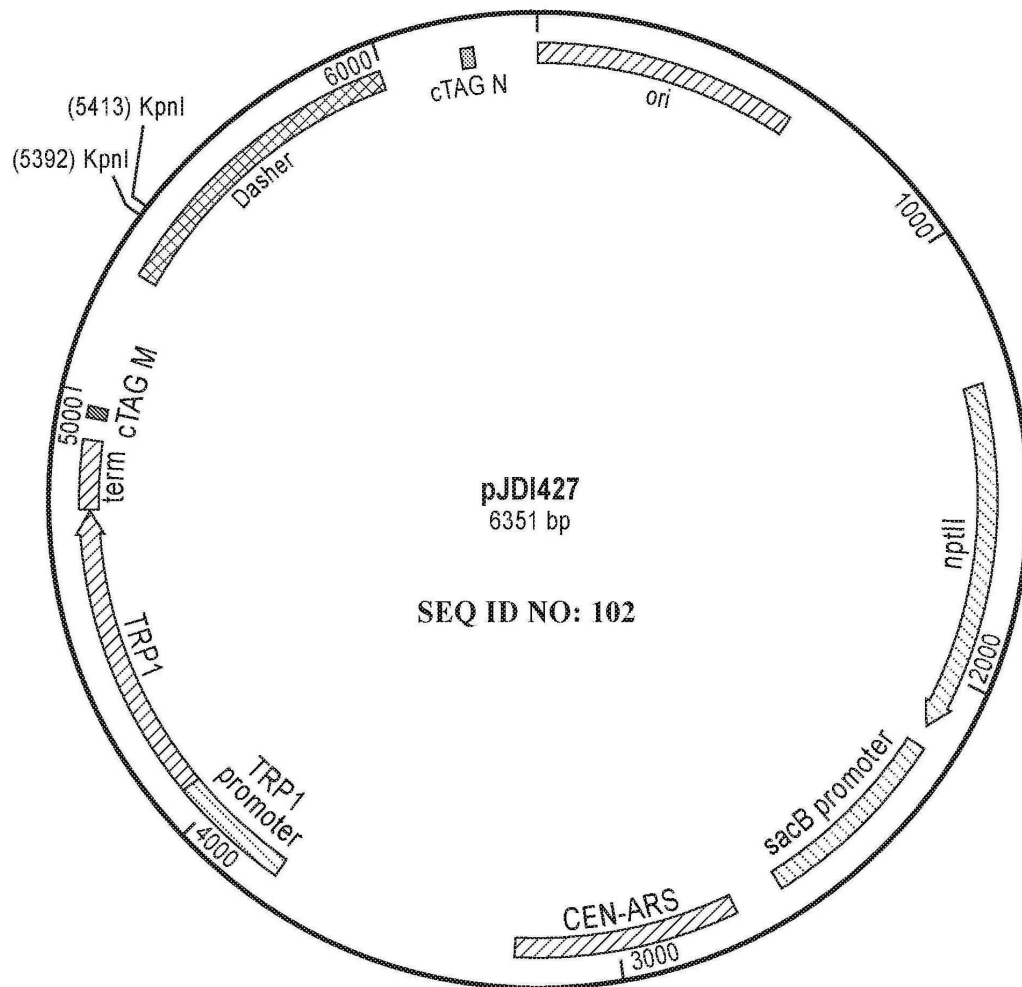
도면5



도면6



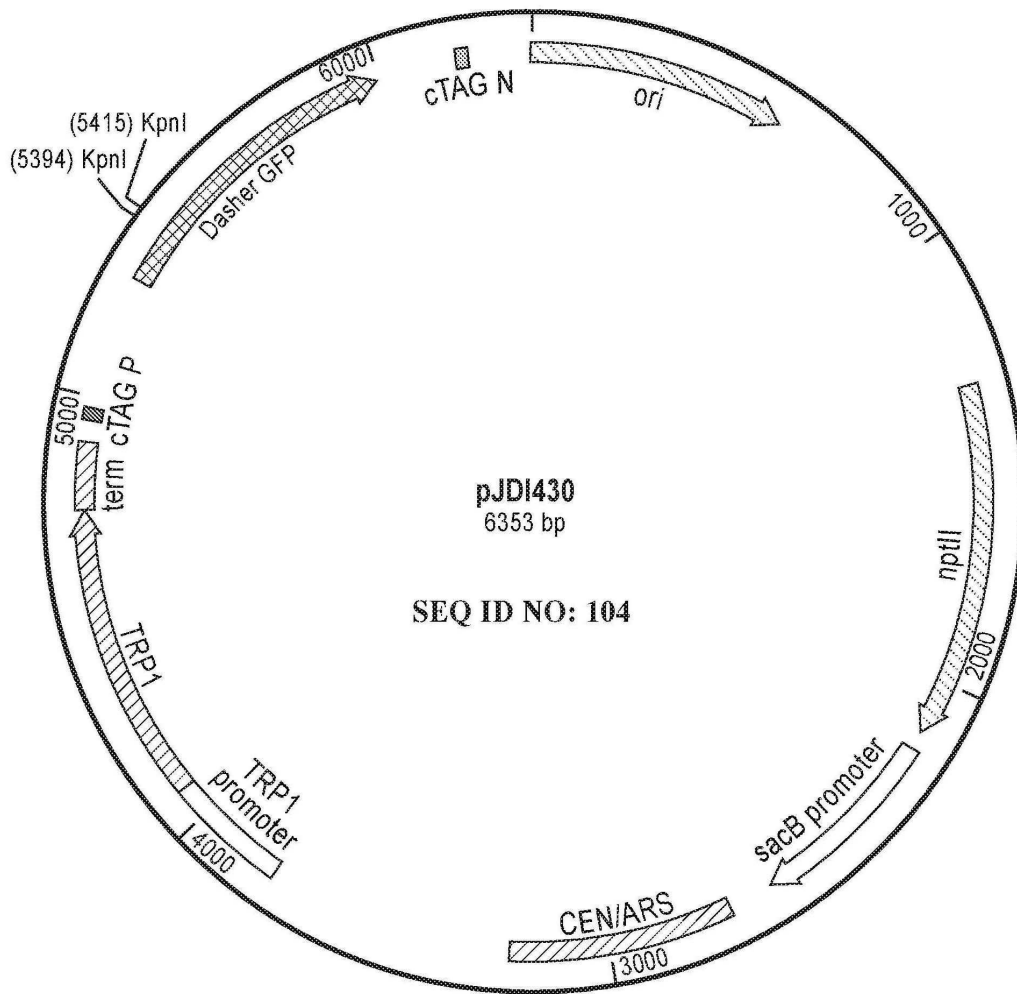
도면7



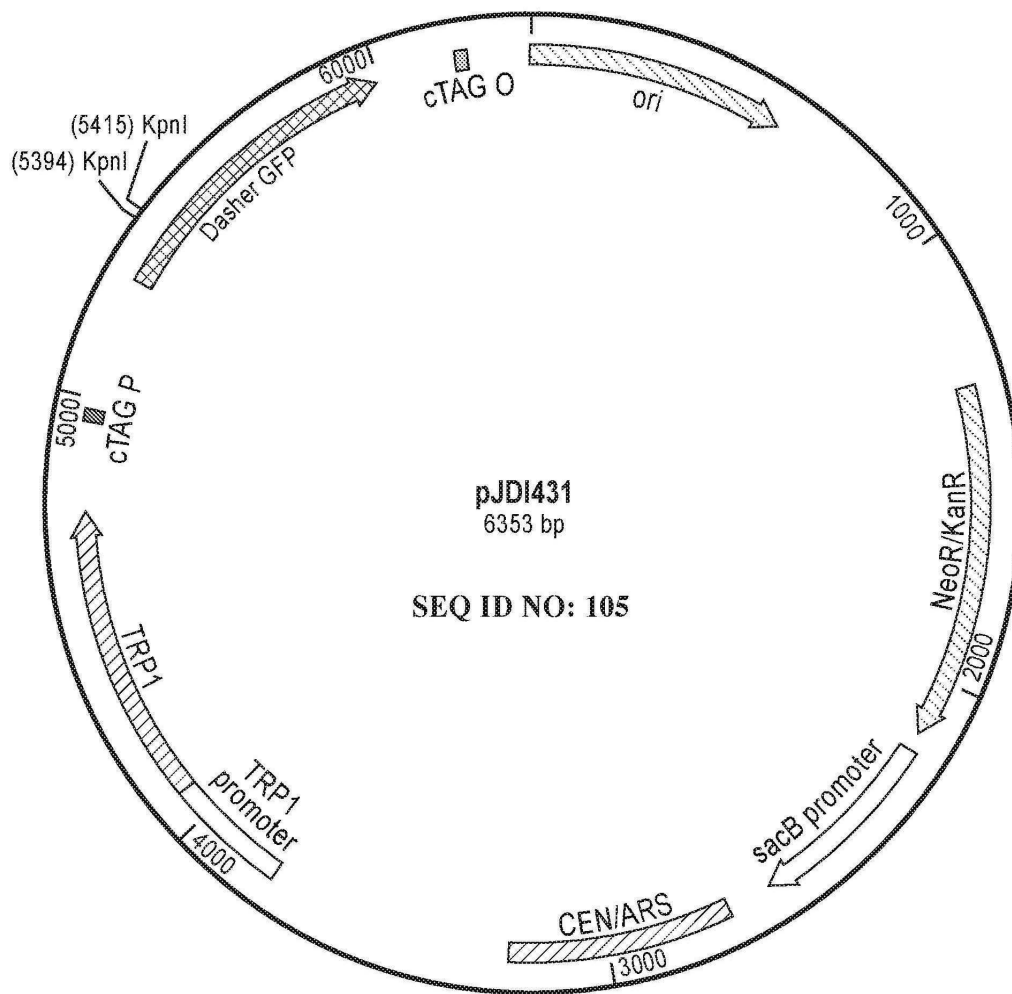
도면8



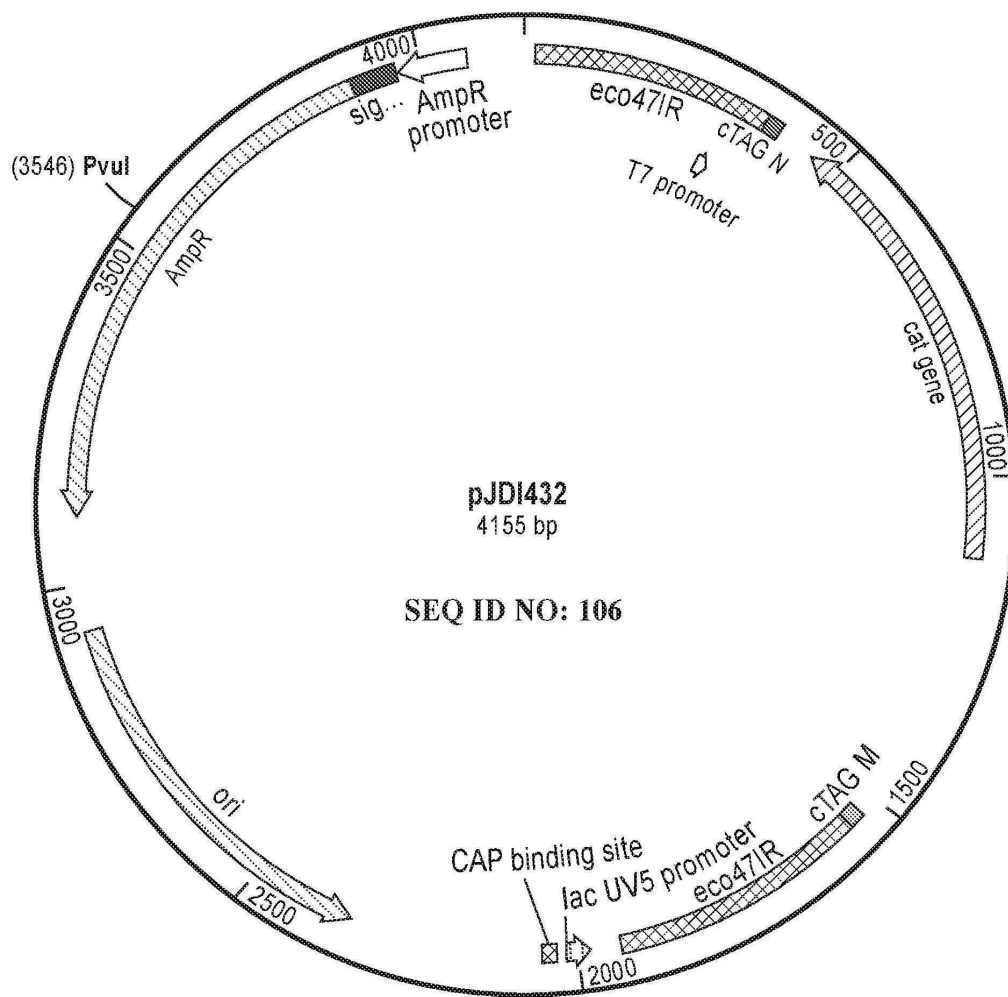
도면9



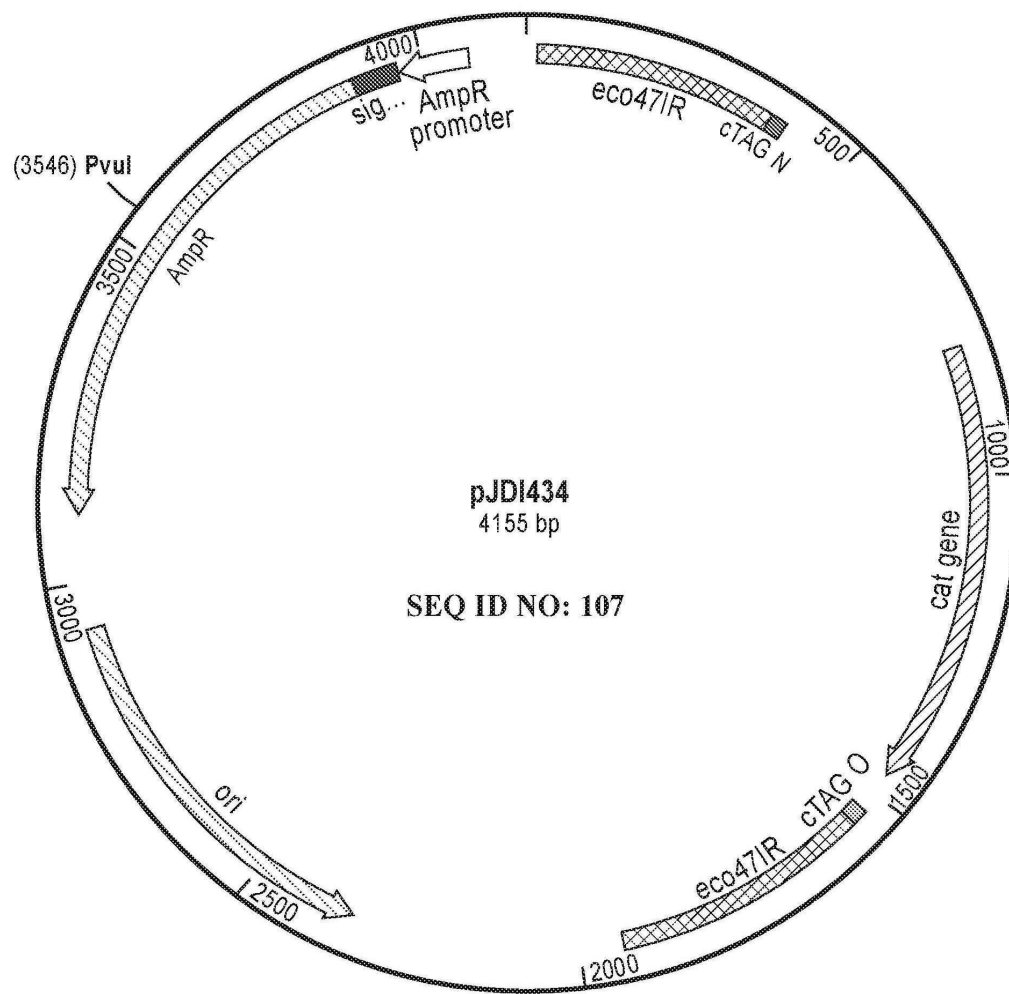
도면10



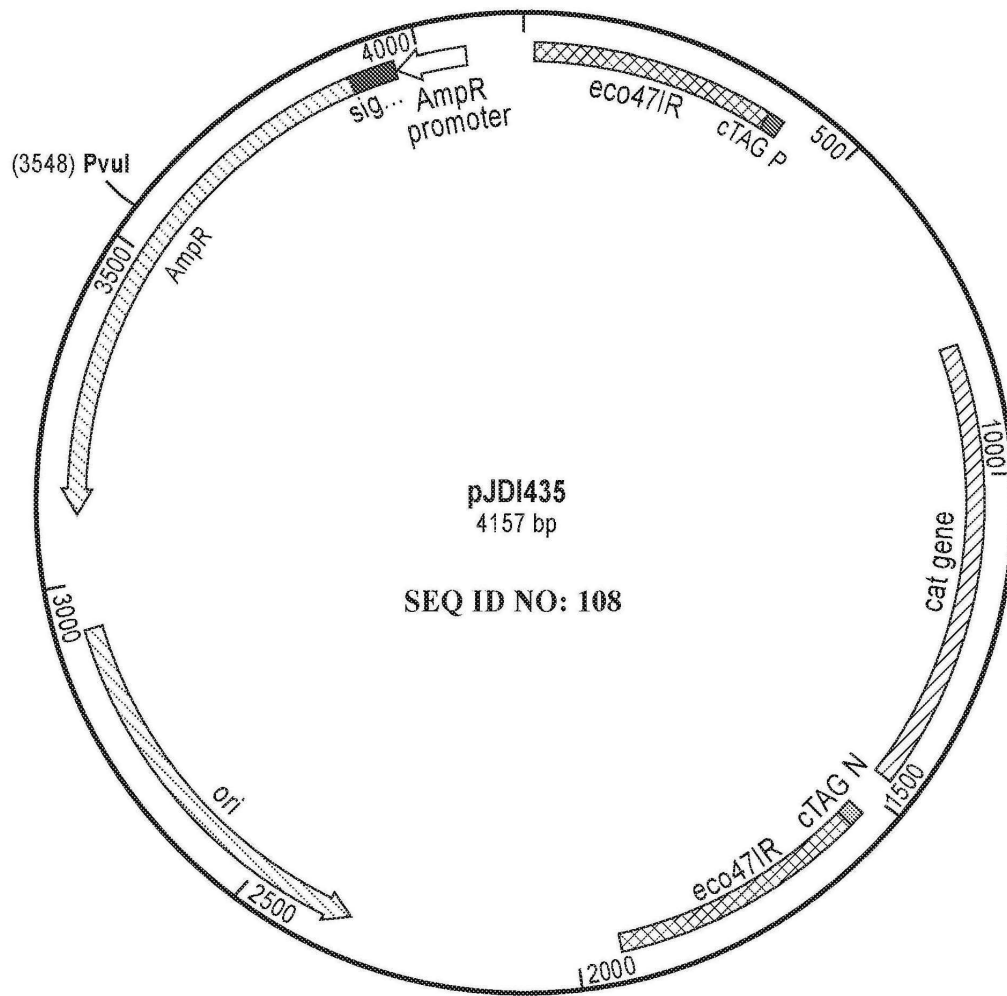
도면11



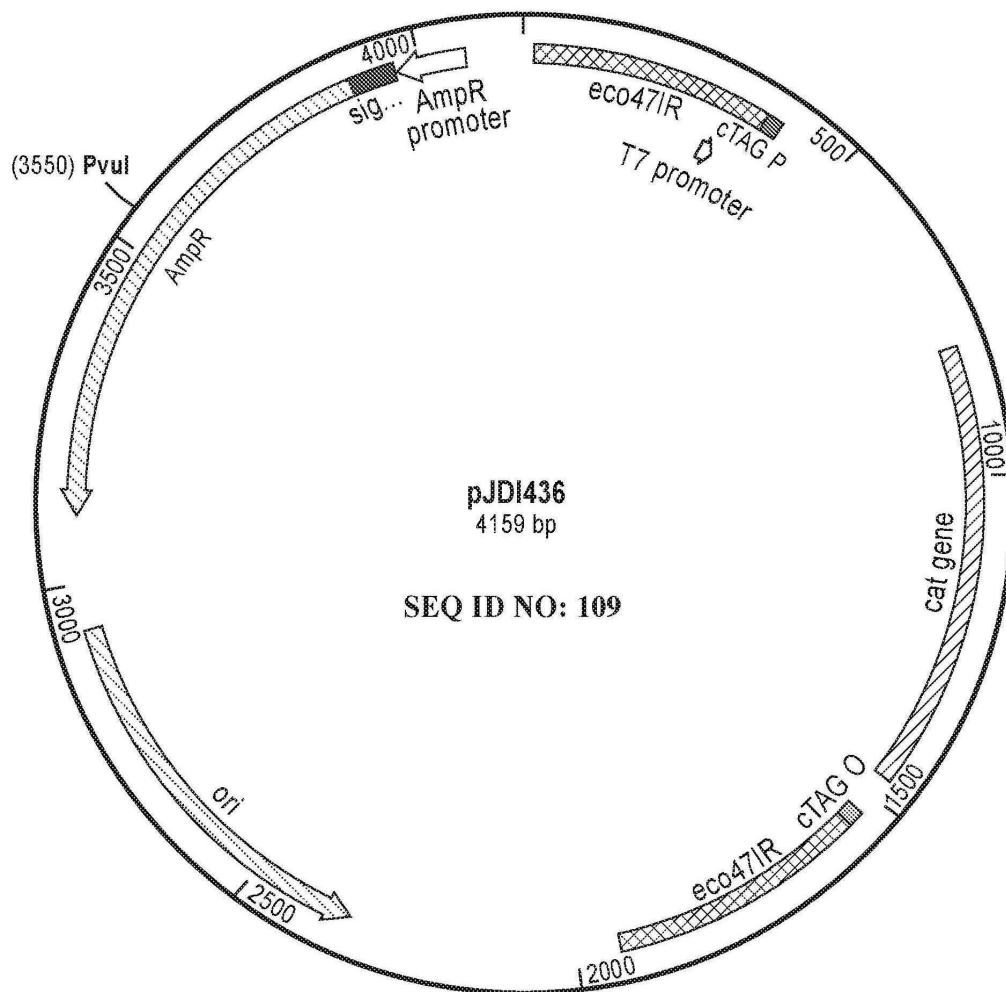
도면12



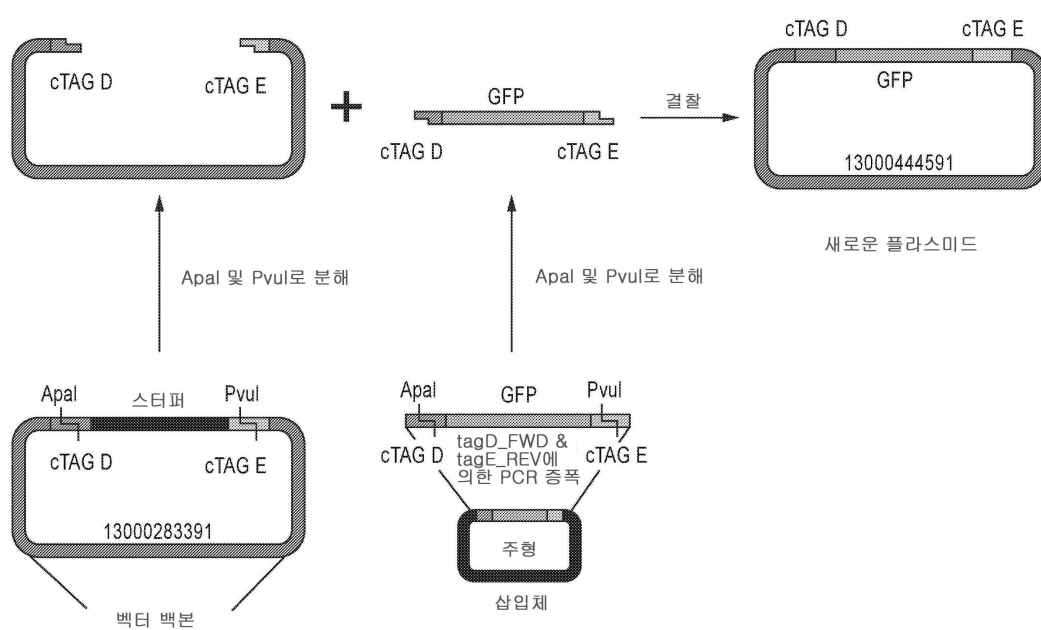
도면13



도면14



도면15



서열 목록

SEQUENCE LISTING

<110> Zymergen Inc.
 Chaikind, Brian
 van Rossum, Hendrik Marinus

<120> A MODULAR UNIVERSAL PLASMID DESIGN STRATEGY FOR THE ASSEMBLY AND
 EDITING OF MULTIPLE DNA CONSTRUCTS FOR MULTIPLE HOSTS

<130> ZYMR-010/01WO 327574-2032

<150> US 62/457,493

<151> 2017-02-10

<160> 113

<170> PatentIn version 3.5

<210> 1

<211> 984

<212> PRT

<213> Campylobacter jejuni

<220><221> MISC_FEATURE

<222> (1)..(984)

<223>

> Cas9

<400> 1

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

Leu Asn His Leu Lys His Leu Ile Ala Asn Glu Phe Lys Leu Asn Tyr
 65 70 75 80
 Glu Asp Tyr Gln Ser Phe Asp Glu Ser Leu Ala Lys Ala Tyr Lys Gly
 85 90 95
 Ser Leu Ile Ser Pro Tyr Glu Leu Arg Phe Arg Ala Leu Asn Glu Leu

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

Asn Glu Ile Ala Lys Asp Ile Thr Leu Ile Lys Asp Glu Ile Lys Leu
 355 360 365
 Lys Lys Ala Leu Ala Lys Tyr Asp Leu Asn Gln Asn Gln Ile Asp Ser
 370 375 380

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

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

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

 Val Phe Thr Lys Gln Asn Gln Glu Lys Leu Asn Gln Thr Pro Phe Glu
 580 585 590
 Ala Phe Gly Asn Asp Ser Ala Lys Trp Gln Lys Ile Glu Val Leu Ala

595 600 605
 Lys Asn Leu Pro Thr Lys Lys Gln Lys Arg Ile Leu Asp Lys Asn Tyr
 610 615 620
 Lys Asp Lys Glu Gln Lys Asn Phe Lys Asp Arg Asn Leu Asn Asp Thr
 625 630 635 640

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

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

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

 Val Asp Ile Phe Lys His Lys Lys Thr Asn Lys Phe Tyr Ala Val Pro
 835 840 845

Ile Tyr Thr Met Asp Phe Ala Leu Lys Val Leu Pro Asn Lys Ala Val
850 855 860
Ala Arg Ser Lys Lys Gly Glu Ile Lys Asp Trp Ile Leu Met Asp Glu
865 870 875 880
Asn Tyr Glu Phe Cys Phe Ser Leu Tyr Lys Asp Ser Leu Ile Leu Ile
885 890 895
Gln Thr Lys Asp Met Gln Glu Pro Glu Phe Val Tyr Tyr Asn Ala Phe
900 905 910
Thr Ser Ser Thr Val Ser Leu Ile Val Ser Lys His Asp Asn Lys Phe
915 920 925
Glu Thr Leu Ser Lys Asn Gln Lys Ile Leu Phe Lys Asn Ala Asn Glu
930 935 940
Lys Glu Val Ile Ala Lys Ser Ile Gly Ile Gln Asn Leu Lys Val Phe
945 950 955 960
Glu Lys Tyr Ile Val Ser Ala Leu Gly Glu Val Thr Lys Ala Glu Phe
965 970 975
Arg Gln Arg Glu Asp Phe Lys Lys
980
<210> 2
<211> 1056
<212> PRT
<213> Pasteurella multocida
<220><221> MISC_FEATURE
<222> (1)..(1056)
<223> Cas9
<400> 2
Met Gln Thr Thr Asn Leu Ser Tyr Ile Leu Gly Leu Asp Leu Gly Ile
1 5 10 15
Ala Ser Val Gly Trp Ala Val Val Glu Ile Asn Glu Asn Glu Asp Pro
20 25 30
Ile Gly Leu Ile Asp Val Gly Val Arg Ile Phe Glu Arg Ala Glu Val
35 40 45

Pro Lys Thr Gly Glu Ser Leu Ala Leu Ser Arg Arg Leu Ala Arg Ser
50 55 60
Thr Arg Arg Leu Ile Arg Arg Arg Ala His Arg Leu Leu Leu Ala Lys
65 70 75 80
Arg Phe Leu Lys Arg Glu Gly Ile Leu Ser Thr Ile Asp Leu Glu Lys
85 90 95
Gly Leu Pro Asn Gln Ala Trp Glu Leu Arg Val Ala Gly Leu Glu Arg
100 105 110
Arg Leu Ser Ala Ile Glu Trp Gly Ala Val Leu Leu His Leu Ile Lys
115 120 125
His Arg Gly Tyr Leu Ser Lys Arg Lys Asn Glu Ser Gln Thr Asn Asn
130 135 140
Lys Glu Leu Gly Ala Leu Leu Ser Gly Val Ala Gln Asn His Gln Leu
145 150 155 160
Leu Gln Ser Asp Asp Tyr Arg Thr Pro Ala Glu Leu Ala Leu Lys Lys
165 170 175
Phe Ala Lys Glu Glu Gly His Ile Arg Asn Gln Arg Gly Ala Tyr Thr
180 185 190
His Thr Phe Asn Arg Leu Asp Leu Leu Ala Glu Leu Asn Leu Leu Phe
195 200 205
Ala Gln Gln His Gln Phe Gly Asn Pro His Cys Lys Glu His Ile Gln
210 215 220
Gln Tyr Met Thr Glu Leu Leu Met Trp Gln Lys Pro Ala Leu Ser Gly
225 230 235 240
Glu Ala Ile Leu Lys Met Leu Gly Lys Cys Thr His Glu Lys Asn Glu
245 250 255
Phe Lys Ala Ala Lys His Thr Tyr Ser Ala Glu Arg Phe Val Trp Leu
260 265 270
Thr Lys Leu Asn Asn Leu Arg Ile Leu Glu Asp Gly Ala Glu Arg Ala
275 280 285
Leu Asn Glu Glu Glu Arg Gln Leu Leu Ile Asn His Pro Tyr Glu Lys

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

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

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

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

 530 535 540

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

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

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

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

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

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

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

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

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

1040

1045

1050

Pro Val Arg

1055

<210> 3

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<220><221> MISC_FEATURE

<222> (1)..(1368)

<223> Cas9

<400> 3

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

1 5 10 15

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

20 25 30

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

35

40

45

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

50 55 60

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

65 70 75 80

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

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100

105

110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

130 135 140

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

145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

165 170 175

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

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

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

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

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

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

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

<210> 4

<211> 1121

<212> PRT

<213> Streptococcus thermophilus

<220><221> MISC_FEATURE

<222> (1)..(1121)

<223> Cas9

<400> 4

Met Ser Asp Leu Val Leu Gly Leu Asp Ile Gly Ile Gly Ser Val Gly

1 5 10 15

Val Gly Ile Leu Asn Lys Val Thr Gly Glu Ile Ile His Lys Asn Ser

20 25 30

Arg Ile Phe Pro Ala Ala Gln Ala Glu Asn Asn Leu Val Arg Arg Thr

35 40 45

Asn Arg Gln Gly Arg Arg Leu Ala Arg Arg Lys Lys His Arg Arg Val

50 55 60

Arg Leu Asn Arg Leu Phe Glu Glu Ser Gly Leu Ile Thr Asp Phe Thr

65 70 75 80

Lys Ile Ser Ile Asn Leu Asn Pro Tyr Gln Leu Arg Val Lys Gly Leu

85 90 95

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

100 105 110

Val Lys His Arg Gly Ile Ser Tyr Leu Asp Asp Ala Ser Asp Asp Gly

115 120 125

Asn Ser Ser Val Gly Asp Tyr Ala Gln Ile Val Lys Glu Asn Ser Lys

130 135 140

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

145 150 155 160

Thr Tyr Gly Gln Leu Arg Gly Asp Phe Thr Val Glu Lys Asp Gly Lys

165 170 175

Lys His Arg Leu Ile Asn Val Phe Pro Thr Ser Ala Tyr Arg Ser Glu

180 185 190

Ala Leu Arg Ile Leu Gln Thr Gln Gln Glu Phe Asn Pro Gln Ile Thr

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

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

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

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

 420 425 430
 Met Glu Leu Ile Pro Glu Leu Tyr Glu Thr Ser Glu Glu Gln Met Thr
 435 440 445

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

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

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

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

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

 930 935 940

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

995 1000 1005
 Lys Lys Glu Gly Val Asp Ser Asp Ser Glu Phe Lys Phe Thr Leu
 1010 1015 1020
 Tyr Lys Asn Asp Leu Leu Leu Val Lys Asp Thr Glu Thr Lys Glu
 1025 1030 1035
 Gln Gln Leu Phe Arg Phe Leu Ser Arg Thr Met Pro Lys Gln Lys
 1040 1045 1050
 His Tyr Val Glu Leu Lys Pro Tyr Asp Lys Gln Lys Phe Glu Gly

1055 1060 1065
 Gly Glu Ala Leu Ile Lys Val Leu Gly Asn Val Ala Asn Ser Gly
 1070 1075 1080
 Gln Cys Lys Lys Gly Leu Gly Lys Ser Asn Ile Ser Ile Tyr Lys
 1085 1090 1095
 Val Arg Thr Asp Val Leu Gly Asn Gln His Ile Ile Lys Asn Glu
 1100 1105 1110
 Gly Asp Lys Pro Lys Leu Asp Phe
 1115 1120

<210> 5

<211> 1082

<212> PRT

<213> Neisseria meningitides

<220><221> MISC_FEATURE

<222> (1)..(1082)

<223> Cas9

<400> 5

Met Ala Ala Phe Lys Pro Asn Pro Ile Asn Tyr Ile Leu Gly Leu Asp
1 5 10 15
Ile Gly Ile Ala Ser Val Gly Trp Ala Met Val Glu Ile Asp Glu Asp
20 25 30
Glu Asn Pro Ile Cys Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg
35 40 45
Ala Glu Val Pro Lys Thr Gly Asp Ser Leu Ala Met Ala Arg Arg Leu
50 55 60
Ala Arg Ser Val Arg Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu
65 70 75 80
Arg Ala Arg Arg Leu Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asp
85 90 95
Phe Asp Glu Asn Gly Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln
100 105 110
Leu Arg Ala Ala Ala Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser
115 120 125
Ala Val Leu Leu His Leu Ile Lys His Arg Gly Tyr Leu Ser Gln Arg
130 135 140
Lys Asn Glu Gly Glu Thr Ala Asp Lys Glu Leu Gly Ala Leu Leu Lys
145 150 155 160
Gly Val Ala Asp Asn Ala His Ala Leu Gln Thr Gly Asp Phe Arg Thr
165 170 175
Pro Ala Glu Leu Ala Leu Asn Lys Phe Glu Lys Glu Ser Gly His Ile
180 185 190
Arg Asn Gln Arg Gly Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu
195 200 205
Gln Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn
210 215 220
Pro His Val Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met
225 230 235 240
Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly

245	250	255
His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr		
260	265	270
Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile		
275	280	285
Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr		
290	295	300
Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala		
305	310	315
Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg		
325	330	335
Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala		
340	345	350
Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys		
355	360	365
Lys Ser Pro Leu Asn Leu Ser Pro Glu Leu Gln Asp Glu Ile Gly Thr		
370	375	380
Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys		
385	390	395
Asp Arg Ile Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser		
405	410	415
Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val		
420	425	430
Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile		
435	440	445
Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu		
450	455	460
Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala		
465	470	475
Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly		
485	490	495

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

 Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Gly
 565 570 575
 Arg Leu Asn Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe
 580 585 590
 Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Val Leu Gly
 595 600 605
 Ser Glu Asn Gln Asn Lys Gly Asn Gln Thr Pro Tyr Glu Tyr Phe Asn
 610 615 620

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

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

740	745	750	
Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Glu Val Leu His Gln			
755	760	765	
Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu Val Met			
770	775	780	
Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala			
785	790	795	800
Asp Thr Pro Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser			
805	810	815	
Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg			
820	825	830	
Ala Pro Asn Arg Lys Met Ser Gly Gln Gly His Met Glu Thr Val Lys			
835	840	845	
Ser Ala Lys Arg Leu Asp Glu Gly Val Ser Val Leu Arg Val Pro Leu			
850	855	860	
Thr Gln Leu Lys Leu Lys Asp Leu Glu Lys Met Val Asn Arg Glu Arg			
865	870	875	880
Glu Pro Lys Leu Tyr Glu Ala Leu Lys Ala Arg Leu Glu Ala His Lys			
885	890	895	
Asp Asp Pro Ala Lys Ala Phe Ala Glu Pro Phe Tyr Lys Tyr Asp Lys			
900	905	910	
Ala Gly Asn Arg Thr Gln Gln Val Lys Ala Val Arg Val Glu Gln Val			
915	920	925	
Gln Lys Thr Gly Val Trp Val Arg Asn His Asn Gly Ile Ala Asp Asn			
930	935	940	
Ala Thr Met Val Arg Val Asp Val Phe Glu Lys Gly Asp Lys Tyr Tyr			
945	950	955	960
Leu Val Pro Ile Tyr Ser Trp Gln Val Ala Lys Gly Ile Leu Pro Asp			
965	970	975	
Arg Ala Val Val Gln Gly Lys Asp Glu Glu Asp Trp Gln Leu Ile Asp			
980	985	990	

Asp Ser Phe Asn Phe Lys Phe Ser Leu His Pro Asn Asp Leu Val Glu
995 1000 1005

Val Ile Thr Lys Lys Ala Arg Met Phe Gly Tyr Phe Ala Ser Cys
1010 1015 1020

His Arg Gly Thr Gly Asn Ile Asn Ile Arg Ile His Asp Leu Asp
1025 1030 1035

His Lys Ile Gly Lys Asn Gly Ile Leu Glu Gly Ile Gly Val Lys
1040 1045 1050

Thr Ala Leu Ser Phe Gln Lys Tyr Gln Ile Asp Glu Leu Gly Lys
1055 1060 1065

Glu Ile Arg Pro Cys Arg Leu Lys Lys Arg Pro Pro Val Arg
1070 1075 1080

<210> 6

<211> 1345

<212> PRT

<213> Streptococcus mutans

<220><221> MISC_FEATURE

<222> (1)..(1345)

<223> Cas9

<400> 6

Met Lys Lys Pro Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Val Thr Asp Asp Tyr Lys Val Pro Ala Lys Lys Met
20 25 30

Lys Val Leu Gly Asn Thr Asp Lys Ser His Ile Glu Lys Asn Leu Leu
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Asn Thr Ala Glu Asp Arg Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Tyr Thr Arg Arg Arg Asn Arg Ile Leu
65 70 75 80

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

85

90

95

Phe Phe His Arg Leu Glu Asp Ser Phe Leu Val Thr Glu Asp Lys Arg

100

105

110

Gly Glu Arg His Pro Ile Phe Gly Asn Leu Glu Glu Glu Val Lys Tyr

115

120

125

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

130

135

140

Asn Pro Glu Lys Val Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala His

145

150

155

160

Ile Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Lys Phe Asp Thr

165

170

175

Arg Asn Asn Asp Val Gln Arg Leu Phe Gln Glu Phe Leu Ala Val Tyr

180

185

190

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

195

200

205

Glu Glu Ile Leu Thr Asp Lys Ile Ser Lys Ser Ala Lys Lys Asp Arg

210

215

220

Val Leu Lys Leu Phe Pro Asn Glu Lys Ser Asn Gly Arg Phe Ala Glu

225

230

235

240

Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp Phe Lys Lys His Phe

245

250

255

Glu Leu Glu Glu Lys Ala Pro Leu Gln Phe Ser Lys Asp Thr Tyr Glu

260

265

270

Glu Glu Leu Glu Val Leu Leu Ala Gln Ile Gly Asp Asn Tyr Ala Glu

275

280

285

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

290

295

300

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

305

310

315

320

Met Ile Gln Arg Tyr Asn Glu His Gln Met Asp Leu Ala Gln Leu Lys

325

330

335

Gln Phe Ile Arg Gln Lys Leu Ser Asp Lys Tyr Asn Glu Val Phe Ser
340 345 350

Asp Val Ser Lys Asp Gly Tyr Ala Gly Tyr Ile Asp Gly Lys Thr Asn
355 360 365

Gln Glu Ala Phe Tyr Lys Tyr Leu Lys Gly Leu Leu Asn Lys Ile Glu
370 375 380

Gly Ser Gly Tyr Phe Leu Asp Lys Ile Glu Arg Glu Asp Phe Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gln Glu Met Arg Ala Ile Ile Arg Arg Gln Ala Glu Phe Tyr Pro Phe
420 425 430

Leu Ala Asp Asn Gln Asp Arg Ile Glu Lys Leu Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Lys Ser Asp Phe Ala Trp
450 455 460

Leu Ser Arg Lys Ser Ala Asp Lys Ile Thr Pro Trp Asn Phe Asp Glu
465 470 475 480

Ile Val Asp Lys Glu Ser Ser Ala Glu Ala Phe Ile Asn Arg Met Thr
485 490 495

Asn Tyr Asp Leu Tyr Leu Pro Asn Gln Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Lys Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

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

Gln Glu Ile Phe Asp Gly Val Phe Lys Val Tyr Arg Lys Val Thr Lys
545 550 555 560

Asp Lys Leu Met Asp Phe Leu Glu Lys Glu Phe Asp Glu Phe Arg Ile
565 570 575

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

580 585 590
 Tyr Gly Thr Tyr His Asp Leu Cys Lys Ile Leu Asp Lys Asp Phe Leu
 595 600 605

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

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

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

 Val Glu Asn Ser Gln Leu Gln Asn Asp Arg Leu Phe Leu Tyr Tyr Leu
 805 810 815
 Gln Asn Gly Arg Asp Met Tyr Thr Gly Glu Glu Leu Asp Ile Asp Tyr
 820 825 830

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

835 840 845

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

850 855 860

Gly Lys Ser Asp Asp Val Pro Ser Lys Asp Val Val Arg Lys Met Lys

865 870 875 880

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

885 890 895

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

900 905 910

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

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

Val Tyr Gly Asp Tyr Pro His Phe His Gly His Lys Glu Asn Lys

1010 1015 1020

Ala Thr Ala Lys Lys Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe

1025 1030 1035

Lys Lys Asp Asp Val Arg Thr Asp Lys Asn Gly Glu Ile Ile Trp

1040 1045 1050

Lys Lys Asp Glu His Ile Ser Asn Ile Lys Lys Val Leu Ser Tyr

1055 1060 1065

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

1070	1075	1080
Phe Ser Lys Glu Ser Ile Leu	Pro Lys Gly Asn Ser	Asp Lys Leu
1085	1090	1095
Ile Pro Arg Lys Thr Lys Lys	Phe Tyr Trp Asp Thr	Lys Lys Tyr
1100	1105	1110
Gly Gly Phe Asp Ser Pro Ile	Val Ala Tyr Ser Ile	Leu Val Ile
1115	1120	1125
Ala Asp Ile Glu Lys Gly Lys	Ser Lys Lys Leu Lys	Thr Val Lys
1130	1135	1140
Ala Leu Val Gly Val Thr Ile	Met Glu Lys Met Thr	Phe Glu Arg
1145	1150	1155
Asp Pro Val Ala Phe Leu Glu	Arg Lys Gly Tyr Arg	Asn Val Gln
1160	1165	1170
Glu Glu Asn Ile Ile Lys Leu	Pro Lys Tyr Ser Leu	Phe Lys Leu
1175	1180	1185
Glu Asn Gly Arg Lys Arg Leu	Leu Ala Ser Ala Arg	Glu Leu Gln
1190	1195	1200
Lys Gly Asn Glu Ile Val Leu	Pro Asn His Leu Gly	Thr Leu Leu
1205	1210	1215
Tyr His Ala Lys Asn Ile His	Lys Val Asp Glu Pro	Lys His Leu
1220	1225	1230
Asp Tyr Val Asp Lys His Lys	Asp Glu Phe Lys Glu	Leu Leu Asp
1235	1240	1245
Val Val Ser Asn Phe Ser Lys	Lys Tyr Thr Leu Ala	Glu Gly Asn
1250	1255	1260
Leu Glu Lys Ile Lys Glu Leu	Tyr Ala Gln Asn Asn	Gly Glu Asp
1265	1270	1275
Leu Lys Glu Leu Ala Ser Ser	Phe Ile Asn Leu Leu	Thr Phe Thr
1280	1285	1290
Ala Ile Gly Ala Pro Ala Thr	Phe Lys Phe Phe Asp	Lys Asn Ile
1295	1300	1305

Asp Arg Lys Arg Tyr Thr Ser Thr Thr Glu Ile Leu Asn Ala Thr
1310 1315 1320

Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp
1325 1330 1335

Leu Asn Lys Leu Gly Gly Asp
1340 1345

<210> 7

<211> 1345

<212> PRT

<213> Francisella tularensis subsp. Novicida U112

<220><221> MISC_FEATURE

<222> (1)..(1345)

<223> FnCpf1; pY004

<400> 7

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 Asn Val 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 Lys Arg Pro Ala Ala Thr Lys Lys

1295 1300 1305

Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val

1310 1315 1320

Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro

1325 1330 1335

Tyr Asp Val Pro Asp Tyr Ala

1340 1345

<210> 8

<211> 1278

<212> PRT

<213> Lachnospiraceae bacterium MC2017

<220><221> MISC_FEATURE

<222> (1)..(1278)

<223> Lb3Cpf1; pY005

<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 Leu 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
Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys
1235 1240 1245
Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr
1250 1255 1260
Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1265 1270 1275

<210> 9
<211> 1345
<212> PRT
<213> Butyrivibrio proteoclasticus
<220><221> MISC_FEATURE
<222> (1)..(1345)
<223> BpCpf1; pY006
<400> 9

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 Asn Val 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 Lys Arg Pro Ala Ala Thr Lys Lys		
1295	1300	1305

Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val

1310 1315 1320

Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro

1325 1330 1335

Tyr Asp Val Pro Asp Tyr Ala

1340 1345

<210> 10

<211> 1522

<212> PRT

<213> Peregrinibacteria bacterium GW2011_GWA_33_10

<220><221> MISC_FEATURE

<222> (1)..(1522)

<223> PeCpf1; pY007

<400> 10

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 Asp Lys Leu Lys Ile Ile Thr Asp Ser Gln
 485 490 495
 Thr Pro Ser Glu Ala Leu Leu Lys Met Ile Phe Asn Asp Ile Glu Lys
 500 505 510
 Asn Met Glu Ser Phe Leu Lys Glu Thr Asn Asp Ile Ile Thr Leu Lys

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

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

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

645 650 655

Leu Phe Gln Lys Glu Trp Thr Glu Gly Arg Gly Lys Asn Leu Thr Lys

660 665 670

Lys Ser Asn Pro Leu Phe Glu Ile Asn Asn Cys Glu Ile Leu Ser Lys

675 680 685

Met Glu Tyr Asp Phe Trp Ala Asp Val Ser Lys Met Ile Pro Lys Cys

690 695 700

Ser Thr Gln Leu Lys Ala Val Val Asn His Phe Lys Gln Ser Asp Asn

705 710 715 720

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

725 730 735

Arg Glu Glu Cys Lys Ile Ser Lys Gln Asp Phe Glu Leu Asn Asn Lys

740 745 750

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

755 760 765

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

770 775 780

Trp Glu Leu Leu Phe Lys Gln Glu Lys Arg Asp Thr Lys Leu Thr Asn

785 790 795 800

Asn Glu Ile Phe Asn Glu Trp Ile Asn Phe Cys Asn Lys Lys Tyr Ser

805 810 815

Glu Leu Leu Ser Trp Glu Arg Lys Tyr Lys Asp Ala Leu Thr Asn Trp

820 825 830

Ile Asn Phe Cys Lys Tyr Phe Leu Ser Lys Tyr Pro Lys Thr Thr Leu

835 840 845

Phe Asn Tyr Ser Phe Lys Glu Ser Glu Asn Tyr Asn Ser Leu Asp Glu

850 855 860

Phe Tyr Arg Asp Val Asp Ile Cys Ser Tyr Lys Leu Asn Ile Asn Thr

865 870 875 880

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

885 890 895
 Tyr Leu Phe Glu Ile Lys Asn Gln Asp Ser Asn Asp Gly Lys Ser Ile

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

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

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

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

Ile Val	Leu Glu Asp Leu Asn	Thr Gly Phe Lys Arg	Ser Arg Gln
1130	1135	1140	
Lys Ile	Asp Lys Ser Val Tyr	Gln Lys Phe Glu Leu	Ala Leu Ala
1145	1150	1155	
Lys Lys	Leu Asn Phe Leu Val	Asp Lys Asn Ala Lys	Arg Asp Glu
1160	1165	1170	
Ile Gly	Ser Pro Thr Lys Ala	Leu Gln Leu Thr Pro	Pro Val Asn
1175	1180	1185	
Asn Tyr	Gly Asp Ile Glu Asn	Lys Lys Gln Ala Gly	Ile Met Leu
1190	1195	1200	
Tyr Thr	Arg Ala Asn Tyr Thr	Ser Gln Thr Asp Pro	Ala Thr Gly
1205	1210	1215	
Trp Arg	Lys Thr Ile Tyr Leu	Lys Ala Gly Pro Glu	Glu Thr Thr
1220	1225	1230	
Tyr Lys	Lys Asp Gly Lys Ile	Lys Asn Lys Ser Val	Lys Asp Gln
1235	1240	1245	
Ile Ile	Glu Thr Phe Thr Asp	Ile Gly Phe Asp Gly	Lys Asp Tyr
1250	1255	1260	
Tyr Phe	Glu Tyr Asp Lys Gly	Glu Phe Val Asp Glu	Lys Thr Gly
1265	1270	1275	
Glu Ile	Lys Pro Lys Lys Trp	Arg Leu Tyr Ser Gly	Glu Asn Gly
1280	1285	1290	
Lys Ser	Leu Asp Arg Phe Arg	Gly Glu Arg Glu Lys	Asp Lys Tyr
1295	1300	1305	
Glu Trp	Lys Ile Asp Lys Ile	Asp Ile Val Lys Ile	Leu Asp Asp
1310	1315	1320	
Leu Phe	Val Asn Phe Asp Lys	Asn Ile Ser Leu Leu	Lys Gln Leu
1325	1330	1335	
Lys Glu	Gly Val Glu Leu Thr	Arg Asn Asn Glu His	Gly Thr Gly
1340	1345	1350	
Glu Ser	Leu Arg Phe Ala Ile	Asn Leu Ile Gln Gln	Ile Arg Asn

1355 1360 1365
Thr Gly Asn Asn Glu Arg Asp Asn Asp Phe Ile Leu Ser Pro Val

1370 1375 1380
Arg Asp Glu Asn Gly Lys His Phe Asp Ser Arg Glu Tyr Trp Asp

1385 1390 1395
Lys Glu Thr Lys Gly Glu Lys Ile Ser Met Pro Ser Ser Gly Asp

1400 1405 1410
Ala Asn Gly Ala Phe Asn Ile Ala Arg Lys Gly Ile Ile Met Asn

1415 1420 1425
Ala His Ile Leu Ala Asn Ser Asp Ser Lys Asp Leu Ser Leu Phe

1430 1435 1440
Val Ser Asp Glu Glu Trp Asp Leu His Leu Asn Asn Lys Thr Glu

1445 1450 1455
Trp Lys Lys Gln Leu Asn Ile Phe Ser Ser Arg Lys Ala Met Ala

1460 1465 1470
Lys Arg Lys Lys Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln

1475 1480 1485
Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr

1490 1495 1500
Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val

1505 1510 1515
Pro Asp Tyr Ala

1520

<210> 11

<211> 1397

<212> PRT

<213> Parcubacteria bacterium GWC2011_GWC2_44_17

<220><221> MISC_FEATURE

<222> (1)..(1397)

<223> PbCpf1; pY008

<400> 11

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 Phe 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 Glu 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 Lys

1340 1345 1350

Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys

1355 1360 1365

Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp

1370 1375 1380

Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala

1385 1390 1395

<210> 12

<211> 1295

<212> PRT

<213> Smithella sp. SC_K08D17

<220><221> MISC_FEATURE

<222> (1)..(1295)

<223> SsCpf1; pY009

<400> 12

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
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
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	Ala 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 Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys
 1250 1255 1260

 Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr
 1265 1270 1275
 Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp
 1280 1285 1290
 Tyr Ala
 1295
 <210> 13
 <211> 1352
 <212> PRT
 <213> Acidaminococcus sp. BV3L6
 <220><221> MISC_FEATURE
 <222> (1)..(1352)
 <223> AsCpf1; pY010
 <400> 13
 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 Lys		
1295	1300	1305
Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys		
1310	1315	1320
Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp		
1325	1330	1335

Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala

1340 1345 1350

<210> 14

<211> 1251

<212> PRT

<213> Lachnospiraceae bacterium MA2020

<220><221> MISC_FEATURE

<222> (1)..(1251)

<223> Lb2Cpf1; pY011

<400> 14

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 Asn 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 Ser Asp Thr 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 Tyr Leu Ser 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
Leu Thr Ser Lys Phe Lys Ser Phe Lys Glu Leu Gly Lys Gln Ser Gly		
	965	970
Val Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile Asp Pro Thr		975
	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 Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala
1205 1210 1215

Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1220 1225 1230

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro
1235 1240 1245

Asp Tyr Ala
1250

<210> 15

<211> 1283

<212> PRT

<213> Candidatus Methanoplasma termitum

<220><221> MISC_FEATURE

<222> (1)..(1283)

<223> CMtCpf1; pY012

<400> 15

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 Glu 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 Lys 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 Pro Arg 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 Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly		
1235	1240	1245
Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp		
1250	1255	1260
Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp		
1265	1270	1275
Val Pro Asp Tyr Ala		
1280		
<210> 16		

<211> 1327

<212> PRT

<213> Eubacterium eligens

<220><221> MISC_FEATURE

<222> (1)..(1327)

<223> EeCpf1; pY013

<400> 16

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 Lys Asp Ile Ile Ile Lys Leu Asn Gly Gln Ala Glu Leu
 740 745 750

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

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

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

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

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

Phe Met Gln Phe Asp Glu Ile Arg Tyr Cys Ala Glu Lys Asp Met
1070 1075 1080
Phe Ser Phe Gly Phe Asp Tyr Asn Asn Phe Asp Thr Tyr Asn Ile
1085 1090 1095
Thr Met Gly Lys Thr Gln Trp Thr Val Tyr Thr Asn Gly Glu Arg
1100 1105 1110
Leu Gln Ser Glu Phe Asn Asn Ala Arg Arg Thr Gly Lys Thr Lys
1115 1120 1125

Ser Ile Asn Leu Thr Glu Thr Ile Lys Leu Leu Leu Glu Asp Asn
1130 1135 1140
Glu Ile Asn Tyr Ala Asp Gly His Asp Ile Arg Ile Asp Met Glu
1145 1150 1155
Lys Met Asp Glu Asp Lys Lys Ser Glu Phe Phe Ala Gln Leu Leu
1160 1165 1170
Ser Leu Tyr Lys Leu Thr Val Gln Met Arg Asn Ser Tyr Thr Glu

1175	1180	1185
Ala Glu Glu Gln Glu Asn Gly Ile Ser Tyr Asp Lys Ile Ile Ser		
1190	1195	1200
Pro Val Ile Asn Asp Glu Gly Glu Phe Phe Asp Ser Asp Asn Tyr		
1205	1210	1215
Lys Glu Ser Asp Asp Lys Glu Cys Lys Met Pro Lys Asp Ala Asp		
1220	1225	1230
Ala Asn Gly Ala Tyr Cys Ile Ala Leu Lys Gly Leu Tyr Glu Val		
1235	1240	1245
Leu Lys Ile Lys Ser Glu Trp Thr Glu Asp Gly Phe Asp Arg Asn		
1250	1255	1260
Cys Leu Lys Leu Pro His Ala Glu Trp Leu Asp Phe Ile Gln Asn		
1265	1270	1275
Lys Arg Tyr Glu Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln		
1280	1285	1290
Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr		
1295	1300	1305
Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val		
1310	1315	1320
Pro Asp Tyr Ala		
1325		
<210> 17		
<211> 1418		
<212> PRT		
<213> Moraxella bovoculi 237		
<220><221> MISC_FEATURE		
<222> (1)..(1418)		
<223> MbCpf1; pY014		
<400> 17		
Met Leu Phe Gln Asp Phe Thr His Leu Tyr Pro Leu Ser Lys Thr Val		
1	5	10
		15

Arg Phe Glu Leu Lys Pro Ile Asp Arg Thr Leu Glu His Ile His Ala

20 25 30

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

35 40 45

Val Lys Val Ile Leu Asp Asp Tyr His Arg Asp Phe Ile Ala Asp Met

50 55 60

Met Gly Glu Val Lys Leu Thr Lys Leu Ala Glu Phe Tyr Asp Val Tyr

65 70 75 80

Leu Lys Phe Arg Lys Asn Pro Lys Asp Asp Glu Leu Gln Lys 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 Asp Lys
 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

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 Ala Gln 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
 Glu 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 Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly
 1370 1375 1380
 Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp
 1385 1390 1395
 Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp

1400 1405 1410
 Val Pro Asp Tyr Ala

1415
 <210> 18
 <211> 1308
 <212> PRT
 <213> Leptospira inadaï
 <220><221> MISC_FEATURE
 <222> (1)..(1308)
 <223> LiCpf1; pY015

<400> 18

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 Lys Asn Leu Phe Ser Lys Glu Leu

130 135 140

Ile Arg Asn Glu Ile Glu Lys Phe Cys Glu Thr Asp Glu Glu Arg Lys

145 150 155 160

Gln Val Ser Asn Phe Lys Ser Phe Thr Thr Tyr Phe Thr Gly Phe His

165 170 175

Ser Asn Arg Gln Asn Ile Tyr Ser Asp Glu Lys Lys Ser Thr Ala Ile

180 185 190

Gly Tyr Arg Ile Ile His Gln Asn Leu Pro Lys Phe Leu Asp Asn Leu

195 200 205

Lys Ile Ile Glu Ser Ile Gln Arg Arg Phe Lys Asp Phe Pro Trp Ser

210 215 220

Asp Leu Lys Lys Asn Leu Lys Lys Ile Asp Lys Asn Ile Lys Leu Thr

225 230 235 240

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

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

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

 Glu Lys Trp Leu Lys Gln Lys Tyr Tyr Thr Ile Ser Phe Leu Asn Asp
 420 425 430
 Ala Ile Glu Ser Tyr Ser Lys Ser Gln Asp Glu Lys Arg Val Lys Ile
 435 440 445
 Arg Leu Glu Ala Tyr Phe Ala Glu Phe Lys Ser Lys Asp Asp Ala Lys
 450 455 460
 Lys Gln Phe Asp Leu Leu Glu Arg Ile Glu Glu Ala Tyr Ala Ile Val
 465 470 475 480

 Glu Pro Leu Leu Gly Ala Glu Tyr Pro Arg Asp Arg Asn Leu Lys Ala

485	490	495
Asp Lys Lys Glu Val Gly Lys Ile Lys Asp Phe Leu Asp Ser Ile Lys		
500	505	510
Ser Leu Gln Phe Phe Leu Lys Pro Leu Leu Ser Ala Glu Ile Phe Asp		
515	520	525
Glu Lys Asp Leu Gly Phe Tyr Asn Gln Leu Glu Gly Tyr Tyr Glu Glu		
530	535	540
Ile Asp Ser Ile Gly His Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr		
545	550	555
Gly Lys Ile Tyr Ser Lys Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser		
565	570	575
Thr Leu Leu Lys Gly Trp Asp Glu Asn Arg Glu Val Ala Asn Leu Cys		
580	585	590
Val Ile Phe Arg Glu Asp Gln Lys Tyr Tyr Leu Gly Val Met Asp Lys		
595	600	605
Glu Asn Asn Thr Ile Leu Ser Asp Ile Pro Lys Val Lys Pro Asn Glu		
610	615	620
Leu Phe Tyr Glu Lys Met Val Tyr Lys Leu Ile Pro Thr Pro His Met		
625	630	635
Gln Leu Pro Arg Ile Ile Phe Ser Ser Asp Asn Leu Ser Ile Tyr Asn		
645	650	655
Pro Ser Lys Ser Ile Leu Lys Ile Arg Glu Ala Lys Ser Phe Lys Glu		
660	665	670
Gly Lys Asn Phe Lys Leu Lys Asp Cys His Lys Phe Ile Asp Phe Tyr		
675	680	685
Lys Glu Ser Ile Ser Lys Asn Glu Asp Trp Ser Arg Phe Asp Phe Lys		
690	695	700
Phe Ser Lys Thr Ser Ser Tyr Glu Asn Ile Ser Glu Phe Tyr Arg Glu		
705	710	715
Val Glu Arg Gln Gly Tyr Asn Leu Asp Phe Lys Lys Val Ser Lys Phe		
725	730	735

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

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

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

 Trp Gly Thr Val Glu Asn Ile Lys Glu Leu Lys Glu Gly Tyr Leu Ser
 930 935 940
 Ile Val Ile His Gln Ile Ser Lys Leu Met Val Glu Asn Asn Ala Ile
 945 950 955 960
 Val Val Leu Glu Asp Leu Asn Ile Gly Phe Lys Arg Gly Arg Gln Lys
 965 970 975
 Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Lys Met Leu Ile Asp Lys

980	985	990
Leu Asn Phe Leu Val Phe Lys Glu	Asn Lys Pro Thr Glu	Pro Gly Gly
995	1000	1005
Val Leu Lys Ala Tyr Gln Leu	Thr Asp Glu Phe Gln	Ser Phe Glu
1010	1015	1020
Lys Leu Ser Lys Gln Thr Gly	Phe Leu Phe Tyr Val	Pro Ser Trp
1025	1030	1035
Asn Thr Ser Lys Ile Asp Pro	Arg Thr Gly Phe Ile	Asp Phe Leu
1040	1045	1050
His Pro Ala Tyr Glu Asn Ile	Glu Lys Ala Lys Gln	Trp Ile Asn
1055	1060	1065
Lys Phe Asp Ser Ile Arg Phe	Asn Ser Lys Met Asp	Trp Phe Glu
1070	1075	1080
Phe Thr Ala Asp Thr Arg Lys	Phe Ser Glu Asn Leu	Met Leu Gly
1085	1090	1095
Lys Asn Arg Val Trp Val Ile	Cys Thr Thr Asn Val	Glu Arg Tyr
1100	1105	1110
Phe Thr Ser Lys Thr Ala Asn	Ser Ser Ile Gln Tyr	Asn Ser Ile
1115	1120	1125
Gln Ile Thr Glu Lys Leu Lys	Glu Leu Phe Val Asp	Ile Pro Phe
1130	1135	1140
Ser Asn Gly Gln Asp Leu Lys	Pro Glu Ile Leu Arg	Lys Asn Asp
1145	1150	1155
Ala Val Phe Phe Lys Ser Leu	Leu Phe Tyr Ile Lys	Thr Thr Leu
1160	1165	1170
Ser Leu Arg Gln Asn Asn Gly	Lys Lys Gly Glu Glu	Glu Lys Asp
1175	1180	1185
Phe Ile Leu Ser Pro Val Val	Asp Ser Lys Gly Arg	Phe Phe Asn
1190	1195	1200
Ser Leu Glu Ala Ser Asp Asp	Glu Pro Lys Asp Ala	Asp Ala Asn
1205	1210	1215

Gly Ala Tyr His Ile Ala Leu Lys Gly Leu Met Asn Leu Leu Val
1220 1225 1230

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

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

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys
1265 1270 1275

Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr
1280 1285 1290

Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1295 1300 1305

<210> 19

<211> 1272

<212> PRT

<213> Lachnospiraceae bacterium ND2006

<220><221> MISC_FEATURE

<222> (1)..(1272)

<223> LbCpf1; pY016

<400> 19

Ser Lys Leu Glu Lys Phe Thr Asn Cys Tyr Ser Leu Ser Lys Thr Leu
1 5 10 15
Arg Phe Lys Ala Ile Pro Val Gly Lys Thr Gln Glu Asn Ile Asp Asn
20 25 30

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

85

90

95

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

100

105

110

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

115

120

125

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

130

135

140

Gly Phe Thr Thr Ala Phe Thr Gly Phe Phe Asp Asn Arg Glu Asn Met

145

150

155

160

Phe Ser Glu Glu Ala Lys Ser Thr Ser Ile Ala Phe Arg Cys Ile Asn

165

170

175

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

180

185

190

Asp Ala Ile Phe Asp Lys His Glu Val Gln Glu Ile Lys Glu Lys Ile

195

200

205

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

210

215

220

Asn Phe Val Leu Thr Gln Glu Gly Ile Asp Val Tyr Asn Ala Ile Ile

225

230

235

240

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

245

250

255

Tyr Ile Asn Leu Tyr Asn Gln Lys Thr Lys Gln Lys Leu Pro Lys Phe

260

265

270

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

275

280

285

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

290

295

300

Asn Thr Leu Asn Lys Asn Ser Glu Ile Phe Ser Ser Ile Lys Lys Leu

305

310

315

320

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

325

330

335

Val Lys Asn Gly Pro Ala Ile Ser Thr Ile Ser Lys Asp Ile Phe Gly
340 345 350

Glu Trp Asn Val Ile Arg Asp Lys Trp Asn Ala Glu Tyr Asp Asp Ile
355 360 365

His Leu Lys Lys Lys Ala Val Val Thr Glu Lys Tyr Glu Asp Asp Arg
370 375 380

Arg Lys Ser Phe Lys Lys Ile Gly Ser Phe Ser Leu Glu Gln Leu Gln
385 390 395 400

Glu Tyr Ala Asp Ala Asp Leu Ser Val Val Glu Lys Leu Lys Glu Ile
405 410 415

Ile Ile Gln Lys Val Asp Glu Ile Tyr Lys Val Tyr Gly Ser Ser Glu
420 425 430

Lys Leu Phe Asp Ala Asp Phe Val Leu Glu Lys Ser Leu Lys Lys Asn
435 440 445

Asp Ala Val Val Ala Ile Met Lys Asp Leu Leu Asp Ser Val Lys Ser
450 455 460

Phe Glu Asn Tyr Ile Lys Ala Phe Phe Gly Glu Gly Lys Glu Thr Asn
465 470 475 480

Arg Asp Glu Ser Phe Tyr Gly Asp Phe Val Leu Ala Tyr Asp Ile Leu
485 490 495

Leu Lys Val Asp His Ile Tyr Asp Ala Ile Arg Asn Tyr Val Thr Gln
500 505 510

Lys Pro Tyr Ser Lys Asp Lys Phe Lys Leu Tyr Phe Gln Asn Pro Gln
515 520 525

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

Ile Leu Arg Tyr Gly Ser Lys Tyr Tyr Leu Ala Ile Met Asp Lys Lys
545 550 555 560

Tyr Ala Lys Cys Leu Gln Lys Ile Asp Lys Asp Asp Val Asn Gly Asn
565 570 575

Tyr Glu Lys Ile Asn Tyr Lys Leu Leu Pro Gly Pro Asn Lys Met Leu

580	585	590
Pro Lys Val Phe Phe Ser Lys Lys Trp Met Ala Tyr Tyr Asn Pro Ser		
595	600	605
Glu Asp Ile Gln Lys Ile Tyr Lys Asn Gly Thr Phe Lys Lys Gly Asp		
610	615	620
Met Phe Asn Leu Asn Asp Cys His Lys Leu Ile Asp Phe Phe Lys Asp		
625	630	635
Ser Ile Ser Arg Tyr Pro Lys Trp Ser Asn Ala Tyr Asp Phe Asn Phe		
645	650	655
Ser Glu Thr Glu Lys Tyr Lys Asp Ile Ala Gly Phe Tyr Arg Glu Val		
660	665	670
Glu Glu Gln Gly Tyr Lys Val Ser Phe Glu Ser Ala Ser Lys Lys Glu		
675	680	685
Val Asp Lys Leu Val Glu Glu Gly Lys Leu Tyr Met Phe Gln Ile Tyr		
690	695	700
Asn Lys Asp Phe Ser Asp Lys Ser His Gly Thr Pro Asn Leu His Thr		
705	710	715
Met Tyr Phe Lys Leu Leu Phe Asp Glu Asn Asn His Gly Gln Ile Arg		
725	730	735
Leu Ser Gly Gly Ala Glu Leu Phe Met Arg Arg Ala Ser Leu Lys Lys		
740	745	750
Glu Glu Leu Val Val His Pro Ala Asn Ser Pro Ile Ala Asn Lys Asn		
755	760	765
Pro Asp Asn Pro Lys Lys Thr Thr Thr Leu Ser Tyr Asp Val Tyr Lys		
770	775	780
Asp Lys Arg Phe Ser Glu Asp Gln Tyr Glu Leu His Ile Pro Ile Ala		
785	790	795
800		
Ile Asn Lys Cys Pro Lys Asn Ile Phe Lys Ile Asn Thr Glu Val Arg		
805	810	815
Val Leu Leu Lys His Asp Asp Asn Pro Tyr Val Ile Gly Ile Asp Arg		
820	825	830

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

835 840 845

Ile Val Glu Gln Tyr Ser Leu Asn Glu Ile Ile Asn Asn Phe Asn Gly

850 855 860

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

865 870 875 880

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

885 890 895

Glu Leu Lys Ala Gly Tyr Ile Ser Gln Val Val His Lys Ile Cys Glu

900 905 910

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

915 920 925

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

930 935 940

Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Met Val Asp Lys Lys

945 950 955 960

Ser Asn Pro Cys Ala Thr Gly Gly Ala Leu Lys Gly Tyr Gln Ile Thr

965 970 975

Asn Lys Phe Glu Ser Phe Lys Ser Met Ser Thr Gln Asn Gly Phe Ile

980 985 990

Phe Tyr Ile Pro Ala Trp Leu Thr Ser Lys Ile Asp Pro Ser Thr Gly

995 1000 1005

Phe Val Asn Leu Leu Lys Thr Lys Tyr Thr Ser Ile Ala Asp Ser

1010 1015 1020

Lys Lys Phe Ile Ser Ser Phe Asp Arg Ile Met Tyr Val Pro Glu

1025 1030 1035

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

1040 1045 1050

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

1055 1060 1065

Asn Arg Ile Arg Ile Phe Arg Asn Pro Lys Lys Asn Asn Val Phe

1070	1075	1080
Asp Trp Glu Glu Val Cys Leu Thr Ser Ala Tyr Lys Glu Leu Phe		
1085	1090	1095
Asn Lys Tyr Gly Ile Asn Tyr Gln Gln Gly Asp Ile Arg Ala Leu		
1100	1105	1110
Leu Cys Glu Gln Ser Asp Lys Ala Phe Tyr Ser Ser Phe Met Ala		
1115	1120	1125
Leu Met Ser Leu Met Leu Gln Met Arg Asn Ser Ile Thr Gly Arg		
1130	1135	1140
Thr Asp Val Asp Phe Leu Ile Ser Pro Val Lys Asn Ser Asp Gly		
1145	1150	1155
Ile Phe Tyr Asp Ser Arg Asn Tyr Glu Ala Gln Glu Asn Ala Ile		
1160	1165	1170
Leu Pro Lys Asn Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Arg		
1175	1180	1185
Lys Val Leu Trp Ala Ile Gly Gln Phe Lys Lys Ala Glu Asp Glu		
1190	1195	1200
Lys Leu Asp Lys Val Lys Ile Ala Ile Ser Asn Lys Glu Trp Leu		
1205	1210	1215
Glu Tyr Ala Gln Thr Ser Val Lys His Lys Arg Pro Ala Ala Thr		
1220	1225	1230
Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr		
1235	1240	1245
Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala		
1250	1255	1260
Tyr Pro Tyr Asp Val Pro Asp Tyr Ala		
1265	1270	
<210> 20		
<211> 1305		
<212> PRT		
<213> Porphyromonas crevioricanis		
<220><221> MISC_FEATURE		

<222> (1)..(1305)

<223> PcCpf1; pY017

<400> 20

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 Lys Arg Pro
 1250 1255 1260
 Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser

1265 1270 1275
 Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro
 1280 1285 1290
 Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
 1295 1300 1305

<210> 21

<211> 1368

<212> PRT

<213> Prevotella disiens

<220><221> MISC_FEATURE

<222> (1)..(1368)

<223> PdCpf1; pY018

<400> 21

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 Lys 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 Asn Ala Glu Leu

370 375 380

Gln Thr Phe Asn Asn Tyr Glu Leu Ile Ala Asn Ala Leu Arg Ala Lys

385 390 395 400

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

405 410 415

Glu Tyr Ile Asp Lys Lys Val Lys Ala Ile Asp Ser Leu Ser Ile Gln

420 425 430

Glu Ile Asn Glu Leu Val Glu Asn Tyr Val Ser Glu Phe Asn Ser Asn

435 440 445

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

450 455 460

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

465 470 475 480

Leu Ser Ala Ala Glu Lys Leu Leu Gly Thr Lys Tyr Gln Glu Thr Ala

485 490 495

Lys Asp Ile Phe Lys Lys Asp Glu Asn Ser Lys Leu Ile Lys Glu Leu

500 505 510

Leu Asp Ala Thr Lys Gln Phe Gln His Phe Ile Lys Pro Leu Leu Gly

515 520 525

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

530 535 540

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

545 550 555 560

Arg Asn Arg Leu Thr Gln Lys Pro Tyr Ser Lys Asp Lys Ile Arg Leu

565 570 575
 Cys Phe Asn Lys Pro Lys Leu Met Thr Gly Trp Val Asp Ser Lys Thr

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

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

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

 770 775 780
 Phe Pro Ile Ser Asn Val Glu Leu Glu Lys Glu Met Gly Asp Lys Asp
 785 790 795 800
 Lys Pro Leu Cys Leu Phe Gln Ile Ser Asn Lys Asp Leu Ser Phe Ala
 805 810 815

Lys Thr Phe Ser Ala Asn Leu Arg Lys Lys Arg Gly Ala Glu Asn Leu
 820 825 830
 His Thr Met Leu Phe Lys Ala Leu Met Glu Gly Asn Gln Asp Asn Leu
 835 840 845
 Asp Leu Gly Ser Gly Ala Ile Phe Tyr Arg Ala Lys Ser Leu Asp Gly
 850 855 860
 Asn Lys Pro Thr His Pro Ala Asn Glu Ala Ile Lys Cys Arg Asn Val
 865 870 875 880
 Ala Asn Lys Asp Lys Val Ser Leu Phe Thr Tyr Asp Ile Tyr Lys Asn
 885 890 895
 Arg Arg Tyr Met Glu Asn Lys Phe Leu Phe His Leu Ser Ile Val Gln
 900 905 910
 Asn Tyr Lys Ala Ala Asn Asp Ser Ala Gln Leu Asn Ser Ser Ala Thr
 915 920 925
 Glu Tyr Ile Arg Lys Ala Asp Asp Leu His Ile Ile Gly Ile Asp Arg
 930 935 940
 Gly Glu Arg Asn Leu Leu Tyr Tyr Ser Val Ile Asp Met Lys Gly Asn
 945 950 955 960
 Ile Val Glu Gln Asp Ser Leu Asn Ile Ile Arg Asn Asn Asp Leu Glu
 965 970 975
 Thr Asp Tyr His Asp Leu Leu Asp Lys Arg Glu Lys Glu Arg Lys Ala
 980 985 990
 Asn Arg Gln Asn Trp Glu Ala Val Glu Gly Ile Lys Asp Leu Lys Lys
 995 1000 1005
 Gly Tyr Leu Ser Gln Ala Val His Gln Ile Ala Gln Leu Met Leu
 1010 1015 1020
 Lys Tyr Asn Ala Ile Ile Ala Leu Glu Asp Leu Gly Gln Met Phe
 1025 1030 1035
 Val Thr Arg Gly Gln Lys Ile Glu Lys Ala Val Tyr Gln Gln Phe
 1040 1045 1050
 Glu Lys Ser Leu Val Asp Lys Leu Ser Tyr Leu Val Asp Lys Lys

1055	1060	1065
Arg Pro Tyr Asn Glu Leu Gly	Gly Ile Leu Lys Ala	Tyr Gln Leu
1070	1075	1080
Ala Ser Ser Ile Thr Lys Asn	Asn Ser Asp Lys Gln	Asn Gly Phe
1085	1090	1095
Leu Phe Tyr Val Pro Ala Trp	Asn Thr Ser Lys Ile	Asp Pro Val
1100	1105	1110
Thr Gly Phe Thr Asp Leu Leu	Arg Pro Lys Ala Met	Thr Ile Lys
1115	1120	1125
Glu Ala Gln Asp Phe Phe Gly	Ala Phe Asp Asn Ile	Ser Tyr Asn
1130	1135	1140
Asp Lys Gly Tyr Phe Glu Phe	Glu Thr Asn Tyr Asp	Lys Phe Lys
1145	1150	1155
Ile Arg Met Lys Ser Ala Gln	Thr Arg Trp Thr Ile	Cys Thr Phe
1160	1165	1170
Gly Asn Arg Ile Lys Arg Lys	Lys Asp Lys Asn Tyr	Trp Asn Tyr
1175	1180	1185
Glu Glu Val Glu Leu Thr Glu	Glu Phe Lys Lys Leu	Phe Lys Asp
1190	1195	1200
Ser Asn Ile Asp Tyr Glu Asn	Cys Asn Leu Lys Glu	Glu Ile Gln
1205	1210	1215
Asn Lys Asp Asn Arg Lys Phe	Phe Asp Asp Leu Ile	Lys Leu Leu
1220	1225	1230
Gln Leu Thr Leu Gln Met Arg	Asn Ser Asp Asp Lys	Gly Asn Asp
1235	1240	1245
Tyr Ile Ile Ser Pro Val Ala	Asn Ala Glu Gly Gln	Phe Phe Asp
1250	1255	1260
Ser Arg Asn Gly Asp Lys Lys	Leu Pro Leu Asp Ala	Asp Ala Asn
1265	1270	1275
Gly Ala Tyr Asn Ile Ala Arg	Lys Gly Leu Trp Asn	Ile Arg Gln
1280	1285	1290

Ile Lys Gln Thr Lys Asn Asp Lys Lys Leu Asn Leu Ser Ile Ser
1295 1300 1305

Ser Thr Glu Trp Leu Asp Phe Val Arg Glu Lys Pro Tyr Leu Lys
1310 1315 1320

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys

1325 1330 1335

Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr
1340 1345 1350

Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1355 1360 1365

<210> 22

<211> 1291

<212> PRT

<213> Porphyromonas macacae

<220><221> MISC_FEATURE

<222> (1)..(1291)

<223> PmCpf1; pY09

<400> 22

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 Glu 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 Ile Ile Gly Ile Asp Arg Gly Glu Arg

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

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

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

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

 1085 1090 1095

Ser Arg Ile Ala Lys Ser Lys Lys Ser Gly Lys Trp Met Val Glu
1100 1105 1110

Arg Ile Glu Asn Leu Ser Leu Cys Phe Leu Glu Leu Phe Glu Gln
1115 1120 1125

Phe Asn Ile Gly Tyr Arg Val Glu Lys Asp Leu Lys Lys Ala Ile
1130 1135 1140

Leu Ser Gln Asp Arg Lys Glu Phe Tyr Val Arg Leu Ile Tyr Leu
1145 1150 1155

Phe Asn Leu Met Met Gln Ile Arg Asn Ser Asp Gly Glu Glu Asp
1160 1165 1170

Tyr Ile Leu Ser Pro Ala Leu Asn Glu Lys Asn Leu Gln Phe Asp
1175 1180 1185

Ser Arg Leu Ile Glu Ala Lys Asp Leu Pro Val Asp Ala Asp Ala
1190 1195 1200

Asn Gly Ala Tyr Asn Val Ala Arg Lys Gly Leu Met Val Val Gln
1205 1210 1215

Arg Ile Lys Arg Gly Asp His Glu Ser Ile His Arg Ile Gly Arg
1220 1225 1230

Ala Gln Trp Leu Arg Tyr Val Gln Glu Gly Ile Val Glu Lys Arg
1235 1240 1245

Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly
1250 1255 1260

Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val
1265 1270 1275

Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1280 1285 1290

<210> 23

<211> 399

<212> PRT

<213> Francisella philomiragia subsp.

<220><221> MISC_FEATURE

<222> (1)..(399)

<223> Genbank ABZ87876 Cpf1

<400> 23

Met Phe Tyr Leu Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser Phe

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr Asp

65 70 75 80

Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Leu Ile Gly Thr Ser Val

85 90 95

Leu Glu Tyr Ile Thr Gln Gln Val Ala Pro Lys Asn Leu Asp Asn Pro

100 105 110

Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala Lys

115 120 125

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

130 135 140

His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Phe Ala Ser

145 150 155 160

Phe Ala Asp Ile Pro Val Leu Phe Asp Glu Ile Ala Gln Asn Lys Asn

165 170 175

Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys Asp

180 185 190

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

195 200 205

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

210 215 220

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

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

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

 340 345 350
 Phe Ser Ala Lys Ser Ile Asn Phe Tyr Asn Pro Ser Glu Asp Ile Leu
 355 360 365
 Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln Lys
 370 375 380
 Asp Met Lys Asn Leu Ser Leu Ile Leu Lys Ile Ala Glu Asn Leu
 385 390 395
 <210> 24
 <211> 485
 <212> PRT
 <213> Francisella philomiragia subsp.

 <220><221> MISC_FEATURE
 <222> (1)..(485)
 <223> Genbank ABZ87877 Cpf1
 <400> 24
 Met Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr Arg Lys Gln
 1 5 10 15
 Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Gln Ala Ile Ala Asn

20 25 30
 Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Phe Phe Glu Tyr Asp Leu
 35 40 45
 Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe His Cys Pro

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

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

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

 245 250 255
 Tyr Val Pro Ala Gly Phe Thr Ser Lys Ile Cys Pro Val Thr Gly Phe
 260 265 270

Val Asn Gln Leu Tyr Pro Lys Tyr Glu Ser Val Ser Lys Ser Gln Glu
275 280 285

Phe Phe Ser Lys Phe Asp Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr
290 295 300

Phe Glu Phe Ser Phe Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys
305 310 315 320

Gly Lys Trp Thr Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg
325 330 335

Asn Ser Asp Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr
340 345 350

Lys Glu Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His
355 360 365

Gly Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
370 375 380

Phe Ala Lys Leu Thr Ser Ile Leu Asn Ser Ile Leu Gln Met Arg Asn
385 390 395 400

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

Val Asn Gly Asn Phe Phe Asp Ser Arg His Ala Pro Lys Asn Met Pro
420 425 430

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

Met Leu Leu Tyr Arg Ile Lys Asn Asn Gln Asp Gly Lys Lys Leu Asn
450 455 460

Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu Phe Val Gln Asn Arg Asn
465 470 475 480

Lys Ser Ser Lys Ile
485

<210> 25

<211> 939

<212> PRT

<213> Francisella philomiragia subsp.

<220><221> MISC_FEATURE

<222> (1)..(939)

<223> Genbank AJ156734 Cpf1

<400> 25

Met Asn Leu Tyr Ser Asn Leu Thr Asn Lys Tyr Ser Leu Ser Lys Thr

1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Glu 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 His Ile Ser Arg Tyr

100 105 110

Ile Asn 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 Asn Val Tyr Ser Ser

180 185 190

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

195 200 205

Pro Lys Phe Ile Glu Asn Lys Ala Lys Tyr Glu Asn 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 Leu 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 Val Ala Pro Lys Asn Leu Asp Asn
 420 425 430
 Pro Ser Lys Lys Glu Gln Asp 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 Tyr Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Phe Ala
 465 470 475 480
 Ser Phe Ala Asp Ile Pro Val Leu Phe Asp Glu Ile Ala Gln Asn Lys
 485 490 495
 Asn Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
 500 505 510

 Asp Leu Leu Gln Thr Ser Ala Glu Val 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 Thr Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
 545 550 555 560
 Phe Tyr Leu Val Phe Asp Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
 565 570 575

 Ala 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
 Val 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 Asn 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 Leu Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe

705 710 715 720
 Ile Asp Phe Tyr Lys His Ser Ile Ser Arg His Pro Glu Trp Lys Asp
 725 730 735
 Phe Gly Phe Arg Phe Ser Asp Thr Lys Lys 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 Leu Val Asp Glu Gly Lys Leu Tyr
 770 775 780
 Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Val Tyr Ser Lys Gly Lys
 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 Ile 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 Cys Lys Asn
 930 935

 <210> 26
 <211> 1261
 <212> PRT

<213> Moraxella bovoculi
 <220><221> MISC_FEATURE
 <222> (1)..(1261)
 <223> Genbank AKG08867 Cpf1
 <400> 26

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

Met Gly Glu Val Lys Leu Thr Lys Leu Ala Glu Phe Tyr Asp Val Tyr
 65 70 75 80
 Leu Lys Phe Arg Lys Asn Pro Lys Asp Asp Gly Leu Gln Lys 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 Ala 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 Arg 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

Gly Val Ser Phe Leu Pro Ser Lys Phe Ala Asp Asp Ser Glu Val 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 Tyr Gln Lys Asp Gly Ile
340 345 350

Tyr Val Glu Tyr 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 Asp Lys Ser Pro Glu Ile Arg Gln
 485 490 495
 Leu Lys Glu Leu Leu Asp Asn Ala Leu Asn Val Ala His Phe Ala Lys
 500 505 510

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

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

 Ser Asn Leu Asp Tyr Tyr Asn Pro Ser Ala Glu Leu Leu Asp Lys Tyr
 645 650 655
 Ala Gln Gly Thr His Lys Lys Gly Asp Asn Phe Asn Leu Lys Asp Cys
 660 665 670
 His Ala Leu Ile Asp Phe Phe Lys Ala Gly Ile Asn Lys His Pro Glu
 675 680 685
 Trp Gln His Phe Gly Phe Lys Phe Ser Pro Thr Ser Ser Tyr Gln Asp
 690 695 700

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

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

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

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

945	950	955	960
Tyr Asn Ala Ile Val Val Leu Glu Asp Leu Asn Phe Gly Phe Lys Arg			
	965	970	975
Gly Arg Phe Lys Val Glu Lys Gln Ile Tyr Gln Asn Phe Glu Asn Ala			
	980	985	990
Leu Ile Lys Lys Leu Asn His Leu Val Leu Lys Asp Lys Ala Asp Asp			
	995	1000	1005
Glu Ile Gly Ser Tyr Lys Asn Ala Leu Gln Leu Thr Asn Asn Phe			
1010	1015	1020	
Thr Asp Leu Lys Ser Ile Gly Lys Gln Thr Gly Phe Leu Phe Tyr			
1025	1030	1035	
Val Pro Ala Trp Asn Thr Ser Lys Ile Asp Pro Glu Thr Gly Phe			
1040	1045	1050	
Val Asp Leu Leu Lys Pro Arg Tyr Glu Asn Ile Ala Gln Ser Gln			
1055	1060	1065	
Ala Phe Phe Gly Lys Phe Asp Lys Ile Cys Tyr Asn Ala Asp Arg			
1070	1075	1080	
Gly Tyr Phe Glu Phe His Ile Asp Tyr Ala Lys Phe Asn Asp Lys			
1085	1090	1095	
Ala Lys Asn Ser Arg Gln Ile Trp Lys Ile Cys Ser His Gly Asp			
1100	1105	1110	
Lys Arg Tyr Val Tyr Asp Lys Thr Ala Asn Gln Asn Lys Gly Ala			
1115	1120	1125	
Thr Ile Gly Val Asn Val Asn Asp Glu Leu Lys Ser Leu Phe Thr			
1130	1135	1140	
Arg Tyr His Ile Asn Asp Lys Gln Pro Asn Leu Val Met Asp Ile			
1145	1150	1155	
Cys Gln Asn Asn Asp Lys Glu Phe His Lys Ser Leu Met Tyr Leu			
1160	1165	1170	
Leu Lys Thr Leu Leu Ala Leu Arg Tyr Ser Asn Ala Ser Ser Asp			
1175	1180	1185	

Glu Asp Phe Ile Leu Ser Pro Val Ala Asn Asp Glu Gly Val Phe
1190 1195 1200

Phe Asn Ser Ala Leu Ala Asp Asp Thr Gln Pro Gln Asn Ala Asp
1205 1210 1215

Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Leu Trp Leu Leu
1220 1225 1230

Asn Glu Leu Lys Asn Ser Asp Asp Leu Asn Lys Val Lys Leu Ala
1235 1240 1245

Ile Asp Asn Gln Thr Trp Leu Asn Phe Ala Gln Asn Arg
1250 1255 1260

<210>

27

<211> 1282

<212> PRT

<213> Eubacterium eligens

<220><221> MISC_FEATURE

<222> (1)..(1282)

<223> Genbank CDA41776 Cpf1

<400> 27

Met Asn Gly Asn Arg Ser Ile Val Tyr Arg Glu Phe Val Gly Val Thr

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 Leu Thr Asp Leu Asp Phe Thr

65 70 75 80

Leu Leu Phe Glu Leu Met Asn Ser Val Gln Ser Ser Leu Ser Lys Asp

85 90 95

Asn Lys Lys Ala Leu Glu Lys Glu His Asn Lys Met Arg Glu Gln Ile

100	105	110	
Cys Thr His Leu Gln Ser Asp Ser Asp Tyr Lys Asn Met Phe Asn Ala			
115	120	125	
Lys Leu Phe 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 Glu His Ile
435 440 445

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

Asp Met Tyr Met Asn Met Tyr His Trp Val 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 Thr 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 Asn 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 Lys Asn Ile Val Ile Lys Leu Asn Gly Gln Ala Glu Leu
 740 745 750

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

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

Thr Ile Asn Tyr Lys Val Thr Ala Arg Asn Asn Val Asn Asp Met Ala
 850 855 860
 Val Lys Tyr Ile Ala Gln Asn Asp Asp Ile His Val Ile Gly Ile Asp
 865 870 875 880

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

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

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

Phe Met Gln Phe Asp Glu Ile Arg Tyr Cys Ala Glu Lys Asp Met
 1070 1075 1080
 Phe Ser Phe Gly Phe Asp Tyr Asn Asn Phe Asp Thr Tyr Asn Ile

1085	1090	1095
Thr Met Gly Lys Thr Gln Trp	Thr Val Tyr Thr Asn	Gly Glu Arg
1100	1105	1110
Leu Gln Ser Glu Phe Asn Asn	Ala Arg Arg Thr Gly	Lys Thr Lys
1115	1120	1125
Ser Ile Asn Leu Thr Glu Thr	Ile Lys Leu Leu Leu	Glu Asp Asn
1130	1135	1140
Glu Ile Asn Tyr Ala Asp Gly	His Asp Val Arg Ile	Asp Met Glu
1145	1150	1155
Lys Met Tyr Glu Asp Lys Asn	Ser Glu Phe Phe Ala	Gln Leu Leu
1160	1165	1170
Ser Leu Tyr Lys Leu Thr Val	Gln Met Arg Asn Ser	Tyr Thr Glu
1175	1180	1185
Ala Glu Glu Gln Glu Lys Gly	Ile Ser Tyr Asp Lys	Ile Ile Ser
1190	1195	1200
Pro Val Ile Asn Asp Glu Gly	Glu Phe Phe Asp Ser	Asp Asn Tyr
1205	1210	1215
Lys Glu Ser Asp Asp Lys Glu	Cys Lys Met Pro Lys	Asp Ala Asp
1220	1225	1230
Ala Asn Gly Ala Tyr Cys Ile	Ala Leu Lys Gly Leu	Tyr Glu Val
1235	1240	1245
Leu Lys Ile Lys Ser Glu Trp	Thr Glu Asp Gly Phe	Asp Arg Asn
1250	1255	1260
Cys Leu Lys Leu Pro His Ala	Glu Trp Leu Asp Phe	Ile Gln Asn
1265	1270	1275
Lys Arg Tyr Glu		
1280		
<210>	28	
<211>	1263	
<212>	PRT	
<213>	Eubacterium rectale	
<220><221>	MISC_FEATURE	

<222> (1)..(1263)

<223> Genbank CUM80100 Cpf1

<400> 28

Met Asn Asn Gly Thr Asn Asn Phe Gln Asn Phe Ile Gly Ile Ser Ser

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

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

35 40 45
Glu Asn Arg Gln Ile Leu Lys Asp Ile Met Asp Asp Tyr Tyr Arg Gly

50 55 60
Phe Ile Ser Glu Thr Leu Ser Ser Ile Asp Asp Ile Asp Trp Thr Ser

65 70 75 80
Leu Phe Glu Lys Met Glu Ile Gln Leu Lys Asn Gly Asp Asn Lys Asp

85 90 95
Thr Leu Ile Lys Glu Gln Ala Glu Lys Arg Lys Ala Ile Tyr Lys Lys

100 105 110
Phe Ala Asp Asp Asp Arg Phe Lys Asn Met Phe Ser Ala Lys Leu Ile

115 120 125
Ser Asp Ile Leu Pro Glu Phe Val Ile His Asn Asn Asn Tyr Ser Ala

130 135 140
Ser Glu Lys Glu Glu Lys Thr Gln Val Ile Lys Leu Phe Ser Arg Phe

145 150 155 160
Ala Thr Ser Phe Lys Asp Tyr Phe Lys Asn Arg Ala Asn Cys Phe Ser

165 170 175
Ala Asp Asp Ile Ser Ser Ser Ser Cys His Arg Ile Val Asn Asp Asn

180 185 190
Ala Glu Ile Phe Phe Ser Asn Ala Leu Val Tyr Arg Arg Ile Val Lys

195 200 205
Asn Leu Ser Asn Asp Asp Ile Asn Lys Ile Ser Gly Asp Met Lys Asp

210 215 220

Ser Leu Lys Glu Met Ser Leu Asp Glu Ile Tyr Ser Tyr Glu Lys Tyr
 225 230 235 240
 Gly Glu Phe Ile Thr Gln Glu Gly Ile Ser Phe Tyr Asn Asp Ile Cys
 245 250 255
 Gly Lys Val Asn Ser Phe Met Asn Leu Tyr Cys Gln Lys Asn Lys Glu
 260 265 270
 Asn Lys Asn Leu Tyr Lys Leu Arg Lys Leu His Lys Gln Ile Leu Cys
 275 280 285
 Ile Ala Asp Thr Ser Tyr Glu Val Pro Tyr Lys Phe Glu Ser Asp Glu
 290 295 300
 Glu Val Tyr Gln Ser Val Asn Gly Phe Leu Asp Asn Ile Ser Ser Lys
 305 310 315 320
 His Ile Val Glu Arg Leu Arg Lys Ile Gly Asp Asn Tyr Asn Gly Tyr
 325 330 335
 Asn Leu Asp Lys Ile Tyr Ile Val Ser Arg Phe Tyr Glu Ser Val Ser
 340 345 350
 Gln Lys Thr Tyr Arg Asp Trp Glu Thr Ile Asn Thr Ala Leu Glu Ile
 355 360 365
 His Tyr Asn Asn Ile Leu Pro Gly Asn Gly Lys Ser Lys Ala Asp Lys
 370 375 380
 Val Lys Lys Ala Val Lys Asn Asp Leu Gln Lys Ser Ile Thr Glu Ile
 385 390 395 400
 Asn Glu Leu Val Ser Asn Tyr Lys Leu Cys Pro Asp Asp Asn Ile Lys
 405 410 415
 Ala Glu Thr Tyr Ile His Glu Ile Ser His Ile Leu Asn Asn Phe Glu
 420 425 430
 Ala Gln Glu Leu Lys Tyr Asn Pro Glu Ile His Leu Val Glu Ser Glu
 435 440 445
 Leu Lys Ala Ser Glu Leu Lys Asn Val Leu Asp Val Ile Met Asn Ala
 450 455 460
 Phe His Trp Cys Ser Val Phe Met Thr Glu Glu Leu Val Asp Lys Asp

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

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

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

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

 705 710 715 720

Thr Met Tyr Leu Lys Asn Leu Phe Ser Glu Glu Asn Leu Lys Asp Ile
725 730 735

Val Leu Lys Leu Asn Gly Glu Ala Glu Ile Phe Phe Arg Lys Ser Ser
740 745 750

Ile Lys Asn Pro Ile Ile His Lys Lys Gly Ser Ile Leu Val Asn Arg
755 760 765

Thr Tyr Glu Ala Glu Glu Lys Asp Gln Phe Gly Asn Ile Gln Ile Val
770 775 780

Arg Lys Thr Ile Pro Glu Asn Ile Tyr Gln Glu Leu Tyr Lys Tyr Phe
785 790 795 800

Asn Asp Lys Ser Asp Lys Glu Leu Ser Asp Glu Ala Ala Lys Leu Lys
805 810 815

Asn Val Val Gly His His Glu Ala Ala Thr Asn Ile Val Lys Asp Tyr
820 825 830

Arg Tyr Thr Tyr Asp Lys Tyr Phe Leu His Met Pro Ile Thr Ile Asn
835 840 845

Phe Lys Ala Asn Lys Thr Ser Phe Ile Asn Asp Arg Ile Leu Gln Tyr
850 855 860

Ile Ala Lys Glu Asn Asp Leu His Val Ile Gly Ile Asp Arg Gly Glu
865 870 875 880

Arg Asn Leu Ile Tyr Val Ser Val Ile Asp Thr Cys Gly Asn Ile Val
885 890 895

Glu Gln Lys Ser Phe Asn Ile Val Asn Gly Tyr Asp Tyr Gln Ile Lys
900 905 910

Leu Lys Gln Gln Glu Gly Ala Arg Gln Ile Ala Arg Lys Glu Trp Lys
915 920 925

Glu Ile Gly Lys Ile Lys Glu Ile Lys Glu Gly Tyr Leu Ser Leu Val
930 935 940

Ile His Glu Ile Ser Lys Met Val Ile Lys Tyr Asn Ala Ile Ile Ala
945 950 955 960

Met Glu Asp Leu Ser Tyr Gly Phe Lys Lys Gly Arg Phe Lys Val Glu

965	970	975
Arg Gln Val Tyr Gln Lys Phe Glu Thr Met Leu Ile Asn Lys Leu Asn		
980	985	990
Tyr Leu Val Phe Lys Asp Ile Ser Ile Thr Glu Asn Gly Gly Leu Leu		
995	1000	1005
Lys Gly Tyr Gln Leu Thr Tyr Ile Pro Glu Lys Leu Lys Asn Val		
1010	1015	1020
Gly His Gln Cys Gly Cys Ile Phe Tyr Val Pro Ala Ala Tyr Thr		
1025	1030	1035
Ser Lys Ile Asp Pro Thr Thr Gly Phe Ala Asn Ile Phe Lys Phe		
1040	1045	1050
Lys Asp Leu Thr Val Asp Ala Lys Arg Glu Phe Ile Lys Lys Phe		
1055	1060	1065
Asp Ser Ile Arg Tyr Asp Ser Glu Lys Asn Leu Phe Cys Phe Thr		
1070	1075	1080
Phe Asp Tyr Asn Asn Phe Ile Thr Gln Asn Thr Val Met Ser Lys		
1085	1090	1095
Ser Ser Trp Ser Val Tyr Thr Tyr Gly Val Arg Ile Lys Arg Arg		
1100	1105	1110
Phe Val Asn Gly Arg Phe Ser Asn Glu Ser Asp Thr Ile Asp Ile		
1115	1120	1125
Thr Lys Asp Met Glu Lys Thr Leu Glu Met Thr Asp Ile Asn Trp		
1130	1135	1140
Arg Asp Gly His Asp Leu Arg Gln Asp Ile Ile Asp Tyr Glu Ile		
1145	1150	1155
Val Gln His Ile Phe Glu Ile Phe Lys Leu Thr Val Gln Met Arg		
1160	1165	1170
Asn Ser Leu Ser Glu Leu Glu Asp Arg Asp Tyr Asp Arg Leu Ile		
1175	1180	1185
Ser Pro Val Leu Asn Glu Asn Asn Ile Phe Tyr Asp Ser Ala Lys		
1190	1195	1200

Ala Gly Asp Ala Leu Pro Lys Asp Ala Asp Ala Asn Gly Ala Tyr

1205 1210 1215

Cys Ile Ala Leu Lys Gly Leu Tyr Glu Ile Lys Gln Ile Thr Glu

1220 1225 1230

Asn Trp Lys Glu Asp Gly Lys Phe Ser Arg Asp Lys Leu Lys Ile

1235 1240 1245

Ser Asn Lys Asp Trp Phe Asp Phe Ile Gln Asn Lys Arg Tyr Leu

1250 1255 1260

<210> 29

<211> 966

<212> PRT

<213> Unknown

<220><223> Uncultured Bacterium (groundwater metagenome) hypothetical
protein ACD_18C00234G0001

<220><221> MISC_FEATURE

<222> (1)..(966)

<223> Genbank EKE06926 Cpf1

<400> 29

Met Glu Asn Tyr Asp Lys Cys Leu Thr Gln Leu Gly Ile Glu Lys Phe

1 5 10 15

Asn Leu Glu Ile Val Gly Glu Ile Asn Lys Glu Leu Asn Leu Phe Ser

20 25 30

Gln Gln Asn Arg Asp Ile Leu Pro Lys Ala Pro Lys Leu Lys Phe Leu

35 40 45

Tyr Lys Gln Ile Gly Cys Gly Lys Arg Ile Phe Asp Leu Phe Ser Ile

50 55 60

Ile Val Gly Asn Glu Trp Lys Glu Leu Lys Asn Leu Gln Asn Asn Lys

65 70 75 80

Asp Asp Gly Lys Ser Phe Ser Gln Lys Glu Leu Leu Glu Lys Ile Arg

85 90 95

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

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

Asn Gly Thr Lys Phe Ser Glu Ser Glu Leu Asn Lys Ile Tyr Asn Ile

355 360 365

Asn Ser Asp Ser Ile Lys Ala Lys Arg Leu Leu Tyr Asn Thr Gln Lys

370 375 380

Ile Asp Asn Lys Asn Pro Pro Arg Trp Phe Ile Arg Ser Lys Lys Thr

385 390 395 400

Thr Phe Ser Pro Met Val Arg Glu Gly Leu Leu Asp Pro Glu Ser Ile

405 410 415

Leu Glu Leu Tyr Asp Lys Lys Leu Tyr Ser Lys Thr Glu Asn Lys Asn

420 425 430

Gly Tyr Lys Glu Tyr Leu Pro Arg Leu Leu Asp Tyr Phe Lys Asp Gly

435 440 445

Phe Leu Lys His Lys Asp Phe Val Gln Phe Lys Glu Asn Phe Lys Trp

450 455 460

Leu Asp Asn Ser Glu Tyr Asp Thr Val Val Asp Phe Tyr Asn His Thr

465 470 475 480

Ala Asp Met Cys Tyr Lys Thr Ser Trp Glu Asp Ile Asn Phe Thr Glu

485 490 495

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

500 505 510

Asn Lys Asp Phe Ala Glu Lys Thr Ser Gly Ile Lys Asn Ser His Thr

515 520 525

Ile Leu Phe Leu Glu Leu Leu Lys Ser Glu Asn Asn Leu Lys Leu Lys

530 535 540

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

545 550 555 560

Glu Ile Asp Lys Glu Arg Ser Phe Lys Thr Asp Lys Phe Glu Ile Ile

565 570 575

Lys Asn Lys Arg Tyr Ser Glu Glu Lys Tyr Phe Leu His Phe Pro Ile

580 585 590

Glu Ile Lys Gly Arg Lys Leu Lys Gly Ser Phe Asn Gln Phe Leu Asn

595 600 605
Lys Glu Ile Ser Lys Lys Glu Ser Val Asn Ile Leu Gly Ile Asp Arg

610 615 620
Gly Glu Lys His Leu Leu Tyr Tyr Ser Leu Val Asn Asn Asn Gly Glu
625 630 635 640

Ile Ile Lys Gln Gly Ser Phe Asn Lys Ile Lys Cys Gly Asn Lys Ile
645 650 655

Val Asp Tyr Asn Glu Leu Leu Ser Lys Arg Ala Lys Glu Met Met Glu
660 665 670

Ala Arg Gln Ser Trp Glu Thr Ile Gly Lys Ile Lys Asp Leu Lys Glu

675 680 685
Gly Tyr Leu Ser Gln Val Ile His Glu Ile Tyr Lys Leu Val Ile Glu

690 695 700
Asn Asn Ala Ile Val Val Leu Glu Asp Leu Asn Thr Glu Phe Lys Ala

705 710 715 720
Lys Arg Thr Ala Lys Val Glu Lys Ser Val Tyr Lys Lys Phe Glu Leu

725 730 735
Ala Leu Val Lys Lys Leu Asn His Leu Ile Leu Lys Glu Lys Lys Ala

740 745 750
Asn Glu Leu Gly Gly Ser Leu Asn Ala Tyr Gln Leu Thr Pro Tyr Ile

755 760 765
Lys Pro Gly Asp Val Asp Lys Phe Glu Lys Ala Lys Gln Trp Gly Ile

770 775 780
Met Phe Tyr Val Arg Pro Asp Tyr Thr Ser Gln Thr Asp Pro Val Thr

785 790 795 800
Gly Trp Arg Lys Thr Ile Tyr Ile Ser Asn Ser Glu Thr Ile Glu Asn

805 810 815
Ile Lys Lys Lys Trp Lys Asp Ala Asn Ile Lys Ile Tyr Phe Asp Ser

820 825 830
Asp Lys Lys Cys Phe Lys Phe Leu Tyr Asp Lys Trp Glu Leu Cys Ala

835 840 845

Tyr Pro Asn Leu Glu Arg Leu Tyr Trp Asn Arg Ser Glu Lys Asn Ala
850 855 860

Glu Gly Lys Phe Gly Asn Met Lys Lys Tyr Ser Leu His Lys Glu Phe

865 870 875 880

Glu Cys Ile Phe Glu Gly Val Asn Lys Ser Lys Asn Ile Ser Asp Gln

885 890 895

Met Phe Asp Lys Glu Asp Phe Asn Trp Lys Ser Phe Ile Phe Tyr Trp

900 905 910

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

915 920 925

Glu Ser Asp Phe Ile Gln Ser Pro Ile Trp Ser Glu Lys Ile Asn Asp

930 935 940

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

945 950 955 960

Asp Ala Asn Gly Ala Tyr

965

<210> 30

<211> 1247

<212> PRT

<213> Unknown

<220><223> Uncultured bacterium (gcode 4) (groundwater metagenome)

hypothetical protein ACD_3C00058G0015

<220><221> MISC_FEATURE

<222> (1)..(1247)

<223> Genbank EKE28449 Cpf1

<400> 30

Met Phe Lys Gly Asp Ala Phe Thr Gly Leu Tyr Glu Val Gln Lys Thr

1 5 10 15

Leu Arg Phe Glu Leu Val Pro Ile Gly Leu Thr Gln Ser Tyr Leu Glu

20 25 30

Asn Asp Trp Val Ile Gln Lys Asp Lys Glu Val Glu Glu Asn Tyr Gly

35 40 45

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

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

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

 Asp Glu Asn Leu Ile Thr Phe Ile Lys Asn Ile Glu Asp Phe Lys Lys
 195 200 205
 Leu Asn Ser Ser Phe Arg Glu Phe Val Asn Asp Asn Phe Ser Glu Glu
 210 215 220
 Asp Lys Gln Ile Phe Glu Ile Asp Phe Tyr Asn Asn Cys Leu Leu Gln
 225 230 235 240
 Pro Trp Ile Asp Lys Tyr Asn Lys Ile Val Trp Trp Tyr Ser Leu Glu
 245 250 255

 Asn Trp Glu Lys Val Gln Trp Leu Asn Glu Lys Ile Asn Asn Phe Lys
 260 265 270
 Gln Asn Gln Asn Lys Ser Asn Ser Lys Asp Leu Lys Phe Pro Arg Met
 275 280 285
 Lys Leu Leu Tyr Lys Gln Ile Leu Gly Asp Lys Glu Lys Lys Val Tyr

290 295 300
 Ile Asp Glu Ile Arg Asp Asp Lys Asn Leu Ile Asp Leu Ile Asp Asn
 305 310 315 320

 Ser Lys Arg Arg Asn Gln Ile Lys Ile Asp Asn Ala Asn Asp Ile Ile
 325 330 335
 Asn Asp Phe Ile Asn Asn Asn Ala Lys Phe Glu Leu Asp Lys Ile Tyr
 340 345 350
 Leu Thr Arg Gln Ser Ile Asn Thr Ile Ser Ser Lys Tyr Phe Ser Ser
 355 360 365
 Trp Asp Tyr Ile Arg Trp Tyr Phe Trp Thr Gly Glu Leu Gln Glu Phe
 370 375 380

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

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

 Glu Tyr Leu Glu Gly Phe Glu Ile Trp Lys Asp Tyr Asn Leu Val Arg
 515 520 525
 Asn Tyr Ile Thr Lys Lys Gln Val Asn Thr Asp Lys Ile Lys Leu Asn
 530 535 540

Phe Asp Asn Ser Gln Phe Leu Thr Trp Trp Asp Lys Asp Lys Glu Asn
 545 550 555 560
 Glu Arg Leu Gly Ile Ile Leu Arg Arg Glu Trp Lys Tyr Tyr Leu Trp
 565 570 575

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

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

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

 Leu Phe Asn Glu Lys Asn Leu Gln Asn Leu Val Val Lys Leu Ser Trp
 770 775 780
 Trp Ala Asp Ile Phe Phe Arg Asp Lys Thr Glu Asn Leu Lys Phe Lys

785 790 795 800
 Lys Asp Lys Asn Gly Gln Glu Ile Leu Asp His Arg Arg Phe Ser Gln
 805 810 815
 Asp Lys Ile Met Phe His Ile Ser Ile Thr Leu Asn Ala Asn Cys Trp
 820 825 830

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

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

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

 Val Thr Cys Pro Asn Cys Trp Leu Arg Lys Asn Leu Tyr Ile Ser
 1025 1030 1035

Asn Ser Ala Thr Lys Glu Asn Gln Lys Lys Ser Leu Asn Ser Ile		
1040	1045	1050
Ala Ile Lys Tyr Asn Asp Trp Lys Phe Ser Phe Ser Tyr Glu Ile		
1055	1060	1065
Asp Asp Lys Ser Trp Lys Gln Lys Gln Ser Leu Asn Lys Lys Lys		
1070	1075	1080
Phe Ile Val Tyr Ser Asp Ile Glu Arg Phe Val Tyr Ser Pro Leu		
1085	1090	1095
Glu Lys Leu Thr Lys Val Ile Asp Val Asn Lys Lys Leu Leu Glu		
1100	1105	1110
Leu Phe Arg Asp Phe Asn Leu Ser Leu Asp Ile Asn Lys Gln Ile		
1115	1120	1125
Gln Glu Lys Asp Leu Asp Ser Val Phe Phe Lys Ser Leu Thr His		
1130	1135	1140
Leu Phe Asn Leu Ile Leu Gln Leu Arg Asn Ser Asp Ser Lys Asp		
1145	1150	1155
Asn Lys Asp Tyr Ile Ser Cys Pro Ser Cys Tyr Tyr His Ser Asn		
1160	1165	1170
Asn Trp Leu Gln Trp Phe Glu Phe Asn Trp Asp Ala Asn Trp Ala		
1175	1180	1185
Tyr Asn Ile Ala Arg Lys Gly Ile Ile Leu Leu Asp Arg Ile Arg		
1190	1195	1200
Lys Asn Gln Glu Lys Pro Asp Leu Tyr Val Ser Asp Ile Asp Trp		
1205	1210	1215
Asp Asn Phe Val Gln Ser Asn Gln Phe Pro Asn Thr Ile Ile Pro		
1220	1225	1230
Ile Gln Asn Ile Glu Lys Gln Val Pro Leu Asn Ile Lys Ile		
1235	1240	1245
<210> 31		
<211> 1250		
<212> PRT		
<213> Smithella sp.		

<220><221> MISC_FEATURE

<222> (1)..(1250)

<223> Genbank KF067989 Cpf1

<400> 31

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 Glu 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 Ile
 340 345 350
 Tyr Phe Arg Ser Gly Thr 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 Val 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 Tyr 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 Glu 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 Ser 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 Glu Asn Glu Thr His Lys Lys Gly Asp Asn
 645 650 655
 Phe Asn Leu Asn His Cys His Gln 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 Ile 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 Val 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 Thr 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 Lys 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 Gly 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 Ala 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 Leu Ala Asp Phe Phe Arg Thr Leu Met Lys Tyr Leu			
	1145	1150	1155
Ser Val Thr Leu Ser Leu Arg His Asn Asn Gly Glu Lys Gly Glu			
	1160	1165	1170
Thr Glu Gln Asp Tyr Ile Leu Ser Pro Val Ala Asp Ser Met Gly			
	1175	1180	1185
Lys Phe Phe Asp Ser Arg Lys Ala Gly Asp Asp Met Pro Lys Asn			

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 Met Gln Thr Leu

1235 1240 1245

Lys Gly

1250

<210> 32

<211> 1477

<212> PRT

<213> Candidatus Peregrinibacteria

<220><221> MISC_FEATURE

<222> (1)..(1477)

<223> Genbank KKP36646 Cpf1

<400> 32

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 Asp Lys Leu Lys Ile Ile Thr Asp Ser Gln
485 490 495

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

500 505 510
Asn Met Glu Ser Phe Leu Lys Glu Thr Asn Asp Ile Ile Thr Leu Lys
515 520 525

Lys Tyr Lys Gly Asn Lys Glu Gly Thr Glu Lys Ile Lys Gln Trp Phe
530 535 540

Asp Tyr Thr Leu Ala Ile Asn Arg Met Leu Lys Tyr Phe Leu Val Lys
545 550 555 560

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

565 570 575
Leu Lys Thr Leu Ile Tyr Ser Asp Asp Ala Glu Trp Phe Lys Trp Tyr
580 585 590

Asp Ala Leu Arg Asn Tyr Leu Thr Gln Lys Pro Gln Asp Glu Ala Lys
595 600 605

Glu Asn Lys Leu Lys Leu Asn Phe Asp Asn Pro Ser Leu Ala Gly Gly
 610 615 620
 Trp Asp Val Asn Lys Glu Cys Ser Asn Phe Cys Val Ile Leu Lys Asp

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

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

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

 820 825 830
 Ile Asn Phe Cys Lys Tyr Phe Leu Ser Lys Tyr Pro Lys Thr Thr Leu
 835 840 845
 Phe Asn Tyr Ser Phe Lys Glu Ser Glu Asn Tyr Asn Ser Leu Asp Glu

850 855 860
 Phe Tyr Arg Asp Val Asp Ile Cys Ser Tyr Lys Leu Asn Ile Asn Thr
 865 870 875 880
 Thr Ile Asn Lys Ser Ile Leu Asp Arg Leu Val Glu Glu Gly Lys Leu

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

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

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

 1070 1075 1080
 Met Ala Ser Asn Arg Asp Met Ala Arg Lys Asn Trp Gln Arg Ile
 1085 1090 1095

Gly Thr	Ile Lys Glu Ala Lys	Asn Gly Tyr Val Ser	Leu Val Ile
1100	1105	1110	
Arg Lys	Ile Ala Asp Leu Ala	Val Asn Asn Glu Arg	Pro Ala Phe
1115	1120	1125	
Ile Val	Leu Glu Asp Leu Asn	Thr Gly Phe Lys Arg	Ser Arg Gln
1130	1135	1140	
Lys Ile	Asp Lys Ser Val Tyr	Gln Lys Phe Glu Leu	Ala Leu Ala
1145	1150	1155	
Lys Lys	Leu Asn Phe Leu Val	Asp Lys Asn Ala Lys	Arg Asp Glu
1160	1165	1170	
Ile Gly	Ser Pro Thr Lys Ala	Leu Gln Leu Thr Pro	Pro Val Asn
1175	1180	1185	
Asn Tyr	Gly Asp Ile Glu Asn	Lys Lys Gln Ala Gly	Ile Met Leu
1190	1195	1200	
Tyr Thr	Arg Ala Asn Tyr Thr	Ser Gln Thr Asp Pro	Ala Thr Gly
1205	1210	1215	
Trp Arg	Lys Thr Ile Tyr Leu	Lys Ala Gly Pro Glu	Glu Thr Thr
1220	1225	1230	
Tyr Lys	Lys Asp Gly Lys Ile	Lys Asn Lys Ser Val	Lys Asp Gln
1235	1240	1245	
Ile Ile	Glu Thr Phe Thr Asp	Ile Gly Phe Asp Gly	Lys Asp Tyr
1250	1255	1260	
Tyr Phe	Glu Tyr Asp Lys Gly	Glu Phe Val Asp Glu	Lys Thr Gly
1265	1270	1275	
Glu Ile	Lys Pro Lys Lys Trp	Arg Leu Tyr Ser Gly	Glu Asn Gly
1280	1285	1290	
Lys Ser	Leu Asp Arg Phe Arg	Gly Glu Arg Glu Lys	Asp Lys Tyr
1295	1300	1305	
Glu Trp	Lys Ile Asp Lys Ile	Asp Ile Val Lys Ile	Leu Asp Asp
1310	1315	1320	
Leu Phe	Val Asn Phe Asp Lys	Asn Ile Ser Leu Leu	Lys Gln Leu

1325 1330 1335
 Lys Glu Gly Val Glu Leu Thr Arg Asn Asn Glu His Gly Thr Gly
 1340 1345 1350
 Glu Ser Leu Arg Phe Ala Ile Asn Leu Ile Gln Gln Ile Arg Asn
 1355 1360 1365
 Thr Gly Asn Asn Glu Arg Asp Asn Asp Phe Ile Leu Ser Pro Val

1370 1375 1380
 Arg Asp Glu Asn Gly Lys His Phe Asp Ser Arg Glu Tyr Trp Asp
 1385 1390 1395
 Lys Glu Thr Lys Gly Glu Lys Ile Ser Met Pro Ser Ser Gly Asp
 1400 1405 1410
 Ala Asn Gly Ala Phe Asn Ile Ala Arg Lys Gly Ile Ile Met Asn
 1415 1420 1425
 Ala His Ile Leu Ala Asn Ser Asp Ser Lys Asp Leu Ser Leu Phe

1430 1435 1440
 Val Ser Asp Glu Glu Trp Asp Leu His Leu Asn Asn Lys Thr Glu
 1445 1450 1455
 Trp Lys Lys Gln Leu Asn Ile Phe Ser Ser Arg Lys Ala Met Ala
 1460 1465 1470
 Lys Arg Lys Lys
 1475

<210> 33

<211> 1285

<212> PRT

<213> Candidatus Roizmanbacteria bacterium

<220><221> MISC_FEATURE

<222> (1)..(1285)

<223> Genbank KKQ38174 Cpf1

<400> 33

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

20 25 30
 Asn Ala Gly Val Phe Glu Lys Asp Lys Leu Ile Gln Lys Lys Tyr Gly
 35 40 45
 Lys Thr Lys Pro Tyr Phe Asp Arg Leu His Arg Glu Phe Ile Glu Glu
 50 55 60

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

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

 Thr Arg Lys Asn Phe Tyr Lys Asn Asp Gly Thr Ser Thr Ala Ile Ala
 195 200 205
 Thr Arg Ile Ile Asp Gln Asn Leu Lys Arg Phe Ile Asp Asn Leu Ser
 210 215 220
 Ile Val Glu Ser Val Arg Gln Lys Val Asp Leu Ala Glu Thr Glu Lys
 225 230 235 240
 Ser Phe Ser Ile Ser Leu Ser Gln Phe Phe Ser Ile Asp Phe Tyr Asn
 245 250 255

 Lys Cys Leu Leu Gln Asp Gly Ile Asp Tyr Tyr Asn Lys Ile Ile Gly
 260 265 270

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

 Glu Ile Lys Asn Asp Thr Glu Leu Ile Glu Ala Leu Ser Gln Phe Ala
 325 330 335
 Lys Thr Ala Glu Glu Lys Thr Lys Ile Val Lys Lys Leu Phe Ala Asp
 340 345 350
 Phe Val Glu Asn Asn Ser Lys Tyr Asp Leu Ala Gln Ile Tyr Ile Ser
 355 360 365
 Gln Glu Ala Phe Asn Thr Ile Ser Asn Lys Trp Thr Ser Glu Thr Glu
 370 375 380

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

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

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

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

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

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

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

Glu Ile His Asn Lys Asp Trp Asn Leu Asp Lys Ala Arg Asp Gly Lys

770 775 780

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

785 790 795 800

Phe Ser Asn Asp Asn Val Val Gln Asn Phe Pro Ile Lys Leu Asn Gly

805 810 815

Gln Ala Glu Ile Phe Tyr Arg Pro Lys Thr Glu Lys Asp Lys Leu Glu

820 825 830

Ser Lys Lys Asp Lys Lys Gly Asn Lys Val Ile Asp His Lys Arg Tyr

835 840 845

Ser Glu Asn Lys Ile Phe Phe His Val Pro Leu Thr Leu Asn Arg Thr

850 855 860

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

865 870 875 880

Asn Asn Lys Asp Ile Asn Ile Ile Gly Val Asp Arg Gly Glu Lys His

885 890 895

Leu Val Tyr Tyr Ser Val Ile Thr Gln Ala Ser Asp Ile Leu Glu Ser

900 905 910

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

915 920 925

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

930 935 940

Gln Gly Ile Lys Asp Leu Lys Lys Gly Tyr Ile Ser Gln Val Val Arg

945 950 955 960

Lys Leu Ala Asp Leu Ala Ile Lys His Asn Ala Ile Ile Ile Leu Glu

965 970 975

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

980 985 990

Ile Tyr Gln Gln Leu Glu Lys Ala Leu Ile Asp Lys Leu Ser Phe Leu

995 1000 1005

Val Asp Lys Gly Glu Lys Asn Pro Glu Gln Ala Gly His Leu Leu

1010	1015	1020
Lys Ala Tyr Gln Leu Ser Ala Pro Phe Glu Thr Phe Gln Lys Met		
1025	1030	1035
Gly Lys Gln Thr Gly Ile Ile Phe Tyr Thr Gln Ala Ser Tyr Thr		
1040	1045	1050
Ser Lys Ser Asp Pro Val Thr Gly Trp Arg Pro His Leu Tyr Leu		
1055	1060	1065
Lys Tyr Phe Ser Ala Lys Lys Ala Lys Asp Asp Ile Ala Lys Phe		
1070	1075	1080
Thr Lys Ile Glu Phe Val Asn Asp Arg Phe Glu Leu Thr Tyr Asp		
1085	1090	1095
Ile Lys Asp Phe Gln Gln Ala Lys Glu Tyr Pro Asn Lys Thr Val		
1100	1105	1110
Trp Lys Val Cys Ser Asn Val Glu Arg Phe Arg Trp Asp Lys Asn		
1115	1120	1125
Leu Asn Gln Asn Lys Gly Gly Tyr Thr His Tyr Thr Asn Ile Thr		
1130	1135	1140
Glu Asn Ile Gln Glu Leu Phe Thr Lys Tyr Gly Ile Asp Ile Thr		
1145	1150	1155
Lys Asp Leu Leu Thr Gln Ile Ser Thr Ile Asp Glu Lys Gln Asn		
1160	1165	1170
Thr Ser Phe Phe Arg Asp Phe Ile Phe Tyr Phe Asn Leu Ile Cys		
1175	1180	1185
Gln Ile Arg Asn Thr Asp Asp Ser Glu Ile Ala Lys Lys Asn Gly		
1190	1195	1200
Lys Asp Asp Phe Ile Leu Ser Pro Val Glu Pro Phe Phe Asp Ser		
1205	1210	1215
Arg Lys Asp Asn Gly Asn Lys Leu Pro Glu Asn Gly Asp Asp Asn		
1220	1225	1230
Gly Ala Tyr Asn Ile Ala Arg Lys Gly Ile Val Ile Leu Asn Lys		
1235	1240	1245

Ile Ser Gln Tyr Ser Glu Lys Asn Glu Asn Cys Glu Lys Met Lys
1250 1255 1260

Trp Gly Asp Leu Tyr Val Ser Asn Ile Asp Trp Asp Asn Phe Val
1265 1270 1275

Thr Gln Ala Asn Ala Arg His
1280 1285

<210> 34

<211> 1403

<212> PRT

<213> Candidatus Falkowbacteria bacterium

<220><221> MISC_FEATURE

<222> (1)..(1403)

<223> Genbank KKR91555 Cpf1

<400> 34

Met Leu Phe Phe Met Ser Thr Asp Ile Thr Asn Lys Pro Arg Glu Lys

1 5 10 15

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

20 25 30

Thr Phe Gly Leu Ile Pro Leu Lys Trp Asp Asp Asn Lys Lys Met Ile

35 40 45

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

50 55 60

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

65 70 75 80

Asn Ile Leu His Arg Glu Leu Ile Gly Lys Val Leu Gly Ser Leu Lys

85 90 95

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

100 105 110

Glu Lys Asn Asn Lys Asn Glu Asn Ile Ser Glu Asp Lys Lys Lys Glu

115 120 125

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

130 135 140

Leu Lys Lys Val Gly Glu Val Phe Glu Ser Asn Gly Ser Gly Ile Leu

145 150 155 160

Ser Ser Lys Asn Cys Leu Asp Glu Leu Thr Lys Arg Phe Thr Arg Gln

165 170 175

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

180 185 190

Pro Asp Val Ala Tyr Arg Glu Lys Asp Gly Lys Glu Glu Thr Lys Ser

195 200 205

Phe Phe Ala Met Asp Val Gly Tyr Leu Asp Asp Phe His Lys Asn Arg

210 215 220

Lys Gln Leu Tyr Ser Val Lys Gly Lys Lys Asn Ser Leu Gly Arg Arg

225 230 235 240

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

245 250 255

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

260 265 270

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

275 280 285

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

290 295 300

Ser Asn Lys Gln Glu Arg Thr Ala Asn Ile His Gly Leu Asn Gln Ile

305 310 315 320

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

325 330 335

Lys Glu Thr Gly Lys Lys Lys Ile Lys Phe Asn Lys Lys Asp Tyr Pro

340 345 350

Thr Phe Thr Cys Leu Gln Lys Gln Ile Leu Ser Gln Val Phe Arg Lys

355 360 365

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

370 375 380

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

385 390 395 400
 Glu Phe Leu Leu Asn His Glu Glu Asn Asp Ile Asp Leu Ala Met Val

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

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

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

 595 600 605
 Asn Tyr Leu Lys Arg Arg Ser Arg Asn Thr Thr Lys Glu Ile Lys Leu
 610 615 620
 His Phe Gly Asn Asp Asp Leu Leu Glu Gly Trp Asp Met Asn Lys Glu
 625 630 635 640

Pro Glu Tyr Trp Ser Phe Ile Leu Arg Asp Arg Asn Gln Tyr Tyr Leu
645 650 655

Gly Ile Gly Lys Lys Asp Gly Glu Ile Phe His Lys Lys Leu Gly Asn
660 665 670

Ser Val Glu Ala Val Lys Glu Ala Tyr Glu Leu Glu Asn Glu Ala Asp
675 680 685

Phe Tyr Glu Lys Ile Asp Tyr Lys Gln Leu Asn Ile Asp Arg Phe Glu
690 695 700

Gly Ile Ala Phe Pro Lys Lys Thr Lys Thr Glu Glu Ala Phe Arg Gln
705 710 715 720

Val Cys Lys Lys Arg Ala Asp Glu Phe Leu Gly Gly Asp Thr Tyr Glu
725 730 735

Phe Lys Ile Leu Leu Ala Ile Lys Lys Glu Tyr Asp Asp Phe Lys Ala
740 745 750

Arg Arg Gln Lys Glu Lys Asp Trp Asp Ser Lys Phe Ser Lys Glu Lys
755 760 765

Met Ser Lys Leu Ile Glu Tyr Tyr Ile Thr Cys Leu Gly Lys Arg Asp
770 775 780

Asp Trp Lys Arg Phe Asn Leu Asn Phe Arg Gln Pro Lys Glu Tyr Glu
785 790 795 800

Asp Arg Ser Asp Phe Val Arg His Ile Gln Arg Gln Ala Tyr Trp Ile
805 810 815

Asp Pro Arg Lys Val Ser Lys Asp Tyr Val Asp Lys Lys Val Ala Glu
820 825 830

Gly Glu Met Phe Leu Phe Lys Val His Asn Lys Asp Phe Tyr Asp Phe
835 840 845

Glu Arg Lys Ser Glu Asp Lys Lys Asn His Thr Ala Asn Leu Phe Thr
850 855 860

Gln Tyr Leu Leu Glu Leu Phe Ser Cys Glu Asn Ile Lys Asn Ile Lys
865 870 875 880

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

885 890 895
 Ile Arg Phe Arg Pro Lys Thr Asp Asp Val Lys Leu Lys Ile Tyr Gln
 900 905 910
 Lys Lys Gly Lys Asp Val Thr Tyr Ala Asp Lys Arg Asp Gly Asn Lys

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

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

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

 1100 1105 1110
 Val Asn Lys Leu Gly Tyr Leu Val Asp Lys Lys Tyr Asp Asn Tyr
 1115 1120 1125

Arg Asn Val Tyr Gln Phe Ala Pro Ile Val Asp Ser Val Glu Glu		
1130	1135	1140
Met Glu Lys Asn Lys Gln Ile Gly Thr Leu Val Tyr Val Pro Ala		
1145	1150	1155
Ser Tyr Thr Ser Lys Ile Cys Pro His Pro Lys Cys Gly Trp Arg		
1160	1165	1170
Glu Arg Leu Tyr Met Lys Asn Ser Ala Ser Lys Glu Lys Ile Val		
1175	1180	1185
Gly Leu Leu Lys Ser Asp Gly Ile Lys Ile Ser Tyr Asp Gln Lys		
1190	1195	1200
Asn Asp Arg Phe Tyr Phe Glu Tyr Gln Trp Glu Gln Glu His Lys		
1205	1210	1215
Ser Asp Gly Lys Lys Lys Lys Tyr Ser Gly Val Asp Lys Val Phe		
1220	1225	1230
Ser Asn Val Ser Arg Met Arg Trp Asp Val Glu Gln Lys Lys Ser		
1235	1240	1245
Ile Asp Phe Val Asp Gly Thr Asp Gly Ser Ile Thr Asn Lys Leu		
1250	1255	1260
Lys Ser Leu Leu Lys Gly Lys Gly Ile Glu Leu Asp Asn Ile Asn		
1265	1270	1275
Gln Gln Ile Val Asn Gln Gln Lys Glu Leu Gly Val Glu Phe Phe		
1280	1285	1290
Gln Ser Ile Ile Phe Tyr Phe Asn Leu Ile Met Gln Ile Arg Asn		
1295	1300	1305
Tyr Asp Lys Glu Lys Ser Gly Ser Glu Ala Asp Tyr Ile Gln Cys		
1310	1315	1320
Pro Ser Cys Leu Phe Asp Ser Arg Lys Pro Glu Met Asn Gly Lys		
1325	1330	1335
Leu Ser Ala Ile Thr Asn Gly Asp Ala Asn Gly Ala Tyr Asn Ile		
1340	1345	1350
Ala Arg Lys Gly Phe Met Gln Leu Cys Arg Ile Arg Glu Asn Pro		

1355 1360 1365
 Gln Glu Pro Met Lys Leu Ile Thr Asn Arg Glu Trp Asp Glu Ala
 1370 1375 1380
 Val Arg Glu Trp Asp Ile Tyr Ser Ala Ala Gln Lys Ile Pro Val
 1385 1390 1395
 Leu Ser Glu Glu Asn
 1400
 <210> 35

 <211> 1352
 <212> PRT
 <213> Parcubacteria group bacterium
 <220><221> MISC_FEATURE
 <222> (1)..(1352)
 <223> Genbank KKT48220 Cpf1
 <400> 35

 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 Phe 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 Glu 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> 36

<211> 1298

<212> PRT

<213> *Thiomicrospira* sp.

<220><221> MISC_FEATURE

<222> (1)..(1298)

<223> Genbank KIJ74576 Cpf1

<400> 36

Met Thr Lys Thr Phe Asp Ser Glu Phe Phe Asn Leu Tyr Ser Leu Gln

1 5 10 15

Lys Thr Val Arg Phe Glu Leu Lys Pro Val Gly Glu Thr Ala Ser Phe

20 25 30

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

35 40 45

Glu Arg Arg Ala Val Asp Tyr Gln Lys Val Lys Glu Ile Ile Asp Asp

50 55 60

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

65 70 75 80

Val Ser Lys Asp Ala Leu Glu Gln Ala Phe His Leu Tyr Gln Lys Leu

85 90 95

Lys Ala Ala Lys Val Glu Glu Arg Glu Lys Ala Leu Lys Glu Trp Glu

100 105 110

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

115 120 125

Ser Asn Lys Ala Arg Phe Ser Arg Ile Asp Lys Lys Glu Leu Ile Lys

130 135 140

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

145 150 155 160

Pro Thr Val Glu Thr Phe Asn Asn Phe Thr Thr Tyr Phe Thr Gly Phe

165 170 175

His Glu Asn Arg Lys Asn Ile Tyr Ser Lys Asp Asp His Ala Thr Ala

	180	185	190
Ile Ser Phe Arg Leu Ile His Glu Asn Leu Pro Lys Phe Phe Asp Asn			
195	200	205	
Val Ile Ser Phe Asn Lys Leu Lys Glu Gly Phe Pro Glu Leu Lys Phe			
210	215	220	
Asp Lys Val Lys Glu Asp Leu Glu Val Asp Tyr Asp Leu Lys His Ala			
225	230	235	240
Phe Glu Ile Glu Tyr Phe Val Asn Phe Val Thr Gln Ala Gly Ile Asp			
245	250	255	
Gln Tyr Asn Tyr Leu Leu Gly Gly Lys Thr Leu Glu Asp Gly Thr Lys			
260	265	270	
Lys Gln Gly Met Asn Glu Gln Ile Asn Leu Phe Lys Gln Gln Gln Thr			
275	280	285	
Arg Asp Lys Ala Arg Gln Ile Pro Lys Leu Ile Pro Leu Phe Lys Gln			
290	295	300	
Ile Leu Ser Glu Arg Thr Glu Ser Gln Ser Phe Ile Pro Lys Gln Phe			
305	310	315	320
Glu Ser Asp Gln Glu Leu Phe Asp Ser Leu Gln Lys Leu His Asn Asn			
325	330	335	
Cys Gln Asp Lys Phe Thr Val Leu Gln Gln Ala Ile Leu Gly Leu Ala			
340	345	350	
Glu Ala Asp Leu Lys Lys Val Phe Ile Lys Thr Ser Asp Leu Asn Ala			
355	360	365	
Leu Ser Asn Thr Ile Phe Gly Asn Tyr Ser Val Phe Ser Asp Ala Leu			
370	375	380	
Asn Leu Tyr Lys Glu Ser Leu Lys Thr Lys Lys Ala Gln Glu Ala Phe			
385	390	395	400
Glu Lys Leu Pro Ala His Ser Ile His Asp Leu Ile Gln Tyr Leu Glu			
405	410	415	
Gln Phe Asn Ser Ser Leu Asp Ala Glu Lys Gln Gln Ser Thr Asp Thr			
420	425	430	

Val Leu Asn Tyr Phe Ile Lys Thr Asp Glu Leu Tyr Ser Arg Phe Ile
435 440 445

Lys Ser Thr Ser Glu Ala Phe Thr Gln Val Gln Pro Leu Phe Glu Leu
450 455 460

Glu Ala Leu Ser Ser Lys Arg Arg Pro Pro Glu Ser Glu Asp Glu Gly
465 470 475 480

Ala Lys Gly Gln Glu Gly Phe Glu Gln Ile Lys Arg Ile Lys Ala Tyr
485 490 495

Leu Asp Thr Leu Met Glu Ala Val His Phe Ala Lys Pro Leu Tyr Leu
500 505 510

Val Lys Gly Arg Lys Met Ile Glu Gly Leu Asp Lys Asp Gln Ser Phe
515 520 525

Tyr Glu Ala Phe Glu Met Ala Tyr Gln Glu Leu Glu Ser Leu Ile Ile
530 535 540

Pro Ile Tyr Asn Lys Ala Arg Ser Tyr Leu Ser Arg Lys Pro Phe Lys
545 550 555 560

Ala Asp Lys Phe Lys Ile Asn Phe Asp Asn Asn Thr Leu Leu Ser Gly
565 570 575

Trp Asp Ala Asn Lys Glu Thr Ala Asn Ala Ser Ile Leu Phe Lys Lys
580 585 590

Asp Gly Leu Tyr Tyr Leu Gly Ile Met Pro Lys Gly Lys Thr Phe Leu
595 600 605

Phe Asp Tyr Phe Val Ser Ser Glu Asp Ser Glu Lys Leu Lys Gln Arg
610 615 620

Arg Gln Lys Thr Ala Glu Glu Ala Leu Ala Gln Asp Gly Glu Ser Tyr
625 630 635 640

Phe Glu Lys Ile Arg Tyr Lys Leu Leu Pro Gly Ala Ser Lys Met Leu
645 650 655

Pro Lys Val Phe Phe Ser Asn Lys Asn Ile Gly Phe Tyr Asn Pro Ser
660 665 670

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

675 680 685
 Thr Pro Gln Lys Gly His Ser Lys Val Glu Phe Asn Leu Asn Asp Cys
 690 695 700

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

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

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

 Pro Asp Val Asn Ile Ile Gly Ile Asp Arg Gly Glu Arg His Leu Leu
 900 905 910
 Tyr Phe Thr Val Val Asn Gln Lys Gly Glu Ile Leu Val Gln Glu Ser
 915 920 925

Leu Asn Thr Leu Met Ser Asp Lys Gly His Val Asn Asp Tyr Gln Gln
 930 935 940
 Lys Leu Asp Lys Lys Glu Gln Glu Arg Asp Ala Ala Arg Lys Ser Trp
 945 950 955 960

Thr Thr Val Glu Asn Ile Lys Glu Leu Lys Glu Gly Tyr Leu Ser His
 965 970 975
 Val Val His Lys Leu Ala His Leu Ile Ile Lys Tyr Asn Ala Ile Val
 980 985 990

Cys Leu Glu Asp Leu Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val
 995 1000 1005
 Glu Lys Gln Val Tyr Gln Lys Phe Glu Lys Ala Leu Ile Asp Lys
 1010 1015 1020

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

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

Tyr Gln Asn Arg Arg Asn Gln Lys Gly His Trp Glu Thr Glu Glu
 1145 1150 1155
 Val Asn Val Thr Glu Lys Leu Lys Ala Leu Phe Ala Ser Asp Ser

1160 1165 1170
 Lys Thr Thr Thr Val Ile Asp Tyr Ala Asn Asp Asp Asn Leu Ile
 1175 1180 1185
 Asp Val Ile Leu Glu Gln Asp Lys Ala Ser Phe Phe Lys Glu Leu
 1190 1195 1200

Leu Trp Leu Leu Lys Leu Thr Met Thr Leu Arg His Ser Lys Ile
 1205 1210 1215
 Lys Ser Glu Asp Asp Phe Ile Leu Ser Pro Val Lys Asn Glu Gln
 1220 1225 1230
 Gly Glu Phe Tyr Asp Ser Arg Lys Ala Gly Glu Val Trp Pro Lys
 1235 1240 1245
 Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Leu
 1250 1255 1260

Trp Asn Leu Gln Gln Ile Asn Gln Trp Glu Lys Gly Lys Thr Leu
 1265 1270 1275
 Asn Leu Ala Ile Lys Asn Gln Asp Trp Phe Ser Phe Ile Gln Glu
 1280 1285 1290
 Lys Pro Tyr Gln Glu
 1295

<210> 37

<211> 1242

<212> PRT

<213> Bacteroidales bacterium

<220><221> MISC_FEATURE

<222> (1)..(1242)

<223> Genbank KXB38146 Cpf1

<400> 37

Met Lys Lys Phe Thr Asn Leu Tyr Pro Val Gln Lys Thr Leu Arg Phe

1 5 10 15
 Glu Leu Ile Pro Gln Gly Asn Thr Ser Lys His Leu Cys Lys Ile Ile
 20 25 30
 Gln Glu Asp Glu Gln Ile Ala Glu Asp Ser Gln Glu Val Lys Lys Leu

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

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

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

 195 200 205
 Glu Ile Pro Gln Leu Glu Lys Glu Leu Val Arg Ala Gly Ala Ser Leu
 210 215 220
 Val Ser His Val Asp Ser Val Phe Thr Ile Asp Phe Phe Leu Glu Val
 225 230 235 240
 Phe Thr Gln Ser Gly Ile Asp Gln Tyr Asn Ala Leu Ile Gly Lys Ile
 245 250 255
 Val Asn Gln Glu Gln Gly Glu Val Lys Gly Leu Asn Glu Arg Ile Asn

 260 265 270
 Leu Tyr Asn Gln Gln His Lys Gln Glu Ala Lys Leu Pro Leu Phe Lys
 275 280 285

Pro Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Gln Leu Ser Trp Leu
290 295 300

Ala Glu Ala Tyr Asn Ser Asp Lys Asp Leu Leu Asp Ser Ile Gln Lys
305 310 315 320

Tyr Tyr Gln Leu Leu Ile Asp Asn Gln Ile Phe Glu Arg Ile Pro Arg
325 330 335

Leu Met His Thr Leu Glu Lys Ala Pro Leu Asp Lys Ile Trp Ile Thr
340 345 350

Tyr Asp Thr Gln Leu Thr Ser Ile Ser Asn Thr Leu Tyr Gly Ser Trp
355 360 365

Arg Val Ile Gly Glu Ala Leu Gly Arg Asn Ala Lys Ser Glu Lys Glu
370 375 380

Arg Lys Ser Ser Gln Lys Lys Ala Leu Asn Tyr Ser Leu Glu Ser Ile
385 390 395 400

Asn Gln Ala Ile Ala Lys Met Pro Ser Asp Glu Glu Leu Pro Pro Ile
405 410 415

Gln Lys Tyr Phe Ile Ala Leu Gly Ser Asn Pro Ser Lys Lys Asp Ala
420 425 430

Ser Ser Gly Thr Leu Val Asp Lys Val Arg Ser Ser Tyr Lys Ala Cys
435 440 445

Gln Asp Ile Leu Thr Asn Pro Asp His Thr Gly Lys Lys Leu Ile Gln
450 455 460

Asp Lys Lys Gln Val Asp Leu Leu Lys Gln Leu Leu Asp Asp Leu Leu
465 470 475 480

Ile Leu Gln Arg Phe Ile Lys Pro Leu Leu Tyr Ser Asn Asn Glu Asn
485 490 495

Glu Thr His Lys Asp Glu Val Phe Tyr Thr Glu Leu Thr Asp Ile Met
500 505 510

Asp Leu Leu Asn Pro Ile Val Gly Leu Tyr Asn Lys Val Arg Asn Tyr
515 520 525

Leu Thr Gln Lys Pro Tyr Ser Thr Glu Lys Phe Lys Ile Asn Phe Lys

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

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

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

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

 770 775 780

Asn Lys Lys Ser Leu Leu Asn Glu Gly Glu Thr Ser Leu Phe Asp Tyr
 785 790 795 800
 Asp Leu Ile Lys Asp Arg Arg Phe Thr Val Asp Lys Phe Gln Phe His
 805 810 815
 Val Pro Ile Thr Met Asn Phe Lys Ala Thr Gln Gly Thr Lys Val Asn
 820 825 830
 Gln Met Val Gln Glu Glu Val Lys Lys Ser Lys Gly Phe His Leu Ile
 835 840 845
 Gly Ile Asp Arg Gly Glu Arg Asn Leu Leu Tyr Ile Val Val Ile Asn
 850 855 860
 Glu Arg Gly Glu Ile Ile Glu Gln Cys Ser Leu Asn Lys Ile Val Asn
 865 870 875 880
 Thr Tyr Gln Glu Lys Glu His Thr Val Asp Tyr Lys Ala Leu Leu Glu
 885 890 895
 Lys Arg Ser Gln Ser Arg Leu Glu Glu Arg Lys Ser Trp Gln Thr Ile
 900 905 910
 Glu Asn Ile Lys Glu Leu Lys Gly Gly Tyr Leu Ser Gln Val Val His
 915 920 925
 Lys Ile Ala Gln Leu Met Ile Lys Tyr Asn Ala Ile Ala Val Leu Glu
 930 935 940
 Asp Leu Asn Phe Gly Phe Ile Arg Thr Arg Lys Lys Phe Glu Phe Ser
 945 950 955 960
 Val Tyr Gln Glu Phe Glu Lys Lys Leu Ile Asp Lys Leu Gly Tyr Val
 965 970 975
 Val Asp Lys Lys Ala Pro Ile Gln Gln Glu Gly Gly Leu Leu Gln Ala
 980 985 990
 Tyr Gln Leu Thr Ala Pro Phe Lys Ser Phe Arg Glu Met Gly Lys Gln
 995 1000 1005
 Asn Gly Phe Leu Phe Tyr Val Pro Ala Trp Asn Thr Ser Ala Ile
 1010 1015 1020
 Asp Pro Arg Thr Gly Phe Val Asn Leu Leu Asp Thr Arg Tyr Glu

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

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

1145 1150 1155
 Ser Leu Ser Asn Ser Leu Ile Asp Tyr Ile Leu Ser Pro Val Ala
 1160 1165 1170
 Asp Glu Asn Gly Glu Phe Phe Asp Ser Arg Lys Ala Leu Ser Asn
 1175 1180 1185
 Glu Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Ala Leu
 1190 1195 1200
 Lys Gly Leu Trp Val Leu Asp Lys Ile Arg Lys Thr Glu Lys Val

1205 1210 1215
 Thr Pro Ala Lys Leu Ala Leu Ser Asn Gln Glu Trp Leu Ser Phe
 1220 1225 1230
 Ala Gln Glu Lys Pro Phe Phe Asn Glu
 1235 1240

<210> 38

<211> 1300

<212> PRT

<213> Francisella tularensis

<220><221> MISC_FEATURE

<222> (1)..(1300)

<223> Genbank WP_003040289 Cpf1

<400> 38

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 Asn Val 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> 39

<211> 1310

<212> PRT

<213> Helcococcus kunzii

<220><221> MISC_FEATURE

<222> (1)..(1310)

<223> Genbank WP_005398606 Cpf1

<400> 39

Met Phe Glu Lys Leu Ser Asn Ile Val Ser Ile Ser Lys Thr Ile Arg
1 5 10 15

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

20 25 30

Gly Lys Leu Glu Lys Asp Phe Glu Arg Ser Asp Phe Tyr Pro Ile Leu
35 40 45

Lys Asn Ile Ser Asp Asp Tyr Tyr Arg Gln Tyr Ile Lys Glu Lys Leu
50 55 60

Ser Asp Leu Asn Leu Asp Trp Gln Lys Leu Tyr Asp Ala His Glu Leu
65 70 75 80
Leu Asp Ser Ser Lys Lys Glu Ser Gln Lys Asn Leu Glu Met Ile Gln
85 90 95
Ala Gln Tyr Arg Lys Val Leu Phe Asn Ile Leu Ser Gly Glu Leu Asp
100 105 110
Lys Ser Gly Glu Lys Asn Ser Lys Asp Leu Ile Lys Asn Asn Lys Ala
115 120 125
Leu Tyr Gly Lys Leu Phe Lys Lys Gln Phe Ile Leu Glu Val Leu Pro
130 135 140
Asp Phe Val Asn Asn Asn Asp Ser Tyr Ser Glu Glu Asp Leu Glu Gly
145 150 155 160
Leu Asn Leu Tyr Ser Lys Phe Thr Thr Arg Leu Lys Asn Phe Trp Glu
165 170 175
Thr Arg Lys Asn Val Phe Thr Asp Lys Asp Ile Val Thr Ala Ile Pro
180 185 190
Phe Arg Ala Val Asn Glu Asn Phe Gly Phe Tyr Tyr Asp Asn Ile Lys
195 200 205
Ile Phe Asn Lys Asn Ile Glu Tyr Leu Glu Asn Lys Ile Pro Asn Leu
210 215 220
Glu Asn Glu Leu Lys Glu Ala Asp Ile Leu Asp Asp Asn Arg Ser Val
225 230 235 240
Lys Asp Tyr Phe Thr Pro Asn Gly Phe Asn Tyr Val Ile Thr Gln Asp
245 250 255
Gly Ile Asp Val Tyr Gln Ala Ile Arg Gly Gly Phe Thr Lys Glu Asn
260 265 270
Gly Glu Lys Val Gln Gly Ile Asn Glu Ile Leu Asn Leu Thr Gln Gln
275 280 285
Gln Leu Arg Arg Lys Pro Glu Thr Lys Asn Val Lys Leu Gly Val Leu
290 295 300
Thr Lys Leu Arg Lys Gln Ile Leu Glu Tyr Ser Glu Ser Thr Ser Phe

305 310 315 320
 Leu Ile Asp Gln Ile Glu Asp Asp Asn Asp Leu Val Asp Arg Ile Asn
 325 330 335
 Lys Phe Asn Val Ser Phe Phe Glu Ser Thr Glu Val Ser Pro Ser Leu

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

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

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

 530 535 540
 Val Asn Phe Asp Phe Pro Thr Leu Ala Asp Gly Trp Ser Glu Ser Lys
 545 550 555 560

Ile Ser Asp Asn Arg Ser Ile Ile Leu Arg Lys Gly Gly Tyr Tyr Tyr
 565 570 575
 Leu Gly Ile Leu Ile Asp Asn Lys Leu Leu Ile Asn Lys Lys Asn Lys
 580 585 590
 Ser Lys Lys Ile Tyr Glu Ile Leu Ile Tyr Asn Gln Ile Pro Glu Phe
 595 600 605
 Ser Lys Ser Ile Pro Asn Tyr Pro Phe Thr Lys Lys Val Lys Glu His
 610 615 620
 Phe Lys Asn Asn Val Ser Asp Phe Gln Leu Ile Asp Gly Tyr Val Ser
 625 630 635 640
 Pro Leu Ile Ile Thr Lys Glu Ile Tyr Asp Ile Lys Lys Glu Lys Lys
 645 650 655
 Tyr Lys Lys Asp Phe Tyr Lys Asp Asn Asn Thr Asn Lys Asn Tyr Leu
 660 665 670
 Tyr Thr Ile Tyr Lys Trp Ile Glu Phe Cys Lys Gln Phe Leu Tyr Lys
 675 680 685
 Tyr Lys Gly Pro Asn Lys Glu Ser Tyr Lys Glu Met Tyr Asp Phe Ser
 690 695 700
 Thr Leu Lys Asp Thr Ser Leu Tyr Val Asn Leu Asn Asp Phe Tyr Ala
 705 710 715 720
 Asp Val Asn Ser Cys Ala Tyr Arg Val Leu Phe Asn Lys Ile Asp Glu
 725 730 735
 Asn Thr Ile Asp Asn Ala Val Glu Asp Gly Lys Leu Leu Leu Phe Gln
 740 745 750
 Ile Tyr Asn Lys Asp Phe Ser Pro Glu Ser Lys Gly Lys Lys Asn Leu
 755 760 765
 His Thr Leu Tyr Trp Leu Ser Met Phe Ser Glu Glu Asn Leu Arg Thr
 770 775 780
 Arg Lys Leu Lys Leu Asn Gly Gln Ala Glu Ile Phe Tyr Arg Lys Lys
 785 790 795 800
 Leu Glu Lys Lys Pro Ile Ile His Lys Glu Gly Ser Ile Leu Leu Asn

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

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

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

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

 1040 1045 1050

Asn Leu	Glu Pro Ser Gly Ile	Leu Asn Pro Ile Gln	Ala Cys Tyr
1055	1060	1065	
Pro Val	Asp Ala Tyr Gln Glu	Leu Gln Gly Gln Asn	Gly Ile Val
1070	1075	1080	
Phe Tyr	Leu Pro Ala Ala Tyr	Thr Ser Val Ile Asp	Pro Val Thr
1085	1090	1095	
Gly Phe	Thr Asn Leu Phe Arg	Leu Lys Ser Ile Asn	Ser Ser Lys
1100	1105	1110	
Tyr Glu	Glu Phe Ile Lys Lys	Phe Lys Asn Ile Tyr	Phe Asp Asn
1115	1120	1125	
Glu Glu	Glu Asp Phe Lys Phe	Ile Phe Asn Tyr Lys	Asp Phe Ala
1130	1135	1140	
Lys Ala	Asn Leu Val Ile Leu	Asn Asn Ile Lys Ser	Lys Asp Trp
1145	1150	1155	
Lys Ile	Ser Thr Arg Gly Glu	Arg Ile Ser Tyr Asn	Ser Lys Lys
1160	1165	1170	
Lys Glu	Tyr Phe Tyr Val Gln	Pro Thr Glu Phe Leu	Ile Asn Lys
1175	1180	1185	
Leu Lys	Glu Leu Asn Ile Asp	Tyr Glu Asn Ile Asp	Ile Ile Pro
1190	1195	1200	
Leu Ile	Asp Asn Leu Glu Glu	Lys Ala Lys Arg Lys	Ile Leu Lys
1205	1210	1215	
Ala Leu	Phe Asp Thr Phe Lys	Tyr Ser Val Gln Leu	Arg Asn Tyr
1220	1225	1230	
Asp Phe	Glu Asn Asp Tyr Ile	Ile Ser Pro Thr Ala	Asp Asp Asn
1235	1240	1245	
Gly Asn	Tyr Tyr Asn Ser Asn	Glu Ile Asp Ile Asp	Lys Thr Asn
1250	1255	1260	
Leu Pro	Asn Asn Gly Asp Ala	Asn Gly Ala Phe Asn	Ile Ala Arg
1265	1270	1275	
Lys Gly	Leu Leu Leu Lys Asp	Arg Ile Val Asn Ser	Asn Glu Ser

1280 1285 1290
 Lys Val Asp Leu Lys Ile Lys Asn Glu Asp Trp Ile Asn Phe Ile
 1295 1300 1305
 Ile Ser
 1310
 <210> 40
 <211> 1257
 <212> PRT
 <213> *Prevotella bryantii*
 <220><221> MISC_FEATURE
 <222> (1)..(1257)
 <223> Genbank WP_006283774 Cpf1
 <400> 40
 Met Gln Ile Asn Asn Leu Lys Ile Ile Tyr Met Lys Phe Thr Asp Phe
 1 5 10 15

 Thr Gly Leu Tyr Ser Leu Ser Lys Thr Leu Arg Phe Glu Leu Lys Pro
 20 25 30
 Ile Gly Lys Thr Leu Glu Asn Ile Lys Lys Ala Gly Leu Leu Glu Gln
 35 40 45
 Asp Gln His Arg Ala Asp Ser Tyr Lys Lys Val Lys Lys Ile Ile Asp
 50 55 60
 Glu Tyr His Lys Ala Phe Ile Glu Lys Ser Leu Ser Asn Phe Glu Leu
 65 70 75 80

 Lys Tyr Gln Ser Glu Asp Lys Leu Asp Ser Leu Glu Glu Tyr Leu Met
 85 90 95
 Tyr Tyr Ser Met Lys Arg Ile Glu Lys Thr Glu Lys Asp Lys Phe Ala
 100 105 110
 Lys Ile Gln Asp Asn Leu Arg Lys Gln Ile Ala Asp His Leu Lys Gly
 115 120 125
 Asp Glu Ser Tyr Lys Thr Ile Phe Ser Lys Asp Leu Ile Arg Lys Asn
 130 135 140

 Leu Pro Asp Phe Val Lys Ser Asp Glu Glu Arg Thr Leu Ile Lys Glu

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

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

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

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

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

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

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

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

	645	650	655
Lys Gly Ala Asn Phe Asn Leu Ala Asp Cys His Asn Leu Ile Asp Phe			
	660	665	670
Phe Lys Ser Ser Ile Ser Lys His Glu Asp Trp Ser Lys Phe Asn Phe			
	675	680	685
His Phe Ser Asp Thr Ser Ser Tyr Glu Asp Leu Ser Asp Phe Tyr Arg			
	690	695	700
Glu Val Glu Gln Gln Gly Tyr Ser Ile Ser Phe Cys Asp Val Ser Val			
705	710	715	720
Glu Tyr Ile Asn Lys Met Val Glu Lys Gly Asp Leu Tyr Leu Phe Gln			
	725	730	735
Ile Tyr Asn Lys Asp Phe Ser Glu Phe Ser Lys Gly Thr Pro Asn Met			
	740	745	750
His Thr Leu Tyr Trp Asn Ser Leu Phe Ser Lys Glu Asn Leu Asn Asn			
	755	760	765
Ile Ile Tyr Lys Leu Asn Gly Gln Ala Glu Ile Phe Phe Arg Lys Lys			
	770	775	780
Ser Leu Asn Tyr Lys Arg Pro Thr His Pro Ala His Gln Ala Ile Lys			
785	790	795	800
Asn Lys Asn Lys Cys Asn Glu Lys Lys Glu Ser Ile Phe Asp Tyr Asp			
	805	810	815
Leu Val Lys Asp Lys Arg Tyr Thr Val Asp Lys Phe Gln Phe His Val			
	820	825	830
Pro Ile Thr Met Asn Phe Lys Ser Thr Gly Asn Thr Asn Ile Asn Gln			
	835	840	845
Gln Val Ile Asp Tyr Leu Arg Thr Glu Asp Asp Thr His Ile Ile Gly			
	850	855	860
Ile Asp Arg Gly Glu Arg His Leu Leu Tyr Leu Val Val Ile Asp Ser			
865	870	875	880
His Gly Lys Ile Val Glu Gln Phe Thr Leu Asn Glu Ile Val Asn Glu			
	885	890	895

Tyr Gly Gly Asn Ile Tyr Arg Thr Asn Tyr His Asp Leu Leu Asp Thr
 900 905 910

Arg Glu Gln Asn Arg Glu Lys Ala Arg Glu Ser Trp Gln Thr Ile Glu
 915 920 925

Asn Ile Lys Glu Leu Lys Glu Gly Tyr Ile Ser Gln Val Ile His Lys
 930 935 940

Ile Thr Asp Leu Met Gln Lys Tyr His Ala Val Val Val Leu Glu Asp
 945 950 955 960

Leu Asn Met Gly Phe Met Arg Gly Arg Gln Lys Val Glu Lys Gln Val
 965 970 975

Tyr Gln Lys Phe Glu Glu Met Leu Ile Asn Lys Leu Asn Tyr Leu Val
 980 985 990

Asn Lys Lys Ala Asp Gln Asn Ser Ala Gly Gly Leu Leu His Ala Tyr
 995 1000 1005

Gln Leu Thr Ser Lys Phe Glu Ser Phe Gln Lys Leu Gly Lys Gln
 1010 1015 1020

Ser Gly Phe Leu Phe Tyr Ile Pro Ala Trp Asn Thr Ser Lys Ile
 1025 1030 1035

Asp Pro Val Thr Gly Phe Val Asn Leu Phe Asp Thr Arg Tyr Glu
 1040 1045 1050

Ser Ile Asp Lys Ala Lys Ala Phe Phe Gly Lys Phe Asp Ser Ile
 1055 1060 1065

Arg Tyr Asn Ala Asp Lys Asp Trp Phe Glu Phe Ala Phe Asp Tyr
 1070 1075 1080

Asn Asn Phe Thr Thr Lys Ala Glu Gly Thr Arg Thr Asn Trp Thr
 1085 1090 1095

Ile Cys Thr Tyr Gly Ser Arg Ile Arg Thr Phe Arg Asn Gln Ala
 1100 1105 1110

Lys Asn Ser Gln Trp Asp Asn Glu Glu Ile Asp Leu Thr Lys Ala
 1115 1120 1125

Tyr Lys Ala Phe Phe Ala Lys His Gly Ile Asn Ile Tyr Asp Asn

1130 1135 1140
 Ile Lys Glu Ala Ile Ala Met Glu Thr Glu Lys Ser Phe Phe Glu
 1145 1150 1155

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

Gly Leu Met Leu Ile Gln Gln Ile Lys Asp Ser Thr Ser Ser Asn
 1220 1225 1230
 Arg Phe Lys Phe Ser Pro Ile Thr Asn Lys Asp Trp Leu Ile Phe
 1235 1240 1245
 Ala Gln Glu Lys Pro Tyr Leu Asn Asp
 1250 1255

<210> 41

<211> 1262

<212> PRT

<213> Bacteroidetes oral

<220><221> MISC_FEATURE

<222> (1)..(1262)

<223> Genbank WP_009217842 Cpf1

<400> 41

Met Arg Lys Phe Asn Glu Phe Val Gly Leu Tyr Pro Ile Ser Lys Thr

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

50 55 60
 Ala Leu Ser Gly Phe Thr Phe Asp Thr Glu Ala Asp Gly Arg Ser Asn

 65 70 75 80
 Asn Ser Leu Ser Glu Tyr Tyr Leu Tyr Tyr Asn Leu Lys Lys Arg Asn

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

 100 105 110
 Gln Ile Val Asn Lys Leu Thr Gln Ser Glu Lys Tyr Lys Arg Ile Asp

 115 120 125
 Lys Lys Glu Leu Ile Thr Thr Asp Leu Pro Asp Phe Leu Thr Asn Glu

 130 135 140
 Ser Glu Lys Glu Leu Val Glu Lys Phe Lys Asn Phe Thr Thr Tyr Phe

 145 150 155 160
 Thr Glu Phe His Lys Asn Arg Lys Asn Met Tyr Ser Lys Glu Glu Lys

 165 170 175
 Ser Thr Ala Ile Ala Phe Arg Leu Ile Asn Glu Asn Leu Pro Lys Phe

 180 185 190
 Val Asp Asn Ile Ala Ala Phe Glu Lys Val Val Ser Ser Pro Leu Ala

 195 200 205
 Glu Lys Ile Asn Ala Leu Tyr Glu Asp Phe Lys Glu Tyr Leu Asn Val

 210 215 220
 Glu Glu Ile Ser Arg Val Phe Arg Leu Asp Tyr Tyr Asp Glu Leu Leu

 225 230 235 240
 Thr Gln Lys Gln Ile Asp Leu Tyr Asn Ala Ile Val Gly Gly Arg Thr

 245 250 255
 Glu Glu Asp Asn Lys Ile Gln Ile Lys Gly Leu Asn Gln Tyr Ile Asn

 260 265 270
 Glu Tyr Asn Gln Gln Gln Thr Asp Arg Ser Asn Arg Leu Pro Lys Leu

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

 290 295 300

Leu Pro Pro Lys Phe Asp Ser Asp Lys Asn Leu Leu Ile Lys Ile Lys
 305 310 315 320
 Glu Cys Tyr Asp Ala Leu Ser Glu Lys Glu Lys Val Phe Asp Lys Leu

 325 330 335
 Glu Ser Ile Leu Lys Ser Leu Ser Thr Tyr Asp Leu Ser Lys Ile Tyr
 340 345 350
 Ile Ser Asn Asp Ser Gln Leu Ser Tyr Ile Ser Gln Lys Met Phe Gly
 355 360 365
 Arg Trp Asp Ile Ile Ser Lys Ala Ile Arg Glu Asp Cys Ala Lys Arg
 370 375 380
 Asn Pro Gln Lys Ser Arg Glu Ser Leu Glu Lys Phe Ala Glu Arg Ile

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

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

 515 520 525
 Leu Asp Gln Val Thr Pro Leu Tyr Asn Met Val Arg Asn Tyr Leu Thr
 530 535 540
 Ser Lys Pro Tyr Ser Thr Lys Lys Ile Lys Leu Asn Phe Glu Asn Ser

545 550 555 560
 Thr Leu Met Asp Gly Trp Asp Leu Asn Lys Glu Pro Asp Asn Thr Thr
 565 570 575
 Val Ile Phe Cys Lys Asp Gly Leu Tyr Tyr Leu Gly Ile Met Gly Lys

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

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

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

 770 775 780
 Ile Thr Val Gln Ser Pro Thr His Pro Ala Asn Ser Pro Ile Lys Asn
 785 790 795 800

Lys Asn Lys Asp Asn Gln Lys Lys Glu Ser Lys Phe Glu Tyr Asp Leu
 805 810 815
 Ile Lys Asp Arg Arg Tyr Thr Val Asp Lys Phe Leu Phe His Val Pro
 820 825 830
 Ile Thr Met Asn Phe Lys Ser Val Gly Gly Ser Asn Ile Asn Gln Leu
 835 840 845
 Val Lys Arg His Ile Arg Ser Ala Thr Asp Leu His Ile Ile Gly Ile
 850 855 860
 Asp Arg Gly Glu Arg His Leu Leu Tyr Leu Thr Val Ile Asp Ser Arg
 865 870 875 880
 Gly Asn Ile Lys Glu Gln Phe Ser Leu Asn Glu Ile Val Asn Glu Tyr
 885 890 895
 Asn Gly Asn Thr Tyr Arg Thr Asp Tyr His Glu Leu Leu Asp Thr Arg
 900 905 910
 Glu Gly Glu Arg Thr Glu Ala Arg Arg Asn Trp Gln Thr Ile Gln Asn
 915 920 925
 Ile Arg Glu Leu Lys Glu Gly Tyr Leu Ser Gln Val Ile His Lys Ile
 930 935 940
 Ser Glu Leu Ala Ile Lys Tyr Asn Ala Val Ile Val Leu Glu Asp Leu
 945 950 955 960
 Asn Phe Gly Phe Met Arg Ser Arg Gln Lys Val Glu Lys Gln Val Tyr
 965 970 975
 Gln Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Leu Val Asp
 980 985 990
 Lys Lys Lys Pro Val Ala Glu Thr Gly Gly Leu Leu Arg Ala Tyr Gln
 995 1000 1005
 Leu Thr Gly Glu Phe Glu Ser Phe Lys Thr Leu Gly Lys Gln Ser
 1010 1015 1020
 Gly Ile Leu Phe Tyr Val Pro Ala Trp Asn Thr Ser Lys Ile Asp
 1025 1030 1035
 Pro Val Thr Gly Phe Val Asn Leu Phe Asp Thr His Tyr Glu Asn

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

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

1145 1150 1155
 Glu Leu Leu Arg Leu Leu Arg Leu Thr Leu Gln Met Arg Asn Ser
 1160 1165 1170
 Met Pro Ser Ser Asp Ile Asp Tyr Leu Ile Ser Pro Val Ala Asn
 1175 1180 1185
 Asp Thr Gly Cys Phe Phe Asp Ser Arg Lys Gln Ala Glu Leu Lys
 1190 1195 1200
 Glu Asn Ala Val Leu Pro Met Asn Ala Asp Ala Asn Gly Ala Tyr

1205 1210 1215
 Asn Ile Ala Arg Lys Gly Leu Leu Ala Ile Arg Lys Met Lys Gln
 1220 1225 1230
 Glu Glu Asn Asp Ser Ala Lys Ile Ser Leu Ala Ile Ser Asn Lys
 1235 1240 1245
 Glu Trp Leu Lys Phe Ala Gln Thr Lys Pro Tyr Leu Glu Asp
 1250 1255 1260

<210> 42

<211> 1318

<212> PRT

<213> Flavobacterium branchiophilum

<220><221> MISC_FEATURE

<222> (1)..(1318)

<223> Genbank WP_014085038 Cpf1

<400> 42

Met Thr Asn Lys Phe Thr Asn Gln Tyr Ser Leu Ser Lys Thr Leu Arg

1 5 10 15

Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Phe Ile Gln Glu Lys

20 25 30

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

35 40 45

Lys Lys Thr Ile Asp Lys Phe His Lys Tyr Phe Ile Asp Leu Ala Leu

50 55 60

Ser Asn Ala Lys Leu Thr His Leu Glu Thr Tyr Leu Glu Leu Tyr Asn

65 70 75 80

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

85 90 95

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

100 105 110

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

115 120 125

Thr Val Glu Leu Glu Lys Trp Phe Glu Asn Asn Glu Gln Lys Asp Ile

130 135 140

Tyr Phe Asp Glu Lys Phe Lys Thr Phe Thr Thr Tyr Phe Thr Gly Phe

145 150 155 160

His Gln Asn Arg Lys Asn Met Tyr Ser Val Glu Pro Asn Ser Thr Ala

165 170 175

Ile Ala Tyr Arg Leu Ile His Glu Asn Leu Pro Lys Phe Leu Glu Asn

180 185 190

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

195 200 205

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

450 455 460
 Lys Lys Glu Asn Glu Thr Asp Lys Thr Phe Asp Phe Ile Ala Asn Ile
 465 470 475 480
 Thr Ala Lys Tyr Gln Cys Ile Gln Gly Ile Leu Glu Asn Ala Asp Gln
 485 490 495
 Tyr Glu Asp Glu Leu Lys Gln Asp Gln Lys Leu Ile Asp Asn Leu Lys

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

 565 570 575
 Ile Lys Leu Asn Phe Glu Asn Ala Gln Leu Leu Asn Gly Trp Asp Ala
 580 585 590
 Asn Lys Glu Gly Asp Tyr Leu Thr Thr Ile Leu Lys Lys Asp Gly Asn
 595 600 605
 Tyr Phe Leu Ala Ile Met Asp Lys Lys His Asn Lys Ala Phe Gln Lys
 610 615 620
 Phe Pro Glu Gly Lys Glu Asn Tyr Glu Lys Met Val Tyr Lys Leu Leu

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

 690 695 700

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

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

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

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

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

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

 1070 1075 1080
 Gly Phe Leu Phe Tyr Val Pro Ala Trp Asn Thr Ser Lys Ile Asp
 1085 1090 1095
 Pro Thr Thr Gly Phe Val Asn Tyr Phe Tyr Thr Lys Tyr Glu Asn
 1100 1105 1110
 Val Asp Lys Ala Lys Ala Phe Phe Glu Lys Phe Glu Ala Ile Arg
 1115 1120 1125
 Phe Asn Ala Glu Lys Lys Tyr Phe Glu Phe Glu Val Lys Lys Tyr

 1130 1135 1140
 Ser Asp Phe Asn Pro Lys Ala Glu Gly Thr Gln Gln Ala Trp Thr
 1145 1150 1155
 Ile Cys Thr Tyr Gly Glu Arg Ile Glu Thr Lys Arg Gln Lys Asp
 1160 1165 1170
 Gln Asn Asn Lys Phe Val Ser Thr Pro Ile Asn Leu Thr Glu Lys
 1175 1180 1185

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

1190 1195 1200

Asn Cys Ile Lys Ser Gln Ile Ala Ser Lys Asp Asp Lys Ala Phe

1205 1210 1215

Phe Glu Thr Leu Leu Tyr Trp Phe Lys Met Thr Leu Gln Met Arg

1220 1225 1230

Asn Ser Glu Thr Arg Thr Asp Ile Asp Tyr Leu Ile Ser Pro Val

1235 1240 1245

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

1250 1255 1260

Leu Glu Asn Pro Thr Leu Pro Lys Asp Ala Asp Ala Asn Gly Ala

1265 1270 1275

Tyr His Ile Ala Lys Lys Gly Leu Met Leu Leu Asn Lys Ile Asp

1280 1285 1290

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

1295 1300 1305

Asp Trp Leu Gln Phe Val Gln Lys Asn Lys

1310 1315

<210> 43

<211> 1230

<212> PRT

<213> Lachnospiraceae bacterium

<220><221> MISC_FEATURE

<222> (1)..(1230)

<223> Genbank WP_016301126 Cpf1

<400> 43

Met His Glu Asn Asn Gly Lys Ile Ala Asp Asn Phe Ile Gly Ile Tyr

1 5 10 15

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

20 25 30

Gln Glu Tyr Ile Glu Lys His Gly Ile Leu Asp Glu Asp Leu Lys Arg

35 40 45
 Ala Gly Asp Tyr Lys Ser Val Lys Lys Ile Ile Asp Ala Tyr His Lys
 50 55 60
 Tyr Phe Ile Asp Glu Ala Leu Asn Gly Ile Gln Leu Asp Gly Leu Lys
 65 70 75 80
 Asn Tyr Tyr Glu Leu Tyr Glu Lys Lys Arg Asp Asn Asn Glu Glu Lys
 85 90 95
 Glu Phe Gln Lys Ile Gln Met Ser Leu Arg Lys Gln Ile Val Lys Arg
 100 105 110

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

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

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

Ser Asp Arg Asp Lys Ile Ser Phe Ile Pro Glu Gln Phe Asp Ser Asp
290 295 300

Thr Glu Val Leu Glu Ala Val Asp Met Phe Tyr Asn Arg Leu Leu Gln
305 310 315 320

Phe Val Ile Glu Asn Glu Gly Gln Ile Thr Ile Ser Lys Leu Leu Thr
325 330 335

Asn Phe Ser Ala Tyr Asp Leu Asn Lys Ile Tyr Val Lys Asn Asp Thr
340 345 350

Thr Ile Ser Ala Ile Ser Asn Asp Leu Phe Asp Asp Trp Ser Tyr Ile
355 360 365

Ser Lys Ala Val Arg Glu Asn Tyr Asp Ser Glu Asn Val Asp Lys Asn
370 375 380

Lys Arg Ala Ala Ala Tyr Glu Glu Lys Lys Glu Lys Ala Leu Ser Lys
385 390 395 400

Ile Lys Met Tyr Ser Ile Glu Glu Leu Asn Phe Phe Val Lys Lys Tyr
405 410 415

Ser Cys Asn Glu Cys His Ile Glu Gly Tyr Phe Glu Arg Arg Ile Leu
420 425 430

Glu Ile Leu Asp Lys Met Arg Tyr Ala Tyr Glu Ser Cys Lys Ile Leu
435 440 445

His Asp Lys Gly Leu Ile Asn Asn Ile Ser Leu Cys Gln Asp Arg Gln
450 455 460

Ala Ile Ser Glu Leu Lys Asp Phe Leu Asp Ser Ile Lys Glu Val Gln
465 470 475 480

Trp Leu Leu Lys Pro Leu Met Ile Gly Gln Glu Gln Ala Asp Lys Glu
485 490 495

Glu Ala Phe Tyr Thr Glu Leu Leu Arg Ile Trp Glu Glu Leu Glu Pro
500 505 510

Ile Thr Leu Leu Tyr Asn Lys Val Arg Asn Tyr Val Thr Lys Lys Pro
515 520 525

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

530 535 540
 Asp Gly Trp Asp Lys Asn Lys Glu Lys Asp Asn Leu Gly Ile Ile Leu
 545 550 555 560

 Leu Lys Asp Gly Gln Tyr Tyr Leu Gly Ile Met Asn Arg Arg Asn Asn
 565 570 575
 Lys Ile Ala Asp Asp Ala Pro Leu Ala Lys Thr Asp Asn Val Tyr Arg
 580 585 590
 Lys Met Glu Tyr Lys Leu Leu Thr Lys Val Ser Ala Asn Leu Pro Arg
 595 600 605
 Ile Phe Leu Lys Asp Lys Tyr Asn Pro Ser Glu Glu Met Leu Glu Lys
 610 615 620

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

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

 Glu Ile Phe Tyr Arg Lys Ala Ser Ile Asn Gln Lys Asp Val Val Val
 755 760 765
 His Lys Ala Asp Leu Pro Ile Lys Asn Lys Asp Pro Gln Asn Ser Lys
 770 775 780

Lys Glu Ser Met Phe Asp Tyr Asp Ile Ile Lys Asp Lys Arg Phe Thr
 785 790 795 800
 Cys Asp Lys Tyr Gln Phe His Val Pro Ile Thr Met Asn Phe Lys Ala
 805 810 815

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

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

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

Ala Trp Asn Thr Ser Lys Leu Asp Pro Thr Thr Gly Phe Val Asn
 1010 1015 1020
 Leu Phe Tyr Thr Lys Tyr Glu Ser Val Glu Lys Ser Lys Glu Phe

1025	1030	1035
Ile Asn Asn Phe Thr Ser Ile	Leu Tyr Asn Gln Glu	Arg Glu Tyr
1040	1045	1050
Phe Glu Phe Leu Phe Asp Tyr	Ser Ala Phe Thr Ser	Lys Ala Glu
1055	1060	1065
Gly Ser Arg Leu Lys Trp Thr	Val Cys Ser Lys Gly	Glu Arg Val
1070	1075	1080
Glu Thr Tyr Arg Asn Pro Lys	Lys Asn Asn Glu Trp	Asp Thr Gln
1085	1090	1095
Lys Ile Asp Leu Thr Phe Glu	Leu Lys Lys Leu Phe	Asn Asp Tyr
1100	1105	1110
Ser Ile Ser Leu Leu Asp Gly	Asp Leu Arg Glu Gln	Met Gly Lys
1115	1120	1125
Ile Asp Lys Ala Asp Phe Tyr	Lys Lys Phe Met Lys	Leu Phe Ala
1130	1135	1140
Leu Ile Val Gln Met Arg Asn	Ser Asp Glu Arg Glu	Asp Lys Leu
1145	1150	1155
Ile Ser Pro Val Leu Asn Lys	Tyr Gly Ala Phe Phe	Glu Thr Gly
1160	1165	1170
Lys Asn Glu Arg Met Pro Leu	Asp Ala Asp Ala Asn	Gly Ala Tyr
1175	1180	1185
Asn Ile Ala Arg Lys Gly Leu	Trp Ile Ile Glu Lys	Ile Lys Asn
1190	1195	1200
Thr Asp Val Glu Gln Leu Asp	Lys Val Lys Leu Thr	Ile Ser Asn
1205	1210	1215
Lys Glu Trp Leu Gln Tyr Ala	Gln Glu His Ile Leu	
1220	1225	1230
<210> 44		
<211> 1246		
<212> PRT		
<213> Porphyromonas macacae		
<220><221> MISC_FEATURE		

<222> (1)..(1246)

<223> Genbank WP_018359861 Cpf1

<400> 44

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 Glu 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 Ile Ile Gly Ile Asp Arg Gly Glu Arg
850 855 860

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

Gln Phe Ser Leu Asn Val Ile Glu Ser Asp Lys Gly Asp Leu Arg Thr
885 890 895

Asp Tyr Gln Lys Ile Leu Gly Asp Arg Glu Gln Glu Arg Leu Arg Arg
900 905 910

Arg Gln Glu Trp Lys Ser Ile Glu Ser Ile Lys Asp Leu Lys Asp Gly
915 920 925

Tyr Met Ser Gln Val Val His Lys Ile Cys Asn Met Val Val Glu His
930 935 940

Lys Ala Ile Val Val Leu Glu Asn Leu Asn Leu Ser Phe Met Lys Gly
945 950 955 960

Arg Lys Lys Val Glu Lys Ser Val Tyr Glu Lys Phe Glu Arg Met Leu

965	970	975
Val Asp Lys Leu Asn Tyr Leu Val	Val Asp Lys Lys Asn Leu Ser Asn	
980	985	990
Glu Pro Gly Gly Leu Tyr Ala Ala	Tyr Gln Leu Thr Asn Pro Leu Phe	
995	1000	1005
Ser Phe Glu Glu Leu His Arg Tyr	Pro Gln Ser Gly Ile Leu Phe	
1010	1015	1020
Phe Val Asp Pro Trp Asn Thr	Ser Leu Thr Asp Pro Ser Thr Gly	
1025	1030	1035
Phe Val Asn Leu Leu Gly Arg Ile	Asn Tyr Thr Asn Val Gly Asp	
1040	1045	1050
Ala Arg Lys Phe Phe Asp Arg Phe	Asn Ala Ile Arg Tyr Asp Gly	
1055	1060	1065
Lys Gly Asn Ile Leu Phe Asp Leu	Asp Leu Ser Arg Phe Asp Val	
1070	1075	1080
Arg Val Glu Thr Gln Arg Lys Leu	Trp Thr Leu Thr Thr Phe Gly	
1085	1090	1095
Ser Arg Ile Ala Lys Ser Lys Lys	Ser Gly Lys Trp Met Val Glu	
1100	1105	1110
Arg Ile Glu Asn Leu Ser Leu Cys	Phe Leu Glu Leu Phe Glu Gln	
1115	1120	1125
Phe Asn Ile Gly Tyr Arg Val Glu	Lys Asp Leu Lys Lys Ala Ile	
1130	1135	1140
Leu Ser Gln Asp Arg Lys Glu Phe	Tyr Val Arg Leu Ile Tyr Leu	
1145	1150	1155
Phe Asn Leu Met Met Gln Ile Arg	Asn Ser Asp Gly Glu Glu Asp	
1160	1165	1170
Tyr Ile Leu Ser Pro Ala Leu Asn	Glu Lys Asn Leu Gln Phe Asp	
1175	1180	1185
Ser Arg Leu Ile Glu Ala Lys Asp	Leu Pro Val Asp Ala Asp Ala	
1190	1195	1200

Asn Gly Ala Tyr Asn Val Ala Arg Lys Gly Leu Met Val Val Gln
1205 1210 1215

Arg Ile Lys Arg Gly Asp His Glu Ser Ile His Arg Ile Gly Arg
1220 1225 1230

Ala Gln Trp Leu Arg Tyr Val Gln Glu Gly Ile Val Glu
1235 1240 1245

<210> 45
<211> 1259
<212> PRT
<213> Synergistes jonesii
<220><221> MISC_FEATURE
<222> (1)..(1259)
<223> Genbank WP_037975888 Type V CV CRISPR-associated protein Cpf1

<400> 45

Met Ala Asn Ser Leu Lys Asp Phe Thr Asn Ile Tyr Gln Leu Ser Lys
1 5 10 15

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

Asn Arg Lys Leu Ile Ile Met His Asp Glu Lys Arg Gly Glu Asp Tyr
35 40 45

Lys Ser Val Thr Lys Leu Ile Asp Asp Tyr His Arg Lys Phe Ile His
50 55 60

Glu Thr Leu Asp Pro Ala His Phe Asp Trp Asn Pro Leu Ala Glu Ala
65 70 75 80

Leu Ile Gln Ser Gly Ser Lys Asn Asn Lys Ala Leu Pro Ala Glu Gln
85 90 95

Lys Glu Met Arg Glu Lys Ile Ile Ser Met Phe Thr Ser Gln Ala Val
100 105 110

Tyr Lys Lys Leu Phe Lys Lys Glu Leu Phe Ser Glu Leu Leu Pro Glu
115 120 125

Met Ile Lys Ser Glu Leu Val Ser Asp Leu Glu Lys Gln Ala Gln Leu
130 135 140

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

 Val Arg Ala Tyr Thr Leu Ile Lys Glu Arg Ala Pro Glu Val Ile Asp
 195 200 205
 Lys Ala Gln Lys Glu Leu Ser Gly Ile Leu Gly Gly Lys Thr Leu Asp
 210 215 220
 Asp Ile Phe Ser Ile Glu Ser Phe Asn Asn Val Leu Thr Gln Asp Lys
 225 230 235 240
 Ile Asp Tyr Tyr Asn Gln Ile Ile Gly Gly Val Ser Gly Lys Ala Gly
 245 250 255

 Asp Lys Lys Leu Arg Gly Val Asn Glu Phe Ser Asn Leu Tyr Arg Gln
 260 265 270
 Gln His Pro Glu Val Ala Ser Leu Arg Ile Lys Met Val Pro Leu Tyr
 275 280 285
 Lys Gln Ile Leu Ser Asp Arg Thr Thr Leu Ser Phe Val Pro Glu Ala
 290 295 300
 Leu Lys Asp Asp Glu Gln Ala Ile Asn Ala Val Asp Gly Leu Arg Ser
 305 310 315 320

 Glu Leu Glu Arg Asn Asp Ile Phe Asn Arg Ile Lys Arg Leu Phe Gly
 325 330 335
 Lys Asn Asn Leu Tyr Ser Leu Asp Lys Ile Trp Ile Lys Asn Ser Ser
 340 345 350
 Ile Ser Ala Phe Ser Asn Glu Leu Phe Lys Asn Trp Ser Phe Ile Glu
 355 360 365
 Asp Ala Leu Lys Glu Phe Lys Glu Asn Glu Phe Asn Gly Ala Arg Ser
 370 375 380

 Ala Gly Lys Lys Ala Glu Lys Trp Leu Lys Ser Lys Tyr Phe Ser Phe

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

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

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

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

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

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

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

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

885

890

895

Asp Phe Asn Ile Ile Gly Gly Met Asp Tyr His Ala Lys Leu Asn Gln

900

905

910

Lys Glu Lys Glu Arg Asp Thr Ala Arg Lys Ser Trp Lys Thr Ile Gly

915

920

925

Thr Ile Lys Glu Leu Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu

930

935

940

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

945

950

955

960

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

965

970

975

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

980

985

990

Phe Lys Asp Ala Gly Tyr Asp Ala Pro Cys Gly Ile Leu Lys Gly Leu

995

1000

1005

Gln Leu Thr Glu Lys Phe Glu Ser Phe Thr Lys Leu Gly Lys Gln

1010

1015

1020

Cys Gly Ile Ile Phe Tyr Ile Pro Ala Gly Tyr Thr Ser Lys Ile

1025

1030

1035

Asp Pro Thr Thr Gly Phe Val Asn Leu Phe Asn Ile Asn Asp Val

1040

1045

1050

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

1055

1060

1065

Ile Arg Phe Asp Ala Lys Arg Asp Met Phe Thr Phe Glu Phe Asp

1070

1075

1080

Tyr Asp Lys Phe Arg Thr Tyr Gln Thr Ser Tyr Arg Lys Lys Trp

1085

1090

1095

Ala Val Trp Thr Asn Gly Lys Arg Ile Val Arg Glu Lys Asp Lys

1100

1105

1110

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

1115

1120

1125

Lys Asn Ile Leu Asn Lys Tyr Ala Leu Ala Tyr Lys Ala Gly Glu
1130 1135 1140

Asp Ile Leu Pro Asp Val Ile Ser Arg Asp Lys Ser Leu Ala Ser
1145 1150 1155

Glu Ile Phe Tyr Val Phe Lys Asn Thr Leu Gln Met Arg Asn Ser
1160 1165 1170

Lys Arg Asp Thr Gly Glu Asp Phe Ile Ile Ser Pro Val Leu Asn
1175 1180 1185

Ala Lys Gly Arg Phe Phe Asp Ser Arg Lys Thr Asp Ala Ala Leu
1190 1195 1200

Pro Ile Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Ala Leu Lys
1205 1210 1215

Gly Ser Leu Val Leu Asp Ala Ile Asp Glu Lys Leu Lys Glu Asp
1220 1225 1230

Gly Arg Ile Asp Tyr Lys Asp Met Ala Val Ser Asn Pro Lys Trp
1235 1240 1245

Phe Glu Phe Met Gln Thr Arg Lys Phe Asp Phe
1250 1255

<210> 46

<211> 1263

<212> PRT

<213> Leptospira inadai

<220><221> MISC_FEATURE

<222> (1)..(1263)

<223> Genbank WP_020988726 Cpf1

<400> 46

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

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

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

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

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

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

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

 Asp Lys Lys Glu Val Gly Lys Ile Lys Asp Phe Leu Asp Ser Ile Lys
 500 505 510
 Ser Leu Gln Phe Phe Leu Lys Pro Leu Leu Ser Ala Glu Ile Phe Asp
 515 520 525
 Glu Lys Asp Leu Gly Phe Tyr Asn Gln Leu Glu Gly Tyr Tyr Glu Glu

530 535 540
 Ile Asp Ser Ile Gly His Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr
 545 550 555 560

 Gly Lys Ile Tyr Ser Lys Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser
 565 570 575
 Thr Leu Leu Lys Gly Trp Asp Glu Asn Arg Glu Val Ala Asn Leu Cys
 580 585 590
 Val Ile Phe Arg Glu Asp Gln Lys Tyr Tyr Leu Gly Val Met Asp Lys
 595 600 605
 Glu Asn Asn Thr Ile Leu Ser Asp Ile Pro Lys Val Lys Pro Asn Glu
 610 615 620

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

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

 Tyr Asn Lys Asp Phe Ser Ile Phe Ser Lys Gly Lys Pro Asn Leu His
 755 760 765
 Thr Ile Tyr Phe Arg Ser Leu Phe Ser Lys Glu Asn Leu Lys Asp Val
 770 775 780

Cys Leu Lys Leu Asn Gly Glu Ala Glu Met Phe Phe Arg Lys Lys Ser
785 790 795 800
Ile Asn Tyr Asp Glu Lys Lys Lys Arg Glu Gly His His Pro Glu Leu
805 810 815

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

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

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

Val Leu Lys Ala Tyr Gln Leu Thr Asp Glu Phe Gln Ser Phe Glu
1010 1015 1020
Lys Leu Ser Lys Gln Thr Gly Phe Leu Phe Tyr Val Pro Ser Trp

1025	1030	1035
Asn Thr Ser Lys Ile Asp Pro	Arg Thr Gly Phe Ile	Asp Phe Leu
1040	1045	1050
His Pro Ala Tyr Glu Asn Ile	Glu Lys Ala Lys Gln	Trp Ile Asn
1055	1060	1065
Lys Phe Asp Ser Ile Arg Phe	Asn Ser Lys Met Asp	Trp Phe Glu
1070	1075	1080
Phe Thr Ala Asp Thr Arg Lys	Phe Ser Glu Asn Leu	Met Leu Gly
1085	1090	1095
Lys Asn Arg Val Trp Val Ile	Cys Thr Thr Asn Val	Glu Arg Tyr
1100	1105	1110
Phe Thr Ser Lys Thr Ala Asn	Ser Ser Ile Gln Tyr	Asn Ser Ile
1115	1120	1125
Gln Ile Thr Glu Lys Leu Lys	Glu Leu Phe Val Asp	Ile Pro Phe
1130	1135	1140
Ser Asn Gly Gln Asp Leu Lys	Pro Glu Ile Leu Arg	Lys Asn Asp
1145	1150	1155
Ala Val Phe Phe Lys Ser Leu	Leu Phe Tyr Ile Lys	Thr Thr Leu
1160	1165	1170
Ser Leu Arg Gln Asn Asn Gly	Lys Lys Gly Glu Glu	Glu Lys Asp
1175	1180	1185
Phe Ile Leu Ser Pro Val Val	Asp Ser Lys Gly Arg	Phe Phe Asn
1190	1195	1200
Ser Leu Glu Ala Ser Asp Asp	Glu Pro Lys Asp Ala	Asp Ala Asn
1205	1210	1215
Gly Ala Tyr His Ile Ala Leu	Lys Gly Leu Met Asn	Leu Leu Val
1220	1225	1230
Leu Asn Glu Thr Lys Glu Glu	Asn Leu Ser Arg Pro	Lys Trp Lys
1235	1240	1245
Ile Lys Asn Lys Asp Trp Leu	Glu Phe Val Trp Glu	Arg Asn Arg
1250	1255	1260

<210> 47

<211> 1260

<212> PRT

<213> Porphyromonas crevioricanis

<220><221> MISC_FEATURE

<222> (1)..(1260)

<223> Genbank WP_023941260 Cpf1

<400> 47

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 Val 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 Arg 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
Ala 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> 48

<211> 1253

<212> PRT

<213> Prevotella albensis

<220><221> MISC_FEATURE

<222> (1)..(1253)

<223> Genbank WP_024988992 Cpf1

<400> 48

Met Asn Ile Lys Asn Phe Thr Gly Leu Tyr Pro Leu Ser Lys Thr Leu
 1 5 10 15

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

Asn Gly Ile Leu Thr Lys Asp Glu Gln Arg Ala Lys Asp Tyr Leu Ile
 35 40 45

Val Lys Gly Phe Ile Asp Glu Tyr His Lys Gln Phe Ile Lys Asp Arg
 50 55 60

Leu Trp Asp Phe Lys Leu Pro Leu Glu Ser Glu Gly Glu Lys Asn Ser
 65 70 75 80

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

Gln Glu Ala Asp Phe Thr Glu Ile Lys Asp Asn Leu Arg Ser Ser Ile
100 105 110

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

Lys Glu Phe Ile Arg Glu Asp Leu Val Asn Phe Leu Glu Asp Glu Lys
130 135 140

Asp Lys Asn Ile Val Lys Gln Phe Glu Asp Phe Thr Thr Tyr Phe Thr
145 150 155 160

Gly Phe Tyr Glu Asn Arg Lys Asn Met Tyr Ser Ser Glu Glu Lys Ser
165 170 175

Thr Ala Ile Ala Tyr Arg Leu Ile His Gln Asn Leu Pro Lys Phe Met
180 185 190

Asp Asn Met Arg Ser Phe Ala Lys Ile Ala Asn Ser Ser Val Ser Glu
195 200 205

His Phe Ser Asp Ile Tyr Glu Ser Trp Lys Glu Tyr Leu Asn Val Asn
210 215 220

Ser Ile Glu Glu Ile Phe Gln Leu Asp Tyr Phe Ser Glu Thr Leu Thr
225 230 235 240

Gln Pro His Ile Glu Val Tyr Asn Tyr Ile Ile Gly Lys Lys Val Leu
245 250 255

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

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

Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Lys Leu Ser Trp Ile Ala
290 295 300

Glu Glu Phe Asp Ser Asp Lys Lys Met Leu Ser Ala Ile Thr Glu Ser
305 310 315 320

Tyr Asn His Leu His Asn Val Leu Met Gly Asn Glu Asn Glu Ser Leu
325 330 335

Arg Asn Leu Leu Leu Asn Ile Lys Asp Tyr Asn Leu Glu Lys Ile Asn

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

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

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

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

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

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

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

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

835 840 845
 Val Ile Arg Gln Asn Gly Ile Thr His Ile Ile Gly Ile Asp Arg Gly
 850 855 860
 Glu Arg His Leu Leu Tyr Leu Ser Leu Ile Asp Leu Lys Gly Asn Ile
 865 870 875 880
 Ile Lys Gln Met Thr Leu Asn Glu Ile Ile Asn Glu Tyr Lys Gly Val
 885 890 895
 Thr Tyr Lys Thr Asn Tyr His Asn Leu Leu Glu Lys Arg Glu Lys Glu
 900 905 910

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

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

 Gly Phe Val Asn Leu Phe Asn Thr Arg Tyr Glu Ser Ile Lys Glu
 1040 1045 1050
 Thr Lys Val Phe Trp Ser Lys Phe Asp Ile Ile Arg Tyr Asn Lys
 1055 1060 1065
 Glu Lys Asn Trp Phe Glu Phe Val Phe Asp Tyr Asn Thr Phe Thr
 1070 1075 1080

Thr Lys Ala Glu Gly Thr Arg Thr Lys Trp Thr Leu Cys Thr His
1085 1090 1095

Gly Thr Arg Ile Gln Thr Phe Arg Asn Pro Glu Lys Asn Ala Gln
1100 1105 1110

Trp Asp Asn Lys Glu Ile Asn Leu Thr Glu Ser Phe Lys Ala Leu
1115 1120 1125

Phe Glu Lys Tyr Lys Ile Asp Ile Thr Ser Asn Leu Lys Glu Ser
1130 1135 1140

Ile Met Gln Glu Thr Glu Lys Lys Phe Phe Gln Glu Leu His Asn
1145 1150 1155

Leu Leu His Leu Thr Leu Gln Met Arg Asn Ser Val Thr Gly Thr
1160 1165 1170

Asp Ile Asp Tyr Leu Ile Ser Pro Val Ala Asp Glu Asp Gly Asn
1175 1180 1185

Phe Tyr Asp Ser Arg Ile Asn Gly Lys Asn Phe Pro Glu Asn Ala
1190 1195 1200

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

Ile Arg Gln Ile Lys Gln Ala Asp Pro Gln Lys Lys Phe Lys Phe
1220 1225 1230

Glu Thr Ile Thr Asn Lys Asp Trp Leu Lys Phe Ala Gln Asp Lys
1235 1240 1245

Pro Tyr Leu Lys Asp
1250

<210> 49

<211> 1231

<212> PRT

<213> Butyrivibrio fibrisolvens

<220><221> MISC_FEATURE

<222> (1)..(1231)

<223> Genbank WP_027216152 Cpf1

<400> 49

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

1 5 10 15

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

20 25 30

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

35 40 45

Val Lys Ala Ile Met Asp Asp Tyr His Lys Arg Leu Ile Asn Glu Ala

50 55 60

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

65 70 75 80

Leu Ser Arg Ser Lys Ser Ala Asp Asp Ile Glu Ser Phe Ser Arg Phe

85 90 95

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

100 105 110

Asn Phe Gly Lys Ile Gly Asn Lys Asp Ile Ile Pro Leu Leu Gln Lys

115 120 125

Leu Ser Glu Asn Glu Asp Asp Tyr Asn Ala Leu Glu Ser Phe Lys Asn

130 135 140

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

145 150 155 160

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

165 170 175

Asn Leu Pro Arg Phe Leu Asp Asn Ile Arg Ala Tyr Asp Ala Val Gln

180 185 190

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

195 200 205

Leu Phe Leu Val Asn Thr Phe Asn Asn Val Leu Ile Gln Asp Gly Ile

210 215 220

Asn Thr Tyr Asn Glu Asp Ile Gly Lys Leu Asn Val Ala Ile Asn Leu

225 230 235 240

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

245 250 255
Met Lys Val Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Glu Ser Phe

260 265 270
Ile Asp Glu Phe Glu Ser Asp Thr Glu Leu Leu Asp Ser Leu Glu Ser

275 280 285
His Tyr Ala Asn Leu Ala Lys Tyr Phe Gly Ser Asn Lys Val Gln Leu

290 295 300
Leu Phe Thr Ala Leu Arg Glu Ser Lys Gly Val Asn Val Tyr Val Lys

305 310 315 320
Asn Asp Ile Ala Lys Thr Ser Phe Ser Asn Val Val Phe Gly Ser Trp

325 330 335
Ser Arg Ile Asp Glu Leu Ile Asn Gly Glu Tyr Asp Asp Asn Asn Asn

340 345 350
Arg Lys Lys Asp Glu Lys Tyr Tyr Asp Lys Arg Gln Lys Glu Leu Lys

355 360 365
Lys Asn Lys Ser Tyr Thr Ile Glu Lys Ile Ile Thr Leu Ser Thr Glu

370 375 380
Asp Val Asp Val Ile Gly Lys Tyr Ile Glu Lys Leu Glu Ser Asp Ile

385 390 395 400
Asp Asp Ile Arg Phe Lys Gly Lys Asn Phe Tyr Glu Ala Val Leu Cys

405 410 415
Gly His Asp Arg Ser Lys Lys Leu Ser Lys Asn Lys Gly Ala Val Glu

420 425 430
Ala Ile Lys Gly Tyr Leu Asp Ser Val Lys Asp Phe Glu Arg Asp Leu

435 440 445
Lys Leu Ile Asn Gly Ser Gly Gln Glu Leu Glu Lys Asn Leu Val Val

450 455 460
Tyr Gly Glu Gln Glu Ala Val Leu Ser Glu Leu Ser Gly Ile Asp Ser

465 470 475 480
Leu Tyr Asn Met Thr Arg Asn Tyr Leu Thr Lys Lys Pro Phe Ser Thr

485 490 495

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

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

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

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

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

 965 970 975
 Ala Leu Gln Leu Thr Pro Asp Phe Lys Ser Phe Lys Val Leu Gly Lys
 980 985 990

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

1220 1225 1230

<210> 50

<211> 1235

<212> PRT

<213> Anaerovibrio sp.

<220><221> MISC_FEATURE

<222> (1)..(1235)

<223> Genbank WP_027407524 Cpf1

<400> 50

Met Val Ala Phe Ile Asp Glu Phe Val Gly Gln Tyr Pro Val Ser Lys

1 5 10 15

Thr Leu Arg Phe Glu Ala Arg Pro Val Pro Glu Thr Lys Lys Trp Leu

20 25 30

Glu Ser Asp Gln Cys Ser Val Leu Phe Asn Asp Gln Lys Arg Asn Glu

35 40 45

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

50 55 60

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

65 70 75 80

Asn Ala Tyr Asp Leu Tyr Cys Asn Arg Asp Thr Asn Ala Phe Ser Ser

85 90 95

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

100 105 110

Lys Asp Tyr Leu Leu Gly Ser Asp Gln Leu Lys Asp Leu Val Lys Leu

115 120 125

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

130 135 140

Val Asp Ser Arg Leu Ile Asn Trp Leu Asn Asn Asn Ala Lys Tyr Ser

145 150 155 160

Ala Glu Asp Arg Glu Lys Tyr Ile Lys Ala Ile Glu Ser Phe Glu Gly

165 170 175

Phe Val Thr Tyr Leu Thr Asn Tyr Lys Gln Ala Arg Glu Asn Met Phe

180 185 190
 Ser Ser Glu Asp Lys Ser Thr Ala Ile Ala Phe Arg Val Ile Asp Gln
 195 200 205
 Asn Met Val Thr Tyr Phe Gly Asn Ile Arg Ile Tyr Glu Lys Ile Lys

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

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

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

 405 410 415
 Asp Lys Phe Asn Leu Lys Asp Tyr Phe Ile Gln Lys Asn Ala Leu Asp
 420 425 430

Ala Thr Leu Pro Asp Leu Asp Lys Ile Thr Gln Tyr Lys Pro His Leu
435 440 445
Asp Ala Met Leu Gln Ala Ile Arg Lys Tyr Lys Leu Phe Ser Met Tyr
450 455 460
Asn Gly Arg Lys Lys Met Asp Val Pro Glu Asn Gly Ile Asp Phe Ser
465 470 475 480
Asn Glu Phe Asn Ala Ile Tyr Asp Lys Leu Ser Glu Phe Ser Ile Leu
485 490 495
Tyr Asp Arg Ile Arg Asn Phe Ala Thr Lys Lys Pro Tyr Ser Asp Glu
500 505 510
Lys Met Lys Leu Ser Phe Asn Met Pro Thr Met Leu Ala Gly Trp Asp
515 520 525
Tyr Asn Asn Glu Thr Ala Asn Gly Cys Phe Leu Phe Ile Lys Asp Gly
530 535 540
Lys Tyr Phe Leu Gly Val Ala Asp Ser Lys Ser Lys Asn Ile Phe Asp
545 550 555 560
Phe Lys Lys Asn Pro His Leu Leu Asp Lys Tyr Ser Ser Lys Asp Ile
565 570 575
Tyr Tyr Lys Val Lys Tyr Lys Gln Val Ser Gly Ser Ala Lys Met Leu
580 585 590
Pro Lys Val Val Phe Ala Gly Ser Asn Glu Lys Ile Phe Gly His Leu
595 600 605
Ile Ser Lys Arg Ile Leu Glu Ile Arg Glu Lys Lys Leu Tyr Thr Ala
610 615 620
Ala Ala Gly Asp Arg Lys Ala Val Ala Glu Trp Ile Asp Phe Met Lys
625 630 635 640
Ser Ala Ile Ala Ile His Pro Glu Trp Asn Glu Tyr Phe Lys Phe Lys
645 650 655
Phe Lys Asn Thr Ala Glu Tyr Asp Asn Ala Asn Lys Phe Tyr Glu Asp
660 665 670
Ile Asp Lys Gln Thr Tyr Ser Leu Glu Lys Val Glu Ile Pro Thr Glu

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

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

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

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

915 920 925

Leu Gln Tyr Asn Ala Val Ile Val Leu Glu Asn Leu Asn Val Gly Phe
 930 935 940
 Lys Arg Gly Arg Thr Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu
 945 950 955 960
 Lys Ala Met Ile Asp Lys Leu Asn Tyr Leu Val Phe Lys Asp Arg Gly
 965 970 975
 Tyr Glu Met Asn Gly Ser Tyr Ala Lys Gly Leu Gln Leu Thr Asp Lys
 980 985 990
 Phe Glu Ser Phe Asp Lys Ile Gly Lys Gln Thr Gly Cys Ile Tyr Tyr
 995 1000 1005
 Val Ile Pro Ser Tyr Thr Ser His Ile Asp Pro Lys Thr Gly Phe
 1010 1015 1020
 Val Asn Leu Leu Asn Ala Lys Leu Arg Tyr Glu Asn Ile Thr Lys
 1025 1030 1035
 Ala Gln Asp Thr Ile Arg Lys Phe Asp Ser Ile Ser Tyr Asn Ala
 1040 1045 1050
 Lys Ala Asp Tyr Phe Glu Phe Ala Phe Asp Tyr Arg Ser Phe Gly
 1055 1060 1065
 Val Asp Met Ala Arg Asn Glu Trp Val Val Cys Thr Cys Gly Asp
 1070 1075 1080
 Leu Arg Trp Glu Tyr Ser Ala Lys Thr Arg Glu Thr Lys Ala Tyr
 1085 1090 1095
 Ser Val Thr Asp Arg Leu Lys Glu Leu Phe Lys Ala His Gly Ile
 1100 1105 1110
 Asp Tyr Val Gly Gly Glu Asn Leu Val Ser His Ile Thr Glu Val
 1115 1120 1125
 Ala Asp Lys His Phe Leu Ser Thr Leu Leu Phe Tyr Leu Arg Leu
 1130 1135 1140
 Val Leu Lys Met Arg Tyr Thr Val Ser Gly Thr Glu Asn Glu Asn
 1145 1150 1155
 Asp Phe Ile Leu Ser Pro Val Glu Tyr Ala Pro Gly Lys Phe Phe

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

 1220 1225 1230
 Asn Gly
 1235
 <210> 51
 <211> 1205
 <212> PRT
 <213> Pseudobutyrvibrio ruminis
 <220><221> MISC_FEATURE
 <222> (1)..(1205)
 <223> Genbank WP_028248456 Cpf1
 <400> 51
 Met Tyr Tyr Gln Asn Leu Thr Lys Met Tyr Pro Ile Ser Lys Thr Leu
 1 5 10 15
 Arg Asn Glu Leu Ile Pro Val Gly Lys Thr Leu Glu Asn Ile Arg Lys
 20 25 30

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

 Gln Asp Lys Leu Arg Lys Glu Ile Val Ser Leu Leu Lys Asn His Glu

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

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

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

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

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

 Arg Lys His Asp Lys Asn Lys Ser Leu Arg Lys Asn Thr Asn Ala Val
 420 425 430
 Glu Ala Ile Lys Ser Tyr Leu Asp Thr Val Lys Asp Phe Glu Arg Asp
 435 440 445
 Ile Lys Leu Ile Asn Gly Ser Gly Gln Glu Val Glu Lys Asn Leu Val
 450 455 460
 Val Tyr Ala Glu Gln Glu Asn Ile Leu Ala Glu Ile Lys Asn Val Asp
 465 470 475 480

 Ser Leu Tyr Asn Met Ser Arg Asn Tyr Leu Thr Gln Lys Pro Phe Ser
 485 490 495
 Thr Glu Lys Phe Lys Leu Asn Phe Asn Arg Ala Thr Leu Leu Asn Gly
 500 505 510
 Trp Asp Lys Asn Lys Glu Thr Asp Asn Leu Gly Ile Leu Phe Glu Lys
 515 520 525
 Asp Gly Met Tyr Tyr Leu Gly Ile Met Asn Thr Lys Ala Asn Lys Ile
 530 535 540

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

595 600 605
 Glu Asp Cys His Ala Leu Ile Asp Phe Phe Lys Ala Ser Ile Glu Lys
 610 615 620
 His Pro Asp Trp Ser Ser Phe Gly Phe Glu Phe Ser Glu Thr Cys Thr
 625 630 635 640
 Tyr Glu Asp Leu Ser Gly Phe Tyr Arg Glu Val Glu Lys Gln Gly Tyr
 645 650 655
 Lys Ile Thr Tyr Thr Asp Val Asp Ala Asp Tyr Ile Thr Ser Leu Val
 660 665 670

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

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

 Ala Leu Arg Asn Asp Glu Lys Val Arg Val Ile Gly Ile Asp Arg Gly
 805 810 815
 Glu Arg Asn Leu Leu Tyr Val Val Val Val Asp Thr Asp Gly Thr Ile
 820 825 830
 Leu Glu Gln Ile Ser Leu Asn Ser Ile Ile Asn Asn Glu Tyr Ser Ile
 835 840 845

Glu Thr Asp Tyr His Lys Leu Leu Asp Glu Lys Glu Gly Asp Arg Asp
850 855 860

Arg Ala Arg Lys Asn Trp Thr Thr Ile Glu Asn Ile Lys Glu Leu Lys
865 870 875 880

Glu Gly Tyr Leu Ser Gln Val Val Asn Val Ile Ala Lys Leu Val Leu
885 890 895

Lys Tyr Asn Ala Ile Ile Cys Leu Glu Asp Leu Asn Phe Gly Phe Lys
900 905 910

Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu Lys
915 920 925

Met Leu Ile Asp Lys Leu Asn Tyr Leu Val Ile Asp Lys Ser Arg Lys
930 935 940

Gln Asp Lys Pro Glu Glu Phe Gly Gly Ala Leu Asn Ala Leu Gln Leu
945 950 955 960

Thr Ser Lys Phe Thr Ser Phe Lys Asp Met Gly Lys Gln Thr Gly Ile
965 970 975

Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile Asp Pro Thr Thr
980 985 990

Gly Phe Ala Asn Leu Phe Tyr Val Lys Tyr Glu Asn Val Glu Lys Ala
995 1000 1005

Lys Glu Phe Phe Ser Arg Phe Asp Ser Ile Ser Tyr Asn Asn Glu
1010 1015 1020

Ser Gly Tyr Phe Glu Phe Ala Phe Asp Tyr Lys Lys Phe Thr Asp
1025 1030 1035

Arg Ala Cys Gly Ala Arg Ser Gln Trp Thr Val Cys Thr Tyr Gly
1040 1045 1050

Glu Arg Ile Ile Lys Phe Arg Asn Thr Glu Lys Asn Asn Ser Phe
1055 1060 1065

Asp Asp Lys Thr Ile Val Leu Ser Glu Glu Phe Lys Glu Leu Phe
1070 1075 1080

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

1085 1090 1095
Lys Ile Met Ser Val Asp Glu Ala Asp Phe Phe Arg Ser Leu Thr
1100 1105 1110

Arg Leu Phe Gln Gln Thr Met Gln Met Arg Asn Ser Ser Asn Asp
1115 1120 1125

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

Glu Phe Phe Asn Ser Glu Ala Cys Asp Ala Ser Lys Pro Lys Asp
1145 1150 1155

Ala Asp Ala Asn Gly Ala Phe Asn Ile Ala Arg Lys Gly Leu Trp
1160 1165 1170

Val Leu Glu Gln Ile Arg Asn Thr Pro Ser Gly Asp Lys Leu Asn
1175 1180 1185

Leu Ala Met Ser Asn Ala Glu Trp Leu Glu Tyr Ala Gln Arg Asn
1190 1195 1200

Gln Ile

1205

<210> 52

<211> 1154

<212> PRT

<213> *Proteocatella sphenisci*

<220><221> MISC_FEATURE

<222> (1)..(1154)

<223> Genbank WP_028830240 Cpf1

<400> 52

Met Glu Asn Phe Lys Asn Leu Tyr Pro Ile Asn Lys Thr Leu Arg Phe

1 5 10 15
Glu Leu Arg Pro Tyr Gly Lys Thr Leu Glu Asn Phe Lys Lys Ser Gly

20 25 30
Leu Leu Glu Lys Asp Ala Phe Lys Ala Asn Ser Arg Arg Ser Met Gln

35 40 45
Ala Ile Ile Asp Glu Lys Phe Lys Glu Thr Ile Glu Glu Arg Leu Lys

50 55 60
 Tyr Thr Glu Phe Ser Glu Cys Asp Leu Gly Asn Met Thr Ser Lys Asp

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

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

 195 200 205
 His Tyr Asn Glu Ile Ile Gly Gly Ile Ser Lys Ser Glu Asn Val Lys
 210 215 220
 Ile Gln Gly Ile Asn Glu Gly Ile Asn Leu Tyr Cys Gln Lys Asn Lys
 225 230 235 240
 Val Lys Leu Pro Arg Leu Thr Pro Leu Tyr Lys Met Ile Leu Ser Asp
 245 250 255
 Arg Val Ser Asn Ser Phe Val Leu Asp Thr Ile Glu Asn Asp Thr Glu

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

Leu Gly Asn Leu Pro Gly Ile Ser Tyr Ser Ser Ile Val Asn Ala Ile
 305 310 315 320
 Cys Ser Asp Tyr Asp Asn Asn Phe Gly Asp Gly Lys Arg Lys Lys Ser
 325 330 335
 Tyr Glu Asn Asp Arg Lys Lys His Leu Glu Thr Asn Val Tyr Ser Ile
 340 345 350
 Asn Tyr Ile Ser Glu Leu Leu Thr Asp Thr Asp Val Ser Ser Asn Ile
 355 360 365
 Lys Met Arg Tyr Lys Glu Leu Glu Gln Asn Tyr Gln Val Cys Lys Glu
 370 375 380
 Asn Phe Asn Ala Thr Asn Trp Met Asn Ile Lys Asn Ile Lys Gln Ser
 385 390 395 400
 Glu Lys Thr Asn Leu Ile Lys Asp Leu Leu Asp Ile Leu Lys Ser Ile
 405 410 415
 Gln Arg Phe Tyr Asp Leu Phe Asp Ile Val Asp Glu Asp Lys Asn Pro
 420 425 430
 Ser Ala Glu Phe Tyr Thr Trp Leu Ser Lys Asn Ala Glu Lys Leu Asp
 435 440 445
 Phe Glu Phe Asn Ser Val Tyr Asn Lys Ser Arg Asn Tyr Leu Thr Arg
 450 455 460
 Lys Gln Tyr Ser Asp Lys Lys Ile Lys Leu Asn Phe Asp Ser Pro Thr
 465 470 475 480
 Leu Ala Lys Gly Trp Asp Ala Asn Lys Glu Ile Asp Asn Ser Thr Ile
 485 490 495
 Ile Met Arg Lys Phe Asn Asn Asp Arg Gly Asp Tyr Asp Tyr Phe Leu
 500 505 510
 Gly Ile Trp Asn Lys Ser Thr Pro Ala Asn Glu Lys Ile Ile Pro Leu
 515 520 525
 Glu Asp Asn Gly Leu Phe Glu Lys Met Gln Tyr Lys Leu Tyr Pro Asp
 530 535 540
 Pro Ser Lys Met Leu Pro Lys Gln Phe Leu Ser Lys Ile Trp Lys Ala

545 550 555 560
 Lys His Pro Thr Thr Pro Glu Phe Asp Lys Lys Tyr Lys Glu Gly Arg
 565 570 575
 His Lys Lys Gly Pro Asp Phe Glu Lys Glu Phe Leu His Glu Leu Ile

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

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

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

 770 775 780
 Arg His Leu Ile Tyr Leu Thr Val Val Asp Val Ser Thr Gly Glu Ile
 785 790 795 800

Val Glu Gln Lys His Leu Asp Glu Ile Ile Asn Thr Asp Thr Lys Gly
805 810 815

Val Glu His Lys Thr His Tyr Leu Asn Lys Leu Glu Glu Lys Ser Lys
820 825 830

Thr Arg Asp Asn Glu Arg Lys Ser Trp Glu Ala Ile Glu Thr Ile Lys
835 840 845

Glu Leu Lys Glu Gly Tyr Ile Ser His Val Ile Asn Glu Ile Gln Lys
850 855 860

Leu Gln Glu Lys Tyr Asn Ala Leu Ile Val Met Glu Asn Leu Asn Tyr
865 870 875 880

Gly Phe Lys Asn Ser Arg Ile Lys Val Glu Lys Gln Val Tyr Gln Lys
885 890 895

Phe Glu Thr Ala Leu Ile Lys Lys Phe Asn Tyr Ile Ile Asp Lys Lys
900 905 910

Asp Pro Glu Thr Tyr Ile His Gly Tyr Gln Leu Thr Asn Pro Ile Thr
915 920 925

Thr Leu Asp Lys Ile Gly Asn Gln Ser Gly Ile Val Leu Tyr Ile Pro
930 935 940

Ala Trp Asn Thr Ser Lys Ile Asp Pro Val Thr Gly Phe Val Asn Leu
945 950 955 960

Leu Tyr Ala Asp Asp Leu Lys Tyr Lys Asn Gln Glu Gln Ala Lys Ser
965 970 975

Phe Ile Gln Lys Ile Asp Asn Ile Tyr Phe Glu Asn Gly Glu Phe Lys
980 985 990

Phe Asp Ile Asp Phe Ser Lys Trp Asn Asn Arg Tyr Ser Ile Ser Lys
995 1000 1005

Thr Lys Trp Thr Leu Thr Ser Tyr Gly Thr Arg Ile Gln Thr Phe
1010 1015 1020

Arg Asn Pro Gln Lys Asn Asn Lys Trp Asp Ser Ala Glu Tyr Asp
1025 1030 1035

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

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

1085 1090 1095
 Asp Ser Arg Glu Asn Ile Lys Asn Leu Pro Ala Asp Ala Asp Ala
 1100 1105 1110
 Asn Gly Ala Tyr Asn Ile Ala Arg Lys Gly Ile Met Ala Ile Glu
 1115 1120 1125
 Asn Ile Met Asn Gly Ile Ser Asp Pro Leu Lys Ile Ser Asn Glu
 1130 1135 1140
 Asp Tyr Leu Lys Tyr Ile Gln Asn Gln Gln Glu
 1145 1150

<210> 53

<211> 1206

<212> PRT

<213> Butyrivibrio sp.

<220><221> MISC_FEATURE

<222> (1)..(1206)

<223> Genbank WP_035798880 Cpf1

<400> 53

Met Tyr Tyr Gln Asn Leu Thr Lys Lys Tyr Pro Val Ser Lys Thr Ile
 1 5 10 15
 Arg Asn Glu Leu Ile Pro Ile Gly Lys Thr Leu Glu Asn Ile Arg Lys
 20 25 30
 Asn Asn Ile Leu Glu Ser Asp Val Lys Arg Lys Gln Asp Tyr Glu His
 35 40 45

 Val Lys Gly Ile Met Asp Glu Tyr His Lys Gln Leu Ile Asn Glu Ala
 50 55 60
 Leu Asp Asn Tyr Met Leu Pro Ser Leu Asn Gln Ala Ala Glu Ile Tyr

65	70	75	80
Leu Lys Lys His Val Asp Val Glu Asp Arg Glu Glu Phe Lys Lys Thr			
	85	90	95
Gln Asp Leu Leu Arg Arg Glu Val Thr Gly Arg Leu Lys Glu His Glu			
	100	105	110
Asn Tyr Thr Lys Ile Gly Lys Lys Asp Ile Leu Asp Leu Leu Glu Lys			
	115	120	125
Leu Pro Ser Ile Ser Glu Glu 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 Glu Glu Lys Ser Ser Thr Val Ala Tyr Arg Leu Ile Asn			
	165	170	175
Glu Asn Leu Pro Lys Phe Leu Asp Asn Ile Lys Ser Tyr Ala Phe Val			
	180	185	190
Lys Ala Ala Gly Val Leu Ala Asp Cys Ile Glu Glu Glu Glu Gln Asp			
	195	200	205
Ala Leu Phe Met Val Glu Thr Phe Asn Met Thr Leu Thr Gln Glu Gly			
	210	215	220
Ile Asp Met Tyr Asn Tyr Gln Ile Gly Lys Val Asn Ser Ala Ile Asn			
225	230	235	240
Leu Tyr Asn Gln Lys Asn His Lys Val Glu Glu Phe Lys Lys Ile Pro			
	245	250	255
Lys Met Lys Val Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Glu Val			
	260	265	270
Phe Ile Gly Glu Phe Lys Asp Asp Glu Thr Leu Leu Ser Ser Ile Gly			
	275	280	285
Ala Tyr Gly Asn Val Leu Met Thr Tyr Leu Lys Ser Glu Lys Ile Asn			
	290	295	300
Ile Phe Phe Asp Ala Leu Arg Glu Ser Glu Gly Lys Asn Val Tyr Val			
305	310	315	320

Lys Asn Asp Leu Ser Lys Thr Thr Met Ser Asn Ile Val Phe Gly Ser
 325 330 335
 Trp Ser Ala Phe Asp Glu 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 Thr Leu Glu Gln Met Ser Asn Leu Ser
 370 375 380
 Lys Glu Asp Ile Ser Pro Ile Glu Asn Tyr Ile Glu Arg Ile Ser Glu
 385 390 395 400
 Asp Ile Glu Lys Ile Cys Ile Tyr Asn Gly Glu Phe Glu Lys Ile Val
 405 410 415
 Val Asn Glu His Asp Ser Ser Arg Lys Leu Ser Lys Asn Ile Lys Ala
 420 425 430

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

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

 Val Asp Tyr Lys Leu Leu Pro Gly Ser Asn Lys Met Leu Pro Lys Val

565 570 575
 Phe Phe Ala Lys Ser Asn Ile Gly Tyr Tyr Asn Pro Ser Thr Glu Leu
 580 585 590
 Tyr Ser Asn Tyr Lys Lys Gly Thr His Lys Lys Gly Pro Ser Phe Ser
 595 600 605
 Ile Asp Asp Cys His Asn Leu Ile Asp Phe Phe Lys Glu Ser Ile Lys
 610 615 620

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

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

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

Gly Glu Arg Asn Leu Leu Tyr Val Val Val Ile Asn Ser Glu Gly Lys
 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 Asn Tyr His Ala Leu Leu Asp Glu Arg Glu Asp Asp Arg
 850 855 860
 Asn Lys Ala Arg Lys Asp Trp Asn Thr Ile Glu Asn Ile Lys Glu Leu
 865 870 875 880

 Lys Thr Gly Tyr Leu Ser 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 Glu Lys Leu Asn Tyr Leu Val Ile Asp Lys Ser Arg
 930 935 940

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

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

1055 1060 1065

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

1070 1075 1080

Phe Lys Gln Tyr Arg Ile Pro Tyr Glu Asn Gly Glu Asp Ile Lys

1085 1090 1095

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

1100 1105 1110

Phe Arg Leu Leu His 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 Asp Arg

1130 1135 1140

Gly Glu Phe Phe Cys Ser Glu Phe Ser Glu Gly Thr Met 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 Asp Glu Gly Glu Lys Val

1175 1180 1185

Asn Leu Ser Met Thr Asn Ala Glu Trp Leu Lys Tyr Ala Gln Leu

1190 1195 1200

His Leu Leu

1205

<210> 54

<211> 1264

<212> PRT

<213> Moraxella caprae

<220><221> MISC_FEATURE

<222> (1)..(1264)

<223> Genbank WP_036388671 Cpf1

<400> 54

Met Leu Phe Gln Asp Phe Thr His Leu Tyr Pro Leu Ser Lys Thr Met

1 5 10 15

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

20 25 30

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

35 40 45

Val Lys Ala Ile Leu Asp Asp Tyr His Arg Asp Phe Ile Ala Asp Met

50 55 60

Met Gly Glu Val Lys Leu Thr Lys Leu Ala Glu Phe Tyr Asp Val Tyr

65 70 75 80

Leu Lys Phe Arg Lys Asn Pro Lys Asp Asp Gly Leu Gln Lys 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 Thr Tyr Arg Leu Ile

180 185 190

His Glu Asn Leu Pro Arg Phe Ile Asp Asn Leu Gln Ile Leu Ala 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 Arg 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
Gly 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 Ala 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 Ile 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 Asn Val Tyr Gln Lys Met
 610 615 620
 Ile Tyr Lys Leu Leu Pro Gly Pro Asn Lys Met Leu Pro Lys Val Phe
 625 630 635 640
 Phe Ala Lys Ser Asn Leu Asp Tyr Tyr Asn Pro Ser Ala Glu Leu Leu
 645 650 655
 Asp Lys Tyr Ala Gln Gly Thr His Lys Lys Gly Asn Asn Phe Asn Leu
 660 665 670
 Lys Asp Cys His Ala Leu Ile Asp Phe Phe Lys Ala Gly Ile Asn Lys
 675 680 685
 His Pro Glu Trp Gln His Phe Gly Phe Lys Phe Ser Pro Thr Ser Ser
 690 695 700
 Tyr Gln Asp Leu Ser Asp Phe Tyr Arg Glu Val Glu Pro Gln Gly Tyr
 705 710 715 720
 Gln Val Lys Phe Val Asp Ile Asn Ala Asp Tyr Ile Asn Glu Leu Val
 725 730 735
 Glu Gln Gly Gln Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser
 740 745 750
 Pro Lys Ala His Gly Lys Pro Asn Leu His Thr Leu Tyr Phe Lys Ala

755 760 765
 Leu Phe Ser Lys Asp Asn Leu Ala Asn Pro Ile Tyr Lys Leu Asn Gly
 770 775 780
 Glu Ala Gln Ile Phe Tyr Arg Lys Ala Ser Leu Asp Met Asn Glu Thr

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

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

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

 980 985 990
 Glu Asn Ala Leu Ile Lys Lys Leu Asn His Leu Val Leu Lys Asp Glu
 995 1000 1005

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

1235 1240 1245
Lys Leu Ala Ile Asp Asn Gln Thr Trp Leu Asn Phe Ala Gln Asn
1250 1255 1260
Arg

<210> 55

<211> 1259

<212> PRT

<213> Synergistes jonesii

<220><221> MISC_FEATURE

<222> (1)..(1259)

<223> Genbank WP_037975888 Cpf1

<400> 55

Met Ala Asn Ser Leu Lys Asp Phe Thr Asn Ile Tyr Gln Leu Ser Lys

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

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

130 135 140
Asp Ala Val Lys Ser Phe Asp Lys Phe Ser Thr Tyr Phe Thr Gly Phe

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

 195 200 205
 Lys Ala Gln Lys Glu Leu Ser Gly Ile Leu Gly Gly Lys Thr Leu Asp
 210 215 220
 Asp Ile Phe Ser Ile Glu Ser Phe Asn Asn Val Leu Thr Gln Asp Lys
 225 230 235 240
 Ile Asp Tyr Tyr Asn Gln Ile Ile Gly Gly Val Ser Gly Lys Ala Gly
 245 250 255
 Asp Lys Lys Leu Arg Gly Val Asn Glu Phe Ser Asn Leu Tyr Arg Gln

 260 265 270
 Gln His Pro Glu Val Ala Ser Leu Arg Ile Lys Met Val Pro Leu Tyr
 275 280 285
 Lys Gln Ile Leu Ser Asp Arg Thr Thr Leu Ser Phe Val Pro Glu Ala
 290 295 300
 Leu Lys Asp Asp Glu Gln Ala Ile Asn Ala Val Asp Gly Leu Arg Ser
 305 310 315 320
 Glu Leu Glu Arg Asn Asp Ile Phe Asn Arg Ile Lys Arg Leu Phe Gly

 325 330 335
 Lys Asn Asn Leu Tyr Ser Leu Asp Lys Ile Trp Ile Lys Asn Ser Ser
 340 345 350
 Ile Ser Ala Phe Ser Asn Glu Leu Phe Lys Asn Trp Ser Phe Ile Glu
 355 360 365
 Asp Ala Leu Lys Glu Phe Lys Glu Asn Glu Phe Asn Gly Ala Arg Ser
 370 375 380
 Ala Gly Lys Lys Ala Glu Lys Trp Leu Lys Ser Lys Tyr Phe Ser Phe

 385 390 395 400

Ala Asp Ile Asp Ala Ala Val Lys Ser Tyr Ser Glu Gln Val Ser Ala
405 410 415

Asp Ile Ser Ser Ala Pro Ser Ala Ser Tyr Phe Ala Lys Phe Thr Asn
420 425 430

Leu Ile Glu Thr Ala Ala Glu Asn Gly Arg Lys Phe Ser Tyr Phe Ala
435 440 445

Ala Glu Ser Lys Ala Phe Arg Gly Asp Asp Gly Lys Thr Glu Ile Ile
450 455 460

Lys Ala Tyr Leu Asp Ser Leu Asn Asp Ile Leu His Cys Leu Lys Pro
465 470 475 480

Phe Glu Thr Glu Asp Ile Ser Asp Ile Asp Thr Glu Phe Tyr Ser Ala
485 490 495

Phe Ala Glu Ile Tyr Asp Ser Val Lys Asp Val Ile Pro Val Tyr Asn
500 505 510

Ala Val Arg Asn Tyr Thr Thr Gln Lys Pro Phe Ser Thr Glu Lys Phe
515 520 525

Lys Leu Asn Phe Glu Asn Pro Ala Leu Ala Lys Gly Trp Asp Lys Asn
530 535 540

Lys Glu Gln Asn Asn Thr Ala Ile Ile Leu Met Lys Asp Gly Lys Tyr
545 550 555 560

Tyr Leu Gly Val Ile Asp Lys Asn Asn Lys Leu Arg Ala Asp Asp Leu
565 570 575

Ala Asp Asp Gly Ser Ala Tyr Gly Tyr Met Lys Met Asn Tyr Lys Phe
580 585 590

Ile Pro Thr Pro His Met Glu Leu Pro Lys Val Phe Leu Pro Lys Arg
595 600 605

Ala Pro Lys Arg Tyr Asn Pro Ser Arg Glu Ile Leu Leu Ile Lys Glu
610 615 620

Asn Lys Thr Phe Ile Lys Asp Lys Asn Phe Asn Arg Thr Asp Cys His
625 630 635 640

Lys Leu Ile Asp Phe Phe Lys Asp Ser Ile Asn Lys His Lys Asp Trp

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

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

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

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

Asp Phe Asn Ile Ile Gly Gly Met Asp Tyr His Ala Lys Leu Asn Gln

900 905 910
Lys Glu Lys Glu Arg Asp Thr Ala Arg Lys Ser Trp Lys Thr Ile Gly

915 920 925
Thr Ile Lys Glu Leu Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu

930 935 940
Ile Val Arg Leu Ala Val Asp Asn Asn Ala Val Ile Val Met Glu Asp

945 950 955 960
Leu Asn Ile Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val

965 970 975
Tyr Gln Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Leu Val

980 985 990
Phe Lys Asp Ala Gly Tyr Asp Ala Pro Cys Gly Ile Leu Lys Gly Leu

995 1000 1005
Gln Leu Thr Glu Lys Phe Glu Ser Phe Thr Lys Leu Gly Lys Gln

1010 1015 1020
Cys Gly Ile Ile Phe Tyr Ile Pro Ala Gly Tyr Thr Ser Lys Ile

1025 1030 1035
Asp Pro Thr Thr Gly Phe Val Asn Leu Phe Asn Ile Asn Asp Val

1040 1045 1050
Ser Ser Lys Glu Lys Gln Lys Asp Phe Ile Gly Lys Leu Asp Ser

1055 1060 1065
Ile Arg Phe Asp Ala Lys Arg Asp Met Phe Thr Phe Glu Phe Asp

1070 1075 1080
Tyr Asp Lys Phe Arg Thr Tyr Gln Thr Ser Tyr Arg Lys Lys Trp

1085 1090 1095
Ala Val Trp Thr Asn Gly Lys Arg Ile Val Arg Glu Lys Asp Lys

1100 1105 1110
Asp Gly Lys Phe Arg Met Asn Asp Arg Leu Leu Thr Glu Asp Met

1115 1120 1125
Lys Asn Ile Leu Asn Lys Tyr Ala Leu Ala Tyr Lys Ala Gly Glu

1130 1135 1140
Asp Ile Leu Pro Asp Val Ile Ser Arg Asp Lys Ser Leu Ala Ser

1145 1150 1155
Glu Ile Phe Tyr Val Phe Lys Asn Thr Leu Gln Met Arg Asn Ser

1160 1165 1170
Lys Arg Asp Thr Gly Glu Asp Phe Ile Ile Ser Pro Val Leu Asn

1175 1180 1185
Ala Lys Gly Arg Phe Phe Asp Ser Arg Lys Thr Asp Ala Ala Leu

1190 1195 1200
Pro Ile Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Ala Leu Lys

1205 1210 1215
Gly Ser Leu Val Leu Asp Ala Ile Asp Glu Lys Leu Lys Glu Asp

1220 1225 1230
Gly Arg Ile Asp Tyr Lys Asp Met Ala Val Ser Asn Pro Lys Trp

1235 1240 1245
Phe Glu Phe Met Gln Thr Arg Lys Phe Asp Phe

1250 1255
<210> 56

<211> 1264

<212> PRT

<213> Prevotella brevis

<220><221> MISC_FEATURE

<222> (1)..(1264)

<223> Genbank WP_044110123 Cpf1

<400> 56

Met Lys Gln Phe Thr Asn Leu Tyr Gln Leu Ser Lys Thr Leu Arg Phe

1 5 10 15

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

20 25 30

Phe Ile Asp Asn Asp Ala His Arg Ala Glu Ser Tyr Lys Lys Val Lys

35 40 45

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

50

55

60

Asn Phe Lys Leu Asn Gly Glu Tyr Leu Gln Ala Tyr Phe Asp Leu Tyr

65

70

75

80

Ser Gln Asp Thr Lys Asp Lys Gln Phe Lys Asp Ile Gln Asp Lys Leu

85

90

95

Arg Lys Ser Ile Ala Ser Ala Leu Lys Gly Asp Asp Arg Tyr Lys Thr

100

105

110

Ile Asp Lys Lys Glu Leu Ile Arg Gln Asp Met Lys Thr Phe Leu Lys

115

120

125

Lys Asp Thr Asp Lys Ala Leu Leu Asp Glu Phe Tyr Glu Phe Thr Thr

130

135

140

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

145

150

155

160

Ala Lys Ser Thr Ala Ile Ala Tyr Arg Leu Ile His Asp Asn Leu Pro

165

170

175

Lys Phe Ile Asp Asn Ile Ala Val Phe Lys Lys Ile Ala Asn Thr Ser

180

185

190

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

195

200

205

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

210

215

220

Val Leu Thr Gln Thr Gln Ile Glu Val Tyr Asn Ser Ile Ile Gly Gly

225

230

235

240

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

245

250

255

Asn Leu Tyr Asn Gln Gln Leu Ala Asn Lys Lys Asp Arg Leu Pro Lys

260

265

270

Leu Lys Pro Leu Phe Lys Gln Ile Leu Ser Asp Arg Val Gln Leu Ser

275

280

285

Trp Leu Gln Glu Glu Phe Asn Thr Gly Ala Asp Val Leu Asn Ala Val

290 295 300
 Lys Glu Tyr Cys Thr Ser Tyr Phe Asp Asn Val Glu Glu Ser Val Lys

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

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

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

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

Glu Asn Lys Glu Ser Thr Asn Ala Ser Ile Ile Leu Arg Lys Asn Gly
 545 550 555 560
 Met Tyr Tyr Leu Gly Ile Met Lys Lys Glu Tyr Arg Asn Ile Leu Thr

 565 570 575
 Lys Pro Met Pro Ser Asp Gly Asp Cys Tyr Asp Lys Val Val Tyr Lys
 580 585 590
 Phe Phe Lys Asp Ile Thr Thr Met Val Pro Lys Cys Thr Thr Gln Met
 595 600 605
 Lys Ser Val Lys Glu His Phe Ser Asn Ser Asn Asp Asp Tyr Thr Leu
 610 615 620
 Phe Glu Lys Asp Lys Phe Ile Ala Pro Val Val Ile Thr Lys Glu Ile

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

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

 755 760 765
 Phe Asp Glu Ser Asn Leu Asn Asp Val Val Tyr Lys Leu Asn Gly Gln
 770 775 780
 Ala Glu Val Phe Tyr Arg Lys Lys Ser Ile Thr Tyr Gln His Pro Thr

785 790 795 800
 His Pro Ala Asn Lys Pro Ile Asp Asn Lys Asn Val Asn Asn Pro Lys
 805 810 815
 Lys Gln Ser Asn Phe Glu Tyr Asp Leu Ile Lys Asp Lys Arg Tyr Thr

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

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

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

 1010 1015 1020
 Leu Pro Asn Lys Asn Ser Lys Gln Ser Gly Phe Leu Phe Tyr Val
 1025 1030 1035

Pro Ala Trp Asn Thr Ser Lys Ile Asp Pro Val Thr Gly Phe Val		
1040	1045	1050
Asn Leu Leu Asp Thr Arg Tyr Glu Asn Val Ala Lys Ala Gln Ala		
1055	1060	1065
Phe Phe Lys Lys Phe Asp Ser Ile Arg Tyr Asn Lys Glu Tyr Asp		
1070	1075	1080
Arg Phe Glu Phe Lys Phe Asp Tyr Ser Asn Phe Thr Ala Lys Ala		
1085	1090	1095
Glu Asp Thr Arg Thr Gln Trp Thr Leu Cys Thr Tyr Gly Thr Arg		
1100	1105	1110
Ile Glu Thr Phe Arg Asn Ala Glu Lys Asn Ser Asn Trp Asp Ser		
1115	1120	1125
Arg Glu Ile Asp Leu Thr Thr Glu Trp Lys Thr Leu Phe Thr Gln		
1130	1135	1140
His Asn Ile Pro Leu Asn Ala Asn Leu Lys Glu Ala Ile Leu Leu		
1145	1150	1155
Gln Ala Asn Lys Asn Phe Tyr Thr Asp Ile Leu His Leu Met Lys		
1160	1165	1170
Leu Thr Leu Gln Met Arg Asn Ser Val Thr Gly Thr Asp Ile Asp		
1175	1180	1185
Tyr Met Val Ser Pro Val Ala Asn Glu Cys Gly Glu Phe Phe Asp		
1190	1195	1200
Ser Arg Lys Val Lys Glu Gly Leu Pro Val Asn Ala Asp Ala Asn		
1205	1210	1215
Gly Ala Tyr Asn Ile Ala Arg Lys Gly Leu Trp Leu Ala Gln Gln		
1220	1225	1230
Ile Lys Asn Ala Asn Asp Leu Ser Asp Val Lys Leu Ala Ile Thr		
1235	1240	1245
Asn Lys Glu Trp Leu Gln Phe Ala Gln Lys Lys Gln Tyr Leu Lys		
1250	1255	1260
Asp		

<210> 57

<211> 1273

<212> PRT

<213> Flavobacterium sp.

<220><221> MISC_FEATURE

<222> (1)..(1273)

<223> Genbank WP_045971446 Cpf1

<400> 57

Met Lys Asn Phe Ser Asn Leu Tyr Gln Val Ser Lys Thr Val Arg Phe

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

Asn Lys Lys Leu Ser Tyr Leu Asn Glu Tyr Ile Ala Leu Tyr Thr Gln

65 70 75 80

Ser Ala Glu Ala Lys Lys Glu Asp Lys Phe Lys Ala Asp Phe Lys Lys

85 90 95

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

100 105 110

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

115 120 125

Ile Glu Leu Glu Lys Trp Lys Asn Glu Asn Asn Leu Ala Val Tyr Leu

130 135 140

Asp Glu Ser Phe Lys Ser Phe Thr Thr Tyr Phe Thr Gly Phe His Gln

145 150 155 160

Asn Arg Lys Asn Met Tyr Ser Ala Glu Ala Asn Ser Thr Ala Ile Ala

165 170 175

Tyr Arg Leu Ile His Glu Asn Leu Pro Lys Phe Ile Glu Asn Ser Lys

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

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

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

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

 420 425 430

Tyr Val Lys Thr Leu Asp Glu Asn Leu Asp Trp Lys Gln Lys Tyr Thr
 435 440 445
 Pro Ser Cys Ile Ala Asp Tyr Phe Thr Thr His Phe Ile Ala Lys Lys
 450 455 460
 Glu Asn Glu Ala Asp Lys Thr Phe Asn Phe Ile Ala Asn Ile Lys Ala
 465 470 475 480
 Lys Tyr Gln Cys Ile Gln Gly Ile Leu Glu Gln Ala Asp Asp Tyr Glu
 485 490 495
 Asp Glu Leu Lys Gln Asp Gln Lys Leu Ile Asp Asn Ile Lys Phe Phe
 500 505 510
 Leu Asp Ala Ile Leu Glu Val Val His Phe Ile Lys Pro Leu His Leu
 515 520 525
 Lys Ser Glu Ser Ile Thr Glu Lys Asp Asn Ala Phe Tyr Asp Val Phe
 530 535 540
 Glu Asn Tyr Tyr Glu Ala Leu Asn Val Val Thr Pro Leu Tyr Asn Met
 545 550 555 560
 Val Arg Asn Tyr Val Thr Gln Lys Pro Tyr Ser Thr Glu Lys Ile Lys
 565 570 575
 Leu Asn Phe Glu Asn Ala Gln Leu Leu Asn Gly Trp Asp Ala Asn Lys
 580 585 590
 Glu Lys Asp Tyr Leu Thr Thr Ile Leu Lys Arg Asp Gly Asn Tyr Phe
 595 600 605
 Leu Ala Ile Met Asp Lys Lys His Asn Lys Thr Phe Gln Gln Phe Thr
 610 615 620
 Glu Asp Asp Glu Asn Tyr Glu Lys Ile Val Tyr Lys Leu Leu Pro Gly
 625 630 635 640
 Val Asn Lys Met Leu Pro Lys Val Phe Phe Ser Asn Lys Asn Ile Ala
 645 650 655
 Phe Phe Asn Pro Ser Lys Glu Ile Leu Asp Asn Tyr Lys Asn Asn Thr
 660 665 670
 His Lys Lys Gly Ala Thr Phe Asn Leu Lys Asp Cys His Ala Leu Ile

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

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

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

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

Glu Ile Ala Arg Asp Lys Ala Arg Lys Asn Trp Gly Ser Ile Glu Ser

930

935

940

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

945

950

955

960

Thr Lys Met Met Ile Glu His Asn Ala Ile Val Val Met Glu Asp Leu

965

970

975

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

980

985

990

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

995

1000

1005

Lys Asp Lys Gln Pro His Glu Leu Gly Gly Leu Tyr Asn Ala Leu

1010

1015

1020

Gln Leu Thr Asn Lys Phe Glu Ser Phe Gln Lys Met Gly Lys Gln

1025

1030

1035

Ser Gly Phe Leu Phe Tyr Val Pro Ala Trp Asn Thr Ser Lys Ile

1040

1045

1050

Asp Pro Thr Thr Gly Phe Val Asn Tyr Phe Tyr Thr Lys Tyr Glu

1055

1060

1065

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

1070

1075

1080

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

1085

1090

1095

Tyr Ser Asp Phe Asn Pro Lys Ala Ala Asp Thr Arg Gln Glu Trp

1100

1105

1110

Thr Ile Cys Thr His Gly Glu Arg Ile Glu Thr Lys Arg Gln Lys

1115

1120

1125

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

1130

1135

1140

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

1145

1150

1155

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

1160 1165 1170
Glu Glu Leu Phe His Leu Leu Lys Leu Thr Leu Gln Met Arg Asn

1175 1180 1185
Ser Glu Ser His Thr Glu Ile Asp Tyr Leu Ile Ser Pro Val Ala

1190 1195 1200
Asn Glu Lys Gly Ile Phe Tyr Asp Ser Arg Lys Ala Thr Ala Ser

1205 1210 1215
Leu Pro Ile Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Ala Lys

1220 1225 1230
Lys Gly Leu Trp Ile Met Glu Gln Ile Asn Lys Thr Asn Ser Glu

1235 1240 1245
Asp Asp Leu Lys Lys Val Lys Leu Ala Ile Ser Asn Arg Glu Trp

1250 1255 1260

Leu Gln Tyr Val Gln Gln Val Gln Lys Lys

1265 1270

<210> 58

<211> 973

<212> PRT

<213> Sneathia amnii

<220><221> MISC_FEATURE

<222> (1)..(973)

<223> Genbank WP_046328599 Cpf1

<400> 58

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

1 5 10 15

Pro Glu Ser Val Ile Lys Thr Ile Asp Asn Phe Ile Asp Ser Ile Ile

20 25 30

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

35 40 45

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

50 55 60

Ser Asn Arg Lys Phe Gly Asp Tyr Asn Ser Ile Asn Met Met Ile Lys

65				70				75				80			
Gln	Ser	Met	Asn	Glu	Lys	Gly	Ile	Leu	Ser	Lys	Lys	Glu	Ile	Asn	Glu
				85				90				95			
Leu	Ile	Pro	Asp	Leu	Glu	Asn	Ile	Asn	Lys	Pro	Lys	Val	Lys	Ser	Phe
				100				105				110			
Asn	Leu	Ser	Phe	Ile	Phe	Glu	Asn	Leu	Thr	Lys	Glu	His	Lys	Glu	Leu
				115				120				125			
Ile	Ile	Asp	Tyr	Ile	Arg	Glu	Asn	Ile	Cys	Asn	Val	Ile	Glu	Asn	Val
130				135				140							
Lys	Ile	Thr	Ile	Glu	Lys	Tyr	Arg	Asn	Ile	Asp	Asn	Lys	Ile	Glu	Phe
145				150				155				160			
Lys	Asn	Asn	Ala	Glu	Lys	Val	Ser	Lys	Ile	Lys	Glu	Met	Leu	Glu	Ser
				165				170				175			
Ile	Asn	Glu	Leu	Cys	Lys	Leu	Ile	Lys	Glu	Phe	Asn	Thr	Asp	Glu	Ile
				180				185				190			
Glu	Lys	Asn	Asn	Glu	Phe	Tyr	Asn	Ile	Leu	Asn	Lys	Asn	Phe	Glu	Ile
195				200				205							
Phe	Glu	Ser	Ser	Tyr	Lys	Val	Leu	Asn	Lys	Val	Arg	Asn	Phe	Val	Thr
210				215				220							
Lys	Lys	Glu	Val	Ile	Glu	Asn	Lys	Met	Lys	Leu	Asn	Phe	Ser	Asn	Tyr
225				230				235				240			
Gln	Leu	Gly	Asn	Gly	Trp	His	Lys	Asn	Lys	Glu	Lys	Asp	Cys	Ser	Ile
				245				250				255			
Ile	Leu	Phe	Arg	Lys	Arg	Asn	Asn	Glu	Arg	Trp	Ile	Tyr	Tyr	Leu	Gly
260				265				270							
Ile	Leu	Lys	His	Gly	Thr	Lys	Ile	Lys	Glu	Asn	Asp	Tyr	Leu	Ser	Ser
275				280				285							
Val	Asp	Thr	Gly	Phe	Tyr	Lys	Met	Asp	Tyr	Tyr	Ala	Gln	Asn	Ser	Leu
290				295				300							
Ser	Lys	Met	Ile	Pro	Lys	Cys	Ser	Ile	Thr	Val	Lys	Asn	Val	Lys	Asn
305				310				315				320			

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

Glu Pro Leu Glu Ile Thr Pro Glu Ile Arg Lys Leu Tyr Gly Asn Asn
340 345 350

Glu His Ile Lys Gly Asp Lys Phe Lys Lys Glu Ser Leu Val Lys Trp
355 360 365

Ile Asp Phe Cys Lys Glu Phe Leu Leu Lys Tyr Lys Ser Phe Glu Lys
370 375 380

Ala Lys Lys Glu Ile Leu Lys Leu Lys Glu Ser Asn Leu Tyr Glu Asn
385 390 395 400

Leu Glu Glu Phe Tyr Ser Asp Ala Glu Glu Lys Ala Tyr Phe Leu Glu
405 410 415

Phe Ile Asn Ile Asp Glu Asp Lys Ile Lys Lys Leu Val Lys Glu Lys
420 425 430

Asn Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser
435 440 445

Thr Gly Asn Lys Asn Leu His Thr Met Tyr Phe Glu Glu Leu Phe Thr
450 455 460

Asp Glu Asn Leu Lys Lys Pro Val Phe Lys Leu Asn Gly Asn Thr Glu
465 470 475 480

Val Phe Tyr Arg Ile Ala Ser Ser Lys Pro Lys Ile Val His Asn Lys
485 490 495

Gly Glu Lys Leu Val Asn Lys Thr Tyr Leu Asp Asp Gly Ile Ile Lys
500 505 510

Thr Ile Pro Asp Ser Val Tyr Glu Glu Ile Ser Glu Lys Val Lys Asn
515 520 525

Asn Glu Asp Tyr Ser Lys Leu Leu Glu Glu Asn Asn Ile Lys Asn Leu
530 535 540

Glu Ile Lys Val Ala Thr His Glu Ile Val Lys Asp Lys Arg Tyr Phe
545 550 555 560

Glu Asn Lys Phe Leu Phe Tyr Leu Pro Ile Thr Leu Asn Lys Lys Val

565 570 575
 Ser Asn Lys Asn Thr Asn Lys Asn Ile Asn Lys Asn Val Ile Asp Glu
 580 585 590

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

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

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

 Lys Lys Ser Asn Asn Asn Glu Glu Ser Ile Val Lys Ala Ile Asp Lys
 785 790 795 800
 Ile Ser Tyr Asp Lys Glu Lys Asp Met Phe Tyr Phe Asp Ile Asn Leu
 805 810 815

Ser Asn Ser Ser Phe Asn Leu Thr Val Lys Asn Val Leu Lys Lys Glu
820 825 830

Trp Arg Ile Tyr Thr Asn Gly Glu Arg Ile Ile Tyr Lys Asp Arg Lys
835 840 845

Tyr Ile Thr Leu Asn Ile Thr Gln Glu Met Lys Asp Ile Leu Ser Lys
850 855 860

Cys Gly Ile Asp Tyr Leu Asn Ile Asp Asn Leu Lys Gln Asp Ile Leu
865 870 875 880

Lys Asn Lys Leu His Lys Lys Val Tyr Tyr Ile Phe Glu Leu Ala Asn
885 890 895

Lys Met Arg Asn Glu Asn Lys Asp Val Asp Tyr Ile Ile Ser Pro Val
900 905 910

Leu Asn Lys Asp Gly Lys Phe Phe Met Thr Gln Glu Ile Asn Glu Leu
915 920 925

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

Gly Lys Leu Met Ile Asp Asn Leu Asn Lys Lys Glu Lys Phe Val Phe
945 950 955 960

Leu Ser Asn Glu Asp Trp Leu Asn Phe Ile Gln Gly Arg
965 970

<210> 59

<211> 1238

<212> PRT

<213> Candidatus Methanoplasma

<220><221> MISC_FEATURE

<222> (1)..(1238)

<223> Genbank WP_048112740 Cpf1

<400> 59

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 Glu 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 Lys 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 Pro Arg 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> 60		
<211> 1222		

<212> PRT

<213> Oribacterium sp.

<220><221

> MISC_FEATURE

<222> (1)..(1222)

<223> Genbank WP_049895985 Cpf1

<400> 60

Met Glu Thr Glu Ile Leu Lys Tyr Asp Phe Phe Glu Arg Glu Gly Lys

1 5 10 15

Tyr Met Tyr Tyr Asp Gly Leu Thr Lys Gln Tyr Ala Leu Ser Lys Thr

20 25 30

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

35 40 45

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

50 55 60

His Val Lys Lys Leu Met Asp Met Tyr His Lys Lys Ile Ile Asn Glu

65 70 75 80

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

85 90 95

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

100 105 110

Ile Gln Asp Lys Leu Arg Lys Glu Ile Val Glu Gln Leu Lys Gly His

115 120 125

Thr Asp Tyr Ser Lys Val Gly Asn Lys Asp Phe Leu Gly Leu Leu Lys

130 135 140

Ala Ala Ser Thr Glu Glu Asp Arg Ile Leu Ile Glu Ser Phe Asp Asn

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205
 Asn Ala Gly Val Ile Ser Gly Asp Met Ser Ile Val Glu Gln Asp Glu
 210 215 220
 Leu Phe Glu Val Asp Thr Phe Asn His Thr Leu Thr Gln Tyr Gly Ile
 225 230 235 240
 Asp Thr Tyr Asn His Met Ile Gly Gln Leu Asn Ser Ala Ile Asn Leu

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

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

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

 435 440 445

Val Gln Ala Ile Lys Asn Met Leu Asp Ser Ile Lys Asp Phe Glu His
450 455 460

Asp Val Met Leu Ile Asn Gly Ser Gly Gln Glu Ile Lys Lys Asn Leu
465 470 475 480

Asn Val Tyr Ser Glu Gln Glu Ala Leu Ala Gly Ile Leu Arg Gln Val
485 490 495

Asp His Ile Tyr Asn Leu Thr Arg Asn Tyr Leu Thr Lys Lys Pro Phe
500 505 510

Ser Thr Glu Lys Ile Lys Leu Asn Phe Asn Arg Pro Thr Phe Leu Asp
515 520 525

Gly Trp Asp Lys Asn Lys Glu Glu Ala Asn Leu Gly Ile Leu Leu Ile
530 535 540

Lys Asp Asn Arg Tyr Tyr Leu Gly Ile Met Asn Thr Ser Ser Asn Lys
545 550 555 560

Ala Phe Val Asn Pro Pro Lys Ala Ile Ser Asn Asp Ile Tyr Lys Lys
565 570 575

Val Asp Tyr Lys Leu Leu Pro Gly Pro Asn Lys Met Leu Pro Lys Val
580 585 590

Phe Phe Ala Thr Lys Asn Ile Ala Tyr Tyr Ala Pro Ser Glu Glu Leu
595 600 605

Leu Ser Lys Tyr Arg Lys Gly Thr His Lys Lys Gly Asp Ser Phe Ser
610 615 620

Ile Asp Asp Cys Arg Asn Leu Ile Asp Phe Phe Lys Ser Ser Ile Asn
625 630 635 640

Lys Asn Thr Asp Trp Ser Thr Phe Gly Phe Asn Phe Ser Asp Thr Asn
645 650 655

Ser Tyr Asn Asp Ile Ser Asp Phe Tyr Arg Glu Val Glu Lys Gln Gly
660 665 670

Tyr Lys Leu Ser Phe Thr Asp Ile Asp Ala Cys Tyr Ile Lys Asp Leu
675 680 685

Val Asp Asn Asn Glu Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe

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

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

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

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

Lys Met Leu Ile Asp Lys Met Asn Tyr Leu Val Leu Asp Lys Ser Arg

945 950 955 960

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

965 970 975

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

980 985 990

Ile Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile Asp Pro Thr

995 1000 1005

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

1010 1015 1020

Lys Ala Arg Asp Phe Phe Ser Lys Phe Asp Phe Ile Arg Tyr Asn

1025 1030 1035

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

1040 1045 1050

Thr Glu Arg Ala Ser Gly Cys Lys Ser Lys Trp Ile Ala Cys Thr

1055 1060 1065

Asn Gly Glu Arg Ile Val Lys Tyr Arg Asn Ser Asp Lys Asn Asn

1070 1075 1080

Ser Phe Asp Asp Lys Thr Val Ile Leu Thr Asp Glu Tyr Arg Ser

1085 1090 1095

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

1100 1105 1110

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

1115 1120 1125

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

1130 1135 1140

Ser Asp Gly Lys Arg Asp Tyr Ile Ile Ser Pro Val Lys Asn Tyr

1145 1150 1155

Arg Glu Glu Phe Phe Cys Ser Glu Phe Ser Asp Asp Thr Phe Pro

1160 1165 1170

Arg Asp Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Arg Lys Gly

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

1190 1195 1200
Ile Asn Leu Ala Met Ser Asn Ser Glu Trp Leu Glu Tyr Ala Gln

1205 1210 1215
Cys Asn Leu Leu

1220

<210> 61

<211> 1326

<212> PRT

<213> Prevotella disiens

<220><221> MISC_FEATURE

<222> (1)..(1326)

<223> Genbank WP_050786240 Cpf1

<400> 61

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

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

Lys Lys His Lys Ile Phe Ile Glu Glu Thr Leu Ser Ser Phe Ser Ile

65 70 75 80

Ser Asn Asp Leu Leu Lys Gln Tyr Phe Asp Cys Tyr Asn Glu Leu Lys

85 90 95

Ala Phe Lys Lys Asp Cys Lys Ser Asp Glu Glu Glu Val Lys Lys Thr

100 105 110

Ala Leu Arg Asn Lys Cys Thr Ser Ile Gln Arg Ala Met Arg Glu Ala

115 120 125

Ile Ser Gln Ala Phe Leu Lys Ser Pro Gln Lys Lys Leu Leu Ala Ile

130 135 140
 Lys Asn Leu Ile Glu Asn Val Phe Lys Ala Asp Glu Asn Val Gln His
 145 150 155 160
 Phe Ser Glu Phe Thr Ser Tyr Phe Ser Gly Phe Glu Thr Asn Arg Glu
 165 170 175
 Asn Phe Tyr Ser Asp Glu Glu Lys Ser Thr Ser Ile Ala Tyr Arg Leu
 180 185 190
 Val His Asp Asn Leu Pro Ile Phe Ile Lys Asn Ile Tyr Ile Phe Glu
 195 200 205

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

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

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

Ala Glu Leu Gln Thr Phe Asn Asn Tyr Glu Leu Ile Ala Asn Ala Leu
385 390 395 400

Arg Ala Lys Ile Lys Lys Glu Thr Lys Gln Gly Arg Lys Ser Phe Glu
405 410 415

Lys Tyr Glu Glu Tyr Ile Asp Lys Lys Val Lys Ala Ile Asp Ser Leu
420 425 430

Ser Ile Gln Glu Ile Asn Glu Leu Val Glu Asn Tyr Val Ser Glu Phe
435 440 445

Asn Ser Asn Ser Gly Asn Met Pro Arg Lys Val Glu Asp Tyr Phe Ser
450 455 460

Leu Met Arg Lys Gly Asp Phe Gly Ser Asn Asp Leu Ile Glu Asn Ile
465 470 475 480

Lys Thr Lys Leu Ser Ala Ala Glu Lys Leu Leu Gly Thr Lys Tyr Gln
485 490 495

Glu Thr Ala Lys Asp Ile Phe Lys Lys Asp Glu Asn Ser Lys Leu Ile
500 505 510

Lys Glu Leu Leu Asp Ala Thr Lys Gln Phe Gln His Phe Ile Lys Pro
515 520 525

Leu Leu Gly Thr Gly Glu Glu Ala Asp Arg Asp Leu Val Phe Tyr Gly
530 535 540

Asp Phe Leu Pro Leu Tyr Glu Lys Phe Glu Glu Leu Thr Leu Leu Tyr
545 550 555 560

Asn Lys Val Arg Asn Arg Leu Thr Gln Lys Pro Tyr Ser Lys Asp Lys
565 570 575

Ile Arg Leu Cys Phe Asn Lys Pro Lys Leu Met Thr Gly Trp Val Asp
580 585 590

Ser Lys Thr Glu Lys Ser Asp Asn Gly Thr Gln Tyr Gly Gly Tyr Leu
595 600 605

Phe Arg Lys Lys Asn Glu Ile Gly Glu Tyr Asp Tyr Phe Leu Gly Ile
610 615 620

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

625 630 635 640
 Tyr Glu Arg Leu Asp Tyr Tyr Gln Pro Lys Ala Asn Thr Ile Tyr Gly
 645 650 655

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

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

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

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

Arg Asn Val Ala Asn Lys Asp Lys Val Ser Leu Phe Thr Tyr Asp Ile
885 890 895

Tyr Lys Asn Arg Arg Tyr Met Glu Asn Lys Phe Leu Phe His Leu Ser
900 905 910

Ile Val Gln Asn Tyr Lys Ala Ala Asn Asp Ser Ala Gln Leu Asn Ser
915 920 925

Ser Ala Thr Glu Tyr Ile Arg Lys Ala Asp Asp Leu His Ile Ile Gly
930 935 940

Ile Asp Arg Gly Glu Arg Asn Leu Leu Tyr Tyr Ser Val Ile Asp Met
945 950 955 960

Lys Gly Asn Ile Val Glu Gln Asp Ser Leu Asn Ile Ile Arg Asn Asn
965 970 975

Asp Leu Glu Thr Asp Tyr His Asp Leu Leu Asp Lys Arg Glu Lys Glu
980 985 990

Arg Lys Ala Asn Arg Gln Asn Trp Glu Ala Val Glu Gly Ile Lys Asp
995 1000 1005

Leu Lys Lys Gly Tyr Leu Ser Gln Ala Val His Gln Ile Ala Gln
1010 1015 1020

Leu Met Leu Lys Tyr Asn Ala Ile Ile Ala Leu Glu Asp Leu Gly
1025 1030 1035

Gln Met Phe Val Thr Arg Gly Gln Lys Ile Glu Lys Ala Val Tyr
1040 1045 1050

Gln Gln Phe Glu Lys Ser Leu Val Asp Lys Leu Ser Tyr Leu Val
1055 1060 1065

Asp Lys Lys Arg Pro Tyr Asn Glu Leu Gly Gly Ile Leu Lys Ala
1070 1075 1080

Tyr Gln Leu Ala Ser Ser Ile Thr Lys Asn Asn Ser Asp Lys Gln
1085 1090 1095

Asn Gly Phe Leu Phe Tyr Val Pro Ala Trp Asn Thr Ser Lys Ile
1100 1105 1110

Asp Pro Val Thr Gly Phe Thr Asp Leu Leu Arg Pro Lys Ala Met

1115	1120	1125
Thr Ile Lys Glu Ala Gln Asp	Phe Phe Gly Ala Phe	Asp Asn Ile
1130	1135	1140
Ser Tyr Asn Asp Lys Gly Tyr	Phe Glu Phe Glu Thr	Asn Tyr Asp
1145	1150	1155
Lys Phe Lys Ile Arg Met Lys	Ser Ala Gln Thr Arg	Trp Thr Ile
1160	1165	1170
Cys Thr Phe Gly Asn Arg Ile	Lys Arg Lys Lys Asp	Lys Asn Tyr
1175	1180	1185
Trp Asn Tyr Glu Glu Val Glu	Leu Thr Glu Glu Phe	Lys Lys Leu
1190	1195	1200
Phe Lys Asp Ser Asn Ile Asp	Tyr Glu Asn Cys Asn	Leu Lys Glu
1205	1210	1215
Glu Ile Gln Asn Lys Asp Asn	Arg Lys Phe Phe Asp	Asp Leu Ile
1220	1225	1230
Lys Leu Leu Gln Leu Thr Leu	Gln Met Arg Asn Ser	Asp Asp Lys
1235	1240	1245
Gly Asn Asp Tyr Ile Ile Ser	Pro Val Ala Asn Ala	Glu Gly Gln
1250	1255	1260
Phe Phe Asp Ser Arg Asn Gly	Asp Lys Lys Leu Pro	Leu Asp Ala
1265	1270	1275
Asp Ala Asn Gly Ala Tyr Asn	Ile Ala Arg Lys Gly	Leu Trp Asn
1280	1285	1290
Ile Arg Gln Ile Lys Gln Thr	Lys Asn Asp Lys Lys	Leu Asn Leu
1295	1300	1305
Ser Ile Ser Ser Thr Glu Trp	Leu Asp Phe Val Arg	Glu Lys Pro
1310	1315	1320
Tyr Leu Lys		
1325		
<210> 62		
<211> 998		
<212> PRT		

<213> Arcobacter butzleri

<220><221> MISC_FEATURE

<222> (1)..(998)

<223> Genbank WP_052943011 Cpf1

<400> 62

Met Gly Leu Leu Glu His Leu Glu Gly Ala Ile Val Glu Asp Met Phe

1 5 10 15

Ser Leu Asp Tyr Phe Ser Leu Thr Leu Ser Gln Arg Tyr Ile Asp Ile

20 25 30

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

35 40 45

Gln Gly Ile Asn Glu Asn Ile Asn Ile Tyr Arg Gln Lys Asn Asn Ile

50 55 60

Asp Arg Lys Asn Leu Pro Thr Leu Lys Pro Leu His Lys Gln Leu Leu

65 70 75 80

Ser Asp Arg Glu Thr Leu Ser Trp Ile Pro Glu Ala Phe Lys Thr Lys

85 90 95

Glu Glu Val Val Gly Ala Ile Glu Asp Phe Tyr Lys Asn Asn Ile Ile

100 105 110

Ser Phe Lys Cys Cys Asp Asn Ile Val Asp Ile Thr Lys Gln Phe Ile

115 120 125

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

130 135 140

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

145 150 155 160

Tyr Arg Ile Ile Lys Glu Ala Leu Trp Gln Lys His Ile Asn Glu Asn

165 170 175

Pro Lys Ala Ala Lys Ser Lys Asp Leu Thr Gly Asp Lys Glu Lys Tyr

180 185 190

Phe Ser Arg Lys Asn Ser Phe Phe Ser Phe Glu Glu Ile Ile Ser Ser

195 200 205

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

450 455 460
 Gln Gln Gly Tyr Lys Ile Thr Phe Lys Asn Ile Asp Glu Glu Phe Ile
 465 470 475 480
 Asn Thr Leu Ile Asn Glu Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn
 485 490 495
 Lys Asp Phe Ser Thr Phe Ser Lys Gly Thr Lys Asn Leu His Thr Leu

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

 565 570 575
 Lys Phe Gln Phe His Val Pro Ile Thr Met Asn Phe Lys Ala Thr Gly
 580 585 590
 Arg Ser Tyr Ile Asn Glu Glu Val Asn Asp Phe Leu Arg Gln Asn Ser
 595 600 605
 Lys Asp Val Lys Ile Ile Gly Ile Asn Arg Gly Glu Arg His Leu Ile
 610 615 620
 Tyr Leu Thr Met Ile Asn Ala Lys Gly Glu Ile Ile Gln Gln Tyr Ser

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

 690 695 700

Ala Ile Val Val Leu Glu Asp Leu Asn Phe Glu Phe Lys Arg Glu Arg
705 710 715 720
Leu Lys Val Glu Lys Ser Ile Tyr Gln Lys Phe Glu Lys Met Leu Ile
725 730 735
Asp Lys Leu Asn Tyr Leu Val Asp Lys Lys Lys Asp Ile Asn Glu Asn
740 745 750
Gly Gly Leu Leu Lys Ala Leu Gln Leu Thr Asn Lys Phe Glu Ser Phe
755 760 765
Glu Lys Ile Gly Lys Gln Asn Gly Phe Leu Phe Phe Val Asn Ala Trp
770 775 780
Asn Ile Thr Lys Ile Cys Pro Val Thr Gly Phe Val Ser Leu Phe Asp
785 790 795 800
Thr Arg Tyr Gln Ser Val Asp Lys Ala Arg Glu Phe Phe Ser Lys Phe
805 810 815
Asp Ser Ile Lys Tyr Asn Glu Glu Lys Glu His Tyr Glu Phe Val Phe
820 825 830
Asp Tyr Ser Asn Phe Thr Asp Lys Ala Lys Asp Thr Lys Thr Lys Trp
835 840 845
Thr Val Cys Ser Tyr Gly Thr Arg Ile Lys Thr Phe Arg Asn Ser Glu
850 855 860
Lys Asn Asn Asn Trp Asp Asn Lys Thr Val Ser Pro Thr Glu Asp Leu
865 870 875 880
Ser Lys Leu Leu Lys Ser Cys Asp Arg Asp Ile Lys Glu Phe Ile Ile
885 890 895
Ser Gln Asp Lys Lys Glu Phe Phe Val Glu Leu Leu Glu Ile Phe Ser
900 905 910
Leu Ile Val Gln Met Lys Asn Ser Ile Ile Asn Ser Glu Ile Asp Tyr
915 920 925
Ile Ile Ser Pro Val Ala Asn Glu Asn Gly Glu Phe Phe Asp Ser Arg
930 935 940
Phe Ala Asn Ser Ser Leu Pro Lys Asn Ala Asp Ala Asn Ala Ala Tyr

945 950 955 960
 Asn Thr Ala Arg Lys Gly Leu Met Leu Leu Glu Lys Ile Arg Asp Ser
 965 970 975
 Glu Ile Gly Lys Lys Ile Asp Met Lys Ile Thr Asn Thr Glu Trp Leu
 980 985 990
 Asn Phe Val Gln Glu Arg
 995
 <210> 63
 <211> 1475
 <212> PRT
 <213> Treponema endosymbiont of Eucomonympha sp.
 <220><221> MISC_FEATURE
 <222> (1)..(1475)

 <223> Genbank WP_062376669 Cpf1
 <400> 63
 Met Thr Ala Phe Glu Glu Leu Thr Asn Leu Tyr Ser Val Ser Lys Thr
 1 5 10 15
 Leu Arg Phe Glu Leu Lys Pro Ala Gly Lys Asp Gly Asn Thr Leu Ser
 20 25 30
 Ala Glu Glu Ser Ala Lys Leu Phe Lys Glu Ile Leu Asp Gln Asp Arg
 35 40 45
 Lys Ile Lys Asp Ala Tyr Leu Ala Leu Lys Pro Val Met Asp Thr Ile
 50 55 60

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

 Glu Lys Gly Lys Pro Ile Phe Lys Lys Lys Lys Gly Lys Asp Val Gly

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

 Leu Ala Val Phe Asp Ser Phe Phe Thr Tyr Phe Gly Gly Tyr Asn Thr
 195 200 205
 Asn Arg Lys Asn Tyr Tyr Thr Glu Asp Lys Glu Lys Ala Thr Ala Val
 210 215 220
 Ala Thr Arg Ile Val His Asp Asn Leu Pro Lys Phe Cys Asp Asn Cys
 225 230 235 240
 Ile Gln Phe Ser Gln Asp Lys Ile Val Thr Lys Arg Lys Ser Lys Glu
 245 250 255

 Thr Ile Pro Ala Arg Lys Asp Glu Tyr Leu Asn Ala Tyr Gln Phe Leu
 260 265 270
 Lys Asp Ser Glu Lys Thr Thr Gln Ile Lys Asp Ala Ala Thr Asn Thr
 275 280 285
 Met Ile Glu Ala Tyr Pro Val Asp Glu Arg Thr Phe Glu Ile Ala Arg
 290 295 300
 Phe Ser Glu Cys Leu Thr Gln Ala Gly Ile Glu Glu Tyr Asn Arg Ile
 305 310 315 320

 Ile Gly His Tyr Asn Ser Leu Ile Asn Leu Tyr Asn Gln Thr Arg Lys
 325 330 335
 Gly Glu Ser Asp Phe Arg Lys Leu Glu Gln Phe Lys Thr Leu His Lys
 340 345 350
 Gln Ile Gly Cys Gly Glu Lys Lys Gly Trp Arg Asp Ala Leu Lys Asp
 355 360 365
 Asn Glu Asp Leu Lys Gly Lys Leu Lys Ala Ile Ser Glu Ala Gly Gln
 370 375 380

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

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

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

 Lys Asp Phe Asp Ile Lys Glu Ser Lys Arg Lys Gly Asn Ala Ile Asn
 580 585 590
 Ala Glu Leu Ser Asn Met Leu Thr Glu Leu Leu His Ala Asp Asp Ala
 595 600 605
 Arg Trp Phe Asp Trp Tyr Asp Leu Val Arg Asn Tyr Leu Thr Lys Lys
 610 615 620
 Pro Gln Asp Asp Ala Lys Ala Asn Lys Leu Lys Leu Asn Phe Gly Tyr

625					630					635					640				
Gly	Lys	Leu	Leu	Asp	Gly	Phe	Val	Asp	Ser	His	Thr	Asp	Glu	Ser	Asp				
645					650					655									
Ala	Ser	Thr	Gln	Tyr	Gly	Gly	Tyr	Leu	Phe	Arg	Lys	Arg	Thr	Asp	Ala				
660					665					670									
Gln	Glu	Lys	Ser	Asp	Phe	Glu	Tyr	Phe	Leu	Gly	Ile	Ser	Lys	Asn	Thr				
675					680					685									
Lys	Leu	Phe	Arg	Cys	His	Leu	Gln	Ser	Thr	Val	Gln	Asn	His	Asp	Lys				
690					695					700									
Ser	Asn	Phe	Glu	Arg	Leu	Glu	Tyr	Tyr	Gln	Ala	Lys	Ser	Thr	Thr	Tyr				
705					710					715					720				
Phe	Asp	Ala	Lys	Tyr	Ser	Glu	Asn	Lys	Ala	Lys	Leu	Val	Glu	Ile	Ile				
725					730					735									
Glu	Ser	Leu	Ile	Asp	Asp	Arg	Ala	Lys	Thr	Asp	Ser	Glu	Leu	Thr	Val				
740					745					750									
Ile	Gly	Asn	Glu	Ile	Lys	Lys	Arg	Asp	Gly	Lys	Gly	Glu	Ile	Thr	Pro				
755					760					765									
Ser	Ala	Leu	Phe	Asp	Arg	Val	Lys	Lys	Asp	Lys	Ala	Phe	Ser	Arg	Ile				
770					775					780									
Leu	Asn	Asp	Asp	Thr	Leu	Leu	Lys	Ala	Val	Thr	Gln	Ala	Ile	Arg	Asp				
785					790					795					800				
Leu	Gln	Asn	Ser	Cys	Asp	Asn	Phe	Ser	Glu	Arg	Ala	Pro	Arg	Leu	Lys				
805					810					815									
Glu	Val	Gln	Lys	Arg	Gly	Tyr	Phe	Gly	Ile	Val	Gly	Phe	Lys	Gln	Ile				
820					825					830									
Val	Glu	Asp	Leu	Gln	Thr	Val	Ala	Lys	Glu	Asn	Lys	Val	Phe	Asn	Phe				
835					840					845									
Phe	Asn	Val	Ser	Gln	Pro	Glu	Phe	Glu	Asn	Ala	Phe	Ile	Ser	Gly	Gly				
850					855					860									
Lys	Arg	Ile	Tyr	Leu	Phe	Lys	Ile	Ser	Asn	Lys	Asp	Leu	Ser	Tyr	Ser				
865					870					875					880				

Glu Thr Ser Gln Lys Asp Glu Asn Gly Gln Gln Lys Arg Asn Phe Lys
885 890 895

Gly Ile Glu Asn Leu His Thr His Tyr Phe Arg Ala Leu Met Arg Glu
900 905 910

Tyr Lys Asn Cys Thr Asn Val Asp Leu Gly Lys Gly Glu Ile Phe Phe
915 920 925

Arg Ala Pro Val Pro Glu Ile Lys Lys Glu Ala Thr His Lys Val Tyr
930 935 940

Asp Lys Met Val Asn Arg Arg Glu Asn Glu Thr His Ile Ala Ile Pro
945 950 955 960

Glu Lys Val His Asn Glu Leu Leu Leu Phe Ala Asn Glu Lys Ile Ala
965 970 975

Val Asn Asn Leu Ser Asn Glu Thr Lys Leu Tyr Leu Asp Gln Asn Lys
980 985 990

Lys Ile Asp Glu Ser Arg Val Lys Ile Lys Asp Val Lys His Asp Ile
995 1000 1005

Ile Lys Asp Lys Arg Phe Thr Glu Ala Lys Tyr Gln Leu His Leu
1010 1015 1020

Ser Ile Leu Leu Asn Phe Thr Pro Thr Lys Glu Glu Val Asn Ala
1025 1030 1035

Lys Ile Asn Asp Thr Phe Thr Lys Ser Asp Asp Ile Gln Phe Leu
1040 1045 1050

Gly Ile Asp Arg Gly Glu Lys His Leu Ile Tyr Tyr Ser Leu Val
1055 1060 1065

Asp Ala Asn Gly Thr Ile Arg Ala Gln Asp His Phe Asp Val Ile
1070 1075 1080

Asn Lys Thr Asp Tyr Leu Gln Lys Ile Thr Glu Ala Ala Lys Ile
1085 1090 1095

Arg Arg Glu Lys Gln Glu Asn Trp Gln Gln Lys Gly Asn Ile Ser
1100 1105 1110

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

1115	1120	1125
Glu Lys Met Lys Asp Glu Asn Gly Ser Phe Lys Pro Met Phe Ile		
1130	1135	1140
Val Leu Glu Asp Leu Asn Thr Gly Phe Lys Arg Ser Arg Gln Lys		
1145	1150	1155
Phe Glu Gln Gln Val Tyr Gln Lys Phe Glu Leu Ala Leu Ala Lys		
1160	1165	1170
Lys Leu Asn Tyr Leu Val Asp Lys Asn Ala Lys Asp Gly Glu Leu		
1175	1180	1185
Ala Ser Val Ser Arg Ala Leu Gln Leu Thr Pro Leu Val Met Asn		
1190	1195	1200
Tyr Gln Asp Ile Glu Asn Arg Lys Gln Val Gly Ile Met Leu Tyr		
1205	1210	1215
Thr Arg Ala Asn Tyr Thr Ser Val Thr Asp Pro Ala Thr Gly Trp		
1220	1225	1230
Arg Lys Thr Val Tyr Leu Lys Pro Gly Ser Glu Glu Ser Ile Lys		
1235	1240	1245
Lys Gln Ile Leu Gly Val Phe Ser Glu Ile Gly Val Asp Glu Lys		
1250	1255	1260
Gly Asp Tyr Phe Phe Gln Tyr Ser Asp Thr Asn Thr Glu Arg Thr		
1265	1270	1275
Trp Arg Leu Trp Ser Ser Lys Asn Gly Lys Ser Leu Glu Arg Tyr		
1280	1285	1290
Arg Gly Arg Arg Asp Lys Asn Thr Asn Ala Phe Thr Val Glu Pro		
1295	1300	1305
Tyr Asp Val Lys Ala Thr Leu Asn Lys Leu Phe Val Gly Phe Asp		
1310	1315	1320
Lys Asp Met Ser Leu Leu Lys Gln Leu Lys Asp Gly Lys Thr Leu		
1325	1330	1335
Ser Lys Ile Asp Gly Arg Lys Glu Thr Ala Trp Glu Ser Leu Arg		
1340	1345	1350

Phe Val Ile Asp Leu Ile Gln Gln Ile Arg Asn Ser Gly Asp Thr
1355 1360 1365

Ser Lys Asn Gln Asp Asp Asn Phe Leu Leu Ser Pro Val Arg Asn
1370 1375 1380

Ala Gln Gly Glu His Phe Asp Ser Arg Leu Tyr Gln Asn Gln Glu
1385 1390 1395

Thr Pro Lys Leu Pro Lys Asp Ala Asp Ala Asn Gly Ala Tyr Asn
1400 1405 1410

Ile Ala Arg Lys Gly Ile Ile Met Tyr Ala His Ile Arg Gln Trp
1415 1420 1425

Ile Asn Asp Gly Gly Gln Arg Phe Glu Lys Ser Ser Asp Leu Asp
1430 1435 1440

Leu Phe Val Ser Asp Asn Glu Trp Asp Leu Trp Leu Phe Asp Ser
1445 1450 1455

Lys Gln Trp Lys Glu Gln Leu Gln Lys Phe Ala Ser Arg Lys Gln
1460 1465 1470

Lys Lys
1475

<210> 64

<211> 1251

<212> PRT

<213> Moraxella lacunata

<220><221> MISC_FEATURE

<222> (1)..(1251)

<223> Genbank WP_062499108 Cpf1

<400> 64

Met Leu Phe Gln Asp Phe Thr His Leu Tyr Pro Leu Ser Lys Thr Val

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

Val Lys Ala Ile Leu Asp Asp Tyr His Arg Asp Phe Ile Thr Lys Met
50 55 60
Met Ser Glu Val Thr Leu Thr Lys Leu Pro Glu Phe Tyr Glu Val Tyr
65 70 75 80
Leu Ala Leu Arg Lys Asn Pro Lys Asp Asp Thr Leu Gln Lys Gln Leu
85 90 95
Thr Glu Ile Gln Thr Ala Leu Arg Glu Glu Val Val Lys Pro Ile Asp
100 105 110
Ser Gly Gly Lys Tyr Lys Ala Gly Tyr Glu 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 Pro Gln Ile 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 Ser Asp 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 Val Thr
195 200 205
Ile Lys Gln Lys His Ser Val Leu Tyr Asp Gln Ile Val Asn Glu Leu
210 215 220
Asn Ala Asn 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 Arg Ile Ile
245 250 255
Gly Glu Val Asn Ser Tyr Thr Asn Lys His Asn Gln Ile Cys His Lys
260 265 270
Ser Glu Arg Ile Ala Lys Leu Arg Pro Leu His Lys Gln Ile Leu Ser
275 280 285
Asp Gly Met Gly Val Ser Phe Leu Pro Ser Lys Phe Ala Asp Asp Ser

290 295 300
 Glu Met Cys Gln Ala Val Asn Glu Phe Tyr Arg His Tyr Ala His Val
 305 310 315 320
 Phe Ala Lys Val Gln Ser Leu Phe Asp Arg Phe Asp Asp Tyr Gln Lys

 325 330 335
 Asp Gly Ile Tyr Val Glu His Lys Asn Leu Asn Glu Leu Ser Lys Gln
 340 345 350
 Ala Phe Gly Asp Phe Ala Leu Leu Gly Arg Val Leu Asp Gly Tyr Tyr
 355 360 365
 Val Asp Val Val Asn Pro Glu Phe Asn Asp Lys Phe Ala Lys Ala Lys
 370 375 380
 Thr Asp Asn Ala Lys Glu Lys Leu Thr Lys Glu Lys Asp Lys Phe Ile

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

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

 515 520 525
 Ala Lys Ile Ala Thr Leu Tyr Asn Lys Val Arg Asp Tyr Leu Ser Gln
 530 535 540

Lys Pro Phe Ser Thr Glu Lys Tyr Lys Leu Asn Phe Gly Asn Pro Thr
 545 550 555 560
 Leu Leu Asn Gly Trp Asp Leu Asn Lys Glu Lys Asp Asn Phe Gly Val
 565 570 575
 Ile Leu Gln Lys Asp Gly Cys Tyr Tyr Leu Ala Leu Leu Asp Lys Ala
 580 585 590
 His Lys Lys Val Phe Asp Asn Ala Pro Asn Thr Gly Lys Ser Val Tyr
 595 600 605
 Gln Lys Met Val Tyr Lys Leu Leu Pro Gly Ser Asn Lys Met Leu Pro
 610 615 620
 Lys Val Phe Phe Ala Lys Ser Asn Leu Asp Tyr Tyr Asn Pro Ser Ala
 625 630 635 640
 Glu Leu Leu Asp Lys Tyr Ala Gln Gly Thr His Lys Lys Gly Asp Asn
 645 650 655
 Phe Asn Leu Lys Asp Cys His Ala Leu Ile Asp Phe Phe Lys Ala Ser
 660 665 670
 Ile Asn Lys His Pro Glu Trp Gln His Phe Gly Phe Glu Phe Ser Leu
 675 680 685
 Thr Ser Ser Tyr Gln Asp Leu Ser Asp Phe Tyr Arg Glu Val Glu Pro
 690 695 700
 Gln Gly Tyr Gln Val Lys Phe Val Asp Ile Asp Ala Asp Tyr Ile Asp
 705 710 715 720
 Glu Leu Val Glu Gln Gly Gln Leu Tyr Leu Phe Gln Ile Tyr Asn Lys
 725 730 735
 Asp Phe Ser Pro Lys Ala His Gly Lys Pro Asn Leu His Thr Leu Tyr
 740 745 750
 Phe Lys Ala Leu Phe Ser Glu Asp Asn Leu Ala Asn Pro Ile Tyr Lys
 755 760 765
 Leu Asn Gly Glu Ala Glu Ile Phe Tyr Arg Lys Ala Ser Leu Asp Met
 770 775 780
 Asn Glu Thr Thr Ile His Arg Ala Gly Glu Val Leu Glu Asn Lys Asn

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

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

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

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

 1025 1030 1035

Pro Val Thr Gly Phe Val Asp Leu Leu Lys Pro Arg Tyr Glu Asn		
1040	1045	1050
Ile Ala Gln Ser Gln Ala Phe Phe Asp Lys Phe Asp Lys Ile Cys		
1055	1060	1065
Tyr Asn Ala Asp Lys Gly Tyr Phe Glu Phe His Ile Asp Tyr Ala		
1070	1075	1080
Lys Phe Thr Asp Lys Ala Lys Asn Ser Arg Gln Ile Trp Thr Ile		
1085	1090	1095
Cys Ser His Gly Asp Lys Arg Tyr Val Tyr Asp Lys Thr Ala Asn		
1100	1105	1110
Gln Asn Lys Gly Ala Thr Ile Gly Ile Asn Val Asn Asp Glu Leu		
1115	1120	1125
Lys Ser Leu Phe Ala Arg Tyr Arg Ile Asn Asp Lys Gln Pro Asn		
1130	1135	1140
Leu Val Met Asp Ile Cys Gln Asn Asn Asp Lys Glu Phe His Lys		
1145	1150	1155
Ser Leu Thr Tyr Leu Leu Lys Ala Leu Leu Ala Leu Arg Tyr Ser		
1160	1165	1170
Asn Ala Ser Ser Asp Glu Asp Phe Ile Leu Ser Pro Val Ala Asn		
1175	1180	1185
Asp Lys Gly Val Phe Phe Asn Ser Ala Leu Ala Asp Asp Thr Gln		
1190	1195	1200
Pro Gln Asn Ala Asp Ala Asn Gly Ala Tyr His Ile Ala Leu Lys		
1205	1210	1215
Gly Leu Trp Leu Leu Asn Glu Leu Lys Asn Ser Asp Asp Leu Asp		
1220	1225	1230
Lys Val Lys Leu Ala Ile Asp Asn Gln Thr Trp Leu Asn Phe Ala		
1235	1240	1245
Gln Asn Arg		
1250		

<210> 65

<211> 50

<212> DNA
 <213> Artificial Sequence
 <220><223> Chemically synthesized cTAG_A sequence
 <400> 65
 actgggtgga atcccttctg cagcacctgg attacctgt tatccctagt 50

<210> 66
 <211> 50
 <212> DNA
 <213> Artificial Sequence
 <220><223> Chemically synthesized cTAG_B sequence
 <400> 66
 taatgagtag tcctcatctc cctcaagcag gcgccggcgg tactgccatc 50

<210> 67
 <211> 50
 <212> DNA
 <213> Artificial Sequence
 <220><223> Chemically synthesized cTAG_C sequence
 <400> 67
 catataatct ccctcaagca ggccccgctg gcgcgcgcga atgttaggaa 50

<210> 68
 <211> 50
 <212> DNA
 <213> Artificial Sequence
 <220><223> Chemically synthesized cTAG_D sequence
 <400> 68
 gcctataatg tgaagagctt cactgagtag ggccccgggct gtaaacggtt 50

<210> 69
 <211> 50
 <212> DNA
 <213> Artificial Sequence
 <220><223> Chemically synthesized cTAG_E sequence
 <400> 69
 attcgctagc agatgtagtg ttccacagg ggcgatcgct gatatgggtc 50

<210> 70
 <211> 50
 <212> DNA
 <213> Artificial Sequence
 <220><223> Chemically synthesized cTAG_F sequence
 <400> 70
 actacctagc tgcattttca ggaggaagcg atgggcggcc gcacaccttc 50
 <210> 71

 <211> 50
 <212> DNA
 <213> Artificial Sequence
 <220><223> Chemically synthesized cTAG_G sequence
 <400> 71
 tgataatggg tgagttagtg tgtgcgtgtg ggcgcgcgca gatgggaaca 50
 <210> 72
 <211> 50
 <212> DNA
 <213> Artificial Sequence
 <220><223> Chemically synthesized cTAG_H sequence
 <400> 72
 actccagtct ttctagaaga tggcaaacag ctattatggg tattatgggt 50
 <210> 73
 <211> 50
 <212> DNA
 <213> Artificial Sequence
 <220><223> Chemically synthesized cTAG_I sequence

 <400> 73
 tagtggacgg ggccactagg gacaggattg gctgcagga ttcccgtaa 50
 <210> 74
 <211> 42
 <212> DNA
 <213> Artificial Sequence
 <220><223> Chemically synthesized cTAG_J sequence

<400> 74
tgaactaagg cggctgcaca accagtggag gcctaaatga tc 42

<210> 75
<211> 25
<212> DNA
<213> Artificial Sequence
<220><223> Chemically synthesized tagD_FWD PCR primer
<400> 75
gcctataatg tgaagagctt cactg 25

<210>
> 76
<211> 19
<212> DNA
<213> Artificial Sequence
<220><223> Chemically synthesized tagE_REV PCR primer
<400> 76
gaccatatac agcgatcgc 19

<210> 77
<211> 9312
<212> DNA
<213> Artificial Sequence
<220><223> Plasmid 13000444591
<400> 77
tccaaacgag agtctaataa aatgaggtcg aaaagtaaat cgcgcggtt tgttactgat 60
aaagcaggca agacataaaa tgtgttaaagg gcaaagtgtg tactttggcg tcacccctta 120
catatcttag gtcttttttt attgtgcgta actaacttgc catcttcaaa caggagggtt 180

ggaagaagca gaccgctaac acagtacata aaaaaggaga catgaacgac tccagtcctt 240
ctagaagatg gcaaacagct attatgggtg ttatgggtcc ccgaagcagg gttatgcagc 300
ggaaaagctc cccgaaaagt gccacctggg tccttttcat cacgtgctat aaaaataatt 360
ataatttaaa ttttttaata taaatatata aattaaaaat agaaagtaaa aaaagaaatt 420
aaagaaaaaa tagtttttgt tttccgaaga tgtaaaagac tctaggggga tcgccaacaa 480
atactacctt ttatcttgct ctctctgctc tcaggtatta atgccgaatt gtttcatctt 540
gtctgtgtag aagaccacac acgaaaatcc tgtgatttta cattttactt atcgtaatac 600

gaatgtatat ctatttaac tgcttttctt gtctaataaa tatatatgta aagtacgctt 660
 ttgttgaaa ttttttaaac ctttgtttat ttttttttc ttcatccgt aactcttcta 720
 ccttctttat ttactttcta aaatccaaat acaaaacata aaaataaata aacacagagt 780
 aaattcccaa attattccat cattaaaaga tacgaggcgc gtgtaagtta caggcaagcg 840
 atccgtctaa gaaaccatta ttatcatgac attaacctat aaaaataggc gtatcacgag 900
 gcccttctgt ctgcgcggtt tcggtgatga cggtgaaaac ctctgacaca tgcagctccc 960
 ggagacggtc acagcttgtc tgtaagcgga tgccgggagc agacaagccc gtcagggcgc 1020

 gtcagcgggt gttggcgggt gtcgggctg gcttaactat gcggcatcag agcagattgt 1080
 actgagagtg caccatacca cagccggaag aggagtaggg aatattactg gctgaaaata 1140
 agtcttgaat gaacgtatac gcgtatattt ctaccaatct ctcaacactg agtaatggta 1200
 gttataagaa agagaccgag ttagggacag ttagaggcgg tggagatatt ctttatggca 1260
 tgtctggcga tgataaaact tttcaaacgg cagccccgat ctaaaagagc tgacagggaa 1320
 atggtcagaa aaagaaacgt gcacccgcc gtctggacgc gccgtcacc cgcacggcag 1380
 agaccaatca gtaaaaatca acggttaacg acattactat atatataata taggaagcat 1440

 ttaatagaac agcatcgtaa tataatgtgta ctttgcagtt atgacgccag atggcagtag 1500
 tggaagatat tctttattga aaaatagctt gtcaccttac gtacaatctt gatccggagc 1560
 ttttcttttt ttgccgatta agaattcggc cgaaaaaaga aaaggagagg gccaaagagg 1620
 agggcatttg tgactattga gcacgtgagt atacgtgatt aagcacacaa aggcagcttg 1680
 gagtatgtct gttattaatt tcacaggtag ttctgggtcca ttggtgaaag tttgcggctt 1740
 gcagagcaca gaggccgcag aatgtgctct agattccgat gctgacttgc tgggtattat 1800
 atgtgtgccc aatagaaaga gaacaattga cccggttatt gcaaggaaaa tttcaagtct 1860

 tgtaaaagca tataaaaaata gttcaggcac tccgaaatac ttggttggcg tgtttcgtaa 1920
 tcaacctaa gaggatgttt tggctctggt caatgattac ggcatgata tcgtccaact 1980
 gcatggagat gagtctggc aagaatacca agagttctc ggtttgccag ttattaaaag 2040
 actcgtattt ccaaaagact gcaacatact actcagtga gcttcacaga aacctattc 2100
 gtttattccc ttgtttgatt cagaagcagg tgggacagg gaacttttgg attggaactc 2160
 gatttctgac tgggttgga ggcaagagag ccccgaaagc ttacatttta tgttagctgg 2220
 tggactgacg ccagaaaatg ttggtgatgc gcttagatta aatggcgta ttggtgtga 2280

 tgtaagcgga ggtgtggaga caaatggtgt aaaagactct aacaaaatag caaatctgt 2340
 caaaaatgct aagaaatagg ttattactga gtagtattta ttttaagtatt gtttgtcac 2400
 ttgcctgcag gccttttgaa aagcaagcat aaaagatcta aacataaaat ctgtaaaata 2460

acaagatgta aagataatgc taaatcattt ggctttttga ttgattgtac aggactgggt 2520
ggaatccctt ctgcagcacc tggattaccc tgttatccct agttttgatg ggtatcggt 2580
ttgctactgg ggaatatcgc gcggcagagg cagcaaaaaa agcaatttcc agcccgttc 2640
ttgaagcggc cattgacggg gcgcaaggcg tcctcatgaa catcactgga ggaacaaacc 2700

tcagcctata tgaggttcag gaagcagcag acattgtcgc ttcggcgtct gatcaagacg 2760
taaacatgat tttcggttct gttattaatg aaaatctaaa agatgagatt gtggtgacag 2820
tgattgcaac cggctttatc gaacaagaga aggacgtgac gaagcctcag cgcccaagct 2880
taaatacaag catcaaaaca cacaatcaaa gtgttccgaa gcgtgacgca aaacgtgagg 2940
aacctcagca gcagaacaca gtaagccgtc atacttcaca gccggctgat gatacgcttg 3000
acatcccgac attcttaaga aaccgtaata aacgcggcta atgtaaagga caaatcggt 3060
ttcgattttg tcttttttgt ttttctcttc acacttcctt cttataaagt ctttttcct 3120

attgctttcc ttcgcttagt aacaaaacag ataattagac ccatattttt ttgtgacatt 3180
tttatcattt tcatatatat ggaaattgaa tgacatgaaa cgacaatatc tgtaattcag 3240
attgtctaca gttaatatac agcgatgttc tgacaaacca ttcattatta aaaggaggga 3300
cgacactttt tttaaaaagc atgttgaaaa aggggggatga aaatgaggaa aaaaacgaaa 3360
aacagactca tcagctctgt ttttaagtaca gttgtcatca gttcactgct gtttccggga 3420
gcagccgggg caagcagtaa agtcacctca cttctgtta aaaaggagct tcaatctgcg 3480
gaatccattc aaacaagat ttcgagttca ttaaagaaaa gctttaaaaa gaaagaaaaa 3540

acgacttttc tgattaaatt taaagatctt aatgagtagt cctcatctcc ctcaagcagg 3600
cgccggcggg actgccatcc aaaaatcaga ccagacaaaa gcggcaaatg aataagcggg 3660
acggggaagg atttgcggtc aagtcttcc cttccgcag tatcaattcg caagcttttc 3720
ctttataata gaatgaatga gaaggatgaa accgaatgaa cacttatgaa caaattaaca 3780
aagtgaaaaa gatcttacgc aaacacttga aaaataactt aattggcaca tacatgttcg 3840
gtagcggagt agaaagcggc cttaaacgga acagcgacct tgacttcctt gttgtgttt 3900
cagagccttt aacggatcaa tctaaagaaa ttttaatcca gaaaattcgc ccaatctcta 3960

aaaaaatcgg agacaaatca aatctgcgct atatgaact tacaattatt attcaacaag 4020
agatggtecc atggaatcac cctccgaaac aagagttcat ttacggtgag tggttgcagg 4080
aactctatga acaaggatat atcccgcaga aggagcttaa ctccgatctt actatcatgc 4140
tttaccaagc gaaacgtaaa aataaacgta tctacggtaa ctacgatctc gaagaacttc 4200
ttctgatata tccattctct gatgttcgcc gcgccatcat ggatagctct gaagaactca 4260
tcgacaacta ccaggacgat gaaacaaact ctattcttac attatgtcgc atgattctta 4320

ctatggatac gggtaaaatc atcccaaagg atattgcagg aaacgcagta gctgaatcct	4380
ccctcttga gcatcgcgaa cgcattcttc ttgctgtacg ttcttacctt ggggaaaata	4440
tcgaatggac aaacgaaaat gttaatTTaa caatcaacta tcttaacaac cgccttaaaa	4500
aactgtgaaa aaaagcgagc ctgaaatagc tgcgcttttt tgtgtcataa catataatct	4560
ccctcaagca ggccccgctg gcgcgcgcga atgttaggaa acgattagtc ttttgactgt	4620
ttgacggtagg tggtagctgg gcctataatg tgaagagctt cactgagtag ggccccggct	4680
glaaacgggtt aattttgtca aaataatttt attgacaacg tcttattaac gttgatataa	4740
tttaaatTTT atttgacaaa aatgggctcg tgtgttaca taaatgttac tagagaaagg	4800
tggtgaatac tagatgacgg cattgacgga aggtgcaaaa ctgtttgaga aagagatccc	4860
gtatatcacc gaactggaag gcgacgtcga aggtatgaaa tttatcatta aaggcgaggg	4920
taccggtgac gcgaccacgg gtaccattaa agcgaaatac atctgcaacta cggcgacct	4980
gccggtcccc tgggcaaccc tggtagcac cctgagctac ggtgttcagt gtttcgcaa	5040
gtacccgagc cacatcaagg atttcttta gagcgccatg ccggaagggtt ataccaaga	5100
gcgtaccatc agcttcgaag gcgacggcgt gtacaagacg cgtgctatgg ttacctacga	5160
acgcggttct atctacaatc gtgtcacgct gactggtagg aactttaaga aagacggtca	5220
cattctgcgt aagaacgttg cattccaatg cccgccaagc attctgtata ttctgcctga	5280
caccgttaac aatggcatcc gcgttgagtt caaccaggcg tacgatattg aaggtgtgac	5340
cgaaaaactg gttaccaaat gcagccaaat gaatcgtccg ttggcgggct ccgcggcagt	5400
gcataccccg cgttatcatc acattacctt ccacacaaa ctgagcaaag accgcgacga	5460
gcgcctgat cacatgtgtc tggtagaggt cgtgaaagcg gttgatctgg acacgtatca	5520
glaaaaaaaa agcgagctg aaatagctgc gctttttgt gtcataaccc tttacagtca	5580
taaaaattat ggtataatca tttctgttgt ctttttaaag acacaagcat gaccattatg	5640
actagattcg ctagcagatg tagtgtttcc acaggggcga tcgctgatat gggccagtt	5700
caaaagggat tatggctgat aaagattttt atctcgtata tatccagtgc aaacctgac	5760
cgtattcacc tggattggca atggatgaaa ccggccagaa ttccggccgc aactggcagt	5820
atatagatgg aaagtggcag ccaggtgaca aagcggatgg caactatatg attcgcgat	5880
tagttgatta tgaagctgct gtacctgaga ttacttcacc gacagacaaa tcatacaca	5940
ataaggatag cgtcactgta aaaggaaacg cttctcctgg cacaacggta cacatttata	6000
atggagagaa agaagcagga gaaacgaaag ctgctgcgga tggcacgttc catgcaggca	6060

tcatactcaa caagggtgaa aatgagctga cggcaactgc atcaactgac aacggaacaa	6120
cagatgcctc cagcccaatc acggtcacgc ttgatcaaga aaagcctgaa ttaacttg	6180
acaatccaaa ggatggcggg aaaacaaata aagaaacgct gactgtcaaa ggggctgtat	6240
ccgatgacaa tcigaaagac gtcaagggtga atggcaaaaa agcaacagta gctgatggtt	6300
catactcagc ccgtattctt ttggaaaatg gaagaaatga aatcaaggta attgctacag	6360
acttggcagg caacaaaacg acgaaaaaga cagtcattga tgtgaacttt gacaagcctg	6420
tcatttccgg cttaattccg ggagaggata aaaacttaaa agccgggtgaa tctgtgaaaa	6480
tcgctttctc aagcgtgag gatttagatg caacgtttac cattcgtatg ccgtgacca	6540
atgcaagagc gagtgtgcaa aatgccaccg aactcccggt aagagaaatc tctccgggga	6600
gatatgaagg ctattggact gccacttctt ctattaaagc aaaaggagca aaagtagaag	6660
tgatcgctcg agatgattat ggaaatgaaa caagaaaaac tgcgaatgga aaacttaata	6720
tgaacacaga aaatactacc tagctgcatt ttcaggagga agcgtatggc gcccgcacac	6780
cttctatcgc gtgtgaaata ccgcatgac ccttaaatat tctgacaaat gctctttccc	6840
taaactcccc ccataaaaaa acccgccgaa gcgggttttt acgttatttg cggattaacg	6900
attactcggt atcagaaccg cccaggatgc ctggcagttc cctactctcg ccgtgcgct	6960
cggtcggttcg gctgcgggac ctcagcgcta gcggagtgtg tactggctta ctatgttggc	7020
actgatgagg gtgtcagtga agtgcattcat gtggcaggag aaaaaaggct gcaccggtgc	7080
gtcagcagaa tatgtgatac aggatataat ccgcttcctc gctcactgac tcgtacgct	7140
cggtcggttcg actgcggcga gcggaaatgg cttacgaacg gggcggagat ttcttgaag	7200
atgccaggaa gatacttaac aggggaagtga gagggccgcg gcaaagccgt tttccatag	7260
gctccgcccc cctgacaagc atcacgaaat ctgacgtca aatcagtggg gccgaaaccc	7320
gacaggacta taaagatacc aggcgttttc cctggcggc tcctcgtgc gctctcctgt	7380
tcctgccttt cggtttaccg gtgtcattcc gctgttatgg ccgcgtttgt ctcattccac	7440
gcctgacact cagttccggg taggcagttc gctccaagct ggactgtatg cacgaacccc	7500
ccgttcagtc cgaccgtgc gccttatecg gtaactatcg tcttgagtcc aaccggaaa	7560
gacatgcaaa agcaccactg gcagcagcca ctggtaatg atttagagga gttagtcttg	7620
aagtcatcgc ccggttaagg ctaaaactgaa aggacaagtt ttggtgactg cgctcctcca	7680
agccagttac ctcggttcaa agagtgggia gctcagagaa ccttcgaaaa accgcctgc	7740
aaggcggttt tttcgttttc agagcaagag attacgcgca gacaaaaacg atctcaagaa	7800
gatcatctta ttaagctca ctcattagc acccaggct ttacactgat aatgggtgag	7860
tgagtgtgtg cgtgtggggc gcgccagatg ggaacagcta gcttcacgct gccgcaagca	7920

ctcagggcgc aagggtgct aaaggaagcg gaacacgtag aaagccagtc cgcagaaacg 7980
 gtgctgaccc cggatgaatg tcagctactg ggctatctgg acaagggaaa acgcaagcgc 8040
 aaagagaaag caggtagctt gcagtgggct tacatggcga tagctagact gggcggtttt 8100
 atggacagca agcgaaccgg aattgccagc tggggcgccc tctggttaagg ttgggaagcc 8160

ctgcaaagta aactggatgg ctttcttgcc gccaaggatc tgatggcgca ggggatcaag 8220
 atctgatcaa gagacaggat gaggatcggt tcgcatgatt gaacaagatg gattgcacgc 8280
 aggttctccg gccgcttggg tggagaggct attcggctat gactgggcac aacagacaat 8340
 cggctgctct gatgccgcgc gtgtccggct gtcagcgagc gggcgcccgg tttttttgt 8400
 caagaccgac ctgtccggtg ccctgaatga actccaagac gaggcagcgc ggctatctg 8460
 gctggccacg acgggcgttc cttgcgcagc tgtgctcgac gttgtcactg aagcggaag 8520
 ggactggctg ctattggcg aagtgccggg gcaggatctc ctgtcatctc accttgctcc 8580

tgccgagaaa gtatccatca tggctgatgc aatgcggcgg ctgcatacgc ttgatccggc 8640
 tactgccc ttcgaccacc aagcgaaca tcgcatcgag cgagcacgta ctcgatgga 8700
 agccggtctt gtcgatcagg atgatctgga cgaagagcat caggggctcg cgccagccga 8760
 actgttcgcc aggtcaagg cgcggatgcc cgacggcgag gatctcgtcg tgacctatgg 8820
 cgatgcctgc ttgccgaata tcatggtgga aaatggccgc ttttctggat tcatcgactg 8880
 tggccggctg ggtgtggcgg accgctatca ggacatagcg ttggctaccc gtgatattgc 8940
 tgaagagctt ggcgcgcaat gggctgaccg cttcctcgtg ctttacggta tcgccgctcc 9000

cgattcgag cgcatcgct tctatcgct tcttgacgag ttcttctgag cgggactctg 9060
 gggttcgcta gaggatcgat ctttttaac ccatcacata tactgcccgt tcactattat 9120
 ttagtgaaat gagatattat gatattttct gaattgtgat taaaaaggca actttatgcc 9180
 catgcaacag aaactataaa aaatacagag aatgaaaaga aacagataga ttttttagtt 9240
 ctttaggcc gtagtctgca aatcctttta tgattttcta tcaaacaaaa gaggaaaata 9300
 gaccagttgc aa 9312

<210> 78
 <211> 50
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Chemically synthesized cloning tag cTag K from pMLN005
 <400> 78

gcagatgtag tgtttcaca gggccccct tattaaatga cttctgaaa 50

<210> 79

<211> 77

<212> DNA

<213> Artificial Sequence

<220>

<223> Chemically synthesized cloning tag cTag L from pMLN005

<400> 79

tttcgtctag agccttttgt ataaaaatta gggaccttgc actgactgcc ccccggtgag 60

tgagtgtgtg cgtgtgg 77

<210> 80

<211> 50

<212> DNA

<213> Artificial Sequence

<220>

<223> Chemically synthesized cloning tag cTag K' from pMAS006

<400> 80

gcagatgtag tgtttcaca gggtttcgag aagtcattta ataagcccc 50

<210> 81

<211> 50

<212> DNA

<213> Artificial Sequence

<220>

<223> Chemically synthesized cloning tag cTag L' from pMAS006

<400> 81

aaaaaataca aaaggctcta gacgaaaggt gaggtagtgt gtgcgtgtgg 50

<210> 82

<211> 2781

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetically constructed plasmid pMLN005 CamR in vitro

<400> 82

gcattttcag gaggaagcga tgggcggccg cacaccttct catgacaaa atcccttaac	60
gtgagttttc gttccactga gcgtcagacc ccgtagaaaa gatcaaagga tcttcttgag	120
atcctttttt tctgcgcgta atctgctgct tgcaaacaaa aaaaccaccg ctaccagcgg	180
tggtttgttt gccggatcaa gagctiacaa ctctttttcc gaaggtaact ggcttcagca	240
gagcgcagat accaaatact gtccttctag tgtagccgta gttaggccac cacttcaaga	300
actctgtagc accgcctaca tacctcgctc tgctaactct gttaccagtg gctgctgcca	360
gtggcgataa gtcgtgtctt accgggttgg actcaagacg atagttaccg gataaggcgc	420
agcggtcggg ctgaacgggg ggttcgtgca cacagcccag cttggagcga acgacctaca	480
ccgaactgag atacctacag cgtgagctat gagaaagcgc cacgcttccc gaaggagaaa	540
aggcggacag gtatccggtg agcggcaggg tcggaacagg agagcgcacg agggagcttc	600
cagggggaaa cgcttggtat ctttatagtc ctgtcgggtt tcgccacctc tgacttgagc	660
gtcgattttt gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc agcaacgcgg	720
cctgataatt gtaggctgga taagatgcgt cagcatcgca tccggcaaag gcagatctcg	780
cagatgtagt gttccacag ggcccccttg atcgggcacg taagaggttc caactttcac	840
cataatgaaa taagatcact accgggcgta ttttttgagt tatcgagatt ttcaggagct	900
aaggaagcta aaatggagaa aaaaatcact ggatatacca ccgttgatat atcccaatgg	960
catcgtaaag aacattttga ggcatctcag tcagttgctc aatgtacctc taaccagacc	1020
gttcagctgg atattacggc ctttttaaag accgtaaaga aaaataagca caagttttat	1080
ccggccttta ttcacattct tgcccgcctg atgaatgctc atccggaatt tcgtatggca	1140
atgaaagacg gtgagctggt gatatgggat agtgttcacc cttgttacac cgttttccat	1200
gagcaaaactg aaacgttttc atcgctctgg agtgaatacc acgacgattt ccggcagttt	1260
ctacacatat attcgcaaga tgtggcgtgt tacggtgaaa acctggccta tttccctaaa	1320
gggtttattg agaatatgtt tttcgtttca gccaatccct gggtgagttt caccagtttt	1380
gatttaaagc tggccaatat ggacaacttc ttcgcccccg ttttcacat gggcaaatat	1440
tatacgcaag gcgacaaggt gctgatgccg ctggcgattc aggttcatca tgccgtttgt	1500
gatggcttcc atgtcggcag aatgcttaat gaattacaac agtactgcga tgagtggcag	1560
ggcggggcgt aatttgatat cgagctcgct tggactcctg ttgatagatc cagtaatgac	1620
ctcagaactc catctggatt tgttcagaac gctcggttgc cgccgggcgt tttttattgg	1680
taaaaattag ggaccttgca ctgactgcc cccggtgagt gagtgtgtgc gtgtggtcag	1740
aattggttaa ttggttgtaa cactgacccc tatttgttta tttttctaaa tacattcaaa	1800
tatgtatccg ctcatgagac aataaccctg ataaatgctt caataatatt gaaaaaggaa	1860

gaatatgagt attcaacatt tccgtgtcgc ccttattccc ttttttgcgg catTTTgcct 1920
 tcctgttttt gtcacccag aaacgtggt gaaagtaaaa gatgctgaag atcagttggg 1980
 tgcacgagtg ggttacatcg aactggatct caacagcggg aagatccttg agagttttcg 2040
 cccgaagaa cgTTTTccaa tgatgagcac ttttaaagtt ctgctatgtg gcgcggtatt 2100

atcccgattt gacgccgggc aagagcaact cggtcgccgc atacactatt ctcagaatga 2160
 ctgtgttag taccacccag tcacagaaaa gcatcttacg gatggcatga cagtaagaga 2220
 attatgcagt gctgccataa ccatgagtga taacactgcg gccaaattac ttctgacaac 2280
 gatcggagga ccgaaggagc taaccgcttt ttgcacaac atggggggtc atgtaactcg 2340
 ccttgatcgt tgggaaccgg agctgaatga agccatacca aacgacgagc gtgacaccac 2400
 gatgcctgta gcgatggcaa caacgttgcg caaactatta actggcgaac tacttactct 2460
 agcttcccg caacaattaa tagactggat ggaggcggat aaagtgcag gaccacttct 2520

gcgctcggcc cticcggctg gctggtttat tgctgataaa tccggagccg gtgagcgtgg 2580
 ttctcgggt atcatcgag cgctggggcc agatggtaag ccctcccgtc tcgtagttat 2640
 ctacacgacg gggagtcagg caactatgga tgaacgaaat agacagatcg ctgagatagg 2700
 tgcctcactg attaagcatt ggtaataatg aagcgatagc gccggccttag tcagatttaa 2760
 tctgcgcgcg tggtagtat t 2781

<210> 83

<211> 3128

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetically constructed plasmid 13000789485 pMLN 005

<400> 83

gcattttcag gaggaagcga tgggcggccg cacaccttct catgacaaa atccctaac 60
 gtgagtttgc gttccactga gcgtcagacc ccgtagaaaa gatcaaagga tcttcttgag 120
 atcctttttt tctgcgcgta atctgctgct tgcaaaaaaaa aaaaccaccg ctaccagcgg 180
 tggttttgtt gccggatcaa gagctaccaa ctctttttcc gaaggtaact ggcttcagca 240
 gagcgcatg accaaatact gtccttctag ttagccgta gttaggccac cacttcaaga 300
 actctgtagc accgcctaca taccctgctc tgctaactct gttaccagtg gctgctgcca 360
 gtggcgataa gtcgtgtctt accgggttgg actcaagacg atagttaccg gataaggcgc 420

agcggtcggg ctgaacgggg ggttcgtgca cacagcccag cttggagcga acgacctaca 480

ccgaactgag atacctacag cgtgagctat gagaaagcgc cacgcttccc gaaggagaaa 540
 aggcggacag gtatccggtg agcggcaggg tcggaacagg agagcgcacg agggagcttc 600
 cagggggaaa cgcctggtat ctttatagtc ctgtcgggtt tcgccacctc tgacttgagc 660
 gtcgattttt gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc agcaacgcgg 720
 cctgataatt gtaggctgga taagatgcgt cagcatcgca tccggcaaag gcagatctcg 780
 cagatgtagt gtttccacag ggccccctt attaatgac ttctcgaaac tgtggataac 840

 cgtattaccg ccggcggigt tgacataaat accactggcg gtgatactga gcacatcagc 900
 aggtcacaca ggaaagtact agatgtcgca tcttcagaa ttagtagctt cagcgaaggc 960
 cgcgatttct caggcgagtg acgtcgcagc actggataat gtacgtgttg agtacctggg 1020
 aaagaaggga caccttactc ttcaaatgac aaccctgcgc gaactgccgc cggaggaacg 1080
 ccccgcgca ggagcggtaa tcaatgagc aaaggagcaa gtacaacagg cactgaacgc 1140
 ccgtaaggct gagttagaat ccgccgcat aaacgcgcgc cttgtcgcgg aaaccattga 1200
 tgtctcgtg cccgggcgcc gcattgagaa tggaggctta caccagtga ctctaccat 1260

 cgaccgtatc gaatctttct ttggcgaact tggcttcact gtggcaactg gaccggagat 1320
 tgaggacgac taccacaatt tcgatgcctt gaacattccc ggtcatcatc ctgcacgcgc 1380
 cgatcatgat acattctggt ttgataccac ccgtttgctt cgtaccaga caagcgggtg 1440
 ccaaatccgt acgatgaagg ctcagcaacc accgatccgt atcattgtc cagggcgcgt 1500
 glaccgtaac gattatgacc agacacatac accgatgttt caccaaatgg aagggttgat 1560
 tgtggatacg aatatctctt tcacgaatct gaagggcacc ttacatgatt tcttacgcaa 1620
 ctttttcgag gaggaccttc aaattcgctt tcgtccatcg tacttccctt ttgcagaacc 1680

 ttcggctgaa gtggatgtaa tggggaaaaa cgtaagtgg ctggaggttt taggttgcgg 1740
 gatggttcat ccaaatgtgc ttcgcaacgt cggcatcgac cccgaagtct acagtggatt 1800
 cggattcggg atgggaatgg aacgtctgac tatgcttcgt tacggcgtaa cggatttgcg 1860
 ctctttttt gagaacgac ttctgtttct gaagcaattc aaataagcat ttttagtacg 1920
 tgcaataacc actctggttt ttccagggtg gttttttgat gccctttttg gagtcttcaa 1980
 ctgctgcgtt atccccgat tctgtgtttc gtctagagcc ttttgtataa aaattaggga 2040
 ccttgcactg actgcccccc ggtgagtgag tgtgtgcgtg tggtcagaat tggttaattg 2100

 gttgtaacac tgaccctat ttgtttatit ttctaaatac attcaaatat gtatccgctc 2160
 atgagacaat aaccctgata aatgcttcaa taatattgaa aaaggaagaa tatgagtatt 2220
 caacatttcc gtgtgcctt tattcccttt ttgcggcat ttgccttcc tgtttttgct 2280
 caccagaaa cgctggtgaa agtaaaagat gctgaagatc agttgggtgc acgagtgggt 2340

tacatcgaac tggatctcaa cagcggtaag atccttgaga gttttcgccc cgaagaacgt	2400
tttccaatga tgagcacttt taaagtcttg ctatgtggcg cggattatc ccgtattgac	2460
gccgggcaag agcaactcgg tcgccgcata cactattctc agaatgactt ggttagtac	2520
tcaccagtca cagaaaagca tcttacggat ggcatgacag taagagaatt atgcagtgt	2580
gccataacca tgagtataa cactcgggcc aacttacttc tgacaacgat cggaggaccg	2640
aaggagctaa ccgtttttt gcacaacatg ggggatcatg taactcgcct tgatcgttgg	2700
gaaccggagc tgaatgaagc cataccaaac gacgagcgtg acaccacgat gcctgtagcg	2760
atggcaacaa cgttgcgcaa actattaact ggcgaaactac ttactctagc ttcccggcaa	2820
caattaatag actggatgga ggcggataaa gttgcaggac cacttctgcg ctcgccctt	2880
ccggctggct ggtttattgc tgataaatcc ggagccggtg agcgtggttc tcgcggtatc	2940
atgcgagcgc tggggccaga tggtaagccc tcccgatcg tagttatcta cagcagggg	3000
agtcaggcaa ctatggatga acgaaataga cagatcgtg agataggtgc ctactgatt	3060
aagcattggt aataatgaag cgatagcgcc ggcttagtca gatttaattc gcgcgcgtgg	3120
tggtatatt	3128
<210> 84	
<211> 2498	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Synthetically constructed plasmid 13000823784 pMAS006	
<400> 84	
acgaccgagc gcagcgagtc agtgagcgag gaagcgggaag agcgcccaat acgcaaaccg	60
cctctccccg cgcgttggcc gattcattaa tgcagctggc acgacaggtt tcccgactgg	120
aaagcgggca gtgagcgcaa cgcaattaat gtgagttagc tctactatta ggcaccccag	180
gctttacact ttatgcttcc ggctcgtatg ttgtgtggaa ttgtgagcgg ataacaattt	240
cacacaggaa acagctatga ccatgattac gccaaagctt catgcctgca ggtcgactct	300
agaggatccc cgcagatgta gtgtttccac agggtttcga gaagtcattt aataagcccc	360
cttgatcggg cagtaagag gttccaactt tcaccataat gaaataagat cactaccggg	420
cgtatTTTTT gagttatcga gattttcagg agctaaggaa gctaaaatgg agaaaaaat	480
cactggatat accacgttg atatatccca atggcatcgt aaagaacatt ttgaggcatt	540
tcagtcagtt gctcaatgta cctataacca gaccgttcag ctggatatta cggccttttt	600

aaagaccgta aagaaaaata agcacaagtt ttatccggcc tttattcaca ttcttgcccc	660
cctgatgaat gctcatccgg aatttcgtat ggcaatgaaa gacggtgagc tggatgatg	720
ggatagtgtt cacccttggt acaccgtttt ccatgagcaa actgaaacgt tttcatcgct	780
ctggagttaa taccacgacg atttccggca gtttctacac atatattcgc aagatgtggc	840
gtgttacggt gaaaacctgg cctatttccc taaagggttt attgagaata tgtttttcgt	900
ttcagccaat ccttgggtga gtttcaccag ttttgattta aacgtggcca atatggacaa	960
cttcttcgcc ccggttttca ccatgggcaa atattatacg caaggcgaca aggtgctgat	1020
gccgtggcg attcaggttc atcatgccgt ttgtgatggc ttccatgtcg gcagaatgct	1080
taatgaatta caacagtact gcgatgagtg gcaggcgagg gcgtaatgtg atatcgagct	1140
cgcttggact cctgttgata gatccagtaa tgacctcaga actccatctg gattttgtca	1200
gaacgctcgg ttcccgccgg gcgtttttta ttggtaaaaa atacaaaagg ctctagacga	1260
aaggtgagtg agtgtgtcgc tgtgggtcga ctctagagga tccccgggta ccgagctcga	1320
attcactggc cgtcgtttta caacgtcgtg actgggaaaa ccctggcgtt acccaactta	1380
atcgcttgc agcacatccc cttttcgcca gctggcgtaa tagcgaagag gcccgcaccg	1440
atcgcccttc ccaacagttg cgcagcctga atggcgaaatg gcgcctgatg cggatatttc	1500
tccttacgca tctgtcgggt atttcacacc gcataatggtg cactctcagt acaatctgct	1560
ctgatgccgc atagttaagc cagcccgac acccgccaac acccgctgac gcgcctgac	1620
gggcttgtct gtcgccgca tccgcttaca gacaagctgt gaccgtctcc gggagctgca	1680
tgtgtcagag gttttaccg tcataccga aacgcgcgag acgaaagggc ctcgtgatac	1740
gcctattttt ataggttaat gtcataataa taatggtttc ttagacgtca ggtggcactt	1800
ttcggggaaa tgtgcgcgga accctattt gtttattttt ctaaatacat tcaaataatgt	1860
atccgctcat gagacaataa ccctgataaa tgcctcaata atattgaaaa aggaagagta	1920
tgagtattca acatttcgt gtcgccctta ttcctttttt tgcggcattt tgccttcctg	1980
tttttgcct cccagaaacg ctgggtgaaag taaaagatgc tgaagatcag ttgggtgcac	2040
gagtgggtta catcgaactg gatctcaaca gcggtaagat ccttgagagt tttcgccccg	2100
aagaacgcc ttgcttgttg tcacaaccg ttaaacctta aaagctttaa aagccttata	2160
tattcttttt tttcttataa aacttaaaac cttagaggct atttaagttg ctgatttata	2220
ttaattttat tgttcaaca tgagagctta gtacgtgaaa catgagagct tagtacgtta	2280
gccatgaggg tttagtctgt tagccatgag ggtttagtgc gttaaacatg agagcttagt	2340
acgttaacaa tgagagctta gtacgtgaaa catgagagct tagtacgtac tatcaacagg	2400
ttgaactgct gatcttcttc tttctgcgt tatccctga ttctgtggat aaccgtatta	2460

ccgcctttga gtgagctgat accgctcgcc gcagccga 2498

<210> 85

<211> 11567

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetically constructed plasmid 13000283399

<400> 85

actacctagc tgcattttca ggaggaagcg atgggcggcc gcacaccttc tatgcggtgt 60

gaaataccgc catgaccctt aaatatcttg acaaatgctc tttccctaaa ctccccccat 120

aaaaaaaaacc gccgaagcgg gtttttacgt tatttgcgga ttaacgatta ctcgttatca 180

gaaccgcccc ggatgcctgg cagttccctc ctctcgccgc tgcgctcggt cgttcggctg 240

cgggacctca gcgctagcgg agtgtatact ggcttactat gttggcactg atgagggtgt 300

cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggctgcgtca gcagaatatg 360

tgatacagga tatattccgc ttctctgctc actgactcgc tacgctcggt cgttcgactg 420

cggcgagcgg aaatggctta cgaacggggc ggagatttcc tggaagatgc caggaagata 480

cttaacaggg aagttagagg gccgcggcaa agccgttttt ccataggctc cgccccctg 540

acaagcatca cgaaatctga cgctcaaatc agtgggtggcg aaaccgaca ggactataaa 600

gataccaggc gtttccccct ggcggtccc tcgtgcgctc tctgttcct gcctttcggt 660

ttaccggtgt cattccgctg ttatggccgc gtttgtctca ttccacgcct gacactcagt 720

tccgggtagg cagttcgctc caagctggac tgtatgcacg aacccccgt tcagtccgac 780

cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc cggaaagaca tgcaaaagca 840

ccactggcag cagccactgg taattgattt agaggagtta gtcttgaagt catgcgccgg 900

ttaaggctaa actgaaagga caagttttgg tgaactgcgt cctccaagcc agttacctcg 960

gttcaaagag ttggtagctc agagaacctt cgaaaaaccg ccctgcaagg cggttttttc 1020

gttttcagag caagagatta cgcgagacc aaaacgatct caagaagatc atcttattaa 1080

gcctcactca ttaggcaccc caggctttac actgataatg ggtgagttag tgtgtgcgtg 1140

tggggcgcgc cagatgggaa cagctagctt cacgctgccg caagcactca gggcgcaagg 1200

gctgctaaag gaagcggaac acgtagaaag ccagtcgca gaaacggtgc tgaccccgga 1260

tgaatgtcag ctactgggct atctggacaa gggaaaaacg aagcgcaaag agaaagcagg 1320

tagcttgcag tgggcttaca tggcgatagc tagactgggc ggttttatgg acagcaagcg 1380

aaccggaatt gccagctggg gcgcctcttg gtaaggttgg gaagccctgc aaagtaaact	1440
ggatggcttt ctgcccga aggatctgat ggcgagggg atcaagatct gatcaagaga	1500
caggatgagg atcgtttcgc atgattgaac aagatggatt gcacgcaggt tctccggccg	1560
cttgggtgga gaggctattc ggctatgact gggcacaaca gacaatcggc tgctctgatg	1620
ccgccgtgtt ccggctgtca gcgcaggggc gcccggttct tttgtcaag accgacctgt	1680
ccggtgccct gaatgaactc caagacgagg cagcgcggt atcgtggctg gccacgacgg	1740
gcgttccttg cgagctgtg ctgcagcttg tcaactgaagc gggaaggac tggctgctat	1800
tgggcgaagt gccggggcag gatctcctgt catctcacct tgctcctgcc gagaaagtat	1860
ccatcatggc tgatgcaatg cggcggtgc atacgttga tccggctacc tgccattcg	1920
accaccaagc gaaacatcg atcgagcag cagtgactcg gatggaagcc ggtcttctcg	1980
atcaggatga tctggacgaa gagcatcagg ggctcgcgcc agccgaactg ttcgccaggc	2040
tcaaggcgcg gatgcccgc gcgcaggatc tcgtcgtgac ccatggcgat gcctgcttgc	2100
cgaatatcat ggtggaaaat ggccgctttt ctggattcat cgactgtggc cggctgggtg	2160
tggcggaccg ctatcaggac atagcgttgg ctaccctga tattgctgaa gagcttggcg	2220
gcgaatgggc tgaccgttc ctctgcttt acggtatcg cgctcccgat tcgcagcgca	2280
tcgcttcta tcgcttctt gacgagttct tctgagcggg actctgggt tcgctagagg	2340
atcgatcctt ttaacctat cacatatacc tgccgttcac tattatttag tgaaatgaga	2400
tattatgata tttctgaat tgtgattaaa aaggcaactt tatgcccatg caacagaaac	2460
tataaaaaat acagagaatg aaaagaaaca gatagatttt ttagttcttt aggcccgtag	2520
tctgcaaatc cttttatgat tttctatcaa acaaaaggagg aaaatagacc agttgcaatc	2580
caaacgagag tctaatagaa tgaggtcgaa aagtaaatcg cgcggtttg ttactgataa	2640
agcaggcaag acctaaaatg tgtaaagggc aaagtgtata ctttggcgtc accccttaca	2700
tattttaggt cttttttat tgtgcgtaac taacttgcca tcttcaaaca ggagggtg	2760
aagaagcaga ccgctaacac agtacataaa aaaggagaca tgaacgactc cagtcttct	2820
agaagatggc aaacagctat tatgggtatt atgggtcccc gaagcagggt tatgcagcgg	2880
aaaagctccc cgaagggtc cacctgggtc cttttcatca cgtgctataa aaataattat	2940
aatttaaat ttttaataa aatatataaa ttaaaaatag aaagtaaaaa aagaaattaa	3000
agaaaaata gttttgttt tccgaagatg taaaagactc tagggggatc gccaacaaat	3060
actaccttt atcttgctct tctgctctc aggtattaat gccgaattgt ttcatttgt	3120
ctgtgtagaa gaccacacac gaaaatcctg tgattttaca ttttacttat cgttaatcga	3180
atgtatatct atttaactcg cttttctgt ctaataaata tataatgtaa gtacgtttt	3240

tggtgaaatt ttttaaacct ttgtttatatt tttttttctt cattccgtaa ctcttctacc	3300
ttctttatatt actttctaaa atccaaatac aaaacataaa aataaataaa cacagagtaa	3360
attcccaaat tattccatca ttaaaagata cgaggcgcgt gtaagttaca ggcaagcgt	3420
ccgtctaaga aaccattatt atcatgacat taacctataa aaataggcgt atcacgaggc	3480
cctttcgtct cgcgcgtttc ggtgatgacg gtgaaaacct ctgacacatg cagctcccgg	3540
agacggtcac agcttgtctg taagcggatg ccgggagcag acaagcccgt cagggcgcgt	3600
cagcgggtgt tggcgggtgt cggggctggc ttaactatgc ggcatcagag cagattgtac	3660
tgagagtga ccataccaca gccggaagag gtagtaggaa tattactggc tgaaaataag	3720
tcttgaatga acgtatacgc gtatatattt accaatctct caacactgag taatggtagt	3780
tataagaaag agaccgagtt agggacagtt agaggcgggt gagatattcc ttatggcatg	3840
tctggcgatg ataaaaattt tcaaacggca gccccgatct aaaagagctg acagggaaat	3900
ggtcagaaaa agaaacgtgc acccgcccgt ctggacgcgc cgctcacccg cacggcagag	3960
accaatcagt aaaaatcaac ggttaacgac attactatat atataatata ggaagcattt	4020
aatagaacag catcgtataa tatgtgtact ttgcagttat gacgccagat ggtagtagtg	4080
gaagatattc ttatttgaaa aatagcttgt caccttacgt acaatcttga tccggagctt	4140
ttcttttttt gccgattaag aattcggctg aaaaaagaaa aggagagggc caagagggag	4200
ggcattgggtg actattgagc acgtgagtat acgtgattaa gcacacaaaag gcagcttgga	4260
gtatgtctgt tattaatttc acaggtagtt ctggctcatt ggtgaaagt tgcggcttgc	4320
agagcacaga ggccgcagaa tgtgtcttag attccgatgc tgacttgctg ggtattatat	4380
gtgtgccc aa tagaaagaga acaattgacc cggttattgc aaggaaaatt tcaagcttg	4440
taaaagcata taaaatagt tcaggcactc cgaaatactt ggttggcgtg tttcgtaac	4500
aacctaaagga ggatgttttg gctctggtca atgattacgg cattgatac gtccaactgc	4560
atggagatga gtcgtggcaa gaataccaag agttcctcgg ttgcccagtt attaaaagac	4620
tcgtatttcc aaaagactgc aacatactac tcagtgcagc ttcacagaaa cctcattcgt	4680
ttattccctt gtttgattca gaagcagggt ggacagggtga acttttggat tggaactcga	4740
tttctgactg ggttggaagg caagagagcc ccgaaagctt acattttatg ttagctggtg	4800
gactgacgcc agaaaatgtt ggtgatgcgc ttagattaaa tggcgttatt ggtgttgatg	4860
taagcggagg tgggagaca aatggtgtaa aagactctaa caaaatagca aatttcgtca	4920
aaaatgctaa gaaataggtt attactgagt agtatttatt taagtattgt ttgtgcactt	4980

gcctgcaggc cttttgaaaa gcaagcataa aagatctaaa cataaaatct gtaaaataac 5040
aagatgtaaa gataatgcta aatcatttgg ctttttgatt gattgtacag gactgggtgg 5100
aatcccttct gcagcacctg gattaccctg ttatccctag taacaaaatt ctccagtctt 5160
cacatcggtt tgaaaggagg aagcggaaga atgaagtaag agggattttt gactccgaag 5220
taagtcttca aaaaatcaaa taaggagtgt caagaatgtt tgcaaaacga ttcaaaacct 5280
ctttactgcc gttattcgct ggatttttat tgcgttttca tttggttctg gcaggaccgg 5340
cggctgcgag tgctgaaacg gcgaacaaat cgaatgagct tacagcaccg tcgatcaaaa 5400

gcggaaccat tcttcatgca tggatttggc cgttcaatac gttaaaacac aatatgaagg 5460
atattcatga tgcaggatat acagccattc agacatctcc gattaaccaa gtaaaggaag 5520
ggaatcaagg agataaaagc atgtcgaact ggtactggct gtatcagccg acatcgtatc 5580
aaattggcaa ccgttactta ggtactgaac aagaatttaa agaaatgtgt gcagccgctg 5640
aagaatatgg cataaaggtc attgttgacg cggatcatca tcataccacc agtgattatg 5700
ccgcgatttc caatgaggtt aagagtattc caaactggac acatggaaac acacaaatta 5760
aaaactggtc tgatcgaaat agtacataat aatgagtagt cctcatctcc ctcaagcagg 5820

cgccggcggc actgccatct catgtttgac agcttatcat cggcaatagt tacccttatt 5880
atcaagataa gaaagaaaag gatttttcgc tacgtcaaaa tcctttaaaa aaacacaaaa 5940
gaccacattt ttaaatgtgg tctttattct tcaactaaag caccatttag ttcaacaaac 6000
gaaaattgga taaagtggga tatttttaaa atatatattt atgttacagt aatattgact 6060
tttaaaaaag gattgattct aatgaagaaa gcagacaagt aagcctcta aattcacttt 6120
agataaaaaa ttaggaggca tatcaaatga actttaataa aattgattta gacaattgga 6180
agagaaaaga gatatttaat cattatttga accaacaacac gacttttagt ataaccacag 6240

aaattgatat tagtgtttta taccgaaaca taaaacaaga aggatataaa ttttaccctg 6300
catttatctt cttagtgaac agggtgataa actcaaatac agcttttaga actggttaca 6360
atagcgacgg agagttaggt tattgggata agttagagcc actttataca atttttgatg 6420
gtgtatctaa aacattctct ggtatttggc ctctgtaaa gaatgacttc aaagagtttt 6480
atgatttata ctttctgat gtagagaaat ataatgggtc ggggaaattg tttcccaaaa 6540
cacctatacc tgaaatgct ttttctcttt ctattattcc atggacttca tttactgggt 6600
ttaacttaaa tatcaataat aatagtaatt accttctacc cattattaca gcaggaaaat 6660

tcattaataa aggttaattca atatatttac cgctatcttt acaggtacat cattctgttt 6720
gtgatggtta tcatgcagga ttgtttatga actctattca ggaattgtca gataggccta 6780
atgactggct tttataatat gagataatgc cgactgtact ttttacagtc ggttttctaa 6840

tgctactaac ctgccccgtt agttgaagaa ggtttttata ttacagctcc agatcctcta	6900
cgccggagcg atcgtggcat ataatctccc tcaagcaggc cccgctggcg cgcgcgaatg	6960
ttaggaaacg attagtcttt tgactgtttg acggtggtgg tactggggcc tataatgtga	7020
agagcttcac tgagtagggc ccgggctgta aacggttgaa ttcgcggcgc cttctagagg	7080
gagttctgag aattggtatg ctttataagt ccaattaaca gttgaaaacc tgcataggag	7140
agctatgcgg gttttttatt ttacataatg atacataatt taccgaaact tgcggaacat	7200
aattgaggaa tcatagaatt ttgtcaaat aattttattg acaacgtctt attaacgttg	7260
atataattta aattttattt gacaaaaatg ggctcgtgtt gtacaataaa tgtagttact	7320
agtagcggcc gctgcaggga tccacagtac ataaaaaagg agacatgacg atggtcgttt	7380
tacaacgtga ctgggtcgac cgggaaaacc ctggcggttac ccaacttaat cgccttcag	7440
cacatcccc tttcgccagc tggcgtaata gcgaagaggc ccgcaccgat cgcccttccc	7500
aacagttgcg cagcctgaat ggcgatggc gctttgcctg gtttccggca ccagaagcgg	7560
tgccgaaaag ctggctggag tgcgatcttc ctgaggccga tactgtcgtc gtccctcaa	7620
actggcagat gcacggttac gatgcgcca tctacacaa cgtaacctat cccattacgg	7680
tcaatccgcc gtttgttccc accgagaatc cgacgggttg ttactcgtc acatttaatg	7740
ttgatgaaag ctggctacag gaaggccaga cgcgaattat ttttgatggc gtttaactcg	7800
cgtttcatct gtggtgcaac gggcgctggg tcggttacgg ccaggacagt cgtttgccgt	7860
ctgaatttga cctgagcgca ttttacgcg ccggagaaaa ccgcctcgcg gtgatggtgc	7920
tgcgttggag tgacggcagt tatctggaag atcaggatat gtggcggatg agcggcattt	7980
tccgtgacgt ctggttctg cataaaccga ctacacaaat cagcgatttc catgttgcca	8040
ctcgttttaa tgatgatttc agccgcgtg tactggaggc tgaagttcag atgtgcggcg	8100
agttgcgtga ctacctagg gtaacagttt ctttatggca gggtgaaacg caggtcgcca	8160
gcggcaccgc gcctttcggc ggtgaaatta tcgatgagcg tgggtggttat gccgatcgcg	8220
tcacactacg tctgaacgtc gaaaaccga aactgtggag cgccgaaatc ccgaatctct	8280
atcgtgcggt ggttgaactg cacaccgccg acggcacgct gattgaagca gaagcctgcg	8340
atgtcggttt ccgcgaggtg cggattgaaa atggtctgct gctgctgaac ggcaagccgt	8400
tgctgattcg aggcgttaac cgctacgagc atcatctct gcattggtcag gtcattgatg	8460
agcagacgat ggtgcaggat atcctgctga tgaagcagaa caactttaac gccgtgcgt	8520
gttcgatta tccgaacat ccgctgtggt acacgctgtg cgaccgtac ggcctgtatg	8580
tgggtgatga agccaatatt gaaaccacg gcattggtgcc aatgaatcgt ctgaccgatg	8640
atccgcgctg gctaccggcg atgagcgaac gcgtaacgcg aatggtgcag cgcatcgta	8700

atcacccgag tgtgatcatc tggctgctgg ggaatgaatc aggccacggc gctaatacacg	8760
acgcgctgta tcgctggatc aaatctgtcg atccttcccg cccggtgcag tatgaaggcg	8820
gcggagccga caccacggcc accgatatta ttgcccgat gtacgcgcgc gtggatgaag	8880
accagccctt cccggctgtg ccgaaatggt ccatcaaaa atggctttcg ctacctggag	8940
agacgcgcc gctgatcctt tgcgaatacg cccacgcgat gggtaacagt ctggcggtt	9000
tcgctaaata ctggcaggcg ttctgtcagt atccccgtt acaggcgccg ttctgtctggg	9060
actgggtgga tcagtgcctg attaaatatg atgaaaacgg caaccctgg tcggcttacg	9120
gcggtgattt tggcgatacg ccgaacgatc gccagtcttg tatgaacggt ctggtctttg	9180
ccgaccgcac gccgcatcca gcgctgacgg aagcaaaaca ccagcagcag tttttccagt	9240
tccgtttatc cgggcaaacc atcgaagtga ccagcgaata cctgttccgt catagcgata	9300
acgagctcct gcactggatg ttggcgctgg atggttaagcc gctggcaagc ggtgaagtgc	9360
ctctggatgt cgctccaca ggtaaacagt tgattgaact gcctgaacta ccgcagccgg	9420
agagcgcgg gcaactctgg ctacacgtac gcgtagtga accgaacgcg accgcatggt	9480
cagaacccgg gcacatcagc gcctggcagc agtggcgtct ggcggaaaac ctcagtgtga	9540
cgctccccgc cgcgtcccac gccatcccc atctgaccac cagcgaaatg gatTTTTTgca	9600
tcgagctggg taataagcgt tggcaattta accgccagtc aggctttctt tcacagatgt	9660
ggattggcga taaaaaacaa ctgctgacgc cgtgcgcga tcagttcacc cgtgcaccgc	9720
tggataacga cattggcgta agtgaagcga cccgcattga ccctaacgcc tgggtcgaac	9780
gctggaaggc ggccggccat taccaggccg aagcagcgtt gttgcagtgc acggcagata	9840
cacttctgta tgcggtgctg attacgaccg ctacgcgtg gcagcatcag gggaaaacct	9900
tatttatcag ccggaaaacc taccggattg atggtagtgg tcaaatggcg attaccgttg	9960
atgttgaagt ggcgagcgat acaccgcac cggcgcggtt tggcctgaac tgccagctgg	10020
cgcaggtagc agagcgggta aactggctcg gattagggcc gcaagaaaac tatcccgacc	10080
gccttactgc cgcctgtttt gaccgtggg atctgccatt gtcagacatg tataccccgt	10140
acgtcttccc gagcgaaaac ggtctgcgt gcgggacgcg cgaattgaat tatggccac	10200
accagtggcg cggcgacttc cagttcaaca tcagccgcta cagtcaacag caactgatgg	10260
aaaccagcca tcgccatctg ctgcacgcgg aagaaggcac atggctgaat atcgacggtt	10320
tccatatggg gattggtggc gacgactcct ggagcccgtc agtatcgcg gaattaattc	10380
cagctgagcg ccggtcgcta ccattaccag ttggtctggt gtcaaaaata ataataaccg	10440

ggcaggccat gtctgccgt atttcgctga aggaaatcca attcgctagc agatgtagt 10500
 tttccacagg ggcgatcgct gatattgggtc gtcgatcgac atggatgagc gatgatgata 10560
 tccgttttagg ctgggcgggtg atagcttctc gttcaggcag tacgcctctt ttcttttcca 10620
 gacctgaggg aggcggaaat ggtgtgaggt tccccagatc tgggggaaaa gccaaatagg 10680
 cgatcgcggg agtgctttat ttgaagatca ggctatcact gcggtcaata gatttcacaa 10740
 tgtgatggct ggacagcctg aggaactctc gaacccgaat ggaaacaacc agatatttat 10800
 gaatcagcgc ggctcacatg gcgttgtgct ggcaaatgca ggttcatcct ctgtctctat 10860

caatacggca acaaaattgc ctgatggcag gtatgacaat aaagctggag cgggttcatt 10920
 tcaagtgaac gatggtaaac tgacaggcac gatcaatgcc aggtctgtag ctgtgcttta 10980
 tcctgatgat attgcaaaag cgcctcatgt tttccttgag aattacaaaa caggtgtaac 11040
 acattctttc aatgatcaac tgacgattac ctgtcggtgca gatcggaata caacaaaagc 11100
 cgtttatcaa atcaataatg gaccagacga caggcgctta aggatggaga tcaattcaca 11160
 atcggaaaag gagatccaat ttggcaaac atacaccatc atgttaaaag gaacgaacag 11220
 tgatggtgta acgaggaccg agaaatacag ttttgttaaa agagatccag cgtcggccaa 11280

aaccatcggc tatcaaaatc cgaatcattg gagccaggta aatgcttata tctataaaca 11340
 tgatgggagc cgagtaattg aattgaccgg atcttggcct ggaaaaccaa tgactaaaaa 11400
 tgcagacgga atttacacgc tgacgtgcc tgcggacacg gatacaacca acgcaaaagt 11460
 gatttttaat aatggcagcg cccaagtgcc cggtcagaat cagcctggct ttgattacgt 11520
 gctaaatggt ttatataatg actcgggctt aagcggttct cttcccc 11567

<210> 86

<211> 79

<212> DNA

<213> Artificial Sequence

<220>

<223> Chemically synthesized oligonucleotide GFP Cpf1 cTAG M fwd

<400> 86

cagcacctgg attaccctgt tatccctagt tttgggttaa agatgggttaa atgattcgaa 60
 aataataaag ggaaaatca 79

<210> 87

<211> 79

<212> DNA

<213> Artificial Sequence

<220>

<223> Chemically synthesized oligonucleotide GFP Cpf1 cTAG N fwd

<400> 87

cagcacctgg attaccctgt taccctagt ttgggatgt taaggtccc tatcttcgaa 60

aataataaag gaaaatca 79

<210> 88

<211> 81

<212> DNA

<213> Artificial Sequence

<220>

<223> Chemically synthesized oligonucleotide GFP Cpf1 cTAG P fwd

<400> 88

cagcacctgg attaccctgt taccctagt ttggagggt gttcagctc cgtgaactcg 60

aaaataataa agggaaaatc a 81

<210> 89

<211> 79

<212> DNA

<213> Artificial Sequence

<220>

<223> Chemically synthesized oligonucleotide GFP Cpf1 cTAG O rvs

<400> 89

cgcttcctcc tgaatatgca gctaggtagt ttgaccgcc ccccccatac cccaatcgac 60

atgccgaact cagaagtga 79

<210> 90

<211> 79

<212> DNA

<213> Artificial Sequence

<220>

<223> Chemically synthesized oligonucleotide GFP Cpf1 cTAG N rvs

<400> 90

cgcttcctcc tgaatatgca gctaggtagt ttgggatgt taaggtccc tatcttcgac 60

atgccgaact cagaagtga 79

<210> 91

<211> 51
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Chemically synthesized oligonucleotide CAT01 Cpf1 cTAG M fwd
 <400> 91
 tttgggttaa agatgggttaa atgattcgac atacacataa agtagcttgc g 51

<210> 92
 <211> 51
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Chemically synthesized oligonucleotide CAT01 Cpf1 cTAG N fwd
 <400> 92
 tttgggatgt taagagtccc tatcttcgac atacacataa agtagcttgc g 51

<210> 93
 <211> 53
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Chemically synthesized oligonucleotide CAT01 Cpf1 cTAG P fwd
 <400> 93
 tttgaggagt gttcagtcctc cgtgaactcg acatacacat aaagtagctt gcg 53

<210> 94
 <211> 49
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Chemically synthesized oligonucleotide CAT01 Cpf1 cTAG N rvs
 <400> 94
 tttgggatgt taagagtccc tatcttcgac tggaaggaca aggggggacc 49

<210> 95
 <211> 49

<212> DNA
 <213> Artificial Sequence
 <220>
 <223> Chemically synthesized oligonucleotide CAT01 Cpf1 cTAG 0 rvs
 <400> 95
 tttagaccgcc cccccatac cccaatcgac tggaaggaca aggggggacc 49
 <210> 96
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Chemically synthesized oligonucleotide Cpf1 cTAG M
 <400> 96
 tttaggggtaa agatgggtaa atgat 25
 <210> 97
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Chemically synthesized oligonucleotide Cpf1 cTAG N
 <400> 97
 tttagggatgt taagagtccc tatct 25
 <210> 98
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Chemically synthesized oligonucleotide Cpf1 cTAG 0
 <400> 98
 tttagaccgcc cccccatac cccaa 25
 <210> 99
 <211> 27
 <212> DNA

<213> Artificial Sequence

<220>

<223> Chemically synthesized oligonucleotide Cpf1 cTAG P

<400> 99

tttgaggagt gttcagtctc cgtgaac 27

<210> 100

<211> 5093

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetically constructed plasmid pZHR039 (Kanamycin resistance, GFP)

<400> 100

actacctagc tgcattttca ggaggaagcg atgggcggcc gcacaccttc tatgcggtgt 60

gaaataccgc catgacaaa atcccttaac gtgagttttc gttccactga gcgtcagacc 120

ccgtagaaaa gatcaaagga tcttcttgag atcctttttt tctgcgcgta atctgctgct 180

tgcaaacaaa aaaaccaccg ctaccagcgg tggtttgttt gccggatcaa gagctaccaa 240

ctcttttttc gaaggttaact ggcttcagca gagcgcagat accaaatact gttcttctag 300

tgtagccgta gttaggccac cacttcaaga actctgtagc accgcctaca tacctcgctc 360

tgctaactct gttaccagtg gctgctgcca gtggcgataa gtcgtgtctt accgggttgg 420

actcaagacg atagttaccg gataaggcgc agcggtcggg ctgaacgggg ggttcgtgca 480

cacagcccag ctiggagcga acgacctaca ccgaactgag atacctacag cgtgagctat 540

gagaaagcgc cagccttccc gaaggagaaa aggcggacag gtatccggtg agcggcaggg 600

tcggaacagg agagcgcacg agggagcttc cagggggaaa cgcctggtat ctttatagtc 660

ctgtcgggtt tcgccacctc tgacttgagc gtcgattttt gtgatgctcg tcaggggggc 720

ggagcctatg gaaaaacgcc agcaacgcgg cttttttacg gttcctggcc ttttgcctggc 780

cttttgctca catgttcttt cctgcgttat ccctgattc tgtggataac cgtattaccg 840

cctttgagtg agctgatacc gctcgccgca gccgaacgac cgagcgcagc gagtcatgta 900

gcgaggaagc ggaagagcgc ccaatacgca aaccgcctct ccccgcgctt tggccgattc 960

attaatgcag ctggcacgac aggtttcccg actggaaagc gggcagtgag cgcaacgcaa 1020

ttaatgtgag ttagctcact cattagggcac ccaggtctt acactgataa tgggtgagtg 1080

agtgtgtgcg tgtggggcgc gccagatggg aacagctagc ttcacgctgc cgcaagcact 1140

cagggcgcaa gggctgctaa aggaagcggg acacgtagaa agccagtccg cagaaacggt 1200

gctgaccccg gatgaatgtc agctactggg ctatctggac aagggaac gcaagcgcaa 1260
agagaaagca ggtagcttgc agtgggctta catggcgata gctagactgg gcggttttat 1320
ggacagcaag cgaaccggaa ttgccagctg gggcgccctc tggtaagggtt gggaagccct 1380

gcaaagtaaa ctggatggct ttcttgccgc caaggatctg atggcgagg ggatcaagat 1440
ctgatcaaga gacaggatga ggatcgtttc gcatgattga acaagatgga ttgcacgcag 1500
gttctccggc cgcttgggtg gagaggctat tcggctatga ctgggcacaa cagacaatcg 1560
gctgctctga tgccgccgtg ttccggctgt cagcgagggg gcgcccgtt cttttgtca 1620
agaccgacct gtccggtgcc ctgaatgaac tccaagacga ggcagcgagg ctatcgagg 1680
tgccacgac gggcggttct tgccgagctg tgctcgagct tgtcactgaa gcgggaaggg 1740
actggctgct attgggcgaa gtgccggggc aggatctct gtcactcac cttgctctg 1800

ccgagaaagt atccatcatg gctgatgcaa tgcggcggct gcatacgtt gatccggcta 1860
cctgcccatt cgaccacaa gcgaacatc gcatcgagcg agcacgtact cggatggaag 1920
ccggtcttgt cgatcaggat gatctggacg aagagcatca ggggctcgcg ccagccgaac 1980
tgttcgcccag gctcaaggcg cggatgcccg acggcgagga tctcgtcgtg acctatggcg 2040
atgcctgctt gccgaatata atggtggaaa atggccgctt ttctggattc atcgactgtg 2100
gccggctggg tgtggcgagc cgctatcagg acatagcgtt ggctaccgt gatattgctg 2160
aagagcttgg cggcgaatgg gctgaccgt tctcgtgct ttacggatc gccgctcccg 2220

attcgagcg catcgcttc tategcttc ttgacgagtt cttctgagcg ggactctggg 2280
gttcgctaga ggatcgatcc ttttaaccc atcacatata cctgccgttc actattattt 2340
agtgaatga gatattatga tattttctga attgtgatta aaaaggcaac ttatgccca 2400
tgcaacagaa actataaaaa atacagagaa tgaaaagaaa cagatagatt ttttagttct 2460
ttaggccgt agtctgcaaa tccttttatg attttctatc aaacaaaaga ggaaaataga 2520
ccagttgcaa tccaaacgag agtctaatag aatgaggtcg aaaagtaaat cgcgcgggtt 2580
tgttactgat aaagcaggca agacctaaaa tgtgtaaagg gcaaagtgt tactttggcg 2640

tcacccctta catatttttag gtcttttttt attgtgcgt actaacttgc catcttcaaa 2700
caggagggtt ggaagaagca gaccgtaac acagtacata aaaaaggaga catgaacgac 2760
tccagctttt ctagaagatg gcaaacagct attatgggta ttatgggtcc ccgaagcagg 2820
gttatgcagc ggaaaagctc ccgaaaagt gccacctggg tccttttcat cacgtgctat 2880
aaaaataatt ataattttta ttttttaata taaatatata aattaaaaat agaaagtaaa 2940
aaaagaaatt aaagaaaaaa tagtttttgt tttccgaaga tgtaaaagac tctaggggga 3000

tcgccaacaa atactacctt ttatcttgct cttcctgctc tcaggtatta atgccgaatt	3060
gtttcatctt gtctgtgtag aagaccacac acgaaaatcc tgtgatttta cattttactt	3120
atcgtaatac gaatgtatat ctatttaac tgcctttctt gtctaataaa tatatatgta	3180
aagtacgctt ttigtgtgaa ttttttaaac ctttgtttat tttttttct tcattccgta	3240
actctttctac cttctttatt tacttttctaa aatccaaata caaaacataa aaataaataa	3300
acacagagta aattcccaaa ttattccatc attaaaagat acgaggcgcg tgtaagttac	3360
aggcaagcga tccgtcctaa gaaaccatta ttatcatgac attaacctat aaaaataggc	3420
gtatcacgag gccctttcgt ctgcgcgctt tcggtgatga cggtgaaaac ctctgacaca	3480
tgcagctccc ggagacggtc acagcttgic tgtaagcgga tgccgggagc agacaagccc	3540
gtcagggcgc gtcagcgggt gttggcgggt gtcggggctg gcttaactat gcggcatcag	3600
agcagattgt actgagagtg caccatacca cagccggaag aggagtaggg aatattactg	3660
gctgaaaata agtcttgaat gaacgtatac gcgtatatat ctaccaatct ctcaacactg	3720
agtaatggta gttataagaa agagaccgag ttagggacag ttagaggcgg tggagatatt	3780
ccttatggca tgtctggcga tgataaaaact tttcaaacgg cagccccgat ctaaaagagc	3840
tgacagggaa atggtcagaa aaagaaacgt gcacccgccc gtcctggacgc gccgctcacc	3900
cgcacggcag agaccaatca gtaaaaatca acggttaacg acattactat atataataa	3960
taggaagcat ttaatagaac agcatcgtaa tatatgtgta ctttgcagtt atgacgccag	4020
atggcagtag tggaagatat tctttattga aaaatagctt gtcaccttac gtacaatctt	4080
gatccggagc ttttcttttt ttgccgatta agaattcggc cgaaaaaaga aaaggagagg	4140
gccaaagagg agggcattgg tgactattga gcacgtgagt atacgtgatt aagcacacaa	4200
aggcagcttg gagtatgtct gttatttaatt tcacagtag ttctgggtcca ttggtgaaag	4260
tttgcggctt gcagagcaca gaggccgcag aatgtgctct agattccgat gctgacttgc	4320
tgggtattat atgtgtgccc aatagaaaga gaacaattga cccggttatt gcaaggaaaa	4380
tttcaagtct tgtaaaagca tataaaaata gttcaggcac tccgaaatac ttggttggcg	4440
tgtttcgtaa tcaacctaaag gaggatgttt tggtcttggt caatgattac ggcatigata	4500
tcgtccaact gcatggagat gagtctggc aagaatacca agagttctc ggtttgccag	4560
ttattaaaag actcgtattt ccaaaaagact gcaacatact actcagtga gcttcacaga	4620
aacctcattc gtttatccccc ttgtttgatt cagaagcagg tgggacaggt gaacttttgg	4680
attggaactc gatttctgac tgggttggaa ggcaagagag ccccgaaagc ttacatttta	4740

tgtttagctgg tggactgacg ccagaaaatg ttggtgatgc gcttagatta aatggcggtta 4800
 ttggtgttga tgtaagcgga ggtgtggaga caaatggtgt aaaagactct aacaaaatag 4860
 caaatttcgt caaaaatgct aagaaaatagg ttattactga gtagtattta ttttaagtatt 4920
 gttttgacac ttgcctgcag gccttttgaa aagcaagcat aaaagatcta aacataaaat 4980
 ctgtaaaata acaagatgta aagataatgc taaatcattt ggctttttga ttgattgtac 5040
 aggactgggt ggaatccctt ctgcagcacc tggattaccc tgttatccct agt 5093

<210> 101

<211> 13737

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetically constructed plasmid 13000223370 Cat01

(Chloramphenicol source)

<400> 101

gaatcttacc ccatggctaa gaaagtagac acctcgaacg ctacccccgc tctagccctt 60
 ctacggaga ggcaaatcc ttttgagctg gatgttcacg atgtagatcc aaaatcatca 120
 aagggtctcg cattggatgc ctctgaagta atgggtgtgg agccggaagt ggtgtttaaa 180
 acgctcatgg cagatattga tggatgaacac gtgggtcgcga ttgttccagc cagcagcacg 240
 ttgaatctca agcagttggc taaggctgga aaaggttaagc atgcaaacat gatggatcgc 300
 agccgtgcac aggtagtcac ggggtatgtc cctgggtggaa tctcacccgat agggcagaag 360

aataagcacc gcgtgttttt ggatgagtct gcaattctcc aggagcgaat ctacgtcagc 420
 gcaggacgcc gaggttggtc actgattatc gccccgatg atgtttctct ggctaccgat 480
 ggggtttacg cggatattgc tgatcattca taaaaggtaa aaccaccccc cgcaggggtg 540
 cgtcgtaaag caagctcctt ttcttagtac ggcgcgatag tggactcgtg cgcgtcgatg 600
 ttgatcgccg cgtgcacgga cgggaatgcg cgtcgcgcgc agttctcgcg ggtgcacact 660
 cggcagccgg acccgatggg ggtggcgggt gataggtcct ggaggttgaa gccgcgggag 720
 tacacggtgc ggtcggcgtg gcgcgcttcg cagccgagcc cgatggcgaa cttttgtcc 780

acttcgccga accgggcttc gtgggtgcgc acggtgcgtg agatccacag gtagttgcgt 840
 ccgtcgggca tttgcgcgaa ttggcggagc acttggccgg ggttgggtgaa ggtttcaaac 900
 acgttccaca gggggcaggt gccgcgtag tgggtgaagt ggaagccggt ggcggattgg 960
 cgttttgaca tgttgccggc gcggtcgacg cgcacgaagg taaacgggat gccgcgcagg 1020
 ttgggacgct gcagggtgga caggcgggtg gcggtcgtct cgtagccac gccgaaaagc 1080

tgacctaggt attcgatgtc gtagccggat ttttcggcct cggagtggaa gatttttag	1140
ggcagcatta cggcggcgcc gaagtaggag gcgacgccgc ggatggctag ggtgcgggct	1200
tcgggggtgg accagatgcc gtcgtcgacg atgccttcga tgaggctgtt ggcttctagg	1260
tagccgagtt cggtagccat gcggaaggcg cgttgtccgg ggttaggcg tgcgtggatt	1320
gtcagtaggc gcgtctcggg gtcaaagtgg tgcagcgtgc cggattcctc tttggaggag	1380
gtgatggta catcgtgac catttgcagg cgcctggcga tcgaatctc catggcgagg	1440
gaatcgtacg gctgccagcc cagttgcgcg gcgatggctt cggcgcggcg gtcgagcgca	1500
tcgaagtagt tttggcgggc gtaaatgaaa tcgcgcacct cttcgtgcgg catgctcacg	1560
gcctccgca tggggcgacg ttcctcaggc gtgttggtgc gattgtccac tgcgatggag	1620
aacttgtcgc gtacgtttcg gtagcgctgg tgcatttcca ccatggcgcg gccagctgc	1680
gggtggttgt acaccatctc cगतagctct tggagctcca cgttcgcggg gttgatctcc	1740
cggccaaca tgacatctg aatttcagct aaaaggcggg aatcatctc gcgggagaaa	1800
aacgttgcgt ctacccgaa cgcctcgggt atgcgcaata acaccgttac ggtgagcggg	1860
cgtacgtcgt gctcaatctg atttacataa ctggcagata agccaagggt tgctgccaac	1920
gacgcctggc tcaggtctct ttcgcggcgc agttggcgca gcctggacct cacatatgtc	1980
tttcccatgg cgcaactata gcgtgatcac catcacctta accactcttg gcattgaggt	2040
gttgcaaac tcagaggcta aagaacaagg gcttactgtg cgcagagggt ccctcaccgg	2100
ggtcaaaaat ttgacatccg gtgagggtaa taccagcaca gcctagtctc gccagaatgc	2160
tggtcagcat acgaaagagc ttaaggcagg ccaattcgca ctgtcagggt cacttgggtg	2220
tttagcacta ccgacaggta cgctagtatg cgttcttct accagaggtc tgtggccgcg	2280
tggtcaaaag tgcggcttct gattttgtg ctctgtttta ctctcacact tagctttgac	2340
ctgcacaaat agttgcaaat tgtccacat acacataaag tagcttgcgt atttaaaatt	2400
atgaacctaa ggggttttag acttcacgt gccgcaagca ctcagggcgc aagggtgct	2460
aaaggaagcg gaacacgtag aaagccagtc cgcagaaacg gtgctgacct cggatgaatg	2520
tcagctactg ggctatctgg acaagggaac acgcaagcgc aaagagaaag caggtagctt	2580
gcagtgggct tacatggcga tagctagact gggcggtttt atggacagca agcgaaccgg	2640
aattgccagc tggggcgccc tctggttaagg ttgggaagcc ctgcaaagta aactggatgg	2700
ctttcttgcc gccaaaggatc tgatggcgca ggggatcaag atctgatcaa gagacaggat	2760
gaggatcgtt tcgcatggag aaaaagatca cgggctacac taccgtggac atctcgcaat	2820

ggcatcgcaa ggaacacttc gaggcatttc aaagcgtggc acaatgtact tacaaccaga 2880

ccgtccagct ggatattacc gcgtttttga agaccgttaa gaaaaacaag cacaagtttt 2940

atccagcctt tatccatatt ctgccccgt tgaatgaatgc gcaccccgaa tttcgtatgg 3000

ccatgaaaga tggtagagctc gttatctggg actcagtcga tccatgctat accgttttcc 3060

acgaacaacac tgaactttt tcttcgctgt ggtccgaata tcacgatgat ttcgcccaat 3120

ttttgcatat ctacagccaa gatgtcgcgt gctatggatga aaacctggct tactttccca 3180

agggatcat cgagaatatg tttttgttt cagcaaaccc ctgggtgtcc ttcacgtcgt 3240

ttgacttgaa cgtggccaat atggataatt tcttcgctcc agttttcacc atgggtaagt 3300

actataccca aggagacaag gtccttatgc cacttgcaat ccaagtacac cacgcagtct 3360

gcgatggttt ccatgtggga cgcatgctta acgaactcca acagtactgt gatgaatggc 3420

aaggcggcgc gtagccccc aaccgaagt tgggggattt ttgaatcctc ggtccccctt 3480

gtccttccag attgatcgac gcgttgttga ttttcgcttt tcgacgcagc ccgccgcat 3540

cgggtgcccg gcgtggtcag gccacatgcg ccccggaac tttttggga cctacgggtc 3600

aacagtgcg aaaaattgtgt cacctgcgca aagccttgct tctattcggg aaattcgggt 3660

gtctaaactt ttgggtgat accaaacggg gttagaaact gttcagatcg gtatcctgtg 3720

aggaagctca ccttggtttt agaattgtga aaaagcctca cgtttccgca ggtagagcac 3780

actcaattaa atgagcgta aacgacaata aagtaaggct accctaataa ctggggtttt 3840

atgcctctaa acagtcatgt gggggcggta ggggagcgtc ccatgactgg ttaatgcctc 3900

gatctgggac gtacagtaac agcgacactg gaggtgcat gactgttaga aatcccgacc 3960

gtgaggcaat ccgtcacgga aaaattacga cggaggcgt gcgtgagcgt cccgcatacc 4020

cgacctgggc aatgaagctg accatggcca tcatggcct aatcttcggt ggcttcgttc 4080

ttgttcacat gatcggaac ctgaaaatct tcatgccgga ctacgcagcc gattctgcgc 4140

atccgggtga agcacaagta gatgtctacg gcgagttcct gcgtgagatc ggatccccga 4200

tcctcccaca cggtcagtc ctctggatcc tacgtattat cctgctggtc gcattggttc 4260

tgcacatcta ctgtgcatc gcattgaccg gccgtttca ccagtcgccg ggaaagtcc 4320

gccgtaccaa cctcgttggc ggcttcaact ccttcgcgac ccgtccatg ctggtgaccg 4380

gaatcgttct ccttgcgttc attatcttcc acatctcga cctgacctg ggtgttgctc 4440

cagcagcccc aacttcattc gagcacggcg aagtatacgc aaatatgggt gcttccttta 4500

gccgctggcc ttagcaatt tggtagatca ttccaacct ggtcctgttc gtccacctgt 4560

ctcacggcat ctggcttgca gtctctgacc tgggaatcac cggacgtcgc tggagggcaa	4620
tcctcctcgc agttgcgtac atcggttcctg cactgggtcct gatcggcaac atcaccattc	4680
cgttcgccat cgtgtttggc tggattgcgt aaaggttagg aagaatttat gagcactcac	4740
tctgaaacca cccgcccaga gtatcatccac ccagtctcag tcctcccaga ggtctcagct	4800
ggtacggtec ttgacgtcgc tgagccagca ggcgttccca ccaaagatat gtgggaatac	4860
caaaaagacc acatgaacct ggtctcccca ctgaaccgac gcaagttccg cgtcctcgtc	4920
gttggcaccg gcctgtccgg tggcgtgca gcagcagccc tcggcgaact cggatacgac	4980
gtcaaggcgt tcacctacca cgacgcacct cgcgtgcgc actccattgc tgcacagggt	5040
ggcgttaact ccgcccgcgg caagaaggta gacaacgacg gcgcataccg ccacgtcaag	5100
gacaccgtca agggcggcga ctaccgtggc cgcgagtcgc actgctggcg tctcgccgtc	5160
gagtcggtcc gcgtcatcga ccacatgaac gccatcggtg caccattcgc ccgcgaatac	5220
ggtggcgcct tggcaaccgg ttccttcggt ggtgtgcagg tctccgtac ctactacacc	5280
cgtggacaaa ccggacagca gctgcagctc tccaccgat ccgcactaca gcgccagatc	5340
cacctcggct ccgtagagat cttcacccat aacgaaatgg ttgacgtaat tgtaccgaa	5400
cgtaatggtg aaaagcgtg cgaaggcctg atcatgcga accgatcac cggcgagctc	5460
accgcacaca ccggccatgc cgttatcctg gcaaccggtg gctacggcaa cgtgtaccac	5520
atgtccacce tggcgaagaa ctccaacgcc tctagggata acagggtaat acgtcgtgac	5580
tgggaaaacc ctggcgttac ccaacttaat cgccttgacg cacatcccc tttcgccagc	5640
tggcgtaata gcgaagagc ccgcaccgat cgccttccc aacagttgcg cagcctgaat	5700
ggcgaatggc gataagctag cttcacgctg ccgaagcac tcagggcgca agggctgcta	5760
aaggaagcgg aacacgtaga aagccagtcc gcagaaacgg tgctgacccc ggatgaatgt	5820
cagctactgg gctatctgga caagggaata cgcaagcgca aagagaaagc aggtagcttg	5880
cagtgggctt acatggcgat agctagactg ggcggtttta tggacagcaa gcgaaccgga	5940
attgccagct ggggcgcctt ctggttaaggt tgggaagccc tgcaaagtaa actggatggc	6000
tttcttgccg ccaaggatct gatggcgcag gggatcaaga tctgatcaag agacaggatg	6060
aggatcgttt cgcatgattg aacaagatgg attgcacgca ggttctccgg ccgcttgggt	6120
ggagaggcta ttcggctatg actgggcaca acagacaatc ggctgctctg atgccccgt	6180
gttccggctg tcagcgcagg ggcgcccgtt tctttttgtc aagaccgacc tgtccggtgc	6240
cctgaatgaa ctccaagacg aggcagcgcg gctatcgtgg ctggccacga cgggcgttcc	6300
ttgcgcagct gtgctcgacg ttgtcactga agcgggaagg gactggctgc tattgggcga	6360
agtgcggggg caggatctcc tgtcatctca ccttgctcct gccgagaaag tatccatcat	6420

ggctgatgca atgcggcggc tgcatacgt tgatccggct acctgccc atcgaccacca 6480
 agcgaaacat cgcacgagc gagcacgtac tcggatggaa gccggtcttg tcgatcagga 6540
 tgatctggac gaagagcatc aggggctcgc gccagccgaa ctgttcgcca ggctcaaggc 6600
 gcggatgccc gacggcgagg atctcgtcgt gacccatggc gatgcctgct tgccgaatat 6660

 catggatggaa aatggccgct tttctggatt catcgactgt ggccggctgg gtgtggcgga 6720
 ccgctatcag gacatagcgt tggctaccgc tgatattgct gaagagcttg gcggcgaatg 6780
 ggctgaccgc ttctctgtgc ttacgggtat cgcgcgtccc gattcgcagc gcatcgcctt 6840
 ctatcgcctt cttgacgagt tctctgagc gggactctgg ggttcgctag aggatcgatc 6900
 ctttttaacc catcacatat acctgccgtt cactattatt tagtgaaatg agatattatg 6960
 atattttctg aattgtgatt aaaaaggcaa ctttatgccc atgcaacaga aactataaaa 7020
 aatacagaga atgaaaagaa acagatagat tttttagtgc tttaggcccg tagtctgcaa 7080

 atccttttat gatcttctat caaacaagaagg aggaataatg accagtgtga atccaacga 7140
 gagtctaata gaatgaggtc gaaaagtaaa tcgcgcgggt ttgttactga taaagcaggc 7200
 aagacctaaa atgtgtaagg ggcaaatgt atactttggc gtcacccctt acatatttta 7260
 ggtctttttt tattgtgcgt aactaacttg ccactttcaa acaggagggc tggaagaagc 7320
 agaccgctaa cacagtacat aaaaaggag acatgaacgt gagctgttta caattaatca 7380
 tcgtgtggta ccatgtgtgg aattggaaag gacatgaaca tcaaaaagt ttgcaaaaca 7440
 gcaacagtat taacctttac taccgcactg ctggcaggag gcgcaactca agcgtttgcg 7500

 aaagaaacga accaaaagcc atataaggaa acatacggca tttcccatat tacacgcat 7560
 gatatgctgc aaatccctga acagcaaaaa aatgaaaaat atcaagtttc tgaattgat 7620
 tcgtccacaa ttaaaaatat ctctctgca aaaggcctgg acgtttggga cagctggcca 7680
 ttacaaaacg ctgacggcac tgctgcaaac taccacggct accacatcgt ctttgcat 7740
 gccggagatc ctaaaaatgc ggatgacaca tcgatttaca tgttctatca aaaagtcggc 7800
 gaaacttcta ttgacagctg gaaaaacgct ggccgcgtct ttaagacag cgacaaatc 7860
 gatgcaaatg attctatcct aaaagaccaa acacaagaat ggtcaggttc agccacattt 7920

 acatctgacg gaaaaatccg tttattctac actgatttct ccggtaaaca ttacggcaaa 7980
 caaactga caactgcaca agttaacgta tcagatcag acagctcttt gaacatcaac 8040
 ggtgttagagg attataaatc aatctttgac ggtgacggaa aaacgtatca aaatgtacag 8100
 cagttcatcg atgaaggcaa ctacagctca ggcgacaacc atacgtgag agatcctcac 8160
 tacgtagaag ataaaggcca caaatactta gtatttgaag caaactgg aactgaagat 8220
 ggctaccaag gcgaagaatc tttatttaac aaagcactat atggcaaaag cacatcattc 8280

ttccgtcaag aaagtcaaaa acttctgcaa agcgataaaa aacgcacggc tgagttagca	8340
aacggcgctc tcggtatgat tgagctaaac gatgattaca cactgaaaaa agtgatgaaa	8400
ccgctgattg catctaacac agtaacagat gaaattgaac gcgcgaacgt ctttaaaatg	8460
aacggcaaat ggtacctgtt cactgactcc cgcggatcaa aaatgacgat tgacggcatt	8520
acgtctaacg atattttacat gcttggttat gtttctaatt ctttaactgg cccatacaag	8580
ccgctgaaca aaactggcct tgtgttaaaa atggatcttg atcctaacga tgtaaccttt	8640
acttactcac acttcgctgt acctcaagcg aaaggaaaca atgtcgtgat tacaagctat	8700
atgacaaaca gaggattcta cgcagacaaa caatcaacgt ttgcgccgag cttcctgctg	8760
aacatcaaag gcaagaaaac atctgttgc aaagacagca tccttgaaca aggacaatta	8820
acagttaaca aataaaaacg caaaagaaaa tgccgatggg taccgagcga aatgaccgac	8880
caagcgacgc ccaacctgcc atcacgagat ttcgattcca ccgccgcctt ctatgaaagg	8940
ttgggcttcg gaatcgtttt ccgggacgcc ctgcggacg tgctcatagt ccaataacta	9000
cattgagcga aatgccaaac acatgtcca tgcttttact aatgtggggt cttagaagaa	9060
agcgaccaat ttaaggagag ttgaatatgt ctgaaatcca gctgacggag gcatcattga	9120
acgaagcagc cgatgctgca attaaagcgt tcgatggagc aaaaaacctc gatgaattgg	9180
ctgctctgcg acgagatcac ctgggtgatg cggcaccaat ccctcaggca cgccgctcgc	9240
ttggaacctat tcaaaaagat cagcgttaagg atgccggacg attcgtaaac atggcgctgg	9300
gccgcgcgga aaagcacttc gccaggtta aggtggttct tgaagaaaag cgaaacgcag	9360
aagtcttgga gctggagcgc gtggatgtta ccgtccctac cacacgtgaa caagtcggcg	9420
cactgcacc aattacgatt ctcaacgaac agatcgcgga catctttgtt ggcatgggct	9480
gggagatcgc agagggcccg gaagttgaag ccgaatactt caatttcgat gcacttaact	9540
ttctcccaga ccacccagcc cgcacctgc aggatactt ccacatcgca cctgaaggat	9600
cgcgccaagt gttgcgcacc catacctctc ctgtccagggt tcgcacgatg ctgaatcgag	9660
aggtacctat ctatatcgcc tgctctgttc gcgtcttcg cactgacgaa ttggatgcta	9720
cccacacccc tgcttttcac cagatcgagg gcctggctgt cgacaaaggc ctgacaatgg	9780
ctcaccttcg cggaactctg gatcacttgg ctaaagaact gttcggaact gagactaaaa	9840
cccgcatcgc ttcaaaactc tteccatttt ctgagcccag cgcggaagtt gatgtctggt	9900
tcccaaataa gaagggcggt gccggctgga tcgaatgggg cggtctcggc atggtcaacc	9960
caaacgtgct ccgcgtgta ggcgtcgacc cggaagaata cactggattc ggcttcggtta	10020

tgggtattga acgcaccttg caattccgaa atggactctc agatatgcmc gatattgtag	10080
aaggcgacat tgcctttacc ctccctttcg gcattcaggc ttaggcattt ttagtacgtg	10140
caataaccac tctggttttt ccagggtggt tttttgatgc cttttttgga gtcttcaact	10200
gagcctcgca gagcaggatt cccgttgagc accgccaggc gcgaataagg gacagtgaag	10260
aaggaacacc cgtcgcggg tgggcctact tcacctatcc tgcccgctg acgccgttgg	10320
atacaccaag gaaagtctac acgaacctt tggcaaaatc ctgtatatcg tgcgaaaaag	10380
gatggatata ccgaaaaaat cgctataatg accccgaagc agggttatgc agcggaaaag	10440
ctccccgaaa agtgccacct gggctctttt catcacgtgc tataaaaata attataattt	10500
aaatttttta atataaatat ataaattaaa aatagaaagt aaaaaagaa attaaagaaa	10560
aaatagtttt tgitttccga agatgtaaaa gactctaggc ggatcgcaa caaatactac	10620
cttttatctt gctcttctg ctctcaggta ttaatgccga attgtttcat ctgtctgtg	10680
tagaagacca cacacgaaaa tcctgtgatt ttacatttta cttatcgta atcgaaatga	10740
tatctattta atctgctttt ctgtctaat aaatatatat gtaaagtacg cttttgttg	10800
aaatttttta aacctttgtt tttttttt tcttatttc gtaactcttc taccttcttt	10860
atttactttc taaaatccaa atacaaaaca taaaaataa taaacacaga gtaaattccc	10920
aaattattcc atcattaaaa gatacaggc gcgtgtaagt tacaggcaag cgatccgtcc	10980
taagaaacca ttattatcat gacattaacc tataaaaata ggctatcac gaggccttt	11040
cgctcgcgc gtttcggtga tgacggtgaa aacctctgac acatgcagct cccggagacg	11100
gtcacagctt gtctgtaagc ggatgccggg agcagacaag cccgtcaggc gcgctcagcg	11160
gggtttggcg gggtcgggg ctggcttaac tatgcggcat cagagcagat tgtactgaga	11220
gtgcaccata ccacagccgg aagaggagta gggaatatta ctggctgaaa ataagtcttg	11280
aatgaacgta tacgcgtata tttctaccaa tctctcaaca ctgagtaatg gtagttataa	11340
gaaagagacc gagttaggga cagttagagg cggtaggat attccttatg gcatgtctgg	11400
cgatgataaa acttttcaaa cggcagcccc gatctaaaag agctgacagg gaaatggtca	11460
gaaaaagaaa cgtgcacccg cccgtctgga cgcgccgctc accgcacgg cagagaccaa	11520
tcagtaaaaa tcaacggta acgacattac tatatatata atataggaag catttaatag	11580
aacagcatcg taatatatgt gtactttgca gttatgacgc cagatggcag tagtggaaga	11640
tattctttat tgaaaaatag ctgtcacct tacgtacaat ctgatccgg agcttttctt	11700
tttttgccga ttaagaattc ggtcgaaaaa agaaaaggag agggccaaga gggaggcgt	11760
tggtgactat tgagcacgtg agtatactg attaaagcaca caaaggcagc ttggagtatg	11820
tctgttatta atttcacagg tagttctggt ccattggtga aagtttgcgg ctgcagagc	11880

acagaggccg cagaatgtgc tctagattcc gatgtgact tgctgggtat tatatgtgtg	11940
cccaatagaa agagaacaat tgacccggtt attgcaagga aaatttcaag tcttgtaaaa	12000
gcataataaa atagttcagg cactccgaaa tacttggttg gcgtgtttcg taatcaacct	12060
aaggaggatg ttttggctct ggtcaatgat tacggcattg atatcgtcca actgcatgga	12120
gatgagtcgt ggcaagaata ccaagagttc ctcggtttgc cagttattaa aagactcgta	12180
tttccaaaag actgcaacat actactcagt gcagcttcac agaaacctca ttcgtttatt	12240
cccttgtttg attcagaagc aggtgggaca ggtgaacttt tggattggaa ctcgatttct	12300
gactgggttg gaaggcaaga gagccccgaa agcttacatt ttatgttagc tgggtgactg	12360
acgccagaaa atgttgggta tgcgcttaga ttaaattggcg ttattgggtg tgatgtaagc	12420
ggagggtgtg agacaaatgg tgtaaaagac tctaacaaaa tagcaaattt cgtcaaaaat	12480
gctaagaaat aggttattac tgagtagtat ttatttaagt attgtttgtg cacttgcctg	12540
caggcctttt gaaaagcaag cataaaagat ctaaacataa aatctgtaa ataacaagat	12600
gtaaagataa tgctaaatca tttggctttt tgattgattg tacaggatat cggtgtgaaa	12660
taccgccatg accaaaaacc cttaacgtga gttttcgttc cactgagcgt cagaccccg	12720
agaaaagatc aaaggatctt cttgagatcc tttttttctg cgcgtaactc gctgcttgca	12780
aacaaaaaaaa ccaccgtac cagcgggtgg ttttttgccg gatcaagagc taccaactct	12840
ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgttc ttctagtga	12900
gccgtagtta ggccaccact tcaagaactc tgtagcaccg cctacatacc tcgctctgct	12960
aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttacg ggttggactc	13020
aagacgatag ttaccggata aggcgcagcg gtcgggctga acgggggggtt cgtgcacaca	13080
gcccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg agctatgaga	13140
aagcgccacg cttcccgaag ggagaaaggc ggacaggtat ccggtaaagc gcagggtcgg	13200
aacaggagag cgcacgaggg agcttccagg gggaaacgcc tggatatctt atagtctgt	13260
cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtcag gggggcggag	13320
cctatggaaa aacgccagca acgcggcctt ttacggttc ctggcctttt gctggccttt	13380
tgctcacatg ttctttcctg cgttatccc tgattctgtg gataaccgta ttaccgcctt	13440
tgagtgagct gataccgctc gccgcagccg aacgaccgag cgcagcaggt cagtgagcga	13500
ggaagcggaa gagcgcccaa tacgcaaacc gcctctcccc gcgcgttggc cgattcatta	13560
atgcagctgg cagcagaggt ttcccgactg gaaagcgggc agtgagcgca acgcaattaa	13620
tgtgagttag ctactcatt aggcacccca ggttttacac tttatgcttc cggctcgtat	13680
gttgtgtgga attgtgagcg gataacaatt taactataac ggtcctaagg tagcgaa	13737

<210> 102

<211> 6351

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetically constructed plasmid pJDI427

<400> 102

ttgagatcct tttttctgc gcgtaatctg ctgcttgcaa acaaaaaaac caccgctacc	60
agcgggtggtt tgittgccgg atcaagagct accaactctt tttccgaagg taactggctt	120
cagcagagcg cagataccaa atactgttct tctagtgtag ccgtagttag gccaccactt	180
caagaactct gtagcaccgc ctacatacct cgtcttgccta atcctgttac cagtggctgc	240
tgccagtggc gataagtcgt gtcttaccgg gttggactca agacgatagt taccggataa	300
ggcgcagcgg tcgggctgaa cgggggggttc gtgcacacag ccagcttgg agcgaacgac	360
ctacaccgaa ctgagatacc tacagcgtga gctatgagaa agcgccacgc ttcccgaagg	420
gagaaaggcg gacaggtatc cggtaagcgg cagggtcgga acaggagagc gcacgaggga	480
gcttccaggg ggaacgcct ggtatcttta tagtctgtc gggtttcgcc acctctgact	540
tgagcgtcga tttttgtgat gctcgtcagg ggggcggagc ctatggaaaa acgccagcaa	600
cgcggccttt ttacggttcc tggccttttg ctggcctttt gctcacatgt tctttcctgc	660
gttatccctt gattctgtgg ataaccgtat taccgccttt gattgagctg ataccgctcg	720
ccgcagccga acgaccgagc gcagcgagtc agtgagcgag gaagcggaag agcgcccaat	780
acgcaaaccg cctctccccg cgcgttggcc gattcattaa tgcagctggc acgacaggtt	840
tcccgactgg aaagcgggca gtgagcgcaa cgcaattaat gtgagttagc tactcatta	900
ggcaccccag gctttacact gataatgggt gagtgagtgt gtgcgtgtgg ggcgcgccag	960
atgggaacag ctagcttcac gctgccgcaa gcactcaggg cgcaagggtt gctaaaggaa	1020
gcggaacacg tagaaagcca gtccgcagaa acgggtgctga ccccgatga atgtcagcta	1080
ctgggctatc tggacaaggg aaaacgcaag cgcaaagaga aagcaggtag cttgcagtgg	1140
gcttacatgg cgatagctag actgggcggt tttatggaca gcaagcgaac cggaattgcc	1200
agctggggcg ccctctggta aggttgggaa gccctgcaaa gtaactgga tggctttctt	1260
gccccaagg atctgatggc gcaggggatc aagatctgat caagagacag gatgaggatc	1320
gtttcgcatt attgaacaag atggattgca cgcaggttct ccggccgctt ggggtggagag	1380
gctattcggc tatgactggg cacaacagac aatcggtgc tctgatgccg ccgtgttccg	1440

gctgtcagcg caggggcgcc cggttctttt tgtaagacc gacctgtccg gtgccctgaa 1500
 tgaactccaa gacgaggcag cgcgctatc gtggctggcc acgacgggcg ttccttgccg 1560
 agctgtgctc gacgttgtca ctgaagcggg aagggaactgg ctgctattgg gcgaagtgcc 1620

ggggcaggat ctctgtcat ctcacattgc tctgcccag aaaglatcca tcatggctga 1680
 tgcaatgcgg cggtgcata cgcttgatcc ggctacctgc ccattcgacc accaagcgaa 1740
 acatcgcatc gagcgagcac gtactcggat ggaagccggt cttgtcgatc aggatgatct 1800
 ggacgaagag catcaggggc tcgcgccagc cgaactgttc gccaggctca aggcgcgat 1860
 gccccagggc gaggatctcg tcgtgacca tggcgatgcc tgcttgccga atatcatggt 1920
 ggaaaatggc cgcttttctg gattcatcga ctgtggccgg ctgggtgtgg cggaccgcta 1980
 tcaggacata gcgttggcta cccgtgatat tgctgaagag cttggcggcg aatgggctga 2040

ccgttctctc gtgttttacg gtatcgccgc tcccattcg cagcgcatcg ctttctatcg 2100
 ctttcttgac gagtcttctc gagcgggact ctggggttcg ctaggagatc gatccttttt 2160
 aacccatcac atatactgc cgttcactat tatttagtga aatgagatat tatgatattt 2220
 tctgaattgt gattaaaaa gcaactttat gcccatgcaa cagaaactat aaaaaataca 2280
 gagaatgaaa agaaacagat agatttttta gtcttttagg cccgtagtct gcaaactctt 2340
 ttatgatattt ctatcaaaaa aaagaggaaa atagaccagt tgcaatccaa acgagagtct 2400
 aatagaatga ggtcgaaaag taaatcgcg cggtttgtta ctgataaagc aggcaagacc 2460

taaaatgtgt aaagggcaaa gtgtatactt tggcgtcacc cttacatat tttaggtctt 2520
 tttttattgt gcgtactaa ctggccatct tcaaacagga gggtggaag aagcagaccg 2580
 ctaacacagt acataaaaa ggagacatga acgactccag tctttctaga agatggcaaa 2640
 cagctattat gggattatg ggtccccgaa gcagggttat gcagcggaag agctccccga 2700
 aaagtgccac ctgggtcctt ttcacacgt gctataaaaa taattataat ttaaattttt 2760
 taatataaat atataaatta aaaatagaaa gtaaaaaaag aaattaaaga aaaaatagtt 2820
 ttgtttttcc gaagatgtaa aagactctag ggggatcgcc aacaataact accttttctc 2880

ttgtcttcc tgctctcagg tattaatgcc gaattgtttc atcttgtctg tgtagaagac 2940
 cacacacgaa aatcctgiga ttttacattt tacttatcgt taatcgaatg tatatctatt 3000
 taatctgctt tcttgtcta ataaatatat atgtaaagta cgctttttgt tgaaattttt 3060
 taaacctttg tttatttttt tttcttcatt ccgtaactct tctaccttct ttatttactt 3120
 tctaaaatcc aaatacaaaa cataaaaata aataaacaca gagtaaattc ccaaattatt 3180
 ccatcattaa aagatacagc gcgcgtgtaa gttacaggca agcgatccgt cctaagaaac 3240

cattattatc atgacattaa cctataaaaa taggcgtatc acgaggccct ttcgtctcgc	3300
gcgtttcggg gatgacgggtg aaaacctctg acacatgcag ctcccggaga cggtcacagc	3360
ttgtctgtaa gcggatgccg ggagcagaca agcccgtcag ggcgcgctcag cgggtgttgg	3420
cgggtgtcgg ggctggctta actatgcggc atcagagcag attgtactga gagtgcacca	3480
taccacagcc ggaagaggag tagggaatat tactggctga aaataagtct tgaatgaacg	3540
tatacgcgta tatttctacc aatctctcaa cactgagtaa tggtagttat aagaaagaga	3600
ccgagttagg gacagttaga ggcggtggag atattcctta tggcatgtct ggcatgata	3660
aaacttttca aacggcagcc ccgatctaaa agagctgaca gggaaatggt cagaaaaaga	3720
aacgtgcacc cgcccgtctg gacgcgccgc tcaccgcac ggacagagacc aatcagtaaa	3780
aatcaacggt taacgacatt actatatata taatatagga agcatttaat agaacagcat	3840
cgtaatatat gtgtactttg cagttatgac gccagatggc agtagtggaa gatattcttt	3900
attgaaaaat agcttgtcac cttacgtaca atcttgatcc ggagcttttc tttttttgcc	3960
gattaagaat tcggtcgaag aaagaaaagg agagggccaa gagggagggc attggtgact	4020
attgagcacg tgagtatagc tgattaagca cacaaggca gcttggagta tgtctgttat	4080
taatttcaca ggtagtctctg gtccattggt gaaagtttgc ggcttgcaga gcacagaggc	4140
cgcagaatgt gctctagatt ccgatgctga ctgtctgggt attatatgtg tgcccaatag	4200
aaagagaaca attgaccggt ttattgcaag gaaaatttca agtcttgtta aagcatataa	4260
aaatagtcca ggcactccga aatacttggg tggcgtgttt cgtaatcaac ctaaggagga	4320
tgttttggct ctggtcaatg attacggcat tgatatcgtc caactgcatg gagatgagtc	4380
gtggcaagaa taccaagagt tcctcggttt gccagttatt aaaagactcg tatttccaaa	4440
agactgcaac atactactca gtgcagcttc acagaaacct cattcgttta ttcccttggt	4500
tgattcagaa gcaggtggga caggtgaact tttggattgg aactcgattt ctgactgggt	4560
tggaaggcaa gagagcccg aaagcttaca ttttatgtta gctggtggac tgacgccaga	4620
aaatgttggt gatgcgctta gattaatgg cgttattggt gttgatgtta gcggagggtg	4680
ggagacaaat gggtgaaaag actctaacaa aatagcaaat ttcgtcaaaa atgctaagaa	4740
ataggttatt actgagtagt atttatataa gtattgtttg tgcacttgcc tgcaggcctt	4800
ttgaaaagca agcataaag atctaacaat aaaatctgta aaataacaag atgtaaagat	4860
aatgctaaat catttggttt tttgattgat tgtacaggac tgggtggaat cccttctgca	4920
gcacctggat taccctgtta tcctagttt tgggttaaag atggttaaat gattcgaaaa	4980

taataaaggg aaaatcagtt ttgatataca aaattataca tgtcaacgat aatacaaaat 5040
 ataatacaaa ctataagatg ttatcagtat ttattatgca tttagaataa attttgtgtc 5100
 gcccttaatt gtgagcggat aacaattacg agcttcatgc acagtgaat catgaaaaat 5160
 ttatttgctt tggagcggga taacaattat aatatgtgga attgtgagcg ctacacaattc 5220
 cacaacggtt tccctctaga aataattttg tttactttt cgagacctta ggaggtaaac 5280
 atatgacggc attgacggaa ggtgcaaac tgtttgagaa agagatcccg tatatcaccg 5340
 aactggaagg cgacgtcgaa ggtatgaaat ttatcattaa aggcgagggt accggtgacg 5400

 cgaccacggg taccattaaa gcgaaataca tctgcactac gggcgacctg ccggtcccg 5460
 gggaaccct ggtgagcacc ctgagctacg gtgttcagtg tttcgccaag taccgagcc 5520
 acatcaagga tttctttaag agcgccatgc cggaagggtta tacccaagag cgtacatca 5580
 gcttcaagg cgacggcgtg tacaagacgc gtgctatggt tacctacgaa cgcggttcta 5640
 tctacaatcg tgcacgctg actggtgaga actttaagaa agacgggtcac attctgcgta 5700
 agaacgttg attccaatgc ccgccaagca tttgtatat tctgcctgac accgttaaca 5760
 atggcatccg cgttgagttc aaccaggcgt acgatattga aggtgtgacc gaaaaactgg 5820

 ttaccaaag cagccaaatg aatcgtccgt tggcgggctc cgcggcagtg catatccgc 5880
 gttatcatca cattacctac cacacaaac tgagcaaaga ccgagcagc cgccgtgatc 5940
 acatgtgtct ggtagaggct gtgaaagcgg ttgatctgga cacgtatcag taatgagaat 6000
 tctgtacct cgagggtctc acccaaggg cgacacccc taattagccc gggcgaaagg 6060
 cccagtctt cgactgagcc tttcgtttta ttgatgcct ggcagttccc tactctcgca 6120
 tggggagtcc ccacactacc atcggcgcta cggcgtttca cttctgagtt cggcatgtcg 6180
 aagataggga ctcttaacat cccaaaacta cctagctgca ttttcaggag gaagcgatgg 6240

 gcggcgcac accttctatg cgggtgtgaaa taccgcatg accaaaatcc cttaacgtga 6300
 gttttcgttc cactgagcgt cagaccccg agaaaagatc aaaggatctt c 6351
 <210> 103
 <211> 6351
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetically constructed plasmid pJDI429
 <400> 103

 ttgagatcct ttttttctgc gcgtaatctg ctgcttgcaa acaaaaaaac caccgtacc 60
 agcggtggtt tgtttgccgg atcaagagct accaactctt tttccgaagg taactggctt 120

cagcagagcg cagataccaa atactgttct tctagtgtag ccgtagttag gccaccactt 180

caagaactct gtagcaccgc ctacatacct cgctctgcta atcctgttac cagtggctgc 240

tgccagtggc gataagtcgt gtcttaccgg gttggactca agacgatagt taccggataa 300

ggcgcagcgg tcgggctgaa cgggggggttc gtgcacacag cccagcttgg agcgaacgac 360

ctacaccgaa ctgagatacc tacagcgtga gctatgagaa agcggccacgc tccccgaagg 420

gagaaaggcg gacaggtatc cggtaagcgg cagggtcgga acaggagagc gcacgaggga 480

gcttccaggg ggaaacgcct ggtatcttta tagtcctgtc gggtttcgcc acctctgact 540

tgagcgtcga tttttgtgat gctcgtcagg ggggcggagc ctatggaaaa acgccagcaa 600

cgcggccttt ttacggttcc tggccttttg ctggcctttt gctcacatgt tctttcctgc 660

gttatccctt gattctgtgg ataaccgtat taccgccttt gattgagctg ataccgctcg 720

ccgcagccga acgaccgagc gcagcagatc agtgagcgag gaagcgggaag agcgcccaat 780

acgcaaaccg cctctccccg cgcgttggcc gattcattaa tgcagctggc acgacaggtt 840

tcccacttgg aaagcgggca gtgagcgcaa cgcaattaat gtgagttagc tcactcatta 900

ggcacccag gctttacact gataatgggt gattgagtgt gtgcgtgtgg ggcgcgccag 960

atgggaacag ctagcttcac gctgccgcaa gcactcaggg cgcaagggtt gctaaaggaa 1020

gcggaacacg tagaaagcca gtccgcagaa acggtgctga ccccgatga atgtcagcta 1080

ctgggctatc tggacaaggg aaaacgcaag cgcaaagaga aagcaggtag cttgcagtgg 1140

gcttacatgg cgatagctag actgggcggt tttatggaca gcaagcgaac cggaattgcc 1200

agctggggcg cctcttggtt aggttgggaa gccctgcaaa gtaaaactgga tggctttctt 1260

gccccaagg atctgatggc gcaggggatc aagatctgat caagagacag gatgaggatc 1320

gtttcgcatg attgaacaag atggattgca cgcaggttct ccggccgctt gggtggagag 1380

gctattcggc tatgactggg cacaacagac aatcggtgc tctgatgccg ccgtgttccg 1440

gctgtcagcg caggggcgcc cggttctttt tgtcaagacc gacctgtccg gtgccctgaa 1500

tgaactcaa gacgaggcag cgcggtatc gtggctggcc acgacgggcg ttccttgccg 1560

agctgtgctc gacgttgtca ctgaagcggg aagggactgg ctgctattgg gcgaagtgcc 1620

ggggcaggat ctctgtcat ctcaccttgc tctgccgag aaagtatcca tcatggctga 1680

tgcaatgcgg cggtgcata cgcttgatcc ggctacctgc ccattcgacc accaagcgaa 1740

acatcgatc gagcgagcac gtactcggat ggaagccggt cttgtcgatc aggatgatct 1800

ggacgaagag catcaggggc tcgcgccagc cgaactgttc gccaggctca aggcgcggat 1860

gccccacggc gaggatctcg tcgtgaccca tggcgatgcc tgcttgccga atatcatggt	1920
ggaaaatggc cgcttttctg gattcatcga ctgtggccgg ctgggtgtgg cggaccgcta	1980
tcaggacata gcgttggtta cccgtgatat tgcgaagag cttggcggcg aatgggctga	2040
ccgcttcctc gtgctttacg gtatcgccgc tcccgatcgc cagcgcatcg ccttctatcg	2100
ccttcttgac gagttcttct gagcgggact ctggggttcg ctaggagatc gatccttttt	2160
aacccatcac atatacctgc cgttcactat tatitagtga aatgagatat tatgatattt	2220
tctgaattgt gattaaaaag gcaactttat gcccatgcaa cagaaactat aaaaaataca	2280
gagaatgaaa agaacagat agatttttta gttctttagg cccgtagtct gcaaatcctt	2340
ttatgatattt ctatcaaca aaagaggaaa atagaccagt tgcaatcaa acgagagtct	2400
aatagaatga ggtcgaaaag taaatcgcg cgggtttgtta ctgataaagc aggcaagacc	2460
taaaaatgtgt aaagggcaaa gtgtatactt tggcgtcacc ccttacatat tttaggtctt	2520
tttttattgt gcgtaactaa cttgccatct tcaaacagga gggctggaag aagcagaccg	2580
ctaacacagt acataaaaaa ggagacatga acgactccag tctttctaga agatggcaaa	2640
cagctattat gggatattat gggtcccgaa gcagggttat gcagcggaaa agctccccga	2700
aaagtccac ctgggtcctt ttcacacgt gctataaaaa taattataat ttaaattttt	2760
taataataat atataaatta aaaatagaaa gtaaaaaaag aaattaaaga aaaaatagtt	2820
tttgttttcc gaagatgtaa aagactctag ggggatcgcc aacaataact accttttacc	2880
ttgctcttcc tgctctcagg tattaatgcc gaattgtttc atcttgtctg tgtagaagac	2940
cacacacgaa aatcctgtga ttttacattt tacttatcgt taatcgaatg tatatctatt	3000
taatctgctt ttcttgtcta ataaatatat atgtaaagta cgctttttgt tgaaattttt	3060
taaacctttg tttatttttt tttcttcatt ccgtaactct tctaccttct ttatttactt	3120
tctaaaatcc aaatacaaaa cataaaaaata aataaacaca gagtaaatc ccaaattatt	3180
ccatcattaa aagatacagg gcgcgtgtta gttacaggca agcgatccgt cctaagaaac	3240
cattattatc atgacattaa cctataaaaa taggcgtatc acgaggccct ttcgtctcgc	3300
gcgtttcggg gatgacggtg aaaacctctg acacatgcag ctcccggaga cggtcacagc	3360
ttgtctgtaa gcggatgccg ggagcagaca agcccgtcag ggcgctcag cgggtgttgg	3420
cgggtgtcgg ggcgtgctta actatcgccg atcagagcag attgtactga gagtgcacca	3480
taccacagcc ggaagaggag tagggaatat tactggctga aaataagtct tgaatgaacg	3540
tatacgctga tatttctacc aatctctcaa cactgagtaa tggtagttat aagaaagaga	3600
ccgagttagg gacagttaga ggcgtgggag atattcctta tggcatgtct ggcatgata	3660
aaacttttca aacggcagcc ccgatctaaa agagctgaca gggaaatggt cagaaaaaga	3720

aacgtgcacc cgcccgctcg gacgcgccgc tcacccgcac ggagagagacc aatcagtaaa 3780
 aatcaacggt taacgacatt actatatata taatatagga agcatttaat agaacagcat 3840
 cgtaatatat gtgtactttg cagttatgac gccagatggc agtagtgga gatattcttt 3900
 attgaaaaat agcttgtcac cttacgtaca atcttgatcc ggagcttttc tttttttgcc 3960

 gattaagaat tcggtcgaaa aaagaaaagg agagggccaa gagggagggc attggtgact 4020
 attgagcacg tgagtatacg tgattaagca cacaaaggca gcttgagta tgtctgttat 4080
 taatttcaca ggtagtctcg gtccattggt gaaagtttgc ggcttgcaga gcacagaggc 4140
 cgcagaatgt gctctagatt cccgatgctga ctgctgggt attatatgtg tgcccaatag 4200
 aaagagaaca attgaccggg ttattgcaag gaaaatttca agtcttgtaa aagcatataa 4260
 aaatagtcca ggcactccga aatacttggg tggcgtgttt cgtaatcaac ctaaggagga 4320
 tgttttggct ctggtcaatg attacggcat tgatatcgtc caactgcatg gagatgagtc 4380

 gtggcaagaa taccaagagt tcctcggttt gccagtattt aaaagactcg tatttcctaaa 4440
 agactgcaac atactactca gtgcagcttc acagaaacct cattcgttta ttccttgtt 4500
 tgattcagaa gcaggtggga caggtgaact tttggattgg aactcgattt ctgactgggt 4560
 tggaaggcaa gagagccccg aaagcttaca ttttatgtta gctggtggac tgacgccaga 4620
 aaatgttggg gatgcgctta gattaaatgg cgttattggg gttgatgtta gcggagggtg 4680
 ggagacaaat ggtgtaaaag actctaacaa aatagcaaat ttcgtcaaaa atgctaagaa 4740
 ataggttatt actgagtagt atttatttaa gtattgtttg tgcacttgcc tgcaggcctt 4800

 ttgaaaagca agcataaaaag atctaacaat aaaatctgta aaataacaag atgtaaagat 4860
 aatgctaaat catttgctt tttgattgat tgtacaggac tgggtggaat cccttctgca 4920
 gcacctggat taccctgtta tccctagttt tgggatgtta agagtcctta tcttcgaaaa 4980
 taataaaggg aaaatcagtt tttgatatca aaattataca tgtcaacgat aatacaaaat 5040
 ataatacaa ctataagatg ttatcagtat ttattatgca tttagaataa attttgtgtc 5100
 gcccttaatt gtgagcggat aacaattacg agcttcatgc acagtgaat catgaaaaat 5160
 ttatttgcct tgtgagcgga taacaattat aatatgtga attgtgagcg ctcacaattc 5220

 cacaacggtt tcctctaga aataattttg tttactttt cgagacctta ggaggtaaac 5280
 atatgacggc attgacggaa ggtgcaaaac tgtttgagaa agagatcccg tatatcaccg 5340
 aactggaagg cgacgtcgaa ggtatgaaat ttatcattaa aggcgagggt accggtgacg 5400
 cgaccacggg taccattaaa gcgaaataca tctgcactac gggcgacctg ccggtcccgt 5460
 gggcaacctt ggtgagcacc ctgagctacg gtgttcagtg tttcgccaag taccgagcc 5520
 acatcaagga tttctttaag agcgccatgc cggaagggtta tacccaagag cgtacatca 5580

gcttcgaagg cgacggcgtg tacaagacgc gtgctatggt tacctacgaa cgcggttcta 5640

tctacaatcg tgtcacgctg actggtgaga actttaagaa agacggtcac attctgcgta 5700

agaacgttgc attccaatgc cgcgaagca tctgtatat tctgcctgac accgttaaca 5760

atggcatccg cgttgagttc aaccaggcgt acgatatga aggtgtgacc gaaaaactgg 5820

ttaccaaagc cagccaaatg aatcgtccgt tggcgggctc cgcggcagtg catatcccgc 5880

gttatcatca cattacctac cacaccaaac tgagcaaaga ccgcgacgag cgccgtgac 5940

acatgtgtct ggtagaggtc gtgaaagcgg ttgatctgga cacglatcag taatgagaat 6000

tctgtacact cgagggtctc accccaaggg cgacaccccc taattagccc gggcgaaagg 6060

cccagtcctt cgactgagcc ttctgtttta ttgatgcct ggcagttccc tactctcgca 6120

tggggagtcc ccacactacc atcggcgcta cggcgtttca cttctgagtt cggcatgtcg 6180

attgggggtat ggggggggcg gtcaaaacta cctagctgca ttttcaggag gaagcgatgg 6240

gcggccgcac accttctatg cgggtgtgaaa taccgccatg accaaaatcc cttaacgtga 6300

gttttcgttc cactgagcgt cagaccccggt agaaaagatc aaaggatctt c 6351

<210> 104

<211> 6353

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetically constructed plasmid pJDI430

<400> 104

ttgagatcct tttttctgc gcgtaatctg ctgcttgcaa acaaaaaaac caccgtacc 60

agcgggtggtt tgtttgccgg atcaagagct accaactctt tttccgaagg taactggctt 120

cagcagagcg cagataccaa atactgttct tctagtgtag ccgtagttag gccaccactt 180

caagaactct gtagcaccgc ctacatacct cgctctgcta atctgttac cagtggctgc 240

tgccagtggc gataagtcgt gtcttaccgg gttggactca agacgatagt taccggataa 300

ggcgcagcgg tcgggctgaa cgggggggtt gtgcacacag ccagcttgg agcgaacgac 360

ctacaccgaa ctgagatacc tacagcgtga gctatgagaa agcggccacgc ttcccgaagg 420

gagaaaggcg gacaggatc cggtaaagcg cagggtcgga acaggagagc gcacagggga 480

gcttcagggg gaaacgcct ggtatcttta tagtcctgtc gggtttcgcc acctctgact 540

tgagcgtcga tttttgtgat gctcgtcagg ggggaggagc ctatggaaaa acgccagcaa 600

cgcggccttt ttacggttcc tggccttttg ctggcctttt gctcacatgt tctttctgc 660

gttatccct gattctgtgg ataaccgtat taccgccttt gagtgagctg ataccgctcg	720
ccgcagccga acgaccgagc gcagcgagtc agtgagcgag gaagcggaag agcgcccaat	780
acgcaaaccg cctctccccg cgcgttggcc gattcattaa tgcagctggc acgacaggtt	840
tcccactgg aaagcgggca gtgagcgcaa cgcaattaat gtgagttagc tcactcatta	900
ggcaccacag gctttacact gataatgggt gagtgagtgt gtgcgtgtgg ggcgcgccag	960
atgggaacag ctagcttcac gctgccgcaa gcactcaggg cgcaagggt gctaaaggaa	1020
gcggaacacg tagaaagcca gtccgcagaa acggtgctga ccccgatga atgtcagcta	1080
ctgggtatc tggacaaggg aaaacgcaag cgcaaagaga aagcaggtag ctgtcagtgg	1140
gcttacatgg cgatagctag actgggcggt tttatggaca gcaagcgac cggaattgcc	1200
agctggggcg ccctctggta aggttgggaa gccctgcaa gtaactgga tggctttctt	1260
gccccaagg atctgatggc gcaggggatc aagatctgat caagagacag gatgaggatc	1320
gtttcgcag attgacaag atggattgca cgcaggttct ccggccgctt ggggtggagag	1380
gctattcggc tatgactggg cacaacagac aatcggtgc tctgatgcc ccgtgttccg	1440
gtgttcagcg caggggcgc cggttctttt tgtcaagacc gacctgtccg gtgccctgaa	1500
tgaactcaa gacgaggcag cgcgctatc gtggctggcc acgacgggcg ttccttgcgc	1560
agctgtgctc gacgttgta ctgaagcggg aagggaactgg ctgctattgg gcgaagtgcc	1620
ggggcaggat ctctgtcat ctacattgc tctgccgag aaagtatcca tcatggctga	1680
tgcaatgcgg cggtgcata cgcttgatcc ggctacctgc ccattcgacc accaagcgaa	1740
acatcgcatc gagcgagcac gtactcggat ggaagccggt cttgtcgatc aggatgatct	1800
ggacgaagag catcaggggc tcgcgccagc cgaactgttc gccaggctca aggcgcggat	1860
gccccagcgc gaggatctcg tcgtgacca tggcgatgcc tgcttgccga atatcatggt	1920
ggaaaatggc cgcttttctg gattcatcga ctgtggccgg ctgggtgtgg cggaccgcta	1980
tcaggacata gcgttggcta ccctgatat tgctgaagag cttggcggcg aatgggctga	2040
ccgttctctc gtgctttacg gtatcgccgc tcccattcg cagcgcatcg cttctatcg	2100
cttcttgac gagtcttct gagcgggact ctggggttcg ctaggatc gatcctttt	2160
aaccatcac atatactgc cgttcactat tatttagtga aatgagatat tatgatattt	2220
tctgaattgt gattaaaaa gcaactttat gcccatgcaa cagaaactat aaaaaatata	2280
gagaatgaaa agaacagat agatttttta gttctttagg cccgtagtct gcaaatcctt	2340
ttatgatatt ctatcaaaaa aaagaggaaa atagaccagt tgcaatcaa acgagagtct	2400
aatagaatga ggtcgaaaa taaatcgcg gggtttgtta ctgataaagc aggcaagacc	2460

taaaatgtgt aaagggcaaa gtgtatactt tggcgtcacc ccttacatat tttaggtctt	2520
tttttattgt gcgtactaa cttgccatct tcaaacagga gggctggaag aagcagaccg	2580
ctaacacagt acataaaaaa ggagacatga acgactccag tctttctaga agatggcaaa	2640
cagctattat gggattatg ggtccccgaa gcagggttat gcagcggaag agtccccga	2700
aaagtccac ctgggtcctt ttcacacgt gctataaaaa taattataat ttaaattttt	2760
taataataat atataaatta aaaatagaaa gtaaaaaaag aaattaaaga aaaaatagtt	2820
tttgttttcc gaagatglaa aagactctag ggggatcgcc aacaataact acctttatc	2880
ttgtcttcc tgctctcagg tattaatgcc gaattgttc atcttgtctg tgtagaagac	2940
cacacacgaa aatcctgtga ttttacattt tacttatcgt taatcgaatg tatatctatt	3000
taatctgctt ttcttgtcta ataaatata atgtaaagta cgctttttgt tgaaattttt	3060
taaacctttg tttatttttt tttcttcatt ccgtaactct tctaccttct ttattttactt	3120
tctaaaatcc aaatacaaaa cataaaaaa aataaacaca gagtaaatc ccaaattatt	3180
ccatcattaa aagatacgag gcgcgtgtaa gttacaggca agcgatccgt cctaagaaac	3240
cattattatc atgacattaa cctataaaaa taggcgtatc acgagccct ttcgtctcgc	3300
gcgtttcggg gatgacggg aaaacctctg acacatgcag ctcccggaga cggtcacagc	3360
ttgtctgtaa gcggatgccg ggagcagaca agcccgtcag ggcgctcag cgggtgttg	3420
cgggtgtcgg ggctggctta actatgcggc atcagagcag attgtactga gagtgcacca	3480
taccacagcc ggaagaggag tagggaatat tactggctga aaataagtct tgaatgaacg	3540
tatacgcta tatttctacc aatctctcaa cactgagtaa tggtagttat aagaaagaga	3600
ccgagtagg gacagtiaga ggcggtggag atattcctta tggcatgtct ggcatgata	3660
aaacttttca aacggcagcc ccgatctaaa agagctgaca gggaaatggt cagaaaaaga	3720
aacgtgcacc cgccgtctg gacgcgccgc tcaccgcac ggcagagacc aatcagtaaa	3780
aatcaacggt taacgacatt actatatata taatatagga agcatttaat agaacagcat	3840
cgtaatatat gtgtactttg cagttatgac gccagatggc agtagtgga gatattcttt	3900
attgaaaaat agcttgtcac cttacgtaca atcttgatcc ggagcttttc ttttttggc	3960
gattaagaat tcggtcgaaa aaagaaaagg agaggccaa gagggagggc attggtgact	4020
attgagcacg tgagtatacg tgattaagca cacaaaggca gcttgagta tgtctgttat	4080
taatttcaca ggiagtctctg gtccattggt gaaagtttgc ggcttgaga gcacagaggc	4140
cgcagaatgt gctctagatt ccgatgctga ctgtctgggt attatatgtg tgcccaatag	4200

aaagagaaca attgaccggt ttattgcaag gaaaatttca agtcttgtaa aagcatataa	4260
aaatagtcca ggcactccga aatacttggg tggcgtgttt cgtaatcaac ctaaggagga	4320
tgttttggct ctggccaatg attacggcat tgatatcgtc caactgcatg gagatgagtc	4380
gtggcaagaa taccaagagt tcctcgggtt gccagttatt aaaagactcg tatttccaaa	4440
agactgcaac atactactca gtgcagcttc acagaaacct cattcgttta ttcccttggt	4500
tgattcagaa gcaggtggga caggtgaact ttggattgg aactcgattt ctgactgggt	4560
tggaaggcaa gagagcccg aaagcttaca ttttatgtta gctgggtggac tgacgccaga	4620
aaatgttggg gatgcgctta gattaaatgg cgttattggg gttgatgtaa gcggaggtgt	4680
ggagacaaat ggtgtaaaag actctaacaa aatagcaaat ttcgtcaaaa atgctaagaa	4740
ataggttatt actgagtagt atttatitaa gtattgtttg tgcacttgcc tgcaggcctt	4800
ttgaaaagca agcataaag atctaacaat aaaatctgta aaataacaag atgtaaagat	4860
aatgctaaat catttggctt ttgtattgat tgtacaggac tgggtggaat cccttctgca	4920
gcacctggat taccctgtta tcctagttt tgaggagtgt tcagtcctcg tgaactcgaa	4980
aataataaag ggaaaatcag tttttgatat caaaattata catgtcaacg ataatacaaa	5040
atataataca aactataaga tgttatcagt atttattatg catttagaat aaattttgtg	5100
tcgcccttaa ttgtgagcgg ataacaatta cgagcttcat gcacagtga atcatgaaaa	5160
atttatttgc ttgtgagcgg gataacaatt ataatatgtg gaattgtgag cgctcacaat	5220
tccacaacgg ttccctcta gaaataattt tgtttaactt ttcgagacct taggaggtaa	5280
acatatgacg gcattgacgg aaggtgcaaa actgtttgag aaagagatcc cgtatatcac	5340
cgaactggaa ggcgacgtcg aaggtatgaa atttatcatt aaaggcgagg gtaccggtga	5400
cgcgaccacg ggtaccatta aagcgaaata catctgcact acgggcgacc tgccggtccc	5460
gtgggcaacc ctggtgagca ccctgagcta cgggtgttcag tgtttcgcca agtaccgag	5520
ccacatcaag gatttcttta agagcgccat gccggaaggt tatacccaag agcgtaccat	5580
cagcttcgaa ggcgacggcg tgtacaagac gcgtgctatg gttacctacg aacgcggttc	5640
tatctacaat cgtgtcacgc tgactgggtga gaactttaag aaagacggtc acattctgcg	5700
taagaacgtt gcattccaat gcccgccaag cattctgtat attctgcctg acaccgttaa	5760
caatggcatc cgcgttgagt tcaaccaggc gtacgatatt gaaggtgtga ccgaaaaact	5820
ggttaccaaa tgcagccaaa tgaatcgtcc gttggcgggc tccgcgagcag tgcatatccc	5880
gcgttatcat cacattacct accacaccaa actgagcaaa gaccgcgacg agcgcctga	5940
tcacatgtgt ctggtagagg tcgtgaaagc ggttgatctg gacacgtatc agtaatgaga	6000
attctgtaca ctgagggtc tcacccaag ggcgacaccc cctaattagc ccgggcgaaa	6060

ggcccagtct ttcgactgag cctttcgttt tatttgatgc ctggcagttc cctactctcg 6120

catggggagt cccacacta ccatcggcgc tacggcgttt cacttctgag ttcggcatgt 6180

cgaagatagg gactcttaac atcccaaac tacctagctg cattttcagg aggaagcgat 6240

ggcgggccgc acaccttcta tgcggtgtga aataccgcca tgaccaaagt cccttaacgt 6300

gagttttcgt tccactgagc gtcagacccc gtagaaaaga tcaaaggatc ttc 6353

<210> 105

<211> 6353

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetically constructed plasmid pJDI431

<400> 105

ttgagatcct tttttctgc gcgtaatctg ctgcttgcaa acaaaaaaac caccgctacc 60

agcgggtggtt tgtttgcgg atcaagagct accaactctt tttccgaagg taactggctt 120

cagcagagcg cagataccaa atactgttct tctagtgtag ccgtagttag gccaccactt 180

caagaactct gtagcaccgc ctacatacct cgctctgcta atcctgttac cagtggctgc 240

tgccagtggc gataagtcgt gtcttaccgg gttggactca agacgatagt taccggataa 300

ggcgcagcgg tcgggctgaa cgggggggttc gtgcacacag ccagcttgg agcgaacgac 360

ctacaccgaa ctgagatacc tacagcgtga gctatgagaa agcgccacgc ttcccgaagg 420

gagaaaggcg gacaggtatc cggtaagcgg cagggtcgga acaggagagc gcacagggga 480

gcttccaggg ggaacgcct ggtatcttta tagtctgtc gggtttcgcc acctctgact 540

tgagcgtcga tttttgtgat gtcgtcagg ggggcggagc ctatggaaaa acgccagcaa 600

cgcggccttt ttacggttcc tggccttttg ctggcctttt gctcacatgt tctttcctgc 660

gttatccctt gattctgtgg ataaccgtat taccgccttt gattgagctg ataccgctcg 720

ccgcagccga acgaccgagc gcagcgagtc agtgagcgag gaagcggaag agcgcccaat 780

acgcaaaccg cctctccccg cgcgttggcc gattcattaa tgcagctggc acgacagggt 840

tcccactgg aaagcgggca gtgagcgcaa cgcaattaat gtgagttagc tactcatta 900

ggcacccag gccttacact gataatgggt gattgagttg gtgcgtgtgg ggccgcccag 960

atgggaacag ctagcttcac gctgccgcaa gcactcaggg cgcaagggt gctaaaggaa 1020

gcggaacacg tagaaagcca gtccgcagaa acggtgctga ccccgatga atgtcagcta 1080

ctgggctatc tggacaaggg aaaacgcaag cgcaaagaga aagcaggtag cttgcagtgg 1140

gcttacatgg cgatagctag actgggcggt tttatggaca gcaagcgaac cggaattgcc 1200
agctggggcg cctcttggtta aggttgggaa gccctgcaaa gtaactgga tggctttctt 1260
gccgccaagg atctgatggc gcaggggatc aagatctgat caagagacag gatgaggatc 1320
gtttcgcattg attgaacaag atggattgca cgcaggttct ccggccgctt ggggtggagag 1380
gctattcggc tatgactggg cacaacagac aatcggtgc tctgatgccg ccgtgttcgg 1440
gctgtcagcg caggggcgcc cggttctttt tgtcaagacc gacctgtccg gtgccctgaa 1500

tgaactccaa gacgaggcag cgcggctatc gtggctggcc acgacgggcg ttccttgcgc 1560
agctgtgctc gacgttgta ctgaagcggg aagggaactgg ctgctattgg gcgaagtgcc 1620
ggggcaggat ctctgtcat ctacacattgc tctgcccag aaagtatcca tcatggctga 1680
tgcaatgcgg cggctgcata cgcttgatcc ggctacctgc ccattcgacc accaagcgaa 1740
acatcgatc gagcgagcac gtactcggat ggaagccggt cttgtcgatc aggatgatct 1800
ggacgaagag catcaggggc tcgcgccagc cgaactgttc gccaggctca aggcgcggat 1860
gcccgcggc gaggatctcg tcgtgacca tggcgatgcc tgcttgccga atatcatggt 1920

ggaaaaatggc cgcttttctg gattcatcga ctgtggccgg ctgggtgtgg cggaccgcta 1980
tcaggacata gcgttggcta cccgtgatat tgcgaagag cttggcggcg aatgggctga 2040
ccgttctctc gtgctttacg gtatcgccgc tcccattcg cagcgcatcg ctttctatcg 2100
ccttcttgac gagtcttct gagcgggact ctggggttcg ctagaggatc gatccttttt 2160
aaccatcac atatactgc cgttcactat tatttagtga aatgagatat tatgatattt 2220
tctgaattgt gattaaaaag gcaactttat gcccatgcaa cagaaactat aaaaaatata 2280
gagaatgaaa agaacagat agatttttta gttctttagg cccgtagtct gcaaactcctt 2340

ttatgatattt ctatcaaaaa aaagaggaaa atagaccagt tgcaatccaa acgagagtct 2400
aatagaatga ggtcgaaaaa taaatcgccg gggtttgtta ctgataaagc aggcaagacc 2460
taaaatgtgt aaagggcaaa gtgtatactt tggcgtcacc cttacatat tttaggtctt 2520
tttttattgt gcgtaactaa cttgccatct tcaaacagga gggctggaag aagcagaccg 2580
ctaacacagt acataaaaaa ggagacatga acgactccag tctttctaga agatggcaaa 2640
cagctattat gggatattatg ggtccccgaa gcagggttat gcagcggaag agtccccga 2700
aaagtgccac ctgggtcctt ttcacacgt gctataaaaa taattataat ttaaattttt 2760

taatataaat atataaatta aaaatagaaa gtaaaaaaag aaattaaaga aaaaatagtt 2820
tttgttttcc gaagatgtaa aagactctag ggggatcgcc aacaataact accttttacc 2880
ttgtcttccc tgcctcagg tattaatgcc gaattgttcc atcttgtctg tgtagaagac 2940
cacacacgaa aatcctgtga ttttacattt tacttatcgt taatcgaatg tatatctatt 3000

taatctgctt ttcttgctta ataaatatat atgtaaagta cgccttttgt tgaaattttt	3060
taaacctttg tttatttttt tttcttcatt ccgtaactct tctaccttct ttatttactt	3120
tctaaaatcc aaatacaaaa cataaaaata aataaacaca gagtaaatc ccaaattatt	3180
ccatcattaa aagatacgag gcgcgtgtaa gttacaggca agcgatccgt cctaagaaac	3240
cattattatc atgacattaa cctataaaaa taggcgtatc acgaggccct ttcgtctcgc	3300
gcgtttcggg gatgacgggtg aaaacctctg acacatgcag ctcccggaga cggtcacagc	3360
ttgtctgtaa gcggatgccg ggagcagaca agcccgctcag ggcgctcag cgggtgttgg	3420
cgggtgtcgg ggctggctta actatgcggc atcagagcag attgtactga gaggtcacca	3480
taccacagcc ggaagaggag tagggaatat tactggctga aaataagtct tgaatgaacg	3540
tatacgcgta tatttctacc aatctctcaa cactgagtaa tggtagtatt aagaaagaga	3600
ccgagttagg gacagttaga ggcgtggag atattcctta tggcatgtct ggcgatgata	3660
aaacttttca aacggcagcc ccgatctaaa agagctgaca gggaaatggt cagaaaaaga	3720
aacgtgcacc cggcgtctg gacgcggcgc tcaccgcac ggcagagacc aatcagtaaa	3780
aatcaacggt taacgacatt actatatata taatatagga agcatttaat agaacagcat	3840
cgtaatatat gtgtactttg cagttatgac gccagatggc agtagtggaa gatattcttt	3900
attgaaaaat agcttgtcac cttacgtaca atcttgatcc ggagcttttc tttttttgcc	3960
gattaagaat tcggtcgaaa aaagaaaagg agagggccaa gagggagggc attggtgact	4020
attgagcacg tgagtatacg tgattaagca cacaaaggca gcttggagta tgtctgttat	4080
taatttcaca ggtagtcttg gtccattggt gaaagtttgc ggcttgcaga gcacagaggc	4140
cgcagaatgt gccttagatt ccgatgctga ctgtctgggt attatatgtg tgcccaatag	4200
aaagagaaca attgaccggg ttattgcaag gaaaatttca agtcttgtaa aagcatataa	4260
aaatagtcca ggcactccga aatacttggg tggcgtgttt cgtaatcaac ctaaggagga	4320
tgttttggct ctggtcaatg attacggcat tgatatcgtc caactgcatg gagatgagtc	4380
gtggcaagaa taccaagagt tctcgggttt gccagttatt aaaagactcg tatttccaaa	4440
agactgcaac atactactca gtgcagcttc acagaaacct cattcgttta ttccttgtt	4500
tgattcagaa gcaggtggga caggtgaact tttggattgg aactcgattt ctgactgggt	4560
tggaggcaaa gagagccccg aaagcttaca ttttatgtta gctggtggac tgacgccaga	4620
aaatgttggg gatgcgctta gattaaatgg cgttattggg gttgatgtaa gcggagggtg	4680
ggagacaaat ggtgtaaaag actctaacaa aatagcaaat ttcgtcaaaa atgctaagaa	4740
ataggttatt actgagtagt atttatttaa gtattgtttg tgcacttggc tgcaggcctt	4800

ttgaaaagca agcataaaag atctaaacat aaaatctgta aaataacaag atgtaaagat 4860

aatgctaaat catttggttt ttgtattgat tgtacaggac tgggtggaat cccttctgca 4920

gcacctggat taccctgtta tccctagttt tgaggagtgt tcagtctccg tgaactcgaa 4980

aataataaag ggaaaatcag tttttgatat caaaattata catgtcaacg ataatacaaa 5040

atataataca aactataaga tgttatcagt atttattatg catttagaat aaattttgtg 5100

tcgcccttaa ttgtgagcgg ataacaatta cgagcttcat gcacagtga atcatgaaaa 5160

atttatttgc ttgtgagcgg gataacaatt ataatatgtg gaattgtgag cgctcacaat 5220

tccacaacgg tttccctcta gaaataattt tgtttaactt ttcgagacct taggaggtaa 5280

acatatgacg gcattgacgg aagggtgcaa actgtttgag aaagagatcc cgtatatcac 5340

cgaactggaa ggcgacgtcg aaggatgaa atttatcatt aaaggcgagg gtaccggtga 5400

cgcgaccacg ggtaccatta aagcgaaata catctgcact acgggcgacc tgccggtccc 5460

gtgggcaacc ctggtgagca ccctgagcta cgggtgttcag tgtttcgcca agtaccgag 5520

ccacatcaag gatttcttta agagcgccat gccggaaggt tatacccaag agcgtaccat 5580

cagcttcgaa ggcgacggcg tgtacaagac gcgtgctatg gttacctacg aacgcggttc 5640

tatctacaat cgtgtcacgc tgactggtga gaactttaag aaagacggtc acattctgcg 5700

taagaacgtt gcattccaat gcccgccaag cattctgtat attctgcctg acaccgttaa 5760

caatggcatc cggtttgagt tcaaccaggc gtacgatatt gaaggtgtga ccgaaaaact 5820

ggttaccaaa tgcagccaaa tgaatcgtcc gttggcgggc tccgcggcag tgcatatccc 5880

gcgttatcat cacattacct accacaccaa actgagcaaa gaccgcgacg agcgccgtga 5940

tcacatgtgt ctggtagagg tcgtgaaagc ggttgatctg gacacgtatc agtaatgaga 6000

attctgtaca ctcgagggtc tcaccccaag ggcgacaccc cctaattagc ccgggcgaaa 6060

ggcccagtct ttcgactgag cctttcggtt tatttgatgc ctggcagttc cctactctcg 6120

catggggagt cccacacta ccatcggcgc tacggcggtt cacttctgag ttcggcatgt 6180

cgattgggtt atgggggggg cggtaaaac tacctagctg cattttcagg aggaagcgat 6240

gggcggccgc acaccttcta tgcggtgtga aataccgcca tgacccaaat cccttaacgt 6300

gagttttcgt tccactgagc gtcagacccc gtagaaaaga tcaaaggatc ttc 6353

<210> 106

<211> 4155

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetically constructed plasmid pJDI432

<400> 106

gcccctgcag ccgaattata ttatTTTTgc caaataattt ttaacaaaag ctctgaagtc	60
ttcttcattt aaattcttag atgatacttc atctggaaaa ttgtcccaat tagtagcatc	120
acgtgtgag taagttctaa accatttttt tatgtttgta ttatctctaa tcttactact	180
cgatgagttt tcggtattat ctctatTTTT aacttggagc aggttccatt cattgttttt	240
ttcatcatag tgaataaaat caactgcttt aacactttgt cctgaacacc atatccatcc	300
ggcgtaatac gactcactat agggagagcg gccgccagat cttccggatg gctcgagttt	360
ttcagcaaga ttttgggtta aagatgggtt aatgattcga catacacata aagtagcttg	420
cgtatttaaa attatgaacc taaggggttt agcacttcac gctgccgcaa gcactcaggg	480
cgcaagggtt gctaaaggaa gcggaacacg tagaaagcca gtccgcagaa acggtgctga	540
ccccggatga atgtcagcta ctgggctatc tggacaaggg aaaacgcaag cgcaaagaga	600
aagcaggtag ctgtcagtgg gcttacatgg cgatagctag actgggcggt tttatggaca	660
gcaagcgaac cggaattgcc agctggggcg cctcttggtt aggttgggaa gccctgcaaa	720
glaaactgga tggcttttct gccccaagg atctgatggc gcaggggatc aagatctgat	780
caagagacag gatgaggatc gtttcgcatg gagaaaaaga tcacgggcta cactaccgtg	840
gacatctcgc aatggcatcg caaggaacac ttcgaggcat ttcaaagcgt ggcacaatgt	900
acttacaacc agaccgtcca gctggatatt accgcgtttt tgaagaccgt taagaaaaac	960
aagcacaagt ttatccagc ctttatccat attctcgccc gcttgatgaa tgcgcacccc	1020
gaatttcgta tggccatgaa agatggtgag ctcgttatct gggactcagt ccatccatgc	1080
tataccgttt tccacgaaca aactgaaact ttttcttcgc tgtgtgccga atatcacgat	1140
gatttcgcc aatttttgca tatctacagc caagatgtcg cgtgctatgg tgaaaacctg	1200
gcttactttc ccaagggtt catcgagaat atgttctttg tttcagcaaa ccctgggtg	1260
tccttcacgt cgtttgactt gaacgtggcc aatatggata atttcttcgc tccagttttc	1320
accatgggtt agtactatac ccaaggagac aaggctcctt tgccacttgc aatccaagta	1380
caccacgcag tctgcgatgg tttccatgtg ggacgcatgc ttaacgaact ccaacagtac	1440
tgtatgaat ggcaaggcgg cgcgtagccc cccaaccgaa gttgagggga tttttgaatc	1500
ctcggtcccc ctgtccttc cagtcgaaga tagggactct taacatccca aaatctttct	1560
agaagatctc ctacaatatt ctacgtgcc atggaaaatc gatgttcttc ttttattctc	1620
tcaagatttt caggctgtat attaaaactt atattaagaa ctatgctaac cacctcatca	1680

ggaaccgttg taggtggcgt gggttttctt ggcaatcgac tctcatgaaa actacgagct	1740
aaatatccaa tatgttcctc ttgaccaact ttattctgca ttttttttga acgaggttta	1800
gagcaagctt caggaaactg agacaggaat ttattataaa atttaatttt tgaagaaagt	1860
tcagggttaa tagcatccat tttttgcttt gcaagtccct cagcattctt aacaaaagac	1920
gtctcttttg acatgtttta agtttaaac tctgtgtga aattattatc cgctcataat	1980
tccacacatt atacgagccg gaagcataaa gtgtaaagcc tggggtgcct aatgagttag	2040
ctaactcaca ttaattgcgt tgcgtcact gccaatgct ttccagtcgg gaaacctgtc	2100
gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcg	2160
ctcttccgct tctcgtcga ctgactcgtc gcgctcggtc gttcggctgc ggcgagcggc	2220
atcagctcac tcaaaggcgg taatacgggt atccacagaa tcaggggata acgcaggaaa	2280
gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggcgg cgttgcctggc	2340
gtttttccat aggtcctccc cccctgacga gcatcacaaa aatcgacgct caagttagag	2400
gtggcgaaac ccgacaggac tataaagata ccaggcggtt ccccttgga gctccctcgt	2460
gcgctctcct gttccgacct tgccgcttac cggatactg tccgccttcc tcccttcggg	2520
aagcgtggcg ctttctcata gctcacgctg taggtatctc agttcgggtg aggtcgttcg	2580
ctccaagctg ggcgtgtgac acgaaccccc cgttcagccc gaccgtgctg cttatccgg	2640
taactatcgt cttaggtcca acccggttaag acacgactta tcgccactgg cagcagccac	2700
tggtaacagg attagcagag cgaggtatgt aggcgggtgct acagagttct tgaagtggg	2760
gcctaactac ggtacacta gaaggacagt atttggtatc tgcgtctgac tgaagccagt	2820
taccttcgga aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg	2880
tggttttttt gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc	2940
tttgatcttt tctacggggt ctgacgctca gtggaacgaa aactcacgtt aagggtttt	3000
ggtcatgaga ttatcaaaaa ggatcttcac ctagatcctt ttaaattaaa aatgaagttt	3060
ttaatcaatc taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag	3120
tgaggcacct atctcagca tctgtctatt tcgttcatcc atagttgcct gactccccgt	3180
cgtgtagata actacgatac gggagggctt accatctggc ccagtgctg caatgatacc	3240
gcgagaccca cgctaccgg ctcagattt atcagcaata aaccagccag ccggaagggc	3300
cgagcgcaga agtggctctg caactttatc cgcctccatc cagtctatta attgttgccg	3360
ggaagctaga gtaagtagtt cgccagttaa tagtttgcgc aacgttgttg ccattgctac	3420

aggcatcgtg gtgtcacgct cgtcgttttg tatggcttca ttcagctccg gttcccaacg 3480
atcaaggcga gttacatgat ccccatgtt gtgcaaaaaa gcggttagct ctttcggtcc 3540
tccgatcgtt gtcagaagta agttggccgc agtggttatca ctcatggtta tggcagcact 3600
gcataattct cttactgtca tgccatccgt aagatgcttt tctgtgactg gtgagtactc 3660
aaccaagtca ttctgagaat agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat 3720
acgggataat accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc 3780
ttcggggcga aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaaccac 3840

tcgtgcacc aactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa 3900
aacaggaagg caaaatgccg caaaaaaggg aataaggcg acacggaaat gttgaatact 3960
catactcttc cttttcaat attattgaag catttatcag gggtattgtc tcatgagcgg 4020
atacatatgt gaattgtatt agaaaaataa acaaataggg gttccgcgca catttccccg 4080
aaaagtgcc cctgacgtct aagaaacat tattatcatg acattaacct ataaaaatag 4140
gcgtatcacg aggcc 4155

<210> 107

<211> 4155

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetically constructed plasmid pJDI434

<400> 107

gcccctgcag ccgaattata ttatttttgc caataattt ttaacaaaag ctctgaagtc 60
ttcttcattt aaattcttag atgatacttc atctggaaaa ttgtcccaat tagtagcatc 120
acgctgtgag taagttctaa accatttttt tattgttgta ttatctctaa tcttactact 180
cgatgagttt tcggtattat ctctattttt aacttgagc aggttcatt cattgttttt 240
ttcatcatag tgaataaaa caactgcttt aacacttggt cctgaacacc atatccatcc 300
ggcgtaatac gactcactat agggagagcg gccgccagat cttccgatg gctcgagttt 360

ttcagcaaga ttttgggatg ttaagagtcc ctatcttcga catacacata aagtagcttg 420
cgtattttaa attatgaacc taagggttt agcacttcac gctgccgcaa gcactcaggg 480
cgcaagggt gctaaaggaa gcggaacacg tagaaagcca gtccgcagaa acggtgctga 540
ccccgatga atgtcagcta ctgggtatc tggacaaggg aaaacgcaag cgcaaagaga 600
aagcaggtag cttgcagtgg gcttacatgg cgatagctag actgggcggt tttatggaca 660
gcaagcgaac cggaattgcc agctggggcg ccctctggta aggttgggaa gcctgcaaa 720

gtaaactgga tggcttttctt gccgccaagg atctgatggc gcaggggatc aagatctgat 780

caagagacag gatgaggatc gtttcgcatg gagaaaaaga tcacgggcta cactaccgtg 840

gacatctcgc aatggcatcg caaggaacac ttcgaggcat ttcaaagcgt ggcacaatgt 900

acttacaacc agaccgtcca gctggatatt accgcgtttt tgaagaccgt taagaaaaac 960

aagcacaagt tttatccagc ctttatccat attctcgccc gcttgatgaa tgcgcacccc 1020

gaatttcgta tggccatgaa agatgggtgag ctcgttatct gggactcagt ccatccatgc 1080

tataccgttt tcacgaaca aactgaaact ttttcttcgc tgtggtcgga atatcacgat 1140

gatttcgccc aatttttgca tatctacagc caagatgtcg cgtgctatgg tgaaaacctg 1200

gcttactttc ccaagggatt catcgagaat atgttctttg tttcagcaaa cccctgggtg 1260

tccttcacgt cgtttgactt gaacgtggcc aatatggata atttcttcgc tccagttttc 1320

accatgggta agtactatac ccaaggagac aaggctctta tgccacttgc aatccaagta 1380

caccacgcag tctcgatgg tttccatgtg ggacgcatgc ttaacgaact ccaacagtac 1440

tgtgatgaat ggcaaggcgg cgcgtagccc cccaaccgaa gttgagggga tttttgaatc 1500

ctcggtcccc cttgtccttc cagtcgattg gggatatgggg ggggcgggtca aaatctttct 1560

agaagatctc ctacaatatt ctacgtgcc atggaaaatc gatgttcttc ttttattctc 1620

tcaagatttt caggctgtat attaaaactt atattaagaa ctatgctaac cacctcatca 1680

ggaaccgttg taggtggcgt gggttttctt ggcaatcgac tctcatgaaa actacgagct 1740

aaatatcaaa tatgttcttc ttgaccaact ttattctgca ttttttttga acgaggttta 1800

gagcaagctt caggaaactg agacaggaat ttatttaaaa atttaaatit tgaagaaagt 1860

tcagggttaa tagcatccat tttttgcttt gcaagttcct cagcattctt aacaaaagac 1920

gtctcttttg acatgtttta agtttaaac tctgtgtga aattattatc cgtcataat 1980

tccacacatt atacgagccg gaagcataaa gtgtaaagcc tggggtgcct aatgagttag 2040

ctaactcaca ttaattgctg tgcgtcact gccaatgct ttccagtcgg gaaacctgtc 2100

gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcg 2160

ctcttccgct tctcgtca ctgactcgtc gcgctcggtc gttcggctgc ggcgagcgg 2220

atcagctcac tcaaaggcgg taatacgtt atccacagaa tcaggggata acgcaggaaa 2280

gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc 2340

gtttttccat aggtctcgcc cccctgacga gcatcacaaa aatcgacgt caagttagag 2400

gtggcgaaac ccgacaggac tataaagata ccaggcgttt cccctggaa gctccctcgt 2460

gcgctctcct gttccgaccc tgccgcttac cggatacctg tccgcctttc tcccttcggg 2520
aagcgtggcg ctttctcata gctcacgctg taggtatctc agttcgggtg aggtcgttcg 2580
ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg ccttatccgg 2640
taactatcgt cttagagcca acccggttaag acacgactta tcgccactgg cagcagccac 2700
tggtaacagg attagcagag cgaggtatgt aggcgggtgct acagagttct tgaagtgggtg 2760
gcctaactac ggctacacta gaaggacagt atttgggtatc tgcgctctgc tgaagccagt 2820
taccttcgga aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg 2880

tggttttttt gtttgaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc 2940
tttgaatctt tctacggggg ctgacgctca gtggaacgaa aactcacgtt aagggttttt 3000
ggatcatgaga ttatcaaaaa ggatcttcac ctatgacctt ttaaattaaa aatgaagttt 3060
taaatcaatc taaagtatat atgagtaaag ttgggtctgac agttaccaat gcttaatcag 3120
tgaggcacct atctcagcga tctgtctatt tcgttcatcc atagtgcct gactccccgt 3180
cgtgtagata actacgatac gggaggggctt accatctggc cccagtgtcg caatgatacc 3240
gcgagaccca cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaagggc 3300

cgagcgcaga agtggctctg caactttatc cgctccatc cagtctatta attgttgccg 3360
ggaagctaga gtaagtagtt cgccagttaa tagtttgccg aacgttgttg ccattgctac 3420
aggcatcgtg gtgtcacgct cgctgttttg tatggcttca ttcagctccg gttcccaacg 3480
atcaaggcga gttacatgat ccccatgtt gtgcaaaaaa gcggttagct ccttcgggtcc 3540
tccgatcgtt gtcagaagta agttggccgc agtgttatca ctcatggtta tggcagcact 3600
gcataattct ctactgtca tgccatccgt aagatgcttt tctgtgactg gtgagtactc 3660
aaccaagtca ttctgagaat agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat 3720

acgggataat accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc 3780
ttcggggcga aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaaccac 3840
tcgtgcaccc aactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa 3900
aacaggaagg caaaatgccg caaaaaaggg aataagggcg acacggaaat gttgaatact 3960
catactcttc ctttttcaat attattgaag catttatcag gggtattgtc tcatgagcgg 4020
atacatattt gaatgtattt agaaaaataa acaaataggg gttccgcgca catttccccg 4080
aaaagtgcc cctgacgtct aagaaacat tattatcatg acattaacct ataaaaatag 4140

gcgtatcacg aggcc 4155

<210> 108

<211> 4157

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetically constructed plasmid pJDI435

<400> 108

```

gccctgcag ccgaattata ttatTTTTgc caaataattt ttaacaaaag ctctgaagtc      60
ttcttcattt aaattcttag atgatacttc atctggaaaa ttgtcccaat tagtagcatc      120
acgctgtgag taagttctaa accatttttt tattgttgta ttatctctaa tcttactact      180
cgatgagttt tcggtattat ctctatTTTT aacttggagc aggttccatt cattgttttt      240

ttcatcatag tgaataaaat caactgcttt aacacttggt cctgaacacc atatccatcc      300
ggcgtaatat gactcactat agggagagcg gccgccagat cttccggatg gctcagagttt      360
ttcagcaaga ttttgaggag tggtcagtct ccgtgaactc gacatacaca taaagtagct      420
tgcgtattta aaattatgaa cctaaggggt ttagcacttc acgctgccgc aagcactcag      480
ggcgcaaggg ctgctaaagg aagcggaaca cgtagaaagc cagtccgcag aaacggtgct      540
gaccccgcat gaatgtcagc tactgggcta tctggacaag ggaaaacgca agcgcaaaga      600
gaaagcaggt agcttgcagt gggcttacat ggcgatagct agactgggcg gttttatgga      660

cagcaagcga accggaattg ccagctgggg cgccctctgg taaggttggg aagccctgca      720
aagtaaaactg gatggctttc ttgccccaa ggatctgatg gcgcagggga tcaagatctg      780
atcaagagac aggatgagga tcgtttcgca tggagaaaaa gatcacgggc tacactaccg      840
tggacatctc gcaatggcat cgcaaggaac acttcgaggc atttcaaagc gtggcacaat      900
gtacttacaa ccagaccgtc cagctggata ttaccgcgtt tttgaagacc gttaagaaaa      960
acaagcacia gttttatcca gcctttatcc atattctcgc ccgcttgatg aatgcgcacc     1020
ccgaatttcg tatggccatg aaagatgggt agctcggtat ctgggactca gtccatccat     1080

gctataccgt tttccacgaa caaactgaaa ctttttcttc gctgtgttcc gaatatcacg     1140
atgatttcgg ccaatttttg catatctaca gccaagatgt cgcgtgctat ggtgaaaacc     1200
tggttacttt tccaaggga ttcatcgaga atatgttctt tgtttcagca aaccctggg      1260
tgtcttctac gtcgtttgac ttgaacgtgg ccaatatgga taatttcttc gctccagttt     1320
tcaccatggg taagtactat acccaaggag acaaggctct tatgccactt gcaatccaag     1380
tacaccacgc agtctgcgat ggtttccatg tgggacgcat gcttaacgaa ctccaacagt     1440
actgtgatga atggcaagcg ggcgcgtagc cccccaaccg aagttgaggg gatttttgaa     1500

tcctcgggtc cccttgctct tccagtcgaa gatagggact cttaacatcc caaaatcttt     1560

```

ctagaagatc tctacaata ttctcagctg ccatggaaaa tcgatgttct tcttttattc 1620
 tctcaagatt ttcaggctgt atattaaaac ttatattaag aactatgcta accacctcat 1680
 caggaaccgt tgtagggtgc gtgggttttc ttggcaatcg actctcatga aaactacgag 1740
 ctaaattatc aatagtgtcc tcttgaccaa ctttattctg cttttttttt gaacgaggtt 1800
 tagagcaagc ttcaggaaac tgagacagga attttattaa aaatttaaat tttgaagaaa 1860
 gttcagggtt aatagcatcc attttttgct ttgcaagtcc ctcagcatc ttaacaaaag 1920

 acgtctcttt tgacatgttt aaagttaaa cctcctgtgt gaaattatta tccgctcata 1980
 attccacaca ttatacgagc cggaagcata aagtgtaaag cctggggtgc ctaatgagt 2040
 agctaactca cattaattgc gttgcgctca ctgccaattg ctttccagtc gggaaacctg 2100
 tcgtgccagc tgcattaatg aatcgcccaa cgcgcgggga gaggcggttt gcgtattggg 2160
 cgctcttccg cttcctcgct cactgactcg ctgcgctcgg tcgttcggtc gggcgagcgc 2220
 gttatcagtc actcaaagc ggtaatacgg ttatccacag aatcagggga taacgcagga 2280
 aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg 2340

 gcgtttttcc ataggtccg cccccctgac gagcatcaca aaaatcgacg ctcaagtcag 2400
 aggtggcgaa acccgacagg actataaaga taccaggcgt tttcccctgg aagctccctc 2460
 gtgcgctctc ctgttccgac cctgccgtt accggatacc tgtccgctt tctcccttcg 2520
 ggaagcgtgg cgcttttca tagctcacgc ttaggtatc tcagttcgtt gtaggtcgtt 2580
 cgctccaagc tgggctgtgt gcacgaacc cccgttcagc ccgaccgtg cgccttatcc 2640
 ggtaactatc gtcttgagtc caaccggta agacacgact tatcgccact ggcagcagcc 2700
 actggtaaca ggattagcag agcgaggatg gtaggcggtg ctacagagtt ctgaagtgg 2760

 tggcctaact acggctacac tagaaggaca gtatttggtg tctgcgctct gctgaagcca 2820
 gttaccttcg gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc 2880
 ggtggttttt ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat 2940
 cctttgatct tttctacggg gtctgacgct cagtggaaacg aaaactcacg ttaagggtt 3000
 ttggtcatga gattatcaaa aaggatcttc acctagatcc ttttaatta aaaatgaagt 3060
 tttaaatcaa tctaaaglat atatgagtaa acttggtctg acagttacca atgcttaatc 3120
 agtgaaggac ctatctcagc gatctgtcta tttcgttcat ccatagttgc ctgactcccc 3180

 gtcgtgtaga taactacgat acgggagggc tiaccatctg gccccagtgc tgcaatgata 3240
 ccgcgagacc cacgtcacc ggctccagat ttatcagcaa taaaccagcc agccggaagg 3300
 gccgagcgca gaagtgttc tgcaacttta tccgctcca tccagtctat taattgttgc 3360
 cgggaagcta gagtaagtag ttcgccagt taaagtgtgc gcaacgttgt tgccattgct 3420

acaggcatcg tgggtgtcacg ctctgtcttt ggatggctt cattcagctc cggttcccaa 3480
cgatcaaggc gaggttacatg atcccccatg ttgtgcaaaa aagcggttag ctcttcggt 3540
cctccgatcg ttgtcagaag taagttggcc gcagtgttat cactcatggt tatggcagca 3600

ctgcataatt ctcttactgt catgccatcc gtaagatgct tttctgtgac tggtagtac 3660
tcaaccaagt cattctgaga atagtgtatg cggcgaccga gttgctcttg cccggcgta 3720
atacgggata ataccgcgcc acatagcaga actttaaaag tgctcatcat tggaaaacgt 3780
tcttcggggc gaaaactctc aaggatctta ccgctgttga gatccagttc gatgtaaccc 3840
actcgtgcac ccaactgac ttacagcatct tttactttca ccagcgttcc tgggtgagca 3900
aaaacaggaa ggcaaaatgc cgcaaaaaag ggaataaggg cgacacggaa atgttgaata 3960
ctcatactct tcctttttca atattattga agcatttacc agggttattg tctcatgagc 4020

ggatacatat ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc 4080
cgaaaagtgc cactgacgt ctaagaaacc attattatca tgacattaac ctataaaaat 4140
aggcgatatca cgaggcc 4157

<210> 109
<211> 4159
<212> DNA
<213> Artificial Sequence
<220>
<223> Synthetically constructed plasmid pJDI436
<400> 109

gcccctgcag ccgaattata ttatttttgc caataattt ttaacaaaag ctctgaagtc 60
ttcttcattt aaattcttag atgatacttc atctggaaaa ttgtcccaat tagtagcatc 120

acgctgtgag taagttctaa accatttttt tattgttgta ttatctctaa tcttactact 180
cgatgagttt tcggtattat ctctattttt aacttgagc aggttccatt cattgttttt 240
ttcatcatag tgaataaaat caactgcttt aacacttggt cctgaacacc atatecatcc 300
ggcgtaatac gactcactat agggagagcg gccgccagat cttccggatg gctcgagttt 360
ttcagcaaga ttttgaggag tgttcagtct ccgtgaactc gacatacaca taaagtagct 420
tgcgtattta aaattatgaa cctaaggggt ttagcacttc acgtgccgc aagcactcag 480
ggcgcaaggg ctgctaaagg aagcggaaca cgtagaaagc cagtccgcag aaacggtgct 540

gaccccggtat gaatgtcagc tactgggcta tctggacaag ggaaaacgca agcgcaaaga 600
gaaagcaggt agcttgacgt gggcttacat ggcgatagct agactgggcg gttttatgga 660

cagcaagcga accggaattg ccagctgggg cgccctctgg taaggttggg aagccctgca	720
aagtaaaactg gatggctttc ttgccgcaa ggatctgatg gcgcagggga tcaagatctg	780
atcaagagac aggatgagga tcgtttcgca tggagaaaaa gatcacgggc tacactaccg	840
tggacatctc gcaatggcat cgcaaggaac acttcgaggc atttcaaagc gtggcacaat	900
gtacttaciaa ccagaccgtc cagctggata ttaccgcgtt tttgaagacc gttaagaaaa	960
acaagcaciaa gttttatcca gcctttatcc atattctcgc ccgcttgatg aatgcgcacc	1020
ccgaatttcg tatggccatg aaagatggtg agctcgttat ctgggactca gtccatccat	1080
gctataccgt tttccacgaa caaactgaaa ctttttcttc gctgtgttcc gaatatcacg	1140
atgatttcgg ccaatttttg catacttaca gccaaagtgt cgcgtgctat ggtgaaaacc	1200
tggcttactt tcccaaggga ttcatcgaga atatgttctt tgtttcagca aaccctggg	1260
tgtctttcac gtcttttgac ttgaacgtgg ccaatatgga taatttcttc gctccagttt	1320
tcaccatggg taagtactat acccaaggag acaaggtcct tatgccactt gcaatccaag	1380
tacaccacgc agtctgcgat ggtttccatg tgggacgcat gcttaacgaa ctccaacagt	1440
actgtgatga atggcaaggc ggcgctagc cccccaaccg aagttgaggg gatttttgaa	1500
tcctcggctc cccttgtct tccagtcgat tggggtatgg gggggcggt caaaaaatct	1560
ttctagaaga tctcctaciaa tattctcagc tgccatggaa aatcgatgtt cttcttttat	1620
tctctcaaga ttttcaggct gtatatataa acttatatta agaactatgc taaccacctc	1680
atcaggaacc gttgtaggtg gcgtgggttt tcttgcaat cgactctcat gaaaactacg	1740
agctaaatat tcaatatgtt cctcttgacc aactttattc tgcatttttt ttgaacgagg	1800
tttagagcaa gcttcaggaa actgagacag gaattttatt aaaaatttaa attttgaaga	1860
aagttcaggg ttaatagcat ccattttttg ctttgcaagt tcctcagcat tcttaacaaa	1920
agacgtctct tttgacatgt ttaaagttta aacctcctgt gtgaaattat tatccgctca	1980
taattccaca cattatacga gccggaagca taaagtgtaa agcctggggg gcctaatgag	2040
tgagctaact cacattaatt gcgttgcgct cactgccaat tgctttccag tcgggaaacc	2100
tgtcgtgcca gctgcattaa tgaatcgcc aacgcgcggg gagaggcggt ttgcgtattg	2160
ggcgctcttc cgcttcctcg ctactgact cgctgcgctc ggtcgttcgg ctgcgcgag	2220
cggtatcagc tcaactcaaag gcggtatcac ggttatccac agaatcaggg gataacgcag	2280
gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc	2340
tggcgttttt ccataggctc cgccccctg acgagcatca caaaaatcga cgctcaagtc	2400
agagggtggc aaacccgaca ggactataaa gataccaggc gtttccccct ggaagctccc	2460
tcgtgcgctc tctgttccg accctgccgc ttaccggata cctgtccgcc tttctccctt	2520

cggaagcgt ggcgctttct catagctcac gctgtaggta tctcagttcg gtgtaggtcg 2580
 ttcgctccaa gctgggctgt gtgcacgaac ccccggttca gcccgaccgc tgcgccttat 2640

ccggtaaacta tcgtcttgag tccaacccgg taagacacga cttatcgcca ctggcagcag 2700
 ccactggtaa caggattagc agagcgaggt atgtaggcgg tgctacagag ttcttgaagt 2760
 ggtggcctaa ctacggctac actagaagga cagtatttgg tatctgcgct ctgctgaagc 2820
 cagttacctt cgaaaaaga gttgtagct cttgatccgg caaacaacc accgctggt 2880
 gcggtggttt ttttgttgc aagcagcaga ttacgcgcag aaaaaagga tctcaagaag 2940
 atcctttgat cttttctacg ggtctgacg ctacgtggaa cgaaaactca cgtaaggga 3000
 ttttggtcat gagattatca aaaaggatct tcacctagat ctttttaaat taaaaatgaa 3060

gttttaaatc aatctaaagt atatatgagt aaacttggc tgacagttac caatgcttaa 3120
 tcagttaggc acctatctca gcgatctgtc tatttcgttc atccatagtt gcctgactcc 3180
 ccgtctgtga gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga 3240
 taccgcgaga cccacgtca ccggtccag atttatcagc aataaaccag ccagccggaa 3300
 gggccgagcg cagaagtgtt cctgcaactt tatccgcctc catccagtct attaattgtt 3360
 gccgggaagc tagagtaagt agttcgccag ttaatagttt gcgcaacgtt gttgccattg 3420
 ctacaggcat cgtggtgtca cgctcgtcgt ttggtatggc ttatttcagc tccggttccc 3480

aacgatcaag gcgagttaca tgatcccca tgttgtgcaa aaaagcggtt agctccttcg 3540
 gtcctccgat cgttgtcaga agtaagttgg ccgcagtgtt atcactcatg gttatggcag 3600
 cactgcataa ttctcttact gtcatgceat ccgtaagatg cttttctgtg actggtgagt 3660
 actcaaccaa gtcatcttga gaatagtga tgccgacgacc gagttgctct tgcggcgct 3720
 caatacggga taataccgcg ccacatagca gaactttaaa agtgctcatc attggaaaac 3780
 gttcttcggg gcgaaaactc tcaaggatct taccgtgtt gagatccagt tcgatgtaac 3840
 ccactcgtgc acccaactga tcttcagcat cttttacttt caccagcgtt tctgggtgag 3900

caaaaacagg aaggcaaat gccgcaaaa aggaataag ggcgacacgg aaatgttgaa 3960
 tactcatact cticcttttt caatattatt gaagcattta tcagggttat tgtctcatga 4020
 gcggatacat atttgaatgt atttagaaaa ataaacaaat aggggttccg cgcacatttc 4080
 cccgaaaagt gccacctgac gtctaagaaa ccattattat catgacatta acctataaaa 4140
 ataggcgtat cacgaggcc 4159

<210> 110
 <211> 41
 <212> RNA

<213> Artificial Sequence

<220>

<223> Synthetically constructed RNA targeting cTAG M

<400> 110

uaauuucuaac ucuuguagau gguuaaagau gguuaaauga u 41

<210> 111

<211> 41

<212> RNA

<213> Artificial Sequence

<220>

<223> Synthetically constructed RNA targeting cTAG N

<400> 111

uaauuucuaac ucuuguagau ggauguuaag agucccuauc u 41

<210> 112

<211> 41

<212> RNA

<213> Artificial Sequence

<220>

<223> Synthetically constructed RNA targeting cTAG O

<400> 112

uaauuucuaac ucuuguagau accgcccccc ccuacccca a 41

<210> 113

<211> 43

<212> RNA

<213> Artificial Sequence

<220>

<223> Synthetically constructed RNA targeting cTAG P

<400> 113

uaauuucuaac ucuuguagau aggaguguuc agucuccgug aac 43