



**(19) 대한민국특허청(KR)**  
**(12) 공개특허공보(A)**

(11) 공개번호 10-2017-0087521  
(43) 공개일자 2017년07월28일

- |   |   |
|---|---|
| <p>(51) 국제특허분류(Int. Cl.)<br/>C12N 15/80 (2006.01) C12N 15/10 (2017.01)<br/>C12N 15/113 (2010.01) C12N 9/22 (2006.01)</p> <p>(52) CPC특허분류<br/>C12N 15/80 (2013.01)<br/>C12N 15/102 (2013.01)</p> <p>(21) 출원번호 10-2017-7019683</p> <p>(22) 출원일자(국제) 2015년12월15일<br/>심사청구일자 없음</p> <p>(85) 번역문제출일자 2017년07월14일</p> <p>(86) 국제출원번호 PCT/US2015/065693</p> <p>(87) 국제공개번호 WO 2016/100272<br/>국제공개일자 2016년06월23일</p> <p>(30) 우선권주장<br/>PCT/CN2014/093914 2014년12월16일 중국(CN)<br/>(뒷면에 계속)</p> | <p>(71) 출원인<br/>다니스코 유에스 인크.<br/>미합중국 캘리포니아 (우편번호 94304) 팔로 알토 페이지 밀 로드 925</p> <p>(72) 발명자<br/>바우어 벤자민 에스<br/>미국 94304 캘리포니아주 팔로 알토 페이지 밀 로드 925<br/>첸 지미<br/>미국 94304 캘리포니아주 팔로 알토 페이지 밀 로드 925<br/>(뒷면에 계속)</p> <p>(74) 대리인<br/>특허법인코리아나</p> |
|---|---|

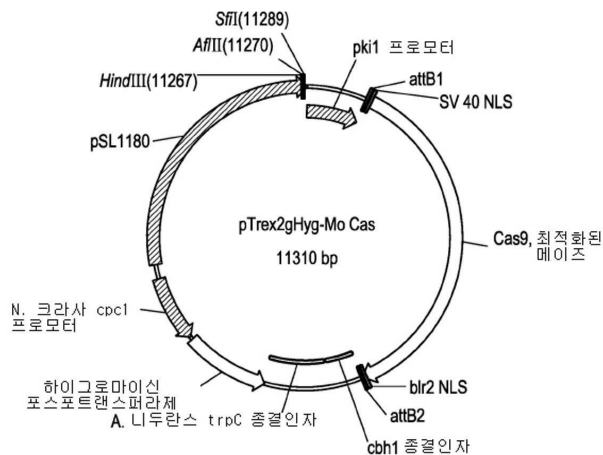
전체 청구항 수 : 총 34 항

(54) 발명의 명칭 **진균 게놈 변형 시스템 및 사용 방법**

**(57) 요약**

사상균 세포의 게놈 내 표적 부위에서 게놈 변형을 위한 조성물 및 방법이 제공된다. 이 방법 및 조성물은 표적 부위를 변형하거나 변경하기 위한 가이드 폴리뉴클레오티드/Cas 엔도뉴클레아제 시스템에 관한 것이다. 변형된 사상균이 결합있는 비상동성 말단 결합 경로를 갖는 양태도 제공된다.

**대표도** - 도2



(52) CPC특허분류

**C12N 15/113** (2013.01)

**C12N 9/22** (2013.01)

(72) 발명자

**거 정**

미국 94304 캘리포니아주 팔로 알토 페이지 밀 로 드 925

**구 샤오강**

미국 94304 캘리포니아주 팔로 알토 페이지 밀 로 드 925

**마드리드 수잔 맘푸스티**

미국 94304 캘리포니아주 팔로 알토 페이지 밀 로 드 925

**송 단평**

미국 94304 캘리포니아주 팔로 알토 페이지 밀 로 드 925

**송 밍민**

미국 94304 캘리포니아주 팔로 알토 페이지 밀 로 드 925

**워드 마이클**

미국 94304 캘리포니아주 팔로 알토 페이지 밀 로 드 925

(30) 우선권주장

PCT/CN2014/093916 2014년12월16일 중국(CN)

PCT/CN2014/093918 2014년12월16일 중국(CN)

## 명세서

### 청구범위

#### 청구항 1

사상균 세포의 게놈좌와 공여체 DNA를 상동성 재조합하기 위한 방법으로서: 상기 방법은:

a) 사상균 세포의 게놈좌에 대한 상동성을 가진 도메인을 포함하는 Cas 엔도뉴클레아제, 가이드 RNA, 및 공여체 DNA를 사상균 세포군에 도입하는 단계로서, Cas 엔도뉴클레아제 및 가이드 RNA는 Cas 엔도뉴클레아제가 사상균 세포의 게놈좌 내 또는 사상균 세포의 게놈좌 근처의 표적 부위에서 작용할 수 있게 하는 복합체를 형성할 수 있는 단계; 및

b) 게놈좌와 공여체 DNA의 상동성 재조합이 나타난 군으로부터 적어도 하나의 사상균 세포를 확인하는 단계를 포함하고,

Cas 엔도뉴클레아제, 가이드 RNA 또는 둘 모두는 사상균 세포군으로 일시적으로 도입되는, 방법.

#### 청구항 2

제1항에 있어서, 사상균 세포들 내 표적 부위에서 비상동성 말단 결합(NHEJ) 메커니즘은 활성화되지 않거나, 비기능적이거나 감소되는, 방법.

#### 청구항 3

제2항에 있어서, 사상균 세포들 내 비상동성 말단 결합(NHEJ) 경로는 하나 이상의 비기능적 또는 활성-감소 성분들을 포함하는, 방법.

#### 청구항 4

제3항에 있어서, 하나 이상의 비기능적 또는 활성-감소 성분은 ku80, ku70, rad50, mre11, xrs2, lig4, xrs, 및 이의 조합들로 구성되는 군으로부터 선택되는, 방법.

#### 청구항 5

제4항에 있어서, 하나 이상의 비기능적 또는 활성-감소 성분은 ku80인, 방법.

#### 청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서, Cas 엔도뉴클레아제는 Cas 니카아제인, 방법.

#### 청구항 7

제1항 내지 제5항 중 어느 한 항에 있어서, Cas 엔도뉴클레아제는 Cas9 엔도뉴클레아제 또는 이의 변이체인, 방법.

#### 청구항 8

제7항에 있어서, Cas9 엔도뉴클레아제 또는 이의 변이체는 스트렙토코쿠스 종(*Streptococcus* sp.), *S. 파이오제네스*(*pyogenes*), *S. 뮤탄스*(*mutans*) 및 *S. 더모필루스*(*thermophilus*), 캄피로박테르 종(*Campylobacter* sp.), *C. 제주니*(*jejuni*), 나이세리아 종(*Neisseria* sp.), *N. 메닝기티데스*(*meningitides*), 프란시셀라 종(*Francisella* sp.), *F. 노비시다*(*novicida*), 및 파스투렐라 종(*Pasteurella* sp.), *P. 물토시다*(*multocida*)로 구성되는 군으로부터 선택되는 종으로부터 전장 Cas9 또는 이의 기능적 단편을 포함하는, 방법.

#### 청구항 9

제8항에 있어서, Cas9 엔도뉴클레아제 또는 이의 변이체는 서열 번호 45 및 서열 번호 48 내지 53 중 어느 하나와 적어도 70% 동일성을 갖는 아미노산 서열을 포함하는, 방법.

**청구항 10**

제1항 내지 제9항 중 어느 한 항에 있어서, 공여체 DNA는 관심 폴리뉴클레오티드 서열을 포함하고 계놈좌에서 상동성 재조합은 계놈좌에서 관심 폴리뉴클레오티드 서열의 삽입을 초래하는, 방법.

**청구항 11**

제1항 내지 제10항 중 어느 한 항에 있어서, 상기 도입 단계는 Cas 엔도뉴클레아제용 발현 카세트를 포함하는 DNA 작제물을 사상균 세포들 내로 도입하는 단계를 포함하는, 방법.

**청구항 12**

제1항 내지 제11항 중 어느 한 항에 있어서, 상기 도입 단계는 가이드 RNA용 발현 카세트를 포함하는 DNA 작제물을 사상균 세포들 내로 도입하는 단계를 포함하는, 방법.

**청구항 13**

제1항 내지 제12항 중 어느 한 항 또는 제48항 내지 제54항 중 어느 한 항에 있어서, 상기 도입 단계는 선택 가능 마커를 인코딩하는 서열을 포함하는 DNA 작제물을 사상균 세포들 내로 도입하는 단계를 포함하는 방법.

**청구항 14**

제13항에 있어서, DNA 작제물은 선택 가능 마커를 인코딩하는 서열 및 공여체 DNA를 모두 포함하는, 방법.

**청구항 15**

제1항 내지 제14항 중 어느 한 항에 있어서, 상기 도입 단계는 Cas 엔도뉴클레아제를 인코딩하는 서열, 가이드 RNA를 인코딩하는 서열, 선택가능 마커를 인코딩하는 서열, 및 공여체 DNA를 포함하는 DNA 작제물을 사상균 세포들 내로 도입하는 단계를 포함하는, 방법.

**청구항 16**

제11항 내지 제15항 중 어느 한 항에 있어서, DNA 작제물은 선형 DNA 작제물인, 방법.

**청구항 17**

제11항 내지 제15항 중 어느 한 항에 있어서, DNA 작제물은 원형 DNA 작제물인, 방법.

**청구항 18**

제11 항 및 제15항 내지 제17항 중 어느 한 항에 있어서, Cas 엔도뉴클레아제용 발현 카세트 또는 Cas 엔도뉴클레아제를 인코딩하는 서열은 사상균 세포에서 발현을 위해 최적화된 Cas 코딩 서열을 포함하는, 방법.

**청구항 19**

제18항에 있어서, Cas 코딩 서열은 서열 번호 44와 적어도 70% 동일한 폴리뉴클레오티드 서열을 포함하는 Cas9 코딩 서열인, 방법.

**청구항 20**

제1항 내지 제10항, 제12항 내지 제14항, 및 제16항 및 제17항 중 어느 한 항에 있어서, 상기 도입 단계는 Cas 엔도뉴클레아제를 사상균 세포들 내로 직접 도입하는 단계를 포함하는, 방법.

**청구항 21**

제1항 내지 제11항, 제13항, 제14항, 및 제16항 내지 제20항 중 어느 한 항에 있어서, 상기 도입 단계는 가이드 RNA를 사상균 세포들 내로 직접 도입하는 단계를 포함하는, 방법.

**청구항 22**

제1항 내지 제21항 중 어느 한 항에 있어서, Cas 엔도뉴클레아제는 핵 위치 신호에 작동 가능하게 연결되는, 방

법.

**청구항 23**

제1항 내지 제22항 중 어느 한 항에 있어서, 사상균 세포는 유미코티나 또는 폐지조미코티나 사상균 세포인, 방법.

**청구항 24**

제1항 내지 제23항 중 어느 한 항에 있어서, 사상균 세포는 트리코테르마, 페니실리움, 아스퍼질러스, 휴미콜라, 크리소스포리움, 푸사리움, 마이셀리오프토라, 뉴로스포라, 하이포크레아, 및 에메리셀라로 구성되는 군으로부터 선택되는, 방법.

**청구항 25**

제1항 내지 제24항 중 어느 한 항에 있어서, 표적 부위는 오픈 리딩 프레임, 프로모터, 조절 서열, 종결인자 서열, 조절 요소 서열, 스플라이스 부위, 코딩 서열, 폴리유비퀴티네이션 부위, 인트론 부위, 및 인트론 증진 모티프로 구성되는 군으로부터 선택되는 관심 유전자의 영역 내에 위치되는, 방법.

**청구항 26**

제1항 내지 제25항 중 어느 한 항에 있어서, 상동성 재조합은 표적 부위에서 또는 표적 부위 근처에서 DNA 서열의 변형을 초래하고, 상기 변형은 하나 이상의 뉴클레오티드의 결실, 하나 이상의 뉴클레오티드의 삽입, 관심 단백질을 인코딩하는 발현 카세트의 삽입, 하나 이상의 뉴클레오티드의 치환, 및 이들의 임의 조합으로 구성되는 군으로부터 선택되는, 방법.

**청구항 27**

제1항 내지 제26항 중 어느 한 항에 있어서, 상기 확인 단계는 상동성 재조합 또는 변형을 선택하거나 선별하는 조건 하에서 단계 (a)로부터 세포군을 배양하는 단계를 포함하는, 방법.

**청구항 28**

제1항 내지 제27항 중 어느 한 항에 있어서, 상기 확인 단계는 불안정한 형질전환체를 선별하는 조건 하에서 단계 (a)로부터 세포군을 배양하는 단계를 포함하는, 방법.

**청구항 29**

제1항 내지 제28항 중 어느 한 항에 있어서, 상기 도입 단계는 선택가능 마커를 인코딩하는 서열 및 공여체 DNA를 포함하는 DNA 작제물을 사상균 세포들 내로 도입하는 단계를 포함하고, 상기 확인 단계는 선택가능 마커를 잃었지만 아직은 공여체 DNA를 보유하고 있는 불안정한 형질전환체를 선별하는 조건 하에서 단계 (a)로부터 세포군을 배양하는 단계를 포함하는, 방법.

**청구항 30**

제1항 내지 제29항 중 어느 한 항의 방법으로 생산된 재조합 사상균 세포.

**청구항 31**

Cas 엔도뉴클레아제용 발현 카세트를 포함하는 제1 재조합 DNA 작제물을 포함하는 재조합 사상균 세포.

**청구항 32**

제30항 또는 제31항에 있어서, 재조합 사상균 세포는 NHEJ 경로에 하나 이상의 비기능성 성분 또는 활성-감소 성분을 포함하는, 재조합 사상균 세포.

**청구항 33**

제32항에 있어서, NHEJ 경로의 하나 이상의 성분은 ku80, ku70, rad50, mre11, xrs2, lig4, 및 xrs로 구성되는 군으로부터 선택되는, 재조합 사상균 세포.

**청구항 34**

제30항 내지 제33항 중 어느 한 항에 있어서, 관심 폴리뉴클레오티드를 포함하는 공여체 DNA를 더 포함하는, 재조합 사상균 세포.

**발명의 설명**

**기술 분야**

[0001] **관련 출원에 대한 상호 참조**

[0002] 본 출원은 모두 2014년 12월 16일에 출원된 PCT 특허 출원 번호 PCT/CN2014/093914, PCT/CN2014/093916, 및 PCT/CN2014/093918에 대한 우선권을 주장하며, 이들의 전체는 참고로 본원에 통합된다.

[0003] **서열 목록**

[0004] 37 C.F.R. § 1.52(e)에 따라 EFS를 통하여 제출된 서열 목록은 본원에 참고로 통합된다. EFS를 통하여 제출된 서열 목록 텍스트 파일은 2015년 12월 3일에 생성되고, 크기가 165 킬로바이트인 파일 "40532-WO-PCT-4\_2015-832 final\_ST25.txt"를 포함한다.

**배경 기술**

[0005] 박테리아와 고세균은, 서열 특이적 방식으로 DNA에 이중 가닥 절단을 도입할 수 있는 무리 지어진 규칙적 공간 사이 짧은 팔린드로믹 반복부들(clustered regularly interspaced short palindromic repeats: CRISPR)/CRISPR-관련(Cas) 시스템으로 불리는 적응적 면역 방어를 진화시켰다. Cas 시스템은 짧은 RNA 서열(tracrRNA 및 crRNA) 및 (가변 표적화 도메인으로 불리는 crRNA의 어떤 부분에 대한 상동성을 통해) 특정 DNA 서열을 표적으로 하는 RNA 의존성 엔도뉴클레아제(Cas 엔도뉴클레아제)를 포함하는 리보뉴클레오티드 단백질 복합체의 활성을 통해 그 기능을 수행하여 표적에서 이중 가닥 절단을 일으킨다. CRISPR 좌는 유사하게 산재된 짧은 서열 반복부들을 갖는 대장균에서 처음 인식되었는데(Ishino 등 (1987) J. Bacteriol. 169:5429-5433; Nakata 등 (1989) J. Bacteriol. 171:3553-3556), 이들 반복부들은 할로페락스 메디테라네이(Haloferax mediterranei), 스트렙토코쿠스 파이오제네스(Streptococcus pyogenes), 아나배나(Anabaena), 및 마이코박테리움 투베르쿨로시스(Mycobacterium tuberculosis)를 포함하지만, 이들에 한정되지 않는 수 많은 세균종에서 나중에 확인되었다(Groenen 등 (1993) Mol. Microbiol. 10:1057-1065; Hoe 등 (1999) Emerg. Infect. Dis. 5:254-263; Masepohl 등 (1996) Biochim. Biophys. Acta 1307:26-30; Mojica 등 (1995) Mol. Microbiol. 17:85-93).

[0006] 게놈 DNA의 특정 표적 부위에서 절단 유도는 그 부위에서 또는 그 부위 근처에서 변형을 도입하기 위해 사용될 수 있다고 잘 알려져있다. 예컨대, 유전자 표적화를 위한 상동성 재조합은 표적이 된 DNA 부위가 이중 가닥 절단을 포함할 때 향상되는 것으로 나타났다(예컨대, Rudin 등, Genetics 122:519-534; Smih 등, Nucl. Acids Res. 23:5012-5019 참조). Cas 시스템의 부위-특정 성질을 고려할 때, 포유류 세포를 포함하여, 이러한 시스템에 기반한 게놈 변형/유전자조작 기술이 설명되었다(예컨대, Hsu 등.; Cell vol. 157, p1262-1278, 5 June 2014 entitled "Development and Applications of CRISPR-Cas9 for Genome Engineering" 참조). Cas-기반 게놈 유전자조작의 힘은 crRNA의 DNA-표적화 영역(가변 표적화 도메인)이 게놈의 원하는 표적 부위에 상동성이어서 이를 (임의의 편리한 수단을 통해) Cas 엔도뉴클레아제와 숙주 세포의 기능성 복합체 내로 결합한 재조합 crRNA를 설계하여 복합체 게놈 내 사실상 모든 특정 위치를 표적화하는 능력에서 온다.

[0007] Cas-기반 게놈 유전자조작 기술이 수 많은 상이한 숙주 세포 유형에 적용되었지만, 진균 세포에서 이러한 시스템의 효율적인 사용은 어려운 것으로 판명되었다. 따라서, 효율적이고 효과적인 Cas-기반 게놈 유전자조작 방법 및 진균 세포의 게놈 표적 부위를 변형/변경하기 위한 조성물을 개발하기 위한 필요성이 여전히 남아있다.

**발명의 내용**

[0008] 진균 세포, 예컨대 사상균 세포 내 게놈좌와 공여체 DNA의 상동성 재조합을 촉진하기 위한 가이드 RNA/Cas 엔도뉴클레아제 시스템을 채용한 조성물 및 방법이 제공된다.

[0009] 본 개시의 양태들은 진균 세포 내 게놈좌와 공여체 DNA의 상동성 재조합을 위한 방법에 관한 것이다. 일부 구현예에서, 이 방법은 a) 진균 세포의 게놈좌에 대한 상동성을 가진 도메인을 포함하는 Cas 엔도뉴클레아제, 가이드 RNA, 및 공여체 DNA를 진균 세포군에 도입하는 단계로서, Cas 엔도뉴클레아제 및 가이드 RNA는 Cas 엔도뉴클

레아제가 진균 세포의 게놈좌 내 또는 진균 세포의 게놈좌 근처의 표적 부위에서 작용할 수 있게 하는 복합체를 형성할 수 있는 단계; 및 b) 게놈좌와 공여체 DNA의 상동성 재조합이 나타난 균으로부터 적어도 하나의 진균 세포를 확인하는 단계를 포함하고, Cas 엔도뉴클레아제, 가이드 RNA 또는 둘 모두는 진균 세포균으로 일시적으로 도입된다.

[0010] 일 구현예에서, 본 개시는 진균 세포 내 게놈좌와 공여체 DNA의 상동성 재조합을 위한 방법에 관한 것으로서, 이 방법은 a) 진균 세포의 게놈좌에 대한 상동성을 가진 도메인을 포함하는 Cas 엔도뉴클레아제, 가이드 RNA, 및 공여체 DNA를 진균 세포에 도입하는 단계로서, Cas 엔도뉴클레아제 및 가이드 RNA는 Cas 엔도뉴클레아제가 진균 세포의 게놈좌 내 또는 진균 세포의 게놈좌 근처의 표적 부위에서 작용할 수 있게 하는 복합체를 형성할 수 있는 단계; 및 b) 게놈좌와 공여체 DNA의 상동성 재조합이 진균 세포에서 나타나는 지를 확인하는 단계를 포함하고, Cas 엔도뉴클레아제, 가이드 RNA 또는 둘 모두는 진균 세포균으로 일시적으로 도입된다.

[0011] 일부 구현예에서, 진균 세포들 내 표적 부위(즉, Cas 엔도뉴클레아제 활성 부위)에서 비상동성 말단 결합(NHEJ) 메커니즘을 억제하거나 비활성화시키면 게놈좌에서 공여체 DNA의 상동성 재조합이 향상되는 것을 발견하였다. 그러므로, 본 발명의 양태들은 진균 세포들 내 표적 부위에서 비상동성 말단 결합(NHEJ) 메커니즘이 활성화되지 않거나, 비기능적이거나 감소되는 조건 하에서 본원에 설명된 상동성 재조합 방법을 수행하는 단계를 포함한다.

[0012] 사상균 세포 내 표적 부위에서 NHEJ 경로가 비기능적(비활성화)이 되게 하거나 NHEJ 경로를 감소시키는 것은 임의의 편리한 방식으로 이루어질 수 있고 숙주 세포의 장기간(또는 안정한) 표현형이거나 숙주 세포의 단기간(또는 일시적) 표현형일 수 있다. 예컨대, NHEJ 경로의 장기간 비활성화는 이의 활성이 숙주 세포로부터 감소되거나 제거되도록(예컨대, NHEJ 경로에서 유전자의 결실) NHEJ 경로에 관련된 하나 이상의 유전자의 염색체 유전자 변경에 의해 달성될 수 있다. 이는 여전히 비기능적/비활성화된 또는 감소된 NHEJ 경로를 갖는 원하는 유전자 변경(원하는 위치에서 공여체 DNA와 게놈 DNA 간 상동성 재조합)을 갖는 자손 세포의 수득을 초래한다. 대안적으로, 숙주 세포의 표적 부위에서 NHEJ 경로의 기능을 차단하거나 NHEJ 경로를 줄이는 것은 일시적으로 수행될 수 있다. 예컨대, NHEJ 경로의 일시적인 비활성화는 이의 발현이 NHEJ 경로의 하나 이상의 특정 성분의 발현 또는 활성을 저해하는 저해 RNA 또는 우성 음성 단백질을 발현하는 일시적 재조합 DNA 작제물을 숙주 세포 내로 도입하여 달성될 수 있다.

[0013] 원하는 유전자 변형을 갖는 자손 세포를 수득한 후, 일시적 재조합 DNA 작제물은, 예컨대 일시적 재조합 DNA 작제물의 유지를 위한 선택 압력을 제거하여 자손 세포로부터 제거될 수 있다. 이런 식으로, 원하는 자손 세포는 정상적으로 기능하는 NHEJ 경로를 가질 것이다. 비기능적이 되거나 활성의 감소를 가질 수 있는 NHEJ 경로 성분들의 예로는 ku80, ku70, rad50, mre11, xrs2, lig4, xrs, 및 임의의 원하는 이의 조합이 있다. 하나의 특별한 구현예에서, 진균 세포는 ku80의 비활성화 또는 발현 및/또는 활성의 감소를 갖는다. 여기서 주목할 것은 용어 “비기능적(non-functional)”은 NHEJ 경로의 특별한 성분에 대한 참조시 이러한 성분이 (예컨대, 유전자 결실로) 세포에 없는 경우뿐만 아니라 이러한 성분이 존재하지만 비기능적(예컨대, 비기능성 돌연변이 단백질)인 경우를 포함한다는 점이다.

[0014] 대안적으로, 이중 가닥 절단 활성보다는 오히려 니킹(nicking) 엔도뉴클레아제 활성(즉, 표적 부위에서 한 가닥의 DNA만을 절단; 본원에서 Cas 니카아제로도 지칭됨)을 갖는 Cas 엔도뉴클레아제를 채용할 수 있다. 표적 부위에서 닉(nick)을 유도하는 것은 이중 가닥 절단에서 있을 수 있는 것처럼 표적 부위에서 NHEJ 경로를 활성화하지는 않지만, 관심 게놈좌(Cas 니카아제용 표적 부위를 포함하거나 Cas 니카아제용 표적 부위에 가까운 것)와 공여체 DNA 사이의 상동성 재조합을 개선한다. Cas 니카아제의 예로는 아래에 기재된 것처럼 Cas 엔도뉴클레아제 변이체가 있다.

[0015] 몇몇 상이한 유형의 CRISPR-Cas 시스템이 기재되었고 유형 I, 유형 II 및 유형 III CRISPR-Cas 시스템으로 분류될 수 있다(예컨대, the description in Liu and Fan, CRISPR-Cas system: a powerful tool for genome editing. *Plant Mol Biol* (2014) 85:209-218 참조). 특정 양태에서, CRISPR-Cas 시스템은 Cas9 엔도뉴클레아제 또는 이의 변이체(예컨대, Cas 니카아제 포함)를 채용한 유형 II CRISPR-Cas 시스템이다. Cas9 엔도뉴클레아제는 다음 세균 종에서 Cas9 엔도뉴클레아제 및 이의 기능성 단편들을 포함하지만, 이들에 한정되지 않는 임의의 편리한 Cas9 엔도뉴클레아제일 수 있다: 스트렙토코쿠스 종(예컨대, S. 파이오제네스, S. 뮤탄스, 및 S. 더모필루스), 캄필로박터 종(예컨대, C. 제주니), 나이세리아 종(예컨대, N. 메닝기티데스), 프란시셀라 종(예컨대, F. 노비시다), 및 파스튜렐라 종(예컨대, P. 물토시다). Cas9의 수많은 기타 종들이 사용될 수 있다. 예를 들어, 서열 번호 45 및 서열 번호 48 내지 53 중 어느 하나와 적어도 70%의 동일성, 예컨대, 서열 번호 45 및 서열 번호 48 내지 53 중 어느 하나와 적어도 80% 동일성, 적어도 90% 동일성, 적어도 95% 동일성, 적어도 96%

동일성, 적어도 97% 동일성, 적어도 98% 동일성, 적어도 99% 동일성 및 100%까지의 동일성을 갖는 아미노산 서열을 포함하는 기능성 Cas9 엔도뉴클레아제 또는 이의 변이체가 채용될 수 있다.

- [0016] 특정 구현예에서, Cas 엔도뉴클레아제 및/또는 가이드 RNA를 진균 세포 내로 도입하는 단계는 Cas 엔도뉴클레아제, 가이드 RNA, 또는 이들 둘을 발현 카세트에 포함하는 하나 이상의 DNA 작제물을 진균 세포 내로 도입하는 단계를 포함한다. 하나 이상의 DNA 작제물은 진균 세포 내에서 한 번 Cas 엔도뉴클레아제 및/또는 가이드 RNA를 발현한다. 특정 구현예에서, DNA 작제물은 Cas 엔도뉴클레아제용 발현 카세트, 가이드 RNA용 발현 카세트, 및 공여체 DNA를 포함하는 원형 DNA 작제물이고, Cas 엔도뉴클레아제는 이중 가닥 절단 Cas 엔도뉴클레아제이거나 Cas 니카아제일 수 있다.
- [0017] 특정 구현예에서, 상기 도입 단계는 Cas 엔도뉴클레아제 폴리펩티드, 가이드 RNA, 또는 이들 둘을 진균 세포 내로 직접 도입하는 단계를 포함한다. 직접 도입 및 DNA 작제물 이용의 임의 조합이 채용될 수 있다(예컨대, 원할 때 동시에 또는 순차적으로, Cas 엔도뉴클레아제용 발현 카세트를 가진 DNA 작제물을 진균 세포 내로 도입하는 단계 및 가이드 RNA를 세포 내로 직접 도입하는 단계).
- [0018] 본원에 기술된 특정 방법들에서, DNA 작제물 내 Cas 발현 카세트는 진균 세포 내에서 발현을 위해 최적화된 Cas 엔도뉴클레아제 인코딩 유전자를 포함한다. 예컨대, 사상균 세포 내 발현을 위해 최적화된 Cas 엔도뉴클레아제 인코딩 유전자는 서열 번호 44와 적어도 70% 서열 동일성을 갖는 서열을 포함한다(S. Pyogenes에서 Cas9를 인코딩; 서열 번호 45).
- [0019] 일부 예에서, Cas 엔도뉴클레아제는 하나 이상의 핵의 표적화 신호(핵 위치 신호/서열(NLS)로도 지칭됨)에 작동 가능하게 연결된다. 서열 번호 7 및 서열 번호 8은 N-말단 및 C-말단에서 NLS 서열 및 인코딩된 아미노산 서열을 각각 갖는 사상균 세포에 최적화된 Cas9 유전자의 예를 제공한다. 많은 상이한 NLS들이 진핵생물에서 알려져 있다. 이들에는 단일파트, 2파트 및 3파트 유형이 있다. 임의의 편리한 NLS가 사용될 수 있는데, SV40 NLS, T. 레에세이 b1r2 (청색광 조절인자 2) 유전자로부터 유래된 NLS, 또는 이들 둘의 조합을 포함하는 예들을 가진 단일파트 유형이 약간 더 편리하다.
- [0020] 특정 구현예에서, 공여체 DNA는 관심 폴리뉴클레오티드 서열을 포함하고 게놈좌에서 상동성 재조합은 게놈좌에서 관심 폴리뉴클레오티드 서열의 삽입을 초래한다.
- [0021] 방법의 일부 구현예에서, 상기 도입 단계는 본원에 기재된 선택가능 마커 또는 표현형 마커를 인코딩하는 서열을 포함하는 DNA 작제물을 진균 세포 내로 도입하는 단계를 포함한다. 특정 구현예에서, DNA 작제물은 선택가능 마커를 인코딩하는 서열 및 공여체 DNA 모두를 포함한다. 일부 구현예에서, DNA 작제물은 Cas 엔도뉴클레아제를 인코딩하는 서열, 선택가능 마커를 인코딩하는 서열 및 공여체 DNA를 포함한다. 일부 구현예에서, DNA 작제물은 가이드 RNA를 인코딩하는 서열, 선택가능 마커를 인코딩하는 서열 및 공여체 DNA를 포함한다. 특별한 구현예에서, DNA 작제물은 Cas 엔도뉴클레아제를 인코딩하는 서열, 가이드 RNA를 인코딩하는 서열, 선택가능 마커를 인코딩하는 서열 및 공여체 DNA를 포함한다. 특정 구현예에서, DNA 작제물은 선형 DNA 작제물이다. 특정 구현예에서, DNA 작제물은 원형 DNA 작제물이다.
- [0022] 본 발명의 방법에서 사용되는 진균 세포는 사상균 세포 중일 수 있다. 특정 구현예에서, 진균 세포는 유미코티나(Eumycotina) 또는 페지조미코티나(Pezizomycotina) 진균 세포이다. 일부 구현예에서, 진균 세포는 트리코테르마, 페니실리움, 아스퍼질러스, 휴미콜라, 크리소스포리움, 푸사리움, 뉴로스포라, 마이셀리오프토라, 더모마이세스, 하이포크레아, 및 에메리셀라로 구성되는 군으로부터 선택된다. 사상균 트리코테르마 레에세이, P. 크리소제넴, M. 더모필라, 더모마이세스 라누지노수스, A. 오리제 및 A. 나이제가 특히 관심이 있다. 효모 종을 포함하는 기타 진균 세포들도 채용될 수 있다.
- [0023] 개시된 방법들의 사용자가 선택한 표적 부위는 오픈 리딩 프레임, 프로모터, 조절 서열, 종결인자 서열, 조절 요소 서열, 스플라이스 부위, 코딩 서열, 폴리유비퀴티네이션 부위, 인트론 부위, 및 인트론 증진 모티프로 구성되는 군으로부터 선택되는 관심 유전자의 영역 내에 위치될 수 있다. 관심 유전자의 예로는 아세틸 에스테라아제, 아미노펩티다아제, 아밀라아제, 아라비아나제, 아라비노푸라노시다아제, 카복시펩티다아제, 카탈라아제, 셀룰라아제, 키티나아제, 큐티나아제, 데옥시리보뉴클레아제, 에피메라아제, 에스테라아제, α-갈락토시다아제, β-갈락토시다아제, α-글루카나아제, 글루카닐라아제, 엔도-β-글루카나아제, 글루코아밀라아제, 글루코스 옥시다아제, α-글루코시다아제, β-글루코시다아제, 글루루로니다아제, 헤미셀룰라아제, 헥소오스 옥시다아제, 하이드로라아제, 인베르타제, 이소메라제, 락카아제, 리파아제, 리아제, 만노시다제, 옥시다아제, 옥시도레덕타아제, 펙테이트 리아제, 펙틴 아세틸 에스테라아제, 펙틴 디폴리머라아제, 펙틴 메틸 에스테라아제, 펙틴 분해

효소, 페록시다아제, 페놀옥시다아제, 피타아제, 폴리갈락투로나아제, 프로테아제, 람노-갈락투로나아제, 리보뉴클레아제, 트랜스퍼라아제, 수송 단백질, 트랜스글루타미나아제, 크실라나아제, 핵소오스 옥시다아제 및 이들의 조합을 인코딩하는 유전자들이 있다. 전사 인자 및 억제자(repressor)와 같은 조절 단백질, 키나아제와 같은 다른 단백질을 변형시키는 단백질, 번역후 변형(예: 글리코실화)에 관여하는 단백질을 인코딩하는 표적 유전자는 Cas 매개된 편집뿐만 아니라 세포 신호 전달, 형태학, 성장율, 및 단백질 분비에 관여하는 유전자를 받을 수 있다. 이 점에 대한 제한은 없다.

[0024] 특정 구현예에서, 게놈좌와 공여체 DNA의 상동성 재조합은 표적 부위에서 또는 표적 부위 근처에서 DNA 서열의 변형을 초래하고, 이러한 변형은 하나 이상의 뉴클레오티드의 결실, 하나 이상의 뉴클레오티드의 삽입, 관심 단백질을 인코딩하는 발현 카세트의 삽입, 하나 이상의 뉴클레오티드의 치환, 및 이들의 임의 조합으로 구성되는 군으로부터 선택된다. 일부 구현예에서, 상기 변형은 공여체 DNA에 원래 존재한다. 특정 구현예에서, 발현 카세트에 의해 인코딩된 관심 단백질은 효소이다. 특별한 구현예에서, 관심 단백질은 헤미셀룰라아제, 페록시다아제, 프로테아제, 셀룰라아제, 크실라나아제, 리파아제, 포스포리파아제, 에스테라아제, 쿠티나아제, 펙티나제, 케라티나아제, 리덕타아제, 옥시다아제, 페놀 옥시다아제, 리폭시게나아제, 리그니나아제, 폴루라나아제, 타나아제, 펜토사나아제, 만나아제, 베타-글루카나아제, 아라미노시다아제, 히알루로니다아제, 콘드로이티나아제, 락카아제, 아밀라아제, 글루코아밀라아제, 이의 변이체, 이의 기능적 단편, 또는 이들의 2종 이상의 하이브리드 또는 혼합물이다. 또 다른 특별한 구현예에서, 관심 단백질은 펩타이드 호르몬, 성장 인자, 클로팅 인자, 케모킨, 사이토킨, 림포킨, 항체, 수용체, 접착 분자, 미생물 항원, 이의 변이체, 이의 기능적 단편, 또는 이들의 2종 이상의 하이브리드 또는 혼합물이다.

[0025] 방법의 일부 구현예에서, 관심 부위에서 또는 관심 부위 근처에서 게놈 변형을 갖는 진균 세포를 확인하는 단계는 상동성 재조합 또는 변형을 선택하거나 선별하는 조건 하에서 단계 (a)로부터 세포군을 배양하는 단계를 포함한다. 이러한 조건은 항생제 선택 조건, 영양 요구성 세포를 선택 또는 선별하는 조건 등을 포함한다. 일부 구현예에서, 상기 확인 단계는 불안정한 형질전환체를 선별하는 조건 하에서 단계 (a)로부터 세포군을 배양하는 단계를 포함한다.

[0026] 상기한 기재의 방법에 있어서, 상기 도입 단계는 선택가능 마커를 인코딩하는 서열 및 공여체 DNA를 포함하는 DNA 작제물을 진균 세포들 내로 도입하는 단계를 포함하고, 상기 확인 단계는 선택가능 마커를 잃었지만 아직은 공여체 DNA를 보유하고 있는 불안정한 형질전환체를 선별하는 조건 하에서 단계 (a)로부터 세포군을 배양하는 단계를 포함한다.

[0027] 본 개시의 다른 양태는 위에서 기재된 방법들에 의해 생산된 재조합 진균 세포들뿐만 아니라 이러한 방법들을 수행함에 있어 부모 숙주세포들로서의 사용을 위한 재조합 진균 세포에 관한 것이다.

[0028] 이처럼, 특정 구현예에서, 본 개시의 양태는 Cas 엔도뉴클레아제용 발현 카세트를 포함하는 제1 재조합 DNA 작제물을 포함하는 재조합 진균 세포들을 포함한다. 특정 구현예에서, 재조합 진균 세포 내 NHEJ 경로는 비기능적(비활성화된)이거나 감소되고, 예를 들어 여기서 NHEJ 경로의 하나 이상의 성분들은 비활성화되거나, 비기능적이거나, 감소된 활성(예컨대, ku80, ku70, rad50, mre11, xrs2, lig4, xrs, 또는 이의 조합)을 갖는다. 예를 들어, 진균 세포는 ku80의 비활성화된/활성이 감소된 형태를 가질 수 있다. 특정한 다른 구현예에서, 재조합 진균 세포 내 NHJE 경로는 기능적이다.

[0029] 특정 구현예에서, 발현 카세트로부터 발현된 Cas 엔도뉴클레아제는 Cas9 엔도뉴클레아제 또는 이의 변이체이다. 대안적으로, 발현 카세트로부터 발현된 Cas 엔도뉴클레아제는 Cas 닉카아제이다.

[0030] 위에서 기재된 것처럼, 일부 경우에서, Cas 엔도뉴클레아제는 Cas9 엔도뉴클레아제(또는 이의 변이체)이다. Cas9 엔도뉴클레아제는 다음 박테리아 종에서 Cas9 엔도뉴클레아제 및 이의 기능성 단편들을 포함하지만, 이들에 한정되지 않는 임의의 편리한 Cas9 엔도뉴클레아제일 수 있다: 스트렙토코쿠스 종(*Streptococcus* sp.)(예컨대, *S. 파이오제네스(pyogenes)*, *S. 뮤탄스(mutans)* 및 *S. 더모필루스(thermophilus)*), 캄피로박테르 종(*Campylobacter* sp.)(예컨대, *C. 제주니(jejuni)*), 나이세리아 종(*Neisseria* sp.)(예컨대, *N. 메닝기티데스(meningitides)*), 프란시셀라 종(*Francisella* sp.)(예컨대, *F. 노비시다(novicida)*) 및 파스튜렐라 종(*Pasteurella* sp.)(예컨대, *P. 물토시다(multocida)*). Cas9의 수많은 기타 종들이 사용될 수 있다. 본원에 기술된 특정 진균 세포들에서, 제1 재조합 DNA 작제물은 진균 세포 내에서 발현을 위해 최적화된 Cas 엔도뉴클레아제 유전자를 포함한다. 예컨대, 사상균 세포 내 발현을 위해 최적화된 Cas 엔도뉴클레아제 인코딩 유전자는 서열 번호 44와 적어도 70% 서열 동일성을 갖는 서열을 포함한다(*S. 파이오제네스*에서 Cas9를 인코딩; 서열 번호 45). 일부 예에서, Cas 엔도뉴클레아제는 하나 이상의 핵의 표적화 신호(핵 위치 신호/서열(NLS)로도

지칭됨)에 작동 가능하게 연결된다. SV40, NLS(서열 번호 46), T. 레에세이 b1r2(청색광 조절인자 2) 유전자(서열 번호 47)로부터 유래된 NLS, 또는 이의 조합을 포함하는 예들을 갖는 임의의 편리한 NLS가 사용될 수 있다. 일부 구현예에서, 재조합 DNA 작제물은 Cas9 엔도뉴클레아제 또는 이의 변이체를 인코딩하는 사상균 세포에 최적화된 폴리뉴클레오티드 서열에 작동 가능하게 연결되는 프로모터를 포함한다.

[0031] 특정 양태에서, 위에서 기재된 사상균 세포는 선택적으로 발현 카세트를 통하여 가이드 RNA를 발현할 수 있는 제2 재조합 DNA 작제물을 더 포함하고, 여기서 가이드 RNA 및 Cas 엔도뉴클레아제는 Cas 엔도뉴클레아제가 재조합 진균 세포의 게놈 내 표적 부위에서 작용할 수 있는 복합체를 형성할 수 있고, 여기서 “작용한다”는 것은 Cas 엔도뉴클레아제가 (이중 가닥 절단 또는 Nick을 만들면서) 예상대로 DNA를 절단한다는 것을 의미한다. 일부 구현예에서, 재조합 DNA 작제물 또는 가이드 RNA용 발현 카세트는 유아스코마이세테(Euascomycete) 또는 페지조마이세테(Pezizomycete)에서 기능적인 DNA 중합효소 III 의존성 프로모터를 포함하고, 여기서 상기 프로모터는 가이드 RNA를 인코딩하는 DNA에 작동 가능하게 연결된다. 일부 구현예에서, 프로모터는 트리코테르마 U6 snRNA 유전자로부터 유래된다. 특정 구현예에서, 프로모터는 서열 번호 40 또는 41과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 뉴클레오티드 서열을 포함한다. 특정 구현예에서, 프로모터는 서열 번호 40 또는 41의 서열을 포함한다. 일부 구현예에서, 재조합 DNA 작제물 또는 가이드 RNA용 발현 카세트는 트리코테르마 U9 snRNA 유전자로부터 인트론 서열을 사용하여 가이드 RNA를 인코딩하는 DNA를 포함한다. 일부 구현예에서, 트리코테르마 U9 snRNA 유전자로부터 유래된 인트론 서열은 서열 번호 42와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 뉴클레오티드 서열을 포함한다. 특정 구현예에서, 트리코테르마 U9 snRNA 유전자로부터 유래된 인트론 서열은 서열 번호 42의 서열을 포함한다.

[0032] 일부 예에서, 재조합 진균 세포는 (상동성 재조합을 통하여 게놈 내 표적 유전자에 또는 표적 유전자 근처에 진균 세포의 게놈으로 삽입되도록 개시된 방법의 사용자에 의해 의도된) 관심 폴리뉴클레오티드를 포함하는 공여체 DNA를 더 포함한다. 이처럼, 특정 구현예에서, 진균 세포는 표적 부위에/표적 부위 근처에 삽입된 관심 폴리뉴클레오티드를 갖는다. 일부 예에서, 게놈좌의 서열에 비하여, 공여체 DNA는 진균 세포의 게놈좌와 상동성을 갖는 도메인 내에 다음 변형들 중 적어도 하나를 포함한다: 하나 이상의 뉴클레오티드의 결실, 하나 이상의 뉴클레오티드의 삽입, 관심 단백질을 인코딩하는 발현 카세트의 삽입, 하나 이상의 뉴클레오티드의 치환, 및 이들의 임의 조합.

[0033] 위에서 주목한 것처럼, Cas-표적화된 상동성 재조합은 Cas 표적 부위에서 NHEJ 경로가 비기능적이거나, 감소되거나 억제되는 세포들에서 향상된다. 그러나, 성공적이거나 심지어 효율적인 Cas-표적화된 상동성 재조합이 일어나기 위해 비기능적이거나, 감소된 또는 억제된 NHEJ 경로를 갖는 것은 필요치 않다.

[0034] 일부 구현예에서, 재조합 진균 세포는 본원에 기재된 선택가능 마커 또는 표현형 마커를 인코딩하는 서열을 포함하는 DNA 작제물을 포함한다. 특정 구현예에서, DNA 작제물은 본원에 기재된 것처럼 선택 가능 마커를 인코딩하는 서열 및 공여체 DNA 모두를 포함한다. 일부 구현예에서, DNA 작제물은 본원에 기재된 Cas 엔도뉴클레아제를 인코딩하는 서열, 선택 가능 마커를 인코딩하는 서열 및 공여체 DNA를 포함한다. 일부 구현예에서, DNA 작제물은 본원에 기재된 가이드 RNA를 인코딩하는 서열, 선택 가능 마커를 인코딩하는 서열 및 공여체 DNA를 포함한다. 특별한 구현예에서, DNA 작제물은 Cas 엔도뉴클레아제를 인코딩하는 서열, 가이드 RNA를 인코딩하는 서열, 선택 가능 마커를 인코딩하는 서열 및 공여체 DNA를 포함한다. 특정 구현예에서, DNA 작제물은 선형 DNA 작제물이다. 특정 구현예에서, DNA 작제물은 원형 DNA 작제물이다. 특정 구현예에서, DNA 작제물은 진균 세포의 게놈 내로 적어도 부분적으로 통합되거나 진균 세포의 게놈과 상동적으로 재조합된다. 특별한 구현예에서, DNA 작제물에 포함된 공여체 DNA의 적어도 일부 또는 전부는 진균 세포의 게놈 내로 통합되거나 진균 세포의 게놈과 상동적으로 재조합되지만, 선택가능 마커-인코딩 서열, Cas 엔도뉴클레아제-인코딩 서열, 또는 가이드 RNA-인코딩 서열은 진균 세포의 게놈 내로 통합되지 않거나 진균 세포의 게놈과 상동적으로 재조합되지 않는다.

[0035] 본 발명의 방법들에서 사용되는 진균 세포는 트리코테르마, 페니실리움, 아스퍼질러스, 휴미콜라, 크리소스포리움, 푸사리움, 뉴로스포라, 마이셀리오프토라, 더모마이세스, 하이포크레아, 및 에메리셀라로부터 선택되는 사상균 세포 종을 포함한다. 사상균 트리코테르마 레에세이, P. 크리소제넴, M. 더모필라, 더모마이세스 라누지노수스, A. 오리제 및 A. 나이거가 특히 관심이 있다. 효모 종을 포함하는 기타 진균 세포들도 사용될 수 있다.

[0036] 개시된 방법들의 사용자가 선택한 표적 부위는 오픈 리딩 프레임, 프로모터, 조절 서열, 종결인자 서열, 조절 요소 서열, 스플라이스 부위, 코딩 서열, 폴리유비퀴티네이션 부위, 인트론 부위, 및 인트론 증진 모티프로 구성되는 균으로부터 선택되는 관심 유전자의 영역 내에 위치될 수 있다. 관심 유전자의 예로는 아세틸 에스테라

아제, 아미노펩티다아제, 아밀라아제, 아라비아아제, 아라비노푸라노시다아제, 카복시펩티다아제, 카탈라아제, 셀룰라아제, 키티나아제, 키모신, 큐티나아제, 데옥시리보뉴클레아제, 에피메라아제, 에스테라아제,  $\alpha$ -갈락토시다아제,  $\beta$ -갈락토시다아제,  $\alpha$ -글루카나아제, 글루카닐라아제, 엔도- $\beta$ -글루카나아제, 글루코아밀라아제, 글루코스 옥시다아제,  $\alpha$ -글루코시다아제,  $\beta$ -글루코시다아제, 글루쿠로니다아제, 헤미셀룰라아제, 핵소오스 옥시다아제, 하이드로라아제, 인베르타제, 이소메라제, 락카아제, 리파아제, 리아제, 만노시다제, 옥시다아제, 옥시도레딕타아제, 펙테이트 리아제, 펙틴 아세틸 에스테라아제, 펙틴 디폴리머라아제, 펙틴 메틸 에스테라아제, 펙틴 분해 효소, 페록시다아제, 페놀옥시다아제, 피타아제, 폴리갈락투로나아제, 프로테아제, 람노-갈락투로나아제, 리보뉴클레아제, 토마틴, 트랜스퍼라아제, 수송 단백질, 트랜스글루타미나아제, 크실라나아제, 핵소오스 옥시다아제 및 이들의 조합을 인코딩하는 유전자들이 있다. 이 점에 대한 제한은 없다.

[0037] 본 발명의 특정 양태는 사상균 세포 최적화 Cas9 엔도뉴클레아제를 인코딩하는 뉴클레오티드 서열에 작동 가능하게 연결된 프로모터 서열을 포함하는 제조용 폴리뉴클레오티드를 포함하고, 여기서 사상균 세포 최적화 Cas9 엔도뉴클레아제는 가이드 RNA와 복합체 형성시 사상균 게놈의 게놈 표적 서열 내 이중 가닥 절단에 결합될 수 있고 이중 가닥 절단을 생성할 수 있다. 사상균 세포에 최적화된 Cas9 엔도뉴클레아제 유전자의 예로는 서열 번호 44 및 서열 번호 7 그리고 이의 부모의 천연 Cas9 인코딩 뉴클레오티드 서열에 비해 사상균 세포에서 발현을 개선한 이의 유사 변이체들이 있다.

[0038] 추가적인 제조용 폴리뉴클레오티드 서열은 이중 유전자에 대한 작동가능한 연결에서 사상균 세포-유래 RNA 폴리머라아제 III(폴(pol) III) 구동 프로모터 서열을 갖는 서열을 포함한다. 일부 구현예에서, 사상균 세포-유래 RNA 폴 III 구동 프로모터 서열은 U6 유전자 프로모터, 예컨대 서열 번호 40, 서열 번호 41, 또는 이의 기능적 변이체를 포함한다. 일부 예에서, 제조용 폴리뉴클레오티드는 RNA 폴 III 전사 유전자(예컨대, U6 유전자, 예를 들어 서열 번호 42)로부터 유래된 이중 서열 내 인트론 및/또는 RNA 폴 III 전사 유전자(예컨대, U6 유전자, 예를 들어 서열 번호 43)로부터 유래된 전사 종결인자를 더 포함한다. 특별한 구현예에서, 이중 서열은 가이드 RNA를 인코딩한다.

[0039] 본 개시의 방법 및 조성물의 추가적인 구현예는 본원에 도시된다.

### 도면의 간단한 설명

[0040] 다음의 상세한 기재 및 본 출원의 일부를 구성하는 첨부 도면들로부터 본 개시를 더 완전하게 이해할 수 있다.

**도 1**은 추정 T. 레에세이 U6 유전자(서열 번호 1)의 뉴클레오티드 서열을 도시한다. TATA 박스(밑줄쳐짐), 전사 시작 부위(하방 화살표), A-박스(밑줄쳐짐), 인트론(상방 화살표), B-박스(밑줄쳐짐; 유전자의 인트론 내), 인간 U6 유전자와 동일한 서열(굵은 이탤릭체), 및 종결인자(밑줄쳐짐)를 포함하는 관심 요소들이 표시된다.

**도 2**는 pTrex2gHyg-Mo Cas 플라스미드의 개략도를 도시한다.

**도 3**은 p219M 플라스미드의 개략도를 도시한다.

**도 4**는 도시된 PCR 프라이머 부위 및 인트론 영역을 갖는 T. 레에세이 ad3A 유전자의 개략도를 도시한다.

**도 5**는 도시된 PCR 프라이머 및 인트론 영역을 갖는 T. 레에세이 글루코아밀라아제 유전자(TrGA)의 개략도를 도시한다.

**도 6**은 텔로미어 서열을 포함하는 pTrex2gHygMoCasgPyr2TS6 플라스미드의 개략도를 도시한다.

**도 7**은 pET30a-Cas9-D10A 닉카아제의 플라스미드 지도를 도시한다.

**도 8a 및 8b**는 pMD18T (T7-Spy-TrGA\_sgF1)(**도 8a**) 및 pMD18T (T7-Spy-TrGA\_sgR1)(**도 8b**)의 플라스미드 지도를 도시한다.

**도 9**는 SpyCas9 뉴클레아제 분석의 결과를 보여주는 아가로스 겔이다. 레인 1, DNA 사다리; 레인 2, 대조; 레인 3 및 4, TrGA\_sgR1의 존재 하에서 SpyCas9 분석; 레인 5 및 6, TrGA\_sgF1의 존재 하에서 SpyCas9 분석.

**도 10**은 SpyCas9(D10A) 평가 결과를 보여주는 아가로스 겔이다. 레인 1, DNA 사다리; 레인 2, TrGA\_sgF1만을 갖는 SpyCas9(D10A); 레인 3, TrGA\_sgR1만을 갖는 SpyCas9(D10A); 레인 4, 양 RNA를 갖는 SpyCas9(D10A).

**도 11**은 TrGA 결실 카세트의 개략도이다.

**도 12**는 보겔의 전분 한천에서 성장하는 형질전환체의 형태와 함께 24-웰 플레이트를 도시한다. 지연된 성장 표

현형을 갖는 형질전환체는 원으로 표시된다. 3개의 대조군의 형태학은 우측 하단에 사각형으로 도시된다. "cellu" = 엔도뉴클레아제-3, 엔도뉴클레아제-4, 엔도뉴클레아제-5, 엔도뉴클레아제-6, 만나아제-1의 추가적인 결실을 갖는 T. 레에세이의 사중(quad)-결실 균주; "ΔAA" = 결실된 알파-아밀라아제 유전자를 갖는 T. 레에세이의 "cellu" 균주; "ΔGA" = 결실된 글루코아밀라아제 유전자를 갖는 T. 레에세이의 "cellu" 균주; 및 "empty" = 셀 없음 (비어있는 웰). 선택된 클론 #1 내지 #9가 표시된다.

**도 13a 및 13b.** T. 레에세이에서 TrGA 결실의 확인. (**도 13a**) 야생형 및 TrGA 결실 균주에서 TrGA 유전자자의 구조, 및 프라이머 Fw1, R3, KOF1(5'-gaacaatcttctttgcaatgttggtc-3')(서열 번호 79), KOR2(5'-ggcagactacaagtctactagtactac-3')(서열 번호 58), R1(5'-gaggaagtctgctttaggcaggc-3')(서열 번호 80) 및 F4(5'-cgacagagcagtcacatgggatacg -3')(서열 번호 81)의 결합 부위를 도시한 개략도. (**도 13b**) PCR의 결과를 도시한 아가로스 겔. 0.8% 아가로스 겔을 사용하여 140 볼트에서 30분 동안 작동하여 PCR 생성물을 분석하였다.

**도 14**는 pTrexMoCasGATS11-HDR의 플라스미드 지도를 도시한다.

### 발명을 실시하기 위한 구체적인 내용

- [0041] 본 개시는 진균 세포 내 게놈좌를 갖는 공여체 DNA의 상동성 재조합을 촉진하는데 사용되는 조성물과 방법을 포함한다. 이러한 방법은 원하는 표적 부위를 인식하여 이 부위에서 이중 가닥 절단 또는 Nick 도입하는 기능적 가이드 RNA/Cas 엔도뉴클레아제 복합체를 채용함으로써, 표적 부위에서 또는 표적 부위 근처에서 상동성 재조합을 촉진하고/촉진하거나 향상시킨다. 특정 양태에서, 진균 세포 내 표적 부위에서 비상동성 말단 결합(NHEJ) 메커니즘은 활성화되지 않거나, 비기능적이거나, 감소되는데, 본원에서 나타내길, 이는 원하는 상동성 재조합 이벤트의 효율을 개선한다.
- [0042] 본 조성물과 방법을 더욱 상세히 설명하기 전에, 본 조성물과 방법은 기재된 특별한 구현예에 한정되지 않으며, 이와 같이 당연히 변화될 수 있음을 이해하여야 한다. 또한, 본 조성물과 방법의 범위는 첨부된 청구항에 의해서만 한정될 것이기 때문에, 본원에서 사용된 용어는 특별한 구현예들만을 기재하기 위한 것이고, 한정하려는 의도가 아니라는 것을 이해하여야 한다.
- [0043] 값의 범위가 제공되는 경우, 문맥상 명확하게 달리 지시되지 않는다면, 이 범위의 상한치와 하한치 및 임의의 달리 언급된 값 또는 이 언급된 범위에서의 사이 값 사이에서 하한치의 단위의 1/10까지, 각각의 사이 값이 본 조성물 및 방법에 포함되는 것이 이해된다. 이들 더 작은 범위들의 상한치 및 하한치는 더 작은 범위에 독립적으로 포함될 수 있고, 이 언급된 범위에서 특별히 제외되는 임의의 제한을 받는 본 조성물 및 방법에도 포함된다. 언급된 범위가 제한들 중 하나 또는 양자 모두를 포함한다면, 이러한 포함된 제한들 중 하나 또는 둘 모두를 제외하는 범위들도 본 조성물 및 방법에 포함된다.
- [0044] 수치값 앞에 “약(about)”이라는 용어가 붙은 특정 범위들이 본원에 제시된다. “약”이라는 용어는 뒤에 놓인 정확한 숫자뿐만 아니라 그 용어의 뒤에 놓인 숫자와 거의 동일한 숫자 또는 근사적인 숫자에 대한 문자적 지원을 제공하기 위해 본원에 사용된다. 숫자가 구체적으로 언급된 숫자와 거의 동일한 숫자 또는 근사적인 숫자인지를 판단함에 있어서, 거의 동일하다고 언급되지 않은 숫자 또는 근사적이라고 언급되지 않은 숫자는 그것이 제시된 문맥에서, 구체적으로 언급된 숫자와 실질적으로 동일한 숫자일 수 있다. 예를 들어, 수치와 관련하여, 용어가 문맥에서 달리 구체적으로 정의되지 않으면, “약”이라는 용어는 이 수치의 -10% 내지 +10%의 범위를 지칭한다. 다른 예에서, pH 값이 구체적으로 달리 정의되지 않으면, “약 6의 pH 값”이라는 문구는 5.4 내지 6.6의 pH 값들을 지칭한다.
- [0045] 본원에 제공된 표제들은 전체적으로 본 명세서에 대한 참조로 가질 수 있는 본 조성물 및 방법의 다양한 양태 및 구현예의 제한은 아니다. 따라서, 바로 아래에 정의된 용어들은 전체적으로 명세서에 대한 참조로 보다 완전하게 정의된다.
- [0046] 본 문서는 읽기 쉽도록 수많은 부분들로 구성되지만, 독자는 하나의 부분에 있는 언급이 다른 부분에도 적용될 수 있다는 것을 이해할 것이다. 이런 방식으로, 본 개시의 다른 부분에 사용된 표제어는 한정하는 것으로 해석되어서는 안된다.
- [0047] 달리 정의되지 않으면, 본원에 사용된 기술적 및 과학적인 모든 용어들은 본 조성물 및 방법이 속하는 기술의 당업자가 일반적으로 이해하는 것과 동일한 의미를 가진다. 본원에 기재된 것과 유사하거나 동일한 임의의 방법들 및 물질들이 본 조성물 및 방법을 실시하거나 시험하는데 사용될 수 있지만, 대표적인 예시 방법 및 물질이 이제 기재된다.

- [0048] 본 명세서에서 언급된 모든 간행물과 특허는, 각 개별적인 간행물 또는 특허가 구체적으로 그리고 개별적으로 참조로 통합되도록 표시되듯이 그리고 이 간행물이 언급되는 것과 관련된 방법 및/또는 물질을 개시하고 기재하기 위해 본원에 참조로 통합되듯이, 본원에 참조로 통합된다. 임의 간행물의 인용은 출원일 전 이의 개시를 위한 것이고 본 조성물 및 방법이 선행 발명에 의해 이러한 간행물에 선행할 자격이 없다는 인정으로 해석되어서는 안된다. 또한, 제공된 공개일은 독립적인 확인이 필요할 수 있는 실제적인 공개일과 상이할 수 있다.
- [0049] 발명을 실시하기 위한 구체적인 내용에 따르면, 다음의 약어 및 정의들이 적용된다. 문맥상 달리 명확하게 지시하지 않으면, 단수형 “a”, “an” 및 “the” 는 복수의 대상들을 포함함에 주목하라. 따라서, 예를 들어, “한 효소(an enzyme)” 에 대한 참조는 다수의 이러한 효소들을 포함하고, “상기 투여량(the dosage)” 에 대한 참조는 하나 이상의 투여량 및 이 기술의 당업자에게 알려진 이의 등가량 등을 포함한다.
- [0050] 또한 주목할 점은 청구항이 임의의 선택적 요소를 제외하도록 작성될 수 있다는 것이다. 따라서, 이 진술은 청구항 요소의 인용 또는 “부정적인” 제한의 사용과 관련하여 “단독으로”, “오직” 등과 같은 배타적인 용어의 사용에 대한 선행 기준으로 작용하기 위한 것이다.
- [0051] 본 개시를 읽을 때 당업자에게 명백한 바와 같이, 본원에 기술되고 예시된 개별적인 구현예들 각각은 본원에 기재된 본 조성물 및 방법의 범위 또는 사상으로 부터 벗어나지 않고서 다른 몇몇 구현예들 중 임의의 구현예의 특징과 쉽게 분리되거나 결합될 수 있는 개별 성분 및 특징을 갖는다. 언급된 임의의 방법은 열거된 사건의 순서로 또는 논리적으로 가능한 다른 순서로 수행될 수 있다.
- [0052] 정의
- [0053] 본원에 사용된, “Cas 엔도뉴클레아제” 또는 “Cas 엔도뉴클레아제 활성” 을 갖는 것으로 언급되는 폴리펩티드는, 하나 이상의 가이드 폴리뉴클레오티드와 기능적으로 결합시, Cas 단백질이 표적 DNA 서열을 절단할 수 있는 Cas 유전자에 의해 인코딩된 CRISPR 관련(Cas) 폴리펩티드에 관한 것이다(“CRISPR-Cas systems and methods for altering expression of gene products”라는 제목을 가진 미국 특허 제8697359 참조). 니킹 엔도뉴클레아제 활성을 갖는 Cas 변이체들, 즉 이중 가닥 DNA 표적 부위에서 단일 가닥 닉을 도입하는 Cas 변이체들을 포함하여, 가이드 폴리뉴클레오티드 유도 엔도뉴클레아제 활성을 보유한 Cas 엔도뉴클레아제의 변이체들도 본 정의에 포함된다(아래의 정의 참조). (주목할 점은 낱말로 확인된 야생형 Cas 엔도뉴클레아제가 표적 부위에서 이중 가닥 절단을 도입한다는 것이다.) Cas 엔도뉴클레아제는, 예컨대 세포의 게놈 내 표적 부위에서 이중 가닥 DNA 내 특정 표적 부위를 인식하여 절단하는 가이드 폴리뉴클레오티드에 의해 안내된다. 몇몇 상이한 유형의 CRISPR-Cas 시스템이 기재되었고 유형 I, 유형 II 및 유형 III CRISPR-Cas 시스템으로 분류될 수 있다(예컨대, Liu and Fan, CRISPR-Cas system: a powerful tool for genome editing. *Plant Mol Biol* (2014) 85:209-218의 설명 참조). 특정 양태에서, CRISPR-Cas 시스템은 Cas9 엔도뉴클레아제 또는 이의 변이체(예컨대, Cas 니카아제 포함)를 채용한 유형 II CRISPR-Cas 시스템이다. Cas9 엔도뉴클레아제는 다음 세균 종에서 Cas9 엔도뉴클레아제 및 이의 기능성 단편들을 포함하지만, 이들에 한정되지 않는 임의의 편리한 Cas9 엔도뉴클레아제일 수 있다: 스트렙토코쿠스 종(예컨대, S. 파이오제네스, S. 뮤탄스, 및 S. 더모필루스), 캄필로박터 종(예컨대, C. 제주니), 나이세리아 종(예컨대, N. 메닝기티데스), 프란시셀라 종(예컨대, F. 노비시다), 및 파스튜렐라 종(예컨대, P. 물토시다). Cas9의 수많은 기타 종들이 사용될 수 있다. 예를 들어, 서열 번호 45 및 서열 번호 48 내지 53 중 어느 하나와 적어도 70 %의 동일성을 갖는 아미노산 서열을 포함하는 기능성 Cas9 엔도뉴클레아제 또는 이의 변이체가 채용될 수 있고, 예컨대, 서열 번호 45 및 서열 번호 48 내지 53 중 어느 하나와 적어도 80 % 동일성, 적어도 90 % 동일성, 적어도 95 % 동일성, 96 % 이상의 동일성, 97 % 이상의 동일성, 98 % 이상의 동일성, 99 % 이상의 동일성 및 100 %까지의 동일성을 갖는 아미노산 서열을 포함하는 기능성 Cas9 엔도뉴클레아제 또는 이의 변이체가 사용될 수 있다. 다른 구현예에서, Cas 엔도뉴클레아제 또는 이의 변이체는 유형 II CRISPR-Cas 시스템의 Cpf1 엔도뉴클레아제이다. Cpf1은 Cas9과 다른 특징을 가진 견고한 DNA 간섭을 매개한다. Cpf1은 tracrRNA가 부족하여 T-풍부 프로토스페이서-인접 모티프를 이용한다. 이는 엇갈린 DNA 이중 가닥 절단을 통해 DNA를 절단한다. 예컨대, Zetsche 등, *Cell* (2015) 163:759-771 참조.
- [0054] 본원에서 사용된 바와 같이, “Cas 니카아제”는, 하나 이상의 가이드 폴리뉴클레오티드와 기능적으로 결합시, 단일 가닥 닉을 표적인 이중 가닥 DNA 서열 내로 도입할 수 있는 Cas 엔도뉴클레아제이다. Cas 니카아제는 (예컨대, 부위-특이적 돌연변이에 의해) 부모 Cas 엔도뉴클레아제 내 2개의 뉴클레아제 도메인 중 하나를 비활성화하여 재조합적으로 생성될 수 있다. Cas 니카아제의 비제한적인 일 예로는 RuvC 도메인이 D10A 돌연변이에 의해 비활성화되는 Sander and Joung(*Nature Biotechnology*, 2013, 1-9)에 기재된 Cas9 니카아제가 있다. 위에서 언급되었듯이, 일반적인 용어 “Cas 엔도뉴클레아제” 는 이중 가닥 절단 및 니킹 Cas 폴리펩티드 둘 모두를 포함

한다. 예컨대, 가이드 RNA가 Cas 엔도뉴클레아제를 원하는 표적 부위에 유도할 수 있는 것으로 기재되면, (아래에 정의된 것처럼) 이중 가닥 절단 Cas 엔도뉴클레아제 및 니킹 Cas 폴리펩티드 둘 모두에 대해 그렇게 할 것이다.

[0055] 본원에 사용된 용어 "가이드 폴리뉴클레오티드"는 Cas 엔도뉴클레아제와 복합체를 형성할 수 있고, 이러한 Cas 엔도뉴클레아제가 DNA 표적 부위를 인식하고 선택적으로 절단할 수 있게 하는 폴리뉴클레오티드 서열에 관한 것이다. 가이드 폴리뉴클레오티드는 단일 분자 또는 이중 분자일 수 있다. 가이드 폴리뉴클레오티드 서열은 RNA 서열, DNA 서열, 또는 이의 조합(RNA-DNA 조합 서열)일 수 있다. 선택적으로, 이러한 가이드 폴리뉴클레오티드는 적어도 하나의 뉴클레오티드, 포스포다이에스테르 결합 또는 연결 변형, 예컨대 고정 핵산(LNA), 5-메틸 dC, 2,6-디아미노퓨린, 2'-플루오로 A, 2'-플루오로 U, 2'-O-메틸 RNA, 포스포로티오에이트 결합, 콜레스테롤 분자에 대한 연결, 폴리에틸렌 글리콜 분자에 대한 연결, 스페이서 18(헥사에틸렌 글리콜 사슬) 분자에 대한 연결, 또는 고리화를 초래하는 5'에서 3'으로의 공유 연결을 포함할 수 있으나, 이에 한정되지 않는다. 단독으로 리보핵산을 포함하는 가이드 폴리뉴클레오티드는 "가이드 RNA"라고도 지칭된다.

[0056] 가이드 폴리뉴클레오티드는 표적 DNA(아래에서 "포토스페이서" 또는 "표적 부위" 로도 불림)의 뉴클레오티드 서열에 상보적인 제1 뉴클레오티드 서열 도메인(가변 표적화 도메인 또는 VT 도메인으로 지칭됨) 및 Cas 엔도뉴클레아제 폴리펩티드와 상호 작용하는 제2 뉴클레오티드 서열 도메인(Cas 엔도뉴클레아제 인식 도메인 또는 CER 도메인으로 지칭됨)을 포함하는 이중 분자(듀플렉스 가이드 폴리뉴클레오티드라고도 지칭됨)일 수 있다. 이중 분자 가이드 폴리뉴클레오티드의 CER 도메인은 상보성 영역을 따라 혼성화된 두 개의 별도의 분자들을 포함한다. 이러한 두 개의 별도의 분자들은 RNA, DNA, 및/또는 RNA-DNA- 조합 서열일 수 있다. 일부 구현예에서, CER 도메인에 연결된 VT 도메인을 포함하는 듀플렉스 가이드 폴리뉴클레오티드의 제1 분자는 "crDNA"(연속된 DNA 뉴클레오티드들로 구성되는 경우) 또는 "crRNA"(연속된 RNA 뉴클레오티드들로 구성되는 경우), 또는 "crDNA-RNA"(DNA와 RNA 뉴클레오티드의 조합으로 구성되는 경우)라 지칭된다. cr뉴클레오티드는 박테리아 및 고세균류에서 자연적으로 발생하는 crRNA의 단편을 포함할 수 있다. 일 구현예에서, 박테리아 및 고세균류에서 자연적으로 발생하고 본원에 기재된 cr뉴클레오티드에 존재하는 crRNA의 단편의 크기는 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개 이상의 뉴클레오티드 범위를 가질 수 있지만, 이들 범위에 제한되지 않는다. 일부 구현예에서, CER 도메인을 포함하는 듀플렉스 가이드 폴리뉴클레오티드의 제2 분자는 "tracrRNA"(연속된 RNA 뉴클레오티드들로 구성되는 경우) 또는 "tracrDNA"(연속된 DNA 뉴클레오티드들로 구성되는 경우) 또는 "tracrDNA-RNA"(DNA와 RNA 뉴클레오티드의 조합으로 구성되는 경우)라 지칭된다. 특정 구현예에서, RNA/Cas9 엔도뉴클레아제 복합체를 가이드하는 RNA는 듀플렉스 crRNA-tracrRNA를 포함하는 듀플렉스화된 RNA이다.

[0057] 또한, 이러한 가이드 폴리뉴클레오티드는 표적 DNA의 뉴클레오티드 서열에 상보적인 (가변 표적화 도메인 또는 VT 도메인으로 지칭되는) 제1 뉴클레오티드 서열 도메인 및 Cas 엔도뉴클레아제 폴리펩티드와 상호 작용하는 (Cas 엔도뉴클레아제 인식 도메인 또는 CER 도메인으로 지칭되는) 제2 뉴클레오티드 도메인을 포함하는 단일 분자일 수 있다. "도메인"은 RNA, DNA, 및/또는 RNA-DNA-조합 서열일 수 있는 뉴클레오티드들이 연속되어 있는 것을 의미한다. 단일 가이드 폴리뉴클레오티드의 VT 도메인 및/또는 CER 도메인은 RNA 서열, DNA 서열, 또는 RNA-DNA-조합 서열을 포함할 수 있다. 일부 구현예에서, 이러한 단일 가이드 폴리뉴클레오티드는 (CER 도메인을 포함하는) tracr뉴클레오티드에 연결된 (CER 도메인에 연결된 VT 도메인을 포함하는) cr뉴클레오티드를 포함하며, 이때, 이러한 연결은 RNA 서열, DNA 서열, 또는 RNA-DNA 조합 서열을 포함하는 뉴클레오티드 서열이다. cr뉴클레오티드 및 tracr뉴클레오티드로부터의 서열들로 구성되는 단일 가이드 폴리뉴클레오티드는 (연속된 RNA 뉴클레오티드들로 구성되는 경우) "단일 가이드 RNA" 또는 (연속된 DNA 뉴클레오티드들로 구성되는 경우) "단일 가이드 DNA" 또는 (RNA와 DNA 뉴클레오티드의 조합으로 구성되는 경우) "단일 가이드 RNA-DNA"라 지칭될 수 있다. 본 개시의 일 구현예에서, 단일 가이드 RNA는 유형 II Cas 엔도뉴클레아제와 복합체를 형성할 수 있는 유형 II CRISPR/Cas 시스템의 crRNA 또는 crRNA 단편 및 tracrRNA 또는 tracrRNA 단편을 포함하고, 여기서 가이드 RNA/Cas 엔도뉴클레아제 복합체는 Cas 엔도뉴클레아제를 진균 세포 게놈 표적 부위에 유도하여, Cas 엔도뉴클레아제가 게놈 타겟 부위 내로 이중 가닥 절단을 도입할 수 있게 한다.

[0058] 단일 가이드 폴리뉴클레오티드 대 듀플렉스 가이드 폴리뉴클레오티드를 사용하는 일 양태는 하나의 발현 카세트만이 표적 세포 내에서 단일 가이드 폴리뉴클레오티드를 발현하기 위해 만들어질 필요가 있다는 것이다.

[0059] 용어 "가변 표적화 도메인" 또는 "VT 도메인"은 본원에서 상호교환적으로 사용되며, 이중 가닥 DNA 표적 부위의 한 가닥(뉴클레오티드 서열)에 상보적인 뉴클레오티드 서열을 포함한다. 제1 뉴클레오티드 서열 도메인(VT 도메인)과 표적 서열 사이의 백분율 보완성은 적어도 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%,

61%, 62%, 63%, 63%, 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% 또는 100% 상보적이다. VT 도메인의 길이는 적어도 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 또는 30개 뉴클레오티드 길이일 수 있다. 일부 구현예에서, 이러한 VT 도메인은 연속된 12개 내지 30개의 뉴클레오티드를 포함한다. VT 도메인은 DNA 서열, RNA 서열, 변형된 DNA 서열, 변형된 RNA 서열, 또는 이들의 임의 조합으로 구성될 수 있다.

[0060] 용어 가이드 폴리뉴클레오티드의 "Cas 엔도뉴클레아제 인식 도메인" 또는 "CER 도메인"은 본원에서 상호교환적으로 사용되며, Cas 엔도뉴클레아제 폴리펩티드와 상호작용하는 뉴클레오티드 서열(예컨대, 가이드 폴리뉴클레오티드의 제2 뉴클레오티드 서열 도메인)을 포함한다. CER 도메인은 DNA 서열, RNA 서열, 변형된 DNA 서열, 변형된 RNA 서열(예컨대, 본원에 기술된 변형 참조), 또는 이들의 임의 조합으로 구성될 수 있다.

[0061] 단일 가이드 폴리뉴클레오티드의 cr뉴클레오티드 및 tracr뉴클레오티드를 연결하는 뉴클레오티드 서열은 RNA 서열, DNA 서열, 또는 RNA-DNA 조합 서열을 포함할 수 있다. 일 구현예에서, 단일 가이드 폴리뉴클레오티드의 cr 뉴클레오티드 및 tracr뉴클레오티드를 연결하는 뉴클레오티드 서열은 그 길이가 적어도 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, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 또는 100 개의 뉴클레오티드일 수 있다. 다른 구현예에서, 단일 가이드 폴리뉴클레오티드의 cr뉴클레오티드 및 tracr뉴클레오티드를 연결하는 뉴클레오티드 서열은 테트라루프 서열, 예컨대 GAAA 테트라루프 서열을 포함할 수 있지만, 이에 한정되지는 않는다.

[0062] 가이드 뉴클레오티드, VT 도메인 및/또는 CER 도메인의 뉴클레오티드 서열 변형은 5' 캡, 3' 폴리아데닐화 테일, 리보스위치 서열, 안정성 조절 서열, dsRNA 듀플렉스를 형성하는 서열, 가이드 폴리뉴클레오티드를 세포내 위치에 표적화하는 변형 또는 서열, 트래킹을 제공하는 변형 또는 서열, 단백질을 위한 결합 부위를 제공하는 변형 또는 서열, 잠금 핵산(Locked Nucleic Acid(LNA)), 5- 메틸 dC 뉴클레오티드, 2,6- 디아미노퓨린 뉴클레오티드, 2'- 플루오로 A 뉴클레오티드, 2'- 플루오로 U 뉴클레오티드; 2' -O-메틸 RNA 뉴클레오티드, 포스포로티오에이트 결합, 콜레스테롤 분자에 대한 연결, 폴리에틸렌 글리콜 분자에 대한 연결, 스페이서 18 분자에 대한 연결, 5 '내지 3'공유 결합, 또는 이들의 임의 조합으로 이루어진 군으로부터 선택될 수 있지만, 이들에 한정되지는 않는다. 이러한 변형은 적어도 하나의 추가적인 유의한 특징을 초래할 수 있고, 여기서 추가적인 유의한 특징은 변형 또는 조절된 안정성, 세포 내 표적화, 트래킹, 형광 표지, 단백질 또는 단백질 복합체에 대한 결합 부위, 상보적인 표적 서열에 대한 변형된 결합 친화도, 세포 분해에 대한 변형된 저항성, 및 증가된 세포 투과성의 군으로부터 선택된다.

[0063] 본원에서 사용된 용어 "가이드 폴리뉴클레오티드/카스 엔도뉴클레아제 시스템" (및 등가물)은 DNA 표적 부위에서 이중 가닥 절단을 도입할 수 있는 Cas 엔도뉴클레아제 및 가이드 폴리뉴클레오티드 (단일 또는 이중)의 복합체를 포함한다. Cas 엔도뉴클레아제는 DNA 표적 부위에 아주 근접하여 DNA 듀플렉스를 풀어주고, 가이드 RNA가 표적 서열을 인식할 때 DNA 가닥 둘 모두를 절단하지만, 올바른 프로토스페이서-인접 모티프(PAM)가 표적 서열의 3' 말단에서 적절하게 배향되는 경우에만 그렇게 한다.

[0064] 용어 "기능적 단편", "기능적으로 동등한 단편(fragment that is functionally equivalent)", "기능적으로 동등한 단편(functionally equivalent fragment)" 등은 상호 교환적으로 사용되며, 부모 폴리펩티드의 정량적 효소 활성을 유지하는 부모 폴리펩티드의 일부 또는 부서열을 지칭한다. 예컨대, Cas 엔도뉴클레아제의 기능적 단편은 가이드 폴리뉴클레오티드와 함께 이중 가닥 절단을 생성하는 능력을 유지한다. 여기서 주목할 점은 기능적 단편이 부모 폴리펩티드에 비해 변경된 양론적 효소 활성을 가질 수 있다는 것이다.

[0065] 용어 "기능적 변이체(functional variant)", "기능적으로 동등한 변이체(variant that is functionally equivalent)", "기능적으로 동등한 변이체(functionally equivalent variant)" 등은 상호 교환적으로 사용되며 부모 폴리펩티드의 정량적 효소 활성을 유지하는 부모 폴리펩티드의 변이체를 지칭한다. 예컨대, Cas 엔도뉴클레아제의 기능적 변이체는 가이드 폴리뉴클레오티드와 함께 (대상 변이체에 따라) 이중 가닥 절단 또는 Nick을 생성하는 능력을 유지한다. 여기서 주목할 점은 기능적 단편이 부모 폴리펩티드에 비해 변경된 양론적 효소 활성을 가질 수 있다는 것이다.

[0066] 단편 및 변이체는 부위-특이적 돌연변이 및 합성 구조를 포함하는 임의의 편리한 방법으로 수득될 수 있다.

- [0067] 진균 세포에 적용시 용어 “게놈”은 핵 내에서 발견되는 염색체 DNA뿐만 아니라 세포의 세포 내 성분(예컨대, 미토콘드리아) 내에서 발견되는 세포소기관 DNA를 포함한다.
- [0068] “코돈 변형 유전자” 또는 “코돈-선택 유전자” 또는 “코돈-최적화 유전자”는 숙주 세포의 선호되는 코돈 사용 빈도를 모방하도록 설계된 코돈 사용 빈도를 갖는 유전자이다. 유전자를 코돈-최적화하도록 이루어진 핵산 변경은 "동의어"이며, 이는 이들이 부모 유전자의 인코딩된 폴리펩티드의 아미노산 서열을 변경시키지 않는다는 것을 의미한다. 그러나, 고유 유전자 및 변이체 유전자 모두는 특정 숙주 세포에 대해 코돈-최적화될 수 있으며, 이와 관련하여 이와 같은 제한은 의도되지 않는다.
- [0069] “코딩 서열”은 특정 아미노산 서열을 코딩하는 폴리뉴클레오티드 서열을 지칭한다. “조절 서열”은 코딩 서열의 상류(5' 비코딩 서열), 내 또는 하류(3' 비코딩 서열)에 위치하며, 관련 코딩 서열의 전사, RNA 가공 또는 안정성, 또는 번역에 영향을 주는 뉴클레오티드 서열을 지칭한다. 조절 서열은 프로모터, 번역 리더 서열, 5' 미번역 서열, 3' 미번역 서열, 인트론, 폴리아데닐화 표적 서열, RNA 가공 부위, 이팩터 결합 부위, 및 스템-루프(stem-loop) 구조를 포함할 수 있지만, 이에 한정되지는 않는다.
- [0070] “프로모터”는 코딩 서열 또는 기능적 RNA의 발현을 조절할 수 있는 DNA 서열을 지칭한다. 프로모터 서열은 근접 상류 요소 및 더 먼 상류 요소로 구성되고, 더 먼 상류 요소는 종종 인헨서(enhancer)로 지칭된다. “인헨서”는 프로모터 활성을 자극할 수 있는 DNA 서열이고, 프로모터의 선천적인 요소 또는 프로모터의 수준 또는 조직 특이성을 향상시키기 위해 삽입된 이종 요소일 수 있다. 프로모터는 고유 유전자로부터 그 전체가 유래될 수도 있거나, 자연에서 발견되는 상이한 프로모터들로부터 유래된 상이한 요소들로 구성될 수 있거나, 심지어 합성 DNA 세그먼트를 포함할 수 있다. 상이한 프로모터가 상이한 조직 또는 세포 유형으로, 또는 발달의 상이한 단계에서 또는 상이한 환경 조건에 반응하여 유전자의 발현을 유도할 수 있음이 당업자에게 이해된다. 또한, 대부분의 경우, 조절 서열의 정확한 경계가 완전히 정의되지 않았기 때문에, 일부 변이체의 DNA 단편이 동일한 프로모터 활성을 가질 수 있음이 또한 인식된다. 이 기술에서 잘 알려진 것처럼, 프로모터들은 이들의 강도 및/또는 이들이 활성인 조건, 예컨대 항시성(constitutive) 프로모터, 강한 프로모터, 약한 프로모터, 유도성/억제성 프로모터, 조직 특이적/발달 조절 프로모터, 세포주기 의존성 프로모터 등에 따라 분류될 수 있다.
- [0071] “RNA 전사물”은 DNA 서열의 RNA 중합효소-효소화 전사로부터 만들어지는 생성물을 지칭한다. “메신저 RNA” 또는 “mRNA”는 인트론이 없고 세포에 의해 단백질로 번역될 수 있는 RNA를 지칭한다. “crRNA”는 역전사 효소를 이용하는 mRNA 템플릿에 상보적이고 그로부터 합성되는 DNA를 지칭한다. “센스” RNA는 mRNA를 포함하고 세포 내 또는 시험관 내에서 단백질로 번역될 수 있는 RNA 전사물을 지칭한다. “안티센스 RNA”는 표적 1차 전사물 또는 mRNA의 전부 또는 일부에게 상보적이고, 특정 조건 하에서, 표적 유전자의 발현을 차단하는 RNA 전사물을 지칭한다(미국 특허 제5,107,065호 참조). 안티센스 RNA의 상보성은 특정 유전자 전사물의 임의의 부분, 즉 5' 비-코딩 서열, 3' 비-코딩 서열, 인트론 또는 코딩 서열과 함께 있을 수 있다. “기능적 RNA”는 폴리펩티드로 번역되지 않을 수 있지만 세포 과정에 영향을 미치는 안티센스 RNA, 리보자임 RNA 또는 기타 RNA를 지칭한다. 용어 “보체(complement)” 및 “역 보체(reverse complement)”는 mRNA 전사체에 대하여 본원에서 상호 교환적으로 사용되며, 메시지의 안티센스 RNA를 정의하기 위한 의미이다.
- [0072] 본원에 사용된, “기능적으로 부착된” 또는 “작동 가능하게 연결된”은 프로모터, 인헨서 부위, 종결인자, 신호 서열, 에피토프 태그 등과 같이 공지된 또는 원하는 활성을 갖는 폴리펩티드 또는 폴리뉴클레오티드 서열의 조절 영역 또는 기능적 도메인이, 조절 영역 또는 기능적 도메인이 그의 공지된 또는 원하는 활성에 따라 그 표적의 발현, 분비 또는 기능을 조절할 수 있는 방식으로 표적(예컨대, 유전자 또는 폴리펩티드)에 부착되거나 연결된다는 것을 의미한다. 예를 들어, 프로모터는, 코딩 서열의 발현을 조절할 수 있는 경우(즉, 코딩 서열이 프로모터의 전사 조절 하에 있을 때), 코딩 서열과 작동 가능하게 연결된다.
- [0073] 본원에서 사용된 표준 재조합 DNA 및 분자 복제 기술은 당업계에 공지되어 있다.
- [0074] “PCR” 또는 “중합 효소 연쇄 반응(polymerase chain reaction)”은 특정 DNA 단편의 합성을 위한 기술이며 일련의 반복적인 변성, 어닐링 및 확장 사이클로 구성되며 당업계에 잘 알려져 있다.
- [0075] 용어 “재조합물”은, 생물학적 성분 또는 조성물(예컨대, 세포, 핵산, 폴리펩티드/효소, 벡터 등)과 관련하여 사용되는 경우, 생물학적 성분 또는 조성물이 자연에서 발견되지 않는 상태로 있다는 것을 나타낸다. 다시 말해서, 생물학적 성분 또는 조성물은 이의 자연 상태에서부터 인간 개입에 의해 변형되었다. 예를 들어, 재조합 세포는 이의 천연 부모(즉, 비-재조합) 세포에서 발견되지 않는 하나 이상의 유전자를 발현하는 세포, 이의 천연 부모 세포와 상이한 양으로 하나 이상의 천연 유전자를 발현하는 세포, 및/또는 천연 부모 세포와 상이한 조

건 하에서 하나 이상의 천연 유전자를 발현하는 세포를 포함한다. 재조합 핵산은 하나 이상의 뉴클레오티드에 의해 천연 서열과 다를 수 있으며, 이중 서열(예컨대, 이중 프로모터, 비-천연 또는 변이체 신호 서열을 인코딩하는 서열 등)에 작동 가능하게 연결될 수 있고, 인트론 서열이 없을 수 있고/있거나 단리된 형태로 존재할 수 있다. 재조합 폴리펩티드/효소는 하나 이상의 아미노산에 의해 천연 서열과 상이할 수 있고, 이중 서열과 융합될 수 있고, 절단될 수 있거나 아미노산의 내부 결실을 가질 수 있으며, 천연 세포에서 발견되지 않는 방식으로 발현될 수 있고/있거나(예컨대, 폴리펩티드를 인코딩하는 발현 벡터의 세포 내 존재로 인해 폴리펩티드를 과발현시키는 재조합 세포로부터), 단리된 형태로 존재할 수 있다. 일부 구현예에서, 재조합 폴리뉴클레오티드 또는 폴리펩티드/효소는 이의 야생형 대응물과 동일하지만 (예를 들어, 단리된 또는 농축된 형태의) 비천연 형태로 존재하는 서열을 갖는다는 것이 강조된다.

[0076] 용어 "플라스미드", "벡터" 및 "카세트"는 관심 폴리뉴클레오티드 서열, 예컨대 세포에서 발현되는 관심 유전자("발현 벡터" 또는 "발현 카세트")를 운반하는 여분의 염색체 요소를 지칭한다. 이러한 요소는 일반적으로 이중 가닥 DNA의 형태를 가지며, 임의의 근원으로부터 유래된 단일 또는 이중 가닥 DNA 또는 RNA의 서열, 게놈 통합 서열, 과자 또는 뉴클레오타이드 서열을 선형 또는 원형 형태로 자동으로 복제할 수 있고, 여기서 다수의 뉴클레오티드 서열이 세포 내로 관심 폴리뉴클레오티드를 도입할 수 있는 독특한 구조로 결합되거나 재조합되었다. 관심 폴리뉴클레오티드 서열은 표적 세포 내에서 발현될 폴리펩티드 또는 기능적 RNA를 인코딩하는 유전자일 수 있다. 발현 카세트/벡터는 일반적으로 숙주 세포 내에서 그 유전자의 발현을 허용하는 작동 가능하게 연결된 요소를 가진 유전자를 포함한다.

[0077] 본원에서 사용된 용어 "발현"은 전구체 또는 성숙한 형태의 기능적 최종 생성물(예컨대, mRNA, 가이드 RNA 또는 단백질)의 생산을 지칭한다.

[0078] 폴리뉴클레오티드 또는 폴리펩티드를 세포(예컨대, 재조합 DNA 작제물/발현 작제물)에 삽입하는 문맥에서 "도입된(introduced)"은 이러한 과제 수행을 위한 임의의 방법을 지칭하며, 원하는 생체 분자의 도입을 달성하기 위한 "형질감염", "형질전환", "형질도입", 물리적 수단 등의 임의의 수단을 포함한다.

[0079] "일시적으로 도입된(introduced transiently)", "일시적으로 도입된(transiently introduced)", "일시적 도입", "일시적으로 발현하는" 등은 비영구적 방식으로 숙주 세포(또는 숙주 세포군)에 생체 분자가 도입되는 것을 의미한다. 이중 가닥 DNA와 관련하여, 일시적인 도입은 도입된 DNA가 숙주 세포의 염색체 내로 통합되지 않아 성장 중에 모든 딸 세포에 전달되지 않는 상황뿐만 아니라 도입되어 염색체로 통합될 수 있었던 DNA 분자가 임의의 편리한 방법(예를 들어, cre-lox 시스템을 채용하고, 에피솜 DNA 작제물에 대한 양성의 선택적인 압력을 제거하고, 선택 매질 등을 사용하여 염색체로부터의 통합된 폴리뉴클레오티드의 전부 또는 일부로부터 루핑(looping)을 촉진시킴으로써)을 사용하여 원하는 시간에 제거되는 상황을 포함한다. 이 점에 대한 제한은 없다. 일반적으로, 숙주 세포에 RNA(예컨대, 가이드 RNA, 메신저 RNA, 리보자임 등) 또는 폴리펩티드(예컨대, Cas 폴리펩티드)를 도입하는 것은 이러한 생체 분자들이 세포 성장 동안 복제되지 않고 무한히 딸 세포들에게 전달된다는 점에서 일시적인 것으로 간주된다. Cas/가이드 RNA 복합체와 관련하여 일시적인 도입은, 두 가지 생체 분자가 표적화된 Cas 엔도뉴클레아제 활성을 발휘하는 데 필요하므로, 어느 한 구성 요소가 일시적으로 도입되는 상황을 다룬다. 따라서, Cas/가이드 RNA 복합체의 일시적 도입은 Cas 엔도뉴클레아제 및 가이드 RNA 중 어느 하나 또는 둘 모두가 일시적으로 도입되는 구현예들을 포함한다. 예를 들어, 기능적 복합체는 일시적 방식으로 숙주 세포에 존재하기 때문에 가이드 RNA가 일시적으로 도입되는 Cas 엔도뉴클레아제를 위한 게놈-통합 발현 카세트(따라서 일시적으로 도입되지 않음)를 갖는 숙주 세포는 일시적으로 도입된 Cas/가이드 RNA 복합체 (또는 시스템을) 가지고 있다고 말할 수 있다. 특정 구현예에서, 상기 도입 단계는: (i) Cas 엔도뉴클레아제를 안정적으로 발현하는 부모 진균 세포군을 수득하는 단계, 및 (ii) 가이드 RNA를 상기 부모 진균 세포군에 일시적으로 도입하는 단계를 포함한다. 역으로, 상기 도입 단계는: (i) 가이드 RNA를 안정적으로 발현하는 부모 진균 세포군을 수득하는 단계, 및 (ii) Cas 엔도뉴클레아제를 상기 부모 진균 세포군에 일시적으로 도입하는 단계를 포함한다.

[0080] "성숙한" 단백질은 번역후 가공된 폴리펩티드(즉, 1차 번역 생성물에 존재하는 임의의 프리폴리펩티드 또는 프로폴리펩티드가 제거된 것)를 지칭한다. "전구체" 단백질은 mRNA의 번역의 1차 생성물(즉, 프리펩티드 및 프로펩티드가 여전히 존재하는)을 지칭한다. 프리펩티드 및 프로펩티드는 세포내 위치 신호일 수 있지만, 이에 한정되지는 않는다.

[0081] "안정한 형질전환"은 핵 및 세포소기관 게놈 둘 모두를 포함하여 숙주 생물의 게놈으로 핵산 단편을 이동시켜 유전학적으로 안정한 유전을 수득하는 것을 지칭한다(수득된 숙주 세포는 때로는 본원에서 "안정한 형질전환

체"라고도 함). 대조적으로, "일시적 형질전환"은 숙주 유기체의 핵산 또는 다른 DNA- 함유 세포소기관으로 핵산 단편을 옮겨서 통합 또는 안정한 유전없이 유전자 발현을 유도하는 것을 의미한다(때때로 본원에서 "불안정한 형질전환"으로 지칭되고, 수득된 숙주 세포는 때때로 본원에서 "불안정한 형질전환체"라 지칭된다). 형질전환된 핵산 단편을 포함하는 숙주 생물은 "형질전환 생물"로 지칭된다.

[0082] 본원에서 사용된 "진균 세포", "곰팡이", "진균 숙주 세포" 등은 필라 아스코마이코타(*phyla Ascomycota*), 바시디오마이코타(*Basidiomycota*), 키트리디오마이코타(*Chytridiomycota*), 및 자이코마이코타(*Zygomycota*)(Hawksworth 등, In, Ainsworth and Bisby 's Dictionary of The Fungi, 8<sup>th</sup> edition, 1995, CAB International, University Press, Cambridge, UK에서 정의됨)뿐만 아니라 오오마이코타(*Oomycota*)(Hawksworth 등, supra에서 인용) 및 모든 불완전 균류(*mitosporic fungi*)(Hawksworth 등, supra)를 포함한다. 특정 구현예에서, 진균 세포는 효모 세포이고, 여기서 "효모"는 자낭홀씨생산 효모(*Endomyceales*), 담자균류 효모, 및 불완전 균류(*Fungi Imperfecti*)(*Blastomycetes*)에 속하는 효모를 의미한다. 이와 같이, 효모 숙주 세포는 칸디다(*Candida*), 한세누라(*Hansenula*), 클루베로마이세스(*Kluyveromyces*), 피치아(*Pichia*), 사카로마이세스(*Saccharomyces*), 스키나사카로마이세스(*Schizosaccharomyces*), 또는 야로위아 세포(*Yarrowia cell*)를 포함한다. 효모 종은 다음을 포함하지만 이에 한정되지 않는다: 사카로마이세스 칼스베르겐시스(*Saccharomyces carlsbergensis*), 사카로마이세스 세레비시애(*Saccharomyces cerevisiae*), 사카로마이세스 디아스타티쿠스(*Saccharomyces diastaticus*), 사카로마이세스 도우글라시(*Saccharomyces douglasii*), 사카로마이세스 클루베리(*Saccharomyces kluyveri*), 사카로마이세스 노르벤시스(*Saccharomyces norbensis*), 사카로마이세스 오비포르미스(*Saccharomyces oviformis*), 클루베로마이세스 락티스(*Kluyveromyces lactis*), 및 야로위아 리폴리티카(*Yarrowia lipolytica*) 세포.

[0083] 용어 "사상균 세포"는 아문 유마이코티나(*Eumycotina*) 또는 페치조마이코티나(*Pezizomycotina*)의 모든 사상 형태를 포함한다. 사상균 속의 적합한 세포들은 아크레모늄(*Acremonium*), 아스퍼질러스(*Aspergillus*), 크리소스포리움(*Chrysosporium*), 코리나스쿠스(*Corynascus*), 채토미움(*Chaetomium*), 푸사리움(*Fusarium*), 기베렐라(*Gibberella*), 후미콜라(*Humicola*), 마그나포르테(*Magnaporthe*), 마이셀리오프토라(*Myceliophthora*), 뉴로스포라(*Neurospora*), 패시로마이세스(*Paecilomyces*), 페니실리움(*Penicillium*), 스키타리디움(*Scytalidium*), 타라로마이세스(*Talaromyces*), 더모아스쿠스(*Thermoascus*), 티에라비아(*Thielavia*), 토리포클라디움(*Tolypocladium*), 하이포크레아(*Hypocrea*), 및 트리코테르마의 세포들을 포함하지만, 이들에 한정되지는 않는다.

[0084] 사상균 종의 적합한 세포들은 아스퍼질러스 아와모리(*Aspergillus awamori*), 아스퍼질러스 푸미가투스(*Aspergillus fumigatus*), 아스퍼질러스 포에티두스(*Aspergillus foetidus*), 아스퍼질러스 자포니쿠스(*Aspergillus japonicus*), 아스퍼질러스 니두란스(*Aspergillus nidulans*), 아스퍼질러스 나이거(*Aspergillus niger*), 아스퍼질러스 오리제(*Aspergillus oryzae*), 크리소스포리움 루크노웬스(*Chrysosporium lucknowense*), 푸사리움 박트리디오데스(*Fusarium bactridioides*), 푸사리움 세레아리스(*Fusarium cerealis*), 푸사리움 크룩웰렌스(*Fusarium crookwellense*), 푸사리움 쿨모룸(*Fusarium culmorum*), 푸사리움 그라미네아룸(*Fusarium graminearum*), 푸사리움 그라미눔(*Fusarium graminum*), 푸사리움 헤테로스포룸(*Fusarium heterosporum*), 푸사리움 네군디(*Fusarium negundi*), 푸사리움 옥시스포룸(*Fusarium oxysporum*), 푸사리움 렉티쿠라툼(*Fusarium reticulatum*), 푸사리움 로세움(*Fusarium roseum*), 푸사리움 삼부시눔(*Fusarium sambucinum*), 푸사리움 사르코크로움(*Fusarium sarcochromum*), 푸사리움 스포로트리키오이데스(*Fusarium sporotrichioides*), 푸사리움 숄푸레움(*Fusarium sulphureum*), 푸사리움 토루로숨(*Fusarium torulosum*), 푸사리움 크리코테시오데스(*Fusarium trichothecioides*), 푸사리움 베네나툼(*Fusarium venenatum*), 후미콜라 인솔렌스(*Humicola insolens*), 후미콜라 라누지노사(*Humicola lanuginosa*), 하이포크레아 제코리나(*Hypocrea jecorina*), 마이셀리오프토라 더모필라(*Myceliophthora thermophila*), 뉴로스포라 크라싸(*Neurospora crassa*), 뉴로스포라 인테르메디아(*Neurospora intermedia*), 페니실리움 푸르푸로게눔(*Penicillium purpurogenum*), 페니실리움 카네센스(*Penicillium canescens*), 페니실리움 소리툼(*Penicillium solitum*), 페니실리움 푸니쿠로숨(*Penicillium funiculosum*), 파네로채테 크리소스포리움(*Phanerochaete chrysosporium*), 타라로마이세스 플라부스(*Talaromyces flavus*), 티에라비아 테레스트리시스(*Thielavia terrestris*), 트리코테르마 하르지아눔(*Trichoderma harzianum*), 트리코테르마 코닝지(*Trichoderma koningii*), 트리코테르마 롱지브라키아툼(*Trichoderma longibrachiatum*), 트리코테르마 레에세이(*Trichoderma reesei*), 및 트리코테르마 비리데(*Trichoderma viride*)를 포함하지만, 이들에 한정되지는 않는다.

[0085] 용어 "표적 부위", "표적 서열", "게놈 표적 부위", "게놈 표적 서열" (및 등가물)은 본원에서 상호 교환 적으로 사용되며, Cas 엔도뉴클레아제 절단이 게놈 변형, 예컨대 공여체 DNA와의 상동성 재조합을 촉진시키도록 요

구되는 진균 세포의 게놈 내 폴리뉴클레오티드 서열을 지칭한다. 그러나, 본 용어가 사용되는 문맥은 이의 의미가 약간 변경될 수 있다. 예를 들어, Cas 엔도뉴클레아제의 표적 부위는 일반적으로 매우 특이적이고 종종 정확한 뉴클레오타이드 서열/위치로 정의될 수 있는 반면, 일부 경우에는 원하는 게놈 변형을 위한 표적 부위가 단지 DNA 절단이 일어나는 부위, 예컨대, 상동성 재조합이 요구되는 게놈좌 또는 영역보다 더 광범위하게 정의될 수 있다. 따라서, 특정 경우에서, Cas/가이드 RNA DNA 절단의 활성을 통하여 일어나는 게놈 변형은 표적 부위"에서" 또는 표적 부위 "근처에서" 일어나는 것으로 기재된다. 표적 부위는 진균 세포 게놈 내의 내인성 부위일 수 있거나 또는 대안적으로 표적 부위는 진균 세포에 이종일 수 있어서 게놈에서 자연 발생하지 않을 수 있거나 또는 자연에서 일어나는 경우에 비해 표적 부위가 이종 게놈 위치에서 발견될 수 있다. 특정한 다른 경우에서, 공여체 DNA가 진균 세포의 게놈좌에 대한 상동성을 가진 도메인을 포함할 때, 진균 세포에 도입된 Cas 엔도뉴클레아제 및 가이드 RNA는 Cas 엔도뉴클레아제가 진균 세포의 게놈좌 내 또는 진균 세포의 게놈좌 근처의 표적 부위에서 작용할 수 있는 복합체를 형성할 수 있다. 일부 구현예에서, 게놈 DNA 위의 Cas 엔도뉴클레아제 절단 부위(또는 표적 부위)는 공여체 DNA 및 게놈좌 사이의 상동성 영역 내에 있고, 여기서 상동성 재조합이 일어날 수 있다. 다른 구현예에서, 절단 부위는 상동성 영역으로부터 약 1 bp 내지 약 10 kb 떨어져 있을 수 있는 (예컨대, 상동성 영역의 부위로부터 1 bp, 2 bp, 5 bp, 10 bp, 20 bp, 50 bp, 100 bp, 250 bp, 500 bp, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb 또는 10 kb 떨어진) 공여체 DNA와 게놈좌 사이의 상동성 영역 근처에 있다.

[0086] 본원에서 사용된 "핵산"은 뉴클레오티드를 의미하고 데옥시리보뉴클레오티드 또는 리보뉴클레오티드 염기들의 단일 가닥 폴리머 또는 이중 가닥 폴리머를 포함한다. 또한 핵산은 단편 및 변형된 뉴클레오티드를 포함할 수 있다. 따라서, 용어 "폴리뉴클레오티드", "핵산 서열", "뉴클레오티드 서열" 및 "핵산 단편"은 단일 또는 이중 가닥인 RNA 및/또는 DNA의 폴리머를 나타내기 위해 교환 가능하게 사용되며, 선택적으로 합성, 비-천연 또는 변형된 뉴클레오티드 염기를 포함한다. (보통 이들의 5' -모노포스페이트 형태로 발견되는) 뉴클레오티드는 다음과 같이 이들의 단일 문자 표시에 의해 지칭된다: "A"는 아데노신 또는 데옥시아데노신(각각 RNA 또는 DNA에 대하여), "C"는 시아노 또는 데옥시시토신, "G"는 구아노신 또는 데옥시구아노신, "U"는 우리딘, "T"는 데옥시티미딘, "R"은 퓨린 (A 또는 G), "Y"는 피리미딘 (C 또는 T), "K"는 G 또는 T, "H"는 A 또는 C 또는 T, "I"는 이노신, 및 "N"은 임의의 뉴클레오티드.

[0087] 용어 "에서 유래된"은 용어 "에서 기원하는", "에서 수득되는", "에서 수득 가능한", "에서 단리된", 및 "에서 생성된"을 포함하고, 일반적으로 하나의 특정된 물질이 이의 기원을 다른 특정된 물질에서 발견하거나 다른 특정된 물질을 참조하여 기재될 수 있는 특징들을 가진다는 것을 나타낸다.

[0088] 본원에서 사용된 용어 "혼성화 조건"은 혼성화 반응이 수행되는 조건을 지칭한다. 이들 조건들은 전형적으로 혼성화가 측정되는 조건의 "엄격도" 정도에 의해 분류된다. 엄격도의 정도는, 예컨대 핵산 결합 복합체 또는 프로브의 용융 온도(T<sub>m</sub>)에 기초할 수 있다. 예를 들어, "최대 엄격도"는 전형적으로 약 T<sub>m</sub> - 5°C(프로브의 T<sub>m</sub> 아래의 5°C)에서 발생하고; "높은 엄격도"는 T<sub>m</sub> 아래의 약 5~10°C에서 발생하고; "중간 엄격도"는 프로브의 T<sub>m</sub> 아래의 약 10~20°C에서 발생하고; 그리고 "낮은 엄격도"는 T<sub>m</sub> 아래의 약 20~25°C에서 발생한다. 대안적으로, 또는 추가적으로, 혼성화 조건은 혼성화의 염 또는 이온 세기 조건 및/또는 한 번 이상의 엄격한 세척(예: 6X SSC = 매우 낮은 엄격도; 3X SSC = 낮은 엄격도 내지 중간 엄격도; 1X SSC = 중간 엄격도; 및 0.5X SSC = 높은 엄격도)에 기초할 수 있다. 기능적으로, 최대 엄격도 조건은 혼성화 프로브와 엄격한 동일성 또는 거의 엄격한 동일성을 갖는 핵산 서열을 확인하는데 사용될 수 있다; 반면 고엄격도 조건은 프로브와 약 80% 이상의 서열 동일성을 갖는 핵산 서열을 확인하는데 사용된다. 높은 선택성을 필요로 하는 적용에 있어서, 하이브리드를 형성하기 위해 비교적 엄격한 조건을 사용하는 것이 전형적으로 바람직하다(예컨대, 비교적 낮은 염 및/또는 고온 조건이 사용됨).

[0089] 본원에 사용된 용어 "혼성화"는 당 업계에 공지된 바와 같이 핵산 가닥이 염기쌍을 통해 상보적 가닥과 결합하는 과정을 지칭한다. 보다 구체적으로, "혼성화"는, 블롯 혼성화 기술 및 PCR 기술 동안 발생하는 것과 같이, 핵산의 한 가닥이 상보적 가닥과 듀플렉스, 즉 염기 쌍을 형성하는 과정을 지칭한다. 핵산 서열은 중등도 내지 고 엄격도 혼성화 및 세척 조건 하에 두 서열이 특이적으로 서로 혼성화되는 경우 기준 핵산 서열에 대해 "선택적으로 혼성화 가능"하다고 간주된다. 혼성화 조건은 핵산 결합 복합체 또는 프로브의 용융 온도(T<sub>m</sub>)에 근거한다. 예를 들어, "최대 엄격도"는 일반적으로 약 T<sub>m</sub> - 5°C(프로브의 T<sub>m</sub> 아래의 5°C)에서 발생한다. 높은 엄격도는 T<sub>m</sub> 아래의 약 5~10°C에서 발생하고; "중간 엄격도"는 프로브의 T<sub>m</sub> 아래의 약 10~20°C에서 발생하고; 그리고 "낮은 엄격도"는 T<sub>m</sub> 아래의 약 20~25°C에서 발생한다. 기능적으로, 최대 엄격도 조건은 엄격한 동일성 또는 혼성화 프로브와 거의 엄격한 동일성을 갖는 서열을 확인하는데 사용될 수 있고; 반면에 중간 또는 낮은 엄격도

혼성화는 폴리뉴클레오티드 서열 동족체를 확인하거나 검출하는데 사용될 수 있다.

- [0090] 중간 및 높은 엄격도 혼성화 조건은 당업계에서 잘 공지되어 있다. 예를 들어, 20% 포름아미드, 5 x SSC (150 mM NaCl, 15 mM 트리소듐 시트레이트), 50 mM 소듐 포스페이트(pH 7.6), 5 x 덴하르트(Denhardt) 용액, 10% 텍스트란 설페이트 및 20 mg/mL 변성 진단 연어 정자 DNA를 포함하는 용액에서 37°C에서 밤새 배양으로 중간 엄격도 혼성화를 수행한 후, 필터를 37~50°C의 1x SSC에서 세척할 수 있다. 고 엄격도 혼성화 조건은 65°C 및 0.1x SSC(1x SSC=0.15 M NaCl, 0.015 M Na 시트레이트, pH 7.0)에서의 혼성화일 수 있다. 대안적으로, 약 42°C에서 50% 포름아미드, 5X SSC, 5X 덴하르트 용액, 0.5% SDS 및 100 µg/mL 변성 캐리어 DNA에서 고 엄격도 혼성화 조건을 수행한 후 상온에서 2X SSC 및 0.5% SDS에서 2회 세척 및 42°C에서 0.1X SSC와 0.5% SDS에서 2회 추가로 세척하였다. 매우 높은 엄격도 혼성화 조건은 68°C 및 0.1X SSC에서의 혼성화일 수 있다. 또한 당업자는 프로브 길이 등과 같은 인자를 수용하기 위해 필요에 따라 온도, 이온 세기 등을 조정하는 방법을 알고있다.
- [0091] 적어도 2 개의 핵산 또는 폴리펩티드와 관련하여 "실질적으로 유사한" 또는 "실질적으로 동일한"이라는 문구는 폴리뉴클레오티드 또는 폴리펩티드가 부모 또는 기준 서열과 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 또는 심지어 적어도 99% 동일하거나, 기능성을 추가하지 않고서 본 설명을 단지 우회하기 위해 이루어진 아미노산 치환, 삽입, 삭제 또는 변형을 포함하지 않는 것을 의미한다.
- [0092] 핵산 또는 폴리펩티드 서열의 문맥에서 "서열 동일성" 또는 "동일성"은, 특정 비교창에 대한 최대 일치율을 위해 정렬될 때, 동일한 두 서열 내의 핵산 염기 또는 아미노산 잔기를 나타낸다.
- [0093] 용어 "서열 동일성 백분율"은 비교창에서 최적으로 정렬된 2개의 서열을 비교하여 결정된 값을 지칭하며, 이때, 비교창 내의 폴리뉴클레오티드 또는 폴리펩티드 서열의 부분은 2개의 서열의 최적 정렬을 위한 (삽입 또는 결실을 포함하지 않는) 기준 서열과 비교하여 삽입 또는 결실(즉, 갭)을 포함할 수 있다. 백분율은, 두 서열에서 동일한 핵산 염기 또는 아미노산 잔기가 나타나는 위치의 개수를 결정하여 일치하는 위치의 개수를 산출하고, 일치하는 위치의 개수를 비교창 내의 위치의 총 개수로 나누고, 그 결과에 100을 곱하여 서열 동일성의 백분율을 산출함으로써 계산한다. 백분율 서열 동일성의 유용한 예는 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% 또는 95% 내지 100%의 임의 정수 백분율을 포함하지만 이들에 한정되지 않는다. 이들 동일성은 본원에 기재된 프로그램중 임의의 것을 사용하여 판단될 수 있다.
- [0094] 서열 정렬 및 백분율 동일성 또는 유사성 계산은 LASERGENE 생물 정보학 컴퓨팅 스위트(DNASTAR Inc., Madison, WI)의 MegAlign™ 프로그램을 포함하지만 이에 한정되지 않는 상동성 서열을 검출하도록 설계된 다양한 비교 방법을 사용하여 결정될 수 있다. 본 출원의 문맥에서, 서열 분석 소프트웨어가 분석에 사용되는 경우, 달리 명시되지 않는 한, 분석 결과는 언급된 프로그램의 "디폴트 값"에 기초할 것이라는 것을 이해할 것이다. 본원에서 사용된 바와 같이, "디폴트 값"은 최초로 초기화 될 때, 원래 소프트웨어로 로딩되는 임의의 값 또는 파라미터 세트를 의미할 것이다.
- [0095] "클러스탈 V 정렬 방법"은 클러스탈 V(Higgins and Sharp, (1989) CABIOS 5 : 151-153, Higgins 등, (1992) Comput Appl Biosci 8 : 189-191에 기재)로 명명되고 LASERGENE 생물 정보학 컴퓨팅 스위트의 MegAlign™ 프로그램 (DNASTAR Inc., Madison, WI)에서 발견되는 정렬 방법에 해당한다. 다중 정렬의 경우, 디폴트 값은 GAP PENALTY=10 및 GAP LENGTH PENALTY=10에 해당한다. 클러스탈 방법을 사용하는 단백질 서열의 백분율 동일성의 계산 및 쌍 정렬을 위한 디폴트 파라미터는 KTUPLE=1, GAP PENALTY=3, WINDOW=5 및 DIAGONALS SAVED=5이다. 핵산의 경우, 이들 파라미터는 KTUPLE=2, GAP PENALTY=5, WINDOW=4 및 DIAGONALS SAVED=4이다. 클러스탈 V 프로그램을 사용하여 서열을 정렬한 후에는, 동일한 프로그램에서 "서열 거리"표를 보고 "백분율 동일성"을 얻을 수 있다.
- [0096] "클러스탈 W 정렬 방법"은 클러스탈 W(Higgins and Sharp, (1989) CABIOS 5 : 151-153, Higgins 등, (1992) Comput Appl Biosci 8 : 189-191에 기재)로 명명되고 LASERGENE 생물 정보학 컴퓨팅 스위트의 MegAlign™ v6.1 프로그램(DNASTAR Inc., Madison, WI)에서 발견되는 정렬 방법에 해당한다. 다중 정렬을 위한 디폴트 파라미터 (GAP PENALTY=10, GAP LENGTH PENALTY=0.2, 지연 발산 서열(%) = 30, DNA 전이 중량 = 0.5, 단백질 무게 매트릭스 = Gonnet 시리즈, DNA 무게 매트릭스 = IUB). 클러스탈 W 프로그램을 사용하여 서열을 정렬한 후에는, 동일한 프로그램에서 "서열 거리"표를 보고 "백분율 동일성"을 얻을 수 있다.
- [0097] 달리 명시되지 않는 한, 본원에 제공된 서열 동일성/유사성 값은 하기 파라미터를 사용하여, GAP 버전 10(GCG,

Accelrys, San Diego, CA)을 사용하여 얻은 값을 나타낸다: 갭 생성 페널티 중량 50 및 갭 길이 연장 페널티 중량 3을 사용하는 뉴클레오타이드 서열에 대한 % 동일성 및 % 유사성, 및 nwsgapdna.cmp 스코어링 매트릭스; GAP 생성 페널티 중량 8 및 갭 길이 연장 페널티 2를 사용하는 아미노산 서열에 대한 % 동일성 및 % 유사성, 및 BLOSUM62 스코어링 매트릭스(Henikoff 및 Henikoff, (1989) Proc. Natl. Acad. Sci.USA89: 10915). GAP는 Needleman 및 Wunsch (1970) J Mol Biol 48 : 443-53의 알고리즘을 사용하여 일치 횟수를 최대화하고 갭의 수를 최소화하는 두 개의 완전한 서열의 정렬을 찾는다. GAP는 가능한 모든 정렬 및 갭 위치를 고려하고, 일치하는 기준 단위로 갭 생성 페널티 및 갭 연장 페널티를 사용하여 가장 많은 수의 일치된 기준과 가장 적은 갭으로 정렬을 생성한다.

[0098] 서열 동일성의 많은 레벨은 다른 종 또는 천연 변형되거나 합성된 종으로부터 폴리펩티드를 확인하는데 유용하고, 이러한 폴리펩티드는 동일 또는 유사한 기능 또는 활성을 갖는다는 것이 당업자에 의해 잘 이해된다. 백분율 동일성의 유용한 예는 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% 또는 95%, 또는 50% 내지 100%의 임의 정수 백분율을 포함하지만 이들에 한정되지 않는다. 실제로, 50% 내지 100%, 예컨대 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% 또는 99%의 임의 정수 아미노산 동일성은 본 개시를 설명하는데 유용할 수 있다.

[0099] "유전자"는, 이들에 한정되지는 않지만 예컨대 특정 폴리펩티드(예컨대, 효소) 또는 기능적 RNA 분자(예컨대, 가이드 RNA, 항-센스 RNA, 리보자임 등)와 같은 기능적 분자를 인코딩하고 발현할 수 있는 핵산 단편을 포함하고, 코딩 서열 앞(5' 비-코딩 서열) 및/또는 코딩 서열 뒤(3' 비-코딩 서열)의 조절 서열을 포함한다. "천연 유전자"는 자신의 조절 서열과 함께 자연에서 발견되는 유전자를 지칭한다. 재조합 유전자는 상이한 유기체 또는 동일한 유기체로부터 유래될 수 있는 상이한 유전자의 조절 서열에 의해 조절되는 유전자를 지칭한다.

[0100] "돌연변이된 유전자"는 인간 개입을 통해 변경된 유전자이다. 이러한 "돌연변이된 유전자"는 적어도 하나의 뉴클레오타이드 삽입, 결실, 또는 치환에 의해 상응하는 비 돌연변이된 유전자의 서열과 상이한 서열을 갖는다. 본 개시의 특정 구현예에서, 돌연변이된 유전자는 본원에 개시된 가이드 폴리뉴클레오타이드/Cas 엔도뉴클레아제 시스템을 이용하여 만들어진 변경을 포함한다. 돌연변이된 진균 세포는 돌연변이된 유전자를 포함하는 진균 세포이다.

[0101] 본원에서 사용된 "표적화된 돌연변이"는, 본원에 개시되거나 당업계에 공지된 바와 같이, 표적 서열의 DNA에서 이중 가닥 절단을 유도할 수 있는 이중 가닥 절단 유도제를 포함하는 방법을 사용하여 천연 유전자 내에서 표적 서열을 변경함으로써 만들어진 천연 유전자의 돌연변이이다.

[0102] 용어 "공여체 DNA" 또는 "공여체 핵산 서열" 또는 "공여체 폴리뉴클레오타이드"는 일반적으로 Cas/가이드 폴리뉴클레오타이드 복합체(위에서 상세하게 기재된 바와 같이, 가이드 폴리뉴클레오타이드는 표적 부위를 정의함)의 활성과 관련하여, 표적 부위에 또는 표적 부위 근처에 삽입되거나 표적 부위에 또는 표적 부위 근처의 영역을 대체할 관심 폴리뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드를 지칭한다. 이와 같이, 공여체 DNA에서 관심 폴리뉴클레오타이드 서열은, 표적 부위에 또는 표적 부위 부근에서 치환되거나/편집될 뉴클레오타이드 서열과 비교할 때, 표적 부위에 또는 표적 부위 근처에 삽입되는 신규한 영역 및/또는 변형된 폴리뉴클레오타이드 서열을 포함할 수 있다. 특정 구현예에서, 공여체 DNA 작제물은 관심 폴리뉴클레오타이드 서열에 인접한 상동성의 제1 및 제2 영역을 더 포함한다. 공여체 DNA의 상동성의 제1 및 제2 영역은 각각 진균 세포 계놈의 표적 부위에 또는 그 측면에 존재하는 제1 및 제2 계놈 부위와 상동성을 공유한다. "상동성"이란 유사한 DNA 서열을 의미한다. 예를 들어, 공여체 DNA에서 발견되는 "계놈 영역에 대한 상동성 영역"은 진균 세포 계놈의 주어진 "계놈 영역"과 유사한 서열을 갖는 DNA 영역이다. 상동성의 영역은 절단된 표적 부위에서 상동성 재조합을 촉진시키기 위해 충분한 임의의 길이일 수 있다. 예를 들어, 상동성의 영역은 해당 계놈 영역과 상동성 재조합을 겪기에 충분한 상동성을 가지도록, 길이가 적어도 5~10, 5~15, 5~20, 5~25, 5~30, 5~35, 5~40, 5~45, 5~50, 5~55, 5~60, 5~65, 5~70, 5~75, 5~80, 5~85, 5~90, 5~95, 5~100, 5~200, 5~300, 5~400, 5~500, 5~600, 5~700, 5~800, 5~900, 5~1000, 5~1100, 5~1200, 5~1300, 5~1400, 5~1500, 5~1600, 5~1700, 5~1800, 5~1900, 5~2000, 5~2100, 5~2200, 5~2300, 5~2400, 5~2500, 5~2600, 5~2700, 5~2800, 5~2900, 5~3000, 5~3100 이상의 염기를 포함할 수 있다. "충분한 상동성"은 2개의 폴리뉴클레오타이드 서열이 상동성 재조합 반응을 위한 기질로서 작용하기에 충분한 구조적 유사성을 갖는다는 것을 나타낸다. 구조적 유사성은 각 폴리뉴클레오타이드 단편의 전체 길이뿐만 아니라 폴리뉴클레오타이드의 서열 유사성을 포함한다. 서열 유사성은 서열의 전체 길이에 걸쳐 백분율 서열 동일성 및/또는 100% 서열 동일성을 갖는 인접 뉴클레오타이드와 같은 국소화된 유사성 및 서열 길이의 일부분에 걸쳐 백

분율 서열 동일성을 포함하는 보존된 영역에 의해 기술될 수 있다.

- [0103] 표적 및 공여체 폴리뉴클레오티드가 공유하는 상동성 또는 서열 동일성의 양은 변할 수 있으며, 약 1~20 bp, 20~50 bp, 50~100 bp, 75~150 bp, 100~250 bp, 150~300 bp, 200~400 bp, 250~500 bp, 300~600 bp, 350~750 bp, 400~800 bp, 450~900 bp, 500~1000 bp, 600~1250 bp, 700~1500 bp, 800~1750 bp, 900~2000 bp, 1~2.5 kb, 1.5~3 kb, 2~4 kb, 2.5~5 kb, 3~6 kb, 3.5~7 kb, 4~8 kb, 5~10 kb, 또는 표적 부위의 전체 길이까지를 포함하는 범위의 단위 적분 값을 갖는 총 길이 및/또는 영역을 포함한다. 이 범위에는 범위 내의 모든 정수가 포함되고, 예를 들어 1~20 bp 범위는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 및 20 bp를 포함한다. 상동성의 양은 적어도 약 50%, 55%, 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100%의 백분율 서열 동일성을 포함하는 2개의 폴리뉴클레오티드의 전체 정렬 길이에 걸쳐 백분율 서열 동일성에 의해 또한 기술될 수 있다. 충분한 상동성은 폴리뉴클레오티드 길이, 전체 백분율 서열 동일성, 및 임의로 보존된 인접 뉴클레오티드 또는 국소 백분율 서열 동일성의 임의의 조합을 포함하며, 예를 들어 충분한 상동성은 표적좌의 영역에 대해 적어도 80% 서열 동일성을 갖는 75 내지 150 bp의 영역으로 기술될 수 있다. 충분한 상동성은 또한 고 엄격도 조건 하에서 특이적으로 혼성화하는 2개의 폴리뉴클레오티드의 예측된 능력에 의해 기술될 수도 있다(예컨대, Sambrook 등, (1989) *Molecular Cloning: A Laboratory Manual*, (Cold Spring Harbor Laboratory Press, NY); *Current Protocols in Molecular Biology*, Ausubel 등, Eds (1994) *Current Protocols*, (Greene Publishing Associates, Inc. and John Wiley & Sons, Inc); and, Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes*, (Elsevier, New York) 참조).
- [0104] 본원에서 사용된 "게놈 영역"은 표적 부위의 어느 한면에 존재하거나 대안적으로 표적 부위의 일부도 포함하는 진균 세포의 게놈에서 염색체의 세그먼트이다. 게놈 영역은 해당 상동성 영역과 상동성 재조합을 겪기에 충분한 상동성을 가지도록, 적어도 5~10, 5~15, 5~20, 5~25, 5~30, 5~35, 5~40, 5~45, 5~50, 5~55, 5~60, 5~65, 5~70, 5~75, 5~80, 5~85, 5~90, 5~95, 5~100, 5~200, 5~300, 5~400, 5~500, 5~600, 5~700, 5~800, 5~900, 5~1000, 5~1100, 5~1200, 5~1300, 5~1400, 5~1500, 5~1600, 5~1700, 5~1800, 5~1900, 5~2000, 5~2100, 5~2200, 5~2300, 5~2400, 5~2500, 5~2600, 5~2700, 5~2800, 5~2900, 5~3000, 5~3100 이상의 염기를 포함할 수 있다.
- [0105] 본원에 사용된 "상동성 재조합"은 상동성 부위에서 2개의 DNA 분자들간 DNA 단편들의 교환을 포함하고, 당업계에 잘 기술되어 있다.
- [0106] 표현형 마커는 시각적 마커 및 그것이 양성의 선택 가능한 마커든 음성의 선택 가능한 마커든 선택 가능한 마커를 포함하는, 선별이 가능하거나 선택 가능한 마커이다. 임의의 표현형 마커가 이용될 수 있다. 구체적으로, 선택 가능하거나 선별 가능한 마커는 종종 특정한 조건 하에서, 그것을 함유하는 분자 또는 세포를 확인할 수 있게 하거나, 그것을 함유하는 분자 또는 세포를 선택 또는 선별할 수 있게 하는 DNA 세그먼트를 포함한다. 이들 마커는 활성, 예컨대, RNA, 펩티드, 또는 단백질의 생산(그러나 이에 한정되지 않음)을 인코딩할 수 있거나, RNA, 펩티드, 단백질, 무기 및 유기 화합물 또는 조성물 등을 위한 결합 자리를 제공할 수 있다.
- [0107] 선택 가능한 마커의 예로는 제한 효소 부위를 포함하는 DNA 세그먼트; 독성 화합물 및 항생제, 예컨대, 클로리무론 에틸, 베노밀, 바스타 및 히그로마이신 포스포트랜스퍼라제(HPT)에 대해 저항성을 제공하는 생성물을 인코딩하는 DNA 세그먼트; 수용 세포에 결합되어 있는 생성물을 인코딩하는 DNA 세그먼트(예컨대, tRNA 유전자, 영양요구성 마커, 우세 이종 마커-amdS); 용이하게 식별될 수 있는 생성물을 인코딩하는 DNA 세그먼트(예를 들어, 표현형 마커, 예컨대,  $\beta$ -갈락토시다제, GUS; 형광 단백질, 예컨대, 녹색 형광 단백질(GFP), 청록색 형광 단백질(CFP), 황색 형광 단백질(YFP), 적색 형광 단백질(RFP) 및 세포 표면 단백질); PCR을 위한 새로운 프라이머 부위의 생성(예컨대, 이전에는 나란히 놓여있지 않았던 두 DNA 서열의 병치), 제한 엔도뉴클레아제 또는 기타 DNA 변형 효소, 화학물질 등에 의해 영향 받지 않거나 영향 받는 DNA 서열의 포함; 및 식별을 가능하게 하는 특이적인 변형(예컨대, 메틸화)에 필요한 DNA 서열의 포함을 들 수 있으나, 이에 한정되지 않는다.
- [0108] 진균 세포 게놈을 변형하기 위한 방법
- [0109] 진균 세포, 예컨대 사상균 세포 내 게놈좌와 공여체 DNA의 상동성 재조합을 촉진하기 위한 가이드 RNA/Cas 엔도뉴클레아제 시스템을 채용한 방법이 제공된다.
- [0110] 본 개시의 양태는 Cas 엔도뉴클레아제/가이드 폴리뉴클레오티드 복합체를 게놈좌와 상동성을 갖는 도메인을 포함하는 공여체 DNA와 함께 세포 내로 일시적으로 도입함으로써 진균 세포의 게놈 내의 게놈좌와 DNA 서열의 상

동성 재조합을 위한 방법을 포함한다. Cas 엔도뉴클레아제/가이드 폴리뉴클레오티드 복합체는 진균 세포의 게놈 내 원하는 표적 부위에서 작용할 수 있고, 여기서 “작용”이란 (위에서 정의된 것처럼) 가이드 폴리뉴클레오티드의 서열에 의해 안내된 Cas 엔도뉴클레아제가 표적 부위에서 DNA의 양 가닥 중 하나를 절단하는 것을 의미한다.

[0111] Cas 엔도뉴클레아제, 가이드 폴리뉴클레오티드, 및 공여체 DNA의 도입은 형질감염, 형질도입, 형질전환, 전기충격유전자전달법(electroporation), 입자 충격, 세포 융합 기술 등을 포함한 임의의 편리한 방식으로 수행될 수 있다. 이들 성분들 각각은 사용자가 원할 시 동시에 또는 순차적으로 도입될 수 있다. 예를 들어, 진균 세포는 처음에 Cas 발현 DNA 작제물로 안정적으로 형질감염될 수 있고, 이어서 가이드 폴리뉴클레오티드 (직접 또는 가이드 폴리뉴클레오티드 발현 DNA 작제물을 사용하여) 안정적인 형질전환체에 도입할 수 있다. 이 설정은 사용자가 상이한 가이드 폴리뉴클레오티드가 독립적으로 도입될 수 있는 안정한 Cas 형질전환체 진균 세포군을 생성할 수 있기 때문에 유리할 수 있다(일부 경우에는 필요하다면 하나 초과 가이드 폴리뉴클레오티드가 동일한 세포 내로 도입될 수 있다). 일부 구현예에서, Cas 발현 진균 세포는 사용자에게 의해 수득되고, 따라서 사용자는 세포 내로 Cas 엔도뉴클레아제를 발현할 수 있는 재조합 DNA 작제물을 도입할 필요가 없으며, 오히려 Cas 발현 세포 내로 가이드 폴리뉴클레오티드를 도입할 필요만 있다.

[0112] 특정 구현예에서, 가이드 폴리뉴클레오티드는 가이드 폴리뉴클레오티드를 인코딩하는 발현 카세트(또는 유전자)를 포함하는 재조합 DNA 작제물을 진균 세포로 도입함으로써 도입된다. 일부 구현예에서, 발현 카세트는 진핵생물의 RNA 폴 III 프로모터에 작동 가능하게 연결된다. 이들 프로모터는 RNA 폴 III에 의한 전사가 RNA 폴 II 의존 프로모터로부터의 RNA 중합효소 II에 의한 전사에 발생하는 5' 캡 구조 또는 폴리아데닐화의 첨가로 이어지지 않기 때문에 특히 흥미롭다. 특정 구현예에서, RNA 폴 III 프로모터는 사상균 세포 U6 중합효소 III 프로모터이다(예컨대, 서열 번호 40 및 이의 기능적 변이체, 예컨대 서열 번호 41; 아래에서 더 상세히 기재됨).

[0113] 이중 가닥 절단이 숙주 세포의 게놈 DNA에서 유도 될 때(예컨대, 표적 부위에서 Cas 엔도뉴클레아제/가이드 RNA 복합체의 활성화, 이중 가닥 엔도뉴클레아제 활성을 갖는 복합체), 세포의 DNA 복구 메커니즘이 활성화되어, 오류가 발생하기 쉬운 성질로 인하여 이중 가닥 절단 부위에서 돌연변이를 일으킬 수 있는 절단을 복구한다. 절단된 말단을 하나로 합치는 가장 일반적인 치료 메커니즘은 비상동성 말단 결합(NHEJ) 경로이다(Bleuyard 등, (2006) DNA Repair 5:1-12). 염색체의 구조적 완전성은 전형적으로 치료에 의해 보존되지만, 결실, 삽입 또는 다른 재배열이 가능하다(Siebert and Puchta, (2002) Plant Cell 14 : 1121-31, Pacher 등, (2007) Genetics 175 : 21 -9).

[0114] 놀랍게도, 사상 곰팡이의 경우 이중 가닥 절단에서 형질전환 된 DNA의 비상동성 삽입이 이중 가닥 절단에서 염색체 DNA의 두 말단 사이의 단순한 말단 결합보다 매우 선호된다는 것을 발견했다. 따라서, Cas 엔도뉴클레아제 또는 가이드 RNA가 DNA 작제물 또는 작제물들을 함유하는 발현 카세트로 형질전환함으로써 제공되는 경우, 이들 DNA 작제물 또는 이의 단편은 높은 빈도로 이중 가닥 절단부에 삽입된다. 이 삽입은 Cas 엔도뉴클레아제의 DNA 서열 또는 가이드 RNA 발현 작제물 및 이중 가닥 절단 주위의 서열 사이에 상동성이 없는 경우에 발생한다. 이 과정은 또한 상동성 재조합보다 전체 공여체 DNA의 삽입이 선호되므로 공여체 DNA와 게놈과 사이의 상동성 재조합이 바람직할 때 문제가 된다. 진균 세포에서 자율적으로 유지될 것으로 예상되는 텔로미어 서열을 포함하는 벡터의 형태일지라도, 형질전환 된 DNA의 바람직하지 않은 삽입이 일어난다는 것을 발견했다.

[0115] 형질전환에 의해 흡수된 DNA는 게놈 내에서 안정한 형태로 통합되거나 일시적으로 유지될 수 있다. 일시적인 유지관리는 불안정한 표현형에 의해 인식될 수 있다. 예를 들어, DNA 흡수는 형질전환 DNA 상에 존재하는 마커 유전자에 대한 선택에 의해 인식될 수 있다. 형질전환 및 삽입 후, 형질전환체는 선택적 조건으로 다시 전이되기 전에 여러 세대 동안 비선택적 조건 하에서 성장될 수 있다. 안정한 형질전환체는 선택적 조건으로 전이한 후에 성장할 수 있지만, 불안정한 형질전환체는 형질전환된 DNA의 손실로 인해 선택적 조건으로 다시 전이한 후에는 성장할 수 없을 것이다. 아래의 실시예 부분에 나타난 바와 같이, 진균 세포/불안정한 형질전환체에서 일시적으로 Cas 엔도뉴클레아제 및/또는 가이드 RNA를 발현하는 것이 가능하다는 것을 나타냈다.

[0116] 불안정한 형질전환체가 요구되는 구현예에서, 자율 복제를 장려하는 텔로미어 서열을 갖는 플라스미드가 사용될 수 있다. 자율 복제 서열, 동위원소 서열 또는 기타 서열을 갖는 것과 같은 자율 복제를 위해 설계된 다른 유형의 플라스미드가 또한 사용될 수 있다. 놀랍게도, 트리코데르마 레에세이에서 복제, 자율 복제 서열, 동위원소 또는 텔로미어 서열의 기원이 알려지지 않은 플라스미드를 사용할 수 있음을 발견했다. 선택 가능한 마커에 대하여 불안정한 표현형을 나타내는 형질전환체를 선별함으로써, 벡터 DNA 삽입없이 효과적인 표적 부위 유전자

변형이 수득된다(예컨대, 공여체 DNA에서 상동성 영역과의 상동성 재조합).

- [0117] 본 개시의 특정 구현에는 반복부에 의해 선택적으로 플랭킹된 Cas 엔도뉴클레아제 발현 카세트 및 제1 선택 가능한 마커를 공팡이의 게놈에 통합하여 발현 카세트 및 제1 선택 가능한 마커의 후속 제거(루프-아웃)를 가능하게 하여 Cas 엔도뉴클레아제를 발현하는 숙주 세포를 생산한다. 이들 세포는 공여체 DNA와의 상동성 재조합을 포함하여 관심 유전적 변형을 얻기 위해 다양한 방법으로 사용될 수 있다.
- [0118] 예를 들어, Cas 엔도뉴클레아제를 발현하는 숙주 세포는 제2 선택 가능 마커 (및 임의로 별도의 공여체 DNA)를 함유하는 가이드 RNA 발현 카세트를 포함하는 DNA 작제물로 형질전환 될 수 있다. 제2 선택 가능 마커를 사용하기 위해 선택된 숙주 세포는 Cas 엔도뉴클레아제 활성을 가능하게 하고 게놈에서 관심 대상이 되는 정의된 목표 부위에 대한 표적화를 가능하게 하는 이 DNA 작제물로부터 가이드 RNA를 발현한다. 제2 선택 가능 마커와 관련하여 불안정한 표현형을 나타내는 형질전환체에 대한 이들 숙주 세포를 선별하면 DNA 작제물 삽입없이 변형된 관심 부위(예컨대, 공여체 DNA와의 상동성 재조합)를 갖는 숙주 세포를 수득할 수 있다.
- [0119] 다른 예로서, Cas 엔도뉴클레아제를 발현하는 숙주 세포를 유도하여 시험관내 합성된 가이드 RNA를 흡수하여 Cas 엔도뉴클레아제 활성을 가능하게 하고 게놈 내 정의된 부위를 표적화할 수 있다. 일부 경우에는, 가이드 RNA 및 선택 가능 마커 유전자를 지닌 분리된 DNA 작제물 모두의 흡수를 유도하여 DNA를 취했고 고주파수에서 가이드 RNA를 동시에 흡수할 것으로 예상되는 세포를 선택하게 하는 것이 바람직할 것이다. 상기와 같이, 벡터 DNA 삽입없이 관심 유전자 변형(예컨대, 공여체 DNA와의 상동성 재조합)에 대한 선택 가능 마커에 대하여 불안정한 표현형을 나타내는 형질전환체를 선별한다.
- [0120] 또 다른 예로서, Cas 엔도뉴클레아제를 발현하는 숙주 세포는 번역시 Cas 엔도뉴클레아제를 “표적 균주”에게 제공할 수 있는 “헬퍼 균주(helper strain)”를 생성하기 위해 사용될 수 있다. 간단히 말해서, 헤테로카리온은 헬퍼 균주 및 표적 균주 사이에서, 예를 들어 각 균주로부터의 원형질체의 융합 또는 사상균의 종에 따른 균사의 문합에 의해 생성될 수 있다. 헤테로카리온의 유지는, 부모 균주가 성장할 수 없지만 헤테로카리온이 보상적으로 인해 성장할 수 있도록 각각의 부모 균주에서 적절한 영양의 또는 다른 마커 유전자 또는 돌연변이 및 적절한 선택 배지에서 성장에 의존할 것이다. 헤테로카리온 형성시 또는 이후에, 가이드 RNA가 형질감염 (및 선택적으로 공여체 DNA)에 의해 도입된다. 가이드 RNA는 직접 도입되거나 Cas 엔도뉴클레아제 발현 카세트 및 선택 가능 마커 유전자를 갖는 DNA 작제물을 통해 도입될 수 있다. Cas 엔도뉴클레아제는 헬퍼 균주 핵 내의 유전자로부터 발현되고 헤테로카리온의 세포질에 존재한다. Cas 엔도뉴클레아제는 게놈에서 원하는 표적 부위(들)를 표적으로 하는 활성 복합체를 생성하기 위해 가이드 RNA와 결합한다. 이어서, 포자는 헤테로카리온으로부터 회수되고, 표적 부위에서 또는 표적 부위 근처에서 변형(예컨대, 게놈좌에서 공여체 DNA와의 상동성 재조합)으로 표적 균주를 회수하기 위해 선택 또는 선별된다. 발현 카세트가 가이드 RNA를 도입하는데 사용되는 경우, 가이드 RNA 발현 작제물이 안정적으로 유지되지 않는 헤테로카리온이 선택된다.
- [0121] 전술한 바와 같이, 본 개시의 방법은 Cas/가이드 RNA 복합체의 표적 부위의 각 측상의 염색체 DNA의 영역과 DNA 서열 상동성을 갖는 DNA 작제물을 세포(또는 공여체 DNA)에 도입하는 단계를 포함한다. 의도는 DNA 단편이 상동성 통합/재조합에 의해 통합되어, 표적 부위에서 DNA의 절단을 복구하고, 대부분의 경우 원하는 위치에서(예컨대, Cas/가이드 RNA 복합체의 표적 부위에서 또는 표적 부위 근처에서) 게놈에 변화를 도입하려는 것이다. 많은 생물체에서, 염색체 DNA의 이중 가닥 절단은 그 부위에서 선형 DNA 단편의 상동성 통합을 자극한다. 놀랍게도, 작용하는 NHEJ 경로를 갖는 사상 균류에서, 공여체 단편이 도입될 때조차, 이중 가닥 절단에서의 비상동성 삽입에 의한 DNA의 삽입이 선형 DNA 단편의 상동성 재조합보다 매우 선호된다는 것을 발견했다.
- [0122] 진균 세포에서의 DNA 복구에 대해서, 기능적 NHEJ 경로의 존재시, 오류가 발생하기 쉬운 복구가 이중 가닥 절단 부위에서 상동성 재조합보다 매우 선호된다는 것을 발견했다. 다시 말해서, 사상균 세포에서 이중 가닥 절단의 DNA 복구에 관해서, 기능적 NHEJ 경로의 존재 하에서, 절단시 DNA의 비상동성 삽입이 (1) DNA 삽입없이 비상동성 말단 결합 및 (2) 공여체 DNA로 이중 가닥 절단 부위에서 상동성 재조합보다 매우 선호된다는 것을 발견했다. 따라서, 본 발명의 특정 양태에서, 집단 내 진균 세포의 표적 부위에서 비상동성 말단 결합(NHEJ) 경로의 기능화는 억제되거나, 활성화되지 않거나, 비기능적이거나, 또는 감소된다. 이는 임의의 편리한 방식으로 달성될 수 있으며, 그 중 일부는 이하에 설명된다.
- [0123] 일부 구현예에서, 진균 세포의 표적 부위에서 비상동성 말단 결합(NHEJ) 경로의 기능화는, NHEJ 경로의 하나 이상의 성분이 비기능적이거나, 감소된 활성을 갖도록(예를 들어, 게놈으로부터 결실되거나 돌연변이되어 비기능적이거나 덜 활성인) 진균 숙주 세포를 변화시킴으로써 억제되거나, 활성화되지 않거나, 비기능적이거나, 또는 감소된다. 진균 세포의 이러한 변경은 유전자 결실, 유전자 돌연변이, 우성 간섭 재조합 단백질의 발현, 유전자

대체, 예컨대 안티센스 RNA/RNAi 방법론 등을 사용하는 유전자 발현 억제 등의 임의의 편리한 수단에 의해 달성될 수 있다. 특정 양태에서, NHEJ 경로의 하나 이상의 성분은 ku80, ku70, rad50, mre11, xrs2, lig4 및 xrs로 이루어진 군으로부터 선택된다. 하나의 예로서, 본 발명의 양태에서 사용되는 진균 세포는 ku80의 발현 및/또는 활성을 저해하는 유전자 변형을 포함한다.

[0124] 추가의 구현예에서, 진균 세포의 표적 부위에서 비상동성 말단 결합(NHEJ) 경로의 기능화는 DNA Nicking 활성을 갖는 Cas 엔도뉴클레아제를 사용함으로써 억제되거나, 활성화되지 않고, 비기능적이거나, 감소된다. 즉, 기능화는 표적 부위(Cas Nickase라고도 부름)에서 DNA의 한 가닥만을 절단한다. DNA의 이중 가닥 절단과는 달리, Nick은 NHEJ 경로를 활성화시키지는 않지만, 하나 이상의 상동성 영역을 갖는 공여체 DNA로 절단 부위에서 또는 절단 부위 근처에서 상동성 재조합을 촉진시키기에 충분하다. 야생형 Cas 엔도뉴클레아제의 재조합 변이체인 다수의 Cas Nickase가 당 기술에 기재되어 있으며(예를 들어 상기 정의 참조), 개시된 방법에서 사용될 수 있다.

[0125] 일부 예에서, 공여체 DNA는 진균 세포의 게놈에서 상응하는 제1 및 제2 영역에 상동인 제1 영역 및 제2 영역을 포함하며, 상동성 영역은 일반적으로 게놈 DNA가 Cas 엔도뉴클레아제에 의해 절단되는 표적 부위를 포함하거나 둘러싼다. 상동성의 이들 영역은 상동성 재조합을 촉진하여 상동성의 상응하는 게놈 영역은 공여체 DNA와 게놈 간의 DNA 교환을 초래한다. 이와 같이, 제공된 방법은 진균 세포 게놈의 표적 부위의 절단 부위에서 또는 절단 부위 근처에서 공여체 DNA의 관심 폴리뉴클레오타이드의 통합을 초래하여, 원래의 표적 부위를 변경시킴으로써 변형된 게놈 표적 부위를 생성한다.

[0126] 주어진 게놈 영역과 공여체 DNA에서 발견되는 해당 상동성 영역 사이의 구조적 유사성은 상동성 재조합이 일어날 수 있게 하는 임의의 서열 동일성의 정도가 될 수 있다. 예를 들어, 공여체 DNA의 "상동성 영역"과 진균 세포 게놈의 "게놈 영역"이 공유하는 상동성 또는 서열 동일성의 양은 적어도 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 95%, 96%, 97%, 98%, 99% 또는 심지어 100%의 서열 동일성을 가져서, 서열이 상동성 재조합을 겪는다.

[0127] 공여체 DNA상의 상동성 영역은 표적 부위에 인접한 임의의 서열과 상동성을 가질 수 있다. 일부 구현예에서 상동성의 영역들은 표적 부위의 바로 옆에 있는 게놈 서열에 대한 중요한 서열 상동성을 공유하지만, 상동성의 영역이 표적 부위에게 5' 또는 3' 더 길 수 있는 영역에 충분한 상동성을 갖도록 설계될 수 있다는 것이 인식된다. 또 다른 구현예에서, 상동성 영역은 하류 게놈 영역과 함께 표적 부위의 단편과 상동성을 가질 수 있다. 하나의 구현예에서, 상동성의 제1 영역은 표적 부위의 제1 단편을 추가로 포함하고, 상동성의 제2 영역은 표적 부위의 제2 단편을 포함하며, 제1 단편 및 제2 단편은 상이하다.

[0128] Cas 엔도뉴클레아제 및 가이드 폴리뉴클레오타이드 발현 작제물과 마찬가지로, 공여체 DNA는 임의의 편리한 수단(본원의 다른 곳에서 논의됨)에 의해 도입될 수 있다.

[0129] 특정 구현예에서, Cas 엔도뉴클레아제는 Cas9 엔도뉴클레아제이다(예컨대, "RNA-directed DNA Cleavage by the Cas9-crRNA Complex"라는 제목의 WO 2013141680 참조). Cas9 엔도뉴클레아제의 예로는 스트렙토코쿠스 종(예컨대, S. 파이오제네스, S. 뮤탄스 및 S. 더모필루스), 캄피로박테르 종(Campylobacter sp.)(예컨대, C. 제주니), 나이세리아 종(Neisseria sp.)(예컨대, N. 메닝기티데스), 프란시셀라 종(Francisella sp.)(예컨대, F. 노비시다) 및 파스튜렐라 종(Pasteurella sp.)(예컨대, P. 물토시다)가 있다(본원에 참고로 통합된 Fonfara 등, Nucleic Acids Res., 2013, pages 1-14에 기재된 Cas9 엔도뉴클레아제를 참조). 일부 구현예에서, Cas 엔도뉴클레아제는 최적화된, 예컨대 진균 세포에서의 발현을 위해 최적화된 Cas9 엔도뉴클레아제 유전자(예를 들어, 하기 기재된 바와 같이 서열 번호 44, 예를 들어, 서열 번호 7을 포함하는 Cas9 인코딩 유전자)에 의해 인코딩된다.

[0130] 특정 예에서, Cas 엔도뉴클레아제 유전자는 세포에서 발현되는 Cas 엔도뉴클레아제/가이드 폴리뉴클레오타이드 복합체가 핵으로 효율적으로 운반되도록 핵 위치 신호를 인코딩하는 하나 이상의 폴리뉴클레오타이드에 작동 가능하게 연결된다. 임의의 편리한 핵 위치 신호, 예를 들어, Cas 코돈 영역의 상류이자 Cas 코돈 영역과 동일 프레임에 존재하는 SV40 핵 위치 신호를 인코딩하는 폴리뉴클레오타이드 및 Cas 코돈 영역의 하류이자 Cas 코돈 영역과 동일 프레임에 존재하는 T. 레에세이 blr2(청색광 조절인자 2)로부터 유도된 핵 위치 신호를 인코딩하는 폴리뉴클레오타이드가 사용될 수 있다. 다른 핵 위치 신호가 사용될 수 있다.

[0131] 본 발명의 특정 구현예에서, 가이드 폴리뉴클레오타이드는 유형 II Cas 엔도뉴클레아제와 복합체를 형성할 수 있는 유형 II CRISPR/Cas 시스템의 crRNA 영역(또는 crRNA 단편) 및 tracrRNA 영역(또는 tracrRNA 단편)을 포함하는 가이드 RNA이다. 상기한 바와 같이, 가이드 RNA/Cas 엔도뉴클레아제 복합체는 Cas 엔도뉴클레아제를 진균

세포 게놈 표적 부위로 향하게 할 수 있어, Cas 엔도뉴클레아제가 게놈 타겟 부위에 이중 가닥 절단을 도입할 수 있게 한다. 일부 경우에서, RNA/Cas9 엔도뉴클레아제 복합체를 안내하는 RNA는 crRNA와 별도의 tracrRNA를 포함하는 듀플렉스이다. 다른 예에서, 가이드 RNA는 (때때로 본원에서 융합된 가이드 RNA로 지칭되는) crRNA 영역 및 tracrRNA 영역 모두를 포함하는 단일 RNA 분자이다. 융합된 가이드 RNA 대 듀플렉스화된 crRNA-tracrRNA를 사용하는 한 가지 장점은 융합된 가이드 RNA를 발현하기 위해 하나의 발현 카세트만을 만들 필요가 있다는 것이다.

[0132] 본원에 개시된 방법에 사용되는 숙주 세포는, 필라 아스코마이코타(*phyla Ascomycota*), 바시디오마이코타 (*Basidiomycota*), 키트리디오마이코타(*Chytridiomycota*), 및 자이코마이코타(*Zygomycota*) (Hawksworth 등, In, Ainsworth and Bisby 's Dictionary of The Fungi, 8<sup>th</sup> edition, 1995, CAB International, University Press, Cambridge, UK에서 정의됨)뿐만 아니라 오오마이코타(*Oomycota*) (Hawksworth 등, supra에서 인용) 및 모든 불완전 균류(*mitosporic fungi*) (Hawksworth 등, supra)에서 유래된 임의의 진균 숙주 세포일 수 있다. 특정 구현예에서, 진균 숙주 세포는 효모 세포, 예컨대, 칸디다, 한세누라, 클루베로마이세스, 피키아, 사카로마이세스, 스키토사카로마이세스 또는 야로위아 세포이다. 효모 종은 다음을 포함하지만 이에 한정되지 않는다: 사카로마이세스 칼스베르겐시스(*Saccharomyces carlsbergensis*), 사카로마이세스 세레비시애(*Saccharomyces cerevisiae*), 사카로마이세스 디아스타티쿠스(*Saccharomyces diastaticus*), 사카로마이세스 도우글라시(*Saccharomyces douglasii*), 사카로마이세스 클루베리(*Saccharomyces kluyveri*), 사카로마이세스 노르벤시스(*Saccharomyces norbensis*), 사카로마이세스 오비포르미스(*Saccharomyces oviformis*), 클루베로마이세스 락티스(*Kluyveromyces lactis*), 및 야로위아 리폴리티카(*Yarrowia lipolytica*) 세포. 추가의 구현예에서, 진균 세포는 트리코데르마, 페니실리움, 아스퍼질러스, 휴미콜라, 크리소스포리움, 푸사리움, 뉴로스포라, 마이셀리오프토라, 하이포크레아 및 에메리셀라의 종을 포함하지만 이에 한정되지 않는 사상균 세포이다. 예를 들어, 사상 곰팡이 T. 레에세이 및 A. 나이저는 개시된 방법의 양태에서 사용된다.

[0133] 표적 부위가 필요로 하는 프로토스페이스 인접 모티프(*protospacer adjacent motif(PAM)*)를 포함하는 한, 사실상 진균 세포 게놈의 임의의 부위는 개시된 방법을 사용하여 표적화될 수 있다. S. 파이로제네스 Cas9의 경우, PAM은 서열 NGG (5' 에서 3'; N은 A, G, C 또는 T)를 가지며, 따라서 게놈의 표적 부위의 선택에 유의한 제한을 부과하지 않는다. 다른 공지된 Cas9 엔도뉴클레아제는 상이한 PAM 부위를 갖는다(예컨대, Fonfara 등, *Nucleic Acids Res.*, 2013, pages 1-14에 기재된 Cas9 엔도뉴클레아제 PAM 부위 참조: 본원에 참고로 통합됨).

[0134] 표적 부위의 길이는 변할 수 있으며, 예를 들어 적어도 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 또는 그 이상의 뉴클레오티드 길이를 갖는 표적 부위를 포함한다. 표적 부위는 팔린드로믹일 수도 있다. 즉, 한 가닥 상에 있는 서열은 상보적 가닥에서 반대 방향으로 동일하게 관독된다. 절단 부위는 표적 서열 내에 존재할 수 있거나 절단 부위는 표적 서열 외부에 존재할 수 있다. 또 다른 변형에서, 절단은 블런트 엔드 컷을 생성하기 위해 서로 바로 마주 보는 뉴클레오티드 위치에서 발생할 수 있거나, 또는 다른 경우에는 "끈끈한 말단(*sticky ends*)"이라고도 부르는, 5' 오버행 또는 3' 오버행일 수 있는 단일 가닥 오버행을 생성하도록 절개가 틀어질 수 있다.

[0135] 일부 경우에, 표적 부위가 (가이드 폴리뉴클레오티드의 crRNA 서열 내에서) 가이드 폴리뉴클레오티드의 관련 서열과 100% 동일하지 않다는 것을 의미하는, 진균 세포의 게놈에서의 활성 변이체 표적 서열도 사용될 수 있다. 이러한 활성 변이체는 주어진 표적 부위와 적어도 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 이상의 서열 동일성을 포함할 수 있고, 활성 변이체 표적 서열은 생물학적 활성을 보유하고, 따라서 Cas 엔도뉴클레아제에 의해 인식되고 절단 될 수 있다. 엔도뉴클레아제에 의한 표적 부위의 이중 가닥 절단을 측정하기 위한 분석은 당업계에 공지되어 있으며, 일반적으로 인식 부위를 포함하는 DNA 기질 상의 작용제의 전체 활성 및 특이성을 측정한다.

[0136] 관심 표적 부위는 관심 유전자의 영역 내에 위치하는 표적을 포함한다. 관심 유전자 내의 영역들의 비제한적 예로는 오픈 리딩 프레임, 프로모터, 전사 조절 요소, 번역 조절 요소, 전사 종결인자 서열, mRNA 스플라이스 부위, 단백질 코딩 서열, 인트론 부위, 및 인트론 증진 모티프가 있다.

[0137] 특정 구현예에서, 진균 세포의 게놈의 변형은 검출될 수 있는 표현형 효과를 초래하고, 많은 예에서 사용자의 바람직한 결과이다. 비 제한적인 예로는 선택 가능한 세포 성장 표현형(예컨대, 항생제에 대한 저항성 또는 민감성, 영양 요구성 특성의 증가 또는 감소, 증가 또는 감소된 성장률 등), 검출 가능한 마커(예컨대, 형광 마커, 세포 표면 분자, 발색 효소 등), 그리고 배양 상등액에서 그 활성이 검출될 수 있는 효소의 분비가 있다.

[0138] 진균 세포 게놈의 변형이 표현형 효과를 가져올 때, 표현형 마커인 관심 폴리뉴클레오티드를 포함하는 공여체

DNA가 종종 사용된다. 흔히 특정 배양 조건 하에서, 그것을 포함하는 진균 세포를 확인하거나 이를 혹은 이에 반하여 선택할 수 있게 하는 선택 가능하거나 선별 가능한 임의의 마커를 포함하여 임의의 편리한 표현형 마커가 사용될 수 있다. 따라서, 본 발명의 일부 양태에서, 원하는 계능 변형을 갖는 진균 세포의 확인은 표적 부위에서 변형을 갖는 세포를 선택하거나 선별하는 조건 하에서 Cas 엔도뉴클레아제 및 가이드 폴리뉴클레오티드(그리고 선택적으로 공여체 DNA)를 받은 진균 세포군을 배양하는 것을 포함한다. 진균 세포(선택 가능 마커라고도 지칭됨)에서의 효소 활성의 증가 또는 감소, 예를 들어 항생제 내성의 획득 또는 영양 요구성 마커의 증가/감소를 평가하는 것을 포함하여 임의 유형의 선택 시스템이 사용될 수 있다.

[0139] 일부 예에서, 진균 세포에서의 계능 변형은 시퀀싱, PCR, 서던 블롯, 제한 효소 분석 등을 포함하고, 이러한 방법들의 조합을 포함하는 임의의 편리한 방법을 사용하여 직접 검출된다.

[0140] 일부 구현예에서, 효소, 예를 들어 아세틸 에스테라아제, 아미노펩티다아제, 아밀라아제, 아라비나아제, 아라비노푸라노시다아제, 카복시펩티다아제, 카탈라아제, 셀룰라아제, 키티나아제, 큐티나아제, 데옥시리보뉴클레아제, 에피메라아제, 에스테라아제,  $\alpha$ -갈락토시다아제,  $\beta$ -갈락토시다아제,  $\alpha$ -글루카나아제, 글루카나아제, 엔도- $\beta$ -글루카나아제, 글루코아밀라아제, 글루코스 옥시다아제,  $\alpha$ -글루코시다아제,  $\beta$ -글루코시다아제, 글루쿠로니다아제, 헤미셀룰라아제, 핵소오스 옥시다아제, 하이드로라아제, 인베르타제, 이소메라제, 락카아제, 리파아제, 리아제, 만노시다아제, 옥시다아제, 옥시도레덕타아제, 펙테이트 리아제, 펙틴 아세틸 에스테라아제, 펙틴 디폴리머라아제, 펙틴 메틸 에스테라아제, 펙틴 분해 효소, 페록시다아제, 페놀옥시다아제, 피타아제, 폴리갈락투로나아제, 프로테아제, 람노 - 갈락투로나아제, 리보뉴클레아제, 트랜스퍼라아제, 수송 단백질, 트랜스글루타미나아제, 크실라나아제, 핵소오스 옥시다아제 및 이들의 조합을 인코딩하는 유전자를 포함하여, 특정 유전자들이 개시된 방법을 사용하여 변형을 위해 표적화된다.

[0141] 본원에 기재된 방법을 구현하기 위한 다양한 변형이 있다. 예를 들어, 진균 숙주 세포에 외인성 서열로서 존재하는 Cas 발현 카세트를 갖는 대신에, 이 카세트는 진균 숙주 세포의 계능에 통합될 수 있다. 이러한 부모 세포주를 생성함으로써, 사용자는 본원의 다른 곳에서 상세히 설명된 바와 같이 관심 계능 부위를 표적으로 하는 원하는 가이드 RNA(예컨대, 가이드 RNA 발현 벡터로서)를 간단히 도입 할 수있게 된다. 이들 구현예들의 일부에서, 통합된 Cas 유전자는 필요한 경우 계능으로부터의 후속 루프-아웃/제거를 위해 거기에 인접한 폴리뉴클레오티드 리피트를 포함하도록 설계될 수 있다.

[0142] 진균 세포의 조성물

[0143] 본 발명의 양태는 변형된 계능을 갖는 수득된 진균 세포뿐만 아니라 상기 기술된 방법을 수행하는데 사용되는 형질전환 진균 세포를 포함한다. 따라서, 본 발명의 구현예는 본원에 기재된 방법의 임의의 양태에 의해 생산된 재조합 진균 세포뿐만 아니라 이들을 생산하는데 사용되는 임의의 부모 진균 세포를 포함한다.

[0144] 본 발명의 특정 구현예는 Cas 엔도뉴클레아제를 포함하는 재조합 진균 세포에 관한 것으로, 여기서 Cas 엔도뉴클레아제는 세포에서 재조합 DNA 작제물(제1 재조합 DNA 작제물)로부터 발현된다. 일부 구현예에서, 재조합 진균 세포는 비기능성 또는 활성이 감소된 NHEJ 경로를 갖는다. Cas 엔도뉴클레아제 및 위에서 상세히 기술된 Cas 엔도뉴클레아제를 인코딩하는 폴리뉴클레오티드의 구현예는 이 부분의 진균 세포 조성물에서 사용된다(그 중 일부는 하기에 기재됨). 이 진균 세포는 관심 대상인 원하는 계능 변형을 갖는 진균 세포를 생성하기 위한 부모로서 사용되며, 계능 변형을 생성하는 것은 세포 내로 가이드 폴리뉴클레오티드(예를 들어, 발현 카세트를 통해)를 도입하여 유전자 변형을 견인하는 Cas/가이드 폴리뉴클레오티드 복합체를 형성할 수 있다(위에서 기술 한 바와 같이). 특정 양태에서, NHEJ 경로의 하나 이상의 성분은 재조합 진균 세포에서 비기능적이거나 감소된 활성, 예컨대, ku80, ku70, rad50, mre11, xrs2, lig4, xrs 및 이들의 조합을 갖는다. 하나의 특별한 구현예에서, 재조합 진균 세포는 ku80의 발현 및/또는 활성을 저해하는 유전자 변형을 갖는다. 유전자 결실, 유전자 돌연변이, 우성 간섭 재조합 단백질의 발현, 유전자 대체, 예컨대 안티센스 RNA/RNAi 방법론 등을 사용하는 유전자 발현 억제제를 포함하지만, 이들에 한정되지 않는 NHEJ 경로 성분(들)의 파괴를 이루기 위한 임의의 편리한 유전자 변형이 채용될 수 있다.

[0145] 특정 양태에서, 재조합 진균 세포는 가이드 RNA를 발현할 수 있는 제2 재조합 DNA 작제물을 추가로 포함하며, 여기서 가이드 RNA 및 Cas 엔도뉴클레아제는 Cas 엔도뉴클레아제가 재조합 진균 세포의 계능 내 표적 부위에서 이중 가닥 절단을 유도할 수 있게 하는 복합체를 형성할 수 있다. 가이드 폴리뉴클레오티드 및 위에서 상세히 기술된 가이드 폴리뉴클레오티드를 인코딩하는 폴리뉴클레오티드의 구현예는 이 부분의 진균 세포 조성물에서 사용된다(그 중 일부는 하기에 기재됨). 가이드 RNA의 발현은 진행 생물 RNA 폴 III 프로모터에 의해 유도될 수 있으며, 특정 구현예에서 RNA 폴 III 프로모터는 사상균 세포 U6 유전자 프로모터(예컨대, 하기에 더 상세히

기술된 바와 같은 서열 번호 40 또는 서열 번호 41) 및 이의 기능적 변이체이다. 이 복합체의 작용은 진균 세포의 표적 부위에서 게놈 DNA 서열을 변형시켜(위에서 기술된 바와 같이), 표적 부위에 (또는 근처에) 변형된 진균 세포를 생성할 수 있다. 변형은 하나 이상의 뉴클레오타이드의 결실, 하나 이상의 뉴클레오타이드의 삽입, 하나 이상의 뉴클레오타이드의 치환, 또는 이들의 임의의 조합을 포함할 수 있다. 일부 구현예에서, 진균 세포는 관심 폴리뉴클레오타이드를 갖는 공여체 DNA를 추가로 포함한다. 공여체 DNA에서 관심 폴리뉴클레오타이드는 표적 부위(예를 들어, 상동성 재조합에 의해 게놈에 삽입)의 게놈에 (또는 근처에) 존재할 수 있으며, 표적 부위에서 Cas/가이드 폴리뉴클레오타이드 복합체의 작용에 의해 유도되는 과정.

[0146] 특정 구현예에서, 재조합 폴리뉴클레오타이드에 의해 인코딩된 Cas 엔도뉴클레아제는 Cas9 엔도뉴클레아제이다. Cas 엔도뉴클레아제 및 이의 기능적 단편을 포함하지만 이들에 한정되지 않는 임의의 Cas9 엔도뉴클레아제가 다음 박테리아 종으로부터 인코딩될 수 있다: 스트렙토코쿠스 종(예컨대, *S. 파이오제네스*, *S. 뮤탄스* 및 *S. 더모필루스*), 캄피로박테르 종(예컨대, *C. 제주니*), 나이세리아 종(예컨대, *N. 메닝기티데스*), 프란시셀라 종(예컨대, *F. 노비시다*) 및 파스튜렐라 종(예컨대, *P. 몰토시다*)가 있다(본원에 참고로 통합된 Fonfara 등, *Nucleic Acids Res.*, 2013, pages 1-14에 기재된 Cas9 엔도뉴클레아제를 참조). 일부 구현예에서, Cas 엔도뉴클레아제 유전자를 인코딩하는 폴리뉴클레오타이드는 사상균 숙주 세포에서 발현을 위해 최적화된 폴리뉴클레오타이드, 예를 들어 서열 번호 44(*S. 파이오제네스* Cas9 엔도뉴클레아제의 사상균 세포 코돈 최적화 버전) 또는 서열 번호 7 (서열 번호 44를 포함하고 또한 N- 및 C- 말단 NLS 서열을 포함)에 나타난 폴리뉴클레오타이드이다. 서열 번호 44 또는 서열 번호 7의 동의어 변이체를 포함하는 추가적인 코돈-최적화된 Ca9 유전자가 사용될 수 있다. 전술한 바와 같이, Cas 엔도뉴클레아제는 Cas 엔도뉴클레아제가 세포질에서 핵으로 그 작용 부위까지, 즉 세포의 핵 내로 이동하는 것을 향상시키는 기능을 하는 하나 이상의 핵 위치 신호에 작동 가능하게 연결될 수 있다. SV40 핵 위치 신호, T. 레에세이 blr2 (청색광 조절인자 2) 유전자로부터 유래된 핵 위치 신호, 또는 이들의 조합을 포함하는 임의의 편리한 핵 위치 신호가 사용될 수 있다.

[0147] 필라 아스코마이코타(phyla Ascomycota), 바시디오마이코타(Basidiomycota), 키트리디오마이코타 Chytridiomycota), 및 자이코마이코타(Zygomycota)(Hawksworth 등, In, Ainsworth and Bisby 's Dictionary of The Fungi, 8<sup>th</sup> edition, 1995, CAB International, University Press, Cambridge, UK에서 정의됨)뿐만 아니라 오오마이코타(Oomycota)(Hawksworth 등, supra에서 인용) 및 모든 불완전 균류(mitosporic fungi)(Hawksworth 등, supra)로부터 유래된 진균 숙주 세포를 포함하는 매우 다양한 사상균 숙주 세포들 중 임의의 것이 본 발명에서 사용된다. 특정 구현예에서, 진균 숙주 세포는 효모 세포, 예컨대, 칸디다, 한세누라, 클루베로마이세스, 피키아, 사카로마이세스, 스킴오사카로마이세스 또는 야로위아 세포이다. 효모 종은 다음을 포함하지만 이에 한정되지 않는다: 사카로마이세스 칼스베르겐시스(*Saccharomyces carlsbergensis*), 사카로마이세스 세레비시애(*Saccharomyces cerevisiae*), 사카로마이세스 디아스타티쿠스(*Saccharomyces diastaticus*), 사카로마이세스 도우글라시(*Saccharomyces douglasii*), 사카로마이세스 클루베리(*Saccharomyces kluyveri*), 사카로마이세스 노르벤시스(*Saccharomyces norbensis*), 사카로마이세스 오비포르미스(*Saccharomyces oviformis*), 클루베로마이세스 락티스(*Kluyveromyces lactis*), 및 야로위아 리폴리티카(*Yarrowia lipolytica*) 세포. 추가의 구현예에서, 진균 세포는 트리코테르마, 페니실리움, 아스퍼질러스, 휴미콜라, 크리소스포리움, 푸사리움, 뉴로스포라, 마이셀리오프토라, 더모마이세스, 하이포크레아 및 에메리셀라의 종을 포함하지만 이에 한정되지 않는 사상균 세포이다. 예를 들어, 사상균 트리코테르마 레에세이, P. 크리소제넘, M. 더모필라, 더모마이세스 라누지노수스, A. 오리제 및 A. 나이거가 본 개시의 양태에서 사용된다.

[0148] 위에서 상세히 설명한 바와 같이, 본 발명은 일반적으로 사상균 세포의 게놈에서 관심 표적 부위를 변형시키는 데 유용한 방법 및 조성물에 관한 것이다. 관심 대상인 특정 표적 부위는 이러한 방법 및 조성물의 사용자에게 의해 결정되며, 프로모터, 조절 서열, 종결인자 서열, 조절 요소 서열, 스플라이스 부위, 코딩 서열, 폴리유비퀴티네이션 부위, 인트론 부위, 및 인트론 증진 모티프를 포함하는 관심 유전자의 영역 내에 위치한 부위를 포함한다. 아울러, 아세틸 에스테라아제, 아미노펩티다아제, 아밀라아제, 아라비나아제, 아라비노푸라노시다아제, 카복시펩티다아제, 카탈라아제, 셀룰라아제, 키티나아제, 큐티나아제, 데옥시리보뉴클레아제, 에피메라아제, 에스테라아제,  $\alpha$ -갈락토시다아제,  $\beta$ -갈락토시다아제,  $\alpha$ -글루카나아제, 글루카닐라아제, 엔도- $\beta$ -글루카나아제, 글루코아밀라아제, 글루코스 옥시다아제,  $\alpha$ -글루코시다아제,  $\beta$ -글루코시다아제, 글루쿠로니다아제, 헤미셀룰라아제, 핵소오스 옥시다아제, 하이드로라아제, 인베르타제, 이소메라제, 락카아제, 리파아제, 리아제, 만노시다제, 옥시다아제, 옥시도레덕타아제, 펙테이트 리아제, 펙틴 아세틸 에스테라아제, 펙틴 디폴리머라아제, 펙틴 메틸 에스테라아제, 펙틴 분해 효소, 페록시다아제, 페놀옥시다아제, 피타아제, 폴리갈락투로나아제, 프로테아제, 람노 - 갈락투로나아제, 리보뉴클레아제, 트랜스피라아제, 수송 단백질, 트랜스글루타미나아제, 크실라

나아제, 핵소오스 옥시다아제 및 이들의 조합을 포함하지만 이들에 한정되지 않는 관심 대상인 임의 유전자들이 사용자에게 의해 선택될 수 있다.

[0149] 재조합 폴리뉴클레오티드

[0150] 본 발명의 양태는 본원에 기재된 방법 및 조성물에 사용되는 재조합 폴리뉴클레오티드에 관한 것이다.

[0151] 본 발명의 구현에는 Cas 엔도뉴클레아제를 인코딩하는 진균 세포에 최적화된 뉴클레오티드 서열에 작동 가능하게 연결된 프로모터 서열을 갖는 재조합 폴리뉴클레오티드 DNA 작제물을 포함한다. 본 개시의 구현에는 Cas 엔도뉴클레아제를 인코딩하는 진균 세포에 최적화된 뉴클레오티드 서열 또는 Cas 엔도뉴클레아제를 인코딩하는 박테리아 세포에 최적화된 뉴클레오티드 서열에 작동 가능하게 연결된 프로모터 서열을 갖는 재조합 폴리뉴클레오티드 DNA 작제물을 포함한다. 전술한 바와 같이, 진균 세포에 최적화된 뉴클레오티드 서열 및 박테리아 세포에 최적화된 뉴클레오티드 서열에 인코딩된 Cas 엔도뉴클레아제는 가이드 RNA와 복합체 형성시에 표적 부위에서 작용할 수 있다. 임의의 Cas 엔도뉴클레아제는 Cas9 엔도뉴클레아제, 예를 들어 S. 피오게네스로부터의 Cas9를 포함하지만 이에 한정되지 않는, 최적화된 뉴클레오티드 서열에 의해 인코딩될 수 있다. 특정 구현예에서, 진균 세포에 최적화된 뉴클레오티드 서열은 사상균 세포에서의 발현에 위해 최적화된다. 예를 들어, 사상균 세포에 최적화된 서열은 Cas9 엔도뉴클레아제를 인코딩할 수 있고 서열 번호 44에 도시된 뉴클레오티드 서열(100% 동일성)을 포함하거나 Cas9 엔도 뉴클레아제를 인코딩할 수 있고, 서열 번호 44와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 동일한 뉴클레오티드 서열을 포함한다. 특정 구현예에서, 박테리아 세포에 최적화된 뉴클레오티드 서열은 대장균 세포에서의 발현을 위해 최적화된다. 예를 들어, 대장균 세포에 최적화된 서열은 Cas9 엔도뉴클레아제를 인코딩할 수 있고 서열 번호 65에 도시된 뉴클레오티드 서열(100% 동일성)을 포함하거나 Cas9 엔도뉴클레아제를 인코딩할 수 있고, 서열 번호 65와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 동일한 뉴클레오티드 서열을 포함한다.

[0152] 본 개시의 구현예들은 추가로 트리코테르마 종 RNA 중합효소 III에 의해 구동된 프로모터에 관한 것이다. RNA 폴 III 유도 프로모터로부터 RNA 폴 III에 의한 유전자의 전사는 RNA 폴 II 의존 프로모터로부터의 RNA 중합효소 II에 의한 전사시에 발생하는 5' 캡 구조 또는 폴리아데닐화의 첨가로 이어지지 않는다. 아래의 실시예에서 기술된 바와 같이, 전사 종결인자 서열뿐만 아니라 U6 유전자와 관련된 T. 레에세이에서 RNA 폴 III에 의해 유도된 프로모터 서열을 확인하였다. 전체 프로모터 서열은 서열 번호 40에 기재되어 있고, 종결인자 서열은 서열 번호 43에 기재되어 있다. 또한, U6 유전자 RNA 폴 III 구동 프로모터의 보다 짧은 버전이 확인되었고 서열 번호 41에 제시되어 있다. 따라서, 본 발명의 양태는 RNA 폴 III 유도된 프로모터로서 기능하고 서열 번호 40 또는 서열 번호 41과 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100%까지 동일한 뉴클레오티드 서열을 갖는 프로모터를 포함한다. 이 RNA 폴 III 유도 프로모터 서열은 임의의 이형 관심 서열을 발현하는데 사용된다. 따라서, 본 개시의 양태는 이종 관심 서열에 작동 가능하게 연결된 T. 레에세이 유래 RNA 폴 III 구동 프로모터 서열을 갖는 재조합 폴리뉴클레오티드 서열을 포함한다. 특정 구현예에서, 이종 서열은 가이드 폴리뉴클레오티드, 예를 들어 tracrRNA, crRNA 또는 단일 가이드 RNA를 인코딩하는 서열이다. 관심 대상인 특정 게놈 부위를 표적으로 삼은 가이드 RNA 인코딩 폴리뉴클레오티드는 실시예에 상세히 기술되어 있고 서열 번호 2 내지 6을 포함한다. 특정 구현예에서, 재조합 폴리뉴클레오티드는, 예를 들어 관심 이종 서열의 하류 부위(여기서, 하류 "는 당해 분야에서 통상적인 전사 방향을 지칭한다)에서 관심 이종 서열(RNA 폴 III 프로모터에 작동 가능하게 연결된)에 작동 가능하게 연결된 전사 종결인자 서열을 포함한다. 일부 구현예에서, 종결인자 서열은 서열 번호 43으로 도시된 폴리뉴클레오티드 서열 또는 이의 기능적 변이체를 포함한다. 따라서, 특정 구현예에서, 재조합 폴리뉴클레오티드는 종결인자에 작동 가능하게 연결된 관심 이종 서열(예를 들어, 가이드 RNA)에 작동 가능하게 연결된 RNA 폴 III 프로모터를 포함한다.

[0153] 본원에 개시된 조성물 및 방법의 비제한적인 예 또는 구현에는 하기와 같다:

[0154] 1. 사상균 세포의 게놈좌와 공여체 DNA를 상동성 재조합하기 위한 방법으로서: 상기 방법은:

[0155] a) 사상균 세포의 게놈좌에 대한 상동성을 가진 도메인을 포함하는 Cas 엔도뉴클레아제, 가이드 RNA, 및 공여체 DNA를 사상균 세포군에 도입하는 단계로서, Cas 엔도뉴클레아제 및 가이드 RNA는 Cas 엔도뉴클레아제가 사상균 세포의 게놈좌 내 또는 사상균 세포의 게놈좌 근처의 표적 부위에서 작용할 수 있게 하는 복합체를 형성할 수 있는 단계; 및

[0156] b) 게놈좌와 공여체 DNA의 상동성 재조합이 나타난 군으로부터 적어도 하나의 사상균 세포를 확인하는 단계를

포함하고,

- [0157] Cas 엔도뉴클레아제, 가이드 RNA 또는 둘 모두는 사상균 세포군으로 일시적으로 도입되는, 방법.
- [0158] 2. 구현예 1에 있어서, 사상균 세포들 내 표적 부위에서 비상동성 말단 결합(NHEJ) 메커니즘은 활성화되지 않거나, 비기능적이거나 감소되는, 방법.
- [0159] 3. 구현예 2에 있어서, 사상균 세포들 내 비상동성 말단 결합(NHEJ) 경로는 하나 이상의 비기능적 또는 활성-감소 성분들을 포함하는, 방법.
- [0160] 4. 구현예 3에 있어서, 하나 이상의 비기능적 또는 활성-감소 성분은 ku80, ku70, rad50, mre11, xrs2, lig4, xrs, 및 이의 조합들로 구성되는 군으로부터 선택되는, 방법.
- [0161] 5. 구현예 4에 있어서, 하나 이상의 비기능적 또는 활성-감소 성분은 ku80인, 방법.
- [0162] 6. 구현예 1 내지 5 중 어느 하나에 있어서, Cas 엔도뉴클레아제는 Cas 니카아제인, 방법.
- [0163] 7. 구현예 1 내지 5 중 어느 하나에 있어서, Cas 엔도뉴클레아제는 Cas9 엔도뉴클레아제 또는 이의 변이체인, 방법.
- [0164] 8. 구현예 7에 있어서, Cas9 엔도뉴클레아제 또는 이의 변이체는 스트렙토코쿠스 종(*Streptococcus* sp.), *S. 파이오제네스(pyogenes)*, *S. 뮤탄스(mutans)* 및 *S. 더모필루스(thermophilus)*, 캄피로박테르 종(*Campylobacter* sp.), *C. 제주니(jejuni)*, 나이세리아 종(*Neisseria* sp.), *N. 메닝기티데스(meningitides)*, 프란시셀라 종(*Francisella* sp.), *F. 노비시다(novicida)*, 및 파스투렐라 종(*Pasteurella* sp.), *P. 몰토시다(multocida)*로 구성되는 군으로부터 선택되는 종으로부터 전장 Cas9 또는 이의 기능적 단편을 포함하는, 방법.
- [0165] 9. 구현예 8에 있어서, Cas9 엔도뉴클레아제 또는 이의 변이체는 서열 번호 45 및 서열 번호 48 내지 53 중 어느 하나와 적어도 70% 동일성을 갖는 아미노산 서열을 포함하는, 방법.
- [0166] 10. 구현예 1 내지 9 중 어느 하나에 있어서, 공여체 DNA는 관심 폴리뉴클레오티드 서열을 포함하고 게놈좌에서 상동성 제조함은 게놈좌에서 관심 폴리뉴클레오티드 서열의 삽입을 초래하는, 방법.
- [0167] 11. 구현예 1 내지 10 중 어느 하나에 있어서, 상기 도입 단계는 Cas 엔도뉴클레아제용 발현 카세트를 포함하는 DNA 작제물을 사상균 세포들 내로 도입하는 단계를 포함하는, 방법.
- [0168] 12. 구현예 1 내지 11 중 어느 하나에 있어서, 상기 도입 단계는 가이드 RNA용 발현 카세트를 포함하는 DNA 작제물을 사상균 세포들 내로 도입하는 단계를 포함하는, 방법.
- [0169] 13. 구현예 1 내지 12 또는 구현예 48 내지 54 중 어느 하나에 있어서, 상기 도입 단계는 선택 가능 마커를 인코딩하는 서열을 포함하는 DNA 작제물을 사상균 세포들 내로 도입하는 단계를 포함하는 방법.
- [0170] 14. 구현예 13에 있어서, DNA 작제물은 선택 가능 마커를 인코딩하는 서열 및 공여체 DNA를 모두 포함하는, 방법.
- [0171] 15. 구현예 1 내지 14 중 어느 하나에 있어서, 상기 도입 단계는 Cas 엔도뉴클레아제를 인코딩하는 서열, 가이드 RNA를 인코딩하는 서열, 선택가능 마커를 인코딩하는 서열, 및 공여체 DNA를 포함하는 DNA 작제물을 사상균 세포들 내로 도입하는 단계를 포함하는, 방법.
- [0172] 16. 구현예 11 내지 15 중 어느 하나에 있어서, DNA 작제물은 선형 DNA 작제물인, 방법.
- [0173] 17. 구현예 11 내지 15 중 어느 하나에 있어서, DNA 작제물은 원형 DNA 작제물인, 방법.
- [0174] 18. 구현예 11 및 구현예 15 내지 17 중 어느 하나에 있어서, Cas 엔도뉴클레아제용 발현 카세트 또는 Cas 엔도뉴클레아제를 인코딩하는 서열은 사상균 세포에서 발현을 위해 최적화된 Cas 코딩 서열을 포함하는, 방법.
- [0175] 19. 구현예 18에 있어서, Cas 코딩 서열은 서열 번호 44와 적어도 70% 동일한 폴리뉴클레오티드 서열을 포함하는 Cas9 코딩 서열인, 방법.
- [0176] 20. 구현예 1 내지 10, 구현예 12 내지 14, 및 구현예 16 및 17 중 어느 하나에 있어서, 상기 도입 단계는 Cas 엔도뉴클레아제를 사상균 세포들 내로 직접 도입하는 단계를 포함하는, 방법.
- [0177] 21. 구현예 1 내지 11, 구현예 13, 14, 및 구현예 16 내지 20 중 어느 하나에 있어서, 상기 도입 단계는 가이드 RNA를 사상균 세포들 내로 직접 도입하는 단계를 포함하는, 방법.

- [0178] 22. 구현에 1 내지 21 중 어느 하나에 있어서, Cas 엔도뉴클레아제는 핵 위치 신호에 작동 가능하게 연결되는, 방법.
- [0179] 23. 구현에 1 내지 22 중 어느 하나에 있어서, 사상균 세포는 유미코티나 또는 페지조미코티나 사상균 세포인, 방법.
- [0180] 24. 구현에 1 내지 23 중 어느 하나에 있어서, 사상균 세포는 트리코테르마, 페니실리움, 아스퍼질러스, 휴미콜라, 크리소스포리움, 푸사리움, 마이셀리오프토라, 뉴로스포라, 하이포크레아, 및 에메리셀라로 구성되는 군으로부터 선택되는, 방법.
- [0181] 25. 구현에 1 내지 24 중 어느 하나에 있어서, 표적 부위는 오픈 리딩 프레임, 프로모터, 조절 서열, 종결인자 서열, 조절 요소 서열, 스플라이스 부위, 코딩 서열, 폴리유비퀴티네이션 부위, 인트론 부위, 및 인트론 증진 모터프로 구성되는 군으로부터 선택되는 관심 유전자의 영역 내에 위치되는, 방법.
- [0182] 26. 구현에 1 내지 25 중 어느 하나에 있어서, 상동성 재조합은 표적 부위에서 또는 표적 부위 근처에서 DNA 서열의 변형을 초래하고, 상기 변형은 하나 이상의 뉴클레오티드의 결실, 하나 이상의 뉴클레오티드의 삽입, 관심 단백질을 인코딩하는 발현 카세트의 삽입, 하나 이상의 뉴클레오티드의 치환, 및 이들의 임의 조합으로 구성되는 군으로부터 선택되는, 방법.
- [0183] 27. 구현에 1 내지 26 중 어느 하나에 있어서, 상기 확인 단계는 상동성 재조합 또는 변형을 선택하거나 선별하는 조건 하에서 단계 (a)로부터 세포군을 배양하는 단계를 포함하는, 방법.
- [0184] 28. 구현에 1 내지 27 중 어느 하나에 있어서, 상기 확인 단계는 불안정한 형질전환체를 선별하는 조건 하에서 단계 (a)로부터 세포군을 배양하는 단계를 포함하는, 방법.
- [0185] 29. 구현에 1 내지 28 중 어느 하나에 있어서, 상기 도입 단계는 선택가능 마커를 인코딩하는 서열 및 공여체 DNA를 포함하는 DNA 작제물을 사상균 세포들 내로 도입하는 단계를 포함하고, 상기 확인 단계는 선택가능 마커를 잃었지만 아직은 공여체 DNA를 보유하고 있는 불안정한 형질전환체를 선별하는 조건 하에서 단계 (a)로부터 세포군을 배양하는 단계를 포함하는, 방법.
- [0186] 30. 구현에 1 내지 29 중 어느 하나의 방법으로 생산된 재조합 사상균 세포.
- [0187] 31. Cas 엔도뉴클레아제용 발현 카세트를 포함하는 제1 재조합 DNA 작제물을 포함하는 재조합 사상균 세포.
- [0188] 32. 구현에 30 또는 31에 있어서, 재조합 사상균 세포는 NHEJ 경로에 하나 이상의 비기능성 성분 또는 활성-감소 성분을 포함하는, 재조합 사상균 세포.
- [0189] 33. 구현에 32에 있어서, NHEJ 경로의 하나 이상의 성분은 ku80, ku70, rad50, mre11, xrs2, lig4, 및 xrs로 구성되는 군으로부터 선택되는, 재조합 사상균 세포.
- [0190] 34. 구현에 33에 있어서, 사상균 세포는 ku80의 발현 및/또는 활성을 저해하는 유전자 변형을 포함하는, 재조합 사상균 세포.
- [0191] 35. 구현에 31에 있어서, Cas 엔도뉴클레아제는 Cas 니카아제인, 재조합 사상균 세포.
- [0192] 36. 구현에 30 내지 35 중 어느 하나에 있어서, 가이드 RNA용 발현 카세트를 포함하는 제2 재조합 DNA 작제물을 추가로 포함하며, 여기서 가이드 RNA 및 Cas 엔도뉴클레아제는 Cas 엔도뉴클레아제가 재조합 사상균 세포의 게놈 내 표적 부위에서 작용할 수 있게 하는 복합체를 형성할 수 있는, 재조합 사상균 세포.
- [0193] 37. 구현에 30 내지 36 중 어느 하나에 있어서, 관심 폴리뉴클레오티드를 포함하는 공여체 DNA를 추가로 포함하는, 재조합 사상균 세포.
- [0194] 38. 구현에 30 내지 37 중 어느 하나에 있어서, Cas 엔도뉴클레아제는 Cas9 엔도뉴클레아제 또는 이의 변이체인, 재조합 사상균 세포.
- [0195] 39. 구현에 38에 있어서, Cas9 엔도뉴클레아제 또는 이의 변이체는 서열 번호 45 및 서열 번호 48 내지 53 중 어느 하나와 적어도 70% 동일성을 갖는 아미노산 서열을 포함하는, 재조합 사상균 세포.
- [0196] 40. 구현에 31 내지 39 중 어느 하나에 있어서, Cas 엔도뉴클레아제용 발현 카세트는 서열 번호 44와 적어도 70% 동일한 폴리뉴클레오티드 서열을 포함하는, 재조합 사상균 세포.
- [0197] 41. 구현에 31 내지 40 중 어느 하나에 있어서, Cas 엔도뉴클레아제용 발현 카세트는 사상균 세포에서의 발현을

위해 최적화된 Cas 엔도뉴클레아제 유전자를 포함하는, 재조합 사상균 세포.

- [0198] 42. 구현예 31 내지 41 중 어느 하나에 있어서, Cas 엔도뉴클레아제는 핵 위치 신호에 작동 가능하게 연결되는, 재조합 사상균 세포.
- [0199] 43. 구현예 42에 있어서, 핵 위치 신호는 SV40 핵 위치 신호(서열 번호 46), T. 레에세이 b1r2(청색광 조절자 2) 유전자로부터 유래된 핵 표적화 신호(서열 번호 47), 및 이 둘의 조합으로 구성되는 군으로부터 선택되는, 재조합 사상균 세포.
- [0200] 44. 구현예 30 내지 43 중 어느 하나에 있어서, 사상균 세포는 트리코데르마, 페니실리움, 아스퍼질러스, 휴미콜라, 크리소스포리움, 푸사리움, 마이셀리오프토라, 뉴로스포라, 하이포크레아, 및 에메리셀라로 구성되는 군으로부터 선택되는 사상균 세포인, 재조합 사상균 세포.
- [0201] 45. Cas9 엔도뉴클레아제 또는 이의 변이체를 인코딩하는 사상균 세포에 최적화된 폴리뉴클레오티드 서열에 작동 가능하게 연결되는 프로모터를 포함하는, 재조합 DNA 작제물.
- [0202] 46. 구현예 45에 있어서, 사상균 세포에 최적화된 폴리뉴클레오티드 서열은 서열 번호 44와 적어도 70% 동일한, 재조합 DNA 작제물.
- [0203] 47. 가이드 RNA용 발현 카세트는 유아스코마이세테(Euascomycete) 또는 페지조마이세테(Pezizomycete)에서 기능적인 DNA 증합효소 III 의존성 프로모터를 포함하고, 여기서 상기 프로모터는 가이드 RNA를 인코딩하는 DNA에 작동 가능하게 연결되는, 구현예 12의 방법.
- [0204] 48. 구현예 47에 있어서, 프로모터는 트리코데르마 U6 snRNA 유전자로부터 유래되는, 방법.
- [0205] 49. 구현예 48에 있어서, 프로모터는 서열 번호 40 또는 41에 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 뉴클레오티드 서열을 포함하는, 방법.
- [0206] 50. 구현예 49에 있어서, 프로모터는 서열 번호 40 또는 41을 포함하는, 방법.
- [0207] 51. 구현예 12 및 구현예 47 내지 50 중 어느 하나에 있어서, 가이드 RNA용 발현 카세트는 트리코데르마 U9 snRNA 유전자로부터의 인트론 서열을 갖는 가이드 RNA-인코딩 DNA를 포함하는, 방법.
- [0208] 52. 구현예 51에 있어서, 트리코데르마 U9 snRNA 유전자로부터 유래된 인트론 서열은 서열 번호 42와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 뉴클레오티드 서열을 포함하는, 방법.
- [0209] 53. 구현예 52에 있어서, 트리코데르마 U9 snRNA 유전자로부터 유래된 인트론 서열은 서열 번호 42의 서열을 포함하는, 방법.

[0210] **실시예**

[0211] 다음 실시예들에서, 달리 언급되지 않으면, 부(parts) 및 백분율은 중량에 의하고, 도(degree)는 섭씨이다. 이 실시예들은 본 개시의 구현예들을 나타내면서 단지 예시로 제공되는 것임을 이해해야 한다. 상기 논의 및 이들 실시예로부터, 당업자는 다양한 용도 및 조건에 적응시키기 위해 본 개시의 다양한 변형 및 수정을 할 수 있다. 이러한 변형은 또한 첨부된 청구항의 범위 내에 속하는 것으로 의도된다.

[0212] **섹션 A: 발현 벡터에 의한 Cas/가이드 RNA의 도입**

[0213] 실시예 1: T. 레에세이 U6 snRNA 유전자의 확인

[0214] RNA 증합 효소 II에 의존적인 프로모터를 사용함으로써 초래될 5' 캡 구조 또는 폴리아데닐화를 부가하지 않고서 T. 레에세이에서 가이드 RNA를 생산하기 위해서는 RNA 증합 효소 III 유도 프로모터가 바람직하다. 그러나, T. 레에세이에서 기능적인 RNA 증합 효소 III 의존성 프로모터는 기술되지 않았다. 사카로마이세스 세레비지에(Saccharomyces cerevisiae) snr52 유전자, 인간 U6 snRNA 유전자 또는 옥수수 U6 snRNA 유전자의 5' 상류 영역을 포함하여 다른 종으로부터 공지된 RNA 증합 효소 III 의존성 프로모터가 T. 레에세이에서 기능하는 능력을 시험받은 것으로 간주하였다.

[0215] 보다 바람직한 것은 RNA 증합 효소 III 의존성 프로모터로서 작용할 천연 T. 레에세이 서열을 확인하는 것이었다. BLAST 알고리즘을 사용하여 T. 레에세이 v2 게놈 서열(www.jgi.doe.gov)을 검색하기 위해 인간 U6 소형 핵 RNA(snRNA, 유전자은행 수탁번호 M14486)를 인코딩하는 DNA 서열을 사용하였다. T. 레에세이 DNA 서열의 짧은

영역이 인간 서열과 유사한 것으로 확인되었다. 주변 DNA 서열의 조사 및 효모, 특히 슈조사카로마이세스 프롬페(Schizosaccharomyces pombe)의 U6 유전자와의 비교(Marck 등, 2006, Nucleic Acids Research 34 : 1816-1835)로 T. 레에세이 U6 유전자의 많은 특징을 추정하여 확인할 수 있었다(서열 번호 1, 하기에 나타냄). 전사된 서열의 시작과 종결인자는 상류 TATA 박스와 동일하게 확인되었다. 인트론은 전사된 영역을 분명히 방해하고, 가능한 A-박스 및 B-박스 프로모터 요소는 전사된 영역 내에서 인지될 수 있으며, 후자는 인트론 내에서 인식될 수 있다. (도 1 참조).

[0216] AAAAAACACTAGTAAGTACTTACTTATGTATTATTAACACTTTAGCTAACTTCTGCAGTACTACCTAAGAGGCTAGGGGTAGTTTTATAGCAGACTTATAGCTATTATTTTTATTAGTAAAGTGCTTTTAAAGTAAGGCTTTTTTATAGCACTTTTTATTATTATAATATATATATATAATAAATTTTAAAGCCCTGGAATAGTAAAGAGGCTTATATAATAATTTATAGTAATAAAAGCTTAGCAGCTGTAATATAATTCCTAAAGAAACAGCATGAAATGGTATTATGTAAGAGCTATAGCTAAAGGCACTCTGCTGGATAAAAATAGTGGCTATAAGTCTGCTGCAAAACTACCCCCAACCTCGTAGGTATATAAGTACTGTTTGATGGTAGTCTATCGCCTCGGGCATTGGTCAATTTATAACGATACAGGTTTCGTTTCGGCTTTTCTCGGAACCCAGAGGTCATCAGTTCGAATCGCTAACAGGTCAACAGAGAAGATTAGCATGGCCCTGCACTAAGGATGACACGCTCAAGAGAAGCTAAACATTTTTTCTCTTCCAAGTCGTGATGGTATCTTTTTGCTTAGAGAATCTATTCTGTGGACGATTAGTATTGGTAAATCCCTGCTGCACATTGCGCGGATGGTCTCAACGGCATAATACCCCATTCGTGATGCAGCGGTGATCTTCAATATGTAGTGAATACGTTGCATACACCACGAGTTCGGTGCCTCCTGTATGTACAGTACTGTAGTTCGACTCCTCCGCGCAGGTGGAACGATTCCCTAGTGGCAGGTATTTGGCGGGTCAAGAA (서열 번호 1)

[0217] 실시예 2: T. 레에세이 유전자를 표적으로 하기 위한 sgRNA 서열

[0218] 단일 가이드 RNA(sgRNA) 분자가 스트렙토코쿠스 파이오제네스 Cas9 단백질과 상호 작용하여 이 엔도뉴클레아제를 생체 내에서 진핵 생물 계놈의 특정 유전자좌에 표적으로 할 수 있는 것으로 나타났다. sgRNA는 스트렙토코쿠스 파이오제네스 II 형 CRISPR-Cas 시스템의 구성 요소인 것으로 자연적으로 관찰되는 tracrRNA와 crRNA 사이의 융합체로 설계된 하이브리드 분자이다(Gasiunas 등 (2012) Proc. Natl. Acad. Sci.USA 109:E2579-86, Jinek 등 (2012) Science 337:816-21, Mali 등 (2013) Science 339:823-26, and Cong 등 (2013) Science 339:819-23). sgRNA의 처음 20개 뉴클레오티드는 계놈의 표적 부위에 상보적이다. 추가 서열인 포토스페이서 인접 모티프(PAM)도 sgRNA 상보 영역에 인접한 계놈의 표적 부위에 존재해야 한다. S. 파이오제네스 Cas9의 경우, PAM은 서열 NGG(N은 A, G, C 또는 T)를 갖는다.

[0219] 이 실험에 사용된 sgRNA의 서열은 하기에 도시되고, 여기서 표적 부위에 상보적인 것으로 설계된 20개의 뉴클레오티드가 N 잔기(서열 번호 2)(N = A, G, C 또는 U)로 도시된다.

[0220] NNNNNNNNNNNNNNNNNNGUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUC

[0221] sgRNA는 T. 레에세이 계놈 내 상이한 유전자좌를 표적으로 하도록 설계되었다. 표적 부위 1(TS 1)로 지정된 부위에서 T. 레에세이 *ad3A* 유전자(포스포리보실아미도이미다졸-숙시노카복사미드 신타아제)를 표적으로 하기 위한 (gAd3A TS1로 불리는) sgRNA의 서열이 아래에 도시된다(서열 번호 3). T. 레에세이 계놈 서열에 상보적인 20개 뉴클레오티드 영역이 소문자로 도시된다.

[0222] guccugagcaaaagugcGUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUC

[0223] 표적 부위 2(TS 2)로 지정된 부위에서 T. 레에세이 *glal*(글루코아밀라아제) 유전자를 표적으로 하기 위한 (gTrGA TS2로 불리는) sgRNA의 서열이 아래에 도시된다(서열 번호 4). T. 레에세이 계놈 서열에 상보적인 20개 뉴클레오티드 영역이 소문자로 도시된다.

[0224] guucagugcaauagcGUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUC

[0225] 표적 부위 11(TS 11)로 지정된 부위에서 T. 레에세이 *glal*(글루코아밀라아제) 유전자를 표적으로 하기 위한 (gTrGA TS11로 불리는) sgRNA의 서열이 아래에 도시된다(서열 번호 5). T. 레에세이 계놈 서열에 상보적인 20개 뉴클레오티드 영역이 소문자로 도시된다.

[0226] gccaauggcgacggcGUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUC

[0227] 표적 부위 6(TS 6)으로 지정된 부위에서 T. 레에세이 *pyr2*(오로테이트 포스포리보실트란스페라아제(ornithine decarboxylase)) 유전자를 표적으로 하기 위한 (gPyr2 TS6으로 불리는) sgRNA의 서열이 아래에 도시된다(서열 번호 6). T. 레에세이 계놈 서열에 상보적인 20개 뉴클레오티드 영역이 소문자로 도시된다.

[0228] gcacagcgggaugccGUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUC

[0229] 실시예 3: T. 레에세이에서 발현을 위한 Cas9 DNA 및 단백질 서열

[0230] *T. 레에세이*에서의 발현을 위하여 NLS 서열을 포함하여 코돈이 최적화된 스트렙토코쿠스 파이오제네스 Cas9-인 코딩 유전자를 설계하고, 합성하고 검사하였다(서열 번호 7). 인코딩된 단백질(서열 번호 8)은 N 말단 SV40 핵 위치 신호(NLS; 서열 번호 46) 및 *T. 레에세이 blr2*(청색 조절인자 2) 유전자로부터 유래된 C 말단 NLS(서열 번호 47; 둘 모두 아래의 서열 번호 8에 밑줄쳐져 있다)를 갖는다.

[0231] 서열 번호 7:

[0232] acgcacgcacctgagccagctggcgaggagacaagaagaagaagctcaagctctag

[0233] 서열 번호 8:

MAPKKKRKVMDKKYSIGLDIGTNSVGVAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGA  
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEE  
 DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLI  
 EGDLPDNDSDVDKLFQILVQTYNQLFEENPINASGVDAKAILSARLSKSRRENLIAQLP  
 GEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDYYDDLDNLLAQIGDQYADL  
 FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI  
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG  
 SIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRK  
 SEETITPWNFEEVVDKGGASQSFIERMTNFDKNLPNEKVLPHKSHLLYEYFTVYNELTKV  
 KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR  
 FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKV  
 MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK  
 EDIQAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMA  
 RENQTTQKGQKNSRERMKRIIEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD  
 MYVDQELDINRLSDYVDHIVPQSFLKDDSIDNKVLRSDKNRSGKSDNVPSEEVVKKM  
 KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR  
 MNTKYDENDKLIREVKVITLKSFLVSDFRKDFQFYKVINNYHHAHDAYLNAVVGTL  
 IKKYPKLESEFVYGDYKVDVVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI  
 RKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRN  
 SDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGGKSKKLKSVKELLGITIMERSSE  
 EKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYV  
 NFLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL  
 SA YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ  
 SITG  
 LYETRIDLSQLGGDKKKKLLK

[0234]

[0235] 실시예 4: Cas9 발현 벡터의 구성

[0236] 플랭킹 *attL1* 및 *attL2* 부위 사이에 있어 Invitrogen™ Gateway® 클로닝 기술(Thermo Fisher Scientific Inc., Grand Island, NY)에 의해 적합한 발현 벡터로의 전사가 가능하도록 상기에 도시된 Cas9를 인코딩하는 합성 DNA 서열을 pENTR/D-TOPO에 삽입하였다. 다음과 같은 특징을 포함하는 게이트웨이 호환 가능 발현 벡터, pTrex2gHyg를 사용할 수 있었다: 게이트웨이 클로닝 부위에 의해 분리된 *T. 레에세이 pki1*(피루베이트 키나아제) 유전자로부터의 프로모터 영역 및 *T. 레에세이 cbh1*(셀로비오하이드롤라아제 1) 유전자로부터의 종결인자 영역, *뉴로스 포라 크라사(Neurospora crassa) cpc1*(교차 경로 조절 1) 프로모터 영역 및 *아스페질러스 니두란스(Aspergillus nidulans) trpC*(글루타민 아미도 트랜스퍼라아제, 인돌글리세롤포스페이트 신타아제 및 포스포리 보실란트라닐레이트 이소머라아제 활성을 갖는 3 기능성 단백질) 종결인자 영역에 기능적으로 연결된 박테리아 하이그로마이신 포스포트랜스퍼라아제 유전자, 및 대장균에서의 선택 및 유지를 위한 박테리아 벡터 서열. pTrex2gHygMoCas를 제공하기 위해 게이트웨이 클로닝 절차를 사용하여 cas9 유전자를 pTrex2gHyg에 복제하였다

(도 2 참조).

[0237] 실시예 5: sgRNA 발현 벡터의 구성

[0238] 상이한 추정 RNA 중합 효소 III 의존성 프로모터 및 종결인자에 의해 플랭킹된 gAd3A TS1 sgRNA를 인코딩하는 합성 DNA 서열을 획득하였다. 이들 합성 DNA 서열 각각은 또한 어느 한쪽 말단에 제한 효소 인식 부위(EcoRI 및 BamHI)를 갖는다.

[0239] 다음 서열은 사카로마이세스 세레비지에 snr52 프로모터 및 S. 세레비지에 sup4 종결인자(gAd3A TS1-1로 표시, 서열 번호 9)로 gAd3A TS1 sgRNA(밑줄친)를 인코딩한다:

gaattcggatccTCTTTGAAAAGATAATGTATGATTATGCTTTCACTCATATTTATACAGA  
 AACTTGATGTTTTCTTTGAGTATATACAAGGTGATTACATGTACGTTTGAAGTACA  
 ACTCTAGATTTTGTAGTGCCCTCTTGGGCTAGCGGTAAAGGTGCGCATTTTTTTCAC  
 ACCCTACAATGTTCTGTTCAAAAGATTTTGGTCAAACGCTGTAGAAGTGAAAGTTG  
 GTGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATCggtcctcg  
agcaaaaggtgccGTTTTAGAGCTAGAAATAGCAAGTAAAATAAGGCTAGTCCGTTATC  
AACTTGAAAAAGTGGCACCCGAGTCGGTGGTGCTTTTTTTGTTTTTATGTCTgaattcg  
 gatcc

[0240]

[0241] 다음 서열은 *T. 레에세이* U6 프로모터 및 종결인자(gAd3A TS1-2로 표시, 서열 번호 10)로 gAd3A TS1 sgRNA(밑줄친)를 인코딩한다:

gaattcggatccAAAAACACTAGTAAGTACTTACTTATGTATTATTAACTACTTTAGCTA  
 ACTTCTGCAGTACTACCTAAGAGGCTAGGGGTAGTTTTATAGCAGACTTATAGCTA  
 TTATTTTTATTTAGTAAAGTGCTTTTAAAGTAAGGTCTTTTTATAGCACTTTTTATTT  
 ATTATAATATATATTATATAATAATTTAAGCCTGGAATAGTAAAGAGGCTTATATAA  
 TAATTTATAGTAATAAAAGCTTAGCAGCTGTAATATAATTCCTAAAGAAACAGCATG  
 AAATGGTATTATGTAAGAGCTATAGTCTAAAGGCACTCTGCTGGATAAAAATAGTG  
 GCTATAAGTCTGCTGCAAAACTACCCCAACCTCGTAGGTATATAAGTACTGTTTG  
 ATGGTAGTCTATCggtcctcgagcaaaaggtgccGTTTTAGAGCTAGAAATAGCAAGTAAA  
TAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCCGAGTCGGTGGTGCTTTTTTT  
 TTCTCTTgaattcggatcc

[0242]

[0243] 다음 서열은 *T. 레에세이* U6 프로모터, 종결인자 및 인트론(이탈릭체)(gAd3A TS1-3으로 표시, 서열 번호 11)로 gAd3A TS1 sgRNA(밑줄친)를 인코딩한다:

gaattcggatccAAAAACACTAGTAAGTACTTACTTATGTATTATTAACTACTTTAGCTA  
 ACTTCTGCAGTACTACCTAAGAGGCTAGGGGTAGTTTTATAGCAGACTTATAGCTA  
 TTATTTTTATTTAGTAAAGTGCTTTTAAAGTAAGGTCTTTTTATAGCACTTTTTATTT  
 ATTATAATATATATTATATAATAATTTAAGCCTGGAATAGTAAAGAGGCTTATATAA  
 TAATTTATAGTAATAAAAGCTTAGCAGCTGTAATATAATTCCTAAAGAAACAGCATG  
 AAATGGTATTATGTAAGAGCTATAGTCTAAAGGCACTCTGCTGGATAAAAATAGTG  
 GCTATAAGTCTGCTGCAAAACTACCCCAACCTCGTAGGTATATAAGTACTGTTTG  
 ATGGTAGTCTATCggtcctcgagcaaaaggtgccGTTTTAGAGCTAGAGTTCGTTTCGGCTTT  
TCCTCGGAACCCCAAGAGGTCATCAGTTCGAATCGCTAACAGAATAGCAAGTAAA  
ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCCGAGTCGGTGGTGCTTTTTT  
 TTTCTCTTgaattcggatcc

[0244]

[0245] 플라스미드 p219M(도 3)은 이의 천연 프로모터 및 종결인자를 포함하는 *T. 레에세이* *pyr4*(오로티딘 모노포스페이트 디카복실라아제(otridine monophosphate decarboxylase)) 유전자를 함유하는 대장균 벡터이다. 이 벡터를 EcoRI 및 BamHI로 절단하고, 말단을 탈인산화시켰다. 상기 합성 DNA 분자 각각을 EcoRI 및 BamHI로 절단하고 절단된 p219M과 연결하여 sgRNA 발현 카세트 및 *pyr4* 유전자를 포함하는 일련의 벡터를 생성하였다. 각 벡터는 암호화된 sgRNA의 이름으로 지정되었다(예컨대, p219M gAd3A TS1-1은 gAd3A 발현 카세트를 *S. 세레비시에* *snr52* 프로모터 및 *sup4* 종결인자와 통합함).

[0246] 보다 짧은 *T. 레에세이* U6 프로모터 영역을 갖는 가이드 RNA 발현 카세트를 합성 DNA로서 획득 하였다. TS11에서의 *T. 레에세이* *glal* 유전자(서열 번호 12, 인트론 서열은 밑줄쳐진다)를 표적으로 하는 sgRNA에 대한 서열을 포함하는 예가 여기에 제공된다.

AATTCCTAAAGAAACAGCATGAAATGGTATTATGTAAGAGCTATAGTCTAAAGGCA  
 CTCTGCTGGATAAAAATAGTGGCTATAAGTCTGCTGCAAACTACCCCCAACCTCG  
 TAGGTATATAAGTACTGTTTGGATGGTAGTCTATCgccaatggcgacggcagcacGTTTTAGA  
GCTAGAGTTCGTTTCGGCTTTTCCTCGGAACCCCCAGAGGTCATCAGTTCGAATC  
GCTAACAGAATAGCAAGTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG  
 CACCGAGTCGGTGGTGCCTTTTTTTTCTCTT

[0247]

[0248] 상기한 gRNA 발현 카세트를 프라이머 gRNA fwd afIII (5'-  
 cgtcagcctaagAATTCCTAAAGAAACAGCATGAAATGG ; 서열 번호 13) 및 gRNA rev sfiI(5'-  
 cgtcaggccacgtgggccAAGAGAAAAAAGCACCACCGACTCGG; 서열 번호 14)를 사용하여 PCR에 의해 증폭시켰다. 이들 프라이머는 *afIII*를 가이드 RNA 발현 카세트의 5' 말단에, *sfiI* 부위를 3' 말단에 부가한다. PCR 생성물을 Qiagen PCR 정제 키트를 사용하여 제조사의 지침에 따라 정제하였다. 그런 다음, PCR 생성물을 *SfiI* 및 *AfIII*로 절단하고 Qiagen PCR 정제 키트로 다시 세척하였다. 플라스미드 pTrex2g / HygMoCas를 *SfiI* 및 *AfIII*로 절단하고 Roche Rapid 알칼라인 포스파타아제 키트(Roche Diagnostics Corp., IN)를 사용하여 탈인산화시켰다. 절단된 플라스미드 및 PCR 생성물을 최종적으로 Rapid DNA 리가아제 키트를 사용하여 결합시켜 pTrex2g/Hyg MoCas gTrGa TS11B를 생성하였다. 다른 sgRNA 발현 카세트를 pTrex2g / Hyg MoCas에 유사한 방식으로 삽입하였다.

[0249] 실시예 6: 트리코테르마 레에세이에서 Cas9-매개된 유전자 불활성화

[0250] *트리코테르마 레에세이* 균주가 두 개의 개별 발현 벡터, 즉 Cas9 생산용과 gRNA 생산용으로 동시에 형질전환되거나 Cas9 및 gRNA 둘 다 발현용 단일 벡터로 형질전환된 일련의 실험이 여기에 기재되어 있다. 이 실험은 U6 인트론이 gRNA 전사 영역 내에도 존재할 때만 *T. 레에세이* U6 유전자의 5' 상류 영역이 gRNA 전사를 촉진한다는 것을 나타낸다. 실험은 또한 표적 유전자 불활성화가 *T. 레에세이* 형질전환체에서 높은 효율로 발생할 수 있음을 나타낸다.

[0251] **ad3A 유전자의 불활성화**

[0252] 4 개의 주요 분비된 셀룰라아제를 인코딩하는 유전자(*cbh1*, *cbh2*, *eg11* 및 *eg12*)가 결실된 공개적으로 이용가능한 RL-P37 균주에서 유래된 *트리코테르마 레에세이* 균주를 사용하였다. 이 균주에는 또한 기능적 *pyr4* 유전자가 결여되어 있다. (US20060003408A1에 기재된 바와 같은) 생물학적 형질전환을 사용하여 동일한 양의 pTrex2gHyg MoCas(도 2) 및 p219M gAd3A TS1-1, p219M gAd3A TS1-2 또는 p219M gAd3A TS1-3 중 어느 하나의 혼합물과 공변형시켰다. 2% 포도당, 100 mg/L 하이그로마이신 B 및 200 mg/L 아데닌을 함유한 보겔(Vogel)의 최소 배지를 갖는 한천 플레이트에서 형질전환체를 선택하였다. 제1 플레이트에서 선택한 후, 형질전환체 콜로니를 동일한 선택 배지의 신선한 플레이트에 집어 넣었다. 제 2 플레이트에서 성장하는 동안, 안정한 및 불안정한 하이그로마이신-내성 형질전환체를 구별하는 것이 가능하였다. 안정한 형질전환체가 더 빨리 자라났고, 콜로니는 윤곽이 부드러웠고 균사체는 더 조밀했다. 불안정한 형질전환체는 느리게 성장하고 밀도가 낮은 균사체가 있었으며 콜로니는 가장자리가 불규칙한 외형을 가졌다. 제2 플레이트에서 성장한 후 형질전환체를 포도당이 있고 하이그로마이신이 없는 보겔(Vogel) 배지에 옮기고 14 mg/L 아데닌을 사용하여 적색/갈색을 나타내는 것들을 선별 검사하여 이들이 아데닌 영양요구성 인자임을 나타냈다. p219M gAd3A TS1-1을 사용하여 5개의 안정한 형질전환체와 23개의 불안정한 형질전환체를 획득했는데, 모두 아데닌(adenine) 원형질체였다. p219M gAd3A TS1-2를 사용하여 11개의 안정한 형질전환체와 38개의 불안정한 형질전환체를 획득했는데, 11개의 안정한 형질전환체 전부 및 불안정한 형질전환체 중 29개는 아데닌 원형질체였다. p219M gAd3A TS1-3으로 19개의 안정한 형질전환체와 2개

의 불안정한 형질전환체를 수득했는데, 모두 아데닌 영양요구인자였다. 분명히, 아데닌 영양요구인자는 sgAd3A TS1의 전사를 조절하는 *T. 레에세이* U6 프로모터, 인트론 및 종결인자를 이용하는 gAd3A TS1-3으로만 수득되었다. 아데닌 영양요구는 천연 *T. 레에세이* *ad3A* 유전자좌에서 표적화된 Cas9 절단을 나타낸다. 시험된 gAd3A TS1-3을 갖는 모든 형질전환체는 아데닌 영양요구인자이므로 Cas9 매개 유전자 불활성화가 효율적이라는 결론을 내릴 수 있다.

- [0253] pTrex2gHyg MoCas 및 p219M gAd3A TS1-3을 갖는 공동 형질전환체의 *ad3A* 유전자좌에서 돌연변이를 판단하기 위해, 10개의 안정한 아데닌 영양요구성 형질전환체로부터 게놈 DNA를 추출 하였다. 이 DNA를 Cas9 표적 부위에 걸쳐 있거나 표적 부위의 상류 또는 하류에 위치하는 생성물을 생성하도록 설계된 몇몇 상이한 프라이머 쌍을 사용하는 PCR을 위한 주형으로 사용하였다. 제조사의 지시에 따라 PfuUltra II Fusion HS DNA 중합 효소 (Agilent Technologies, Santa Clara, CA)를 PCR에 사용하였다. 각각의 경우에 연장 시간은 아래에 설명된 것처럼 PCR 제품의 예상 크기에 대해 제조업체가 제안한 시간이었다. PCR 생성물의 크기는 아가로스 겔 전기 영동으로 평가하였다.
- [0254] TS1 표적 부위의 5' 측의 영역을 증폭시키는 Ad3 5' fwd + Ad3 5' rev 프라이머(각각 5'-tgaacacagccaccgacatcagc [서열 번호 15] 및 5'-gctggtgagggtttgtgctattg [서열 번호 16])를 사용하여 모든 형질전환체에서 예상 크기(872 bp)의 PCR 생성물을 수득하였다.
- [0255] TS1 표적 부위의 5' 측의 영역을 증폭시키는 Ad3 5' fwd + Ad3a 5005 rev 프라이머(각각 5'-tgaacacagccaccgacatcagc [서열 번호 15] 및 5'-gattgcttgggaggaggacat [서열 번호 17])를 사용하여 모든 형질전환체에서 예상 크기(1214 bp)의 PCR 생성물을 수득하였다.
- [0256] TS1 표적 부위의 3' 측의 영역을 증폭시키는 Ad3 3' fwd + Ad3 3' rev 프라이머(각각 5'-cgaggccactgatgaagtgttc [서열 번호 18] 및 5'-cagttttccaaggctgccaacgc [서열 번호 19])를 사용하여 모든 형질전환체에서 예상 크기(904 bp)의 PCR 생성물을 수득하였다.
- [0257] TS1 표적 부위의 3' 측의 영역을 증폭시키는 Ad3a 5003 fwd + Ad3mid rev 프라이머(각각 5'-ctgatcttgaccctggaatc [서열 번호 20] 및 5'-ctctctatcatttgccaccctcc [서열 번호 21])를 사용하여 모든 형질전환체에서 예상 크기(757 bp)의 PCR 생성물을 수득하였다.
- [0258] 상기 PCR 결과는 게놈 DNA 제제가 Cas9 표적 부위의 상류 또는 하류에서 PCR 생성물을 수득하기에 충분한 품질임을 나타내었다.
- [0259] ad3A에서 TS1 표적 부위에 걸친 Adfrag fwd + Adfrag rev 프라이머(각각 5'-ctccattcaccctcaattctcc [서열 번호 22] 및 5'-gttccttggcggtgcttgatc [서열 번호 23])를 사용하여 임의의 형질전환체에 대한 PCR 생성물을 수득할 수 없었다. Cas9 활성으로 인한 큰 크기 변화가 없다고 가정할 때 이 PCR 생성물의 예상 크기는 약 764 bp였다.
- [0260] ad3A에서 TS1 표적 부위에 걸친 Adfrag fwd + Ad3 3 'rev 프라이머(각각 5'-ctccattcaccctcaattctcc [서열 번호 22] 및 5'-cagttttccaaggctgccaacgc [서열 번호 19])를 사용하여 임의의 형질전환체에 대해 PCR 생성물을 수득할 수 없었다. Cas9 활성으로 인한 큰 크기 변화가 없다고 가정할 때 이 PCR 생성물의 예상 크기는 약 2504 bp였다.
- [0261] ad3A에서 TS1 표적 부위에 걸친 Ad3a 2k fwd + Ad3a 2k rev 프라이머(각각 5'-ctccattcaccctcaattctcc [서열 번호 24] 및 5'-cagttttccaaggctgccaacgc [서열 번호 25])를 사용하여 임의의 형질전환체에 대해 PCR 생성물을 수득할 수 없었다. Cas9 활성으로 인한 큰 크기 변화가 없다고 가정할 때 이 PCR 생성물의 예상 크기는 약 1813 bp였다.
- [0262] 형질전환체 중 5개는 또한 TS1 표적 부위에 걸친 Adfrag fwd + Ad3 mid rev 프라이머(각각 5'-ctccattcaccctcaattctcc [서열 번호 22] 및 5'-ctctctatcatttgccaccctcc [서열 번호 21])를 사용하여 PCR 생성물을 제공하지 못했다. Cas9 활성으로 인한 큰 크기 변화가 없다고 가정할 때 이 PCR 생성물의 예상 크기는 약 1438 bp였다.
- [0263] 공개된 데이터를 기초로, Cas9가 매개하는 유전자의 불활성화는 일반적으로 표적 부위의 DNA에서 이중 가닥 절단의 오류가 발생하기 쉬운 복구를 포함한다. 최종 결과는 표적 부위에서 작은 결실 또는 삽입(삽입 결실)이다. PCR 분석의 상기 결과는 표적 부위에 걸친 예상 크기의 PCR 생성물을 얻는 것이 불가능하다는 점에서 놀라웠으며, 이는 ad3A의 불활성화가 표적 부위에서 작은 삽입 또는 결실(삽입 결실)에 기인한 것이 아니라는 것을 시사

한다. 대신, 이들 데이터는 ad3A의 불활성화가 표적 부위에서의 염색체 재배치 또는 큰 삽입에 의해 야기되었을 가능성과 일치한다.

[0264] **글루코아밀라아제(GA) 유전자의 불활성**

[0265] 4개의 주요 분비된 셀룰라아제를 인코딩하는 유전자(*cbh1*, *cbh2*, *eg11* 및 *eg12*)가 결실된 공개적으로 이용가능한 RL-P37 균주에서 유래된 *트리코테르마 레에세이* 균주를 사용하였다. 이 균주에는 기능적 *pyr4* 유전자가 결여되어 있다. 동량의 pTrex2gHyg MoCas 및 p219M gTrGATM TS2의 혼합물을 이용한 생물학적 방법을 사용하여 이 균주를 동시에 형질전환하였다. 1% 포도당, 100 µg/ml 하이그로마이신 B 및 2 mg/ml 우리딘을 함유하는 보겔(Vogel)의 최소 배지를 사용하여 한천 플레이트 상에서 형질전환체를 선택하였다. 제1 플레이트에서 선택한 후, 형질전환체 콜로니를 동일한 선택 배지의 새로운 플레이트에 집어 넣었다. 제2 플레이트에서 성장하는 동안, 안정한 및 불안정한 하이그로마이신-내성 형질전환체를 구별하는 것이 가능하였다. 17개의 안정한 형질전환체 및 4개의 불안정한 형질전환체를 수득하였다. 분비된 글루코아밀라아제 존재 여부를 선별하기 위해, 이들 형질전환체를 글루코스가 없고 1% 불용성 전분을 함유한 보겔(Vogel) 한천 플레이트로 옮겼다. 글루코아밀라아제를 분비할 수 있는 콜로니는 잘 자라며 포자를 형성한다. 글루코아밀라아제를 분비할 수 없는 콜로니는 매우 드문 균사체로 자라며 명확하게 구별할 수 있다. 17개의 안정한 형질전환체 중 14개는 글루코아밀라아제를 분비할 수 없었고, 4개의 모든 불안정한 형질전환체는 글루코아밀라아제를 분비하지 않았다.

[0266] pTrex2gHyg MoCas 및 p219M gTrGA TS2와의 공동 형질전환체에서 *gla1*(글루코아밀라아제) 유전자좌에서 돌연변이를 판단하기 위해, 5개의 안정한 글루코아밀라아제 비생산 형질전환체로부터 게놈 DNA를 추출하였다. 이 DNA는 Cas9 표적 부위에 걸쳐 있거나 표적 부위의 상류 또는 하류에 위치하는 생성물을 생성하도록 설계된 상이한 프라이머 쌍을 사용하여 PCR을 위한 주형으로 사용되었다. 제조사의 지시에 따라 PfuUltra II Fusion HS DNA 중합효소(Agilent Technologies)를 PCR에 사용하였다. 각각의 경우에 연장 시간은 아래에 설명된 것처럼 PCR 제품의 예상 크기에 대해 제조업체가 제안한 시간이었다. PCR 생성물의 크기는 아가로스 겔 전기 영동으로 평가하였다.

[0267] *gla 1*에서 TS2 표적 부위에 걸친 *glaA* + *glaB* 프라이머(5'- ccgtagttgaagatccttgccc [서열 번호 26] 및 5'- gtcgaggatttgcttcacac [서열 번호 27])를 사용하여 임의의 형질전환체에 대한 PCR 생성물을 수득할 수 없었다. Cas9 활성으로 인한 큰 크기 변화가 없다고 가정할 때 이 PCR 생성물의 예상 크기는 약 1371 bp였다.

[0268] TS2 표적 부위의 5' 측의 영역을 증폭시키는 *glaA* + *glaJ* 프라이머(각각 5'-ccgtagttgaagatccttgccc [서열 번호 26] 및 5'-tgccgactttgtccagtattcg [서열 번호 30])를 사용하여 모든 형질전환체에서 예상 크기(364 bp)의 밴드를 수득하였다.

[0269] TS2 표적 부위의 3' 측의 영역을 증폭시키는 *glaK* + *glaB* 프라이머(각각 5'-ttacatgtggacgcgagatagcg [서열 번호 31] 및 5'-gtcgaggatttgcttcacac [서열 번호 27])를 사용하여 형질전환체 중 4개에서 예상 크기(520 bp)의 밴드를 수득하였다. 형질전환체들 중 하나는 이 프라이머 쌍으로 PCR 생성물을 제공하지 못했다.

[0270] 표적 Cas9 작용에 의해 *gla1* 유전자의 불활성화를 입증하고자 하는 별도의 실험을 RL-P37 유래되고 불활성화 *pyr4* 유전자를 갖는 *T. 레에세이* 균주를 사용하여 수행하였다. 이 균주의 원형질을 폴리에틸렌 글리콜 매개 절차를 사용하여 pTrex2gHyg MoCas gTrGA TS11로 형질전환시켰다(하기에 기술). 2% 포도당, 2 mg/ml 우리딘, 1.1M 소르비톨 및 100 ug/ml 하이그로마이신 B를 함유하는 Vogel 최소 배지의 한천 평판에서 형질전환체를 선택하였다. 제1 플레이트에서 선택한 후, 형질전환체 콜로니를 소르비톨이 없는 동일한 선택 배지의 새로운 플레이트에 집어 넣었다. 제 2 플레이트에서 성장하는 동안, 안정한 및 불안정한 하이그로마이신-내성 형질전환체를 구별하는 것이 가능하였다. 분비된 글루코아밀라아제 존재 여부를 선별하기 위해, 이들 형질전환체를 글루코스가 없고 1% 불용성 전분을 함유한 보겔(Vogel) 한천 플레이트로 옮겼다. 글루코아밀라아제를 분비하지 않는 B#1, B#2, B#4, B#5 및 B#6으로 명명된 5개의 안정한 형질전환체를 추가 분석을 위해 선택하였다. 이들 형질전환체 각각으로부터 게놈 DNA를 추출하였다.

[0271] 게놈 DNA를 주형으로써 그리고 TS11 표적 부위에 걸쳐 야생형 *gla 1* 유전자좌로부터 983 bp의 생성물을 생성하는 *gla1repF* 및 *gla1repR* 프라이머(각각 5'- gtgtgtctaatagcctccaccac [서열 번호 32] 및 5'- gatcgtgctagcgtgctgttg [서열 번호 23])를 사용하여 PCR을 수행하였다. PCR 조건은 매 PCR 사이클마다 프라이머 어닐링 온도를 점차적으로 낮추고 표적 부위에 큰 삽입이 있었는지를 결정하기 위해 긴 연장 시간을 포함시켰다. 구체적인 PCR 조건은 다음과 같다.

[0272] 단계 1: 94°C에서 1분

- [0273] 단계 2: 94°C에서 25초
- [0274] 단계 3: 63°C에서 30초(온도는 사이클마다 0.2°C씩 감소)
- [0275] 단계 4: 70°C에서 8분
- [0276] 단계 2~4를 24회 더 반복
- [0277] 단계 5: 4°C에서 유지
- [0278] 형질전환체 중 2개(B#1 및 B#6)로부터 12 kb를 초과하는 명확한 PCR 생성물을 획득하였는데, 이는 표적 부위에 걸친 DNA 영역에서 11 kb를 초과하는 증가를 나타낸다. 다른 3개의 형질전환체는 아가로스 겔 전기영동에 저 강도 밴드로 나타난 비특이적 PCR 생성물만을 제공했다. B#6로부터 12 kb 초과 PCR 생성물의 서열 분석은 플라스미드 pTrex2gHyg MoCas gTrGA TS11에서 유래된 DNA가 TS11 표적 부위에 삽입되었음을 입증했다.
- [0279] 게놈 DNA 샘플 B#2, B#4 및 B#5 및 프라이머 쌍 1553R 및 1555F(5'-CCGTGATGGAGCCCGTCTTCT [서열 번호 34] 및 5'-CGCGGTGAGTTCAGGCTTTTC [서열 번호 35])를 각각 사용하여 PCR을 수행하였다. 프라이머 1553R은 표적 부위 11의 3' 측의 glal1 유전자에 결합한다. 프라이머 1555F는 플라스미드 pTrex2gHyg MoCas gTrGA TS11 상의 하이그로마이신 포스포트랜스피라제(hygB) 유전자의 개시 코돈 근처에 결합한다. 위의 것과 동일한 PCR 조건을 사용하였다. 형질전환체 B#4 및 B#5 각각에 대해 4.5 kb 및 6.5의 PCR 생성물이 획득되었다. PCR 생성물은 hygB 유전자를 가진 플라스미드가 glal1 유전자에 삽입된 경우에만 획득되어야 한다. 추측컨대, 형질전환체 B#4 및 B#5 내 삽입된 플라스미드 DNA는 너무 커서 프라이머 glalrepF 및 glalrepR을 사용하여 PCR 생성물을 획득할 수 없었다.
- [0280] 종합하면, PCR 데이터는 글루코아밀라아제 불활성화를 갖는 안정한 하이그로마이신 내성 형질전환체가 glal1 유전자의 표적 부위에 Cas9 및 가이드 RNA 발현 벡터의 큰 세그먼트의 삽입을 통해 생성되었음을 입증하였다.
- [0281] **pyr2 유전자의 불활성화**
- [0282] *T. 레에세이* pyr2 유전자 내의 상이한 위치를 표적으로 하는 가이드 RNA 발현 카세트를 포함하는 플라스미드 pTrex2gHyg MoCas의 유도체로 원형질체의 PEG-매개된 형질전환에 의해 *T. 레에세이* 균주의 형질전환체 QM6a 또는 RL-P37이 생성되었다. 이 유전자의 불활성화는 우리딘 영양요구성을 부여하고 5-플루오로오로트산(5-fluoroorotic acid(FOA))에 대한 내성을 부여한다. 형질전환체는 초기에 하이그로마이신 B를 함유하는 배지에서 선별되었다. 하이그로마이신 B를 함유하는 신선한 한천 플레이트로 옮길 때, 이들은 안정적이거나 불안정한 것으로 기록되었다. 이어서, 형질전환체를 2 mg/ml의 우리딘 및 1.2 mg/ml의 FOA를 갖는 보겔 최소 배지의 한천 플레이트로 옮겼다. FOA의 존재하에 성장할 수 있는 능력은 pyr2 유전자의 Cas9-매개 불활성화로 인한 우리딘 영양요구성의 지표이다.
- [0283] PCR 분석을 위해 FOA 내성 하이그로마이신 안정 및 불안정 형질전환체의 일부로부터 게놈 DNA를 추출하였다. 이 분석을 위해 사용된 프라이머는 표적 부위에 걸친 pyr2 유전자좌의 영역을 증폭시키기 위해 설계된 pyr2F(5'-gtataagagcaggaggaggag [서열 번호 36]) 및 pyr2R(5'- gaacgcctcaatcagtcagtcg [서열 번호 37])였고 약 0.8 kb의 길이를 가졌다.
- [0284] FOA 내성이 있는 것으로 나타난 QM6a 형질전환체 중, 안정한 18개 및 불안정한 5개의 하이그로마이신 내성 형질전환체를 야생형 균주와 유사한 크기로 추정하여 pyr2 유전자좌의 영역을 증폭시키기에 충분한 연장 시간을 갖는 PCR 프로토콜을 사용하여 시험하였다. 안정한 형질전환체 중 어느 것도 이 짧은 연장 시간을 갖는 PCR 생성물을 생성하지 않는 반면, 불안정한 형질전환체 중 2개는 PCR 생성물을 생성하였다. 이들 두 PCR 생성물에 대한 DNA 서열 분석 결과, 하나가 단일 뉴클레오타이드 결실을 가지며, 다른 하나는 예상된 표적 부위에서 111 nt 결실을 가졌다.
- [0285] FOA 내성이 있는 것으로 나타난 RL-P37 형질전환체들 중, 안정한 4개 및 불안정한 4개의 하이그로마이신 내성 형질전환체를 짧은 연장 시간을 갖는 PCR 프로토콜을 사용하여 시험하였다. 안정한 형질전환체 중 어느 것도 이 짧은 연장 시간을 갖는 PCR 생성물을 생성하지 않는 반면, 불안정한 형질전환체 중 둘 모두는 PCR 생성물을 생성하였다. 이들 두 PCR 생성물에 대한 DNA 서열 분석 결과, 하나가 단일 뉴클레오타이드 결실을 가지며, 다른 하나는 예상된 표적 부위에서 134 nt의 삽입을 가졌다. 이 삽입은 pTrex2gHyg 벡터의 두 작은 단편으로 구성된다.
- [0286] 큰 DNA 단편이 pyr2 유전자좌의 표적 부위에 삽입된 것으로 추정하고서, pyr2 유전자좌 영역의 증폭을 가능하게



aggagagcggcataggtattgtaatagaccagctttacataataatcgccctgttgcactgactgatgaccttctccctaac  
cagtttctaattaccactgcagtgaggataaccctaactcgctctgggggtattattatactgattagcaggtggcttatatagt  
gctgaagtaactataagagtttctgcggggagggtggaaggactataaactggacacagttagggatagagtgatgaca  
agacctgaatgttatcctccgggtggtatagcgaattggctgacctgacagatggttaatggtttaggcaggggttttcagag  
gggacgagaacgcgttctcgaattaacggctctgccccaagctttacggtttctaatagggcggccgc

(서열 번호 38)

[0291]

[0292]

이 박터는 원형질체의 PEG-매개 형질전환에 의해 *T. 레에세이* 균주 RL-P37에 삽입된다. 형질전환체는 하이그로마이신 저항성을 위해 선택되었고 하이그로마이신이 함유된 신선한 한천 플레이트로 옮겨졌다. 대다수의 형질전환체는 불안정한 하이그로마이신 내성 표현형을 보였다. 개별적으로 형질전환된 콜로니를 2 mg/ml 우리딘 및 1.2 mg/ml 5-플루오로오르트산이 함유된 최소 배지로 옮겨서 성장할 수 있는 개체를 선택하여 Pyr-마이너스 표현형을 얻었다. 142개의 불안정한 형질전환체 중 8개(6%)가 Pyr-마이너스였다. pyr2 유전자좌의 PCR 및 이들 형질전환체 3종의 서열 분석 결과, 2개는 표적 부위에서 작은 결실(각각 1 bp 및 27 bp)을 나타내었고, 하나는 pTrex2gHyg MoCAS gPyr2 TS6의 박테리아 벡터 부분에서 유래된 68 bp의 삽입과 결합된 1 bp 결실을 가졌다. 다른 5개의 형질전환체는 큰 DNA 단편을 증폭하기 위해 설계된 PCR 조건을 사용함에도 불구하고 PCR 생성물을 제공하지 않았다[PCR 조건: 단계 1: 94°C에서 1 분; 단계 2: 94°C에서 25 초; 단계 3: 63°C에서 30초(온도가 사이클마다 0.2°C씩 감소); 단계 4: 70°C에서 8 분; 단계 2-4는 24회 더 반복. 단계 5: 4°C에서 유지. 폴리머라아제: PfuUltra II Fusion HS DNA 중합 효소 (Agilent Technologies)].

[0293]

이러한 결과는 자율적으로 복제하는 벡터로부터의 Cas9 및 가이드 RNA의 발현이 특정 유전자좌(이 경우 pyr2)에 대한 Cas9 표적화를 가능케 함을 나타낸다. 수득된 유전자 불활성화는 벡터 DNA를 표적 부위에 삽입하지 않고 일어날 수 있다.

[0294]

실시예 8: 상동성 통합에 의한 유전자 편집

[0295]

다음 실험을 위해 트리코테르마 레에세이 균주 T4(1)7을 사용했다. 이는 증가된 셀룰라아제 생산성을 선별함으로써 RL-P37로부터 유래된 균주로, pyr2 유전자를 불활성화하여 균주를 우리딘 영양요구인자로 만드는 단일 점 돌연변이를 갖는다.

[0296]

하기 서열(서열 번호 39)을 갖는 Glalrep로 불리는 합성 DNA 단편을 설계하고 수요자 맞춤형으로 만들었다.

gtgtgtctaattgcctccaccacaggaaccaaaccggctttgacctctgggaagaagtcaatgggagctcattcttactgtt  
gccaaccagcaccgaggtatgaagcaaactcctcgacattcgctgctactgacacatgagcattgttactgaccagctctac  
agcactgtcgagggcgccactctgtgcccactctggccagctcggaagcgcctattcatctgttctccccaggtttgtg  
ctttctccaacgattctgggtgctgctggtggatagctgactccaacagatgtctttcactgtttatagagattggccaata  
ctgatagctcgctctagtcaacaccaacgagggcaggactggcaaggatgtcaactccgtcctgacttccatccacacc  
ttcgatccaaccttggtgtgacgcaggcacctccagccatgacgtgacaaagcgcctccaacctcaaggttgttctg  
actcctccgctccatctacggcgtgaacaagggcattcctgCGgtgctgcccctgcgcattggcggatgacagaggatgt  
gtactacaacggcaacccttggtatctgtactcatttctgctgctgcccagcagctgtacgatgcatctacgtctggaagaag  
acgggctccatcacgggtgaccgccacctccctggccttctccaggagctgttcttctggcgtgacggccgggacctactcc  
agcagctctcgacctttaccaacatcatcaacgccgtctcgacatacgccgatggcttctcagcggaggctgccaagtac  
gtccccgcgacgggtctgctgcccagcagtttgaccgcaacagcggcactccgctgtctgctgcttccactgacgtggtc  
gtacgctcgttcttgacagccacggcccgtcgggctggcatcgtgccccctcgtgggccaacagcagcgttagcagc  
atc

[0297]

[0298]

Glalrep 서열은 ORF 내 gla1 유전자좌의 982 bp이다. 이 서열은 TS11 표적 부위(밑줄)에 걸쳐 있다. Glal 코딩 서열에서 프레임 쉬프트를 생성하고 TS11에 인접한 PAM을 파괴하기 위해, TS11 표적 부위의 바로 상류의 야생형 Gla1 유전자의 “CCG” PAM 서열 내 단일 “C” 뉴클레오티드를 결실시켜, Cas9에 의한 절단을 방지한다. PAM의 나머지 두 뉴클레오티드는 대문자 볼드체로 표시된다.

[0299]

프라이머 glalrep F 및 glalrep R(각각 5'-gtgtgtctaattgcctccaccac [서열 번호 32] 및 5'-

gatcgtgctagcgctgctgttg [서열 번호 33])을 사용하는 형질전환에 사용하기 위해 Glalrep 단편을 PCR로 증폭시켰다.

[0300] T. 레에세이 균주 T4(1) 7의 원형질은 pTrex2gHyg MoCas GTrGA TS11B (2 μg) + Glalrep (8 μg)로 PEG-매개 방법에 의해 공동 형질전환되었다. 50 ug/ml 하이그로마이신 B, 2 mg/ml 우리딘 및 1.1 M 소르비톨을 함유한 보겔의 최소 배지를 사용하여 한천 플레이트에서 형질전환체를 선택하였다. 플라스미드 pTrex2gHyg MoCas gTrGA TS11B는 TS11 가이드 RNA의 발현 카세트가 나머지 플라스미드에 대해 반대 방향으로 있다는 것을 제외하면 pTrex2gHyg MoCas gTrGA TS11과 동일하다.

[0301] 형질전환체를 우리딘과 하이그로마이신을 갖는 보겔의 최소 배지의 새로운 한천 플레이트에 집어 넣었고, 안정한 및 불안정한 하이그로마이신 내성 표현형을 구별하는 것이 가능했다. 글루코아밀라아제 양성 또는 음성 표현형에 대해 점수를 매기기 위해 형질전환체를 유일한 탄소원으로서 우리딘 및 1% 불용성 전분을 갖는 보겔 최소 배지의 한천 플레이트로 옮겼다. 안정한 하이그로마이신 저항성 형질전환체의 약 83%는 글루코아밀라아제 생산에 대해 음성이었지만, 불안정한 하이그로마이신 내성 형질전환체의 15%는 글루코아밀라아제 생산에 대해 음성이었다. 글루코아밀라아제-마이너스 표현형을 갖는 7개의 불안정한 형질전환체를 비선택성 한천 배지(보겔의 배지 + 우리딘)로 옮기고 1주 동안 성장시켰다. 그 후 보겔의 배지 + 우리딘 + 하이그로마이신의 플레이트에 집어 넣을 때 이들은 모두 하이그로마이신에 민감하여 pTrex2gHyg MoCas gTrGA TS11B와 연관된 하이그로마이신 내성 유전자의 손실을 나타냈다.

[0302] pTrex2gHyg MoCas GTrGA TS11B 및 Glalrep(형질전환체 #31, 107, 114, 118 및 120)로 얻은 5개의 불안정한 하이그로마이신-민감성 및 글루코아밀라아제-음성 형질전환체로부터 게놈 DNA를 단리하여 약 3.2 kb에 걸친 TS11 또는 glaK[서열 번호 31](하기 참조) 및 glaH 5' -tgccgtgggtcattggcatattc[서열 번호 29]를 증폭하도록 설계된 프라이머 glaA 및 glaD(각각 5'-ccgttagttgaagatccttgccc [서열 번호 26] 및 5'-gagagacgcaggatgactcaaag [서열 번호 28])를 사용하는 PCR(위에서 기재된 프로그램)에서 주형으로 사용하였다. glalrep F 및 glalrep R(각각 5'-gtgtgtctaatgcctccaccac [서열 번호 32] 및 5'-gatcgtgctagcgctgctgttg [서열 번호 33])을 프라이머로 사용하여 PCR 생성물을 서열 분석하여 표적 부위 TS11에서의 변이를 결정하였다. 형질전환체 중 하나는 TS11과 관련된 PAM에서 단일 bp 결실을 도입하고 gla1 유전자를 불활성화시킨 gla1 유전자좌에서 Glalrep의 상동종 재조합과 일치하는 PCR 및 서열 분석 결과를 나타내었다. 형질전환체 중 2개는 TS11 표적 부위에서 작은 삽입 결실을 가지지만 다른 2개는 이 부위를 통한 상동성 통합보다는 Glalrep 단편의 Cas9 절단 부위 내로의 삽입을 보였다.

[0303] pTrex2gHyg MoCas GTrGA TS11A(가이드 RNA 발현 카세트가 벡터 내에서 반대 방향으로 존재하는 것을 제외하고 pTrex2gHyg MoCas GTrGA TS11B와 동일함) + gla 1 유전자좌에서 상동성 재조합에 의한 통합을 하도록 설계된 선형 DNA 단편을 사용하여 원형질체를 동시에 형질전환시킨 위의 실험을 반복하였다. 그러나, 글루코아밀라아제 유전자 내의 표적 부위 TS11에서 상동종 재조합을 위한 공여체로서 982 bp Glalrep DNA 단편을 사용하는 대신, GlalrepL이라 불리는 더 길고 약 2 kb의 단편을 사용하였다. GlalrepL의 중앙 부분은 Glalrep와 동일한 서열이었지만 단편의 5' 및 3' 말단은 gla1 유전자좌의 상류 및 하류 부분을 더 포함하도록 연장되었다. 상기에서 사용된 균주 T4(1)7 대신 트리코테르마 레에세이 균주 RL-P37을 이 실험에 사용하였다. 대조군으로 GlalrepL과 pTrex2gHyg MoCas로 원형질체를 공동 형질전환시켜 활성 Cas9가 없는 gla1 유전자좌에서 GlalrepL이 통합되는 빈도를 결정하였다. 형질전환 및 표현형 선별에 이어 형질전환체를 다음 범주에 할당할 수 있었다.

DNA 형질전환	Hyg <sup>R</sup> 형질전환체가 없음	글루코아밀라아제-마이너스 형질전환체가 없음(각각의 Hyg <sup>R</sup> 형질전환체 중 %)
pTrex2gHyg MoCAS + GlalrepL	26 개 불안정 19 개 안정	0 2(10%)
pTrex2gHyg MoCAS gTrGA TS11A + GlalrepL	52 개 안정 16 개 불안정	46(88%) 14(87%)

[0304]

[0305] pTrex2gHyg MoCas gTrGA TS11A 및 GlalrepL(안정한 형질전환체 #51, 52, 60, 61 및 67, 불안정한 형질전환체

338, 41, 65, 66 및 67)로 얻은 5개의 안정한 및 5개의 불안정한 하이그로마이신-민감성 및 글루코아밀라아제-음성 형질전환체로부터 게놈 DNA를 단리하여 프라이머 glaA 및 glaD를 사용한 PCR에서 주형으로 사용하였다(상기 참조). gla1 유전자의 표적 부위 TS11에 삽입이 없거나 큰 결실이 발생하지 않으면 PCR 생성물은 3.2 kb가 될 것으로 예상된다.

[0306] 5개의 안정한 형질전환체 중 3개(#52, 60 및 61)는 약 3.2 kb의 PCR 생성물을 제공하였고, 다른 2개는 TS11에서 DNA 삽입을 나타내는 더 큰 생성물을 제공하였다. glarepF를 프라이머로서 사용하여 약 3.2kb의 3개의 PCR 생성물을 서열분석하였다. 2개의 형질전환체(#52 및 #61)의 경우 서열분석 결과는 gla1 유전자좌에서 상동성 재조합에 의한 GlarepL의 통합과 일치하고 다른 하나는 쉽게 해석할 수 없는 혼합 신호를 가졌다.

[0307] 5개의 불안정한 형질전환체 중 하나(#66)만이 약 3.2 kb의 PCR 생성물을 제공하였고, 나머지 4개는 TS11에서 DNA 삽입을 나타내는 더 큰 생성물을 제공하였다. glarepF를 프라이머로 사용하여 약 3.2kb의 하나의 PCR 생성물을 서열분석하였고 결과는 gla1 유전자좌에서 상동성 재조합에 의한 GlarepL의 통합과 일치하였다.

[0308] 종합하면, 이들 결과는 선형 DNA 단편의 상동성 통합이 표적 부위에서 Cas9 절단에 의해 자극될 수 있음을 보여준다. 그러나, 비상동성 말단 결합(NHEJ)에 의한 DNA의 작은 삽입 결실 또는 큰 삽입은 또한 일반적인 발생이다. 더 큰 상동성 선형 DNA 단편의 사용은 다른 경우에 비해 표적 부위의 상동성 통합의 빈도를 향상시키는 데 도움이 된다. 하이그로마이신이 없는 배지에서 성장을 허용함으로써 pTrex2g MoCas-기질 벡터가 후속적으로 제거될 수 있는 불안정한 하이그로마이신 내성 형질전환체 내 표적 부위에서 상동성 통합을 얻는 것이 가능하다.

[0309] 실시예 9: T. 레에세이의 NHEJ-결핍 균주에서 유전자 편집

[0310] *트리코테르마 레에세이*(RL-P37에서 유래되고 결실된 셀로바이오하이드로라아제 1, 셀로바이오하이드로라아제 2, 엔도글루카나아제 1, 및 엔도글루카나아제 2 유전자( $\Delta cbh1$ ,  $\Delta cbh2$ ,  $\Delta egl1$ , 및  $\Delta egl2$  균주; WO 92/06184 및 WO 05/001036 참조))의 "사중-결실 균주"로부터 유래되고 천연 엔도글루카나아제-3 및 베타글루카나아제-1 유전자에서 결실을 갖는 균주(MAD6)를 Cas9 표적 부위에서 비상동성 말단 결합(NHEJ) DNA 삽입의 역할을 판단하기 위해 설계된 실험을 위해 사용하였다. 또한 DNA 재조합을 위한 주요 NHEJ 경로에 필수적인 인간 ku80에 대해 병렬상동(orthologous)인 천연 유전자에 대해 MAD6 균주는 결실되었다(MAD6 균주가 제조되는 방법의 설명을 위한 US20130149742 A1 "Filamentous fungal host strains and DNA constructs, and methods of use thereof" 참조). 상기 균주를 위에서 기술된 pTrex2gHyg MoCAS gTrGA TS11B + 공여체 Glarep 단편으로 상기 균주를 공동 형질전환시켰다. gla1 유전자좌에서 상동성 재조합에 의한 이 단편의 통합은 PAM 서열로부터 하나의 bp를 결실시킴으로써 gla1 유전자를 불활성화시키고 TS11 표적 부위를 제거할 것이다. 형질전환체는 PEG-매개 방법에 의해 원형질체로부터 얻어졌다. 1.1M 소르비톨 및 100 ug/mL 하이그로마이신 B를 함유하는 보겔의 최소 배지에서 형질전환체에 대한 선택을 수행하였다. 하이그로마이신만을 함유한 최소 배지의 새로운 한천 플레이트로 옮긴 91개의 형질전환체 중에서 4개는 안정한 하이그로마이신 내성 표현형을 가졌다(비선택적 조건 하에서의 성장 기간 후 하이그로마이신을 갖는 배지상에서 재도포시 성장하는 이들의 능력으로 확인됨). 유일한 탄소원으로서 1% 불용성 전분을 갖는 보겔1의 최소 배지로 모든 형질전환체를 옮기고, 4개의 안정한 형질전환체를 포함하여 17개(18%)가 글루코아밀라아제에 음성으로 나타났다. PCR 및 DNA 서열 분석 결과, 공여체 Glarep이 gla1 좌에서 상동성 재조합에 의해 통합된 경우, 13개의 불안정한 형질전환체 중 12개와 안정한 형질전환체 중 하나가 TS11 PAM에서 단일 bp 결실을 갖는 것으로 나타났다. 다른 불안정한 형질전환체는 글루코아밀라아제 음성 표현형을 가짐에도 불구하고 야생형 gla1 서열을 가졌다. 다른 3개의 안정한 형질전환체는 gla1 유전자좌에 대해 예상되는 크기의 명확한 PCR 생성물을 제공하지 않았고, 이것들에서 벡터 또는 공여체 Glarep 삽입이 일어날 수 있다. 모든 불안정한 글루코아밀라아제 음성 형질전환체를 하이그로마이신이 없는 배지에서 성장시켰고 하이그로마이신을 함유하는 배지로 다시 옮겼다. 어느 것도 성장할 수 없었는데, 이는 이들이 pTrex2gHyg MoCAS gTrGA TS11B 벡터를 잃었음을 나타낸다.

[0311] 이러한 결과는 NHEJ가 결핍된 균주에서 Cas9 표적 부위에 벡터 또는 공여체 DNA 단편 삽입이 최소화되었음을 명확하게 보여준다. 결과적으로, Cas9 및 가이드 RNA의 일시적인 발현과 함께 불안정한 형질전환체에서 공여체 DNA 단편의 상동성 재조합을 통해 매우 특이적인 유전자 편집(단일 bp의 결실)의 높은 빈도가 가능하다.

[0312] **섹션 B: Cas 니카아제/가이드 RNA의 직접 도입**

[0313] 실시예 10: 대장균에서 CRISPR SpyCas9-D10A 니카아제의 이중 발현

[0314] Generay(중국, 상해)는 대장균 코돈에 최적화된 스트렙토코쿠스 파이오제네스 Cas9-D10A(SpyCas9-D10A) 니카아

제 유전자를 합성하여 NcoI 및 HindIII 부위에서 발현 벡터 pET30a에 삽입하여 플라스미드 pET30a-SpyCas9-D10A 닉카아제를 수득하였다(도 7). 도 7의 플라스미드 지도에 표시된 바와 같이, 발현 카세트의 전체 코딩 서열은 5'에서 3'방향으로, 모든 작동가능한 연결에서 N-말단 His6 태그/트롬빈/S · Tag<sup>TM</sup>/엔테로키나아제 영역을 인코딩하는 서열(서열 번호 68, 메티오닌을 위한 개시 코돈을 포함), SV40 핵 위치 신호를 인코딩하는 서열(서열 번호 69), SpyCas9-D10A 닉카아제를 인코딩하는 서열(서열 번호 65), 및 BLR2 핵 위치 신호를 인코딩하는 서열(서열 번호 70)을 포함한다. 이 전체 코딩 서열을 서열 번호 54에 나타낸다. 서열 번호 68에 의해 인코딩되는 N-말단 His6 태그/트롬빈/S · 태그<sup>TM</sup>/엔테로키나아제 영역의 아미노산 서열은 서열 번호 67(위치 1의 메티오닌을 포함함)에 나타나고, 서열 번호 69에 의해 인코딩되는 SV40 핵 위치 신호의 아미노산 서열은 서열 번호 46에 나타나고, 서열 번호 65에 의해 인코딩되는 SpyCas9-D10A 닉카아제의 아미노산 서열은 서열 번호 66에 나타나고, 서열 번호 70에 의해 인코딩되는 BLR2 핵 위치 신호의 아미노산 서열은 서열 번호 47에 나타낸다. 서열 번호 54에 의해 인코딩되는 아미노산 서열은 서열 번호 55에 나타낸다.

[0315] pET30a-SpyCas9-D10A 닉카아제 플라스미드를 Rosetta2 (De3)plysS 대장균 균주(독일, 다름슈타트의 Novagen®, EMD Biosciences, Inc., Merck KGaA)로 형질전환시키고, 형질전환 생성물을 34 ppm 클로로암페니콜(Chloramphenicol) 및 50 ppm 카나마이신(Kanamycin)으로 보충된 Luria 한천 플레이트 위에 펼쳤다. 콜로니를 집어 넣고 250 ml 셰이크 플라스크에서 25 ml의 Invitrogen MagicMedia<sup>TM</sup>(Thermo Fisher Scientific Inc.)으로 24시간 동안 30°C에서 300 rpm으로 발효시켰다.

[0316] SpyCas9-D10A 서열이 유래된 스트렙토코쿠스 피오제네스 유래의 야생형 Cas9 단백질의 아미노산 서열을 서열 번호 45로 나타낸다.

[0317] 실시예 11: SpyCas9-D10A의 정제

[0318] SpyCas9 (D10A)의 정제를 위해 친화성, 소수성 상호작용 및 크기 배제 크로마토그래피 단계들의 조합을 적용하였다. 2리터의 미정제 원액을 채취하여 원심분리 하였다. 세포를 펠릿화하고 400 ml 용해 완충액(20 mM HEPES, pH7.5, 500 mM NaCl, 0.1% Triton X-100, 1 mM DTT 및 1 mM TCEP, Roche에서 구입한 프로테아제 억제제 콕테일)에서 재현탁하였고, 초음파 분쇄기(35% 파워, 20분, 2초 온/3초 오프)(중국, 제정의 SCIENT2-II D, Ningbo Scientz Biotechnology Co., LTD.)로 세포를 용해시켰다. 용해물을 20,000 g에서 40분 동안 원심분리하여 제거하였다.

[0319] 정화된 용해물을 롤링 인큐베이터(중국, 하이멘의 Kylin-Bell Lab Instruments Co., Ltd.)에서 4°C에서 30 rpm으로 Ni-NTA 수지(GE Healthcare)로 밤새 배양하였다. 원심 분리 후, 수지를 XK26/20 컬럼(GE Healthcare)에 옮기고 AKTA Explorer 시스템(GE Healthcare)에 연결하였다. 평형 완충액(20 mM HEPES, pH 7.5, 300 mM NaCl, 0.1 % Triton X-100) 및 세척 완충액(평형 완충액 내 25 mM 이미다졸)으로 광범위하게 세척한 후, 표적 단백질을 평형 완충액 내 50, 250 및 500 mM 이미다졸로 용출시켰다. 원하는 단백질은 50 및 250 mM 이미다졸 용출액에서 상대적으로 높은 순도로 발견되었고, 이들을 모으고 별도로 추가처리 하였다.

[0320] 친화성 단계로부터 수집된 활성 분획에 0.6 M까지 황산암모늄을 첨가하고 20 ml 페닐-세파로스 HP 칼럼(GE Healthcare)에 로딩하였다. 칼럼을 HEPES 완충액(pH 7.5)에서 0.6 M 내지 0.0 M 황산암모늄의 구배로 용출시켰다. 각 분획의 순도는 SDS-PAGE 겔에 의해 평가되었으며, 관심 단백질이 주로 유류(flow-through) 분획에 존재함이 밝혀졌다.

[0321] 최종적으로, 20 mM HEPES pH 7.5, 150 mM KCl 및 10% 글리세롤 중의 Superdex 200 16/60 칼럼(GE Healthcare) 상의 크기 배제 크로마토그래피로 단백질을 추가로 정제하였다. 단백질-함유 순수 분획들을 모으고 Amicon 30 KDa 멤브레인 필터(Millipore)를 사용하여 농축시켰다. 정제 단백질(각각 50 mM 이미다졸 및 250 mM 이미다졸 용출액)의 두 배치를 사용하기 전까지 150 mM KCl 및 40% 글리세롤을 함유한 20 mM HEPES 완충액에 pH7.5에서 -20°C에서 보관하였다.

[0322] 실시예 12: 닉카아제의 시험관내 분석

[0323] **시험관내 닉카아제 절단 분석을 위한 기질 DNA 단편의 제조**

[0324] RL-P37에서 유래된 *트리코테르마 레에세이* 균주에서 추출되었으며 결실된 셀로바이오하이드로라아제 1, 셀로바이오하이드로라아제 2, 엔도글루카나아제 1 및 엔도글루카나아제 2 유전자( $\Delta cbh1$ ,  $\Delta cbh2$ ,  $\Delta egl1$  및  $\Delta egl2$  균주; “사중-결실 균주”로도 불림; WO 92/06184 및 WO 05/001036 참조)를 갖는 게놈 DNA를 Zymo Research사(Irvine, CA)의 ZF Fungal/Bacterial DNA miniprep 키트를 사용하여 추출하였다. 1 ng의 추출된 게놈 DNA를

주형으로 하여, 트리코테르마 레에세이 글루코아밀라아제(TrGA) 유전자(유전자 ID: 18483895) 및 이의 부분 5'-UTR (서열 번호 56)을 포함하는 DNA 단편을 KOD-Plus PCR 키트(일본, Toyobo사) 및 순방향 및 역방향 프라이머 각각 0.4 μM을 사용하여 PCR로 증폭시켰다: 5'-gactgtctccaccatgtaatttttc-3'(서열 번호 57) 및 5'-ggcagactacaagtctactagtagtac-3'(서열 번호 58). PCR 생성물을 Zymo Research사의 DNA Clean & Concentrator™-5 키트를 사용하여 정제 및 농축하고 DNA 농도를 NanoDrop™ 분광광도계(Thermo Fisher Scientific Inc.)로 측정하였다.

[0325] 서열 번호 56(하기)은 기질 DNA 단편의 뉴클레오티드 서열을 나타낸다. UTR 서열은 소문자로 나타나고 TrGA 유전자는 대문자로 나타난다. 두 개의 선택된 VT 도메인인 TrGA\_sgF1과 TrGA\_sgR1은 각각 굵은 글씨와 밑줄로 나타난다(이들 순서가 겹쳐있음에 주목).

[0326] gactgtctccaccatgtaattttccctgcgactccatataacgccggatcgtgaaattttcttcttttctctctcaacaa  
acaacggatctgtgctttggcgtcccctgcgttcacgcgcagggctgactgctctgcagctcgataactccatggagccat  
caactgtctatgggtcaatcatcctatcgacaggccaagaacaagccggctccggctcctcattcgtctgcgaaga

cggttgagtggttggtggaggattcggggcccaatattccaaccctttttccaagccgtcggccgggtgaggttgag  
gaaaaccatgggtgctacatattatcgatgctgggtttgtagtagcaatgttgccggagcttgagccgagcctcgt  
ctgggctctgaccaggcaacgccactgactgagcggcgaaggaggaatgattctgacgacgacgagcaatg  
gaacttcaagtaaaagcccagcaaccgaccatgacagatcagaattctctggctggtggggtggttgagactgc  
ttacggagtcgatgcctcgtgactgtcagccgcctcagcctcctggactctgcatattatgacacgagtaaagcc  
tgcatgatgctgctgtcgtctcatgtcgagaacaacacacctggtgctacataggcaatactacctgtagctcaaa  
gttgactgtttgctttgatgtctttgatcatgccatccatccctgtcttcagtgcatgtggatctctacgctcagacggggag  
aaagcttgctgataaagtagcatgatgctgctggtgctacggccctttatccccatcgtcatgctctatatta  
atccaggagactctcctcctggcatgggtgagtagcaagtgcagaggacatgagaagcagagccacgcaacgcttga  
catctgacctatgtgggcaaaaaatcgagacccaccagctcgtcctacattacatgtgaagatcttagcccacaatcctac  
tgtttactagtagtactgacagctgctacacagctcctcgttctgtgaaaccagctcagctcctgagcacatgcagt  
aacgcccagctcggcgtcatttgcacacccaatttgacactgagggatgctggaagctgctgagcagatcccgtaccg  
attcatggcactactacatccatacgcagcaaacatgggctgggctggcttctcaatgcaaaatgcccgcaaaagtc  
cggcattgctgatgcagagatgcagattcagcgggagatttagggtagggcgactactactactaataccacclagtca  
gtatgtatctagcaccggaggttagggcttagtgacgggaacctggctcattccatcgcaaccaggatcccgcactcgt  
tgcgctctgccccacggggcgggagttggcagaggcagaatgaggagcagccccttctctgcccctggcggggcct  
gtgaaagcaagcagagcagagcagagcgggtgagaagcgggttgacgctgacgggtacgaagcagagcgagaat  
cccgttaagccgagggctgggctccccccccgctcatcatgcccacactcctcctccagcccactcgtctccctgcctcgt  
cgctccccctccccgattagctgcatgttctcctgacagcgtgactaatgacgctgtccagcccattcgcctga  
cgatccccggcatctgagctagctcgtcagctggcaatctggcccagcagagcagcaagcggcgggcatgattg  
ggcctgcccctggcggcagcagctgcccacccgagaccgcatcaccgactgtcggatctctccgagc  
agcaggaggctgatcctggccggcagacgattgaaaagggctgcccggcccggagcaggacagcggcgagagc  
gagcgagagagaggaaaagaagaaggtgactgtctttttcagccagccccggctcaacagaagcagaggagaa  
ggcgaacgactcaacgacgacgacgacgacgaagacgggaagtcctgttagttgaagatccttgcctcaca  
caccatctcgtggatattgcttcccctgcccgtgctgtgcccactgttcccctcttctcctcattccgagcgt  
actggttctactccgagcctcgggtgtgcttctcttctgtcaccatgacccgcccctcgcggcacttgggccccggag  
aattcggcccttgcagcattttggccctcagttcccctgaggagcgtccacacttctcttggccctgcagacctttgt  
cgtcgtccgagtcggaagaagctcagcttgagcgttgagtagcatctacgcggaatcactggacaaagtcggcaa  
gacgaagccgctcgtcgtcgtcgtcgtcttactgcagagcggcctccgactggggcatcggcatalataaaaagat  
gcccgccttgcctgacactggccatgagccactcggcctcctctctcaacgcttctctcacacatcctcctctat  
tcccaccatcATGCACGTCCTGTCGACTGCGGTGCTGCTCGGCTCCGTTGCCGTTCAA  
**AAGGTCCTGGAAGACCAGGATCAAGCGGTCTGTCCGACGTCACCAAGAGGTCT**  
**GTTGACGACTTCATCAGCACCGAGACGCCTATTGCACTGAACAATCTTCTTTGCAA**  
**TGTTGGTCCTGATGGATGCCGTGCATTCGGCACATCAGCTGGTGCGGTGATTGCA**

[0327]



[0332] 상기 특정 PCR 생성물을 주형으로 하여, VT 도메인인 TrGA\_sgF1 또는 TrGA\_sgR1에 대한 RNA를 Invitrogen, Thermo Fisher Scientific Inc.의 MEGAshortscript™ T7 전사 키트를 사용하여 제조사의 지침에 따라 시험관내 전사에 의해 생성시켰다. 전사된 RNA를 Invitrogen, Thermo Fisher Scientific Inc.의 MEGAclean™ 전사 클리닝 키트를 사용하여 정제하였다. RNA 농도를 NanoDrop™에 의해 측정하였다.

[0333] 합성된 단일 가이드 RNA의 기능을 확인하기 위해 SpyCas9 시험관내 분석을 수행했다. 분석을 시작하기 위해, 1 µg의 정제된 SpyCas9, 200 ng의 기질 DNA 단편 및 200 ng의 단일 가이드 RNA(또는 대조물로서의 물)를 50 mM HEPES pH 7.3, 150 mM KCl, 0.5 mM DTT 및 10 mM MgCl2를 함유하는 15 µl의 반응 완충액에서 함께 혼합하였다. 37°C에서 20 분간 분석을 수행한 후, 프로테아제 K(Sigma, Cat No. P6556) 2 µg를 첨가하였다. 40°C에서 20분 동안 반응을 계속하였고 80°C에서 20분 동안 추가로 배양하여 반응을 종료하였다. 0.8% 아가로스 겔을 사용하여 140 볼트에서 30분 동안 반응 결과를 분석하였고, 분석한 결과를 도 9에 나타내었다. 도 9에서, 레인 1은 DNA 사다리(분자량은 좌측에 도시)이고, 레인 2는 대조 SpyCas9 반응(가이드 RNA 없음)이고, 레인 3 및 4는 TrGa\_sgR1의 존재 하에 SpyCas9를 나타내고, 레인 5 및 6은 TrGA\_sgF1의 존재 하에서 SpyCas9를 나타낸다. 손상되지 않은 DNA 기질(서열 번호 56)의 크기는 4.9 Kb(레인 2)인 반면, sgRNA 존재 하에서 절단된 생성물의 크기는 2.8 Kb 및 2.1 Kb이다. 도 9에 도시된 바와 같이, 특정 단일 가이드 RNA의 존재 하에서, SpyCas9는 기질 DNA 단편을 원하는 크기로 성공적으로 절단하여 합성된 RNA의 올바른 기능을 확인할 수 있다.

[0334] 서열 번호 59(하기)는 T7 프로모터, CER 도메인, 및 VT 도메인 TrGA\_sgF1의 전사를 위한 올리고뉴클레오티드 서열을 나타낸다. VT 도메인은 대문자로 표시되어 있지만, T7 프로모터와 CER 도메인 영역은 각각 굵은 글씨와 소문자로 표시되어 있다.

**taatacgaactcactataGGGAAGACCAGGATCAAGgttttagagctagaaatagcaagttaaaataaggct**  
**agtcggttatcaactgaaaaagtggcaccgagtcggtgc**

[0335]

[0336] 서열 번호 60(하기)은 T7 프로모터, CER 도메인, 및 VT 도메인 TrGA\_sgR1의 전사를 위한 올리고뉴클레오티드 서열을 나타낸다. VT 도메인은 대문자로 표시되어 있지만, T7 프로모터와 CER 도메인 영역은 각각 굵은 글씨와 소문자로 표시되어 있다.

**taatacgaactcactataGGACAGACCGCTTGATCCgttttagagctagaaatagcaagttaaaataaggcta**  
**gtccggttatcaactgaaaaagtggcaccgagtcggtgc**

[0337]

[0338] 정제된 SpyCas9(D10A)를 이용한 시험관내 니카아제 절단 분석

[0339] 시험관 내 니카아제 절단 분석은 2단계 반응이다. 첫 번째 단계를 위해, 1 µg의 정제된 SpyCas9(D10A), 200 ng의 기질 DNA 단편 및 200 ng의 단일 가이드 RNA(또는 대조물로서의 물)를 50 mM HEPES pH 7.3, 150 mM KCl, 0.5 mM DTT 및 10 mM MgCl2를 함유하는 15 µl의 반응 완충액에서 함께 혼합하였다. 반응은 SpyCas9 핵산 분해 효소 분석법에 기재된 대로 수행하였다. 제1 단계 반응의 종료 후, 1 µg의 SpyCas9(D10A)와 200 ng의 특정 단일 가이드 RNA를 첨가하였다. 제1 단계 반응을 반복한 후, 그 후 반응 결과를 140 볼트에서 30분 동안 작동하는 0.8% 아가로스 겔을 사용하여 분석하였고 결과를 도 10에 나타내었다. 도 10에서, 레인 1은 DNA 사다리(분자량은 좌측에 도시)이고, 레인 2는 SpyCas9(D10A)와 기질 DNA 및 TrGA\_sgF1만의 반응을 나타내고, 레인 3은 SpyCas9(D10A)와 기질 DNA 및 TrGA\_sgR1만의 반응을 나타내고, 레인 4는 SpyCas9 (D10A)와 기질 DNA 및 두 sgRNA의 반응을 나타낸다. 손상되지 않은 DNA 기질(서열 번호 56)의 크기는 4.9 Kb인 반면, 두 sgRNA를 갖는 절단 생성물의 크기는 2.8 Kb 및 2.1 Kb가 될 것이다. 도 10에 도시된 바와 같이, 기질 DNA 단편은 두 RNA(레인 4)의 존재 하에서 절단만 되는데, 이는 SpyCas9(D10A)가 활성 니카아제라는 것을 제시한다.

[0340] 실시예 13: 생체내 니카아제 흡수 실험

[0341] 원형질체 준비

[0342] 원형질체 준비를 위해, 엔도글루카나아제-3, 엔도글루카나아제-4, 엔도글루카나아제-5, 엔도글루카나아제-6, 만나아제-1, 및 알파-아밀라아제 유전자의 추가 결실을 갖지만, 정상 NHEJ 메카니즘(30°C에서 5일 동안 PDA 플레이트에서 성장한)을 이용하여, *T. 레예세이*(위에서 기재된)의 사중-결실 균주의  $5 \times 10^8$ 개 포자를 4개의 배플을 갖는 250 ml 웨이크 플라스크에서 50 ml 발아 배지(미국 특허 제8,679,815호에 기술된 방법)에 접종하였고 27°C에서 170 rpm으로 17시간 동안 배양하였다. 액체 부피를 50 ml의 원뿔형 튜브로 옮기고 3000 rpm으로 10분 동안

회전시킴으로써 균사체를 회수하였다. 상층액을 따라 내고, 균체 펠릿을 1.2 M MgSO<sub>4</sub> -10 mM Na<sup>-</sup> 인산염 완충액을 사용하여 2회 세척하고, 15 ml 용해 효소 완충액(1.2 M MgSO<sub>4</sub>-10 mM Na<sup>-</sup>인산염 완충액(pH 5.8), 50 mg/ml을 사용하는 *트리카테르마 하르시아눔*(Sigma 카탈로그 #L1412)으로부터 용해 효소를 용해)에 재현탁시켰다. 세포 현탁액을 4개의 배플을 갖는 250 ml 셰이크 플라스크로 옮기고 상온에서 200 rpm으로 적어도 2 시간 동안 흔들었다. 유리 깔때기에서 Greiner 튜브로 집힌 Miracloth(Calbiochem Art. No. 475855)를 통한 여과에 의해 원형질체를 수확하였다. 0.6 M 소르비톨-0.1 M Tris-HCl 완충액을 여과된 원형질체 위에 조심스럽게 첨가하였다. 원형질체를 4000 rpm에서 15분 동안 원심분리하여 수집하였다. 원형질체를 함유하는 중층을 새로운 튜브로 옮기고 적어도 같은 부피의 1.2 M 소르비톨-10 mM 트리스-염산 완충액을 첨가하였다. 원형질체를 4000 rpm에서 5분간 원심분리하여 수집하고, 1.2 M 소르비톨, 10 mM 트리스-염산 완충액을 사용하여 2회 세척하였다. 적어도 1 ml 1.2 M 소르비톨-10 mM Tris-HCl pH 7.5-10 mM CaCl<sub>2</sub> 완충액에 펠릿을 재현탁시키고, 원형질체의 수를 현미경으로 계수하였다. 1.2 M 소르비톨-10 mM Tris-HCl-10 mM CaCl<sub>2</sub> 4부 및 25% PEG6000-50 mM CaCl<sub>2</sub>-10 mM Tris-HCl 1부를 사용하여 원형질체 현탁액을 장래의 형질전환을 위해 ml당 5x10<sup>8</sup>까지 희석시켰다.

[0343] **결실 카세트의 준비**

[0344] TrGA 결실 카세트를 구성하여 도 11에 개략적으로 도시하였다. 이는 *pyr2* 프로모터, *pyr2* CDS 및 *pyr2* 종결인자를 포함하는 *pyr2* 발현 카세트를 포함하고, 이는 5' 및 3' TrGA-상동성 영역에 의해 플랭킹된 후속 루프 아웃을 위한 500 bp 반복 서열이다.

[0345] TrGA 녹아웃 형질전환체를 포도당이 없고 1% 불용성 전분(보겔의 전분 배지)이 있는 보겔 한천 접시에서 선별할 수 있다. 손상되지 않은 TrGA 유전자를 갖는 균주와 비교하여 TrGA 녹아웃 형질전환체는 이 배지에서 불충분하게 성장한다. TrGA 녹아웃 카세트의 뉴클레오티드 서열은 길이가 4248 염기 쌍이고: 1 내지 1000 염기는 TrGA 5' 상동성 영역에 상응하고; 1001 내지 2730 염기는 *pyr2* 발현 카세트에 상응하고; 2739 내지 3248 염기는 500 bp 리피트에 상응하고; 3249 내지 4248 염기는 TrGA 3' 상동성 영역에 상응한다. TrGA 녹아웃 카세트의 뉴클레오티드 서열은 서열 번호 61(아래에 표시)로 제공된다.

[0346] agatgccggccttgaagcagtaggtcgagaggctccacttcatggcgccgttgccgatcatgggtgtgatgatgccggtcgtagctctcgatggcgccgtagtagtcgcccgtcgagggcgaggctcggcgtactgcgtccagagctt

[0347] **형질전환**

[0348] Spycas9 Nickase 프리믹스와 결실 카세트를 사용하여 *T. 레에세이*에서 Trga 유전자를 녹아웃하기 위해, 저장 완충액(20 mM HEPES pH 7.5, 150 mM KCl, 40% 글리세롤) 내 Spycas9 Nickase 단백질 25 µg을 3 µl NEB 완충액 3(뉴 잉글랜드 바이오랩스)의 뉴클레아제가 없는 물에 용해된 20 µg sgRNA(TrGA\_sgR1)와 부드럽게 혼합하여 30 µl 프리믹스를 얻었고, 상온에서 30분간 배양했다. 20 µg(20 µl)의 결실 카세트와 함께 200 µl의 원형질체(1 × 10<sup>8</sup>)에 30 µl의 프리믹스를 첨가하고 30분간 얼음 위에 유지시켰다. 얼음 위에서 30분 동안 배양한 후, 최소 보겔 한천 플레이트의 최상층으로 사용되도록, 냉각된 용융 소르비톨/보겔 한천(2% 글루코오스를 함유한 최소 보겔 한천에서 1.1 M 소르비톨)에 원형질체를 첨가했다(Davis 등 (1970) *Methods in Enzymology* 17A, pgs. 79-143 및 Davis, Rowland, *NEUROSPORA, CONTRIBUTIONS OF A MODEL ORGANISM*, Oxford University Press, (2000)). 플레이트를 30°C에서 1주간 배양하였다. 상세한 단계는 미국 특허 제8,679,815호(본원에 참고로 통합됨)에 기재되어 있다.

[0349] 90개의 형질전환체를 선택하여 3개의 대조군과 함께 웰당 1 µg의 신선한 보겔 한천을 갖는 4개의 24- 웰 플레이트에 접종하였다: 엔도글루카나아제-3, 엔도글루카나아제-4, 엔도글루카나아제-5, 엔도글루카나아제-6, 만나아제-1의 추가 결실을 갖는 사중-결실 균주(위에 기재된 *T. 레에세이* 균주; 도 12의 "셀루(cellu)"), 알파-아밀라아제 유전자의 추가 결실(AA 결실, 도 12에서 ΔAA)을 갖는 "셀루"균주, 및 추가의 글루코아밀라아제 유전자 결실(GA 결실; 도 12에서 ΔGA)을 갖는 "셀루" 균주. 일주일 후, 형질전환체를 웰 당 1 ml의 신선한 보겔-전분 한천을 갖는 또 다른 4개의 24- 웰 플레이트로 옮겼다. 일주일 후, 이들 90개 균주의 형태를 관찰하였다. 47개의 형질전환체는 "셀루" 균주 및 AA 결실(ΔAA)을 갖는 것들과 비교하여 지연된 성장을 보였지만 보겔의 전분 플레이트에서 GA 결실(ΔGA)을 갖는 "셀루" 균주의 형질전환체와 유사하게 보였다(도 12). 보겔의 전분 한천에서, 탄소원으로 포도당을 함유한 정규 보겔 한천에 비해 유일한 탄소원은 전분이다. 글루코아밀라아제는 전분의 비환원 말단으로부터 α-1,4 글리코시드 결합을 연속적으로 가수분해하여 곰팡이 균주에 의해 이용될 수 있는 글루코오스를 생산한다. 글루코아밀라아제가 결실된 경우, 글루코오스의 이용가능성이 제한되어 지체된 성장이 관

찰될 수 있다.

[0350] **균주 검증**

[0351] 보겔의 전분 한천에서 90개의 균주 중 47개의 균주가 성장 지연을 나타내었다. 이들중 9개(도 12의 #1 내지 #9 참조)를 무작위로 선택하여 Fw1(5'-cactactacatccatagcagcaaacatgg-3'(서열 번호 62)) 및 R3(5'-gggtcaagaagcacatgccagagtctcg -3'(서열 번호 63))을 프라이머로 사용하는 PCR 선별을 수행하였다. 도 13a는 프라이머 어닐링 부위뿐만 아니라 야생형 및 TrGA 결실 균주들의 게놈 유전자좌를 나타낸다(PCR 반응에서 사용시 화살표는 중합의 방향을 나타낸다).

[0352] 프라이머 쌍 FW1 및 R3을 갖는 PCR로 9개의 균주를 예비 선별하였다. PCR 생성물을 증폭하기 위한 PCR 조건은 다음과 같다: 단계 1: 95° C에서 5분. 단계 2: 95° C에서 30초. 단계 3: 60° C에서 30초. 단계 4: 68° C에서 3분. 단계 2, 3, 및 4를 추가 29 사이클 동안 반복. 단계 5: 68° C에서 10분. 이상적인 조건에서, 2개의 PCR 단편(1.9-kb 및 5.2-kb)이 PCR 예비 선별에서 예상될 것이다. 이 결과에서 5.2-kb 생성물을 명확하게 관찰할 수 없었다. 그러나 셀루라이터(C1) 포자의 5.1-kb 생성물 및 셀루라이터 + TrGA 결실 카세트(C2) 포자의 PCR 생성물과 비교시, 1.9-kb 단편은 TrGA 결손 카세트가 상동종 재조합에 의한 TrGA 유전자좌(도 13b, 상부 겔)에서 통합되었음을 확실히 확인할 수 있었다.

[0353] 프라이머 쌍 FW1/R1, F4/R3 및 KOF1/KOR2를 사용하여 추가적인 PCR 확인을 각각 수행하기 위해 9개의 균주 중 2개(#1 및 #3)를 선택했다. C2(셀루라이터 포자 + TrGA 결실 카세트)의 결과와 비교하여 #1 및 #3의 포자로부터 2.0-kb 및 2.2-kb 단편이 수득되었을 때(도 13b, 좌측 하단 겔). 프라이머 쌍 FW1/R1 및 F4/R3을 사용한 PCR로 상동성 재조합에 의해 TrGA 유전자좌에 통합된 TrGA 결실 카세트를 추가로 확인하였다. 2.1 kb의 생성물이 셀루라이터(C1)로부터 증폭되는 동안 PCR 생성물이 #1 및 #3의 포자로부터 수득되지 않았을 때, 또 다른 한 쌍의 프라이머인 KOF1 및 KOR2를 사용한 PCR도 결과를 확인하였다(도 13b, 우측 하단 겔, 분자량 마커를 포함하여 처음 4 레인).

[0354] 프라이머 TrGAF2(5'gactgtctccaccatgtaatttttc3' (서열 번호 64)) 및 R3(서열 번호 63)을 사용하여 TrGA 유전자좌의 전체 영역에 대한 #1 및 #3 포자로부터 PCR 생성물을 수득하였고, 이들의 DNA 서열이 결정되었다. 서열 분석 결과는 TrGA 유전자가 *pyr2* 발현 카세트에 의해 대체되었음을 보여주었다(도 13b, 우측 하단 겔, 마지막 2 레인의 PCR 결과 참조; 서열분석 결과에 대한 데이터는 나타내지 않음).

[0355] 위의 결과는 SpyCas9 니카아제와 sgRNA가 사상균류에서 상동종 재조합을 촉진하여 사상균류 T. 레에세이에 상동성 재조합에 기초한 유전자 결실을 허용할 수 있다는 것을 나타낸다. 구체적으로, 90개의 형질전환체 중 47개(약 52%)는 보겔의 전분 플레이트에서 성장 지연된 표현형을 보였는데, 이는 TrGA 유전자가 파괴되었음을 나타낸다. 의도한 상동성 재조합 사건은, 결손 카세트가 상동성 재조합에 의해 숙주 세포에서 TrGA 유전자좌로 성공적으로 혼입되고 결과적으로 TrGA 유전자가 *pyr2* 발현 카세트로 대체되었다는 것을 확인하는 PCR 생성물 서열분석에 의해 입증되었다. 우리의 데이터는 기능적 SpyCas9 니카아제 복합체를 공여체 DNA (즉, 관심 게놈 유전자좌에서 상동성 재조합될 DNA) 이외에 표적 진핵 세포에 직접 도입할 때, 균류 내 상동성 재조합률이 현저히 증가될 수 있다는 것을 나타낸다.

[0356] **섹션 C: 단일 발현 벡터 위에 Cas/가이드 RNA 및 공여체 DNA의 도입**

[0357] 실시예 14: 유전자 편집을 위해 유도된 상동성 공여체 단편을 포함하는 cas9 발현 플라스미드 사용에 의한 *glal* 유전자에서 단일 염기의 결실

[0358] 이 실시예에서, 단일 염기 결실을 함유하는 상동성 *glal* 공여체 단편을 플라스미드 pTrex2gHyg MoCas GTrGA TS11B(실시예 8에 기재)에 혼입시킴으로써 유전자 파괴를 유도하였다. 약 1 kb의 합성 DNA 단편 Glalrep(서열 번호 39, 도 14에서 "Tr-gla 1 kb 상동성 단편"으로 표지됨)을 플라스미드 pTrex2gHygMoCasGTrGATS11B의 유일한 EcoRV 제한 부위에 삽입함으로써 플라스미드 pTrexMoCasGATS11-HDR(도 14)를 생성하였다. EcoRV 부위는 하이그로마이신 내성 마커 (*hph*)의 pSL1180 서열과 N. 크라사(*crassa*) *cpc1* 프로모터 사이의 폴리링커 서열에 존재한다. Glalrep를 증폭시키고 말단 상동성 테일을 첨가하여 pTrex2gHygMoCasGTrGATS11B에서 EcoRV 부위의 양측을 증첩하여 EcoRV-절단 pTrex2gHygMoCasGTrGATS11B 플라스미드로 단편의 길은 어셈블리를 용이하게 하기 위해(매사추세츠, 베버리, 뉴잉글랜드 바이오랩, NEB Gibson Assembly Master Mix를 사용하여), 프라이머 1556F(5'-ATGCGCAAATTTAAAGCGCTGATgtgtgtctaatgcctccaccac[서열 번호 73]) 및 1557R(5'-ATATGGATCTGCGCGCATGCATgatcgtgctagcgtgctgttg[서열 번호: 74])를 사용하였다.

[0359] pTrexMoCasGATS11-HDR 플라스미드를 사용하여 원형질체 및 PEG 매개된 DNA 흡수를 이용하여 정상적인 NHEJ 메카

니즘을 갖는 균주인 T. 레에세이 균주 RL-P37을 형질전환시켰다. 이 균주는 이 실험을 위해 부수적이지만 우리 단(2 g/L)이 모든 성장 배지에 포함되는 것을 필요로 하는 *pyr2*에 대해 영양요구성이었다.

- [0360] 보겔의 최소 배지 + 1.2M 소르비톨 + 75 ppm 하이그로마이신의 선택성 배지 한천 플레이트에서 형질전환체를 초기에 성장시켰다. 성장 후, 안정형 및 불안정한 표현형 모두의 80 가지 형질전환체를 보겔의 전분 배지의 제2 한천 플레이트로 옮겼다. 제2 플레이트에서 성장 후, *gla1* 유전자가 파괴되었는지를 결정하기 위해 전분 제거 구역의 존재 또는 부재를 리포터로 사용하였다. 형질전환체를 제2 플레이트에서 보겔의 최소 배지의 비선택적 제3 플레이트로 옮겼다. 성장 며칠 후, 형질전환체를 보겔의 최소 배지의 제3 플레이트에서 보겔의 최소 배지 + 75 ppm의 하이그로마이신 B의 제4 플레이트로 옮겼다. 제2 및 제3 플레이트의 비선택성 배지에서 성장한 후 형질전환체가 하이그로마이신-함유 배지 상에서 성장할 수 있는 능력을 유지하는지 여부를 결정하기 위해, 이러한 제4 플레이트 상에서 성장에 대해 형질전환체를 평가할 수 있었다.
- [0361] 보겔의 전분 배지의 제거 구역에서 생산된 80개의 형질전환체 중 단지 11개만이 *gla1* 파괴를 나타내며, 보겔의 최소 배지 + 75 ppm 하이그로마이신의 제4 하이그로마이신 선택성 플레이트에서는 성장하지 않는다는 것이 밝혀졌다. 이 형질전환체들은 아마도 *gla1* 유전자의 변이를 가지고 있었고 플라스미드에서 이들의 게놈으로 하이그로마이신 마커를 혼입하지 못했다.
- [0362] 제3 보겔의 최소 배지 플레이트의 균사체에서 이들 11개의 *gla1*- 형질전환체로부터 게놈 DNA를 단리하였다. 프라이머 *glaK*(서열 번호 31) 및 *glaH*(서열 번호 29)를 사용하여 게놈 DNA로부터 *gla1* 유전자를 PCR 증폭하여 TS11 유전자좌의 상태를 결정하였다. 프라이머 1538F 5'-CCACCACAGGAACCAACC(서열 번호 75), 1539R 5'-CTGCGACGGAGGAATGACG(서열 번호 76), 1540F 5'-GGGCGACTGGCAAGGATGT(서열 번호 77) 및 1541R 5'-GCCGTACGCCAGGAACAAG(서열 번호 78)를 사용하여 PCR 생성물을 단리 및 서열분석하였다.
- [0363] 서열 분석 결과, 이들 11개의 *gla1*- 형질전환체 중 8개가 *Gla1rep* 서열의 단일 염기 결실을 포함하는 것으로 밝혀졌다. 2개의 *gla1*-형질전환체는 9 및 100 염기의 TS11 유전자좌에서 *gla1* 결실을 포함하였다. 1개의 *gla1*-형질전환체는 TS11에서 470 염기의 삽입을 포함하였다.
- [0364] 공여체 *Gla1rep* 단편을 플라스미드 pTrex2gHyg MoCas TrGA TS11에 혼입하여 유전자 편집을 유도한 단일 플라스미드를 생성하였다. 하이그로마이신-불안정 형질전환체의 높은 빈도는 공여체 *Gla1rep* 단편 내에서 조작된 *gla1* 유전자에서 단일 염기 결실을 포함하였다. 불안정한 형질전환체가 제1 선택적 한천 배지인 보겔의 최소 배지 + 75 ppm 하이그로마이신 B에서 생성되었으며, 이후 변형된 *cas9* 플라스미드를 잃어 버렸다. 이러한 균주는 이들이 유전자 편집 플라스미드를 게놈에 NHEJ 혼입시키는 불편함을 갖고 있지 않기 때문에, 보다 유리하다.
- [0365] 데이터가 여기에 도시되지 않은 또 다른 실험에서, pTrex2gHyg MoCas gTrGA TS11B에 삽입된 결실 대신에 치환 뉴클레오티드를 포함하는 *gla 1* 상동성 DNA 단편을 사용하여, 본원에 기술된 것과 동일한 상동성 통합 방법을 사용하여 *gla1*의 치환 돌연변이를 T. 레에세이 균주 P37에 성공적으로 도입하였다.
- [0366] 이러한 결과는 *cas9* 및 표적 부위 가이드 RNA와 함께 상동성 단편을 혼입하는 단일 *cas9* 편집 플라스미드를 생성하는 유용성을 나타낸다. 표적 유전자좌에서 특정 *cas9* 표적 절단 및 유전자 편집 상동성 재조합 모두를 유도하는 하나의 플라스미드 벡터가 생성된다. 불안정한 하이그로마이신 저항성 형질전환체를 선별함으로써, 목표 유전자좌로 플라스미드 DNA의 NHEJ 삽입의 발생이 최소화된다. 또한, 본 실시예에서 불안정한 하이그로마이신 내성 형질전환체를 선별함으로써 상동성 재조합에 의해 상동성 공여체 단편을 표적 유전자좌로 혼입시킨 형질전환체를 발견할 수 있다. 그러나, 이러한 형질전환체는 표적화된 재조합을 유도한 플라스미드를 편리하게 잃어 버렸다.
- [0367] 전문적인 조성물 및 방법이 이해의 명료성을 위해 예시 및 실시예로써 일부 상세히 기술되었지만, 본원의 교시에 비추어 첨부된 청구범위의 사상 또는 범위를 벗어나지 않으면서 특정 변경 및 변형이 이루어질 수 있다는 것이 당업자에게는 쉽게 명백하다.
- [0368] 따라서, 선행 기술은 단지 본 발명의 조성물 및 방법의 원리를 설명한다. 당업자는 본원에 명시적으로 기재되거나 도시되지는 않았지만 본 조성물 및 방법의 원리를 구현하고 이의 사상 및 범위 내에 포함되는 다양한 장치를 고안할 수 있음을 이해할 것이다. 또한, 본원에 인용된 모든 실시예 및 조건부 언어는 주로 본 발명의 조성물 및 방법의 원리 및 발명자가 당해 기술을 발전시키는 데 기여한 개념을 독자가 이해하도록 돕는 것을 목적으로 하며, 이렇게 구체적으로 열거된 실시예와 조건에 제한하지 않는 것으로 해석되어야 한다. 또한, 본 조성물 및 방법의 원리, 양태 및 실시예뿐만 아니라 이의 특정 실시예를 기재한 본원의 모든 기재 내용은 이의 구조적 및 기능적 등가물 모두를 포함하는 것으로 의도된다. 또한, 이러한 등가물은 현재 알려진 등가물 및 미래에 개발되

는 등가물, 즉 구조와 상관없이 동일한 기능을 수행하는 임의의 개발된 요소를 모두 포함하는 것으로 의도된다. 따라서, 본 조성물 및 방법의 범위는 본원에 도시되고 기재된 예시적인 실시예에 한정되는 것으로 의도되지 않는다.

[0369] 서열:

[0370] 서열 번호 1:

[0371] 추정 *T. 레에세이* U6 유전자

[0372] AAAAACTAGTAAAGTACTTACTTATTATGATTATTAAGTACTTTAGCTAACTTCTGCAGTACTACCTAAGAGGCTAGGGGTAGTTTTATAGCAGACTTATAGCTATTATTTTTATTAGTAAAGTGCTTTTAAAGTAAGGCTTTTTTATAGCACTTTTTATTATTATAATATATATATAATAAATTTAAGCCTGGAATAGTAAAGAGGCTTATATAATAATTTATAGTAATAAAAGCTTAGCAGCTGTAATATAATTCCTAAAGAAACAGCATGAAATGGTATTATGTAAGAGCTATAGCTTAAAGGCACTCTGCTGGATAAAAATAGTGGCTATAAGTCTGCTGCAAACTACCCCAACCTCGTAGGTATATAAGTACTGTTTGTAGGTAGTCTATCGCCTTCGGGCATTTGGTCAATTTATAACGATACAGGTTTCGTTTCGGCTTTCTCCTCGGAACCCAGAGGTCATCAGTTCGAATCGCTAACAGGTCAACAGAGAAGATTAGCATGGCCCTGCACTAAGGATGACACGCTCACTCAAAGAGAAGCTAAACATTTTTTCTCTCCAAGTCGTGATGGTTATCTTTTTGCTTAGAGAATCTATTCTGTGGACGATTAGTATTGGTAAATCCCTGCTGCACATTGCGGCGGATGGTCTCAACGGCATAATACCCCATTCGTGATGCAGCGGTGATCTTCAATATGTAGTGAATACGTTGCATACACCACCAGGTTCCGGTGCCTCCTGTATGTACAGTACTGTAGTTCGACTCCTCCGCGCAGGTGAAACGATTCCCTAGTGGCAGGATTTTGGCGGGTCAAGAA

[0373] 서열 번호 2:

[0374] sgRNA의 서열(N은 표적 부위에 상보적인 서열)

[0375] NNNNNNNNNNNNNNNNNNGUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUGC

[0376] 서열 번호 3:

[0377] sgRNA: gAd3A TS1

[0378] guccucgagcaaaaguguccGUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUGC

[0379] 서열 번호 4:

[0380] sgRNA: gTrGA TS2

[0381] guucagugcaauagggcugucGUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUGC

[0382] 서열 번호 5:

[0383] sgRNA: gTrGA TS11

[0384] gccaauggcgacggcagcacGUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUGC

[0385] 서열 번호 6:

[0386] sgRNA: gPyr2 TS6

[0387] gcacagcgggaugcccuuguGUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUGC

[0388] 서열 번호 7:

[0389] 코돈 최적화된 *스트렙토코쿠스 피오제네스* Cas9-인코딩 유전자; N- 및 C-말단 NLS 서열을 가짐

[0390] acgcatcgacctgagccagctggcggagacaagaagaagaagctcaagctctag

[0391] 서열 번호 8:

[0392] 서열 번호 7에 의해 인코딩된 *스트렙토코쿠스 피오제네스* Cas9; N- 및 C-말단 NLS 서열을 가짐

[0393] DVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI VWDKGRDFATVRKVL SMPQVNI VVKTEVQTGGFSKESILPKRNSDKL IARKKDWDPKYYGFDSP TVAVSVLVVAKVEKSKLKVSKELLGITIMERSFEKNPIDFLEAKGYKEVKKDLI IKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFVLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANL DKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGDKKKKLLK

[0394] 서열 번호 9:

- [0395] 합성 DNA: gAd3A TS1-1(사카로마이세스 세레비지에 snr52 프로모터 및 S. 세레비지에 sup4 종결인자를 갖는 gAd3A TS1 sgRNA(서열 번호 3))
- [0396] gaattcggatccTCTTTGAAAAGATAATGTATGATTATGCTTTCCTCATATTTATACAGAACTTGATGTTTTCTTTTCGAGTATATACAAGGTGATTACATGTACGTTTGAAGTACAACCTCTAGATTTTGTAGTGCCTCTTGGGCTAGCGGTAAAGGTGCGCATTTTTTTCACACCCTACAATGTTCTGTTCAAAAGATTTGGTCAAACGCTGTAGAAGTAAAGTTGGTGCATGTTTCGGCGTTCGAACTTCTCCGAGTGAAAGATAAATGATCgtcctcgagcaaaaggtgccGTTTTAGAGCTAGAAAATAGCAAGTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGGTGTCTTTTTTTGTTTTTTATGTCTgaattcggatccatcc
- [0397] 서열 번호 10:
- [0398] 합성 DNA: gAd3A TS1-2(T. 레에세이 U6 프로모터 및 종결인자를 갖는 gAd3A TS1 sgRNA(서열 번호 3))
- [0399] gaattcggatccAAAAAACAAGTAAAGTACTTACTTATGTATTATTAAGTACTTTAGCTAACTTCTGCAGTACTACCTAAGAGGCTAGGGGTAGTTTTATAGCAGACTTATAGCTATTATTTTTATTTAGTAAAGTGCTTTTAAAGTAAGGCTTTTTTATAGCAGCTTTTTATTATTATAATATATATATAATAATTTTAAAGCTGGAATAGTAAAGAGGCTTATATAATAATTTATAGTAATAAAAAGCTTAGCAGCTGTAATATAATTCCTAAAGAAACAGCATGAAATGGTATTATGTAAGAGCTATAGTCTAAAGGCACTCTGCTGGATAAAAAATAGTGGCTATAAGTCTGCTGCAAACTACCCCCAACCTCGTAGGTATATAAGTACTGTTTGATGGTAGTCTATCgtcctcgagcaaaaggtgccGTTTTAGAGCTAGAAAATAGCAAGTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGGTGTCTTTTTTTCTCTTgaattcggatcc
- [0400] 서열 번호 11:
- [0401] 합성 DNA: gAd3A TS1-3(T. 레에세이 U6 프로모터, 종결인자 및 인트론을 갖는 gAd3A TS1 sgRNA(서열 번호 3))
- [0402] gaattcggatccAAAAAACAAGTAAAGTACTTACTTATGTATTATTAAGTACTTTAGCTAACTTCTGCAGTACTACCTAAGAGGCTAGGGGTAGTTTTATAGCAGACTTATAGCTATTATTTTTATTTAGTAAAGTGCTTTTAAAGTAAGGCTTTTTTATAGCAGCTTTTTATTATTATAATATATATATAATAATTTTAAAGCTGGAATAGTAAAGAGGCTTATATAATAATTTATAGTAATAAAAAGCTTAGCAGCTGTAATATAATTCCTAAAGAAACAGCATGAAATGGTATTATGTAAGAGCTATAGTCTAAAGGCACTCTGCTGGATAAAAAATAGTGGCTATAAGTCTGCTGCAAACTACCCCCAACCTCGTAGGTATATAAGTACTGTTTGATGGTAGTCTATCgtcctcgagcaaaaggtgccGTTTTAGAGCTAGAGTTCGTTTCGGCTTTTCTCGGAACCCAGAGTTCATCAGTTCGAATCGCTAACAGAATAGCAAGTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGGTGTCTTTTTTTCTCTTgaattcggatcc
- [0403] 서열 번호 12:
- [0404] 보다 짧은 T. 레에세이 U6 프로모터 영역을 갖는 가이드 RNA 발현 카세트를 합성 DNA로서 수득 하였다. TS11에 서의 T. 레에세이 gla1 유전자를 표적으로 하는 sgRNA에 대한 서열을 포함하는 예가 여기에 제공된다.
- [0405] AATTCCTAAAGAAACAGCATGAAATGGTATTATGTAAGAGCTATAGTCTAAAGGCACTCTGCTGGATAAAAAATAGTGGCTATAAGTCTGCTGCAAACTACCCCAACTCGTAGGTATATAAGTACTGTTTGATGGTAGTCTATCgccaatggcgacggcagcacGTTTTAGAGCTAGAGTTCGTTTCGGCTTTTCTCGGAAACCCAGAGGTCATCAGTTCGAATCGCTAACAGAATAGCAAGTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGGTGTCTTTTTTCTCTT
- [0406] 서열 번호 13:
- [0407] 프라이머: gRNA fwd af111
- [0408] cgtcagcttaagAATTCCTAAAGAAACAGCATGAAATGG
- [0409] 서열 번호 14:
- [0410] 프라이머: gRNA rev sf11
- [0411] cgtcaggccacgtgggccAAGAGAAAAAAGCACCACCGACTCGG
- [0412] 서열 번호 15:
- [0413] 프라이머: Ad3 5' fwd
- [0414] tgaacacagccaccgacatcagc
- [0415] 서열 번호 16:
- [0416] 프라이머: Ad3 5' rev

[0417] gctggtgagggtttgtgctattg  
 [0418] 서열 번호 17:  
 [0419] 프라이머: Ad3a 5005 rev  
 [0420] gattgcttgggaggagacat  
 [0421] 서열 번호 18:  
 [0422] 프라이머: Ad3 3' fwd  
 [0423] cgaggccactgatgaagtgttc  
 [0424] 서열 번호 19:  
 [0425] 프라이머: Ad3 3' rev  
 [0426] Cagttttccaaggctgccaacgc  
 [0427] 서열 번호 20:  
 [0428] 프라이머: Ad3a 5003 fwd  
 [0429] ctgatcttgaccctggaatc  
 [0430] 서열 번호 21:  
 [0431] Ad3mid rev  
 [0432] ctctctatcatttgccaccctcc  
 [0433] 서열 번호 22:  
 [0434] 프라이머: Adfrag fwd  
 [0435] ctccattcacctcaattctcc  
 [0436] 서열 번호 23:  
 [0437] 프라이머: Adfrag rev  
 [0438] gttcccttgccggtgcttggatc  
 [0439] 서열 번호 24:  
 [0440] 프라이머: Ad3a 2k fwd  
 [0441] caatagcacaaccctcaccagc  
 [0442] 서열 번호 25:  
 [0443] Ad3a 2k rev  
 [0444] gaacaacttcatcagtggcctcg  
 [0445] 서열 번호 26:  
 [0446] 프라이머: glaA  
 [0447] ccgttagttgaagatccttgccg  
 [0448] 서열 번호 27:  
 [0449] 프라이머: glaB  
 [0450] gtcgaggatttgcttcatacctc  
 [0451] 서열 번호 28:  
 [0452] 프라이머: glaD

- [0453] gagagacgcaggatgactcaaag
- [0454] 서열 번호 29:
- [0455] 프라이머: glaH
- [0456] tgccgtgggtcattggcatattc
- [0457] 서열 번호 30:
- [0458] 프라이머: glaJ
- [0459] tgccgactttgtccagtgattcg
- [0460] 서열 번호 31:
- [0461] 프라이머: glaK
- [0462] ttacatgtggacgcgagatagcg
- [0463] 서열 번호 32:
- [0464] 프라이머: glalrepF
- [0465] gtgtgtctaatagcctccaccac
- [0466] 서열 번호 33:
- [0467] 프라이머: glalrepR
- [0468] gatcgtgctagcgctgctgttg
- [0469] 서열 번호 34:
- [0470] 프라이머: 1553R
- [0471] CCGTGATGGAGCCCGTCTTCT
- [0472] 서열 번호 35:
- [0473] 프라이머: 1555F
- [0474] CGCGGTGAGTTCAGGCTTTTTC
- [0475] 서열 번호 36:
- [0476] 프라이머: pyr2F
- [0477] gtataagagcaggaggaggag
- [0478] 서열 번호 37:
- [0479] 프라이머: pyr2R
- [0480] gaacgcctcaatcagtcagtcg
- [0481] 서열 번호 38
- [0482] 트리코테르마 레에세이 텔로미어 서열간 박테리아 카나마이신 저항 유전자(프로모터 및 종결 인자를 가짐)
- [0483] atagaccagctttacataataatcgctgttgctactgactgatgaccttcttcctaaaccagtttctaat taccactgcagtgaggataaacctaacctcg  
ctctggggttattatataactgat tagcaggtggcttatatagtgtgaagtactataagagtttctgcgaggagggtggaaggactataaactggacaca  
gttagggatagagtgatgacaagacctgaatgttatcctccggtgtggtatagcgaattggctgaccttgcagatggtaatggttttaggcagggttttgca  
gagggggacgagaacgcgttctgcgat ttaacggctgctgccgccaagctttacggttctctaatgggcggccgc
- [0484] 서열 번호 39:
- [0485] 합성 DNA: Glalrep
- [0486] gtgtgtctaatagcctccaccacaggaaccaaacggctttgacctctgggaagaagtcaatgggagctcattctttactgttgccaaccagcaccgaggtat

gaagcaaatcctcgacattcgctgctactgcacatgagcatgttactgaccagctctacagcacttgtcgagggcgccactcttgcctgccactcttggcca  
 gtcgggaagcgcttattcatctgttgcctcccaggttttgccttctccaacgattctgggtgtcgctctggaggatactgcactccaacagatgtcttt  
 tcaactgtttatagagattggccaatactgatagctcgctctagtcaacaccaacgagggcaggactggcaaggatgtcaactccgtcctgacttccatcc  
 acaccttcgatcccaaccttggctgtgacgcaggcaccttccagccatgcagtgacaagcgctctccaacctcaaggttgttgcctgactccttccgctcca  
 tctacggcgtgaacaaggcattcctgCGgtgctgccgtcgccattggccggatgacagaggatgtgactacaacggcaaccttggatcttgcctacatt  
 tgctgctgccgagcagctgtacgatgccatctacgtctggaagaagacgggctccatcacggtgaccgccacctccctggccttcttccaggagcttgttcc  
 tggcgtgacggccgggacctactccagcagctcttcgaccttaccacatcatcaacgccgtctcgacatagccgatggcttctcagcgagctgcca  
 gtacgtccccgccgaggttcgctggccgagcagttgaccgcaacagcggcactccgctgtctgccttccactgacgtggtcgtacgcctcgttcttgac  
 agccacggcccgctcgggctggcatcgtgccccctcgtgggccaacagcagcgcctagcacgac

[0487]

서열 번호 40:

[0488]

전체 U6 유전자 프로모터 서열(전사 개시 부위를 포함하지 않음)

[0489]

AAAAAACAAGTAACTTACTTATGTATTATTAAGTACTTTAGCTAACTTCTGCAGTACTACCTAAGAGGCTAGGGGTAGTTTTATAGCAGACTTATAG  
 CTATTATTTTTATTAGTAAAGTGTCTTTTAAAGTAAAGTCTTTTTTATAGCACTTTTTATTATTATAATATATATAATAAATTTTAAAGCCTGGAATA  
 GTAAAGAGGCTTATATAATAATTTATAGTAATAAAAGCTTAGCAGCTGTAATATAATTCCTAAAGAAACAGCATGAAATGGTATTATGTAAGAGCTATAGTC  
 TAAAGCACTCTGCTGGATAAAAATAGTGGCTATAAGTCTGCTGCAAACTACCCCCAACCTCGTAGGTATATAAGTACTGTTTGATGGTAGTCTATC

[0490]

서열 번호 41:

[0491]

절단된/더 짧은 U6 프로모터 서열(전사 시작 부위를 포함하지 않음)

[0492]

AATTCCTAAAGAAACAGCATGAAATGGTATTATGTAAGAGCTATAGTCTAAAGCACTCTGCTGGATAAAAATAGTGGCTATAAGTCTGCTGCAAACTACC  
 CCCAACCTCGTAGGTATATAAGTACTGTTTGATGGTAGTCTATC

[0493]

서열 번호 42:

[0494]

U6 유전자 인트론

[0495]

GTTCGTTTCGGCTTTTCTCGGAACCCAGAGGTCATCAGTTCGAATCGCTAA

[0496]

CAG

[0497]

서열 번호 43:

[0498]

U6 유전자 전사 종결인자 서열

[0499]

TTTTTTTCTCTT

[0500]

서열 번호 44:

[0501]

사상균 세포 코돈에 최적화된 스트렙토코쿠스 파이오제네스 Cas9-인코딩 유전자; NLS 없음

[0502]

agac

[0503]

서열 번호 45:

[0504]

서열 번호 44에 의해 인코딩된 스트렙토코쿠스 파이오제네스 Cas9; NLS 없음

[0505]

EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNI VVKTEVQTGGFSKESILPKRNSDKLI ARKKDWD  
 PKKYGGFDSPTVAYSVLVAKVEKGSKKLKSVELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLI IKLPKYSLFELENGRKRMLASAGELQKGNELALP  
 SKYVNFYLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEI IEQISEFSKRIVILADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTI  
 DRKRYTSTKEVLDTL IHQSITGLYETRIDLSQLGGD

[0506]

서열 번호 46:

[0507]

SV40 NLS

[0508]

PKKKRKY

[0509]

서열 번호 47

[0510]

*T. 레예세이 blr2* (청색광 조절인자 2) 유전자 NLS

- [0511] KKKKLLK
- [0512] 서열 번호 48:
- [0513] 스트렙토코쿠스 더모필루스 LMD-9 Cas9
- [0514] FRERKSATEKVYFYNSIMNIFKKSISLADGRVIERPLIEVNEETGESVWNKESDLATVRRVLSYPQVNVVKKVVEEQNHGLDRGKPKGLFNANLSSKPKPNSN  
ENLVGAKEYLDPKKYGGYAGISNSFTVLVKGTIEKGAKKKITNVLEFQGISILDRINRDKLNFLEKGYKDIELIIELPKYSLFELSDGSRRMLASILST  
NNKRGEIHKGNQIFLSQKFVKLLYHAKRISNTINENHRKYVENHKKFEELFYIIEFNENYVGAKKNGKLLNSAFQSWQNHSIDELCSSF IGPTGSEKGL  
FELTSRGSAADEFELGVKIPRYRDYTPSSLLKDATLIHQSVTGLYETRIDLAKLGE
- [0515] 서열 번호 49:
- [0516] 스트렙토코쿠스 뮤탄스 UA159 Cas9
- [0517] AKKFFYSNIMNFFKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQVNI VKKVVEEQGGFSKESILPKGNSDKLIPRKTCKFYWDTKKYGDFDSPIVAYSILV  
IADIEKGSKKLKTVKALVGVTIMEKMTFERDPVAFLEKGRYRNVQEENIKLPKYSLFLENGRKRLLASARELQKNEIVLPNHLGTLTYHAKNIHKVDE  
PKHLDYVDKHKDEFKELLDVVSNSKTYLAEGNLEKIKELYAQNGEDLKEASSFINLLTFTAIGAPATFKFFDKNIDRKRYTSTTEILNATLIHQSI  
LYETRIDLNKLG
- [0518] 서열 번호 50:
- [0519] 캄피로박테르 제주니 Cas9
- [0520] MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSARKRLARRKARLNHLKHLIANEFKLNIEDYQSFDES  
ELRFRALNELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYSVGEYLKEYFQKFKENSKEFTNVRNKKESYERCI  
AQSFKDEL  
KLIFKKQREFGFSFKFEEVLSVAFYKRALKDFSHLVGNCSSFTDEKRAPKNSPLAFMFVALTRIIINLLNKNTEGILYTKDDL  
NALLNEVLKNGTLTY  
KQTKLLGLSDDYEFKGEKGYFIEFKYKEFIKALGEHNSQDDLNEIAKDIITLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLN  
ISFKALKLVTPLM  
LEGKKYDEACNELNLKVAI  
NEDKDFLPAFNETYKDEVNPPVLRRAIKYRVLNALLKYGKVKHINI  
ELAREVGNHSQRAKIEKEQENYKAKKDAEL  
ECEKGLKINSKILKRLFKQKFCAYSGEKIKISDLQDEKMLEIDHIYPYRSFDDSYMKNVLFVTKQNEKLNQTPFEAFGND  
SAKWQKIEVLAKNLP  
TKKQKRILDKNYKDEKQNFKDRNLNDRYIARLVNLYTKDYLDLPLSDDENTKLN  
DTQKQSKVHVEAKSGMLTSALRHTWGFSAKDRNHLHHAIDAVII  
AYANNSIVKAFSDFKKEQESNSAELIYAKKISELDYKNRKFPEFSGFRQVLDKIDEIFVSKPERKKPSGALHEETFRKEEFYQSYGGKEGVLKALEL  
GKIRKVNKIVKNGDMFRVDIFKHKKTNFYAVPIYTMDFALKVLPNKAVARSKKEIKDWILMDENYEFCSLYKDSLILIQTKDMQEP  
EYVYNAFTSSTVS  
LIVSKHDNKFETLSKNQILFKNANEKEVIAKSIGIQNLKVFKEYIVSALGEVTKAEFRQREDFKK
- [0521] 서열 번호 51:
- [0522] 나이세리아 메닝기티데스 Cas9
- [0523] GTGNINIRIHDLHDKIGKNGILEGIGVKTALSFQKYQIDELGKEIRPCRLKKRPPVR
- [0524] 서열 번호 52:
- [0525] 프란시셀라 투라렌시스 아종 노비시다 Cas9
- [0526] EKKIADTIWDANKDFKFGNYRSFINLTPQEQKAFRHALFLADENPIKQAVIRAINNRNRTFVNGTQRYFAEVLANNIYLRAKKENLNTDKISFDYFGIPTI  
GNGRGAIEIRQLYEKVDSDIQAYAKGDKPQASYSHLIDAMLAFCAADEHRNDGSGLEIDKNYSYPLDKNTGEVFTKIDFSQIKITDNEFS  
DKKLVKKA  
IEGFNTHRQMTDRGIIYAENYLPILIHKELNEVRKGYTWKNSEEIKIFKGGKYDIQQLNLLVYCLKFVDPKIPISIDIQISTLEELRNIL  
TTNIAATAEYIIIN  
LKTQKLHEYYIENYNTALGYKYSKEMEFLRSLAYRSERVKIKSIDDVKQVLDKDSNFIGKITLPPFKKEWQRLYREWQNTTIKDDY  
EFLKSFFNVKSI  
TKLHKKVVRKDFSLPISTNEGKFLVKKRTWNNFIYQILNDSRSDGATKPFIPAFDISKNEIVEAIDSFTSKNIFWLPKNIELQKVDNKNIFA  
IDTSKWFVET  
PSDLRDIIGIATIQYIDNNSRPKVRVLDYVIDDDSKINYFMNHSLLKSRYPDKVLEILKQSTIEFESSGFNKTIKEMLGMLAGIYNETSNN
- [0527] 서열 번호 53:
- [0528] 파스튜렐라 물토시다 Cas9
- [0529] VKLALSFEKYQVDELGKNRQICRPQQRQ
- [0530] PVR
- [0531] 서열 번호 54
- [0532] 플라스미드 pET30a-Cas9-D10A 니카아제 내 합성된 Cas9-D10A 니카아제 유전자의 뉴클레오티드 서열. N-말단

His6 태그를 인코딩하는 올리고뉴클레오티드는 밑줄이 쳐져 이탤릭체로 표기되고, SV40 핵 위치 신호를 인코딩하는 올리고뉴클레오티드는 이탤릭체 및 회색 하이라이트로 표시되고, BLR 핵 위치 신호를 인코딩하는 올리고뉴클레오티드는 밑줄이 쳐져 회색으로 강조 표시된다.

atg*caccatcatcatcatca*ttcttctggtctggtgccacgcggttctggtatgaaagaaccgctgctgctaaattcgaac  
 gccagcacatggacagcccagatctgggtaccgacgacgacgacaaggccatggcccaaaaaagaaacgcaag  
 gttatggataaaaaatacagcattggtctggccatcggaaccaacagcgttgggtggcagtaataacagatgaataca  
 aagtccgtcaaaaaathtaaggttctggggaatacagatcgccacagcataaaaaagaatctgattggggcattgctg  
 ttgattcgggtgagacagctgaggccacgctgtaaactgacagcaagaagacgttacacacgtcgtaaaaatcgtat

[0533]

ttgctacttacaggaaattttctaacgaaatggccaaggtagatgatagtttctccatcgtctcgaagaatctttctggtga  
 ggaagataaaaaacacgaacgtcacctatctttgcaataatcgtggatgaagtggcctatcatgaaaaatccctacga  
 ttatcatcttcgcaagaagtgggtgatagtagcggacaaagcggatctgcgttaatctatcttgcgttagcgcacatgatca  
 aatttcgtggtcatttctaaggtgatctgaatcctgataactctgatgtggacaaattgtttatacaattagtcaaacct  
 ataatcagctgttcgaggaaaacccattaatgcctctggagttgatgccaagcgttttaagcgcgagactttctaagtc  
 ccggcgtctggagaatctgatcgccagttaccaggggaaaaagaaaaatggctgtttgtaatctgattgccctcagctc  
 gggcttaccocgaactcaaatcaattttgacctggctgaggacgcaaagctgcagctgagcaaaagatacttatgatgat  
 gacctcgacaatctgctgccagattggtgaccaataatgctggatctgttctgagcgaagaatcttccgatgctatctt  
 gctgctggatattctgctgtaataaccgaaatcaccaaagcgcctctgctgcaagtatgatcaagagatacgcgagca  
 ccaccaggacctgactcttctaaggcactggtagcacaacagcttccggagaaatacaaaagaatattctcgaccagt  
 ccaagaatggttacgcggtacatcgatggtggtgcatcacaggaagagttctataaatttataaaccaatcctgaga  
 aaatggatggcacggaagagttactgttaacttaaccgcaagactgcttagaaagcaacgtacattcgacaacgg  
 ctccatcccacaccagattcatttaggtgaacttcacgccaatcttgcgcagacaagaagalltctatcccttctaaaagaca  
 atcgggagaaaaatcagagaagatcctgacgttccgacttccctattatgctggctccctggcagctggtaattctcggttgct  
 ggtgatgcgcgcaaaagtgggaaacctaccccttgaacttgaagaagctggtgataaagggtgctagcgcgcagct  
 ctttatagaaagaatgacgaacttcgataaaaaactgccaacgaaaaagtcctgccaagcactctctttatagatga  
 ctctactgtgtacaacgaactgactaaagtgaatacgttacggaaggtatgcgcaaacctgccttcttagtgccgagcag  
 aaaaaagcaattgtcgtactctttaaaccgaatcgcaaggttaactgtaaaacagctgaaggaagattttcaaaaag  
 atcgaatgctttagtctgagatctcgggtgctgaagatcgtttcaacgcttcttagggacctatcatgattgctgaagat  
 aataaaagacaaagacttctcgacaatgaagaaaatgaagatattcggaggatattgtttgacctgacctattcgaa  
 gatagagagatgatcggagcgcctaaaaacctatgccacctgttgatgacaaagtcataagcaattaaagcgc  
 gcagataatcgggtggggcgcttagcgcgaagtgtgataacggtattgagacaagcagagcggaaaaactatcc  
 tggatttcccaaatctgacggatttgcaaccgcaattttatgacgcttatacatgatgattcgttacattcaaagaggat  
 tcagaaggctcaggtgctggcaaggtgattcactccacgaacataatgcaaaattggcggctctcctcgattaagaa  
 ggggatcctgcaaacagttaaagttggtgataactgttaaagtaatggcgccacaagccggagaatatcgtgata  
 gaaatggcgcgagaaatcaacgacacaaaaaggtaaaagaactcaagagagagaatgaagcgcattgaggga  
 ggggataaaggaacttgatctcaattctgaaagaacatccagttgaaaacactcagctgcaaatgaaaaattgtac  
 ctgtactacctgcagaatggaagagacatgtacgtggatcaggaattggatataatagactctcgactatgacgtgat  
 cacattgtccctcagacttctcaaggatgattctatagataataaagtaacttacgagatcggcaaaaaatcgggtaaat  
 cggataacgtccatcggaggaagctgttaaaaagatgaaaaactattggcgtcaactgctgaacccaagctgatcac  
 acagcgtaaagttgataatctgactaaagccgaacgcggtggtcttagtgaactcgataaagcaggatttataaacggc  
 agttagtagaaacgcgcaaaattacgaaacacgtggctcagatcctcagattctagaatgaatacaaaagtacgatgaaaa  
 cgataaactgatccgtgaagtaaaagtcattacctttaaacttaactgtgtccgattccgcaaaagatttcagtttaca  
 ggtccgggaaatcaataactatcaccatgcacatgatgcataathtaagtcgggttaggcacggcccttataagaatac

[0534]

cctaaactcgaaagtgtgattgtttatggggattataaaagtgtatgacgttcgcaaatgatcgcgaaatcagaacaggaa  
atcggtaaggctaccgctaaatactttttattccaacattatgaatttttaagaccgaaataactctcgcgaatggtgaaat  
ccgtaaacggcctcttatagaaccaatggtgaaacgggagaaatcglttgggataaaggctgacttggccaccgttcg  
taaagtctctcaatgcccaagtaacattgtcaagaagacggaagtcaaacagggggatttccaagaatctatcct  
gccgaagcgtaacagtataaacttattgccagaaaaaagattgggatccaaaaaacggaggcttggattcccct  
accgtcgcgtatagtgtgctggtgtgctaagtcgagaaagggaaaagcaagaattgaaatcagtaaaagaactgc  
tgggtattacaattatggaagatcgtcctttgagaaaaatccgatcgacttttagaggccaaggggtataaggaagta  
aaaaagatctcatcaaataccgaagtatagtcttttgagctggaaaacggcagaaaaagaatgctggcctccg  
ggcgagttacagaagggaaatgagctggcgctcctccaaatatgtaattttctgtacctgcccagtcattatgagaaact  
gaagggcagccccgaagataacgaacagaacaattatcgtggaacagcataagcactatttagatgaaattataga  
gcaaattagtgaaatttctaagcgcgttatcctcgcggatgctaatttagacaaagtactgtcagctataataaacatggg  
ataagccgattagagaacagccgaaaatatcattcattgtttacctaaccaacctggagcaccagctgcctcaaata  
ttcgataccacaattgatcgtaaacgggtatacaagtacaaaagaagcttggacgcaacctcattcatcaatctattactg  
gattatatgagacacgcattgatctttcacagctggcgagacaagaagaaaaaactgaaactg

[0535]

[0536]

서열 번호 55:

[0537]

플라스미드 pET30a-Cas9-D10A 다크아제(서열 번호 54에 의해 인코딩됨)에서 발현된 Cas9-D10A 다크아제 단백질의 아미노산 서열. N-말단 His6 태그는 밑줄과 이탤릭체로 표시되고, 돌연변이된 아미노산은 굵은 밑줄(D10A)로 표시되고, SV40 핵 위치 신호는 이탤릭 및 회색으로 표시되고, BLR 핵 위치 신호는 밑줄 쳐져 회색으로 표시된다.

mhhhhhssglvprgsgmketaaakferqhmdspdlgtdddkamapkkkrkvm<sup>dkysigla</sup>igntnsvgwavitd  
eykvpsskfkvlngntdrhsiknligallfdsgetaetrkrtrrrytrkrnricylqeifsnemakvddsffhrleesflveedk  
kherhpifgnivdevayhekyptiyhrkkldvdstdkadlrliylalahmikfrghfliegdlndnsdvdkflqvlvtynqlfee  
npinasgvdakailsarlsksrrlenliaqlpgekknglfnlialslgtpnfksnfdlaedaklqskdydddlnllaigdq  
yadflaaknlsdailrvnteitkapsasmikrydehhdllkalvrqqlpekykeiffdqskngyagyidggasqee  
fykfiikiplekmdgteellvknredllrkqrtfdngsiphqihlgelhalrrqedfypflkdnrekiekillfripyyvgplargnsrf  
awmtrkseeitpwnfeevvdkgasaqsfiermtnfdknlpnekvlpkhsllieyftvyneltkvkvytegmrkpafisgeq  
kkaivdlflktrnkvtvkqlkedyfkkiecdfsveisgvedrfnaslgythdllkiikdkdflneenediledivltlftfedremie  
erlkyahlfddkvmkqlkrryrtgwgrlsrklngirdkqsgktildfklsdganrnmqlihdslftkediqaqvsgqgdsi  
hehianlagspaikkilqvtkvvdvdelvkmgrhkeniviemarenqtqkgqknsrermkrieegikelgsqilkehpv  
entqlqneklylylqngdrmyvdqeldinrldydvdhivpqsflkddsidnkvltrsdknrgksdnvpseevkkmkny  
wrqlnaklitqrkfdnltaerggseldkagfikrqlvetrqitkhvaqildsrmntkydendklirevkvitlksklvsdfrkdfq  
ykvreinnyhahdaylnavvgtalikkypklesefvygdykvydvrmkiakseigekatakayffysnimnffkteitlan

[0538]

geirkprietngetgeivwdkgrdfatvrkvlsmqvni<sup>kktevtg</sup>ggfskesilpk<sup>rnsdkliarkkdwdp</sup>kkyggsdpt  
vaysvlvvakvekgkskklsvkellgitimerssfeknpidfleakgykevkkdliiklpkyslfelengrkrmlasagelqk  
gnelalpskyvnflylashyeklkgspedneqkqlfveqhkhyldeieiqisefskrviladanldkvl<sup>saynkh</sup>rdkpireq  
aeniihlftltnlgapaafkyfdttidrkyrststkevdatl<sup>ihqsitglyetridlsq</sup>lggdkkkkklk

[0539]

[0540]

서열 번호 56

[0541]

기질 DNA 단편의 뉴클레오티드 서열. UTR 서열은 소문자로 나타나고 TrGA 유전자는 대문자로 나타난다. 두 개의 선택된 VT 도메인인 TrGA\_sgF1과 TrGA\_sgR1은 각각 굵은 글씨와 밑줄로 나타난다.

gactgtctccaccatgtaatTTTCCTGCGACTCCATATAACGCCGGATCGTGAATTTCTTCTTTCTTCTCTCTCAACA  
 acaacggatctgtgctttgCGGTCCTCGGTCACGCGTCAGGGTCGACTGCTGCGACTCGATAACTCCATGGAGCCAT  
 caacttgctatgggtgcaatcatcctatcgacaggccaagaacaagccggcctccggctgcctcattcgctgcaaga  
 cggcttgagtggtggctggaggatcggggccccatattccaaccctTTTTCCAAGCCGTCGGCCGGTGAGGTTGAG  
 gaaaaccatgggtgctacatattatcgatgctgggtttgtagtagcaatgttgcggtggcagttgagccgagcctcgt  
 ctgggctctgaccaggcaacgccatctgactagctgcgccgaaggaaggatgaltcattgtacgacgccagcaatg  
 gaatctcaagtaaaagcccgacgaaccgacctgacagatacagaattcctggctgggtgggttgggtggagactgc  
 ttacggagtcgctgcctcgactgctatggccgctccagccctggactctgccatattatgacacgagtaaagcc  
 tgcagtgatgctgctgtcctatgctgagagaacaacacacctggtgctacataggcaatactacctgtagctcaaa  
 gttgactgtttgcttgatgctttgatcatgccatccatccctgtcttgcagtgcatggtgatctctacgccagcggggag  
 aaagctgtctgataaaagtagcagtgatgctgctggctacggccctttatccccatgctatgcatctctatatta  
 atccaggagactctcctcctggcatgggtgagtagacaagtacgaggacatgtagaagcagagccacgcaacgcttga  
 catctgtacctatTTTGGGCCAAAAATCGAGACCCACCAGCTCCTACCTACATGTGAAGATCTTAGCCACAATCCTAC  
 TTTTACTAGTATTACTGCACAGCTGCATCAGAGTCTCGGTTGCTGTGAAACCCAGCTCAGCTCCTGAGCACATGCAGT  
 AACCCGACTCGGCTCATTTCCACACCCAATTTGGACCTGAGGGATGCTGGAAGCTGCTGAGCAGATCCCGTTACCG  
 ATTCATGGCACTACTACATCCATACGCAGCAAAATGGGCTGGGCTGGCTTCTCAATGCAAAATGCCGCAAAAGTCC  
 CGGCATTGCTGATGCAGAGATGCAGATTTAGCGGGCGATTCTAGGGTAGGGCGACTACTACTAATACCACCTAGTCA  
 GTATGTCTAGCACCGGAGGCTAGGCGGTTAGTGGACGGGAACCTGGTCACTCATCGCAACCAGGATCCCGCACTCGT  
 TGGCTCTGCCCCACGGGGCGGGAGTTGGCAGAGGCAGAATGCGGAGCAGCCCCCTGCTGCCCTGGCCGGGGCCT  
 GTGAAGCAAGCAGACGAGAGCAGAGCGGTTGAGAAGCGGTGGTGACGCTGACGGTACGAAGACGAGCGGAGAAAT  
 CCCGTTAAGCCGAGGCTGGGCTCCCCCCCCGTCATCATGCCCCATCCTGCTCTCCAGCCACTCGTCTCCCTGCCTCGT  
 CGCTCCCCCTCCCCGATTAGCTGCGCATGTTCTCCTGACAGCGTGAATGACGCGTTGCCAGCCATTGCTGAG  
 CGCATCCCGCATCTGAGTCTAGCTGTCACGCTGGCAATCTGGCCAGGCAGAGCAGCAAGACGGCGGGCATGATTG  
 GCGCGTCCCTGGCGGGCATCAGCTGGCCATCCGCTGCCACCCGAGACCGCATCACCGACTGTGCGGATCTCTCCGAGC

[0542]

agcaggaggctgatcctggccggcgagacgattgaaaaggctgccgggcccggagcaggacagcggcgagagc  
 gagcgagagagaggaaaagaagaaggctgactgtcttattttcagccagccccggctcaacagaagcagaggagaa  
 ggcaacgacgtcaacgacgacgacgacgacgaagacggggaagtcggttagtgaagatccttgccgtcaca  
 caccatcctggtgatattgclttcccctgccgttgctgcccactgttccctcttcttcccccttcttctcattccgagcgt  
 actggttctactccgagccttcggtgtgcttcttcttctgaccattgcaccgcccgtcgggactgggccccggag  
 aattcggccctttcgcagcatttggccctcagttcccatggggacgggtccacacttctcttggccctgcagacctttgt  
 cgtcgggtccgagtcggaagaagctcagttgagcgttgagtagcatctacgcgcgaatcactggacaaagtggcaa  
 gacgaagccgtcgtcgcctgctgctgctgttactgacgacagggcctccgactggggcatcggcataataaaaagat  
 gcccgcttcgcatggacctggccatgagccactcggcatcggctctctctcaacgcttctctcacacatcctcctcat  
 tccgcccata**ATGCACGTCCTGTCTGACTGCGGTGCTGCTCGGCTCCGTTGCCGTTCAA**  
**AAGGTCCTGGGAAGACCAGGATCAAGCGGTCTGTCCGACGTCACCAAGAGGTCT**  
**GTTGACGACTTCATCAGCACCGAGACGCCTATTGCACTGAACAATCTTCTTTGCAA**  
**TGTTGGTCCTGATGGATGCCGTGCATTCGGCACATCAGCTGGTGCGGTGATTGCA**  
**TCTCCAGCACAATTGACCCGGACTGTAAGTTGGCCTTGATGAACCATATCATATA**  
**TCGCCGAGAAGTGGACCGCGTGTGAGACTGAGACAGACTATTACATGTGGACG**  
**CGAGATAGCGCTCTTGTCTTCAAGAACCTCATCGACCGCTTACCGAAACGTACG**  
**ATGCGGGCCTGCAGCGCCGCATCGAGCAGTACATTACTGCCAGGTCACTCTCCA**  
**GGGCCTCTCTAACCCCTCGGGCTCCCTCGCGGACGGCTCTGGTCTCGGCGAGCC**  
**CAAGTTTGAGTTGACCCTGAAGCCTTTCACCGGCAACTGGGGTCGACCGCAGCG**  
**GGATGGCCCAGCTCTGCGAGCCATTGCCTTGATTGGATACTCAAAGTGGCTCATC**  
**ACAACAATATCAGTCGACTGTGTCCAACGTCATCTGGCCTATTGTGCGCAACGA**  
**CCTCAACTATGTTGCCAGTACTGGTCAGTGCTTGCTTGCTCTTGAATTACGTCTT**  
**TGCTTGCTGTCTAATGCCTCCACCACAGGAACCAAACCGGCTTTGACCTCTGGG**  
**AAGAAGTCAATGGGAGCTCATTCTTACTGTTGCCAACCAGCACCGAGGTATGAA**  
**GCAAATCCTCGACATTCGCTGCTACTGCACATGAGCATTGTTACTGACCAGCTCTA**  
**CAGCACTTGTCGAGGGCGCCACTCTTGCTGCCACTCTTGCCAGTCGGGAAGCG**  
**CTTATTCATCTGTTGCTCCCCAGGTTTTGTGCTTTCTCCAACGATTCTGGGTGTCG**  
**TCTGGTGGATACGTCTGACTCCAACAGTATGTCTTTTCACTGTTTATATGAGATTGG**  
**CCAATACTGATAGCTCGCCTCTAGTCAACACCAACGAGGGCAGGACTGGCAAGGA**  
**TGTCAACTCCGTCCTGACTTCCATCCACACCTTCGATCCCAACCTTGGCTGTGACG**  
**CAGGCACCTTCCAGCCATGCAGTGACAAAGCGCTCTCCAACCTCAAGGTTGTTGT**  
**CGACTCCTTCCGCTCCATCTACGGCGTGAACAAGGGCATTCTGCCGGTGCTGCC**  
**GTCGCCATTGGCCGGTATGCAGAGGATGTGTACTACAACGGCAACCCTTGGTATC**  
**TTGCTACATTTGCTGCTGCCGAGCAGCTGTACGATGCCATCTACGTCTGGAAGAA**

[0543]

GACGGGCTCCATCACGGTGACCGCCACCTCCCTGGCCTTCTTCCAGGAGCTTGTT  
 CCTGGCGTGACGGCCGGGACCTACTCCAGCAGCTCTTCGACCTTACCAACATCA  
 TCAACGCCGTCTCGACATACGCCGATGGCTTCCTCAGCGAGGCTGCCAAGTACGT  
 CCCCCCGACGGTTCGCTGGCCGAGCAGTTTGACCGCAACAGCGGCACTCCGCT  
 GTCTGCGCTTACCTGACGTGGTTCGTACGCCTCGTTCTTGACAGCCACGGCCCGT  
 CGGGCTGGCATCGTGCCCCCTCGTGGGCCAACAGCAGCGCTAGCACGATCCCC  
 TCGACGTGCTCCGGCGCGTCCGTGGTCCGATCCTACTCGCGTCCCACCGCCACG  
 TCATTCCCTCCGTGCGCAGACGCCAAGCCTGGCGTGCCTTCCGGTACTCCCTACA  
 CGCCCCTGCCCTGCGCGACCCCAACCTCCGTGGCCGTACCTTCCACGAGCTCG  
 TGTCGACACAGTTTGGCCAGACGGTCAAGGTGGCGGGCAACGCCGCGGCCCTGG  
 GCAACTGGAGCACGAGCGCCGCGTGGCTCTGGACGCCGTCAACTATGCCGATA  
 ACCACCCCCTGTGGATTGGGACGGTCAACCTCGAGGCTGGAGACGTCGTGGAGT  
 ACAAGTACATCAATGTGGGCCAAGATGGCTCCGTGACCTGGGAGAGTGATCCCAA  
 CCACACTTACACGGTTCCTGCGGTGGCTTGTGTGACGCAGGTTGTCAAGGAGGAC  
 ACctggcagtcgtaatgaatcggcaaggggtagtactagtagactttagtctgcc

[0544]

[0545] 서열 번호 57:

[0546] 서열 번호 56을 위한 정방향 프라이머:

[0547] 5'- gactgtctccaccatgtaatttttc-3'

[0548] 서열 번호 58:

[0549] TrGA 프라이머 KOR2 ; 서열 번호 56을 위한 역방향 프라이머:

[0550] 5'-ggcagactacaagtctactagtactac-3'

[0551] 서열 번호 59:

[0552] T7 프로모터, CER 도메인, 및 VT 도메인 TrGA\_sgF1의 전사를 위한 올리고뉴클레오티드 서열. VT 도메인은 대문자로 표시되어 있지만, T7 프로모터와 CER 도메인 영역은 각각 굵은 글씨와 소문자로 표시되어 있다.

**taatacgaactactataGGGAAGACCAGGATCAAGgtttagagctagaaatagcaagttaaaataaggct**  
**agtcggttatcaactgaaaaagtgccaccgagtcggtgc**

[0553]

[0554] 서열 번호 60:

[0555] T7 프로모터, CER 도메인, 및 VT 도메인 TrGA\_sgR1의 전사를 위한 올리고뉴클레오티드 서열. VT 도메인은 대문자로 표시되어 있지만, T7 프로모터와 CER 도메인 영역은 각각 굵은 글씨와 소문자로 표시되어 있다.

**taatacgaactactataGGACAGACCGCTTGATCCgtttagagctagaaatagcaagttaaaataaggcta**  
**gtccggttatcaactgaaaaagtgccaccgagtcggtgc**

[0556]

[0557] 서열 번호 61:

[0558] TrGA 녹아웃 카세트의 뉴클레오티드 서열:

[0559] agatgccggccttgaagcagtaggtcgagaggctccacttcatggtgccggtgccgatcatggtgttgatgatgcggtcgtagctctcgatggcggcgtagt  
 agtcgcccgtcgagggcggcgaggtcggcgtactgcgtccagagctt

[0560] 서열 번호 62:

[0561] Fw1 (5'-cactactacatccatcgcagcaaacatgg-3') 및

[0562] 서열 번호 63:

[0563] R3 (5'-ggtcaagaagcacatgccagagttcg -3')

- [0564] 서열 번호 64:
- [0565] TrGAF2 (5' gactgtctccaccatgtaatttttc 3')
- [0566] 서열 번호 65:
- [0567] 대장균 코돈에 최적화된 Cas9-D10A 니카아제 유전자(중지 코돈 없음) agac
- [0568] 서열 번호 66:
- [0569] 스트렙토코쿠스 파이오제네스로부터의 Cas9-D10A 니카아제의 아미노산 서열; 서열 번호 65에 의해 인코딩됨
- [0570] eqeigkatakyffysnimnfketeitlangeirkprietngetgeivwdkgrdfatvrkvlsmqvni vkkt evqtggf skesilpkrnskliarkkdwd  
pkkygfgdsptvaysvlvkvkvekgkskklksvkelgitimerssfeknpidfleakgykevkkdliiklpkyslfelengrkrmlasagelqkgnelalp  
skyvfnlylasyeklkgspedneqqlfveqhkhyldiei eqisefskrviladanldkvl saynkhrdkpireqaeni ihlftltnlgapaafkyfdtti  
drkrytstkevldatlihsitglyetridlsqggd
- [0571] 서열 번호 67:
- [0572] (시작 메티오닌을 갖는) N-말단 His6 tag/트롬빈/S·Tag™ /엔테로키나아제 영역 아미노산 서열
- [0573] mhhhhhssglvprgsmketaaakferqhmdspdlgtdddkama
- [0574] 서열 번호 68:
- [0575] (시작 코돈을 갖는) N-말단 His6 tag/트롬빈/S·Tag™ /엔테로키나아제 영역 폴리뉴클레오티드 서열; 서열 번호 67을 인코딩
- [0576] atgcaccatcatcatcatcatctcttctggctggcgccacggcttctggatgaaagaaaccgctgctgctaatcgaacgccagcacatggacagccca  
gatctgggtaccgacgacgacgacaaggccatggcc
- [0577] 서열 번호 69:
- [0578] SV40 NLS 코딩 서열(서열 번호 46을 인코딩)
- [0579] ccaaaaaagaacgcaaggtt
- [0580] 서열 번호 70:
- [0581] BLR 핵 위치 신호 코딩 서열(서열 번호 47을 인코딩)
- [0582] Aagaagaaaaactgaaactg
- [0583] 서열 번호 71:
- [0584] tgatgacggtgaaaacctc
- [0585] 서열 번호 72:
- [0586] aaaagcaccgactcgg
- [0587] 서열 번호 73:
- [0588] Glalrep 프라이머 1556F
- [0589] ATGCGCAAATTTAAAGCGCTGATgtgtgtctaatgcctccaccac
- [0590] 서열 번호 74:
- [0591] Glalrep 프라이머 1557R
- [0592] ATATGGATCTGCGCGCATCGATgatcgtgctagcgtgctgttg
- [0593] 서열 번호 75:
- [0594] 서열분석 프라이머 1538F
- [0595] CCACCACAGGAACCAACC

- [0596] 서열 번호 76:
- [0597] 서열분석 프라이머 1539R
- [0598] CTGCGACGGAGGAATGACG
- [0599] 서열 번호 77:
- [0600] 서열분석 프라이머 1540F
- [0601] GGGCAGGACTGGCAAGGATGT
- [0602] 서열 번호 78:
- [0603] 서열분석 프라이머 1541R
- [0604] GCCGTCACGCCAGGAACAAG
- [0605] 서열 번호 79:
- [0606] TrGA 프라이머 KOF1
- [0607] gaacaatcttctttgcaatgttggtc
- [0608] 서열 번호 80:
- [0609] Pyr2 프라이머 R1
- [0610] gaggaagtctctgttaggcaggc
- [0611] 서열 번호 81:
- [0612] Pyr2 프라이머 F4
- [0613] cgacagagcagtcataatgggat acg

**도면**

**도면1**

T. 레에세이 U6 유전자 서열 번호 1:

AAAAAACAAGTACTTACTTATGTATTATTAACACTCTTCTGCAGTACTACCT  
 AAGAGGCTAGGGGTAGTTTTATAGCAGACTTATAGCTATATTTTTATTTAGTAAAGTGCTTTTAAA  
 GTAAGGCTTTTTTATAGCACTTTTTATTTATATAATATATATATAAATTTTAAAGCCTGGAATA  
 GTAAAGAGGCTTATATAATAATTTATAGTAATAAAAAGCTTAGCAGCTGTAATATAATTCCTAAAGAA  
 ACAGCATGAAATGGTATTATGTAAGAGCTATAGTCTAAAGGCACCTCTGCTGGATAAAAAATAGTGG

“TATA” 박스

CTATAAGTCTGCTGCAAAACTACCCCAACCTCGTAGG TATATAA GTACTGTTTGATGGTAGTCTA

번역 시작

인트론

TC G CCTTCGGGCATT TGGTCAATTTA TAACGATACAG GTTCGTTTC GGCTTTTCCTCGGAACCC

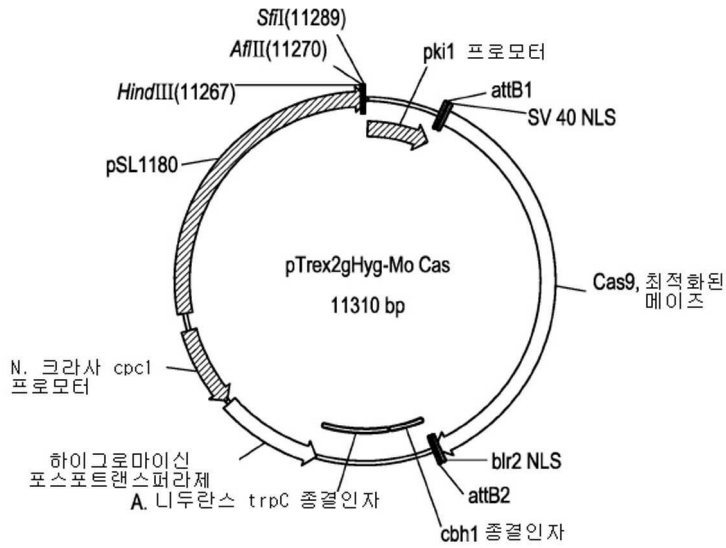
CCAGAGGTCATCAGTTCGAATCGCTAACAG GTCACAGAGAAGATTAGCATGGCCCTGCACT

종결인자

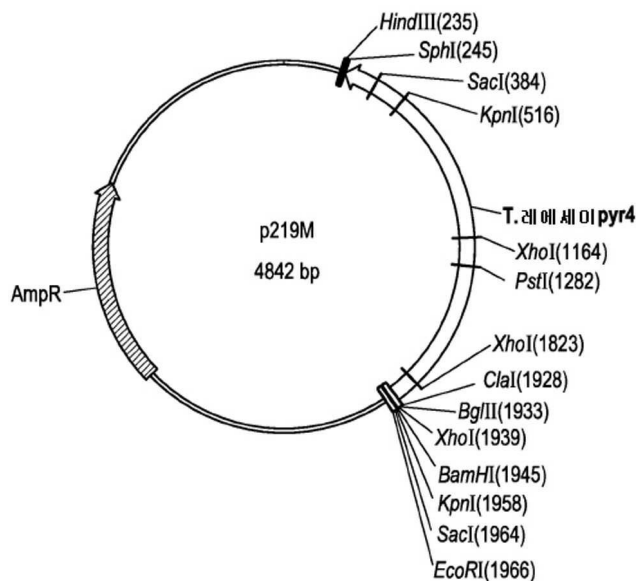
**AAGGATGACACGCT**CACTCAAAGAGAAGCTAAACA TTTTTTTCTCTT CCAAGTCGTGATGGTTA

TCTTTTGGCTTAGAGAATCTATTCTTGTGGACGATTAGTATTGGTAAATCCCTGCTGCACATTGGCGG  
 GGATGGTCTCAACGGCATAATACCCCATTCGTGATGCAGCGGTGATCTTCAATATGTAGTGAATACG  
 TTGCATACACCACAGGTTCCGGTGCCTCCTGTATGTACAGTACTGTAGTTCGACTCCTCCGCGCAG  
 TGGAAACGATTCCCTAGTGGCAGGTATTTTGGCGGGTCAAGAA

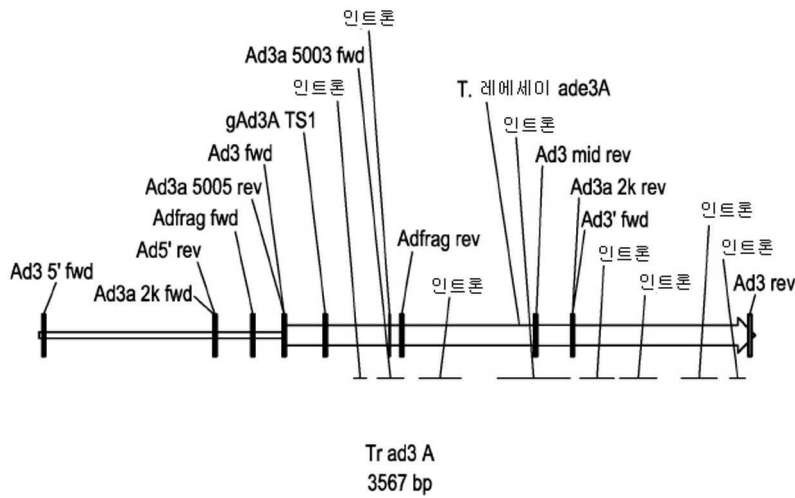
도면2



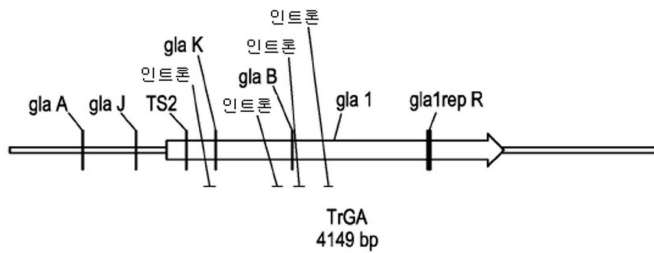
도면3



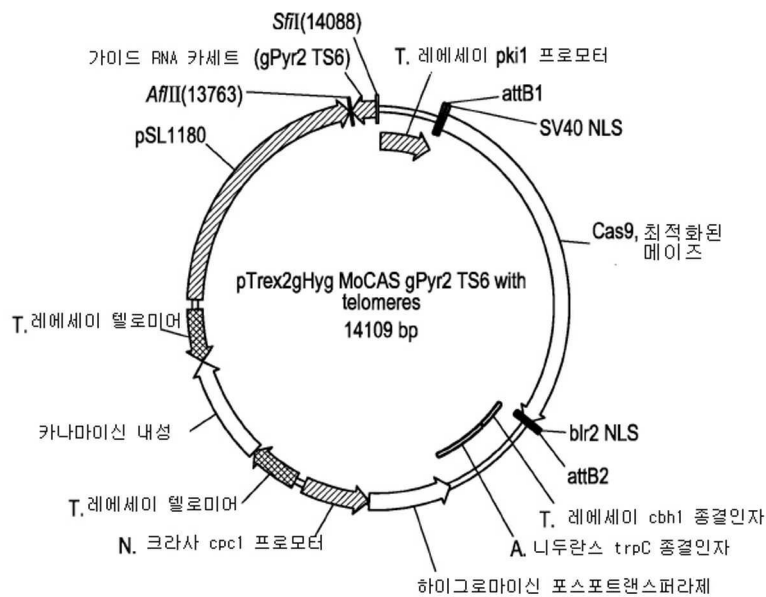
도면4



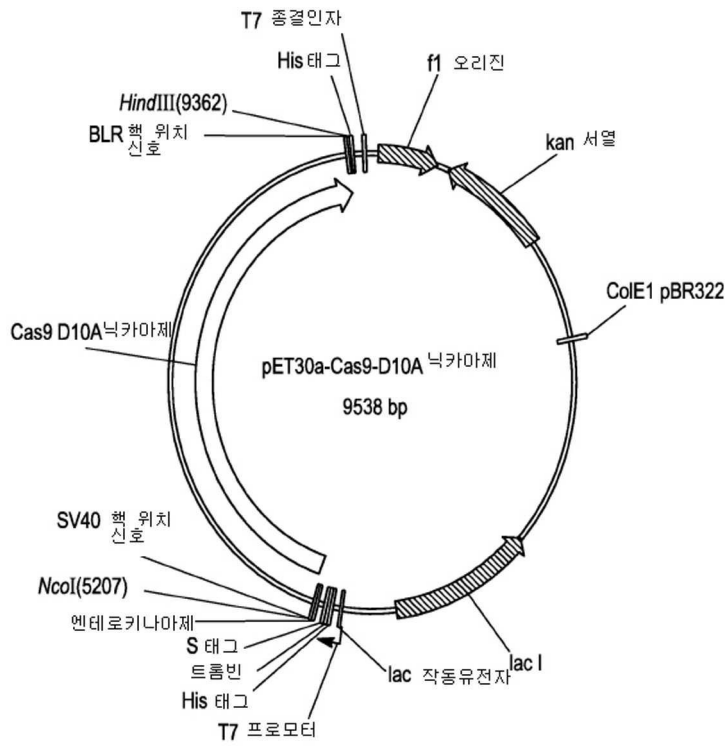
도면5



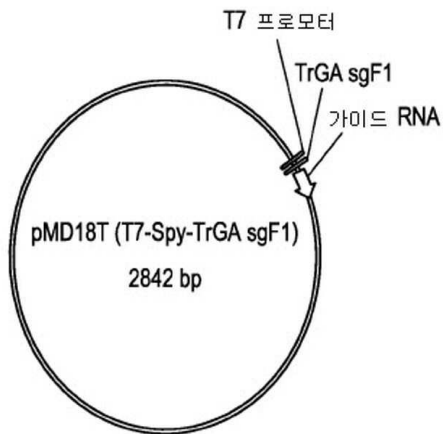
도면6



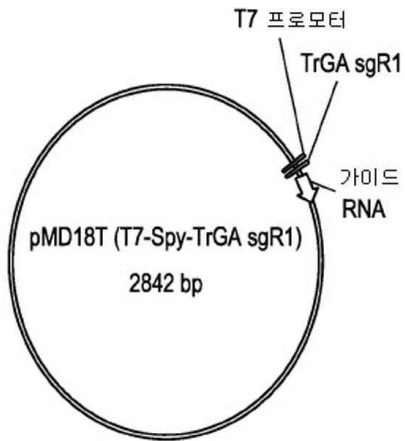
도면7



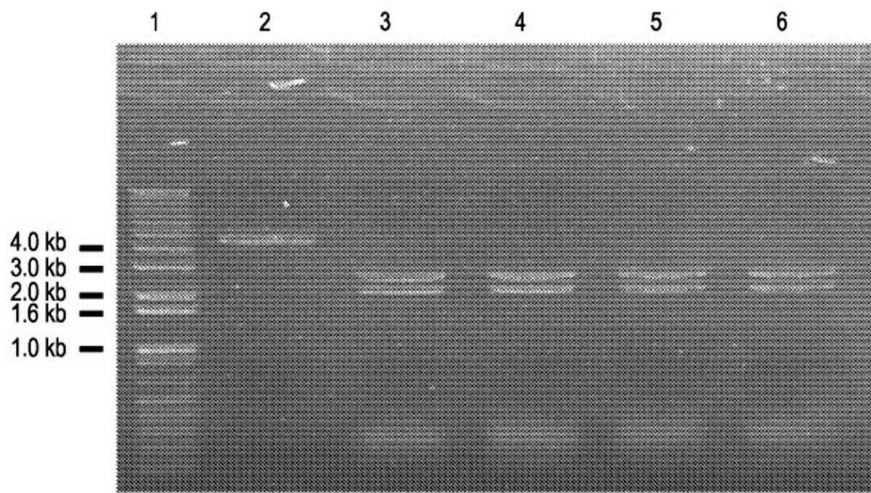
도면8a



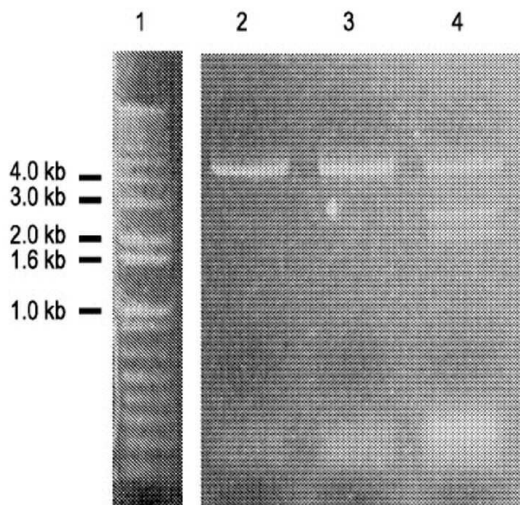
도면8b



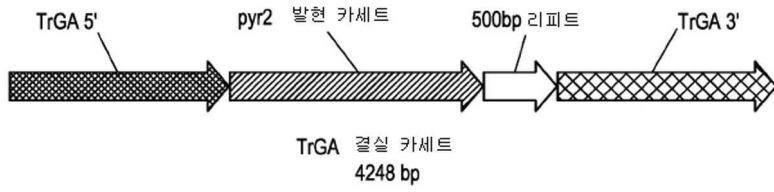
도면9



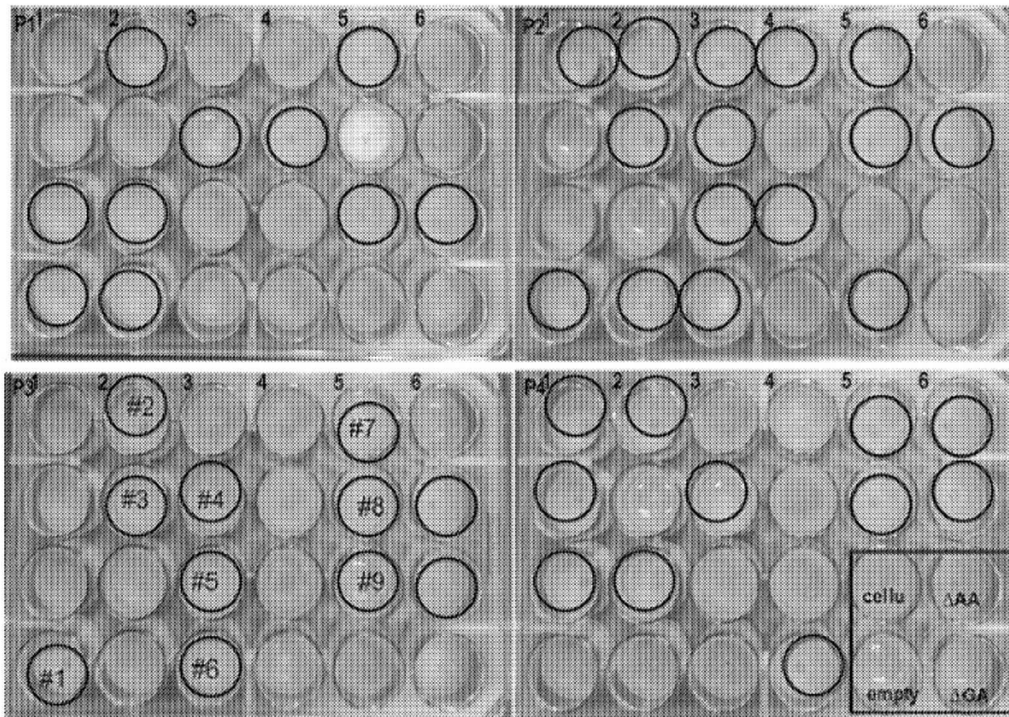
도면10



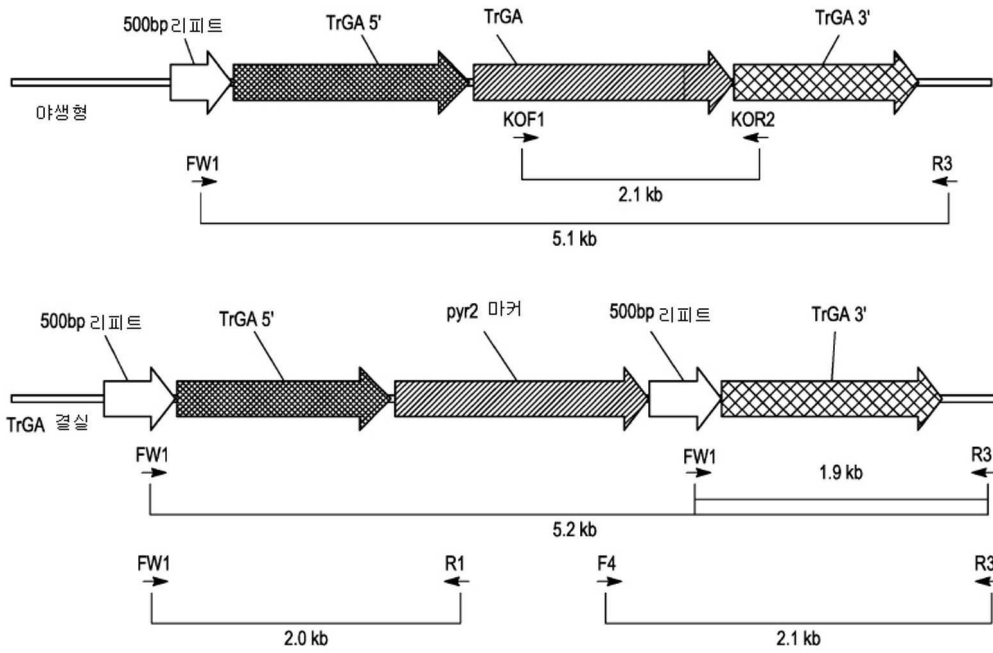
도면11



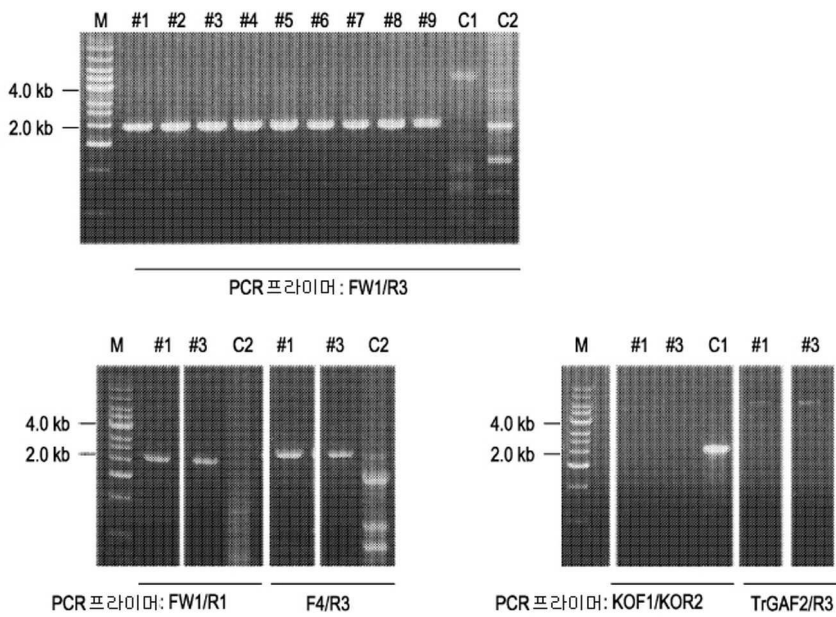
도면12



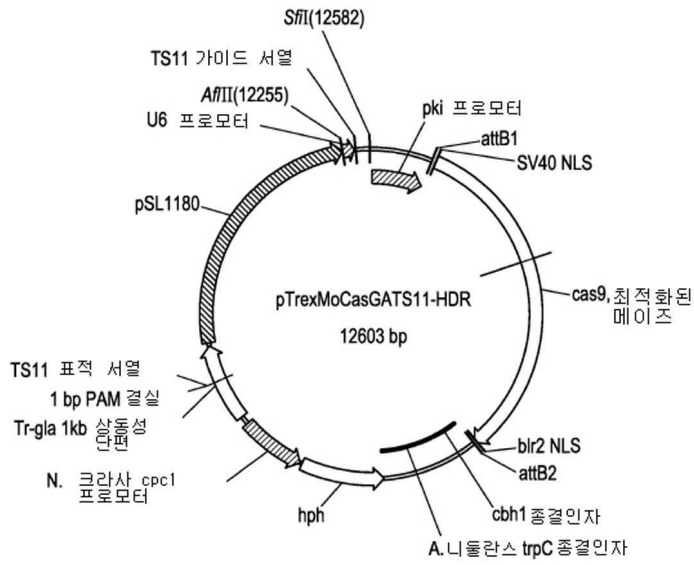
도면13a



도면13b



도면14



서열 목록

- <110> DANISCO US INC.  
 Ward, Michael  
 Chan, Jimmy  
 Bower, Benjamin S.  
 Madrid, Susan Mampusti  
 Ge, Jing  
 Gu, Xiaogang  
 Song, Danfeng  
 Song, Mingmin
- <120> FUNGAL GENOME MODIFICATION SYSTEMS AND METHODS OF USE
- <130> 40532-WO-PCT-4
- <150> PCT/CN2014/093914
- <151> 2014-12-16
- <150> PCT/CN2014/093916
- <151> 2014-12-16
- <150> PCT/CN2014/093918
- <151> 2014-12-16
- <160> 81
- <170> PatentIn version 3.5

<210> 1

<211> 842

<212> DNA

<213> T. reesei

<400> 1

aaaaaacact agtaagtact tactttatgta ttattaacta ctttagctaa ctctgcagt 60

actacctaag aggctagggg tagttttata gcagacttat agctattatt tttatttagt 120

aaagtgcttt taaagtaagg tcttttttat agcacttttt atttattata atatatatta 180

tataataatt ttaagcctgg aatagtaaag aggccttatat aataatttat agtaataaaa 240

gcttagcagc tgtaataata ttcctaaaga aacagcatga aatggatta tgtaagagct 300

atagtctaaa ggcactctgc tggataaaaa tagtggctat aagtctgctg caaaactacc 360

cccaacctcg taggtatata agtactgttt gatggtagtc tatcgcttc gggcatttgg 420

tcaatttata acgatacagg ttcgtttcgg cttttcctcg gaacccccag aggtcatcag 480

ttcgaatcgc taacaggtca acagagaaga ttagcatggc ccctgcaacta aggatgacac 540

gctcactcaa agagaagcta aacatttttt ttctcttcca agtcgtgatg gttatctttt 600

tgcttagaga atctattctt gtggacgatt agtattggta aatccctgct gcacattgcg 660

gcggatggtc tcaacggcat aatacccat tctgatgca gcggtgatct tcaatatgta 720

gtgtaatacg ttgcatacac caccaggttc ggtgcctcct gtatgtacag tactgtagtt 780

cgactcctcc ggcaggtgg aaacattcc ctagtgggca ggtatttttg cggggtcaag 840

aa 842

<210> 2

<211> 99

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<220><221> misc\_feature

<222> (1)..(20)

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

<400> 2

nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaaagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cgguggugc 99

<210> 3

<211> 99  
 <212> RNA  
 <213> Artificial Sequence  
 <220><223> Synthetic sequence  
 <400> 3  
 guccucgagc aaaaggugcc guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60  
 cguaaucaac uugaaaaagu ggcaccgagu cgguggugc 99  
 <210> 4  
 <211> 99  
 <212> RNA  
 <213> Artificial Sequence  
 <220><223> Synthetic sequence  
 <400> 4  
 guucagugca auaggcgucu guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60  
 cguaaucaac uugaaaaagu ggcaccgagu cgguggugc 99  
  
 <210> 5  
 <211> 99  
 <212> RNA  
 <213> Artificial Sequence  
 <220><223> Synthetic sequence  
 <400> 5  
 gccaauggcg acggcagcac guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60  
 cguaaucaac uugaaaaagu ggcaccgagu cgguggugc 99  
 <210> 6  
 <211> 99  
 <212> RNA  
 <213> Artificial Sequence  
 <220><223> Synthetic sequence  
 <400> 6  
 gcacagcggg augcccuugu guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60  
 cguaaucaac uugaaaaagu ggcaccgagu cgguggugc 99  
  
 <210> 7  
 <211> 4155

<212> DNA

<213> Artificial Sequence

<220><223> codon optimized seqs

<400> 7

```

atggcaccga agaagaagcg caaggtgatg gacaagaagt acagcatcgg cctcgacatc      60
ggcaccaact cggtagggctg ggccgtcatc acggacgaat ataaggtccc gtcgaagaag      120
ttcaaggtcc tcggcaatac agaccgccac agcatcaaga aaaacttgat cggcgcctc      180
ctgttcgata gggcgagac gcgggagcgc accaggctca agaggaccgc caggagacgg      240
tacactagge gcaagaacag gatctgctac ctgcaggaga tcttcagcaa cgagatggcg      300

aaggtggacg actccttctt ccaccgcctg gaggaatcat tcctggtgga ggaggacaag      360
aagcatgagc ggcaccaat cttcggcaac atcgtcgacg aggtggccta ccacgagaag      420
taccgacaa tctaccacct ccggaagaaa ctggtggaca gcacagacaa ggcgacctc      480
cggctcatct accttgcctt cgcgcatatg atcaagtcc gcggccactt cctcatcgag      540
ggcgacctga acccgacaa ctcgcagctg gacaagctgt tcatccagct cgtgcagacg      600
tacaatcaac tgttcgagga gaacccata aacgctagcg gcgtggacgc caaggccatc      660
ctctcggcca ggctctcgaa atcaagaagg ctggagaacc ttatcgcga gttgccaggc      720

gaaaagaaga acggcctctt cggcaacctt attgcgetca gcctcggcct gacgecgaac      780
ttcaaatcaa acttcgacct cgcggaggac gccaagctcc agctctcaa ggacacctac      840
gacgacgacc tcgacaacct cctggcccag ataggagacc agtacgcgga cctcttctc      900
gccgccaaga acctctccga cgctatcctg ctgagcgaca tccttcgggt caacaccgaa      960
attaccaagg caccgctgtc gccagcatg attaaacgt acgacgagca ccatcaggac      1020
ctcacgctgc tcaaggcact cgtccgccag cagctccccg agaagtacaa ggagatcttc      1080
ttcgaccaat caaaaaacgg ctacgcggga tatatcgacg gcggtgccag ccaggaagag      1140

ttctacaagt tcatcaaac aatcctggag aagatggacg gcaccgagga gttgctggtc      1200
aagctcaaca gggaggacct cctcaggaag cagaggacct tcgacaacgg ctccatcccg      1260
catcagatcc acctgggcga actgcatgcc atcctgcggc gccaggagga cttctaccg      1320
ttctgaagg ataaccggga gaagatcgag aagatcttga cgttcgcat cccatactac      1380
gtgggcccgc tggtcgcgg caactcccgg ttcgcctgga tgacccggaa gtcggaggag      1440
accatcacac cctggaactt tgaggagggt gtcgataagg gcgctagcgc tcagagcttc      1500
atcgagcgca tgaccaactt cgataaaaac ctgccaatg aaaaagtctt cccaagcac      1560

tcgctgctct acgagtactt caccgtgtac aacgagctca ccaaggtcaa atacgtcacc      1620

```

gagggcatgc ggaagccggc gttcctgagc ggcgagcaga agaaggcgat agtggacctc 1680  
ctcttcaaga ccaacagaa ggtgaccgtg aagcaattaa aagaggacta cttcaagaaa 1740  
atagagtgtc tcgactccgt ggagatctcg ggcgtggagg atcggttcaa cgctcactc 1800  
ggcacgtatc acgacctcct caagatcatt aaagacaagg acttctcga caacgaggag 1860  
aacgaggaca tcctcgagga catcgtctc acctgacct tgttcgagga ccgcgaaatg 1920  
atcgaggaga ggcggaagac ctacgcgac ctgttcgacg acaaggatcat gaaacagctc 1980  
  
aagaggcgcc gctacactgg ttggggaagg ctgtcccga agctcattaa tggcatcagg 2040  
gacaagcaga gcgcaagac catcctggac ttctcaagt ccgacgggtt cgccaaccgc 2100  
aacttcatgc agctcattca cgacgactcg ctcacgttca aggaagacat ccagaaggca 2160  
caggtgagcg ggcagggtga ctccctccac gaacacatcg ccaacctggc cggctcgccg 2220  
gccattaaaa agggcatcct cgacacggtc aaggtcgtcg acgagctcgt gaaggtgatg 2280  
ggccggcaca agcccgaaaa tatcgtcata gagatggcca gggagaacca gaccacccaa 2340  
aaagggcaga agaactcgcg cgagcggatg aaacggatcg aggaggcat taaagagctc 2400  
  
gggtcccaga tectgaagga gcacccctg gaaaataccc agctccagaa tgaanaagctc 2460  
tacctctact acctgcagaa cggcccgac atgtactgg accaggagct ggacattaat 2520  
cggtatcgg actacgacgt cgaccacatc gtcccgagt cgttctcaa ggacgatagc 2580  
atcgacaaca aggtgctcac ccggtcggat aaaaatcggg gcaagagcga caacgtgccc 2640  
agcaggaggg tcgtgaagaa gatgaaaaac tactggcgcc agctcctcaa cgcgaaactg 2700  
atcaccagc gcaagttcga caacctgacg aaggcggaac gcggtggctt gagcgaactc 2760  
gataaggcgg gcttcataaa aaggcagctg gtcgagacgc gccagatcac gaagcatgtc 2820  
  
gcccagatcc tggacagccg catgaatact aagtacgatg aaaacgaca gctgatccgg 2880  
gaggtgaagg tgatcacgct gaagtccaag ctcgtgtcgg acttccgcaa ggacttcag 2940  
ttctacaagg tcccgagat caacaactac caccacgcc acgacgcta cctgaatgcg 3000  
gtggtcggga ccgccctgat caagaagtac ccgaagctgg agtcggagt cgtgtacggc 3060  
gactacaagg tctacgacgt gcgcaaatg atcgccaagt ccgacagga gatcggaag 3120  
gccacggcaa aatcttctt ctactgaac atcatgaact tcttcaagac cgagatcacc 3180  
ctcggaacg gcgagatccg caagcggcgt ctcatgaaa ccaacggcga gacgggcgag 3240  
  
atcgtctggg ataaggccg ggatttcgcg acggtccgca aggtgctctc catgccgcaa 3300  
gtcaatatcg taaaaagac ggaggtccag acggcgggt tcagcaagga gtccatctc 3360  
ccgaagcga actccgaca gctcatcgcg aggaagaagg attgggacct gaaaaatat 3420  
ggcggcttcg acagcccgac cgtcgcatc agcgtcctcg tcgtggcgaa ggtggagaag 3480

ggcaagtcaa agaagctcaa gtccgtgaag gagctgctcg ggatcacgat tatggagcgg 3540  
 tcctccttcg agaagaacct gatcgacttc ctgaggcca agggatataa ggaggtcaag 3600  
 aaggacctga ttattaaact gccgaagtac tcgctcttcg agctggaaaa cggccgcaag 3660

aggatgctcg cctccgcagg cgagttgcag aagggaacg agctcgccct cccgagcaaa 3720  
 tacgtcaatt tctgtacct cgctagccac tatgaaaagc tcaagggcag cccggaggac 3780  
 aacgagcaga agcagctctt cgtggagcag cacaagcatt acctggacga gatcatcgag 3840  
 cagatcagcg agttctcga gcggtgatc ctgccgacg cgaacctgga caaggtgctg 3900  
 tcggcatata acaagcaccg cgacaaacca atacgagcagc aggccgaaaa tatcatccac 3960  
 ctcttcaccc tcaccaacct cggcgctccg gcagccttca agtacttga caccacgatt 4020  
 gaccggaagc ggtacacgag cacgaaggag gtgctcgatg cgacgctgat ccaccagagc 4080

atcacagggc tctatgaaac acgcatcgac ctgagccagc tgggcggaga caagaagaag 4140  
 aagctcaagc tctag 4155

<210> 8

<211> 1384

<212> PRT

<213> Artificial Sequence

<220><223> protein expressed from synthetic construct

<400> 8

Met Ala Pro Lys Lys Lys Arg Lys Val Met Asp Lys Lys Tyr Ser Ile

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg

65 70 75 80

Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser

85 90 95

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

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



Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile  
 610 615 620  
 Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met  
 625 630 635 640  
 Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val  
 645 650 655  
 Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser  
 660 665 670  
 Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile  
 675 680 685  
 Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln  
 690 695 700  
 Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala  
 705 710 715 720  
 Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu  
 725 730 735  
 Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val  
 740 745 750  
 Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile  
 755 760 765  
 Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys  
 770 775 780  
 Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu  
 785 790 795 800  
 Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln  
 805 810 815  
 Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr  
 820 825 830  
 Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp  
 835 840 845  
 His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys  
 850 855 860

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro  
 865 870 875 880

Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu  
 885 890 895

Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala  
 900 905 910

Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg  
 915 920 925

Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu  
 930 935 940

Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg  
 945 950 955 960

Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg  
 965 970 975

Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His  
 980 985 990

Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys  
 995 1000 1005

Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val  
 1010 1015 1020

Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys  
 1025 1030 1035 1040

Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys  
 1045 1050 1055

Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile  
 1060 1065 1070

Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp  
 1075 1080 1085

Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val  
 1090 1095 1100

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

1105                    1110                    1115                    1120  
 Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp  
                                  1125                    1130                    1135  
 Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val  
                                  1140                    1145                    1150  
 Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser  
                                  1155                    1160                    1165  
 Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu  
                                  1170                    1175                    1180  
  
 Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys  
 1185                    1190                    1195                    1200  
 Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu  
                                  1205                    1210                    1215  
 Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly  
                                  1220                    1225                    1230  
 Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala  
                                  1235                    1240                    1245  
 Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys  
  
                                  1250                    1255                    1260  
 Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu  
 1265                    1270                    1275                    1280  
 Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu  
                                  1285                    1290                    1295  
 Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg  
                                  1300                    1305                    1310  
 Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly  
                                  1315                    1320                    1325  
  
 Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg  
                                  1330                    1335                    1340  
 Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser  
 1345                    1350                    1355                    1360  
 Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly

1365                      1370                      1375

Asp Lys Lys Lys Lys Leu Lys Leu

1380

<210> 9

<211> 412

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 9

gaattcggat cctctttgaa aagataatgt atgattatgc tttcactcat atttatacag 60

aaacttgatg ttttctttcg agtatataca aggtgattac atgtacgttt gaagtacaac 120

tctagatfff gtagtgcctt cttgggctag cggtaaaggt gcgcattfff tcacacccta 180

caatgttctg ttcaaaagat tttggtcaaa cgctgtagaa gtgaaagttg gtgcgcatgt 240

ttcggcgctt gaaacttctc cgcagtgaaa gataaatgat cgtcctcgag caaaaggtgc 300

cgttttagag ctagaaatag caagttaaaa taaggctagt ccgttatcaa cttgaaaaag 360

tggcaccgag tcggtggtgc tttttttggt ttttatgtct gaattcggat cc 412

<210> 10

<211> 540

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 10

gaattcggat ccaaaaaaca ctagtaagta cttacttatg tattattaac tacttttagct 60

aacttctgca gtactaccta agaggctagg ggtagtffta tagcagactt atagctatta 120

tttttattta gtaaagtgct tttaaagtaa ggtctttfff atagcactff ttattttatta 180

taatataat tatataataa ttttaagcct ggaatagtaa agaggcttat ataataattt 240

atagtaataa aagcttagca gctgtaatat aattcctaaa gaaacagcat gaaatggtat 300

tatgtaagag ctatagtcta aaggcactct gctggataaa aatagtggtc ataagctgct 360

tgcaaaacta cccccaacct cgtaggtata taagtactgt ttgatggtag tctatcgtcc 420

tcgagcaaaa ggtgccgttt tagagctaga aatagcaagt taaaataagg ctagtccgtt 480

atcaacttga aaaagtgcca ccgagtcggt ggtgctffff tttctcttga attcggatcc 540

540

<210> 11  
 <211> 597  
 <212> DNA

<213> Artificial Sequence  
 <220><223> Synthetic sequence

<400> 11

gaattcggat ccaaaaaaca ctagtaagta cttacttatg tattattaac tactttagct 60  
 aacttctgca gtactaccta agaggctagg ggtagtttta tagcagactt atagctatta 120  
 tttttattta gtaaagtgct tttaaagtaa ggtctttttt atagcacttt ttatttatta 180  
 taatatatat tatataataa ttttaagcct ggaatagtaa agaggcttat ataataattt 240  
 atagtaataa aagcttagca gctgtaatat aattcctaaa gaaacagcat gaaatggat 300  
 tatgtaagag ctatagtcta aaggcactct gctggataaa aatagtggct ataagctgc 360

tgcaaaaacta cccccaacct cgtaggtata taagtactgt ttgatggtag tctatcgtcc 420  
 tcgagcaaaa ggtgcgcttt tagagctaga gttcgtttcg gcttttcctc ggaaccccca 480  
 gaggtcatca gttcgaatcg ctaacagaat agcaagttaa aataaggcta gtccgttadc 540  
 aacttgaaaa agtggcaccg agtcggtggg gctttttttt ctcttgaatt cggatcc 597

<210> 12  
 <211> 315  
 <212> DNA

<213> Artificial Sequence  
 <220><223> Synthetic sequence

<400> 12

aattcctaaa gaaacagcat gaaatggat tatgtaagag ctatagtcta aaggcactct 60  
 gctggataaa aatagtggct ataagctctgc tgcaaaaacta cccccaacct cgtaggtata 120  
 taagtactgt ttgatggtag tctatcgcca atggcgacgg cagcacgttt tagagctaga 180  
 gttcgtttcg gcttttcctc ggaaccccca gaggtcatca gttcgaatcg ctaacagaat 240  
 agcaagttaa aataaggcta gtccgttadc aacttgaaaa agtggcaccg agtcggtggg 300  
 gctttttttt ctctt 315

<210> 13  
 <211> 39  
 <212> DNA

<213> Artificial Sequence

<220><223> Primer  
 <400> 13  
 cgtcagctta agaattccta aagaaacagc atgaaatgg 39

<210> 14  
 <211> 47  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 14  
 cgtcagggcc acgtgggcca agagaaaaa aagcaccacc gactcgg 47

<210> 15  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 15  
 tgaacacagc caccgacatc agc 23

<210> 16  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 16  
 gctggtgagg gtttgctta ttg 23

<210> 17  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 17  
 gattgcttgg gaggaggaca t 21

<210> 18  
 <211> 23

<212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 18  
 cgaggccact gatgaagttg ttc 23  
 <210> 19  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 19  
 cagttttcca aggctgcca cgc 23  
  
 <210> 20  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 20  
 ctgatcttgc accctggaaa tc 22  
 <210> 21  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 21  
 ctctctatca ttgcccacc tcc 23  
 <210> 22  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 22  
 ctccattcac cctcaattct cc 22

<210> 23  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 23  
 gttcccttgg cgggtgcttgg atc 23  
 <210> 24  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 24  
 caatagcaca aaccctcacc agc 23  
 <210> 25  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 25  
 gaacaacttc atcagtggcc tcg 23  
  
 <210> 26  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 26  
 ccgtagttg aagatccttg ccg 23  
 <210> 27  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 27

gtcaggatt tgcttcatac etc	23
<210> 28	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Primer	
<400> 28	
gagagacgca ggatgactca aag	23
<210> 29	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Primer	
<400> 29	
tgccgtgggt cattggcata ttc	23
<210> 30	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Primer	
<400> 30	
tgccgacttt gtccagtgat tcg	23
<210> 31	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Primer	
<400> 31	
ttacatgtgg acgcgagata gcg	23
<210> 32	
<211> 22	
<212> DNA	
<213> Artificial Sequence	

<220><223> Primer  
 <400> 32  
 gtgtgtctaa tgcctccacc ac 22  
 <210> 33  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 33  
 gatcgtgcta gcgctgctgt tg 22  
 <210> 34  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 34  
 ccgtgatgga gcccgctctt c 21  
  
 <210> 35  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 35  
 cgcggtgagt tcaggctttt tc 22  
 <210> 36  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 36  
 gtataagagc aggaggaggg ag 22  
 <210> 37  
 <211> 22  
 <212> DNA

<213> Artificial Sequence

<220><223> Primer

<400> 37

gaacgcctca atcagtcagt cg 22

<210> 38

<211> 2431

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 38

tcaggaata gctttaagta gcttattaag tattaaaatt atatatat ttaatataac 60

tatatttctt taataaatag gtattttaag ctttatatat aaatataata ataaaataat 120

atattatata gctttttatt aataaataaa atagctaaaa atataaaaaa aatagcttta 180

aaatacttat ttttaattag aattttatat atttttaata tataagatct tttacttttt 240

tataagcttc ctaccttaaa ttaaattttt actttttttt actattttac tataatcttaa 300

ataaaggctt taaaaatata aaaaaaatct tcttatatat tataagctat aaggattata 360

tatatat tttttaattt ttaaagtaag tattaaagct agaattaaag ttttaatttt 420

ttaaggcttt atttaaaaaa aggcagtaat agcttataaa agaaatttct ttttctttta 480

tactaaaagt actttttttt taataagggt agggttaggg tttactcaca cggaccatcc 540

caaccacatc ttagggttag ggtaggggtt agggttaggg ttagggttag ggtagggta 600

agggtttaa caaagccagc ttgtgtctca aaatctctga tgttacattg cacaagataa 660

aaatatatca tcatgaacaa taaaactgtc tgcttacata aacagtaata caaggggtgt 720

tatgagccat attcaacggg aaacgtcttg ctcgaggccg cgattaaatt ccaacatgga 780

tgctgattta tatgggtata aatgggctcg cgataatgtc gggcaatcag gtgcgacaat 840

ctatcgattg tatgggaagc ccgatgccc agagttgttt ctgaaacatg gcaaaggtag 900

cgttgccaat gatgttacag atgagatggt cagactaac tggctgacgg aatttatgcc 960

tcttccgacc atcaagcatt ttatccgtac tctgatgat gcatggttac tcaccactgc 1020

gateccccggg aaaacagcat tccaggtatt agaagaatat cctgattcag gtgaaaatat 1080

tgttgatgcg ctggcagtg tctgcgccg gttgcattcg attcctgttt gtaattgtcc 1140

ttttaacage gatcgcgtat ttcgtctcgc tcaggcgcaa tcacgaatga ataacggttt 1200

ggttgatgcg agtgattttg atgacgagcg taatggctgg cctgttgaac aagtctggaa 1260

agaaatgcat aagcttttgc cattctcacc ggattcagtc gtcactcatg gtgatttctc 1320  
 acttgataac cttatTTTTg acgaggggaa attaataaggt tgtattgatg ttggacgagt 1380  
 cggaatcgca gaccgatacc aggatcttgc catcctatgg aactgcctcg gtgagttttc 1440  
 tccttcatta cagaaacggc tttttcaaaa atatggtatt gataatcctg atatgaataa 1500  
 attgcagttt catttgaigc tcgatgagtt tttctaataca gaattggtta attggttga 1560  
  
 acactggcag agcattacgc tgacttgacg ggacggcggc tttgttgaat aaategaact 1620  
 tttgctgagt tgaaggatca gatcacgcat cttcccgaca acgcagaccg ttccgtggca 1680  
 aagcaaaagt tcaaatcac caactggtcc acctacaaca aagctctcat caaccgtggc 1740  
 tcctcactt tctggctgga tgatggggcg attcaggcct ggtatgagtc agcaacacct 1800  
 tcttcacgag gcagacctca gcggtttaa cctaacccta accctaacc taacctaac 1860  
 cctaacccta accctaacc taacctaac cctaacccta accctaacc taacctaac 1920  
 ctaatggggt cgatctgaac cgaggatgag ggttctatag actaatctac aggccgtaca 1980  
  
 tgggtgtgatt gcagatgcga cgggcaaggt gtacagtgtc cagaaggagg agagcggcat 2040  
 aggtatttga atagaccagc tttacataat aatcgctgt tgctactgac tgatgacctt 2100  
 cttccctaac cagtttctta attaccactg cagtgaggat aaccctaact cgctctgggg 2160  
 ttattattat actgattagc aggtggctta tatagtctg aagtactata agagtttctg 2220  
 cgggaggagg tggaaggact ataaactgga cacagttagg gatagagtga tgacaagacc 2280  
 tgaatgttat cctccggtgt ggtatagcga attggctgac cttgcagatg gtaatggttt 2340  
 aggcagggtt tttgcagagg gggacgagaa cgcgttctgc gatttaacgg ctgctgccgc 2400  
  
 caagctttac ggttctctaa tgggcggccg c 2431  
 <210> 39  
 <211> 982  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic sequence  
 <400> 39  
 gtgtgtctaa tgcctccacc acaggaacca aaccggcttt gacctctggg aagaagtcaa 60  
 tgggagctca ttctttactg ttgccaacca gcaccgaggt atgaagcaaa tctctgacat 120  
 tcgctgctac tgcacatgag cattgttact gaccagctct acagcacttg tcgagggcgc 180  
 cactcttget gccactcttg gccagtggg aagcgttat tcactctgtg ctccccaggt 240  
  
 tttgtgcttt ctccaacgat tctgggtgtc gtctgggtga tacgtcgact ccaacagtat 300

gtctttttcac tgtttatatg agattggcca atactgatag ctgcctcta gtcaacacca 360  
 acgagggcag gactggcaag gatgtcaact cgcctctgac ttccatccac accttcgate 420  
 ccaaccttgg ctgtgacgca ggcaccttcc agccatgcag tgacaaagcg ctctccaacc 480  
 tcaaggttgt tgcgactcc ttccgctcca tctacggcgt gaacaagggc attcctgcgg 540  
 tgctgccgtc gccattggcc ggtatgcaga ggatgtgtac tacaacggca accttggtgta 600  
 tcttgctaca tttgctgctg ccgagcagct gtacgatgcc atctacgtct ggaagaagac 660

gggctccatc acggtgaccg ccacctccct ggcttcttc caggagcttg ttcttggcgt 720  
 gacggccggg acctactcca gcagctcttc gacctttacc aacatcatca acgccgtctc 780  
 gacatacgcc gatggcttcc tcagcgaggc tgccaagtac gtccccgccc acggttcgct 840  
 ggccgagcag ttgaccgca acagcggcac tccgctgtct gcgcttcacc tgacgtggtc 900  
 gtacgctcgc ttcttgacag ccacggcccc tcgggctggc atcgtgcccc cctcgtgggc 960  
 caacagcagc gctagcacga tc 982

<210> 40  
 <211> 404  
 <212> DNA

<213> T. reesei  
 <400> 40

aaaaaacact agtaagtact tacttatgta ttattaacta ctttagctaa cttctgcagt 60  
 actacctaag aggctagggg tagttttata gcagacttat agctattatt tttatttagt 120  
 aaagtgcttt taaagtaagg tcttttttat agcacttttt atttattata atatatatta 180  
 tataataatt ttaagcctgg aatagtaaag aggcttatat aataatttat agtaataaaa 240  
 gcttagcagc tgtaataataa ttctaaaga aacagcatga aatggtatta tgtaagagct 300  
 atagtctaaa ggcactctgc tggataaaaa tagtggctat aagtctgctg caaaactacc 360  
 cccaactcgc taggtatata agtactgttt gatggtagtc tate 404

<210> 41  
 <211> 146  
 <212> DNA  
 <213> T. reesei  
 <400> 41

aattcctaaa gaaacagcat gaaatggtat tatgtaagag ctatagtcta aaggcactct 60  
 gctggataaa aatagtggtc ataagtctgc tgcaaaaacta cccccaacct cgtaggtata 120  
 taagtactgt ttgatggtag tctatc 146

<210> 42  
 <211> 57  
 <212> DNA  
 <213> T. reesei  
 <400> 42  
 gttcgtttcg gcttttctc ggaaccccca gaggtcatca gttcgaatcg ctaacag 57  
 <210> 43  
 <211> 13  
 <212> DNA  
 <213> T. reesei  
 <400> 43  
 ttttttttct ctt 13  
 <210> 44  
 <211> 4104  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> codon optimized seqs  
 <400> 44  
 atggacaaga agtacagcat cggcctcgac atcggcacca actcgggtgg ctgggccgtc 60  
 atcacggacg aatataaggt cccgtcgaag aagttcaagg tcctcggaac tacagaccgc 120  
 cacagcatca agaaaaactt gatcggcgcc ctctgttcg atagcggcga gaccgcggag 180  
 gcgaccaggc tcaagaggac cgccaggaga cgttacctg ggcgcaagaa caggatctgc 240  
 tacctgcagg agatcttcag caacgagatg gcgaaggtgg acgactcctt ctccaccgc 300  
 ctggaggaat cattcctggt ggaggaggac aagaagcatg agcggcaccc aatcttcggc 360  
 aacatcgteg acgaggtggc ctaccacgag aagtaccga caatctacca cctccggaag 420  
 aaactggtgg acagcacaga caaggcggac ctccggctca tctaccttgc cctcgcgcat 480  
 atgatcaagt tccgcgcca cttctcatc gagggcgacc tgaacccgga caactccgac 540  
 gtggacaagc tgttcatcca gctcgtgcag acgtacaatc aactgttcga ggagaacccc 600  
 ataaacgcta gcggcgtgga cgccaaggcc atcctctcgg ccaggctctc gaaatcaaga 660  
 aggctggaga accttatcgc gcagttgccg ggcgaaaaga agaacggcct ctccggcaac 720  
 cttattgcgc tcagcctcgg cctgacgccg aacttcaat caaacttca cctcgcggag 780  
 gacgccaagc tccagctctc aaaggacacc tacgacgacg acctcgacaa cctcctggcc 840

cagataggag accagtacgc ggacctcttc ctgccgcca agaacctctc cgacgctatc 900  
 ctgctcagcg acatccttcg ggtcaacacc gaaattacca aggcaccgct gtccgccagc 960  
 atgattaaac gctacgacga gcaccatcag gacctcacgc tgctcaaggc actcgtccgc 1020  
 cagcagctcc ccgagaagta caaggagatc ttcttcgacc aatcaaaaaa cggttacgcg 1080  
  
 ggatatatcg acggcgggtgc cagccaggaa gatttctaca agttcatcaa accaatcctg 1140  
 gagaagatgg acggcaccga ggagtgtctg gtcaagetca acaggaggga cctcctcagg 1200  
 aagcagagga ccttcgacaa cggctccatc ccgcatcaga tccacctggg ggaactgcat 1260  
 gccatcctgc ggcgccagga ggacttctac ccgttctga aggataaccg ggagaagatc 1320  
 gagaagatct tgacgttccg catcccatac tacgtgggcc cgctggctcg cggcaactcc 1380  
 cggttcgctt ggatgaccgg gaagtccggag gagaccatca caccctggaa ctttgaggag 1440  
 gtggtcgata agggcgctag cgctcagagc ttcatcgagc gcatgaccaa cttcgataaa 1500  
  
 aacctgcca atgaaaaagt cctcccaag cactcgctgc tctacgagta cttcacctg 1560  
 tacaacgagc tcaccaaggt caaatcgtc accgagggca tgcggaagcc ggcgttctg 1620  
 agcggcgagc agaagaaggc gatagtggac ctctcttca agaccaacag gaaggtgacc 1680  
 gtgaagcaat taaaagagga ctacttcaag aaaatagagt gcttcgactc cgtggagatc 1740  
 tcgggcgtgg aggatcggtt caacgcctca ctccggcacgt atcacgacct cctcaagatc 1800  
 attaaagaca aggacttctt cgacaacgag gagaacgagg acatcctcga ggacatcgtc 1860  
 ctcacctga cctgttctga ggaccgcgaa atgatcgagg agaggctgaa gacctacgcg 1920  
  
 cacctgttcg acgacaaggt catgaaacag ctcaagaggc gccgctacac tggttgggga 1980  
 aggctgtccc gcaagctcat taatggcatc agggacaagc agagcggcaa gaccatcctg 2040  
 gacttctca agtccgacgg gttcgcacac cgcaacttca tgcagctcat tcacgacgac 2100  
 tcgctcacgt tcaaggaaga catccagaag gcacaggtga gcgggcaggg tgactccctc 2160  
 cacgaacaca tcgccaacct ggccggctcg ccggccatta aaaaggcat cctgcagacg 2220  
 gtcaaggtcg tcgacgagct cgtgaaggtg atgggccggc acaagcccga aaatatcgtc 2280  
 atagagatgg ccaggagaaa ccagaccacc caaaaagggc agaagaactc gcgcgagcgg 2340  
  
 atgaaacgga tcgaggagg cattaaagag ctccgggtccc agatcctgaa ggagcacccc 2400  
 gtggaaaata cccagctcca gaatgaaaag ctctacctct actacctgca gaacggccgc 2460  
 gacatgtacg tggaccagga gctggacatt aatcggctat cggactacga cgtcgaccac 2520  
 atcgtgccgc agtcgttctt caaggacgat agcatcgaca acaaggtgct caccgggtcg 2580  
 gataaaaaatc ggggcaagag cgacaacgtg cccagcgagg aggtcgtgaa gaagatgaaa 2640  
 aactactggc gccagctcct caacgcgaaa ctgatcacc agcgcaagtt cgacaacctg 2700

acgaaggcgg aacgcggtgg cttgagcga ctcgataagg cgggcttcat aaaaaggcag 2760

ctggtcgaga cgcgccagat cacgaagcat gtcgccaga tcctggacag ccgcatgaat 2820

actaagtacg atgaaaacga caagetgac cgggaggtga aggtgatcac getgaagtec 2880

aagctcgtgt cggacttccg caaggacttc cagttctaca aggtccgcga gatcaacaac 2940

taccaccacg cccacgacg ctacctgaat gcggtggtcg ggaccgcct gatcaagaag 3000

taccgaage tggagtcgga gttcgtgtac ggcgactaca aggtctacga cgtgcgcaaa 3060

atgatcgcca agtccgagca ggagatcggc aaggccacgg caaaatactt cttctactcg 3120

aacatcatga acttcttcaa gaccgagatc accctcgcga acggcgagat ccgcaagcgc 3180

ccgctcatcg aaaccaacgg cgagacgggc gagatcgtct gggataaggg cgggatttc 3240

gcgacggtcc gcaaggtgct ctccatgccg caagtcaata tcgtgaaaa gacggaggtc 3300

cagacgggcg ggttcagcaa ggagtcctac ctcccgaagc gcaactccga caagctcatc 3360

gcgaggaaga aggattggga cccgaaaaaa tatggcggtc tcgacagccc gaccgtcgca 3420

tacagcgtcc tcgtcgtggc gaaggtggag aaggcgaagt caaagaagct caagtccgtg 3480

aaggagctgc tcgggatcac gattatggag cggtcctcct tcgagaagaa cccgatcgac 3540

ttcctagagg ccaagggata taaggaggtc aagaaggacc tgattattaa actgccgaag 3600

tactcgtct tcgagctgga aaacggccgc aagaggatgc tcgcctccgc aggcgagttg 3660

cagaagggca acgagctcgc cctcccagc aaatacgtca atttctgta cctcgctagc 3720

cactatgaaa agctcaaggg cagcccggag gacaacgagc agaagcagct cttcgtggag 3780

cagcacaagc attacctgga cgagatcatc gagcagatca gcgagttctc gaagcgggtg 3840

atcctcgccg acgcgaacct ggacaagggt ctgtcggcat ataacaagca ccgcgacaaa 3900

ccaatcgcg agcaggccga aaatatcatc cacctcttca ccctaccaa cctcggcgtc 3960

ccggcagcct tcaagtactt cgacaccacg attgaccgga agcggtagac gagcacgaag 4020

gagtgctcg atgcgacgt gatccaccag agcatcacag ggctctatga aacacgcatc 4080

gacctgagcc agctgggagg agac 4104

<210> 45

<211> 1368

<212> PRT

<213> Artificial Sequence

<220><223> protein expressed from synthetic construct

<400> 45

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



1010                      1015                      1020  
 Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser  
 1025                      1030                      1035                      1040  
  
 Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu  
                                  1045                      1050                      1055  
 Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile  
                                  1060                      1065                      1070  
 Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser  
                                  1075                      1080                      1085  
 Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly  
                                  1090                      1095                      1100  
 Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile  
  
 1105                      1110                      1115                      1120  
 Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser  
                                  1125                      1130                      1135  
 Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly  
                                  1140                      1145                      1150  
 Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile  
                                  1155                      1160                      1165  
 Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala  
                                  1170                      1175                      1180  
  
 Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys  
 1185                      1190                      1195                      1200  
 Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser  
                                  1205                      1210                      1215  
 Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr  
                                  1220                      1225                      1230  
 Val 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 His  
  
 1250                      1255                      1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val  
 1265                      1270                      1275                      1280  
 Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys  
                                  1285                      1290                      1295  
 His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu  
                                  1300                      1305                      1310  
 Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp  
                                  1315                      1320                      1325

Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp  
                                  1330                      1335                      1340  
 Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile  
 1345                      1350                      1355                      1360  
 Asp Leu Ser Gln Leu Gly Gly Asp  
                                  1365

- <210> 46
- <211> 7
- <212> PRT
- <213> SV40
- <400> 46

Pro Lys Lys Lys Arg Lys Val  
 1                      5

- <210> 47
- <211> 7
- <212> PRT
- <213> T. reesei
- <400>

> 47

Lys Lys Lys Lys Leu Lys Leu  
 1                      5

- <210> 48
- <211> 1388
- <212> PRT
- <213> Streptococcus thermophilus
- <400> 48

Met Thr Lys Pro Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val  
 1                      5                      10                      15  
 Gly Trp Ala Val Thr Thr Asp Asn Tyr Lys Val Pro Ser Lys Lys Met  
                     20                      25                      30  
 Lys Val Leu Gly Asn Thr Ser Lys Lys Tyr Ile Lys Lys Asn Leu Leu  
                     35                      40                      45  
  
 Gly Val Leu Leu Phe Asp Ser Gly Ile Thr Ala Glu Gly Arg Arg Leu  
                     50                      55                      60  
 Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Arg Asn Arg Ile Leu  
                     65                      70                      75                      80  
 Tyr Leu Gln Glu Ile Phe Ser Thr Glu Met Ala Thr Leu Asp Asp Ala  
                     85                      90                      95  
 Phe Phe Gln Arg Leu Asp Asp Ser Phe Leu Val Pro Asp Asp Lys Arg  
                     100                      105                      110  
 Asp Ser Lys Tyr Pro Ile Phe Gly Asn Leu Val Glu Glu Lys Ala Tyr  
  
                     115                      120                      125  
 His Asp Glu Phe Pro Thr Ile Tyr His Leu Arg Lys Tyr Leu Ala Asp  
                     130                      135                      140  
 Ser Thr Lys Lys Ala Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala His  
                     145                      150                      155                      160  
 Met Ile Lys Tyr Arg Gly His Phe Leu Ile Glu Gly Glu Phe Asn Ser  
                     165                      170                      175  
 Lys Asn Asn Asp Ile Gln Lys Asn Phe Gln Asp Phe Leu Asp Thr Tyr  
                     180                      185                      190  
  
 Asn Ala Ile Phe Glu Ser Asp Leu Ser Leu Glu Asn Ser Lys Gln Leu  
                     195                      200                      205  
 Glu Glu Ile Val Lys Asp Lys Ile Ser Lys Leu Glu Lys Lys Asp Arg  
                     210                      215                      220  
 Ile Leu Lys Leu Phe Pro Gly Glu Lys Asn Ser Gly Ile Phe Ser Glu  
                     225                      230                      235                      240  
 Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp Phe Arg Lys Cys Phe  
                     245                      250                      255

Asn Leu Asp Glu Lys Ala Ser Leu His Phe Ser Lys Glu Ser Tyr Asp  
 260 265 270  
 Glu Asp Leu Glu Thr Leu Leu Gly Tyr Ile Gly Asp Asp Tyr Ser Asp  
 275 280 285  
 Val Phe Leu Lys Ala Lys Lys Leu Tyr Asp Ala Ile Leu Leu Ser Gly  
 290 295 300  
 Phe Leu Thr Val Thr Asp Asn Glu Thr Glu Ala Pro Leu Ser Ser Ala  
 305 310 315 320  
 Met Ile Lys Arg Tyr Asn Glu His Lys Glu Asp Leu Ala Leu Leu Lys  
 325 330 335  
 Glu Tyr Ile Arg Asn Ile Ser Leu Lys Thr Tyr Asn Glu Val Phe Lys  
 340 345 350  
 Asp Asp Thr Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Lys Thr Asn  
 355 360 365  
 Gln Glu Asp Phe Tyr Val Tyr Leu Lys Lys Leu Leu Ala Glu Phe Glu  
 370 375 380  
 Gly Ala Asp Tyr Phe Leu Glu Lys Ile Asp Arg Glu Asp Phe Leu Arg  
 385 390 395 400  
 Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro Tyr Gln Ile His Leu  
 405 410 415  
 Gln Glu Met Arg Ala Ile Leu Asp Lys Gln Ala Lys Phe Tyr Pro Phe  
 420 425 430  
 Leu Ala Lys Asn Lys Glu Arg Ile Glu Lys Ile Leu Thr Phe Arg Ile  
 435 440 445  
 Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Asp Phe Ala Trp  
 450 455 460  
 Ser Ile Arg Lys Arg Asn Glu Lys Ile Thr Pro Trp Asn Phe Glu Asp  
 465 470 475 480  
 Val Ile Asp Lys Glu Ser Ser Ala Glu Ala Phe Ile Asn Arg Met Thr  
 485 490 495  
 Ser Phe Asp Leu Tyr Leu Pro Glu Glu Lys Val Leu Pro Lys His Ser



755                                      760                                      765  
 Gln Tyr Thr Asn Gln Gly Lys Ser Asn Ser Gln Gln Arg Leu Lys Arg  
 770                                      775                                      780  
 Leu Glu Lys Ser Leu Lys Glu Leu Gly Ser Lys Ile Leu Lys Glu Asn  
 785                                      790                                      795                                      800  
 Ile Pro Ala Lys Leu Ser Lys Ile Asp Asn Asn Ala Leu Gln Asn Asp  
 805                                      810                                      815  
 Arg Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Lys Asp Met Tyr Thr Gly  
 820                                      825                                      830  
 Asp Asp Leu Asp Ile Asp Arg Leu Ser Asn Tyr Asp Ile Asp His Ile  
 835                                      840                                      845  
 Ile Pro Gln Ala Phe Leu Lys Asp Asn Ser Ile Asp Asn Lys Val Leu  
 850                                      855                                      860  
 Val Ser Ser Ala Ser Asn Arg Gly Lys Ser Asp Asp Val Pro Ser Leu  
 865                                      870                                      875                                      880  
 Glu Val Val Lys Lys Arg Lys Thr Phe Trp Tyr Gln Leu Leu Lys Ser  
 885                                      890                                      895  
 Lys Leu Ile Ser Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg  
 900                                      905                                      910  
 Gly Gly Leu Ser Pro Glu Asp Lys Ala Gly Phe Ile Gln Arg Gln Leu  
 915                                      920                                      925  
 Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Arg Leu Leu Asp Glu  
 930                                      935                                      940  
 Lys Phe Asn Asn Lys Lys Asp Glu Asn Asn Arg Ala Val Arg Thr Val  
 945                                      950                                      955                                      960  
 Lys Ile Ile Thr Leu Lys Ser Thr Leu Val Ser Gln Phe Arg Lys Asp  
 965                                      970                                      975  
 Phe Glu Leu Tyr Lys Val Arg Glu Ile Asn Asp Phe His His Ala His  
 980                                      985                                      990  
 Asp Ala Tyr Leu Asn Ala Val Val Ala Ser Ala Leu Leu Lys Lys Tyr  
 995                                      1000                                      1005

Pro Lys Leu Glu Pro Glu Phe Val Tyr Gly Asp Tyr Pro Lys Tyr Asn  
 1010                      1015                      1020

Ser Phe Arg Glu Arg Lys Ser Ala Thr Glu Lys Val Tyr Phe Tyr Ser  
 1025                      1030                      1035                      1040

Asn Ile Met Asn Ile Phe Lys Lys Ser Ile Ser Leu Ala Asp Gly Arg  
                             1045                      1050                      1055

Val Ile Glu Arg Pro Leu Ile Glu Val Asn Glu Glu Thr Gly Glu Ser  
                             1060                      1065                      1070

Val Trp Asn Lys Glu Ser Asp Leu Ala Thr Val Arg Arg Val Leu Ser  
                             1075                      1080                      1085

Tyr Pro Gln Val Asn Val Val Lys Lys Val Glu Glu Gln Asn His Gly  
                             1090                      1095                      1100

Leu Asp Arg Gly Lys Pro Lys Gly Leu Phe Asn Ala Asn Leu Ser Ser  
 1105                      1110                      1115                      1120

Lys Pro Lys Pro Asn Ser Asn Glu Asn Leu Val Gly Ala Lys Glu Tyr  
                             1125                      1130                      1135

Leu Asp Pro Lys Lys Tyr Gly Gly Tyr Ala Gly Ile Ser Asn Ser Phe  
                             1140                      1145                      1150

Thr Val Leu Val Lys Gly Thr Ile Glu Lys Gly Ala Lys Lys Lys Ile  
                             1155                      1160                      1165

Thr Asn Val Leu Glu Phe Gln Gly Ile Ser Ile Leu Asp Arg Ile Asn  
                             1170                      1175                      1180

Tyr Arg Lys Asp Lys Leu Asn Phe Leu Leu Glu Lys Gly Tyr Lys Asp  
 1185                      1190                      1195                      1200

Ile Glu Leu Ile Ile Glu Leu Pro Lys Tyr Ser Leu Phe Glu Leu Ser  
                             1205                      1210                      1215

Asp Gly Ser Arg Arg Met Leu Ala Ser Ile Leu Ser Thr Asn Asn Lys  
                             1220                      1225                      1230

Arg Gly Glu Ile His Lys Gly Asn Gln Ile Phe Leu Ser Gln Lys Phe  
                             1235                      1240                      1245

Val Lys Leu Leu Tyr His Ala Lys Arg Ile Ser Asn Thr Ile Asn Glu  
                             1250                      1255                      1260

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

1265                      1270                      1275                      1280

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

1285                      1290                      1295

Asn Gly Lys Leu Leu Asn Ser Ala Phe Gln Ser Trp Gln Asn His Ser

1300                      1305                      1310

Ile Asp Glu Leu Cys Ser Ser Phe Ile Gly Pro Thr Gly Ser Glu Arg

1315                      1320                      1325

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

1330                      1335                      1340

Phe Leu Gly Val Lys Ile Pro Arg Tyr Arg Asp Tyr Thr Pro Ser Ser

1345                      1350                      1355                      1360

Leu Leu Lys Asp Ala Thr Leu Ile His Gln Ser Val Thr Gly Leu Tyr

1365                      1370                      1375

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

1380                      1385

<210> 49

<211> 1345

<212> PRT

<213> Streptococcus mutans

<400> 49

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 Arg Tyr Thr Arg Arg Arg Asn Arg Ile Leu

65                      70                      75                      80



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



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 Ala  
  
 1010                                      1015                                      1020  
 Thr Ala Lys Lys Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Lys  
 1025                                      1030                                      1035                                      1040  
 Asp Asp Val Arg Thr Asp Lys Asn Gly Glu Ile Ile Trp Lys Lys Asp  
                                     1045                                      1050                                      1055  
 Glu His Ile Ser Asn Ile Lys Lys Val Leu Ser Tyr Pro Gln Val Asn  
                                     1060                                      1065                                      1070  
 Ile Val Lys Lys Val Glu Glu Gln Thr Gly Gly Phe Ser Lys Glu Ser  
                                     1075                                      1080                                      1085

Ile Leu Pro Lys Gly Asn Ser Asp Lys Leu Ile Pro Arg Lys Thr Lys  
 1090                      1095                      1100

Lys Phe Tyr Trp Asp Thr Lys Lys Tyr Gly Gly Phe Asp Ser Pro Ile  
 1105                      1110                      1115                      1120

Val Ala Tyr Ser Ile Leu Val Ile Ala Asp Ile Glu Lys Gly Lys Ser  
                                  1125                      1130                      1135

Lys Lys Leu Lys Thr Val Lys Ala Leu Val Gly Val Thr Ile Met Glu  
                                  1140                      1145                      1150

Lys Met Thr Phe Glu Arg Asp Pro Val Ala Phe Leu Glu Arg Lys Gly  
                                  1155                      1160                      1165

Tyr Arg Asn Val Gln Glu Glu Asn Ile Ile Lys Leu Pro Lys Tyr Ser  
 1170                      1175                      1180

Leu Phe Lys Leu Glu Asn Gly Arg Lys Arg Leu Leu Ala Ser Ala Arg  
 1185                      1190                      1195                      1200

Glu Leu Gln Lys Gly Asn Glu Ile Val Leu Pro Asn His Leu Gly Thr  
                                  1205                      1210                      1215

Leu Leu Tyr His Ala Lys Asn Ile His Lys Val Asp Glu Pro Lys His  
                                  1220                      1225                      1230

Leu 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 Leu  
 1250                      1255                      1260

Glu Lys Ile Lys Glu Leu Tyr Ala Gln Asn Asn Gly Glu Asp Leu Lys  
 1265                      1270                      1275                      1280

Glu Leu Ala Ser Ser Phe Ile Asn Leu Leu Thr Phe Thr Ala Ile Gly  
                                  1285                      1290                      1295

Ala Pro Ala Thr Phe Lys Phe Phe Asp Lys Asn Ile Asp Arg Lys Arg  
                                  1300                      1305                      1310

Tyr Thr Ser Thr Thr Glu Ile Leu Asn Ala Thr Leu Ile His Gln Ser  
                                  1315                      1320                      1325

Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Asn Lys Leu Gly Gly  
                                  1330                      1335                      1340

Asp

134

<210> 50

<211> 984

<212> PRT

<213> *Campylobacter jejuni*

<400> 50

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



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>    51  
 <211>    1082  
 <212>    PRT  
 <213>    Neisseria meningitides  
 <400>    51  
 Met Ala Ala Phe Lys Pro Asn Ser 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 Glu  
                           20                    25                    30  
 Glu Asn Pro Ile Arg 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 Thr Arg Arg Leu Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asn  
                           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 Gly 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 Ser 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  
 305 310 315 320

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  
 385 390 395 400

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  
 465 470 475 480

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 Leu 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 His  
                                  1010                      1015                      1020  
 Arg Gly Thr Gly Asn Ile Asn Ile Arg Ile His Asp Leu Asp His Lys  
  
 1025                      1030                      1035                      1040  
 Ile Gly Lys Asn Gly Ile Leu Glu Gly Ile Gly Val Lys Thr Ala Leu  
                                  1045                      1050                      1055  
 Ser Phe Gln Lys Tyr Gln Ile Asp Glu Leu Gly Lys Glu Ile Arg Pro  
                                  1060                      1065                      1070  
 Cys Arg Leu Lys Lys Arg Pro Pro Val Arg  
                                  1075                      1080  
 <210>    52  
 <211>    1629  
 <212>    PRT  
 <213>    Francisella tularensis  
 <400>    52  
 Met Asn Phe Lys Ile Leu Pro Ile Ala Ile Asp Leu Gly Val Lys Asn  
  
                                  1                      5                      10                      15  
 Thr Gly Val Phe Ser Ala Phe Tyr Gln Lys Gly Thr Ser Leu Glu Arg  
                                  20                      25                      30  
 Leu Asp Asn Lys Asn Gly Lys Val Tyr Glu Leu Ser Lys Asp Ser Tyr  
                                  35                      40                      45  
 Thr Leu Leu Met Asn Asn Arg Thr Ala Arg Arg His Gln Arg Arg Gly  
                                  50                      55                      60



305                    310                    315                    320  
 Asn Val Leu Asp Glu Asn Asn His Gln Glu Gly Tyr Leu Lys Asn Phe  
    325                    330                    335  
 Cys Glu Asn Leu His Asn Lys Lys Tyr Ser Asn Leu Ser Val Lys Asn  
    340                    345                    350  
 Leu Val Asn Leu Ile Gly Asn Leu Ser Asn Leu Glu Leu Lys Pro Leu  
    355                    360                    365  
  
 Arg Lys Tyr Phe Asn Asp Lys Ile His Ala Lys Ala Asp His Trp Asp  
    370                    375                    380  
 Glu Gln Lys Phe Thr Glu Thr Tyr Cys His Trp Ile Leu Gly Glu Trp  
 385                    390                    395                    400  
 Arg Val Gly Val Lys Asp Gln Asp Lys Lys Asp Gly Ala Lys Tyr Ser  
    405                    410                    415  
 Tyr Lys Asp Leu Cys Asn Glu Leu Lys Gln Lys Val Thr Lys Ala Gly  
    420                    425                    430  
 Leu Val Asp Phe Leu Leu Glu Leu Asp Pro Cys Arg Thr Ile Pro Pro  
  
    435                    440                    445  
 Tyr Leu Asp Asn Asn Asn Arg Lys Pro Pro Lys Cys Gln Ser Leu Ile  
    450                    455                    460  
 Leu Asn Pro Lys Phe Leu Asp Asn Gln Tyr Pro Asn Trp Gln Gln Tyr  
 465                    470                    475                    480  
 Leu Gln Glu Leu Lys Lys Leu Gln Ser Ile Gln Asn Tyr Leu Asp Ser  
    485                    490                    495  
 Phe Glu Thr Asp Leu Lys Val Leu Lys Ser Ser Lys Asp Gln Pro Tyr  
    500                    505                    510  
  
 Phe Val Glu Tyr Lys Ser Ser Asn Gln Gln Ile Ala Ser Gly Gln Arg  
    515                    520                    525  
 Asp Tyr Lys Asp Leu Asp Ala Arg Ile Leu Gln Phe Ile Phe Asp Arg  
    530                    535                    540  
 Val Lys Ala Ser Asp Glu Leu Leu Leu Asn Glu Ile Tyr Phe Gln Ala  
 545                    550                    555                    560  
 Lys Lys Leu Lys Gln Lys Ala Ser Ser Glu Leu Glu Lys Leu Glu Ser



Cys Ser Ala Asp Asn Ala His Arg Met Gln Gln Ile Lys Ile Thr Glu  
 820 825 830  
 Pro Val Glu Asp Asn Lys Asp Lys Ile Ile Leu Ser Ala Lys Ala Gln  
 835 840 845  
 Arg Leu Pro Ala Ile Pro Thr Arg Ile Val Asp Gly Ala Val Lys Lys  
 850 855 860  
 Met Ala Thr Ile Leu Ala Lys Asn Ile Val Asp Asp Asn Trp Gln Asn  
 865 870 875 880  
 Ile Lys Gln Val Leu Ser Ala Lys His Gln Leu His Ile Pro Ile Ile  
 885 890 895  
 Thr Glu Ser Asn Ala Phe Glu Phe Glu Pro Ala Leu Ala Asp Val Lys  
 900 905 910  
 Gly Lys Ser Leu Lys Asp Arg Arg Lys Lys Ala Leu Glu Arg Ile Ser  
 915 920 925  
 Pro Glu Asn Ile Phe Lys Asp Lys Asn Asn Arg Ile Lys Glu Phe Ala  
 930 935 940  
 Lys Gly Ile Ser Ala Tyr Ser Gly Ala Asn Leu Thr Asp Gly Asp Phe  
 945 950 955 960  
 Asp Gly Ala Lys Glu Glu Leu Asp His Ile Ile Pro Arg Ser His Lys  
 965 970 975  
 Lys Tyr Gly Thr Leu Asn Asp Glu Ala Asn Leu Ile Cys Val Thr Arg  
 980 985 990  
 Gly Asp Asn Lys Asn Lys Gly Asn Arg Ile Phe Cys Leu Arg Asp Leu  
 995 1000 1005  
 Ala Asp Asn Tyr Lys Leu Lys Gln Phe Glu Thr Thr Asp Asp Leu Glu  
 1010 1015 1020  
 Ile Glu Lys Lys Ile Ala Asp Thr Ile Trp Asp Ala Asn Lys Lys Asp  
 1025 1030 1035 1040  
 Phe Lys Phe Gly Asn Tyr Arg Ser Phe Ile Asn Leu Thr Pro Gln Glu  
 1045 1050 1055  
 Gln Lys Ala Phe Arg His Ala Leu Phe Leu Ala Asp Glu Asn Pro Ile  
 1060 1065 1070

Lys Gln Ala Val Ile Arg Ala Ile Asn Asn Arg Asn Arg Thr Phe Val  
 1075 1080 1085

Asn Gly Thr Gln Arg Tyr Phe Ala Glu Val Leu Ala Asn Asn Ile Tyr  
 1090 1095 1100

Leu Arg Ala Lys Lys Glu Asn Leu Asn Thr Asp Lys Ile Ser Phe Asp  
 1105 1110 1115 1120

Tyr Phe Gly Ile Pro Thr Ile Gly Asn Gly Arg Gly Ile Ala Glu Ile  
 1125 1130 1135

Arg Gln Leu Tyr Glu Lys Val Asp Ser Asp Ile Gln Ala Tyr Ala Lys  
 1140 1145 1150

Gly Asp Lys Pro Gln Ala Ser Tyr Ser His Leu Ile Asp Ala Met Leu  
 1155 1160 1165

Ala Phe Cys Ile Ala Ala Asp Glu His Arg Asn Asp Gly Ser Ile Gly  
 1170 1175 1180

Leu Glu Ile Asp Lys Asn Tyr Ser Leu Tyr Pro Leu Asp Lys Asn Thr  
 1185 1190 1195 1200

Gly Glu Val Phe Thr Lys Asp Ile Phe Ser Gln Ile Lys Ile Thr Asp  
 1205 1210 1215

Asn Glu Phe Ser Asp Lys Lys Leu Val Arg Lys Lys Ala Ile Glu Gly  
 1220 1225 1230

Phe Asn Thr His Arg Gln Met Thr Arg Asp Gly Ile Tyr Ala Glu Asn  
 1235 1240 1245

Tyr Leu Pro Ile Leu Ile His Lys Glu Leu Asn Glu Val Arg Lys Gly  
 1250 1255 1260

Tyr Thr Trp Lys Asn Ser Glu Glu Ile Lys Ile Phe Lys Gly Lys Lys  
 1265 1270 1275 1280

Tyr Asp Ile Gln Gln Leu Asn Asn Leu Val Tyr Cys Leu Lys Phe Val  
 1285 1290 1295

Asp Lys Pro Ile Ser Ile Asp Ile Gln Ile Ser Thr Leu Glu Glu Leu  
 1300 1305 1310

Arg Asn Ile Leu Thr Thr Asn Asn Ile Ala Ala Thr Ala Glu Tyr Tyr

1315                      1320                      1325  
 Tyr Ile Asn Leu Lys Thr Gln Lys Leu His Glu Tyr Tyr Ile Glu Asn  
 1330                      1335                      1340  
 Tyr Asn Thr Ala Leu Gly Tyr Lys Lys Tyr Ser Lys Glu Met Glu Phe  
 1345                      1350                      1355                      1360  
 Leu Arg Ser Leu Ala Tyr Arg Ser Glu Arg Val Lys Ile Lys Ser Ile  
                             1365                      1370                      1375  
  
 Asp Asp Val Lys Gln Val Leu Asp Lys Asp Ser Asn Phe Ile Ile Gly  
                             1380                      1385                      1390  
 Lys Ile Thr Leu Pro Phe Lys Lys Glu Trp Gln Arg Leu Tyr Arg Glu  
                             1395                      1400                      1405  
 Trp Gln Asn Thr Thr Ile Lys Asp Asp Tyr Glu Phe Leu Lys Ser Phe  
                             1410                      1415                      1420  
 Phe Asn Val Lys Ser Ile Thr Lys Leu His Lys Lys Val Arg Lys Asp  
 1425                      1430                      1435                      1440  
 Phe Ser Leu Pro Ile Ser Thr Asn Glu Gly Lys Phe Leu Val Lys Arg  
  
                             1445                      1450                      1455  
 Lys Thr Trp Asp Asn Asn Phe Ile Tyr Gln Ile Leu Asn Asp Ser Asp  
                             1460                      1465                      1470  
 Ser Arg Ala Asp Gly Thr Lys Pro Phe Ile Pro Ala Phe Asp Ile Ser  
                             1475                      1480                      1485  
 Lys Asn Glu Ile Val Glu Ala Ile Ile Asp Ser Phe Thr Ser Lys Asn  
                             1490                      1495                      1500  
 Ile Phe Trp Leu Pro Lys Asn Ile Glu Leu Gln Lys Val Asp Asn Lys  
 1505                      1510                      1515                      1520  
  
 Asn Ile Phe Ala Ile Asp Thr Ser Lys Trp Phe Glu Val Glu Thr Pro  
                             1525                      1530                      1535  
 Ser Asp Leu Arg Asp Ile Gly Ile Ala Thr Ile Gln Tyr Lys Ile Asp  
                             1540                      1545                      1550  
 Asn Asn Ser Arg Pro Lys Val Arg Val Lys Leu Asp Tyr Val Ile Asp  
                             1555                      1560                      1565  
 Asp Asp Ser Lys Ile Asn Tyr Phe Met Asn His Ser Leu Leu Lys Ser

1570                      1575                      1580  
 Arg Tyr Pro Asp Lys Val Leu Glu Ile Leu Lys Gln Ser Thr Ile Ile  
  
 1585                      1590                      1595                      1600  
 Glu Phe Glu Ser Ser Gly Phe Asn Lys Thr Ile Lys Glu Met Leu Gly  
  
                             1605                      1610                      1615  
 Met Lys Leu Ala Gly Ile Tyr Asn Glu Thr Ser Asn Asn  
  
                             1620                      1625  
 <210>    53  
 <211>    1056  
 <212>    PRT  
 <213>    Pasteurella multocida  
 <400>    53  
 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



aacgaaatgg ccaaggtaga tgatagtttc ttccatcgtc tcgaagaatc ttttctggtt 480  
 gaggaagata aaaaacacga acgtcacctc atctttggca atatcgtgga tgaagtggcc 540  
  
 tatcatgaaa aataccctac gatttateat cttcgcaaga agttggttga tagtacggac 600  
 aaagcggatc tgcgtttaat ctatcttgcg ttagcgcaca tgatcaaatt tcgtggtcat 660  
 ttcttaattg aaggtgatct gaatcctgat aactctgatg tggacaaatt gtttatacaa 720  
 ttagtgcaaa cctataatca gctgttcgag gaaaacceca ttaatgcctc tggagtggat 780  
 gccaaagcga ttttaagcgc gagactttct aagtcccggc gtctggagaa tctgatcgcc 840  
 cagttaccag gggaaaagaa aaatggtctg tttggtaatc tgattgccct cagtctgggg 900  
 cttaccccgga acitcaaatc caattttgac ctggctgagg acgcaaagct gcagctgagc 960  
  
 aaagatactt atgatgatga cctcgacaat ctgctcgccc agattggtga ccaatatgcg 1020  
 gatctgtttc tggcagcgaa gaatctttcg gatgctatct tgctgtcgga tattctgcgt 1080  
 gttaataccg aatcaccaa agcgcctctg tctgcaagta tgatcaagag atacgacgag 1140  
 caccaccagg acctgactct tcttaaggca ctggtacgcc aacagcttcc ggagaaatac 1200  
 aaagaaatat tcttcgacca gtccaagaat ggttacgagg gctacatcga tgggtgtgca 1260  
 tcacaggaag agttctataa atttattaaa ccaatccttg agaaaatgga tggcacggaa 1320  
 gagttacttg ttaaacftaa ccgcaagac ttgcttagaa agcaacgtac attcgacaac 1380  
  
 ggctccatcc cacaccagat tcatttaggt gaacttcacg ccatcttgcg cagacaagaa 1440  
 gatttctate ctttcttaaa agacaatcgg gagaaaatcg agaagatcct gacgttccgc 1500  
 attcctatt atgtcggfcc cctggcacgt ggtaatctc ggtttgcctg gatgacgcgc 1560  
 aaaagtgagg aaaccatcac ccttgggaac ttgaagaag tcgtggataa aggtgctagc 1620  
 gcgcagctct ttagaagaa aatgacgaac ttcgataaaa acttgcccaa cgaaaaagtc 1680  
 ctgccaage actctctttt atatgagtac ttactgtgt acaacgaact gactaaagtg 1740  
 aaatacgtta cggaaggat gcgcaaacct gcctttctta gtggcgagca gaaaaagca 1800  
  
 attgtcgatc ttctctttaa aacgaatcgc aaggtaactg taaaacagct gaaggaagat 1860  
 tatttcaaaa agatcgaatg ctttgattct gtcgagatct cgggtgtcga agatcgtttc 1920  
 aacgcttctc tagggaccta tcatgatttg ctgaagataa taaaagacaa agactttctc 1980  
 gacaatgaag aaaaatgaaga tattctggag gatattgttt tgaccttgac ctatttcgaa 2040  
 gatagagaga tgatcgagga gcgcttaaaa acctatgccc acctgtttga tgacaaagtc 2100  
 atgaagcaat taaagcggc cagatatacg ggggtggggc gcttgagccg caagttgatt 2160

aacggtatta gagacaagca gagcggaaaa actatcctgg atttctctcaa atctgacgga 2220  
  
 ttgcgaacc gcaatthtat gcagcttata catgatgatt cgcttacatt caaagaggat 2280  
 attcagaagg ctcagggtgc tgggcaaggt gattcactcc acgaacatat agcaaatttg 2340  
 gccggctctc ctgcgattaa gaaggggatc ctgcaaacag ttaaagttgt ggatgaactt 2400  
 gtaaaagtaa tgggccgcca caagccggag aatatcgtga tagaaatggc gcgcgagaat 2460  
 caaacgacac aaaaaggcca aaagaactca agagagagaa tgaagcgcat tgaggagggg 2520  
 ataaaggaac ttggatctca aattctgaaa gaacatccag ttgaaaacac tcagctgcaa 2580  
 aatgaaaaat tgtacctgta ctacctgcag aatggaagag acatgtacgt ggatcaggaa 2640  
  
 ttggatatca atagactctc ggactatgac gtagatcaca ttgtccctca gagcttctc 2700  
 aaggatgatt ctatagataa taaagtactt acgagatcgg acaaaaatcg cggtaaatcg 2760  
 gataacgtcc catcggagga agtcgttaaa aagatgaaaa actatggcgc tcaactgctg 2820  
 aacgccaagc tgatcacaca gcgtaagttt gataatctga ctaaagccga acgcggtggt 2880  
 cttagtgaac tcgataaagc aggatttata aaacggcagt tagtagaaac gcgccaatt 2940  
 acgaaacacg tggtcagat cctcgattct agaatgaata caaagtacga tgaaaacgat 3000  
 aaactgatcc gtgaagtaaa agtcattacc ttaaaatcta aacttgtgtc cgatttccgc 3060  
  
 aaagatthtc agthtttaca ggtccgggaa atcaataact atcacatgc acatgatgca 3120  
 tatttaaatg cggttgtagg cacggccctt attaagaaat accctaaact cgaaagtgag 3180  
 ttgtttatg gggattataa agtgtatgac gttcgcaaaa tgatcgcgaa atcagaacag 3240  
 gaaatcggtg aggctaccgc taaatacttt thttattcca acattatgaa thtttttaag 3300  
 accgaaataa ctctcgcgaa tgggtgaaatc cgtaaacggc ctcttataga aaccaatggt 3360  
 gaaacgggag aaatcgtttg ggataaaggt cgtgactttg ccaccgttcg taaagtcctc 3420  
 tcaatgccgc aagthaacat tgtcaagaag acggaagttc aaacaggggg atttctccaaa 3480  
  
 gaatctatcc tgccgaagcg taacagtgat aaacttattg ccagaaaaaa agattgggat 3540  
 ccaaaaaaat acggaggctt tgattcccct accgtcgcgt atagtgtgct ggtggttgct 3600  
 aaagtcgaga aagggaaaag caagaaattg aatcagttg aagaactgct gggatttaca 3660  
 attatggaaa gatcgtcctt tgagaaaaat ccgatcgact thtttagagc caaggggtat 3720  
 aaggaagtga aaaaagatct catcatcaaa ttaccgaagt atagtctttt tgagctggaa 3780  
 aacggcagaa aaagaatgct ggcctccgcg ggcgagttac agaagggaaa tgagctggcg 3840  
 ctgccttcca aatatgttaa thttctgtac cttgccagtc attatgagaa actgaagggc 3900

agccccgaag ataacgaaca gaaacaatta ttcgtggaac agcataagca ctatttagat 3960  
 gaaattatag agcaaattag tgaattttct aagcgcgta tcctcgcgga tgctaattta 4020  
 gacaaagtac tgtcagctta taataaacat cgggataagc cgattagaga acaggccgaa 4080  
 aatatcattc atttgtttac cttaaccaac cttggagcac cagctgcctt caaatatttc 4140  
 gataccacaa ttgatcgtaa acggtataca agtacaaaag aagtcttgga cgcaaccctc 4200  
 attcatcaat ctattactgg attatatgag acacgcattg atctttcaca gctgggcgga 4260  
 gacaagaaga aaaaactgaa actg 4284

<210> 55  
 <211> 1428  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> protein expressed from synthetic construct  
 <400> 55

Met His His His His His His Ser Ser Gly Leu Val Pro Arg Gly Ser  
 1 5 10 15  
 Gly Met Lys Glu Thr Ala Ala Ala Lys Phe Glu Arg Gln His Met Asp  
 20 25 30  
 Ser Pro Asp Leu Gly Thr Asp Asp Asp Asp Lys Ala Met Ala Pro Lys  
 35 40 45  
  
 Lys Lys Arg Lys Val Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile  
 50 55 60  
 Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val  
 65 70 75 80  
 Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile  
 85 90 95  
 Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala  
 100 105 110  
 Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg  
 115 120 125  
 Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala  
 130 135 140  
 Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val

145                      150                      155                      160  
 Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val  
    165                      170                      175  
 Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg  
    180                      185                      190  
  
 Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr  
    195                      200                      205  
 Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu  
    210                      215                      220  
 Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln  
 225                      230                      235                      240  
 Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala  
    245                      250                      255  
 Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser  
  
    260                      265                      270  
 Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn  
    275                      280                      285  
 Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn  
    290                      295                      300  
 Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser  
 305                      310                      315                      320  
 Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly  
    325                      330                      335  
  
 Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala  
    340                      345                      350  
 Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala  
    355                      360                      365  
 Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp  
    370                      375                      380  
 Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr  
 385                      390                      395                      400

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

645                                  650                                  655  
 Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile  
 660                                  665                                  670  
 Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg  
 675                                  680                                  685  
 Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu  
  
 690                                  695                                  700  
 Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile  
 705                                  710                                  715                                  720  
 Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu  
 725                                  730                                  735  
 Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp  
 740                                  745                                  750  
 Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly  
 755                                  760                                  765  
  
 Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro  
 770                                  775                                  780  
 Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu  
 785                                  790                                  795                                  800  
 Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met  
 805                                  810                                  815  
 Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu  
 820                                  825                                  830  
 Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile  
  
 835                                  840                                  845  
 Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu  
 850                                  855                                  860  
 Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu  
 865                                  870                                  875                                  880  
 Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro  
 885                                  890                                  895  
 Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg



Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn  
 1155 1160 1165

Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr  
 1170 1175 1180

Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala  
 1185 1190 1195 1200

Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu  
 1205 1210 1215

Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile  
 1220 1225 1230

Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile  
 1235 1240 1245

Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys  
 1250 1255 1260

Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala  
 1265 1270 1275 1280

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu  
 1285 1290 1295

Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val  
 1300 1305 1310

Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu  
 1315 1320 1325

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu  
 1330 1335 1340

Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu  
 1345 1350 1355 1360

Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala  
 1365 1370 1375

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr  
 1380 1385 1390

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu  
 1395 1400 1405

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

1410	1415	1420	
Lys Leu Lys Leu			
1425			
<210>	56		
<211>	4889		
<212>	DNA		
<213>	Artificial Sequence		
<220><223>	Synthetic sequence		
<400>	56		
gactgtctcc accatgtaat ttttcctgc gactccatat aacgccggat cgtgaaattt		60	
tcttctttct tttccttct tctcaacaaa caacggatct gtgctttgcg gtccccctgcg		120	
ttcacgcgtc agggctgact gctctgcagc tcgataactc catggagcca tcaacttgct		180	
atgggtgtaa tcatcctatc gacaggtcca agaacaagcc ggctcgcgc tgcctcattc		240	
gctgtcgcaa gacggcttga gtgtttggc tggaggattc gggggcccca tattccaacc		300	
cttttttcca aggccgtcgg ccggtgaggt tgaggaaaac catgggttgc ctacatatta		360	
tcgatgctgg tgtttgtag tagcaatgtt tgcggtggca gtttgagccg agcctcgtct		420	
tgggcttctg acccaggcaa cgccatctga ctagctgcgc cgaaggaagg atgattcatt		480	
gtacgagcc agtcaatgga atcttcaagt aaaagcccga cgaaccgacc atgtcagata		540	
tcagaattct cctggctggt ggggttgggt ggagactgct tacggagtgc atgcctcgtg		600	
actgtcatgg ccgcgtccag cctcctggga ctctgtccga tattatgaca cgagtaaagc		660	
ctgcatgatg tcagtttct gcgtctcatg tcgagaacaa cacacctggt getacatagg		720	
caatactacc tcgtagcttc aaagttgact gttttgcttt gatgtctttg atcatgcca		780	
tccatccctt gtcttgcagt gcatgtggat ctctacgtcc agacggggag aaagcttgtc		840	
tgtgataaag tacgatgatg cattgatgcc tgtggctacg gcccttttat ccccatcgtc		900	
atgcatctct atattaatcc aggagactct cctcctggca tgggtgagta caagtacga		960	
ggacatgtag aagcagagcc acgcaacgtc ttgacatctg tacctatttt gggccaaaaa		1020	
tcgagaccca ccagctcgtc ctacettaca tgtgaagatc ttageccaca atcctactgt		1080	
tttactagta ttactgcaca gctgtcatca cgagtctctg gttgcttgtg aaaccagct		1140	
cagctcctga gcacatgcag taacgccgac tcggcgtcat ttcgccacac ccaatttga		1200	
cctgagggat gctggaagct gctgagcaga tcccgttacc gattcatggc actactacat		1260	

ccatacgcag caaacatggg cttgggcttg gcttctcaat gcaaaattgc cgcmetaagt 1320  
cccggcattg tcgatgcaga gatgcagatt tcagcgggcg attctagggt agggcgacta 1380  
ctactactaa taccacctag tcagtatgta tctagcaccg gaggctaggc ggttagtgga 1440  
cggaacctg gtcattccat cgcaaccagg atcccgcact tcgttgcgct tctgccccca 1500

cggggcggga gttggcagag gcagaatgcg gacagacccc ttgtctgccc tggccggggc 1560  
ctgttgaagc aagcagacga gagcagagcg gttgagaagc ggtggttgac gettgacggt 1620  
acgaagacga gcgagaatcc cgtaagccg aggctgggct ccccccccg tcatcatcat 1680  
gccccctg ctcttcacg ccaactgtct cctgcctcg tcgctcccc tccctcccc 1740  
gattagctgc gcatgttctc ctgacagcgt gactaatgac gcgttgccag cccattcgcc 1800  
tgacgcatcc cggcatctga gtctagctcg tcacgctggc aatcttggcc caggcagagc 1860  
agcaagacgg cgggcatgat tgggcccgtc cctggcgggc atcagctggc catccgctgc 1920

cacccgagac cgcataccg acttgtcggg tctctccgag cagcaggagg ctgatcctgg 1980  
ccggcgagac gattgaaaag ggctcccggg cccggagcag gacagcggcg agagcgagcg 2040  
agagagagga aaagaagaag gtcgactgtc ttatcttcag ccagccccgg ctcaacagaa 2100  
gcagaggaga aggcgaacga cgtcaacgac gacgacgacg acgacgaaga cggtagaagtc 2160  
cgttagttga agatcctgc cgtcaacaac ccatctctgt gatattgctt tcccctgccg 2220  
ttgcttgc accgttccc tctttctctt cccccctt tctcattcc gagcgtact 2280  
ggttctact ccgagcctt cggttgtgcc tttctcttg tcgaccattg caccgcccgt 2340

cgggcactt gggccccgga gaattcggcc ctttcgacg attttggccc tcagttcccc 2400  
atggggacgg tccacacttc ctctcttggc cctgcagacc ttttgcgtc ggtccgagtc 2460  
ggaagaagct cagctcttag cgcttgagta gcatctacgc gcgaatcact ggacaaagtc 2520  
ggcaagacga agccgtcgtc gcctgctgct gctgctgtta ctgcgacagg cgctccgact 2580  
gggggcatcg gcataataaa aagatgcccc ccttcgcat ggacctggcc atgagccact 2640  
cggcatcggc tctctctctc aacgttctct ctacacatc ctcttcatt ccgccatca 2700  
tgacgtcct gtcgactcgg gtgctgctcg gctcccgtgc cgttcaaaag gtcctgggaa 2760

gaccaggatc aagcggctg tccgacgtca ccaagaggtc tgttgacgac ttcacagca 2820  
ccgagacgcc tattgactg aacaatcttc tttgcaatgt tggctctgat ggatgccgtg 2880  
cattcggcac atcagctggt gcggtgattg catctcccag cacaattgac cggactgta 2940  
agttgcctt gatgaacat atcatatac gccgagaagt ggaccgctg ctgagactga 3000  
gacagactat tacatgtgga cgcgagatag cgtcttctg tcaagaacc tcatgaccg 3060  
cttcaccgaa acgtacgat cgggctgca gcgcccatc gagcagtaca ttactgcccc 3120

ggtcactctc cagggcctct ctaacccctc gggctccctc gcggacggct ctggtctcgg 3180  
  
 cgagcccaag tttagattga cctgaagcc tttaccggc aactggggtc gaccgcagcg 3240  
 ggatggccca gctctgcgag ccattgcctt gattggatac tcaaagtggc tcatcaacaa 3300  
 caactatcag tcgactgtgt ccaacgtcat ctggcctatt gtgcgcaacg acctcaacta 3360  
 tgttgccag tactggtcag tgcctgcttg ctcttgaatt acgtctttgc ttgtgtgtct 3420  
 aatgcctcca ccacaggaac caaacggct ttgacctctg ggaagaagtc aatgggagct 3480  
 cattctttac tgttgccaac cagcaccgag gtatgaagca aatcctcgac attcgctgct 3540  
 actgcacatg agcattgtta ctgaccagct ctacagcact tgtcgagggc gccactcttg 3600  
  
 ctgccactct tggccagtcg ggaagcgtt attcatctgt tgctccccag gttttgtgct 3660  
 ttctccaacg attctgggtg tcgtctgggtg gatacgtcga ctccaacagt atgtcttttc 3720  
 actgtttata tgagattggc caatactgat agctcgcctc tagtcaacac caacgagggc 3780  
 aggactggca aggatgtcaa ctccgtcctg acttccatcc acaccttcga tcccaacctt 3840  
 ggctgtgacg caggcacctt ccagccatgc agtgacaaag cgctctcaa cctcaaggtt 3900  
 gttgtcgact ccttccgctc catctacggc gtgaacaagg gcattcctgc cgggtgctgcc 3960  
 gtcgccattg gccggtatgc agaggatgtg tactacaacg gcaacccttg gtatcttgct 4020  
  
 acatttgctg ctgccgagca gctgtacgat gccatctacg tctggaagaa gacgggctcc 4080  
 atcacggtga ccgccacctc cctggccttc ttccaggagc ttgttcttgg cgtgacggcc 4140  
 gggacctact ccagcagctc ttgaccttt accaacatca tcaacgccgt ctgcacatac 4200  
 gccgatggct tectcagcga ggctgccaag tacgtcccg ccgacggttc gctggccgag 4260  
 cagtttgacc gcaacagcgg cactccgctg tctgcgcttc acctgacgtg gtcgtacgcc 4320  
 tcgttcttga cagccacggc ccgtcgggct ggcacgtgtc cccctcgtg ggccaacagc 4380  
 agcgttagca cgatccctc gacgtgctcc ggcgcgtccg tggtcggatc ctactcgcgt 4440  
  
 cccaccgcca cgtcattccc tccgtcgcag acgccaagc ctggcgtgcc ttccggtact 4500  
 ccctacagc cctgcctctg cgcgaccca acctccgtgg ccgtcacctt ccacgagctc 4560  
 gtgtcgacac agtttgcca gacggtcaag gtggcgggca acgcccgggc cctgggcaac 4620  
 tggagcacga gcgccccgt ggctctggac gccgtcaact atgccgataa ccacccctg 4680  
 tggattggga cggtaacct cgaggctgga gacgtcgtgg agtacaagta catcaatgtg 4740  
 ggccaagatg gctccgtgac ctgggagagt gatcccaacc aacttacac ggttcctgcg 4800  
 gtagcttgtg tgacgcaggt tgtcaaggag gacacctggc agtcgtaatg aatcggaag 4860

gggtagtact agtagacttg tagtctgcc 4889

<210> 57

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Primer

<400> 57

gactgtctcc accatgtaat ttttc 25

<210> 58

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Primer

<400> 58

ggcagactac aagtctacta gtactac 27

<210> 59

<211> 111

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 59

taatacgact cactataggg aagaccagga tcaaggtttt agagctagaa atagcaagtt 60

aaaataaggc tagtccgtta tcaacttgaa aaagtggcac cgagtcggtg c 111

<210> 60

<211> 111

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 60

taatacgact cactatagga cagaccgctt gatccgtttt agagctagaa atagcaagtt 60

aaaataaggc tagtccgtta tcaacttgaa aaagtggcac cgagtcggtg c 111

<210> 61

<211> 4248

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic sequence

<400> 61

ccctgctcg tcgctcccc tcctcccc gattagctgc gcatgttctc ctgacagcgt	60
gactaatgac gcgttgccag cccattcgcc tgacgcatcc eggcatctga gtctagctcg	120
tcacgctggc aatcttgcc caggcagagc agcaagacgg cgggcatgat tgggccgtgc	180
cctggcgggc atcagctggc catccgctgc caccgagac cgcacaccg acttgtcgga	240
tctctccgag cagcaggagg ctgatcctgg ccggcgagac gattgaaaag ggctgccggg	300
cccggagcag gacagcggcg agagcgagcg agagagagga aaagaagaag gtcgactgtc	360
ttattttcag ccagccccgg ctcaacagaa gcagaggaga aggcgaacga cgtcaacgac	420
gacgacgacg acgacgaaga cggatgaagtc cgttagttga agatccttgc cgtcacaaca	480
ccatctcgtg gatattgctt tcccctgccg ttgcgttgcc acctgttccc tctttctctt	540
cccccttct tctcattcc gacgctact ggttctact ccgcagcctt cggttgtgcc	600
tttctctttg tcgaccattg caccgcccgt cgcggcactt gggccccgga gaattcggcc	660
ctttcgcage attttgccc tcagttccc atggggacgg tccacacttc ctctcttggc	720
cctgcagacc tttgtcgtc ggtccgagtc ggaagaagct cagtcttgag cgcttgagta	780
gcatctacgc gcgaatcact ggacaaagtc ggcaagacga agccgtcgtc gctgctgct	840
gctgctgtta ctgcgacagg cgctccgact gggggcatcg gcataataaa aagatgcccc	900
ccttcgcat ggacctggcc atgagccact cggcatcggc tctctctctc aacgttctct	960
ctcacacatc ctcttcatt ccgcccatac tggtttaaac ctcgagtta taagtgacaa	1020
catgctctca aagcgtcat ggctggcaca agcctggaaa gaaccaacac aaagcact	1080
gcagcaaadc agctgaattc gtcaccaatt aagtgaacat caacctgaag gcagagtatg	1140
aggccagaag cacatctgga tcgcagatca tggattgccc ctcttggtga agatgagaat	1200
ctagaaagat ggcggggtat gagataagag cgatgggggg gcacatcacc ttccaagaca	1260
aacaaccttt gcagagtcag gcaattttc gtataagagc aggaggaggg agtccagtca	1320
tttcatcage ggtaaaatca ctctagacaa tcttcaagat gatttctgcc ttgggtgact	1380
tatagccatc atcataccta gacagaagct tggggatc taagaccaac gtacaagctc	1440
gcactgtacg ctttgacttc catgtgaaaa ctcgatcgg cgcgcctcta aatttatag	1500
ctcaacct ccaatccaac ctctgcatcc ctctcactcg tctgatcta ctgttcaat	1560
cagagaataa ggacactatc caaatccaac agaatggcta ccacctcca gctgcctgcc	1620

tacaagcagg acttctcaa atccccatc gacggcggcg tcctcaagtt tggcagcttc 1680  
gagctcaagt ccaagcggat atccccctac ttcttcaacg cgggcgaatt ccacacggcg 1740  
cgctcgccg gcgccatcg ctcgcccttt gcaaagacca tcatcgaggc ccaggagaag 1800  
gccggcctag agttcgacat cgtcttcggc ccggcctaca agggcatccc gctgtgctcc 1860  
gccatcacca tcaagctcgg cgagctggcg cccagaacc tggaccgct ctctactcg 1920  
tttgaccga aggaggccaa ggaccacggc gagggcggca acatcgctgg cgcttcgctc 1980  
aagggaaga ggtcctgat tgtcgacgac gtcatcaccg ccggcaccgc caagaggac 2040

gccattgaga agatcaccaa ggagggcggc atcgtcgccg gcatcgctgt ggcctggac 2100  
cgcatggaga agctccccgc tgcggatggc gacgactcca agcctggacc gagtgcatt 2160  
ggcgagctga ggaaggagta cggcatcccc atctttgcca tcctactct ggatgacatt 2220  
atcgatggca tgaagggtt tgcctaccct gaggatatca agaacacgga ggattaccgt 2280  
gccaagtaca aggcgactga ctgattgagg cgttcaatgt cagaaggag agaaagactg 2340  
aaaagtgga aagaagaggc aaattgttgt tattattatt attctatctc gaatcttcta 2400  
gatcttgctg taaataaaca agcgtaaacta gctagcctcc gtacaactgc ttgaattga 2460

taccgtatg gagggcagtt attttattt gttttcaag attttcatt cgccgtgaa 2520  
ctcgtctcac atcgctgta ttgcccgtt gccatgtgt tctcacta cccaagtec 2580  
ctcacgggtt gtctcacttt ctttctctt tctctcct atttttttc aagtcagca 2640  
cagagcagtc atatgggat acgtgcaact gggactcaca acaggccatc ttatggccta 2700  
atagccggcg ttggatccac tagtcaattg gtttaaacag cacatgcagt aacccgact 2760  
cggcgtcatt tcgccacacc caattggac ctgagggatg ctggaagctg ctgagcagat 2820  
ccggtaccg atcatggca ctactacatc catacgcgc aaacatgggc ttggccttg 2880

cttctcaatg caaattgcc cgcaaaagtc ccggcattgt cgatgcagag atgcagattt 2940  
cagcggcgca ttctaggta gggcgactac tactactaat accacctagt cagtatgat 3000  
ctagcaccgg aggctaggcg gttagtggac gggaaacctg tcattccatc gcaaccagga 3060  
tcccgcactt cgttgcctt ctgccccac gggcggggag ttggcagagg cagaatgcgg 3120  
agcagccct tgcctgcct ggccggggcc tgttgaagca agcagacgag agcagagcgg 3180  
ttgagaagcg gtggttgacg cttgacggta cgaagacgag cgagaatccc gttagccga 3240  
ggctgggcta attaattaat gaatcgcaaa gggtagtac tagtagactt gtagtctgcc 3300

ggattattga ttggagtgg tcagtagaat gaaccacggg aatattcgtt caccgggaca 3360  
tttgggatat agcgtttcga gaagctgctg gttgcagcac attggagaag gatgcccttt 3420  
tacgacttat accgctatgc cgggtatatt aatttagccg ttatgaaact caaagagacg 3480

atgataatga tgacgagtaa ttgttcgttt caatttcgaa agctgactcc cacgaagaat 3540  
 atgccaatga cccacggcat gaagcctgaa ctgggcgtgt gtaacacttt aatttgctg 3600  
 acggcggaca aaacaaaggc ggcagcaatg ttgagaccgt gtgataaacc aaggttcccc 3660  
 agggagagag agagagagag agagagagag ctaggtgaaa gaatgagtcc gcctttgagt 3720

catcctgctg ctctctctcc cctctctca ctctctgtat cccatcaacc tcttccctgt 3780  
 tccttctect atcgcatcca tgcgtttgca tcttccattt cattctttcc cettgagccc 3840  
 catctatgca aactcatcat ccggcgcctc gatggaatcc ttgaccttga tgagaatcgc 3900  
 cgtcatcaa ggtccagcc tgcctcgtgc gtcgaactgg aacagcagct cgctaaactc 3960  
 atcctggctg tggttgtcga cggcgttgca caggctctcg agcagcttgt acttgtattg 4020  
 agaggagaac tcggggtcct ttggcggta ggactcgacg gcgcgccggg tgccgacat 4080  
 gtcgcccgtg gcgaggtggc agatgccggc ctggaagcag taggtcgaga ggctccactt 4140

catggtgccg ttgccgatca tgggtttgat gatgcggtcg tacgtctcga tggcgccgta 4200  
 gtagtgcgcc tcgagggcgg cgaggtcggc gtactgcgtc cagagctt 4248

<210> 62

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Primer

<400> 62

cactactaca tccatacgca gcaaacatgg 30

<210> 63

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Primer

<400> 63

ggtcaagaag cacatgccag agttcg 26

<210> 64

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Primer

<400> 64  
gactgtctcc accatgtaat ttttc 25

<210> 65

<211> 4104

<212> DNA

<213> Artificial Sequence

<220><223> codon optimized seqs

<400> 65

atggataaaa aatacagcat tggctcggcc atcggaacca acagcgttgg gtgggcagta 60  
ataacagatg aatacaaaagt gccgtcaaaa aaatttaagg ttctggggaa tacagatcgc 120  
cacagcataa aaaagaatct gattggggca ttgctgtttg attcgggtga gacagctgag 180

gccacgcgtc tgaaacgtac agcaagaaga cgttacacac gtcgtaaaaa tcgtatttgc 240  
tacttacagg aaatTTTTTc taacgaaatg gccaaagtag atgatagttt cttccatcgt 300  
ctcgaagaat cttttctggt tgaggaagat aaaaaacacg aacgtcaccc tatctttggc 360  
aatatcgtgg atgaagtggc ctatcatgaa aaatacccta cgatttatca tcttcgcaag 420  
aagttggttg atagtacgga caaagcggat ctgcgtttaa tctatcttgc gttagcgcac 480  
atgatcaaat ttcgtggtca tttcttaatt gaaggtgata tgaatcctga taactctgat 540  
gtggacaaat tgtttataca attagtcaa acctataatc agctgttcga ggaaaacccc 600

attaatgcct ctggagtga tgccaaagcg attttaagcg cgagactttc taagtcccgg 660  
cgtctggaga atctgatcgc ccagttacca ggggaaaaga aaaatggtct gtttggtaat 720  
ctgattgcc tcaagtctggg gcttaccg aacttcaat ccaatTTTga cctggctgag 780  
gacgcaaagc tgcagctgag caaagatact tatgatgatg acctcgacaa tctgctgcc 840  
cagattggtg accaatatgc ggatctgttt ctggcagcga agaactttc ggatgctatc 900  
ttgctgtcgg atattctgcg tgtaataacc gaaatcacca aagcgcctct gctgcaagt 960  
atgatcaaga gatacgacga gcaccaccag gacctgactc ttcttaaggc actggtacgc 1020

caacagcttc cggagaaata caaagaaata ttcttcgacc agtccaagaa tggttacgcg 1080  
ggctacatcg atggtggtgc atcacaggaa gaggttctata aatttattaa accaatcctt 1140  
gagaaaaatgg atggcacgga agagttactt gttaaactta accgcaaga cttgcttaga 1200  
aagcaacgta cttcgacaa cggctccatc ccacaccaga ttcatttagg tgaactcac 1260  
gccatcttgc gcagacaaga agatttctat cccttcttaa aagacaatcg ggagaaaatc 1320  
gagaagatcc tgacgttccg cattccctat tatgtcggtc ccctggcacg tggtaattct 1380

cggtttgcct ggatgacgcg caaaagtgag gaaacatca ccccttgaa ctttgaagaa 1440  
  
 gtcgtggata aagtgctag cgcgcagtct tttatagaaa gaatgacgaa cttcgataaa 1500  
 aacttgccca acgaaaaagt cctgeccaag cactctcttt tatatgagta ctttactgtg 1560  
 tacaacgaac tgactaaagt gaaatcgtt acggaaggta tgcgcaaacc tgcctttctt 1620  
 agtggcgagc agaaaaagc aattgtcgat cttctcttta aaacgaatcg caagtaact 1680  
 gtaaaacagc tgaaggaaga ttatttcaaa aagatcgaat gctttgattc tgtcgagatc 1740  
 tcgggtgtcg aagatcgttt caacgcttc ttagggacct atcatgattt gctgaagata 1800  
 ataaaagaca aagactttct cgacaatgaa gaaaatgaag atattctgga ggatattgtt 1860  
  
 ttgacctga ccttattcga agatagagag atgatcagg agcgcttaa aacctatgcc 1920  
 cacctgtttg atgacaaagt catgaagcaa ttaaagcgcc gcagatatac ggggtgggpc 1980  
 cgcttgagcc gcaagttgat taacggtatt agagacaagc agagcggaaa aactatcctg 2040  
 gatttctca aatctgacgg atttgcgaac cgcaatttta tgcagcttat acatgatgat 2100  
 tcgcttacat tcaaagagga tattcagaag gctcaggtgt ctgggcaagg tgattcactc 2160  
 cacgaacata tagcaaatit ggccggctct cctgcgatta agaaggggat cctgcaaaaa 2220  
 gttaaagtig tggatgaact tgtaaaagta atgggccgcc acaagccgga gaatatcgtg 2280  
  
 atagaaatgg cgcgcgagaa tcaaacgaca caaaaaggtc aaaagaactc aagagagaga 2340  
 atgaagcgca ttgaggaggg gataaaggaa cttggatctc aaattctgaa agaacatcca 2400  
 gttgaaaaca ctacgtgca aatgaaaaa ttgtacctgt actacctgca gaatggaaga 2460  
 gacatgtacg tggatcagga attggatata aatagactct cggactatga cgtagatcac 2520  
 attgtccctc agagcttctt caaggatgat tctatagata ataaagtact tacgagatcg 2580  
 gacaaaaatc gcggtaaatc ggataacgtc ccatcggagg aagtcgttaa aaagatgaaa 2640  
 aactattggc gtcaactgct gaacccaag ctgatcacac agcgttaagt tgataatctg 2700  
  
 actaaagccg aacgcggtgg tcttagtgaa ctcgataaag caggatttat aaaacggcag 2760  
 ttagtagaaa cgcgccaat tacgaaacac gtggctcaga tctctgattc tagaatgaat 2820  
 acaaagtacg atgaaaacga taaactgac cgtgaagtaa aagtcattac cttaaaatct 2880  
 aaacttgtgt ccgatttccg caaagatttt cagttttaca aggtccggga aatcaataac 2940  
 tatcaccatg cacatgatgc atatttaaat gcggtttag gcacggccct tattaagaaa 3000  
 tacctaaac tcgaaagtga gtttgtttat ggggattata aagtgtatga cgttcgcaaa 3060  
 atgatcgca aatcagaaca ggaaatcgtt aaggctaccg ctaaatactt tttttattcc 3120

aacattatga attttttta gaccgaaata actctcgcga atggtgaaat ccgtaaacgg 3180  
 cctcttatag aaaccaatgg tgaaacggga gaaatcgttt gggataaagg tcgtgacttt 3240  
 gccaccgttc gtaaagtcct ctcaatgccg caagttaaca ttgtcaagaa gacggaagtt 3300  
 caaacagggg gattctccaa agaatctatc ctgccgaagc gtaacagtga taaacttatt 3360  
 gccagaaaaa aagattggga tccaaaaaaa tacggaggct ttgattcccc taccgtcgcg 3420  
 tatagtgtgc tgggtggttc taaagtcgag aaagggaaaa gcaagaaatt gaaatcagtt 3480  
 aaagaactgc tgggtattac aattatggaa agatcgtcct ttgagaaaaa tccgatcgac 3540

tttttagagg ccaaggggta taaggaagtg aaaaaagatc tcatcatcaa attaccgaag 3600  
 tatagctttt ttgagctgga aaacggcaga aaaagaatgc tggcctccgc gggcgagtta 3660  
 cagaagggaa atgagctggc gctgccttcc aaatatgtta atttctgta ccttgccagt 3720  
 cattatgaga aactgaaggg cagccccgaa gataacgaac agaaacaatt attcgtggaa 3780  
 cagcataagc actatctaga tgaattata gagcaaatta gtgaattttc taagcgcgtt 3840  
 atcctcgcgg atgctaattt agacaaagta ctgtcagctt ataataaaca tcgggataag 3900  
 ccgattagag aacaggccga aaatatcatt catttgitta ccttaaccaa ccttgagca 3960

ccagctgcct tcaaatattt cgataccaca attgatcgta aacgtatac aagtacaaaa 4020  
 gaagtcttgg acgcaacct cattcatcaa tctattactg gattatatga gacacgcatt 4080  
 gatctttcac agctgggcgg agac 4104

<210> 66

<211> 1368

<212> PRT

<213> Artificial Sequence

<220><223> protein expressed from synthetic construct

<400> 66

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala 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



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



1075                      1080                      1085  
 Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly

1090                      1095                      1100  
 Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile  
 1105                      1110                      1115                      1120

Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser  
 1125                      1130                      1135

Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly  
 1140                      1145                      1150

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile  
 1155                      1160                      1165

Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala  
 1170                      1175                      1180

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys  
 1185                      1190                      1195                      1200

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser  
 1205                      1210                      1215

Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr  
 1220                      1225                      1230

Val 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 His

1250                      1255                      1260  
 Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val

1265                      1270                      1275                      1280  
 Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys

1285                      1290                      1295  
 His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu

1300                      1305                      1310

Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp  
 1315                      1320                      1325

Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp  
 1330 1335 1340

Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile  
 1345 1350 1355 1360

Asp Leu Ser Gln Leu Gly Gly Asp  
 1365

<210> 67  
 <211> 46  
 <212> PRT  
 <213> Artificial Sequence  
 <220>  
 ><223> Synthetic sequence  
 <400> 67

Met His His His His His His Ser Ser Gly Leu Val Pro Arg Gly Ser  
 1 5 10 15

Gly Met Lys Glu Thr Ala Ala Ala Lys Phe Glu Arg Gln His Met Asp  
 20 25 30

Ser Pro Asp Leu Gly Thr Asp Asp Asp Asp Lys Ala Met Ala  
 35 40 45

<210> 68  
 <211> 138  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic sequence

<400> 68  
 atgcaccatc atcatcatca ttcttctggt ctggtgccac gcggttctgg tatgaaagaa 60  
 accgctgctg ctaaattcga acgccagcac atggacagcc cagatctggg taccgacgac 120  
 gacgacaagg ccatggcc 138

<210> 69  
 <211> 21  
 <212> DNA  
 <213> SV40  
 <400> 69  
 ccaaaaaaga aacgcaaggt t 21

<210> 70  
 <211> 21  
 <212> DNA  
 <213> T. reesei  
 <400> 70  
 aagaagaaaa aactgaaact g 21

<210> 71  
 <211> 19  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 71  
 tgatgacggt gaaaacctc 19

<210> 72  
 <211> 16  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 72  
 aaaagcaccg actcgg 16

<210> 73  
 <211> 45  
 <212> DNA  
 <213> T. reesei  
 <400> 73  
 atgcgcaaat ttaaagcgct gatgtgtgic taatgcctcc accac 45

<210> 74  
 <211> 45  
 <212> DNA  
 <213> T. reesei  
 <400> 74  
 atatggatct gcgcgcatc gatgatcgtg ctagcgctgc tgttg 45

<210> 75

<211> 19  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 75  
 ccaccacagg aaccaaacc 19  
 <210> 76  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 76  
 ctgcgacgga gggaatgacg 20  
  
 <210> 77  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 77  
 gggcaggact ggcaaggatg t 21  
 <210> 78  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 78  
 gccgtcacgc caggaacaag 20  
 <210> 79  
 <211> 26  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Primer  
 <400> 79

gaacaatctt ctttgcaatg ttggtc

26

<210> 80

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Primer

<400> 80

gaggaagtcc tgctttagg caggc

25

<210> 81

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Primer

<400> 81

cgacagagca gtcatatggg gatacg

26