(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0036792
(43) 공개일자 2017년04월03일

(51) 국제특허분류(Int. Cl.)
C12N 15/10 (2017.01) *C12N 15/113* (2010.01)
C12N 15/81 (2006.01) *C12N 15/90* (2006.01)
C12N 9/16 (2006.01)
(52) CPC특허분류
C12N 15/102 (2013.01)
C12N 15/113 (2013.01)
(21) 출원번호 10-2017-7006429
(22) 출원일자(국제) 2015년07월21일
심사청구일자 없음
(85) 번역문제출일자 2017년03월08일
(86) 국제출원번호 PCT/US2015/041256
(87) 국제공개번호 WO 2016/025131
국제공개일자 2016년02월18일
(30) 우선권주장
62/036,652 2014년08월13일 미국(US)

(71) 출원인
이 아이 듀폰 디 네모아 앤드 캄파니
미국 19805 델라웨어주 윌밍톤 피.오. 박스 2915 센터 로드 974 체스트넛 런 플라자
(72) 발명자
프리슈, 라이언
미국 19702 델라웨어주 뉴어크 버클리 드라이브 27
판, 시아오춘
미국 19382 펜실베이니아주 웨스트 체스터 프랫 레인 175
홍, 승-표
미국 19707 델라웨어주 호케신 피어슨'즈 리지 9
(74) 대리인
양영준, 심미성

전체 청구항 수 : 총 20 항

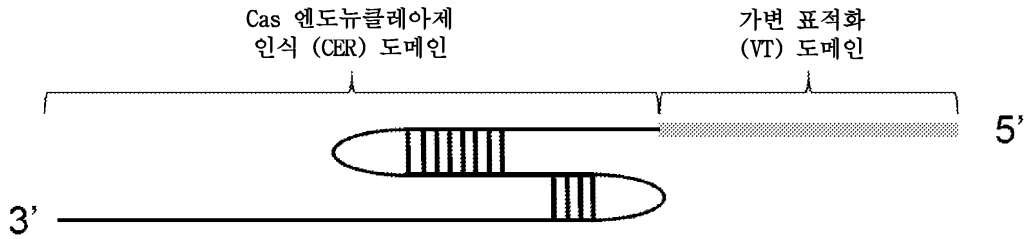
(54) 발명의 명칭 **RNA-유도 엔도뉴클레아제를 사용한 통상적이지 않은 효모에서의 유전자 표적화**

(57) 요 약

5'-캡을 갖지 않는 적어도 하나의 RNA 성분을 포함하는 적어도 하나의 RNA-유도 엔도뉴클레아제 (RGEN)를 포함하는 통상적이지 않은 효모가 본 명세서에 개시되어 있다. 이러한 캡핑되지 않은 RNA 성분은 효모 내의 염색체 또는 에피솜의 표적 부위 서열에 상보적인 서열을 포함한다. RGEN은 표적 부위에서 서열의 하나 또는 두 DNA 가닥에 결합하고, 임의로 절단할 수 있다. 본 명세서의 RGEN의 예는 Cas9 단백질과 가이드 RNA의 복합체이다. 소정 실시 형태에서 리보자임이 사용되어 5'-캡이 없는 RNA 성분을 제공한다. 통상적이지 않은 효모에서 유전자를 표적화하는 방법이 추가로 개시되어 있다.

대 표 도 - 도1

단일 가이드 폴리뉴클레오티드 (예를 들어, sgRNA)



(52) CPC특허분류
C12N 15/81 (2013.01)
C12N 15/905 (2013.01)
C12N 9/16 (2013.01)
C12N 2310/121 (2013.01)

명 세 서

청구범위

청구항 1

5'-캡(cap)을 갖지 않는 적어도 하나의 RNA 성분을 포함하는 적어도 하나의 RNA-유도 엔도뉴클레아제 (RGEN)를 포함하는 통상적이지 않은 효모로서, 여기서 RNA 성분은 효모 내의 염색체 또는 에피솜 상에 표적 부위 서열에 상보적인 서열을 포함하고, RGEN은 표적 부위 서열에 결합할 수 있는, 통상적이지 않은 효모.

청구항 2

제1항에 있어서, RGEN이 표적 부위 서열에 결합하여 절단할 수 있는, 통상적이지 않은 효모.

청구항 3

제1항에 있어서, 상기 효모가 야로위아(*Yarrowia*), 피키아(*Pichia*), 쉬반니오마이세스(*Schwanniomyces*), 클루이베로마이세스(*Kluyveromyces*), 아르술라(*Arxula*), 트리코스포론(*Trichosporon*), 칸디다(*Candida*), 유스틸라고(*Ustilago*), 토룰롭시스(*Torulopsis*), 자이고사카로마이세스(*Zygosaccharomyces*), 트리고놉시스(*Trigonopsis*), 크립토코커스(*Cryptococcus*), 로도토룰라(*Rhodotorula*), 파피아(*Phaffia*), 스포로볼로마이세스(*Sporobolomyces*) 및 파치솔렌(*Pachysolen*)으로 이루어진 군으로부터 선택되는 속의 구성원인, 통상적이지 않은 효모.

청구항 4

제1항에 있어서, RGEN이 크리스퍼(CRISPR; 규칙적으로 사이 간격을 두고 분포하는 짧은 회문구조 반복 서열)-관련 (Cas) 단백질-9 (Cas9) 아미노산 서열을 포함하는, 통상적이지 않은 효모.

청구항 5

Cas 엔도뉴클레아제 및 적어도 하나의 뉴클레오티드 서열에 작동가능하게 연결된 프로모터를 포함하는 폴리뉴클레오티드 서열을 포함하는 통상적이지 않은 효모로서, 여기서 상기 뉴클레오티드 서열은 RNA 성분을 암호화하는 DNA 서열의 상류에 리보자임을 암호화하는 DNA 서열을 포함하고, 상기 RNA 성분은 효모 내의 염색체 또는 에피솜 상의 표적 부위 서열에 상보적인 가변 표적화 도메인을 포함하며, RNA 성분과 Cas 엔도뉴클레아제는 표적 부위 서열에 결합할 수 있는 RNA-유도 엔도뉴클레아제 (RGEN)를 형성할 수 있는, 통상적이지 않은 효모.

청구항 6

제5항에 있어서, RGEN이 표적 부위 서열에 결합하여 절단할 수 있는, 통상적이지 않은 효모.

청구항 7

제5항에 있어서, 뉴클레오티드 서열로부터 전사된 RNA가 자체 촉매 작용으로(autocatalytically) 리보자임을 제거하여 상기 RNA 성분을 생성하며, 상기 RNA 성분이 5' 캡을 포함하지 않는, 통상적이지 않은 효모.

청구항 8

제7항에 있어서, 리보자임이 망치머리(hammerhead) 리보자임, 델타 간염 바이러스(hepatitis delta virus) 리보자임, 그룹 I 인트론 리보자임, RnaseP 리보자임 또는 헤어핀 리보자임인, 통상적이지 않은 효모.

청구항 9

제5항에 있어서, 뉴클레오티드 서열로부터 전사된 RNA가 5' 캡이 없는 리보자임-RNA 성분 융합 분자를 생성하기 위해 자체 촉매 작용으로 리보자임을 제거하지 않는, 통상적이지 않은 효모.

청구항 10

통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 표적 부위를 변형시키는 방법으로서, 상기 방법은 통상적

이지 않은 효모에 Cas 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 제1 재조합 DNA 구조물 및 RNA 성분의 상류에 리보자임을 암호화하는 DNA 서열을 포함하는 제2 재조합 DNA 구조물을 제공하는 단계를 포함하고, 여기서 제2 재조합 DNA 구조물로부터 전사된 RNA는 자체 촉매 작용으로 리보자임을 제거하여 상기 RNA 성분을 생성하며, Cas9 엔도뉴클레아제는 상기 표적 부위에서 단일 또는 이중 가닥 절단을 도입하는, 방법.

청구항 11

통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 표적 부위를 변형시키는 방법으로서, 상기 방법은 통상적이지 않은 효모에 Cas 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 제1 재조합 DNA 구조물 및 리보자임-RNA 성분 융합 분자를 암호화하는 DNA 서열을 포함하는 제2 재조합 DNA 구조물을 제공하는 단계를 포함하고, 여기서 상기 리보자임-RNA 성분 융합 분자와 Cas9 엔도뉴클레아제는 상기 표적 부위에서 단일 또는 이중 가닥 절단을 도입하는 RGEN을 형성할 수 있는, 방법.

청구항 12

제10항 또는 제11항에 있어서, 상기 표적에서 변형을 갖는 적어도 하나의 통상적이지 않은 효모 세포를 동정하는 단계를 추가로 포함하며, 여기서 변형은 상기 표적 부위에서 하나 이상의 뉴클레오티드의 적어도 하나의 결실, 첨가 또는 치환을 포함하는, 방법.

청구항 13

제10항 또는 제11항에 있어서, 상기 효모에 공여 DNA를 제공하는 단계를 추가로 포함하며, 여기서 상기 공여 DNA는 관심 대상의 폴리뉴클레오티드를 포함하는, 방법.

청구항 14

제13항에 있어서, 상기 표적 부위에 통합된 관심 대상의 폴리뉴클레오티드를 이의 염색체 또는 에피솜에 포함하는 적어도 하나의 효모 세포를 동정하는 단계를 추가로 포함하는, 방법.

청구항 15

통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 뉴클레오티드 서열을 편집(editing)하는 방법으로서, 상기 방법은 통상적이지 않은 효모에 폴리뉴클레오티드 변형 주형 DNA, Cas 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 제1 재조합 DNA 구조물 및 RNA 성분의 상류에 리보자임을 암호화하는 DNA 서열을 포함하는 제2 재조합 DNA 구조물을 제공하는 단계를 포함하고, 여기서 제2 재조합 DNA 구조물로부터 전사된 RNA는 자체 촉매 작용으로 리보자임을 제거하여 상기 RNA 성분을 생성하며, Cas9 엔도뉴클레아제는 상기 효모의 염색체 또는 에피솜의 표적 부위에 단일 또는 이중 가닥 절단을 도입하고, 상기 폴리뉴클레오티드 변형 주형 DNA는 상기 뉴클레오티드 서열의 적어도 하나의 뉴클레오티드 변형을 포함하는, 방법.

청구항 16

통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 뉴클레오티드 서열을 침묵(silencing)시키는 방법으로서, 상기 방법은 통상적이지 않은 효모에 불활성화된 Cas9 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 적어도 하나의 제1 재조합 DNA 구조물 및 적어도 하나의 폴리뉴클레오티드에 작동가능하게 연결된 프로모터를 포함하는 적어도 하나의 제2 재조합 DNA 구조물을 제공하는 단계를 포함하고, 여기서 상기 적어도 하나의 폴리뉴클레오티드는 리보자임-RNA 성분 융합 분자를 암호화하며, 상기 리보자임-RNA 성분 융합 분자와 불활성화된 Cas9 엔도뉴클레아제는 상기 효모의 염색체 또는 에피솜 내의 상기 뉴클레오티드 서열에 결합하는 RGEN을 형성함으로써 상기 뉴클레오티드 서열의 전사를 차단할 수 있는, 방법.

청구항 17

통상적이지 않은 효모에서 유전자 변형을 위한 다수의 가이드 RNA를 고효율(high throughput)로 생성하는 방법으로서, 상기 방법은 하기 단계를 포함하는, 방법:

a) 5'에서 3' 순서로 리보자임을 암호화하는 제1 DNA 서열, 역선별제(counterselection agent)를 암호화하는 제2 DNA 서열, 가이드 RNA의 CER 도메인을 암호화하는 제3 DNA 서열 및 종결 서열에 작동가능하게 연결된 프로모터를 포함하는 재조합 DNA 구조물을 제공하는 단계;

b) (a)의 재조합 DNA 구조물에 적어도 하나의 올리고뉴클레오티드 듀플렉스를 제공하는 단계로서, 상기 올리고뉴클레오티드 듀플렉스는 가이드 RNA 표적 서열의 가변 표적화 도메인 (VT)을 암호화할 수 있는 DNA 서열을 포함하는 제1 단일 가닥 올리고뉴클레오티드와 가변 표적화 도메인을 암호화하는 DNA 서열에 상보적인 서열을 포함하는 제2 단일 가닥 올리고뉴클레오티드를 조합하는 것으로부터 유래되는, 단계;

c) (a)의 역선별제를 적어도 하나의 (b)의 올리고듀플렉스와 교환함으로써, 각각이 가이드 RNA의 가변 표적화 도메인을 암호화할 수 있는 DNA 서열을 포함하는 재조합 DNA 구조물의 라이브러리를 생성하는 단계; 및

d) (c)의 재조합 DNA 구조물의 라이브러리를 전사함으로써, 리보자임-가이드 RNA 분자의 라이브러리를 생성하는 단계.

청구항 18

제17항에 있어서, 리보자임-가이드 RNA 분자가 리보자임과 리보자임 상류에 임의의 RNA 서열을 자체 촉매 작용으로 제거하도록 상기 분자의 라이브러리를 유도하여, 5' 캡을 함유하지 않는 가이드 RNA 분자의 라이브러리를 생성하는 단계를 추가로 포함하는, 방법.

청구항 19

제17항에 있어서, 리보자임-가이드 RNA 분자가 리보자임 상류에 임의의 RNA 서열을 절단하도록 상기 분자의 라이브러리를 유도하여, 5' 캡을 함유하지 않는 리보자임-gRNA 융합 분자를 생성하는 단계를 추가로 포함하는, 방법.

청구항 20

(ii) 리보자임과 가이드 RNA를 암호화하는 뉴클레오티드 서열에 작동가능하게 연결된 (i) 폴리머라제-II 프로모터를 포함하는 재조합 DNA 서열로서, 여기서 상기 리보자임은 상기 가이드 RNA의 상류이고, (ii)의 뉴클레오티드 서열로부터 전사된 RNA는 자체 촉매 작용으로 리보자임을 제거하여 상기 가이드 RNA를 생성하며, 상기 가이드 RNA는 통상적이지 않은 효모의 게놈 내의 표적 부위를 인식하고, 결합하며, 임의로 절단할 수 있는 RGEN을 형성할 수 있는, 재조합 DNA 서열.

발명의 설명

기 술 분 야

[0001] 본 출원은 전체적으로 본 명세서에서 참고로 포함되는 2014년 8월 13일자로 출원된 미국 가출원 제62/036652호의 이익을 주장한다.

[0002] 기술분야

[0003] 본 발명은 분자생물학 분야이다. 구체적으로, 본 발명은 RNA-유도 엔도뉴클레아제(RGEN)를 사용하는 통상적이지 않은 효모에서의 유전 표적화에 관한 것이다.

[0004] 전자 제출된 서열 목록에 대한 참조

[0005] 본 서열 목록의 공식 사본은 2015년 7월 21일자로 작성되고 크기가 411KB이며 본 명세서와 함께 제출된, 파일명 20150721_CL6272WOPCT_SequenceListing_ST25.txt의 ASCII 형식의 서열 목록으로서 EFS-웹을 통해 전자적으로 제출되었다. 상기 ASCII 형식의 서류에 포함된 서열 목록은 본 명세서의 일부이며 전체적으로 본 명세서에 참조로 포함된다.

배 경 기 술

[0006] 유기체 내에서 유전자의 기능을 이해하는 강력한 방법은 이의 발현을 억제하는 것이다. 유전자 발현의 억제는, 예를 들어, 유전자의 "녹아웃"을 유발하는 유전자의 DNA 서열을 중단시키거나 결실시켜 달성될 수 있다 (문헌 [Austin et al., *NaT. Genetics* 36:921-924]). 유전자 녹아웃 대부분 박테리아에서 포유동물까지의 다양한 유기체의 적용할 수 있는 기술인 상동 재조합 (HR)을 통해 수행되어 왔다. 유전자 기능을 연구하기 위한 또 다른 도구는 유전자 "녹인(knock-in)"을 통한 것일 수 있으며, 이는 보통 HR에 의해 또한 수행된다. 유전자 표적화 (녹아웃 또는 녹인)를 위한 HR은 표적 부위와 상동성을 갖는 외인성으로 공급된 DNA의 존재를 사용할 수 있다.

[0007] HR에 의한 유전자 표적화는 강력한 도구이지만, 복잡하고, 노동 집약적인 절차일 수 있다. HR은 일반적으로 비용 효과적인 방식으로 규모를 확대하기 어렵기 때문에, HR을 사용하는 대부분의 연구는 일반적으로 경로에서 다중 유전자보다는 단일 유전자의 녹아웃으로 제한되어 왔다. 이러한 어려움은 HR이 효율적이지 않은 유기체에서 가중된다. 이러한 낮은 효율은 전형적으로 실무자가 원하는 HR 이벤트가 발생하는 세포를 확인하는 데 도움을 주는 선택가능한 표현형 또는 외인성 마커에 의존하게 한다.

[0008] 표적화된 DNA 부위가 이중 가닥 절단을 함유하는 경우, 유전자 표적화를 위한 HR이 강화되는 것으로 나타났다 (문헌[Rudin et al., *Genetics* 122:519-534; Smih et al., *NucL. Acids Res.* 23:5012-5019]). 따라서, HR-매개 DNA 표적화를 촉진하기 위해 이중 가닥 절단을 도입하기 위한 전략이 개발되어 왔다. 예를 들어, 징크 핑거 뉴클레아제는 공여 DNA가 존재하는 경우 특정 부위에서 HR의 수준을 향상시키는 특정 DNA 부위를 절단하도록 유전자 조작되어 왔다 (문헌[Bibikova et al., *Science* 300:764; Bibikova et al., *MoL. CelL. BioL.* 21:289-297]). 유사하게, 인공 메가뉴클레아제 (호밍(homing) 엔도뉴클레아제) 및 전사 활성인자-유사 이펙터 (TALE) 뉴클레아제는 또한 HR-매개 DNA 표적화에 사용하기 위해 개발되어 왔다 (문헌[Epinat et al., *Nucleic Acids Res.* 31: 2952-2962; Miller et al., *NaT. Biotech.* 29:143-148]).

[0009] 크리스퍼(CRISPR; 규칙적으로 사이 간격을 두고 분포하는 짧은 회문구조 반복 서열) DNA 절단 시스템을 암호화하는 유전자좌는 약 40%의 박테리아 게놈과 대부분의 고세균 게놈에서만 발견되어 왔다 (문헌[Horvath and Barrangou, *Science* 327:167-170; Karginov and Hannon, *MoL. Cell* 37:7-19]). 특히, 타입 II CRIPSR 시스템의 크리스퍼-관련 (Cas) RNA-유도 엔도뉴클레아제 (RGEN)인 Cas9는 부위 특이적 DNA 가닥 절단을 도입하기 위한 수단으로서 개발되어 왔다 (본 명세서에 그 전체가 참고로 포함되는, 2015년 3월 19일자로 출원된 미국 특허 출원 제2015-0082478 A1호 및 2015년 2월 26일자로 출원된 제2015-0059010 A1호). Cas9의 RNA 성분의 서열은 Cas9가 (i) RNA 성분의 일부에 상보적인 서열 및 (ii) 프로토스페이서 인접 모티프 (PAM) 서열을 함유하는 DNA를 인식하고 절단하도록 설계될 수 있다.

[0010] 고유 Cas9/RNA 복합체는 크리스퍼 RNA (crRNA)와 트랜스-활성화 크리스퍼 RNA (tracrRNA)의 두 가지의 RNA 서열을 포함한다. crRNA는 5'에서 3' 방향으로, 표적 DNA 부위, 및 crRNA가 유래된 크리스퍼 유전자좌의 반복 영역에 의해 암호화되는 서열의 일부에 상보적인 특유의 서열을 함유한다. tracrRNA는 5'에서 3' 방향으로 crRNA의 반복 영역과 어닐링하는 서열 및 스템 루프(stem loop)를 함유 부분을 함유한다. 최근 연구는 tracrRNA에 연결된 crRNA를 5'에서 3' 방향으로 함유하는 키메라 서열인 가이드 RNA (gRNA)의 개발로 이어졌다 (2013년 8월 22일자로 출원된 미국 가출원 제61/868,706호).

[0011] Cas9-매개 DNA 표적화를 수행하기 위해 진핵 세포에서 gRNA와 같은 RNA 성분을 발현하는 방법은 정확하게 한정되고, 개질되지 않은 5'- 및 3'-말단을 갖는 RNA의 전사를 가능하게 하는 RNA 폴리머라제 III (Pol III) 프로모터를 사용하는 것이었다 (문헌[DiCarlo et al., *Nucleic Acids Res.* 41: 4336-4343; Ma et al., *MoL. TheR. Nucleic Acids* 3:e161]). 이러한 전략은 옥수수와 대두 (2013년 8월 22일자로 출원된 미국 가출원 제61/868,706호) 뿐만 아니라 인간, 마우스, 제브라피쉬, 트리코데르마(*Trichoderma*) 및 사카로마이세스 세레비시아(*Sacchromyces cerevisiae*)를 비롯한 여러 상이한 종의 세포에서 성공적으로 적용되어 왔다.

[0012] 그럼에도 불구하고, 이제 본 출원에 기재된 바와 같이, Pol III 프로모터-전사된 gRNA를 사용하여 통상적이지 않은 효모, 예컨대 야로위아 리폴리티카(*Yarrowia lipolytica*)에서 Cas9-매개 DNA 표적화를 수행하는 것이 어려운 것으로 입증되었다. 따라서, Cas9에 대한 RNA 성분을 생성하는 다른 방법은 통상적이지 않은 효모에서 Cas9-매개 DNA 표적화를 제공하는 데 관심이 있다.

발명의 내용

[0013] 일 실시 형태에서, 본 발명은 5'-캡을 갖지 않는 적어도 하나의 RNA 성분을 포함하는 적어도 하나의 RNA-유도 엔도뉴클레아제 (RGEN)를 포함하는 통상적이지 않은 효모에 관한 것이며, 여기서 RNA 성분은 효모 내의 염색체 또는 에피솜 상에 표적 부위 서열에 상보적인 서열을 포함하고, RGEN은 표적 부위 서열에 결합할 수 있다. RGEN은 또한 표적 부위에 결합하여 이를 절단할 수 있다.

[0014] 일 실시 형태에서, 통상적이지 않은 효모는 야로위아, 피키아(*Pichia*), 쉬반니오마이세스(*Schwanniomyces*), 클루이베로마이세스(*Kluyveromyces*), 아르술라(*Arxula*), 트리코스포론(*Trichosporon*), 칸디다(*Candida*), 유스틸라고(*Ustilago*), 토룰롭시스(*Torulopsis*), 자이고사카로마이세스(*Zygosaccharomyces*), 트리고놉시스(*Trigonopsis*), 크립토코커스(*Cryptococcus*), 로도토룰라(*Rhodotorula*), 파피아(*Phaffia*), 스포로볼로마이세스(*Sporobolomyces*) 및 파치솔렌(*Pachysolen*)으로 이루어진 군으로부터 선택되는 속의 구성원이다.

[0015] 일 실시 형태에서, RGEN은 크리스퍼 (규칙적으로 사이 간격을 두고 분포하는 짧은 회문구조 반복 서열)-관련 (Cas) 단백질-9 (Cas9) 아미노산 서열을 포함한다. Cas9 단백질은 스트렙토코커스(*Streptococcus*) Cas9 단백질일 수 있지만, RNA 성분은 트랜스-활성화 크리스퍼 RNA (tracrRNA)에 작동가능하게 연결된 크리스퍼 RNA (crRNA)를 포함하는 가이드 RNA (gRNA)를 포함할 수 있다. PAM (프로토스페이서-인접 모티프) 서열은 표적 부위 서열에 인접할 수 있다. RGEN은 또한 표적 부위에 결합하여 이를 절단할 수 있다. 뉴클레오티드 서열로부터 전사된 RNA는 자체 촉매 작용으로 리보자임을 제거하여 상기 RNA 성분을 생성할 수 있으며, 상기 RNA 성분은 5' 캡을 포함하지 않는다. 이러한 리보자임은 망치머리(hammerhead) 리보자임, 델타 간염 바이러스(hepatitis delta virus) 리보자임, 그룹 I 인트론 리보자임, RnaseP 리보자임 또는 헤어핀 리보자임을 포함할 수 있다. 뉴클레오티드 서열로부터 전사된 RNA는 5' 캡이 없는 리보자임-RNA 성분 융합 분자를 생성하기 위해 자체 촉매 작용으로 리보자임을 제거하지 않는 RNA 분자일 수 있다.

[0016] 일 실시 형태에서, 본 발명은 Cas 엔도뉴클레아제 및 적어도 하나의 뉴클레오티드 서열에 작동가능하게 연결된 프로모터를 포함하는 폴리뉴클레오티드 서열을 포함하는 통상적이지 않은 효모에 관한 것이며, 여기서 상기 뉴클레오티드 서열은 RNA 성분을 암호화하는 DNA 서열의 상류에 리보자임을 암호화하는 DNA 서열을 포함하고, 상기 RNA 성분은 효모 내의 염색체 또는 에피솜 상의 표적 부위 서열에 상보적인 가변 표적화 도메인을 포함하며, RNA 성분은 Cas 엔도뉴클레아제와 함께, 표적 부위 서열에 결합할 수 있는 RNA-유도 엔도뉴클레아제 (RGEN)를 형성할 수 있다.

[0017] 일 실시 형태에서, 본 명세서에 기재된 방법은 통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 표적 부위를 변형시키는 방법을 포함하며, 상기 방법은 통상적이지 않은 효모에 Cas 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 제1 재조합 DNA 구조물 및 RNA 성분의 상류에 리보자임을 암호화하는 DNA 서열을 포함하는 제2 재조합 DNA 구조물을 제공하는 단계를 포함하고, 여기서 제2 재조합 DNA 구조물로부터 전사된 RNA는 자체 촉매 작용으로 리보자임을 제거하여 상기 RNA 성분을 생성하며, Cas9 엔도뉴클레아제는 상기 표적 부위에서 단일 또는 이중 가닥 절단을 도입한다.

[0018] 일 실시 형태에서, 본 명세서에 기재된 방법은 통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 표적 부위를 변형시키는 방법을 포함하며, 상기 방법은 통상적이지 않은 효모에 Cas 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 제1 재조합 DNA 구조물 및 리보자임-RNA 성분 융합 분자를 암호화하는 DNA 서열을 포함하는 제2 재조합 DNA 구조물을 제공하는 단계를 포함하고, 여기서 상기 리보자임-RNA 성분 융합 분자와 Cas9 엔도뉴클레아제는 상기 표적 부위에서 단일 또는 이중 가닥 절단을 도입하는 RGEN을 형성할 수 있다.

[0019] 방법은 상기 표적에서 변형을 갖는 적어도 하나의 통상적이지 않은 효모 세포를 동정하는 단계를 추가로 포함할 수 있으며, 여기서 변형은 상기 표적 부위에서 하나 이상의 뉴클레오티드의 적어도 하나의 결실 또는 치환을 포함한다. 방법은 상기 효모에 공여 DNA를 제공하는 단계를 추가로 포함할 수 있으며, 여기서 상기 공여 DNA는 관심 대상의 폴리뉴클레오티드를 포함한다.

[0020] 일 실시 형태에서, 본 명세서에 기재된 방법은 통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 뉴클레오티드 서열을 편집(editing)하는 방법을 포함하며, 상기 방법은 통상적이지 않은 효모에 폴리뉴클레오티드 변형 주형 DNA, Cas 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 제1 재조합 DNA 구조물 및 RNA 성분의 상류에 리보자임을 암호화하는 DNA 서열을 포함하는 제2 재조합 DNA 구조물을 제공하는 단계를 포함하고, 여기서 제2 재조합 DNA 구조물로부터 전사된 RNA는 자체 촉매 작용으로 리보자임을 제거하여 상기 RNA 성분을 생성하며, Cas9 엔도뉴클레아제는 상기 효모의 염색체 또는 에피솜 내의 표적 부위에서 단일 또는 이중 가닥 절단을 도입하고, 상기 폴리뉴클레오티드 변형 주형 DNA는 상기 뉴클레오티드 서열의 적어도 하나의 뉴클레오티드 변형을 포함한다.

[0021] 일 실시 형태에서, 본 명세서에 기재된 방법은 통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 뉴클레오티드 서열을 침묵(silencing)시키는 방법을 포함하며, 상기 방법은 통상적이지 않은 효모에 불활성화된 Cas9 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 적어도 하나의 제1 재조합 DNA 구조물 및 적어도 하나의 폴리뉴클레오티드에 작동가능하게 연결된 프로모터를 포함하는 적어도 하나의 제2 재조합 DNA 구조물을 제공하는 단계를 포함하고, 여기서 상기 적어도 하나의 폴리뉴클레오티드는 리보자임-RNA 성분 융합 분자를 암호화하며, 상기 리보자임-RNA 성분 융합 분자와 불활성화된 Cas9 엔도뉴클레아제는 상기 효모의 염색체 또는 에피솜 내의 상기 뉴클레오티드 서열에 결합하는 RGEN을 형성함으로써 상기 뉴클레오티드 서열의 전사를 차단할 수 있다.

[0022] 일 실시 형태에서, 본 명세서에 기재된 방법은 통상적이지 않은 효모에서 유전자 변형을 위한 다수의 가이드 RNA를 고효율(high throughput)로 생성하는 방법을 포함하며, 상기 방법은 하기 단계를 포함한다: a) 5'에서 3'

순서로 리보자임을 암호화하는 제1 DNA 서열, 역선별제(counterselection agent)를 암호화하는 제2 DNA 서열, 가이드 RNA의 CER 도메인을 암호화하는 제3 DNA 서열 및 종결 서열에 작동가능하게 연결된 프로모터를 포함하는 재조합 DNA 구조물을 제공하는 단계; b) (a)의 재조합 DNA 구조물에 적어도 하나의 올리고뉴클레오티드 듀플렉스를 제공하는 단계로서, 상기 올리고뉴클레오티드 듀플렉스는 가이드 RNA 표적 서열의 가변 표적화 도메인(VT)을 암호화할 수 있는 DNA 서열을 포함하는 제1 단일 가닥 올리고뉴클레오티드와 가변 표적화 도메인을 암호화하는 DNA 서열에 상보적인 서열을 포함하는 제2 단일 가닥 올리고뉴클레오티드를 조합하는 것으로부터 유래되는 단계; c) (a)의 역선별제를 (b)의 적어도 하나의 올리고듀플렉스와 교환함으로써, 각각이 가이드 RNA의 가변 표적화 도메인을 암호화할 수 있는 DNA 서열을 포함하는 재조합 DNA 구조물의 라이브러리를 생성하는 단계; 및

[0023] d) (c)의 재조합 DNA 구조물의 라이브러리를 전사함으로써, 리보자임-가이드 RNA의 라이브러리를 생성하는 단계.

도면의 간단한 설명

[0024] 도면 및 서열의 간단한 설명

도 1: 단일 가이드 폴리뉴클레오티드, 예컨대 단일 가이드 RNA (sgRNA)의 구조 모델. 가변 표적화 (VT) 도메인은 회색으로 도시되어 있다. Cas9 엔도뉴클레아제 인식 (CER) 도메인은 검정색으로 도시되어 있다.

도 2a: 야로위아 코돈-최적화된 Cas9 발현 카세트. FBA1 프로모터는 검정색으로 도시되어 있고, C-말단 SV40 핵 국재화 신호 (NLS)를 갖는 Cas9를 암호화하는 오픈 리딩 프레임(open reading frame)은 연회색으로 도시되어 있다.

도 2b: 야로위아-최적화된 예비-sgRNA(pre-sgRNA) RGR 발현 카세트 (RGR, 리보자임-sgRNA-리보자임). FBA1 프로모터는 검정색으로 도시되어 있고, 망치머리 (HH) 리보자임은 진회색으로 도시되어 있으며, 단일 가이드 RNA (sgRNA)는 연회색으로 도시되어 있고, HDV 리보자임은 세로 줄무늬로 도시되어 있다.

도 2c: 야로위아-최적화된 예비-sgRNA RG 발현 카세트 (RG, 리보자임-sgRNA). FBA1 프로모터는 검정색으로 도시되어 있고, 망치머리 (HH) 리보자임은 진회색으로 도시되어 있으며, 단일 가이드 RNA (sgRNA)는 연회색으로 도시되어 있고, Sup4 종결자는 세로 줄무늬로 도시되어 있다.

도 3a: pZUFCas9 (서열 번호: 14) 플라스미드는 도 2a에 나타낸 야로위아 코돈-최적화된 Cas9 발현 카세트를 함유한다. 복제 기점 (ARS 18, f1 ori, ColE1)은 크로스-해칭(cross-hatch)으로, 선택가능한 마커 (Ura3, Amp)는 회색으로 도시되어 있다.

도 3b: pZUFCas9/PolIII-sgRNA 플라스미드는 도 2a에 나타낸 야로위아 코돈-최적화된 Cas9 발현 카세트 및 야로위아 내의 Leu2-3을 표적화하기 위한 Y1 Snr52 ("Y152"로 나타낸 Pol III 프로모터)-sgRNA 발현 카세트 함유한다. 도시되어 있지는 않지만, sgRNA 카세트는 또한 사카로마이세스 세레비시아 Sup4 유전자 전사 종결 서열을 함유하였다. 복제 기점 (ARS 18, f1 ori, ColE1)은 크로스-해칭으로, 선택가능한 마커 (Ura3, Amp)는 회색으로 도시되어 있다.

도 3c: pRF38 플라스미드 (서열 번호:19)는 야로위아 리폴리티카 내의 CAN1 유전자를 표적화하기 위한 서열 번호:18의 야로위아-최적화된 예비-sgRNA 발현 카세트 (FBA1 프로모터는 흰색으로, RGR 예비-sgRNA는 대각선으로 도시됨)를 함유한다. 복제 기점 (ARS 18, f1 ori, ColE1)은 크로스-해칭으로, 선택가능한 마커 (Ura3, Amp)는 회색으로 도시되어 있다.

도 4b: (i) pZUFCas9 (서열 번호:14) 단독으로 또는 (ii) pZUFCas9 및 서열 번호:18의 야로위아-최적화된 예비-sgRNA 발현 카세트를 포함하는 선형 DNA (실시예 3 참조)로 형질전환된 야로위아 리폴리티카 세포 내의 일과성 표적화 효율(Transient targeting efficiency). y축은 카나바닌-내성 (Can^R)이기도 한, pZUFCas9로 형질전환된 세포 (즉, Ura^+ 세포)의 빈도를 나타낸다. 오차 막대는 표준 편차를 나타낸다.

도 5: pZUFCas9 (서열 번호:14) 및 서열 번호:18의 야로위아-최적화된 예비-sgRNA 발현 카세트를 포함하는 선형 DNA (실시예 3 참조)로 형질전환된 야로위아 리폴리티카 세포의 CAN1 암호화 영역 내의 Cas9/sgRNA 절단 부위의 서열 지도. 야생형 (WT) CAN1 서열과 관련하여, Can1-1 표적 부위 서열은 볼드로 도시되어 있고, PAM 서열은 밑줄이 쳐져 있다. 예측되는 절단 부위는 PAM의 상류에서 세 번째 뉴클레오티드의 5' 부근(immediately 5') 이다. 삽입된 뉴클레오티드는 이탤릭체이다. 돌연변이체 (1 내지 18)의 각각의 클래스의 수와 빈도가 우측에 나타나 있다. 이 도면에 도시된 서열은 도면에 번호가 매겨진 바와 같이 서열 번호:71 내지 89로 서열 목록에 포

함되어 있다.

도 6: (i) pZUFCas9 (서열 번호:14) 단독으로, (ii) pZUFCas9 및 서열 번호:18 (RGR)의 야로위아-최적화된 예비-sgRNA 발현 카세트를 포함하는 선형 DNA 또는 (iii) pZUFCas9 및 서열 번호:25 (RG)의 야로위아-최적화된 예비-sgRNA 발현 카세트를 포함하는 선형 DNA (실시예 4 참조)로 형질전환된 야로위아 리폴리티카 세포 내의 일과성 표적화 효율. y축은 카나바닌-내성 (Can^R)이기도 한, pZUFCas9로 형질전환된 세포 (즉, Ura^+ 세포)의 빈도를 나타낸다. 오차 막대는 표준 편차를 나타낸다.

도 7: HR 및 NHEJ DNA 수복 경로에 의한 돌연변이 빈도의 비교. 형질전환에서 폴리뉴클레오티드 변형 주형 DNA 서열이 제공된 경우의 HR (진회색) 및 NHEJ (연회색)에 의한 Cas9/sgRNA-매개 DNA 이중 가닥 절단 수복의 총 빈도가 결정되었다 (실시예 5 참조). 오차 막대는 표준 편차를 나타낸다.

도 8: 폴리뉴클레오티드 변형 주형 DNA 서열의 유형에 의한 Cas9/sgRNA-매개 DNA 이중 가닥 절단 부위에서의 HR의 빈도. 점 돌연변이 주형 DNA(진회색), 프레임시프트 주형 DNA (연회색) 및 큰 결실 주형 DNA (흰색)을 사용한 HR 빈도가 도시되어 있다 (실시예 5 참조). 오차 막대는 표준 편차를 나타낸다.

도 9: 야로위아의 CAN1 유전자좌 (Cas9/sgRNA에 의해 절단된 Can1-1 부위에서의 수복)에서의 돌연변이 빈도는 폴리뉴클레오티드 변형 주형 DNA의 존재에 영향을 받지 않는다. 폴리뉴클레오티드 변형 주형 DNA를 포함하지 않거나 (진회색, 주형 DNA 없음), 폴리뉴클레오티드 변형 주형 DNA를 포함하는 (연회색, 주형 DNA를 포함함) 형질전환으로부터 생성되는 세포의 카나바닌-내성 빈도 (두 형질전환 군은 모두 pZUFCas9 (서열 번호:14) 및 RGR 발현 카세트 [서열 번호:18]를 포함함) (실시예 5 참조). y축은 카나바닌-내성 (Can^R)이기도 한, pZUFCas9로 형질전환된 세포 (즉, Ura^+ 세포)의 빈도를 나타낸다. 오차 막대는 표준 편차를 나타낸다.

도 10a: pRF84 플라스미드 (서열 번호:41)는 도 2a에 나타낸 야로위아 코돈-최적화된 Cas9 발현 카세트 및 서열 번호:18의 야로위아-최적화된 RGR 예비-sgRNA 카세트 (대각선으로 도시된 RGR 예비-sgRNA 암호화 영역 ["Can1 RGR"])를 함유한다. 복제 기점 (ARS 18, f1 ori, ColE1)은 크로스-해칭으로, 선택가능한 마커 (Ura3, Amp)는 회색으로 도시되어 있다.

도 10b: pRF85 플라스미드 (서열 번호:42)는 도 2a에 나타낸 야로위아 코돈-최적화된 Cas9 발현 카세트 및 서열 번호:25의 야로위아-최적화된 RG 예비-sgRNA 카세트 (대각선으로 도시된 RG 예비-sgRNA 암호화 영역 ["Can1 RG"])를 함유한다. 복제 기점 (ARS 18, f1 ori, ColE1)은 크로스-해칭으로, 선택가능한 마커 (Ura3, Amp)는 회색으로 도시되어 있다.

도 11: Cas9 단독으로 (pZUFCas9, 서열 번호:14) 발현, 또는 (i) Cas9 및 (ii) RGR 예비-sgRNA (pRF84) 또는 RG sgRNA (pRF85)를 발현에 의한 야로위아의 CAN1 유전자좌에서의 돌연변이 빈도 (실시예 6 참조). y축은 카나바닌-내성 (Can^R)이기도 한, 각각의 벡터로 형질전환된 세포 (즉, Ura^+ 세포)의 빈도를 나타낸다. 오차 막대는 표준 편차를 나타낸다.

도 12a 및 도 12b: HDV-sgRNA 융합 발현 카세트를 제작하기 위한 고효율 클로닝 카세트의 예. 도 12-a는 프로모터 서열을 검정색 박스로, HDV 리보자임을 암호화하는 DNA 서열을 회색 박스로, 타입 II 제한 부위에 의해 플랭킹된(flanked) 클로닝 균주에 대해 역선별가능한 마커를 수평으로 해칭된 박스로, Cas9와 상호작용하기 위한 sgRNA의 CER 도메인을 검은 점선 박스로, 전사 종결자를 대각선으로 해칭된 박스로 도시한다. 가변 표적화 도메인을 암호화하는 DNA 서열 및 타입 II 제한 부위 (수직으로 해칭된 박스 VT)에 적절한 돌출부를 함유하는 DNA 듀플렉스가 플라스미드, DNA 리가제 및 타입 II 효소와 혼합되는 경우, 가변 표적화 도메인 (VT)을 암호화하는 DNA 서열이 역선별가능한 마커를 대체함으로써, HDV-sgRNA 발현 카세트 (프로모터-HDV-VT-CER-종결자)를 생성한다. HDV-sgRNA 발현 카세트가 전사되는 경우, 이는 HDV 리보자임이 임의의 5' 서열 절단하는 RNA 전사물 (HDV-VT-CER 전사물)을 생성한다. 도 12-b는 Can1-1 표적 부위를 암호화하는 DNA 서열 및 플라스미드 pRF291로의 클로닝에 적절한 돌출부를 함유하는 이중 가닥 DNA 분자의 예를 (서열 번호: 99 및 서열 번호: 100의 올리고듀플렉스)를 도시한다.

도 13a 및 도 13b: HH-sgRNA 발현 카세트를 제작하기 위한 고효율 클로닝 카세트의 예. 도 13은 프로모터 서열을 검정색 박스로; 타입 II 제한 부위에 의해 플랭킹된 클로닝 균주에 대해 역선별가능한 마커를 수평으로 해칭된 박스로; Cas9와 상호작용하기 위한 sgRNA의 CER 도메인을 검은 점선 박스로, 전사 종결자를 대각선으로 해칭된 박스로 도시한다. DNA를 암호화하는 표적-부위 특이적 망치머리 리보자임을 함유하는 DNA 듀플렉스 (수직으로 해칭된 박스 HH), 표적화 서열 및 타입 II 부위에 적절한 돌출부 (점선 박스 TS)가 플라스미드, DNA 리가제

및 타입-II 효소와 혼합되는 경우, HH-표적 부위 듀플렉스가 역선별가능한 마커를 대체하여, HH-sgRNA 발현 카세트를 생성한다. 발현 카세트가 전사되는 경우, 이는 전사물을 생성하고, HH 리보자임은 그 자체와 임의의 5' 서열을 절단한다. 도 13b는 ds-주형(temp)-1 표적 부위 (VT) 및 서열 특이적 HH 리보자임을 암호화하는 DNA (HH)를 표적화하기 위한 가변 표적화 도메인 및 플라스미드 pRF291로의 클로닝에 적절한 돌출부를 함유하는 (서열 번호: 162 및 서열 번호: 163의) 이중 가닥 DNA 분자의 예를 도시한다.

도 14: pRF303 (서열 번호: 103) 및 Can1 짧은 편집 주형 (서열 번호: 157)으로 형질전환된 세포로부터의 Can1 유전자좌의 겔 전기영동의 예. MW로 표시된 레인은 분자량 마커이다. 레인 1 내지 레인 16은 도말 정제된 (streak purified) 형질전환체로부터의 개별 콜로니를 나타낸다. 더 높은 MW 밴드는 WT Can1 유전자좌 (서열 번호: 160) 또는 작은 삽입-결실(indel) 돌연변이를 갖는 Can1 유전자좌에 대한 정확한 크기이다. 더 적은 분자량 밴드는 짧은 Can1 편집 주형 (서열 번호: 157)으로 편집된 Can1 유전자좌 (서열 번호: 161)에 대한 정확한 크기이다.

도 15는 콜로니 PCR로부터의 플라스미드 및 게놈 URA3 유전자의 대표적인 서열 분석 결과와 그들의 정렬을 도시한다. 대시와 볼드는 각각 결실과 삽입을 나타낸다. PAM 서열은 밑줄이 쳐져 있다.

도 16-a는 야로위아 URA3 유전자 내에서 RGR-URA3.1, RGR-URA3.2 및 RGR-URA3.3에 대한 표적 서열의 상대적인 위치를 도시한다.

도 16-b는 5-FOA를 함유하는 SC 배지에서 성장한 pYRH222 형질전환체의 콜로니 PCR의 서열 분석 결과 및 서열 정렬을 도시한다. 볼드는 삽입을 나타낸다. PAM 서열은 밑줄이 쳐져 있다. "N"은 혼합된 서열을 나타낸다.
도 16-c는 5-FOA를 함유하는 SC 배지에서 성장한 pYRH282 형질전환체의 콜로니 PCR의 서열 분석 결과 및 서열 정렬을 도시한다. 점선은 결실을 나타낸다. PAM 서열은 밑줄이 쳐져 있다. "N"은 혼합된 서열을 나타낸다.
도 16-d는 5-FOA를 함유하는 SC 배지에서 성장한 pYRH283 형질전환체의 콜로니 PCR의 서열 분석 결과 및 서열 정렬을 도시한다. 점선은 결실을 나타낸다. PAM 서열은 밑줄이 쳐져 있다. "N"은 혼합된 서열을 나타낸다.

도 17은 pYRH282 (콜로니 번호 23 및 24) 및 pYRH283 (콜로니 번호 27 및 36) 형질전환체로부터의 PCR 생성물 상이한 이동을 도시한다. 래더(ladder)의 DNA 크기는 우측에 나타나 있다.

도 18은 Can1 표적 서열의 대표적인 서열 분석 결과를 도시한다. 대시는 각각 결실을 나타낸다. PAM 서열은 볼드로 나타나 있다.

[표 1]

핵산 및 단백질 서열 번호의 요약

설명	핵산 서열 번호	단백질 서열 번호
gRNA 의 Cas9 엔도뉴클레아제 인식 (CER) 도메인.	1 (80 개의 염기)	
야로위아 리폴리티카 Leu2-1 표적 부위 또는 대안적으로 gRNA 의 Leu2-1 가변 표적 도메인을 암호화하는 DNA.	2 (20 개의 염기)	
야로위아 리폴리티카 Leu2-2 표적 부위 또는 대안적으로 gRNA 의 Leu2-2 가변 표적 도메인을 암호화하는 DNA.	3 (20 개의 염기)	
야로위아 리폴리티카 Leu2-3 표적 부위 또는 gRNA 의 Leu2-2 가변 표적 도메인을 암호화하는 DNA.	4 (20 개의 염기)	
S. 세레비시아 Snr52 프로모터.	5 (300 개의 염기)	
S. 세레비시아 Rpr1 프로모터.	6 (300 개의 염기)	
야로위아 리폴리티카 Snr52 프로모터.	7 (300 개의 염기)	
S. 세레비시아 Sup4 종결자.	8 (20 개의 염기)	
스트렙토코커스 피오제네스(*Streptococcus pyogenes*) Cas9; 야로위아 리폴리티카에서 발현에 대해 코돈-최적화된 오픈 리딩 프레임.	9 (4107 개의 염기)	
C-말단 링커 및 SV40 NLS (“Cas9-NLS”)를 포함하는 스트렙토코커스 피오제네스 Cas9; 야로위아 리폴리티카에서 발현에 대해 코돈-최적화된 오픈 리딩 프레임.	10 (4140 개의 염기)	11 (1379 개의 aa)
야로위아 리폴리티카 FBA1 프로모터.	12 (543 개의 염기)	
Cas9-NLS 발현 카세트 (프로모터 및 Cas9-NLS 오픈 리딩 프레임).	13 (4683 개의 염기)	
pZUFCas9 플라스미드.	14 (10706 개의 염기)	
망치머리 (HH) 리보자임.	15 (43 개의 염기)	
HDV 리보자임.	16 (68 개의 염기)	
야로위아 리폴리티카 Can1-1 표적 부위 또는 대안적으로 gRNA 의 Can1-1 가변 표적 도메인을 암호화하는 DNA.	17 (20 개의 염기)	
FBA1 프로모터:HH-sgRNA-HDV (RGR) 예비-sgRNA 발현 카세트 또는 대안적으로, “RGR” 발현 카세트.	18 (760 개의 염기)	
pRF38 플라스미드.	19 (6793 개의 염기)	

RGR 정방향 PCR 프라이머.	20 (19 개의 염기)	
RGR 역방향 PCR 프라이머.	21 (19 개의 염기)	
CAN1 정방향 PCR 프라이머.	22 (20 개의 염기)	
CAN1 역방향 PCR 프라이머.	23 (21 개의 염기)	
CAN1 서열 분석 프라이머.	24 (21 개의 염기)	
FBA1 프로모터:HH-sgRNA-Sup4 종결자 (RG) 예비-sgRNA 발현 카세트 또는 대안적으로, “RG” 발현 카세트.	25 (709 개의 염기)	
폴리-A.	26 (10 개의 염기)	
폴리-T.	27 (10 개의 염기)	
CAN1 프레임시프트 주형 DNA.	28 (100 개의 염기)	
CAN1 프레임시프트 주형 DNA 보체.	29 (100 개의 염기)	
CAN1 점 돌연변이 주형 DNA.	30 (106 개의 염기)	
CAN1 점 돌연변이 주형 DNA 보체.	31 (106 개의 염기)	
CAN1 상류 주형 암(arm).	32 (655 개의 염기)	
CAN1 상류 주형 암을 증폭시키기 위한 정방향 PCR 프라이머.	33 (29 개의 염기)	
CAN1 상류 주형 암을 증폭시키기 위한 역방향 PCR 프라이머.	34 (37 개의 염기)	
CAN1 하류 주형 암.	35 (658 개의 염기)	
CAN1 하류 주형 DNA 암을 증폭시키기 위한 정방향 PCR 프라이머.	36 (37 개의 염기)	
CAN1 하류 주형 DNA 암을 증폭시키기 위한 역방향 PCR 프라이머.	37 (22 개의 염기)	
CAN1 큰 결실 주형 DNA.	38 (1276 개의 염기)	
RG/RGR 정방향 PCR 프라이머.	39 (31 개의 염기)	
RG/RGR 역방향 PCR 프라이머.	40 (29 개의 염기)	
pRF84 플라스미드.	41 (11568 개의 염기)	
pRF85 플라스미드.	42 (11507 개의 염기)	
RNA 루프 형성 서열 (GAAA).	43 (4 개의 염기)	

RNA 루프 형성 서열 (CAAA).	44 (4 개의 염기)	
RNA 루프 형성 서열 (AAAG).	45 (4 개의 염기)	
Cas9 표적 부위의 예: PAM 서열.	46 (23 개의 염기)	
PAM 서열 NGG.	47 (3 개의 염기)	
PAM 서열 NNAGAA.	48 (6 개의 염기)	
PAM 서열 NNAGAAW.	49 (7 개의 염기)	
PAM 서열 NGGNG.	50 (5 개의 염기)	
PAM 서열 NNNNGATT.	51 (8 개의 염기)	
PAM 서열 NAAAAC.	52 (6 개의 염기)	
PAM 서열 NG.	53 (2 개의 염기)	
TracrRNA 메이트 서열(mate sequence) 예 1.	54 (22 개의 염기)	
TracrRNA 메이트 서열 예 2.	55 (15 개의 염기)	
TracrRNA 메이트 서열 예 3.	56 (12 개의 염기)	
TracrRNA 메이트 서열 예 4.	57 (13 개의 염기)	
TracrRNA 실시예 1.	58 (60 개의 염기)	
TracrRNA 실시예 2.	59 (45 개의 염기)	
TracrRNA 실시예 3.	60 (32 개의 염기)	
TracrRNA 실시예 4.	61 (85 개의 염기)	
TracrRNA 실시예 5.	62 (77 개의 염기)	
TracrRNA 실시예 6.	63 (65 개의 염기)	
gRNA 실시예 1.	64 (131 개의 염기)	
gRNA 실시예 2.	65 (117 개의 염기)	
gRNA 실시예 3.	66 (104 개의 염기)	
gRNA 실시예 4.	67 (99 개의 염기)	
gRNA 실시예 5.	68	

	(81 개의 염기)	
gRNA 실시예 6.	69 (68 개의 염기)	
gRNA 실시예 7.	70 (100 개의 염기)	
도 5 에 도시된 WT 서열.	71	
도 5 에 도시된 서열 1.	72	
도 5 에 도시된 서열 2.	73	
도 5 에 도시된 서열 3.	74	
도 5 에 도시된 서열 4.	75	
도 5 에 도시된 서열 5.	76	
도 5 에 도시된 서열 6.	77	
도 5 에 도시된 서열 7.	78	
도 5 에 도시된 서열 8.	79	
도 5 에 도시된 서열 9.	80	
도 5 에 도시된 서열 10.	81	
도 5 에 도시된 서열 11.	82	
도 5 에 도시된 서열 12.	83	
도 5 에 도시된 서열 13.	84	
도 5 에 도시된 서열 14.	85	
도 5 에 도시된 서열 15.	86	
도 5 에 도시된 서열 16.	87	
도 5 에 도시된 서열 17.	88	
도 5 에 도시된 서열 18.	89	
프라이머 AarI-제거-1	90	
프라이머 AarI-제거-2	91	
플라스미드 pRF109	92	
변형된 Aar1- Cas9 유전자	93	
플라스미드 pRF141	94	
고효율 클로닝 카세트	95	
yl52 프로모터	96	
대장균(*Escherichia coli*) 역선별 카세트 rpsL	97	
플라스미드 pRF291	98	
올리고뉴클레오티드 Can1-1F	99	
올리고뉴클레오티드 Can1-1R	100	
Can1-1 표적 부위 및 PAM 서열	101	
Can1-1 을 표적화하기 위한 재조합 HDV-sgRNA 발현 카세트	102	
플라스미드 pRF303	103	
HDV 리보자임-가이드 RNA	104	
야로위아 리폴리티카로부터의 Can1 유전자	105	
Can1-2 표적 부위	106	
Sou2-1 표적 부위	107	
Sou2-2 표적 부위	108	
Can1-2 의 가변 표적화 도메인	109	

Sou2-1 의 가변 표적화 도메인	110	
Sou2-2 의 가변 표적화 도메인	111	
Tgl1-1 표적 부위	112	
Acos10-1 표적 부위	113	
Fat1-1 표적 부위	114	
ura3-1 의 가변 표적화 도메인	115	
URa3-1 표적 부위	116	
Cas9-SV40 NLS D10A H840A		117
프라이머 D10AF	118	
프라이머 D10AR	119	
야로위아 최적화된 Cas9 D10A 유전자	120	
플라스미드 pRF111	121	
프라이머 H840A1	122	
프라이머 H840A2	123	
야로위아 코돈 최적화된 불활성화된 Cas9 유전자	124	
pRF143	125	
야로위아 최적화된 dsRED 익스프레스(express) ORF	126	
야로위아 최적화된 dsRED 익스프레스 클로닝 단편	127	
FBA1-dsRED 익스프레스 발현 카세트	128	
pRF165	129	
PmeI NotI 단편 상의 pRF165 로부터의 FBA1 야로위아 dsRED 익스프레스 카세트	130	
p2PO69 통합 벡터	131	
pRF201	132	
pRF201 로부터의 AscI/SphI 통합 단편	133	
HY026	134	
HY027	135	
pRF169	136	
GPD 프로모터	137	
GPD 프로모터-역선별가능한 마커-CER-종결자	138	
ds-주형-1 표적 부위	139	
ds-주형-2 표적 부위	140	
ds-비주형-3 표적 부위	141	
망치머리 리보자임-VTD 융합체	142	
망치머리 리보자임-VTD 융합체	143	
ds-주형-1F	144	
ds-주형-1R	145	
ds-주형-2F	146	
ds-주형-2R	147	
ds-비주형-1F	148	
ds-비주형-1R	149	
pRF296	150	
pRF298	151	
pRF300	152	
pRF339	153	
pRF341	154	
pRF343	155	

pRF80	156	
짧은 Can1 결실 편집 주형	157	
프라이머 80F	158	
프라이머 80R	159	
Can1 유전자좌 WT (야생형)	160	
Can1 유전자좌 결실 균주	161	
도 13-B 의 정방향 올리고뉴클레오티드	162	
도 13-B 의 역방향 올리고뉴클레오티드	163	
예비-sgRNA URA3.1 (RGR-URA3.1)	164	
URA3.1 표적 서열	165	
예비-sgRNA URA3.2 (RGR-URA3.2	166	
URA3.2 표적 서열	167	
FBA1L 프로모터	168	
아세토하이드록시산 합성효소 유전자	169	
프라이머 RHO705	170	
프라이머 RHO719	171	
프라이머 RHO733	172	
프라이머 RHO734	173	
프라이머 RHO707	174	
야생형 URA3 서열의 단편	175	
콜로니 1 로부터의 플라스미드 URA3 의 단편	176	
콜로니 2 로부터의 플라스미드 URA3 의 단편	177	
콜로니 3 으로부터의 플라스미드 URA3 의 단편	178	
콜로니 5 로부터의 플라스미드 URA3 의 단편	179	
콜로니 6 으로부터의 플라스미드 URA3 의 단편	180	
콜로니 1 로부터의 게놈 URA3 의 단편	181	
콜로니 2 로부터의 게놈 URA3 의 단편	182	
콜로니 3 으로부터의 게놈 URA3 의 단편	183	
콜로니 5 로부터의 게놈 URA3 의 단편	184	
콜로니 6 으로부터의 게놈 URA3 의 단편	185	
하이그로마이신 항생제 내성 선택 마커	186	
TDH1 또는 GPD 프로모터	187	
프라이머 RHO804	188	
프라이머 RHO805	189	
TDH1 프로모터-RGR-URA3.3 융합	190	
예비-sgRNA URA3.3 (RGR-URA3.3)	191	
프라이머 RHO610	192	
프라이머 RHO611	193	
프라이머 RHO704	194	
야생형 URA3 서열의 단편	195	
콜로니 3 으로부터의 URA3 서열의 단편	196	
콜로니 4 로부터의 URA3 서열의 단편	197	
콜로니 5 로부터의 URA3 서열의 단편	198	
콜로니 6 으로부터의 URA3 서열의 단편	199	
콜로니 9 로부터의 URA3 서열의 단편	200	
콜로니 10 으로부터의 URA3 서열의 단편	201	

야생형 URA3 서열의 단편	202	
콜로니 23으로부터의 URA3 서열의 단편	203	
콜로니 24로부터의 URA3 서열의 단편	204	
야생형 URA3 서열의 단편	205	
콜로니 27로부터의 URA3 서열의 단편	206	
콜로니 36으로부터의 URA3 서열의 단편	207	
ARS18 서열	208	
야로위아 코돈 최적화된 녹농균(*P. aeruginosa*) Csy4	209	
야로위아 FBA1 프로모터	210	
TDH1:28bp-gCAN1-28bp	211	
Csy4 인식 서열	212	
Csy4 인식 서열 플랭킹된 sgRNA	213	
CAN1 표적 서열	214	
야생형 CAN1 서열의 단편	215	
콜로니 14로부터의 CAN1의 단편	216	
콜로니 16으로부터의 CAN1의 단편	217	
콜로니 18로부터의 CAN1의 단편	218	
콜로니 19로부터의 CAN1의 단편	219	
콜로니 24로부터의 CAN1의 단편	220	
콜로니 25로부터의 CAN1의 단편	221	
Csy4에 의해 처리된 sgRNA	222	
Csy4 절단 후 5'-플랭킹 서열	223	
Csy4 절단 후 3'-플랭킹 서열	224	

발명을 실시하기 위한 구체적인 내용

[0025] 모든 인용된 특허 및 비특허 문헌의 개시 내용은 전체적으로 본 명세서에 참고로 포함된다.

[0026] 본 명세서에 사용되는 바와 같이, 용어 "발명" 또는 "개시된 발명"은 제한적인 것으로 의미되지 않고 청구범위에 정의되거나 본 명세서에 기재된 발명들 중 임의의 것에 일반적으로 적용된다. 이들 용어는 본 명세서에서 상호교환가능하게 사용된다.

[0027] 본 명세서에서 용어 "통상적이지 않은 효모"는 사카로마이세스 (예를 들어, S. 세레비시아) 또는 스키조사카로마이세스(*Schizosaccharomyces*) 효모 종이 아닌 임의의 효모를 지칭한다. 통상적이지 않은 효모는, 본 명세서에 참고로 포함되는 문헌[Non-Conventional Yeasts in Genetics, Biochemistry and Biotechnology: Practical Protocols (K. Wolf, K.D. Breunig, G. Barth, Eds., Springer-Verlag, Berlin, Germany, 2003)]에 기재되어 있다. 소정 실시 형태에서 통상적이지 않은 효모는 추가로 (또는 대안적으로) 상동 재조합 (HR)에 의해 매개되는 수복 공정보다 비상동 말단 결합 (non-homologous end-joining; NHEJ) DNA 수복 공정을 선호하는 효모일 수 있다. 이러한 맥락에서 -HR보다 NHEJ를 선호- 통상적이지 않은 효모의 정의가 본 명세서에 참고로 포함되는 문헌 [Chen et aL. (*PLoS ONE* 8:e57952)]에 또한 개시되어 있다. 본 명세서에서 바람직한 통상적이지 않은 효모는 야로위아 속 (예를 들어, 야로위아 리폴리티카)의 것들이다. 본 명세서에서 용어 "효모"는 주로 단세포 형태로 존재하는 진균종을 지칭한다. 본 명세서에서 효모는 대안적으로 "효모 세포"로 지칭될 수 있다.

[0028] 본 명세서에서 용어 "RNA-유도 엔도뉴클레아제" (RGEN)는 적어도 하나의 크리스퍼 (규칙적으로 사이 간격을 두고 분포하는 짧은 회문구조 반복 서열)-관련 (Cas) 단백질 및 적어도 하나의 RNA 성분을 포함하는 복합체를 지칭한다. 간략하게, RGEN의 RNA 성분은 표적 부위 서열의 DNA 서열에 상보적인 서열을 함유한다. 이러한 상보성에 기초하여, RGEN은 특정 DNA 표적 부위는 서열을 특이적으로 인식하여 절단할 수 있다. 본 명세서에서 RGEN은 4개의 알려진 크리스퍼 시스템, 예컨대 타입 I, II 또는 III 크리스퍼 시스템 (문헌[Horvath and Barrangou, *Science* 327:167-170]) 중 어느 하나의 Cas 단백질(들)과 적합한 RNA 성분(들)을 포함할 수 있다. 바람직한 실시 형태의 RGEN은 Cas9 엔도뉴클레아제 (크리스퍼 II 시스템)와 적어도 하나의 RNA 성분 (예를 들어, crRNA 및 tracrRNA, 또는 gRNA)을 포함한다.

[0029] 용어 "크리스퍼" (규칙적으로 사이 간격을 두고 분포하는 짧은 회문구조 반복 서열)는, 예를 들어, 박테리아 및 고세균 세포에 의해 외래 DNA를 파괴하는데 사용되는 유형 I, II 또는 III DNA 절단 시스템의 인자를 암호화하는 소정 유전자좌(genetic loci)를 지칭한다 (문헌[Horvath and Barrangou, *Science* 327:167-170]). 크리스퍼 시스템의 성분은 본 명세서에서 통상적이지 않은 효모 세포에서 DNA를 표적화하기 위해 이용된다.

[0030] 용어 "타입 II 크리스퍼 시스템" 및 "타입 II 크리스퍼-Cas 시스템"은 본 명세서에서 상호교환가능하게 사용되며, 적어도 하나의 RNA 성분과 복합체인 Cas9 엔도뉴클레아제를 이용하는 DNA 절단 시스템을 지칭한다. 예를

들어, Cas9는 크리스퍼 RNA (crRNA) 및 트랜스-활성화 크리스퍼 RNA (tracrRNA)와 복합체일 수 있다. 또 다른 예에서, Cas9는 가이드 RNA와 복합체일 수 있다. 따라서, crRNA, tracrRNA 및 가이드 RNA는 본 명세서에서 RNA 성분의 비제한적인 예이다.

[0031] 용어 크리스퍼-관련 ("Cas") 엔도뉴클레아제는 본 명세서에서 Cas 유전자에 의해 암호화되는 Cas 단백질을 지칭한다. 적합한 RNA 성분과 복합체인 경우, Cas 엔도뉴클레아제는 소정 실시 형태에서 특정 DNA 표적 서열의 전부 또는 일부를 절단할 수 있다. 예를 들어, 특정 DNA 표적 서열에 단일- 또는 이중 가닥 절단을 도입할 수 있거나; 대안적으로 특정 DNA 표적 서열의 하나 또는 두 가닥을 절단할 수 있는 것을 특징으로 할 수 있다. Cas 엔도뉴클레아제는 표적 서열에서 DNA 이중 가닥을 풀고, Cas와 복합체인 crRNA 또는 가이드 RNA에 의한 표적 서열의 인식에 의해 매개되는 적어도 하나의 DNA 가닥을 절단한다. Cas 엔도뉴클레아제에 의한 이러한 표적 서열의 인식 및 절단은 전형적으로 정확한 프로토스페이서-인접 모티프 (PAM)가 DNA 표적 서열의 3' 말단에 위치하거나 인접한 경우 발생한다. 대안적으로, 본 명세서에서 Cas 단백질은 DNA 절단 또는 닉킹(nicking) 활성이 부족할 수 있지만, 적합한 RNA 성분과 복합체를 형성한 경우 여전히 DNA 표적 서열에 특이적으로 결합할 수 있다. 본 명세서에서 바람직한 Cas 단백질은 Cas9이다.

[0032] 본 명세서에서 "Cas9" (이전에 Cas5, Csn1 또는 Csx12로 지칭됨)는 DNA 표적 서열의 전부 또는 일부를 특이적으로 인식하고 절단하기 위해, crRNA 및 tracrRNA와, 또는 가이드 RNA와 복합체를 형성하는 타입 II 크리스퍼 시스템의 Cas 엔도뉴클레아제를 지칭한다. Cas9 단백질은 RuvC 뉴클레아제 도메인 및 HNH (H-N-H) 뉴클레아제 도메인을 포함하며, 이들 각각은 표적 서열에서 단일 DNA 가닥을 절단한다 (두 도메인의 협동 작용은 DNA 이중 가닥 절단을 유도하는 반면, 하나의 도메인의 활성은 닉(nick)을 유도한다). 일반적으로, RuvC 도메인은 서브도메인(subdomain) I, II 및 III을 포함하며, 여기서 도메인 I은 Cas9의 N-말단 근처에 위치하고, 서브도메인 II 및 III은 HNH 도메인에 플랭킹한 단백질의 중간에 위치한다 (문헌[Hsu et al, *Cell* 157:1262-1278]). "Apo-Cas9"는 RNA 성분과 복합체가 형성되지 않은 Cas9를 지칭한다. Apo-Cas9는 DNA를 결합할 수 있지만, 비특이적 방식으로 결합하여, DNA를 절단할 수 없다 (문헌[Sternberg et al., *Nature* 507:62-67]).

[0033] 일부 실시 형태에서, Cas 엔도뉴클레아제는 Cas9 폴리펩티드의 변형된 형태를 포함할 수 있다. Cas9 폴리펩티드의 변형된 형태는 Cas9 단백질의 자연 발생 뉴클레아제 활성을 감소시키는 아미노산 변화 (예를 들어, 결실, 삽입 또는 치환)를 포함할 수 있다. 예를 들어, 일부 경우에, Cas9 단백질의 변형된 형태는 상응하는 야생형 Cas9 폴리펩티드의 뉴클레아제 활성의 50% 미만, 40% 미만, 30% 미만, 20% 미만, 10% 미만, 5% 미만 또는 1% 미만을 갖는다 (2014년 3월 6일자로 출원된 미국 특허 출원 제US20140068797 A1호). 일부 경우에서, Cas9 폴리펩티드의 변형된 형태는 실질적인 뉴클레아제 활성을 갖지 않으며, 촉매적으로 "불활성화된 Cas9" 또는 "비활성화된(deactivated) cas9 (dCas9)"로 지칭된다. 촉매적으로 불활성화된 Cas9 변이체는 HNH 및 RuvC 뉴클레아제 도메인에서 돌연변이를 함유하는 Cas9 변이체를 포함한다. 이러한 촉매적으로 불활성화된 Cas9 변이체는 sgRNA와 상호작용하고 생체 내 표적 부위에 결합할 수 있지만, 표적 DNA의 어느 한 가닥을 절단할 수는 없다. DNA에 결합하지만 절단하지 않는 이러한 작용 방식은 영구적인 유전적 변화를 일으키지 않으면서, 염색체의 특정 유전자좌의 발현을 일시적으로 감소시키는 데 사용될 수 있다.

[0034] 촉매적으로 불활성인 Cas9는 이종 서열에 융합될 수 있다 (2014년 3월 6일자로 출원된 미국 특허 출원 제US20140068797 A1호). 적합한 융합 파트너에는 표적 DNA 상에 또는 표적 DNA와 결합한 폴리펩티드 (예를 들어, 히스톤 또는 다른 DNA-결합 단백질) 상에 직접 작용함으로써 간접적으로 전사를 증가시키는 활성을 제공하는 폴리펩티드가 포함되나, 이에 한정되지 않는다. 추가의 적합한 융합 파트너에는 메틸트랜스퍼라제 활성, 데메틸라제 활성, 아세틸트랜스퍼라제 활성, 데아세틸라제 활성, 키나제 활성, 포스파타제 활성, 유비퀴틴 리가제 활성, 탈유비퀴틴화(deubiquitinating) 활성, 아데닐화 활성, 탈아데닐화 활성, 수모화(SUMOylating) 활성, 탈수모화 활성, 리보실화 활성, 탈리보실화 활성, 미리스토일화 활성 또는 탈미리스토일화 활성을 제공하는 폴리펩티드가 포함되나, 이에 한정되지 않는다 추가로 적합한 융합 파트너에는 표적 핵산의 증가된 전사를 직접적으로 제공하는 폴리펩티드 (예를 들어, 전사 활성 인자 또는 이의 단편, 전사 활성 인자를 동원하는(recruit) 단백질 또는 이의 단편, 소분자/약물-반응성 전사 조절자 등)가 포함되나, 이에 한정되지 않는다. 촉매적으로 불활성인 Cas9는 또한 FokI 뉴클레아제에 융합되어 이중 가닥 절단을 생성할 수 있다 (문헌[Guilinger et aL. Nature biotechnology, volume 32, number 6, June 2014]).

[0035] 본 명세서에서 용어 "RNA 성분"은 DNA 표적 서열의 가닥에 상보적인 리보핵산 서열을 함유하는 RGEN의 RNA 성분을 지칭한다. 본 명세서에서 이러한 상보적인 서열은 "가이드 서열" 또는 "가변 표적화 도메인" 서열로 지칭된다. 본 명세서에서 적합한 RNA 성분의 예는 crRNA 및 가이드 RNA를 포함한다. 또한, 본 명세서에서 RNA 성분

은 5'-캡을 가지지 않는다.

[0036] 본 명세서에서 용어 "크리스퍼 RNA" (crRNA)는 하나 이상의 Cas 단백질 (예를 들어, Cas9)과 복합체를 형성할 수 있어, 복합체에 DNA 결합 특이성을 제공하는 RNA 서열을 지칭한다. crRNA는 DNA 표적 서열의 가닥에 상보적인 "가이드 서열" ("가변 표적화 도메인" [VT])을 함유하기 때문에 DNA 결합 특이성을 제공하다. crRNA는 crRNA가 유래된 크리스퍼 유전자좌의 반복 영역에 의해 암호화되는 "반복 서열" ("tracr RNA 메이트 서열")을 추가로 포함한다. crRNA의 반복 서열은 tracrRNA의 5'-말단에서 서열에 어닐링할 수 있다. 고유 크리스퍼 시스템의 crRNA는 크리스퍼 유전자좌로부터 전사된 "예비-crRNA"로부터 유래된다. 예비-crRNA는 스페이서 영역과 반복 영역을 포함하고; 스페이서 영역은 DNA 표적 부위는 서열에 상보적인 특유의 서열을 함유한다. 고유 시스템의 예비-crRNA는 다수의 상이한 crRNA로 처리되며, 각각은 반복 서열의 일부와 함께 가이드 서열을 갖는다. 크리스퍼 시스템은, 예를 들어, DNA 표적화 특이성을 위해 crRNA를 이용한다.

[0037] 본 명세서에서 용어 "트랜스-활성화 크리스퍼 RNA" (tracrRNA)는 타입 II 크리스퍼 시스템에서 사용되며, 5'에서 3' 방향으로 (i) 크리스퍼 타입 II crRNA의 반복 영역과 어닐링하는 서열 및 (ii) 스템 루프-함유하는 부분을 함유하는 비암호화 RNA를 지칭한다 (문헌[Deltcheva et al., *Nature* 471:602-607]).

[0038] 용어 "가이드 RNA" (gRNA) 및 "단일 가이드 RNA" (sgRNA)는 본 명세서에서 상호교환가능하게 사용된다. 본 명세서에서 gRNA는 tracrRNA에 작동가능하게 연결된 crRNA를 함유하는 키메라 서열을 지칭할 수 있다. 대안적으로, gRNA는, 예를 들어 crRNA와 tracrRNA의 합성 융합을 지칭할 수 있다. 문헌[Jinek et aL. (*Science* 337:816-821)]은 일부 gRNA 특징을 개시한다. gRNA는 또한 가이드 서열 (가변 표적화 도메인) 다음에 Cas 엔도뉴클레아제 인식 (CER) 도메인을 갖는 것을 특징으로 할 수 있다 [모두 본 명세서에 전체적으로 참고로 포함되는 2015년 2월 26일자로 출원된 WO2015026883, 2015년 3월 19일자로 출원된 미국 특허 출원 제US 2015-0082478 A1호 및 2015년 2월 26일자로 출원된 제US 2015-0059010 A1호]. CER 도메인은 tracrRNA 메이트 서열 다음에 tracrRNA 서열을 포함한다.

[0039] 본 명세서에서 용어 "표적 부위 서열", "표적 부위", "표적 서열", "표적 DNA", "DNA 표적 서열", "표적 유전자좌", "게놈 표적 부위", "게놈 표적 서열", "게놈 표적 유전자좌" 및 "프로토스페이서"는 상호교환가능하게 사용된다. 표적 부위 서열은 본 명세서의 RGEN가 인식하고, 결합하며, 임의로 닉을 형성하거나 절단하는 통상적이지 않은 효모의 게놈 내의 염색체, 에피솜 또는 임의의 다른 DNA 분자 상의 폴리뉴클레오티드 서열을 지칭한다. 표적 부위는 (i) 효모 내의 내인성/고유 부위이거나, (ii) 효모에 이종성이어서 게놈에서 자연적으로 발생하지 않거나 (iii) 이가 고유하게 발생하는 위치에 비교하여 이종성인 게놈 위치에서 발견되는 것일 수 있다.

[0040] 본 명세서에서 표적 부위 서열은 길이가 적어도 13 뉴클레오티드이고, 가이드 서열과 혼성화되고, (소정 실시 형태에서, 적합한 PAM이 표적 서열에 인접한 경우) 표적 서열에 Cas 단백질 또는 Cas 단백질 복합체를 직접 서열-특이적으로 결합시킬 수 있는 (crRNA 또는 gRNA의) 가이드 서열에 대한 충분한 상보성을 갖는 가닥을 갖는다. (엔도뉴클레오리틱(endonucleolytic) 또는 닉킹 Cas와 적용가능한) 절단/닉 부위는 (예를 들어, Cas9를 사용) 표적 서열 내에 있거나, 절단/닉 부위는 (예를 들어, FokI 효소로부터 유래된 것과 같은 이종 엔도뉴클레아제 도메인에 융합된 Cas9를 사용) 표적 서열 외부에 있을 수 있다.

[0041] 본 명세서에서 "인공 표적 부위" 또는 "인공 표적 서열"은 통상적이지 않은 효모의 게놈에 도입된 표적 서열을 지칭한다. 일부 실시 형태에서, 인공 표적 서열은 효모의 게놈 내의 고유 표적 서열과 서열이 동일할 수 있지만, 게놈의 상이한 위치 (이종 위치)에 위치할 수 있거나, 효모의 게놈 내의 동일한 위치에 위치하는 경우 고유 표적 서열과 상이할 수 있다.

[0042] 본 명세서에서 "에피솜"은 효모 세포의 염색체와 별도로 효모 세포에 자율적으로 존재할 수 있는 (복제할 수 있고 딸 세포로 전달할 수 있는) DNA 분자를 지칭한다. 에피솜 DNA는 효모세포에 고유하거나 이종성일 수 있다. 본 명세서에서 고유 에피솜의 예에는 미토콘드리아 DNA (mtDNA)가 포함된다. 본 명세서에서 이종 에피솜의 예에는 플라스미드 및 효모 인공 염색체 (YAC)가 포함된다.

[0043] 본 명세서에서 "프로토스페이서 인접 모티프" (PAM)는 본 명세서의 RGEN에 의해 인식되는 짧은 서열을 지칭한다. 본 명세서의 PAM의 서열과 길이는 사용되는 Cas 단백질 또는 Cas 단백질 복합체에 따라 다를 수 있지만, 전형적으로는, 예를 들어 2, 3, 4, 5, 6, 7 또는 8개의 뉴클레오티드 길이이다.

[0044] 용어 "5'-캡" 및 "7-메틸구아닐레이트 (m^7G) 캡"은 본 명세서에서 상호교환가능하게 사용된다. 7-메틸구아닐레이트 잔기는 진핵생물에서 메신저 RNA (mRNA)의 5′ 말단 상에 위치한다. RNA 폴리머라제 II (Pol II)는 진핵생물에서 mRNA를 전사한다. 메신저 RNA 캡핑은 일반적으로 다음과 같이 발생한다: mRNA 전사물의 최말단 5' 포

스페이트기는 RNA 말단 포스파타제에 의해 제거되어, 2개의 말단 포스페이트를 남긴다. 구아노신 모노포스페이트 (GMP)는 구아닐릴 트랜스퍼라제에 의해 전사물의 말단 포스페이트에 첨가되어, 전사물 말단에 5′-5′ 트라이포스페이트-연결된 구아닌을 남긴다. 마지막으로, 이러한 말단 구아닌의 7-질소가 메틸 트랜스퍼라제에 의해 메틸화된다.

[0045] 본 명세서에서 용어 "5'-캡을 갖지 않는"은, 예를 들어, 5'-캡 대신에 5'-하이드록실기를 갖는 RNA를 지칭하는 데 사용된다. 이러한 RNA는, 예를 들어 "캡핑되지 않은 RNA"로 지칭될 수 있다. 캡핑되지 않은 RNA는 5'-캡핑된 RNA가 핵외수송(nuclear export)의 대상이기 때문에 전사 후 핵에 더 잘 축적될 수 있다. 본 명세서에서 하나 이상의 RNA 성분이 캡핑되지 않는다

[0046] 용어 "리보자임" 및 "리보핵산 효소"는 본 명세서에서 상호교환가능하게 사용된다. 리보자임은 특정 부위에서 RNA를 절단할 수 있는 2차, 3차, 및/또는 4차 구조(들)를 형성하는 하나 이상의 RNA 서열을 지칭한다. 리보자임은 리보자임 서열에 대하여 시스-부위에서 RNA를 절단할 수 있는 "자가 절단(self-cleaving) 리보자임"을 포함한다 (즉, 자가 촉매적(auto-catalytic) 또는 자가 절단). 리보자임 핵산 분해(nucleolytic) 활성의 일반적인 성질이 기재되어 있다 (예를 들어, 문헌[Lilley, *BiocheM. SoC. TranS.* 39:641-646]). 본 명세서에서 "망치머리 리보자임" (HHR)은 촉매 작용에 관여하는 3개의 염기쌍 형성된 스템 및 고도로 보존된, 비상보적 뉴클레오티드의 코어로 구성되는 작은 촉매 RNA 모티프를 포함할 수 있다. 본 명세서에 참고로 포함되는 문헌[Pley et aL. (*Nature* 372:68-74) and Hammann et aL. (*RNA* 18:871-885)]은 망치머리 리보자임 구조와 활성을 개시하고 있다. 본 명세서에서 망치머리 리보자임은, 예를 들어, 본 명세서에 참고로 포함되는 문헌[by Scott et aL. (*Cell* 81:991-1002])에 개시된 바와 같은 "최소 망치머리" 서열을 포함할 수 있다.

[0047] 본 발명의 일 실시 형태에서, 방법은 통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 표적 부위 서열로 RNA-유도 엔도뉴클레아제 (RGEN)를 표적화하는 방법을 포함하며, 상기 방법은 상기 효모에 Cas 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 제1 재조합 DNA 구조물 및 RNA 성분의 상류에 리보자임을 암호화하는 DNA 서열을 포함하는 적어도 하나의 제2 재조합 DNA 구조물을 제공하는 단계를 포함하고, 여기서 제2 재조합 DNA 구조물로부터 전사된 RNA는 자체 촉매 작용으로 리보자임을 제거하여 상기 RNA 성분을 생성하며, RNA 성분과 Cas9 엔도뉴클레아제는 표적 부위 서열의 전부 또는 일부에 결합할 수 있는 RGEN을 형성할 수 있다.

[0048] 본 발명의 일 실시 형태에서, 통상적이지 않은 효모는 적어도 하나의 뉴클레오티드 서열에 작동가능하게 연결된 프로모터를 포함하는 폴리뉴클레오티드 서열을 포함하며, 여기서 상기 뉴클레오티드 서열은 RNA 성분을 암호화하는 DNA 서열의 상류에 리보자임을 암호화하는 DNA 서열을 포함하고, 상기 RNA 성분은 효모 내의 염색체 또는 에피솜 상의 표적 부위 서열에 상보적인 가변 표적화 도메인을 포함하며, RNA 성분은 RNA-유도 엔도뉴클레아제 (RGEN)를 형성할 수 있고, 상기 RGEN는 표적 부위 서열의 전부 또는 일부에 결합할 수 있으며, 뉴클레오티드 서열로부터 전사된 RNA는 자체 촉매 작용으로 리보자임을 제거하여 상기 RNA 성분을 생성하고, 상기 RNA 성분은 5' 캡을 포함하지 않는다.

[0049] 리보자임은 또한 임의의 선행 전사물을 제거하지만 리보자임 서열은 온전히 남겨두는 자체 서열의 5'을 절단하는 리보자임을 포함한다.

[0050] 본 발명의 일 실시 형태에서, 통상적이지 않은 효모는 적어도 하나의 뉴클레오티드 서열에 작동가능하게 연결된 프로모터를 포함하는 폴리뉴클레오티드 서열을 포함하며, 여기서 상기 뉴클레오티드 서열은 RNA 성분을 암호화하는 DNA 서열의 상류에 리보자임을 암호화하는 DNA 서열을 포함하고, 상기 RNA 성분은 효모 내의 염색체 또는 에피솜 상의 표적 부위 서열에 상보적인 가변 표적화 도메인을 포함하며, RNA 성분은 RNA-유도 엔도뉴클레아제 (RGEN)를 형성할 수 있고, 상기 RGEN는 표적 부위 서열의 전부 또는 일부에 결합할 수 있으며, 뉴클레오티드 서열로부터 전사된 RNA는 자체 촉매 작용으로 리보자임을 제거하여 상기 RNA 성분을 생성하고, 뉴클레오티드 서열로부터 전사된 RNA는 자체 촉매 작용으로 리보자임을 제거하지 않아, 5' 캡이 없이 리보자임-RNA 성분 융합 분자를 생성한다.

[0051] 용어 "표적화", "유전자 표적화", "DNA 표적화", "편집", "유전자 편집" 및 "DNA 편집"은 본 명세서에서 상호교환가능하게 사용된다. 본 명세서에서 DNA 표적화는 통상적이지 않은 효모의 염색체 또는 에피솜에서와 같은 특정 DNA 서열에서의 삽입-결실, 녹아웃 또는 녹인의 특이적 도입일 수 있다. 일반적으로, DNA 표적화는 본 명세서에서 적합한 RNA 성분과 결합한 Cas 단백질을 사용하여 통상적이지 않은 효모의 특정 DNA 서열에서 하나 또는 두 가닥을 절단함으로써 수행될 수 있다. 이러한 DNA 절단은, 이중 가닥 절단 (DSB)의 경우 표적 부위에서 삽입-결실을 형성할 수 있는 NHEJ 공정을 유도할 수 있다. 또한, 절단이 단일-가닥 절단 (SSB) 또는 DSB인지 여부에 관계없이, DNA 닉 또는 절단 부위에 적합한 공여 DNA 폴리뉴클레오티드가 제공되는 경우 HR 공정은 유도될

수 있다. 이러한 HR 공정은 공여 DNA 폴리뉴클레오티드의 서열에 따라 표적 부위에서 녹아웃 또는 녹인을 도입하는 데 사용될 수 있다.

[0052] 대안적으로, 본 명세서에서 DNA 표적화는 본 명세서의 Cas/RNA 성분 복합체와 표적 DNA 서열의 특정 결합을 지칭할 수 있으며, 여기서 Cas 단백질은 DNA 가닥을 절단하거나 절단하지 않는다 (Cas 단백질의 엔도뉴클레오리틱 도메인의 상태에 따라).

[0053] 본 명세서에서 용어 "삽입-결실"은 염색체 또는 에피솜 내의 표적 DNA 서열에서의 뉴클레오티드 염기의 삽입 또는 결실을 지칭한다. 이러한 삽입 또는 결실은, 예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 그 이상의 염기일 수 있다. 소정 실시 형태에서, 삽입-결실은 심지어 더 클 수 있고, 적어도 약 20, 30, 40, 50, 60, 70, 80, 90 또는 100개의 염기일 수 있다. 삽입-결실이 유전자의 오픈 리딩 프레임 (ORF) 내에 도입되는 경우, 종종 삽입-결실은 프레임시프트 돌연변이를 생성함으로써 ORF에 의해 암호화되는 단백질의 야생형 발현을 방해한다.

[0054] 용어 "녹아웃", "유전자 녹아웃" 및 "유전적 녹아웃"은 본 명세서에서 상호교환가능하게 사용된다. 본 명세서에서 녹아웃은 Cas 단백질로 표적화함으로써 부분적으로 또는 완전히 작동하지 않게 된 통상적이지 않은 효모의 DNA 서열을 나타내며; 녹아웃 이전의 이러한 DNA 서열은, 예를 들어, 아미노산 서열을 암호화할 수 있거나 조절 기능 (예를 들어, 프로모터)을 가질 수 있다. 녹아웃은 (NHEJ에 의한) 삽입-결실에 의해 또는, 표적화 부위 또는 그 부근에서 서열의 기능을 감소시키거나 완전히 파괴시키는 서열의 특이적 제거에 의해 생성될 수 있다. 본 명세서에서 녹아웃된 DNA 폴리뉴클레오티드 서열은 대안적으로, 예를 들어, 부분적으로 또는 완전히 파괴되거나 하향조절되는 것을 특징으로 할 수 있다.

[0055] 일 실시 형태에서, 본 발명은 Cas9 엔도뉴클레아제 및 적어도 하나의 뉴클레오티드 서열에 작동가능하게 연결된 프로모터를 포함하는 폴리뉴클레오티드 서열을 포함하는 통상적이지 않은 효모에 관한 것이며, 여기서 상기 뉴클레오티드 서열은 RNA 성분을 암호화하는 DNA 서열의 상류에 리보자임을 암호화하는 DNA 서열을 포함하고, 상기 RNA 성분은 효모 내의 염색체 또는 에피솜 상의 표적 부위 서열에 상보적인 가변 표적화 도메인을 포함하며, RNA 성분은 Cas 엔도뉴클레아제와 함께, 표적 부위 서열에 결합할 수 있는 RNA-유도 엔도뉴클레아제 (RGEN)를 형성할 수 있다. Cas9 엔도뉴클레아제는 효모에서 단백질로서 도입될 수 있거나 재조합 DNA 구조물을 통해 도입될 수 있다. Cas9 엔도뉴클레아제는 본 기술 분야에 알려진 임의의 방법에 의해 안정적이거나 일과성 방식으로 발현될 수 있다.

[0056] 용어 "녹인", "유전자 녹인" 및 "유전적 녹인"은 본 명세서에서 상호교환가능하게 사용된다. 녹인은 Cas 단백질을 사용하여 표적화함으로써 통상적이지 않은 효모의 특정 DNA 서열에서 DNA 서열을 대체하거나 삽입하는 것을 나타낸다. 녹인의 예는 유전자의 암호화 영역에 이종 아미노산 암호화 서열의 특이적 삽입 또는 유전자좌에 전사 조절 요소의 특이적 삽입이다.

[0057] 용어 "공여 폴리뉴클레오티드", "공여 DNA", "표적화 폴리뉴클레오티드" 및 "표적화 DNA"는 본 명세서에서 상호교환가능하게 사용된다. 공여 폴리뉴클레오티드는 DNA 표적 부위 (예를 들어, 본 명세서의 Cas 단백질에 의해 특이적으로 표적화된 서열)에 또는 그 부근의 서열에 상동성인 적어도 하나의 서열을 포함하는 DNA 서열을 지칭한다. 편집될 뉴클레오티드 서열과 비교하는 경우, 적어도 하나의 뉴클레오티드 변형을 포함하는 폴리뉴클레오티드를 포함하는 공여 DNA 폴리뉴클레오티드는 "폴리뉴클레오티드 변형 주형", "폴리뉴클레오티드 변형 주형 DNA" 또는 "주형 DNA"로도 지칭된다. 뉴클레오티드 변형은 적어도 하나의 뉴클레오티드 치환, 첨가 또는 결실일 수 있다. 임의로, 폴리뉴클레오티드 변형 주형은 적어도 하나의 뉴클레오티드 변형에 플랭킹된 상동성 뉴클레오티드 서열을 추가로 포함할 수 있으며, 여기서 플랭킹 상동성 뉴클레오티드 서열은 편집될 원하는 뉴클레오티드 서열에 대한 충분한 상동성을 제공한다.

[0058] 본 명세서에서 공여 폴리뉴클레오티드 내의 "상동성 서열"은 표적 부위 또는 그 부근의 서열과 100% 동일성, 또는 표적 부위 또는 그 부근의 서열과 적어도 약 95%, 96%, 97%, 98% 또는 99% 동일성을 갖는 적어도 약 25개의 뉴클레오티드의 서열을 포함하거나 이로 구성될 수 있다.

[0059] 소정 실시 형태에서, 공여 DNA 폴리뉴클레오티드는 표적 부위에서 서열과 이종성인 서열에 의해 분리된 2개의 상동성 서열을 가질 수 있다. 이러한 공여 폴리뉴클레오티드의 이들 2개의 상동성 서열은 이종 서열에 플랭킹된 "상동성 암"으로 지칭될 수 있다. 2개의 상동성 암을 갖는 표적 부위와 공여 폴리뉴클레오티드 사이의 HR은 전형적으로 표적 부위에서 서열을 공여 폴리뉴클레오티드의 이종 서열로 대체시킨다 (공여 폴리뉴클레오티드의 상동성 암에 상동성인 DNA 서열 사이에 위치한 표적 부위 서열은 공여 폴리뉴클레오티드의 이종 서열에 의해 대

체된다). 2개의 상동성 암을 갖는 공여 폴리뉴클레오티드에서, 암은 1개 이상의 뉴클레오티드에 의해 분리될 수 있다 (즉, 공여 폴리뉴클레오티드의 이종 서열은 길이가 적어도 1개의 뉴클레오티드일 수 있다). 본 명세서에서 통상적이지 않은 효모에서 수행될 수 있는 다양한 HR 절차가, 예를 들어, 본 명세서에 참고로 포함되는 문헌[DNA Recombination: Methods and Protocols: 1st Edition (H. Tsubouchi, Ed., Springer-Verlag, New York, 2011)]에 개시되어 있다.

[0060] 일 실시 형태에서, 공여 DNA 구조물은 Cas 엔도뉴클레아제의 표적 부위로 삽입될 관심 폴리뉴클레오티드를 포함하며, 여기서 공여 DNA 구조물은 관심 폴리뉴클레오티드에 플랭킹된 제1 및 제2 상동성 영역을 추가로 포함한다. 공여 DNA의 제1 및 제2 상동성 영역은 각각 식물 게놈의 표적 부위에 존재하거나 플랭킹된 제1 및 제2 게놈 영역과 상동성을 공유한다.

[0061] 용어 "부피 %", "부피 백분율", "vol %" 및 "v/v %"는 본 명세서에서 상호교환가능하게 사용된다. 용액 중의 용질의 부피%는 하기 식을 사용하여 결정될 수 있다: [(용질의 부피)/(용액의 부피)] × 100%.

[0062] 용어 "중량%", "중량 백분율 (wt%)" 및 "중량-중량 백분율(% w/w)"은 본 명세서에서 상호교환가능하게 사용된다. 중량%는 재료가 조성물, 혼합물 또는 용액 중에 포함될 때 질량을 기준으로 한 재료의 백분율을 지칭한다.

[0063] 용어 "폴리뉴클레오티드", "폴리뉴클레오티드 서열" 및 "핵산 서열"은 본 명세서에서 상호교환가능하게 사용된다. 이들 용어는 뉴클레오티드 서열 등을 포함한다. 폴리뉴클레오티드는 합성된, 비천연 뉴클레오티드 염기 또는 변경된 뉴클레오티드 염기를 임의로 함유하는 단일- 또는 이중-가닥인 DNA 또는 RNA 의 중합체일 수 있다. 폴리뉴클레오티드는 cDNA, 게놈 DNA, 합성 DNA, 또는 이의 혼합물의 하나 이상의 세그먼트로 이루어질 수 있다. 뉴클레오티드 (리보뉴클레오티드 또는 데옥시리보뉴클레오티드)는 다음과 같이 단일 문자 표기로 지칭될 수 있다: (각각 RNA 또는 DNA에 있어서) 아데닐레이트 또는 데옥시아데닐레이트는 "A", (각각 RNA 또는 DNA에 있어서) 시티딜레이트 또는 데옥시시티딜레이트는 "C", (각각 RNA 또는 DNA에 있어서) 구아닐레이트 또는 데옥시구아닐레이트는 "G", (RNA에 있어서) 유리딜레이트는 "U", (DNA에 있어서) 데옥시티미딜레이트는 "T", 퓨린 (A 또는 G)은 "R", 피리미딘 (C 또는 T)은 "Y", G 또는 T는 "K", A 또는 C 또는 T는 "H", 이노신은 "I", A 또는 T는 "W" 및 임의의 뉴클레오티드는 "N" (예를 들어, DNA 서열을 말하는 경우 N은 A, C, T 또는 G일 수 있고; RNA 서열을 말하는 경우 N은 A, C, U 또는 G일 수 있다). 본 명세서에 개시된 임의의 RNA 서열 (예를 들어, crRNA, tracrRNA, gRNA)은 적합한 DNA 서열에 의해 암호화될 수 있다.

[0064] 본 명세서에서 사용되는 바와 같은 용어 "단리된"은 이의 천연 공급원으로부터 완전히 또는 부분적으로 정제된 폴리뉴클레오티드 또는 폴리펩티드 분자를 지칭한다. 일부 경우에, 단리된 폴리뉴클레오티드 또는 폴리펩티드 분자는 더 큰 조성물, 완충 시스템 또는 시약 믹스의 일부이다. 예를 들어, 단리된 폴리뉴클레오티드 또는 폴리펩티드 분자는 이종성 방식으로 세포 또는 유기체 내에 포함될 수 있다.

[0065] 본 명세서에서 사용되는 바와 같은 용어 "유전자"는 암호화 영역으로부터 RNA (RNA는 DNA 폴리뉴클레오티드 서열로부터 전사됨)를 발현하는 DNA 폴리뉴클레오티드 서열을 지칭하며, RNA는 (단백질을 암호화하는) 메신저 RNA 또는 비-단백질-암호화 RNA (예를 들어, 본 명세서에서 crRNA, tracrRNA 또는 gRNA)일 수 있다. 유전자는 암호화 영역만을 지칭할 수 있거나, 암호화 영역 (예를 들어, 프로모터, 5'-비번역 영역, 3'-전사 종결 영역)의 상류 및/또는 하류의 조절 서열을 포함할 수 있다. 단백질을 암호화하는 암호화 영역은 본 명세서에서 대안적으로 "오픈 리딩 프레임" [ORF]으로 지칭될 수 있다. "고유" 또는 "내인성"인 유전자는 자연에서 그 자체의 조절 서열을 갖는 것으로 발견되는 유전자를 지칭하며; 이러한 유전자는 숙주 세포의 게놈에서 이의 자연적 위치에 위치한다. "키메라" 유전자는 자연에서 함께 발견되지 않는 조절 및 암호화 서열을 포함하는, 고유 유전자가 아닌 임의의 유전자를 말한다 (즉, 조절 영역과 암호화 영역이 서로 이종성이다). 따라서, 키메라 유전자는 상이한 공급원으로부터 유래된 조절 서열과 암호화 서열, 또는 동일한 공급원으로부터 유래되었지만, 자연에서 발견되는 것과 상이한 방식으로 배열된 조절 서열과 암호화 서열을 포함할 수 있다. "외래" 또는 "이종" 유전자는 유전자 전달에 의해 숙주 유기체 내로 도입된 유전자를 지칭한다. 외래 유전자는 비-천연 유기체에 삽입된 천연 유전자, 천연 숙주 내의 새로운 위치로 도입된 천연 유전자, 또는 키메라 유전자를 포함할 수 있다. 본 명세서에 개시된 소정 실시 형태에서, 폴리뉴클레오티드 서열은 이종성이다. "트랜스유전자(transgene)"는 형질전환 절차에 의해 게놈 내로 도입된 유전자이다. "코돈-최적화된" 오픈 리딩 프레임은 이의 코돈 사용 빈도가 숙주 세포의 바람직한 코돈 사용 빈도를 모방하도록 설계된 유전자이다.

[0066] 고유 아미노산 서열 또는 폴리뉴클레오티드 서열은 자연 발생적인 반면, 비천연 아미노산 서열 또는 폴리뉴클레오티드 서열은 자연에서 발생하지 않는다.

[0067] 본 명세서에서 사용되는 바와 같은 "조절 서열"은 유전자의 전사 시작 부위 (예를 들어, 프로모터), 5' 비번역 영역 및 3' 비암호화 영역의 상류에 위치하며, 전사, 프로세싱 또는 안정성 또는 유전자로부터 전사된 RNA의 번역에 영향을 줄 수 있는 뉴클레오티드 서열을 지칭한다. 본 명세서에서 조절 서열은 프로모터, 인핸서, 사일런서(silencer), 5' 비번역 선도 서열, 인트론, 폴리아데닐화 인식 서열, RNA 프로세싱 부위, 이펙터 결합 부위, 스템-루프 구조 및 유전자 발현의 조절에 관여하는 다른 요소를 포함할 수 있다. 본 명세서에서 하나 이상의 조절 요소는 본 명세서의 암호화 영역에 이종성일 수 있다.

[0068] 본 명세서에서 사용되는 바와 같은 "프로모터"는 유전자로부터의 RNA의 전사를 조절할 수 있는 DNA 서열을 지칭한다. 일반적으로, 프로모터 서열은 유전자의 전사 시작 부위의 상류에 있다. 프로모터는 고유 유전자로부터 그 전체가 유래될 수도 있거나, 자연에서 발견되는 상이한 프로모터로부터 유래된 상이한 요소로 구성될 수도 있거나, 심지어 합성 DNA 세그먼트를 포함할 수도 있다. 대부분의 세포 유형에서 대부분의 경우에 유전자가 발현되도록 하는 프로모터는 통상적으로 "항시성 프로모터(constitutive promoter)"라고 지칭된다. 본 명세서에서 하나 이상의 프로모터는 본 명세서의 암호화 영역에 이종성일 수 있다.

[0069] 본 명세서에서 사용되는 바와 같은 "강력한 프로모터"는 단위 시간당 상대적으로 많은 수의 생산 개시를 이끌 수 있는 프로모터를 지칭하고/하며 효모 내 유전자의 평균 전사 수준보다 더 높은 수준의 유전자 전사를 유도하는 프로모터이다.

[0070] 본 명세서에서 사용되는 바와 같은 용어 "3' 비암호화 서열", "전사 종결자" 및 "종결자"는 암호화 서열의 하류에 위치한 DNA 서열을 지칭한다. 이는 폴리아데닐화 인식 서열 및 mRNA 프로세싱 또는 유전자 발현에 영향을 줄 수 있는 조절 신호를 암호화하는 다른 서열을 포함한다.

[0071] 본 명세서에서 사용되는 바와 같은 용어 "카세트"는 단백질 또는 비-단백질-암호화 RNA를 암호화하는 서열에 작동가능하게 연결된 프로모터를 지칭한다. 카세트는 임의로 3' 비암호화 서열에 작동가능하게 연결될 수 있다.

[0072] 폴리뉴클레오티드와 관련하여 본 명세서에서 사용되는 바와 같은 용어 "상류" 및 "하류"는 각각 "이의 5'" 및 "이의 3'"을 지칭한다.

[0073] 본 명세서에서 사용되는 바와 같은 용어 "발현"은 (i) 암호화 영역으로부터 RNA의 전사 (예를 들어, mRNA, 또는 crRNA, tracrRNA 또는 gRNA와 같은 비-단백질 암호화 RNA) 또는 (ii) mRNA로부터의 폴리펩티드의 번역을 지칭한다.

[0074] 유전자 또는 폴리뉴클레오티드 서열의 발현을 설명하는 데 사용되는 경우, 용어 "하향-조절", "파괴", "억제", "불활성화" 및 "침묵화"는 폴리뉴클레오티드 서열의 전사가 감소되거나 제거되는 경우를 지칭하도록 본 명세서에서 상호교환가능하게 사용된다. 이는 폴리뉴클레오티드 서열로부터의 RNA 전사물의 감소 또는 제거를 초래하며, 이는 (유전자가 ORF를 포함하는 경우) 폴리뉴클레오티드 서열로부터 유래된 단백질 발현의 감소 또는 제거를 초래한다. 대안적으로, 하향-조절은 폴리뉴클레오티드 서열에 의해 생성된 전사물로부터의 단백질 번역이 감소되거나 제거된 경우를 지칭할 수 있다. 또한 대안적으로, 하향-조절은 폴리뉴클레오티드 서열에 의해 발현된 단백질의 활성이 감소된 경우를 지칭할 수 있다. 세포에서 상기 공정 (전사, 번역, 단백질 활성) 중 어느 하나의 감소는 적합한 대조 세포의 전사, 번역 또는 단백질 활성에 대하여 약 40%, 50%, 60%, 70%, 80%, 90%, 95% 또는 100%일 수 있다. 하향-조절은, 예를 들어, 본 명세서에 개시된 바와 같은 표적화 이벤트 (예를 들어, 삽입-결실, 녹아웃)의 결과일 수 있다.

[0075] 용어 "대조 세포" 및 "적합한 대조 세포"는 본 명세서에서 상호교환가능하게 사용되며, 특정 변형 (예를 들어, 폴리뉴클레오티드의 과발현, 폴리뉴클레오티드의 하향-조절)이 만들어진 세포 (즉, "실험 세포")와 관련하여 언급될 수 있다. 대조 세포는 실험 세포의 특정 변형을 가지지 않거나 발현하지 않는 임의의 세포일 수 있다. 따라서, 대조 세포는 형질전환되지 않은 야생형 세포일 수 있거나, 유전적으로 변형될 수 있지만 유전적 형질전환을 발현하지 않는다. 예를 들어, 대조 세포는 실험 세포의 직접적인 부모일 수 있으며, 직접적인 부모 세포는 실험 세포에 있는 특정 변형을 가지지 않는다. 대안적으로, 대조 세포는 하나 이상의 세대에 의해 제거되는 실험 세포의 부모일 수 있다. 또한 대안적으로, 대조 세포는 실험 세포의 형제일 수 있으며, 형제 세포는 실험 세포에 존재하는 특정 변형을 포함하지 않는다.

[0076] 본 명세서에서 사용되는 바와 같은 용어 "증가된"은 증가된 양 또는 활성이 비교되는 양 또는 활성보다 적어도 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19% 또는 20% 초과인 양 또는 활성을 지칭할 수 있다. 용어 "증가된", "보다 많은" 및 "개선된"은 본 명세서에서 상호교환가능하게 사용된다. 용어 "증가된"은 단백질을 암호화하는 폴리뉴클레오티드의 발현을 특성화하는 데 사용될 수 있는데,

예를 들어, "증가된 발현"은 또한 "과발현"을 의미한다.

[0077] 본 명세서에서 사용되는 바와 같은 용어 "작동가능하게 연결된"은 하나의 기능이 다른 것에 의해 영향을 받는 둘 이상의 핵산 서열의 결합을 지칭한다. 예를 들어, 프로모터는 그것이 암호화 서열의 발현에 영향을 미칠 수 있는 경우에 암호화 서열에 작동가능하게 연결된다. 다시 말하면, 암호화 서열은 프로모터의 전사적 조절 하에 있다. 암호화 서열은, 예를 들어, 조절 서열에 작동가능하게 연결될 수 있다. 또한, 예를 들어, crRNA는 crRNA의 tracrRNA 메이트 서열이 tracrRNA의 5' 서열과 어닐링하도록 본 명세서의 tracrRNA에 작동가능하게 연결 (융합)될 수 있다. 이러한 작동가능한 연결은 적합한 루프 형성 서열, 예컨대 GAAA (서열 번호:43), CAAA (서열 번호:44) 또는 AAAG (서열 번호:45)를 포함할 수 있다.

[0078] 본 명세서에서 사용되는 바와 같은 용어 "재조합"은, 예를 들어, 화학적 합성에 의한 또는 유전자 조작 기술에 의해 분리된 핵산 세그먼트의 조작에 의한, 그렇지 않다면 분리되어 있는 2개의 서열 세그먼트의 인공적인 조합을 지칭한다. 용어 "재조합", "트랜스제닉(transgenic)", "형질전환된", "유전자 조작된" 또는 "외인성 유전자 발현을 위해 변형된"은 본 명세서에서 상호교환가능하게 사용된다.

[0079] 본 명세서의 재조합 구조물/벡터 (예를 들어, 본 명세서의 리보자임-RNA 성분 카세트를 암호화하는 DNA 폴리뉴클레오티드 또는 본 명세서의 Cas 단백질을 암호화하는 DNA 폴리뉴클레오티드)를 제조하는 방법은 문헌[by J. Sambrook and D. Russell (Molecular Cloning: A Laboratory Manual, 3rd Edition, Cold Spring Harbor Laboratory PRess, Cold Spring Harbor, NY, 2001)]; 문헌[T.J. Silhavy et al. (Experiments with Gene Fusions, Cold Spring Harbor Laboratory PRess, Cold Spring Harbor, NY, 1984)]; 및 문헌[F.M. Ausubel et al. (Short Protocols in Molecular Biology, 5th Ed. Current Protocols, John Wiley and Sons, Inc., NY, 2002)]에 기재된 바와 같은 표준 재조합 DNA 및 분자 클로닝 기술을 따를 수 있다.

[0080] 본 명세서에서 사용되는 바와 같은 용어 "형질전환"은 숙주 유기체 또는 숙주 세포로의 핵산 분자의 전달을 지칭한다. 예를 들어, 핵산 분자는 세포에서 자율적으로 복제하거나, 숙주 유기체/세포의 게놈에 통합되거나, 또는 복제되거나 통합되지 않고 세포에 일시적으로 존재하는 것일 수 있다. 플라스미드 및 선형 DNA 분자와 같은 형질전환에 적합한 핵산 분자의 비제한적인 예가 본 명세서에 개시되어 있다. 형질전환된 핵산 단편을 함유하는 숙주 유기체/세포 (예를 들어, 본 명세서의 통상적이지 않은 효모)는 "트랜스제닉", "재조합", "형질전환된" 또는 "형질전환체"로 지칭될 수 있다.

[0081] 폴리뉴클레오티드 또는 폴리펩티드 서열과 관련하여 본 명세서에서 사용되는 바와 같은 용어 "서열 동일성" 또는 "동일성"은 특정 비교창 상에서 최대 상응도로 정렬되는 경우에 동일한 두 서열 내의 핵산 잔기 또는 아미노산 잔기를 지칭한다. 따라서, "서열 동일성 %" 또는 "동일성 %"은 비교창 상에서 최적으로 정렬된 2개의 서열을 비교함으로써 결정된 값을 지칭하며, 여기서 비교창 내의 폴리뉴클레오티드 또는 폴리펩티드 서열의 부분은 2개의 서열의 최적 정렬을 위한 기준 서열 (삽입 또는 결실을 포함하지 않음)과 비교하여 삽입 또는 결실 (즉, 갭)을 포함할 수 있다. 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 백분율을 계산한다. DNA 서열과 RNA 서열 사이의 서열 동일성을 계산하는 경우, DNA 서열의 T 잔기가 RNA 서열의 U 잔기와 정렬되며, 이와 "동일"하게 간주될 수 있음을 이해할 것이다. 제1 폴리뉴클레오티드와 제2 폴리뉴클레오티드의 % 상보성을 결정하기 위해, 예를 들어 (i) 제1 폴리뉴클레오티드와 제2 폴리뉴클레오티드의 보체 서열 사이의 % 동일성 (또는 그 반대) 및/또는 (ii) 표준 (canonical) 왓슨과 크릭 염기쌍을 생성할 제1 폴리뉴클레오티드와 제2 폴리뉴클레오티드 사이의 염기 %를 결정함으로써 이를 얻을 수 있다.

[0082] 예를 들어, 국립 생물공학 정보 센터 (National Center for Biotechnology Information; NCBI) 웹사이트에서 온라인으로 이용가능한 기본 국소 정렬 검색 도구 (Basic Local Alignment Search Tool; BLAST) 알고리즘을 사용하여, 본 명세서에 개시된 둘 이상의 폴리뉴클레오티드 서열 (BLASTN 알고리즘) 또는 폴리펩티드 서열 (BLASTP 알고리즘) 사이의 % 동일성을 측정할 수 있다. 대안적으로, 서열 사이의 % 동일성은 클러스탈 (Clustal) 알고리즘 (예를 들어, 클러스탈W 또는 클러스탈V)을 사용하여 수행될 수 있다. 클러스탈 정렬 방법을 사용하는 다중 정렬의 경우, 디폴트(default) 값은 갭 페널티(GAP PENALTY)=10 및 갭 길이 페널티(GAP LENGTH PENALTY)=10에 상응할 수 있다. 클러스탈 방법을 사용하는 단백질 서열의 % 동일성의 계산 및 쌍정렬을 위한 디폴트 파라미터는 케이터플 (KTUPLE)=1, 갭 페널티=3, 윈도우 (WINDOW)=5 및 다이아고날스 세이브드 (DIAGONALS SAVED)=5일 수 있다. 핵산의 경우, 이들 파라미터는 케이터플=2, 갭 페널티=5, 윈도우=4 및 다이아고날스 세이브드=4일 수 있다. 또한 대안적으로, 서열 사이의 % 동일성은 블로섬 (BLOSUM) 매트릭스 (예를 들

어, 블로섬62)을 사용하는 갭 오픈(GAP OPEN)=10, 갭 연장(GAP EXTEND)=0.5, 엔드 갭 페널티(END GAP PENALTY)=false, 엔드 갭 오픈=10, 엔드 갭 연장=0.5와 같은 파라미터를 갖는 엠보스(EMBOSS) 알고리즘 (예를 들어, 니들(needle))을 사용하여 수행될 수 있다.

[0083] 본 명세서에서, 제2 서열에 "상보적인" 제1 서열은, 대안적으로 제2 서열에 대해 "안티센스(antisense)" 배향인 것으로 지칭될 수 있다.

[0084] 다양한 폴리펩티드 아미노산 서열 및 폴리뉴클레오티드 서열은 개시된 본 발명의 특정 실시 형태의 특징으로서 본 명세서에 개시되어 있다. 본 명세서에 개시된 서열과 적어도 약 70 내지 85%, 85 내지 90% 또는 90% 내지 95% 동일한 이들 서열의 변이체가 사용될 수 있다. 대안적으로, 변이체 아미노산 서열 또는 폴리뉴클레오티드 서열은 본 명세서에 개시된 서열과 적어도 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% 동일성을 가질 수 있다. 변이체 아미노산 서열 또는 폴리뉴클레오티드 서열은 개시된 서열과 동일한 기능/활성, 또는 개시된 서열의 기능/활성의 적어도 약 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% 또는 99%를 갖는다.

[0085] 본 명세서의 Cas9 단백질의 각각의 아미노산 위치에서의 본 명세서에 개시된 모든 아미노산 잔기는 예이다. 특정 아미노산이 서로 유사한 구조 및/또는 전하 특징을 공유한다고 (즉, 보존된) 고려하면, Cas9의 각 위치에서의 아미노산은 개시된 서열에서 제공된 것과 같거나, 다음과 같이 보존된 아미노산 잔기로 치환될 수 있다 ("보존적 아미노산 치환"):

[0086] 1. 다음과 같은 작은 지방족의 비극성 또는 약간 극성인 잔기가 서로를 대체할 수 있다: Ala (A), Ser (S), Thr (T), Pro (P), Gly (G);

[0087] 2. 다음과 같은 극성의 음으로 하전된 잔기 및 그들의 아미드가 서로를 대체할 수 있다: Asp (D), Asn (N), Glu (E), Gln (Q);

[0088] 3. 다음과 같은 극성의 양으로 하전된 잔기가 서로를 대체할 수 있다: His (H), Arg (R), Lys (K);

[0089] 4. 다음과 같은 지방족의 비극성 잔기가 서로를 대체할 수 있다: Ala (A), Leu (L), Ile (I), Val (V), Cys (C), Met (M); 및

[0090] 5. 다음과 같은 큰 방향족 잔기가 서로를 대체할 수 있다: Phe (F), Tyr (Y), Trp (W).

[0091] 하기 실시예 1에 나타낸 바와 같이, Pol III 프로모터-전사된 gRNA를 사용하여 통상적이지 않은 효모, 예컨대 야로위아 리폴리티카에서 Cas9-매개 DNA 표적화를 수행하는 것이 어려운 것으로 입증되었다. 따라서, Cas9에 대한 RNA 성분을 생성하는 다른 방법은 통상적이지 않은 효모에서 Cas9-매개 DNA 표적화를 제공하는 데 관심이 있다.

[0092] 개시된 발명의 실시 형태는 5'-캡을 갖지 않는 적어도 하나의 RNA 성분을 포함하는 적어도 하나의 RNA-유도 엔도뉴클레아제 (RGEN)를 포함하는 통상적이지 않은 효모에 관한 것이다. 이러한 캡핑되지 않은 RNA 성분은 효모 내의 염색체 또는 에피솜의 표적 부위 서열에 상보적인 서열을 포함한다. RGEN은 표적 부위 서열의 전부 또는 일부에 결합하고, 임의로 절단할 수 있다.

[0093] 의미 있게도, RGEN-매개 DNA 표적화는 RGEN 표적 부위 서열과 외인성으로 공급된 공여 DNA 서열 사이의 상동 재조합 (HR) 수준의 증가 또는 삽입-결실 형성에 의해 명시되는 바와 같이 이러한 통상적이지 않은 효모에서 발생한다. 본 발명 이전에, 통상적이지 않은 효모는 일반적으로 HR에 의한 유전자 표적화로 다루기 힘들며, 전형적으로 표적 부위에 랜덤하고 드문 DNA 절단에 의존하여 공여 DNA와 이의 HR을 촉진하였다. 이는 낮은 HR 활성을 가지며 대신에 비상동 말단 결합 (NHEJ) 활성을 선호하는 통상적이지 않은 효모에 기인한다. 따라서, 통상적이지 않은 효모에서 HR에 의한 유전 표적화는 이제 NHEJ 공정보다 HR을 선호하는 S. 세레비시아와 같은 통상적인 효모에서와 같이 실현 가능할 수 있다. 임의의 이론에 구속되기를 바라지는 않지만, 통상적이지 않은 효모 세포에서 5'-캡이 없는 적어도 하나의 RNA 성분을 제공하는 것은 RGEN-매개 DNA 표적화에 참여할 수 있는 핵에서 RNA 성분을 더 잘 축적시킨다고 여겨진다.

[0094] Csy4 (Cas6)-기반 RNA 프로세싱 도구와 같은 RNA 프로세싱 도구가 기재되어 있다 (문헌[Nissim et aL. 2014.분자 세포 54:698-710]). Csy4는 예비-crRNA 스템-루프 반복 서열을 결합하고, 이의 동족 기질을 특이적으로 절단하여 반복 서열의 단편에 의해 플랭킹된 스페이서 서열을 포함하는 성숙 crRNA를 생성한다 (문헌[Sternberg et aL. 2012. RNA,18(4):661-72]). 본 명세서에 개시된 (실시예 12)는 5' 캡을 갖지 않는 RNA 성분 (가이드

RNA)을 생성하도록 Csy4를 사용하여 가이드 RNA를 처리하여, 여기서 RNA 성분은 통상적이지 않은 효모의 게놈에서 표적 부위에 결합하여 절단할 수 있는 RGEN을 형성할 수 있다.

[0095] 본 명세서의 통상적이지 않은 효모는 "통상적인" ("모델") 효모, 예컨대 사카로마이세스 (예를 들어, 출아 효모, 빵 효모 및/또는 맥주 효모로도 알려진 S. 세레비시아) 또는 스키조사카로마이세스 (예를 들어, 분열 효모로도 알려진 S. 폼베(*pombe*)) 종이 아니다. 소정 실시 형태에서, 통상적인 효모는 NHEJ에 매개되는 수복 공정보다 HR DNA 수복 공정을 선호하는 효모이다.

[0096] 소정 실시 형태에서, 통상적이지 않은 효모는 HR에 의해 매개되는 수복 공정보다 NHEJ DNA 수복 공정을 선호하는 효모일 수 있다. 통상적인 효모, 예컨대 사카로마이세스 세레비시아 및 스키조사카로마이세스 폼베는 전형적으로 짧은 플랭킹 상동성 암 (30 내지 50 bp)과 일반적으로 70% 초과의 효율로 공여 DNA의 특이적 통합을 나타내는 반면, 통상적이지 않은 효모, 예컨대 피키아 파스토리스(*Pichia pastoris*), 한세눌라 폴리모르파(*Hansenula polymorpha*), 야로위아 리폴리티카, 피키아 스티피티스(*Pichia stipitis*) 및 클루이베로마이세스 락티스(*Kluyveromyces lactis*)는 보통 유사한 구조의 공여 DNA와 1% 미만의 효율로 특이적 통합을 나타낸다 (문헌[Chen et al., *PLoS ONE* 8:e57952]). 따라서, HR 공정에 대한 선호도는, 예를 들어, 효모를 적합한 공여 DNA로 형질전환시키고 공여 DNA에 의해 표적화될 것으로 예측되는 게놈 부위와 특이적으로 재조합되는 정도를 측정함으로써 측정될 수 있다. 예를 들어, 이러한 검정이 효모 게놈에서 공여 DNA의 고도의 랜덤 통합을 산출하는 경우, NHEJ에 대한 선호도 (또는 HR에 대한 낮은 선호도)가 명백할 것이다. 효모에서 DNA의 특이적 (HR-매개) 및/또는 랜덤 (NHEJ-매개) 통합의 비율을 측정하기 위한 검정이 본 기술 분야에 알려져 있다 (예를 들어, 문헌[Ferreira and Cooper, *Genes Dev.* 18:2249-2254]; 문헌[Corrigan et al., *PLoS ONE* 8:e69628]; 문헌[Weaver et al., *ProC. NatL. AcaD. Sci. U.S.A.* 78:6354-6358]; 문헌[Keeney and Boeke, *Genetics* 136:849-856]).

[0097] 낮은 수준의 HR 활성을 고려하면, 본 명세서의 통상적이지 않은 효모는 (i) 예를 들어, 약 1%, 2%, 3%, 4%, 5%, 6%, 7% 또는 8% 미만의 30 내지 50 bp의 플랭킹 상동성 암을 갖는 적합한 공여 DNA에 의한 특이적인 표적화 비율을 나타내고/내거나 (ii) 예를 들어, 약 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74% 또는 75% 초과의 상기 공여 DNA의 랜덤 통합 비율을 나타낼 수 있다. 적합한 공여 DNA의 (i) 특이적 표적화 및/또는 (ii) 랜덤 통합의 이러한 비율은 본 명세서에 개시된 바와 같은 RGEN이 제공되기 전에 존재하는 그대로의 통상적이지 않은 효모를 특성화할 수 있다. 소정 실시 형태에서, 통상적이지 않은 효모에 RGEN을 제공하는 목적은 특정 부위에서 효모를 HR쪽으로 편향시키기 위한 부위 특이적 DNA 단일 가닥 절단 (SSB) 또는 이중 가닥 절단 (DSB)을 생성하는 것이다. 따라서, 본 명세서의 적합한 RGEN을 포함하는 통상적이지 않은 효모는 전형적으로 특정 공여 DNA와의 HR 비율의 증가를 나타내야 한다. 이러한 증가된 비율은 적합한 대조 (예를 들어, 동일한 공여 DNA로 형질전환되었지만, 적합한 RGEN가 결여된 동일한 통상적이지 않은 효모)에서의 HR 비율보다 적어도 약 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9- 또는 10-배 더 높을 수 있다.

[0098] 본 명세서의 소정 측면에서, 통상적이지 않은 효모는 무성으로 (무성생식형(anamorphic)) 또는 유성으로 (유성생식형(teleomorphic)) 생식할 수 있다. 본 명세서의 통상적이지 않은 효모는 전형적으로 단세포 형태로 존재하지만, 이들 효모의 특정 유형은 임의로 가성균사(pseudohyphae) (연결된 발아 세포의 줄)를 형성할 수 있다. 또 다른 측면에서, 통상적이지 않은 효모는 반수체 또는 이배체일 수 있고/있거나 이들 배수성 형태 중 어느 하나로 존재하는 능력을 가질 수 있다.

[0099] 본 명세서의 통상적이지 않은 효모는 이들 모두가 본 명세서에 참고로 포함되는 문헌[Non-Conventional Yeasts in Genetics, Biochemistry and Biotechnology: Practical Protocols (K. Wolf, K.D. Breunig, G. Barth, Eds., Springer-Verlag, Berlin, Germany, 2003)], 문헌[Yeasts in Natural and Artificial Habitats (J.F.T. Spencer, D.M. Spencer, Eds., Springer-Verlag, Berlin, Germany, 1997)], 및/또는 문헌[Yeast Biotechnology: Diversity and Applications (T. Satyanarayana, G. Kunze, Eds., Springer, 2009)]에 기재된 바와 같은 본 기술 분야에 알려진 임의의 방법에 따라 배양될 수 있다.

[0100] 본 명세서의 통상적이지 않은 효모의 비제한적인 예는 다음과 같은 속의 효모를 포함한다: 야로위아, 피키아, 쉬반니오마이세스, 클루이베로마이세스, 아르술라, 트리코스포론, 칸디다, 유스틸라고, 토룰롭시스, 자이고사카로마이세스, 트리고놉시스, 크립토코커스, 로도토룰라, 파피아, 스포로볼로마이세스 및 파치솔렌. 야로위아 종의 적합한 예는 야로위아 리폴리티카이다. 피키아 종의 적합한 예에는 P. 파스토리스, P. 메타놀리카(*methanolica*), P. 스티피티스, P. 아노말라(*anomala*) 및 P. 앙구스타(*angusta*)가 포함된다. 쉬반니오마이세스 종의 적합한 예에는 *S.* 카스텔리이(*castellii*), S. 알루비우스(*alluvius*), S. 호미니스 (*hominis*), S. 옥시

덴탈리스(*occidentalis*), S. 카프리오티이(*capriottii*), S. 에트첼시이(*etchellsii*), S. 폴리모르푸스(*polymorphus*), S. 슈도폴리모르푸스(*pseudopolymorphus*), S. 반리자에(*vanrijiae*) 및 S. 야마다에(*yamadae*)가 포함된다. 루이베로마이세스 종의 적합한 예에는 K. 락티스, K. 마르시아누스(*marxianus*), K. 프라길리스(*fragilis*), K. 드로소필라룸(*drosophilarum*), K. 써모톨레란스(*thermotolerans*), K. 파세올로스포루스(*phaseolosporus*), K. 바누데니이(*vanudenii*), K. 왈티이(*waltii*), K. 아프리카누스(*africanus*) 및 K. 폴리스포루스(*polysporus*)가 포함된다. 아르술라 종의 적합한 예에는 A. 아데니니보란스(*adeninivorans*) 및 A. 테레스트레(*terRestre*)가 포함된다. 트리코스포론 종의 적합한 예에는 T. 쿠타네움(*cutaneum*), T. 카피타툼(*capitatum*), T. 인킨(*inkin*) 및 T. 비메리(*beemeri*)가 포함된다. 칸디다 종의 적합한 예에는 C. 알비칸스(*albicans*), C. 아스칼라피다움(*ascalaphidarum*), C. 암피시애(*amphixiae*), C. 앤트아티카(*antarctica*), C. 아르젠티아(*argentea*), C. 아틀란티카(*atlantica*), C. 아트모스패리카(*atmosphaerica*), C. 블랏타에(*blattae*), C. 브로멜리아세아룸(*bromeliacearum*), C. 카르포필리아(*carpophila*), C. 카르바잘리스(*carvajalis*), C. 세람비시다룸(*cerambycidarum*), C. 차우리오데스(*chauliodes*), C. 코리달리(*corydali*), C. 도스세이(*dosseyi*), C. 듀블리니엔시스(*dubliniensis*), C. 얼가텐시스(*ergatensis*), C. 프룩투스(*fructus*), C. 글라브라타(*glabrata*), C. 퍼멘타티(*fermentati*), C. 귈리어몬디이(*guilliermondii*), C. 해뮬로니이(*haemulonii*), C. 인섹타멘스(*insectamens*), C. 인섹토룸(*insectorum*), C. 인터메디아(*intermedia*), C. 제프레시이(*jeffResii*), C. 케피르(*kefyr*), C. 케로세네아에(*keroseneae*), C. 크루세이(*krusei*), C. 루시타니애(*lusitaniae*), C. 릭소소필리아(*lyxosophila*), C. 말토사(*maltosa*), C. 마리나(*marina*), C. 멤브라니파시엔스(*membranifaciens*), C. 밀러리(*milleri*), C. 모기이(*mogii*), C. 올레필리아(*oleophila*), C. 오레고네시스(*oregonensis*), C. 파라프실로시스(*parapsilosis*), C. 퀴어시트루사(*quercitrusa*), C. 루고사(*rugosa*), C. 사케(*sake*), C. 쉐하테아(*shehatea*), C. 템노칠라에(*temnochilae*), C. 테누이스(*tenuis*), C. 테아에(*theae*), C. 톨러란스(*tolerans*), C. 트로피칼리스(*tropicalis*), C. 츠시이아에(*tsuchiyae*), C. 시놀라보란티움(*sinolaborantium*), C. 소자에(*sojae*), C. 수브하쉬이(*subhashii*), C. 비스와나치이(*viswanathii*), C. 유틸리스(*utilis*), C. 우바투벤시스(*ubatubensis*) 및 C. 젬플리니나(*zemplinina*)가 포함된다. 유스틸라고 종의 적합한 예에는 U. 아베나에(*avenae*), U. 에스쿨렌타(*esculenta*), U. 호르데이(*hordei*), U. 마이디스(*maydi*s), U. 누다(*nuda*) 및 U. 트리티치(*tritici*)가 포함된다. 토룰롭시스 종의 적합한 예에는 T. 게오차레스(*geochaRes*), T. 아지마(*azyma*), T. 글라브라타(*glabrata*) 및 T. 칸디다가 포함된다. 자이고사카로마이세스 종의 적합한 예에는 Z. 바일리이(*bailii*), Z. 비스포루스(*bisporus*), Z. 시드리(*cidri*), Z. 퍼멘타티(*fermentati*), Z. 플루오렌티누스(*florentinus*), Z. 콤부차엔시스(*kombuchaensis*), Z. 렌투스(*lentus*), Z. 멜리스(*mellis*), Z. 미크로엘리프소이데스(*microellipsoides*), Z. 엠라키이(*mrakii*), Z. 슈도로추시이(*pseudorouxii*) 및 Z. 로우시이가 포함된다. 트리고놉시스 종의 적합한 예에는 T. 바리아빌리스(*variabilis*)가 포함된다. 크립토코커스 종의 적합한 예에는 C. 라우렌티이(*laurentii*), C. 알비두스(*albidus*), C. 네오포르만스(*neoformans*), C. 가티이(*gattii*), C. 유니구툴라투스 (*uniguttulatus*), C. 아델리엔시스(*adeliensis*), C. 아에리우스(*aerius*), C. 알비도시밀리스(*albidosimilis*), C. 안타르크티쿠스(*antarcticus*), C. 아쿠아티쿠스(*aquaticus*), C. 아테르(*ater*), C. 부타네시스(*bhutanensis*), C. 콘소르티오니스(*consortionis*), C. 쿠르바투스(*curvatus*), C. 페놀리쿠스(*phenolicus*), C. 스킨네리(*skinneri*), C. 테레우스(*terreus*) 및 C. 비쉬니마치(*vishniacci*)가 포함된다. 로도토룰라 종의 적합한 예에는 R. 아체니오룸(*acheniorum*), R. 툴라(*tula*), R. 아쿠타(*acuta*), R. 아메리카나(*americana*), R. 아라우차리아에(*araucariae*), R. 아르티카(*arctica*), R. 아르메니아카(*armeniaca*), R. 아우란티아카(*aurantiaca*), R. 아우리쿨라리아에(*auriculariae*), R. 바카룸(*bacarum*), R. 벤티카(*benthica*), R. 비오우르게이(*biourgei*), R. 보고리엔시스(*bogoriensi*s), R. 브론치알리스(*bronchialis*), R. 부포니이(*buffonii*), R. 칼립토게나에(*calyptogenae*), R. 충나멘시스(*chungnamensis*), R. 클란디엔시스(*cladiensis*), R. 코랄리나(*corallina*), R. 크레솔리카(*cResolica*), R. 크로세아(*crocea*), R. 사이클로클라스티카(*cycloclastica*), R. 다이레넨시스(*dairenensis*), R. 디플루덴스(*diffluens*), R. 에베르글라디엔시스(*evergladiensis*), R. 페룰리카(*ferulica*), R. 폴리오룸(*foliorum*), R. 프라가리아(*fragaria*), R. 푸지사넨시스(*fujisanensis*), R. 푸트로넨시스(*futronensis*), R. 겔라티노사(*gelatinosa*), R. 글라시알리스(*glacialis*), R. 글루티니스(*glutinis*), R. 그라실리스(*gracilis*), R. 그라미니스(*graminis*), R. 그린베르그시이(*grinbergsii*), R. 히말라엔시스(*himalayensis*), R. 힌눌레아(*hinnulea*), R. 히스톨리티카(*histolytica*), R. 힐로필리아(*hylophila*), R. 인카르나타(*incarnata*), R. 인게니오사(*ingeniosa*), R. 자바니카(*javanica*), R. 코이쉬카웬시스(*koishikawensis*), R. 락토사(*lactosa*), R. 라멜리브라치아에(*lamellibrachiae*), R. 라린기스(*laryngis*), R. 리그노필라(*lignophila*), R. 리니(*lini*), R. 론기스시마(*longissima*), R. 루드위기이(*ludwigii*), R. 리시노필라(*lysinophila*), R. 마리나(*marina*), R. 마르티니아에-프라간티스(*martyniae-fragantis*), R. 마트리텐시스(*matritensis*), R. 멜리(*meli*), R. 미누타(*minuta*), R. 무실라기노사(*mucilaginosa*), R. 니텐스(*nitens*), R.

노쏘파기(*nothofagi*), R. 오리자에(*oryzae*), R. 파시피카(*pacifica*), R. 팔리다(*pallida*), R. 페네아우스(*peneaus*), R. 필릴라(*philyla*), R. 필로플라나(*phylloplana*), R. 필라티이(*pilatii*), R. 필리마나에(*pilimanae*), R. 피니콜라(*pinicola*), R. 플리카타(*plicata*), R. 폴리모르파(*polymorpha*), R. 사이츠로페놀리카(*psychrophenolica*), R. 사이츠로필라(*psychrophila*), R. 푸스툴라(*pustula*), R. 레티노필라(*retinophila*), R. 로사세아(*rosacea*), R. 로술라타(*rosulata*), R. 루베파시엔스(*rubefaciens*), R. 루벨라(*rubella*), R. 루베센스(*rubescens*), R. 루브라(*rubra*), R. 루브로루고사(*rubrorugosa*), R. 루풀라(*rufula*), R. 루틸라(*rutila*), R. 산구이네아(*sanguinea*), R. 산니에이(*sanniei*), R. 사르토리이(*sartoryi*), R. 실베스트리스(*silvestris*), R. 실플렉스(*simplex*), R. 시넨시스(*sinensis*), R. 슬루피아에(*slooffiae*), R. 순키이(*sonckii*), R. 스트라미네아(*straminea*), R. 수베리콜라(*subericola*), R. 수가니이(*suganii*), R. 타이와넨시스(*taiwanensis*), R. 타이와니아나(*taiwaniana*), R. 테르페노이달리스(*terpenoidalis*), R. 테레아(*terrea*), R. 텍센시스(*texensis*), R. 토쿄엔시스(*tokyoensis*), R. 울자마에(*ulzamae*), R. 바닐리카(*vanillica*), R. 부일레미니이(*vuilleminii*), R. 야로위이, R. 윤나넨시스(*yunnanensis*) 및 R. 졸티이(*zsoltii*)가 포함된다. 파피아 종의 적합한 예에는 P. 로도지마(*rhodozyma*)가 포함된다. 스포로볼로마이세스 종의 적합한 예에는 S. 알보루베센스(*alborubescens*), S. 반나엔시스(*bannaensis*), S. 베이징엔시스(*beijingensis*), S. 비스초리아에(*bischofiae*), S. 클라바투스(*clavatus*), S. 코프로스마에(*coprosmae*), S. 코프로미콜라(*coprosmicola*), S. 코랄리누스(*corallinus*), S. 디메나에(*dimmenae*), S. 드라코필리(dracophylli), S. 에론가투스(elongatus), S. 그라실리스(*gracilis*), S. 이노시토필루스(*inositophilus*), S. 존소니이(*johnsonii*), S. 코알라에(*koalae*), S. 마그니스포루스(*magnisporus*), S. 노보제알란이쿠스(*novozealandicus*), S. 오도루스(*odorus*), S. 파타고니쿠스(*patagonicus*), S. 프로덕투스(*productus*), S. 로세우스(*roseus*), S. 사시콜라(*sasicola*), S. 쉴바타누스(*shibatanus*), S. 신굴라이스(*singularis*), S. 수브룬네우스(*subbrunneus*), S. 시메트리쿠스(*symmetricus*), S. 시지기이(*syzygii*), S. 타우포엔시스(*taupoensis*), S. 수가에(*tsugae*), S. 잔투스(*xanthus*) 및 S. 윤나넨시스가 포함된다. 파치솔렌 종의 적합한 예에는 P. 탄노필루스(*tannophilus*)가 포함된다.

[0101] 야로위아 리폴리티카는 본 명세서에 개시된 소정 실시 형태에서 바람직하다.

[0102] 적합한 야로위아 리폴리티카의 예에는 다음과 같은 아메리칸 타입 컬쳐 컬렉션(American Type Culture Collection) (버지니아주 매나사스 소재의 ATCC)로부터 입수가능한 분리주가 포함된다: 균주명 ATCC #20362, #8862, #8661, #8662, #9773, #15586, #16617, #16618, #18942, #18943, #18944, #18945, #20114, #20177, #20182, #20225, #20226, #20228, #20327, #20255, #20287, #20297, #20315, #20320, #20324, #20336, #20341, #20346, #20348, #20363, #20364, #20372, #20373, #20383, #20390, #20400, #20460, #20461, #20462, #20496, #20510, #20628, #20688, #20774, #20775, #20776, #20777, #20778, #20779, #20780, #20781, #20794, #20795, #20875, #20241, #20422, #20423, #32338, #32339, #32340, #32341, #34342, #32343, #32935, #34017, #34018, #34088, #34922, #34922, #38295, #42281, #44601, #46025, #46026, #46027, #46028, #46067, #46068, #46069, #46070, #46330, #46482, #46483, #46484, #46436, #60594, #62385, #64042, #74234, #76598, #76861, #76862, #76982, #90716, #90811, #90812, #90813, #90814, #90903, #90904, #90905, #96028, #201241, #201242, #201243, #201244, #201245, #201246, #201247, #201249 및/또는 #201847.

[0103] 본 명세서 임의의 다른 통상적이지 않은 효모뿐만 아니라 야로위아 리폴리티카는 유지성일 수 있고/있거나 (예를 들어, 이의 건조 세포 중량의 적어도 25%를 오일로서 생성함), 하나 이상의 다중불포화 지방산 (예를 들어, 오메가-6 또는 오메가-3)을 생성할 수 있다. 이러한 유지성은 효모가 이의 야생형 형태와 비교하여 증가된 양의 지질을 생산하도록 유전적으로 조작된 결과일 수 있다. 유지성 야로위아 리폴리티카 균주의 예가 본 명세서에 참고로 포함되는 미국 특허 출원 공개 제2009/0093543호, 제2010/0317072호, 제2012/0052537호 및 제2014/0186906호에 개시되어 있다.

[0104] 통상적이지 않은 효모에 대해 본 명세서에 개시된 실시 형태는 또한 다른 미생물, 예컨대 진균류에 적용될 수 있다. 소정 실시 형태에서, 진균류는 HR에 의해 매개되는 수복 공정보다 NHEJ DNA 수복 공정을 선호하는 진균류일 수 있다. 본 명세서의 진균은 담자균(*Basidiomycetes*), 접합균(*Zygomycetes*), 호상균(*Chytridiomycetes*) 또는 자낭균 진균류일 수 있다. 본 명세서의 사상진균의 예에는 트리코데르마, 크리소스포리움(*Chrysosporium*), 티엘라비아(*Thielavia*), 뉴로스포라(*Neurospora*) (예를 들어, N. 크라사(*crassa*), N. 시토실라(*sitophila*)), 크리포넥트리아(*Cryphonectria*) (예를 들어, C. 파라시티카(*parasitica*)), 아우레오바시디움(*Aureobasidium*) (예를 들어, A. 풀루란스(*pullulans*)), 필리바시디움(*Filibasidium*), 피로마이세스(*Piromyces*), 크립토콕쿠스(*Cryplococcus*), 아크레모니움(*Acremonium*), 톨리포클라디움(*Tolypocladium*), 신탈

리디움(*Scytalidium*), 스키조필룸(*Schizophyllum*), 스포로트리쿰(*Sporotrichum*), 페니실리움(Penicillium) (예를 들어, P. 빌라이아에(*bilaiae*), P. 카멤베르티(*camemberti*), P. 칸디둠(*candidum*), P. 크리소게눔(*chrysogenum*), P. 엑스펜숨(*expansum*), P. 퍼니쿨로섬(*funiculosum*), P. 글라우컴(*glaucum*), P. 마네페이(*marneffei*), P. 로커포르티(*roqueforti*), P. 베르루코숨(*verrucosum*), P. 비리디카툼(*viridicatum*)), 기베렐라(*Gibberella*) (예를 들어, G. 아쿠미나타(*acuminata*), G. 아베나세아(*avenacea*), G. 바카타(*baccata*), G. 키르키나타(*circinata*), G. 시아노게나(*cyanogena*), G. 푸지쿠로이(*fujikuroi*), G. 인트리칸스(*intricans*), G. 폴리카리스(*pulicaris*), G. 스틸보이데스(*stilboides*), G. 트리킨크타(*tricincta*), G. 제에(*zeae*)), 미셀리오프토라(*Myceliophthora*), 뮤코(*Mucor*) (예를 들어, M. 로우시(rouxii), M. 시르시넬로이데스(*circinelloides*)), 아스퍼질러스(*Aspergillus*) (예를 들어, A. 니거(*niger*), A. 오리자에(*oryzae*), A. 니둘란스(*nidulans*), A. 플라버스(*flavus*), A. 렌툴루스(*lentulus*), A. 테레우스, A. 클라바투스, A. 푸미가투스(*fumigatus*)), 푸사리움(*Fusarium*) (예를 들어, F. 그라미네아룸(*graminearum*), F. 옥시스포룸(*oxysporum*), F. 부비게눔(*bubigenum*), F. 솔라니(*solani*), F. 옥시스포룸, F. 버티실리오이데스(*verticillioides*), F. 프로리페라툼(*proliferatum*), F. 베네나툼(*venenatum*)) 및 후미콜라(*Humicola*) 속의 것들 및 이들의 무성생식형과 유성생식형이 포함된다. 본 명세서에 진균류의 속 및 종은 문헌[Barnett and Hunter (Illustrated Genera of Imperfect Fungi, 3rd Edition, Burgess Publishing Company, 1972)]에 개시된 바와 같은 형태에 의해 원하는 경우 정의될 수 있다. 진균은 임의로 해충/병원균, 예컨대 동물 (예를 들어, 인간)의 해충/병원균으로 특징지어질 수 있다.

[0105] 본 명세서의 소정 측면에서 트리코데르마 종에는 T. 아그레시범(*aggRessivum*), T. 아마조니쿰(*amazonicum*), T. 아스페렐룸(*asperellum*), T. 아트로비리데(*atroviride*), T. 아우레오비리데(*aureoviride*), T. 아우스트로코닌기이(*austrokoningii*), T. 브레비콤팩툼(*brevicompactum*), T. 칸디둠, T. 카리 배움(*caribbaeum*), T. 카톱트론(*catoptron*), T. 크레메움(*cremeum*), T. 세라미쿰(*ceramicum*), T. 세리넘(*cerinum*), T. 클로로스포룸(*chlorosporum*), T. 크로모스페르뭄(*chromospermum*), T. 신나모메움(*cinnamomeum*), T. 시트리노비리데(*citrinoviride*), T. 크라숨(*crassum*), T. 크레메움(*cremeum*), T. 딩글레예아(*dingleyeae*), T. 도로테아에(*dorotheae*), T. 에푸숨(*effusum*), T. 에리나세움(*erinaceum*), T. 에스토니쿰(*estonicum*), T. 페르틸레(*fertile*), T. 겔라티노수스(*gelatinosus*), T. 하넨세(*ghanense*), T. 하마툼(*hamatum*), T. 하르지아눔(*harzianum*), T. 헬리쿰(*helicum*), T. 인트리칸스, T. 코니랑브라(*konilangbra*), T. 코닌기이(*koningii*), T. 코닌지옵시스(*koningiopsis*), T. 론지브라치아툼(*longibrachiatum*), T. 론지필레(*longipile*), T. 미누티스포룸(*minutisporum*), T. 오블론지스포룸(*oblongisporum*), T. 오발리스포룸(*ovalisporum*), T. 페테르세니(*petersenii*), T. 필로스타히디스(*phyllostahydis*), T. 필루리페룸 (*piluliferum*), T. 플레우로티콜라(*pleuroticola*), T. 플레우로툼(*pleurotum*), T. 폴리스포룸(*polysporum*), T. 슈도코닌기(*pseudokoningii*), T. 푸베센스(*pubescens*), T. 레세이(*reesei*), T. 로게르소니(*rogersonii*), T. 로시쿰(*rossicum*), T. 사투르니스포룸(*saturnisporum*), T. 시넨시스(*sinensis*), T. 시누오숨(*sinuosum*), T. 스피랄레(*spirale*), T. 스트라미네움(*stramineum*), T. 스트리고숨(*strigosum*), T. 스트로마티쿰(*stromaticum*), T. 수로툰둠(*surrotundum*), T. 타이와넨세(*taiwanense*), T. 타일랜디쿰(*thailandicum*), T. 텔레포리콜룸(*thelephoricolum*), T. 테오브로미콜라(*theobromicola*), T. 토멘토숨(*tomentosum*), T. 벨루티눔(*velutinum*), T. 베렌스(*virens*), T. 비리데(*viride*) 및 T. 비리데넨스(*viridescens*)가 포함된다. 본 명세서의 트리코데르마 종은, 예를 들어, 본 명세서에 참고로 포함되는 문헌[Trichoderma: Biology and Applications (P.K. Mukherjee et al., Eds., CABI, Oxfordshire, UK, 2013)]에 기재된 바와 같이 배양되고/거나 조작될 수 있다.

[0106] 소정 실시 형태에서, 미생물 세포는 조류 세포이다. 예를 들어, 조류 세포는 다음 중 어느 하나의 유래일 수 있다: 녹조류(green algae), 홍조류(red algae), 갈조류(brown algae), 규조류(diatoms) 및 와편모충류 (와편모조류). 다른 측면에서, 조류 세포는 미세조류 (예를 들어, 식물성 플랑크톤, 미소식물 또는 플랑크톤 조류) 또는 거대조류 (켈프, 해초)일 수 있다. 추가의 예로서, 본 명세서의 조류 세포는 포프리라(*Porphyra*) (김), 팔마리아(*Palmaria*) 종, 예컨대 P. 팔마타(*palmata*) (덜스(*dulse*)), 아르트로스피라(*Arthrospira*) 종, 예컨대 A. 플라텐시스(*platensis*) (스피룰리나(*spirulina*)), 클로렐라(*Chlorella*) (예를 들어, C. 프로토테코이데스(*protothecoides*)), 콘드루스(*Chondrus*) 종, 예컨대 C. 크리스푸스(*crispus*) (아이리쉬 모스(*Irish moss*)), 아파니조메논(*Aphanizomenon*), 사르가쑴(*Sargassum*), 코차유요(*Cochayuyo*), 보트리코쿠스(*Botryococcus*) (예를 들어, B. 브라우니이(*braunii*)), 두날리엘라(*Dunaliella*) (예를 들어, D. 테르티올렉타(*tertiolecta*), 그라실라리아(*Gracilaria*), 플레우로크리시스(*Pleurochrysis*) (예를 들어, P. 카르테아에(*carterae*)), 안키스트로데스무스(*Ankistrodesmus*), 시클로텔라(*Cyclotella*), 한츠슈이아(*Hantzschia*), 난노클로리스(*Nannochloris*), 난노클로롭시스(*Nannochloropsis*), 니츠키아(*Nitzschia*), 패오닥틸룸(*Phaeodactylum*) (예를 들어, P. 트리코누툼

(*tricornutum*)), 세네데스무스(*Scenedesmus*), 스티코코커스(*Stichococcus*), 테트라셀미스(*Tetraselmis*) (예를 들어, T. 수에시카(*suecica*)), 탈라시오시라(*Thalassiosira*) (예를 들어, T. 슈도나나((*pseudonana*)), 크립테코디니움(*Crypthecodinium*) (예를 들어, C. 코흐니(*cohnii*)), 네오클로리스(*Neochloris*) (예를 들어, N. 올레오아분단스(*oleoabundans*)) 또는 치오키트리움(*Schiochytrium*)일 수 있다. 본 명세서의 조류 종은, 예를 들어, 본 명세서에 참고로 포함되는 문헌[Thompson (Algal Cell CulturE. Encyclopedia of Life Support System (EOLSS), Biotechnology Vol 1, available at eolss.net/sample-chapters internet site)]에 기재된 바와 같이 배양되고/거나 조작될 수 있다.

[0107] 본 명세서의 5'-캡을 갖지 않는 적어도 하나의 RNA 성분을 포함하는 적어도 하나의 RGEN을 포함하는 통상적이지 않은 효모는 자연계에서 발생하지 않는다. 임의의 특정 이론에 구애되고자 함이 없이, 본 명세서의 RGEN이, 예를 들어 원핵생물에서만 발생하는 것으로 알려져왔기 때문에 이러한 효모는 자연적으로 발생하지 않는 것으로 여겨진다. 또한, 여겨진다 효모의 소정 실시 형태는 crRNA와 tracrRNA의 이종 결합을 나타내는 gRNA를 포함하는 RNA 성분을 갖는 RGEN을 포함함으로 인해 자연적으로 발생하지 않는다.

[0108] 본 명세서에서 RGEN은 적어도 하나의 Cas 단백질과 적어도 하나의 RNA 성분을 포함하는 복합체를 지칭한다. 적합한 Cas 단백질의 예에는 하나 이상의 타입 I, II 또는 III 크리스퍼 시스템의 Cas 엔도뉴클레아제가 포함된다(본 명세서에 참고로 포함되는 문헌[(Bhaya et al., *Annu. Rev. GeneT.* 45:273-297]). 타입 I 크리스퍼 Cas 단백질은, 예를 들어 Cas3 또는 Cas4 단백질일 수 있다. 타입 II 크리스퍼 Cas 단백질은, 예를 들어 Cas9 단백질일 수 있다. 타입 III 크리스퍼 Cas 단백질은, 예를 들어 Cas10 단백질일 수 있다. Cas9 단백질이 바람직한 실시 형태에서 사용된다. 소정 실시 형태에서, Cas 단백질은 박테리아 또는 고세균 단백질일 수 있다. 본 명세서에서, 타입 I 내지 타입 III 크리스퍼 Cas 단백질은 전형적으로 원핵생물 기원이며; 예를 들어, 타입 I 및 타입 III Cas 단백질은 박테리아 또는 고세균 종으로부터 유래될 수 있는 반면, II Cas 단백질은 (즉, Cas9) 박테리아 종으로부터 유래될 수 있다. 다른 실시 형태에서, 적합한 Cas 단백질은 Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm^3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, 이들의 상동체 또는 이들의 변형된 버전 중 하나 이상을 포함한다.

[0109] 개시된 발명의 다른 측면에서, 본 명세서의 Cas 단백질은 다음의 속 중 어느 하나의 유래일 수 있다: 아에로피룸(*Aeropyrum*), 피로바쿨룸(*Pyrobaculum*), 술폴로부스(*Sulfolobus*), 아캐오글로부스(*Archaeoglobus*), 할로카르쿨라(*Haloarcula*), 메타노박테리움(*Methanobacteriumn*), 메타노코커스(*Methanococcus*), 메타노사르시나(*Methanosarcina*), 메타노피러스(*Methanopyrus*), 피로코커스(*Pyrococcus*), 피크로필러스(*Picrophilus*), 써모플라스마(*Therninoplasnia*), 코리네박테리움(*Corynebacterium*), 마이코박테리움(*Mycobacterium*), 스트렙토마이세스(*Streptomyces*), 아퀴펙스(*Aquifrx*), 포르피로모나스(*Porphvromonas*), 클로로비움(*Chlorobium*), 써머스(*Thermus*), 바실러스, 리스테리아(*Listeria*), 스타필로코커스(*Staphylococcus*), 클로스트리디움(*Clostridium*), 써모아나에로박터(*Thermoanaerobacter*), 마이코플라스마(*Mycoplasma*), 푸소박테리움(*Fusobacterium*), 아자쿠스(*Azarcus*), 크로모박테리움(*Chromobacterium*), 네이세리아(*Neisseria*), 니트로소모나스(*Nitrosomonas*), 디설포비브리오(*Desulfovibrio*), 게오박터(*Geobacter*), 믹소코커스(*Myrococcus*), 캄필로박터(*Campylobacter*), 볼리넬라(*Wolinella*), 아시네토박터(*Acinetobacter*), 에르위니아(*Erwinia*), 에스케리키아(*Escherichia*), 레지오넬라(*Legionella*), 메틸로코커스(*Methylococcus*), 파스퇴렐라(*Pasteurella*), 포토박테리움(*Photobacterium*), 살모넬라(*Salmonella*), 잔토모나스(*Xanthomonas*), 예르시니아(*Yersinia*), 스트렙토코커스, 트레포네마(*Treponema*), 프란시셀라(*Francisella*) 또는 써모토가(*Thermotoga*). 대안적으로, 본 명세서의 Cas 단백질은, 예를 들어, 본 명세서에 참고로 포함되는 미국 특허 출원 공개 제2010/0093617호에 개시된 바와 같이 서열 번호:462-465, 467-472, 474-477, 479-487, 489-492, 494-497, 499-503, 505-508, 510-516 또는 517-521 중 어느 하나에 의해 암호화될 수 있다.

[0110] 소정 실시 형태에서 RGEN은 Cas9 아미노산 서열을 포함한다. 본 명세서의 Cas9 단백질의 아미노산 서열과 본 명세서의 소정 다른 Cas 단백질은, 예를 들어, 스트렙토코커스 (예를 들어, S. 피오제네스, S. 뉴모니애(*pneumoniae*), S. 써모필러스(*thermophilus*), S. 아갈락티아(*agalactiae*), S. 파라상귀니스(*parasanguinis*), S. 오랄리스(*oralis*), S. 살리바리우스(*salivarius*), S. 마카카(*macacae*), S. 디스갈락티애(*dysgalactiae*), S. 안지노서스(*anginosus*), S. 콘스텔라투스(*constellatus*), S. 슈도포르시누스(*pseudoporcinus*), S. 뮤탄스(*mutans*)), 리스테리아 (예를 들어, L. 인노쿠아(*innocua*)), 스피로플라즈마(*Spiroplasma*) (예를 들어, S. 아피스(*apis*), S. 시르피디콜라(*syrphidicola*)), 펩토스트렙토코카세(*Peptostreptococcaceae*), 아토포비움

(*Atopobium*), 포르피로모나스(*Porphyromonas*) (예를 들어, P. 카토니아(*catoniae*)), 프레보텔라(*Prevotella*) (예를 들어, P. 인테르메디아(*intermedia*)), 베일로넬라(*Veillonella*), 트레포네마(*Treponema*) (예를 들어, T. 소크란스키(*socranskii*), T. 티콜라(*denticola*)), 카프노사이토파가(*Capnocytophaga*), 피네골디아(*Finegoldia*) (예를 들어, F. 마그나(*magna*)), 코리오박테리아세아에(*Coriobacteriaceae*) (예를 들어, C. 박테리움(*bacterium*)), 올스넬라(*Olsenella*) (예를 들어, O. 프로푸사(*profusa*)), 헤모필루스(*Haemophilus*) (예를 들어, H. 스푸토룸(*sputorum*), H. 피트마니아에(*pittmaniae*)), 파스튜렐라(*Pasteurella*) (예를 들어, P. 베티아에(*bettyae*)), 올리비박터(*Olivibacter*) (예를 들어, O. 시티엔시스(*sitiensis*)), 에필리토니모나스(*Epilithonimonas*) (예를 들어, E. 테낙스(*tenax*)), 메소니아(*Mesonia*) (예를 들어, M. 모빌리스(*mobilis*)), 락토바실루스(*Lactobacillus*) (예를 들어, L. 플란타룸(*plantarum*)), 바실루스 (예를 들어, B. 세레우스(*cereus*)), 아퀴마리나(*Aquimarina*) (예를 들어, A. 무엘레리(*muelleri*)), 크리세오박테리(*Chryseobacterium*) (예를 들어, C. 파루스트레(*palustre*)), 박테로이데스(*Bacteroides*) (예를 들어, B. 그라미니솔벤스(*graminisolvens*)), 네이세리아 (예를 들어, N. 메닝기티디스(*meningitidis*)), 프란시셀라(*Francisella*) (예를 들어, F. 노비시다(*novicida*)) 또는 플라보박테리움(*Flavobacterium*) (예를 들어, F. 프리기다리움(*frigidarium*), F. 솔리(*soli*)) 종으로부터 유래될 수 있다. 본 명세서의 소정 측면에서 S. 피오제네스 Cas9가 바람직하다. 또 다른 예로서, Cas9 단백질은 본 명세서에 참고로 포함되는 문헌[Chylinski et aL. (RNA Biology 10:726-737)]에 개시된 Cas9 단백질 중 어느 하나 일 수 있다.

[0111] 따라서, 본 명세서의 Cas9 단백질의 서열은, 예를 들어, 참고로 포함되는 진뱅크(GenBank) 기탁 번호 G3ECR1 (S. 써모필러스), WP_026709422, WP_027202655, WP_027318179, WP_027347504, WP_027376815, WP_027414302, WP_027821588, WP_027886314, WP_027963583, WP_028123848, WP_028298935, Q03JI6 (S. 써모필러스), EGP66723, EGS38969, EGV05092, EHI65578 (S. 슈도포르시누스), EIC75614 (S. 오랄리스), EID22027 (S. 콘스텔라투스), EIJ69711, EJP22331 (S. 오랄리스), EJP26004 (S. 안지노서스), EJP30321, EPZ44001 (S. 피오제네스), EPZ46028 (S. 피오제네스), EQL78043 (S. 피오제네스), EQL78548 (S. 피오제네스), ERL10511, ERL12345, ERL19088 (S. 피오제네스), ESA57807 (S. 피오제네스), ESA59254 (S. 피오제네스), ESU85303 (S. 피오제네스), ETS96804, UC75522, EGR87316 (S. 디스갈락티애), EGS33732, EGV01468 (S. 오랄리스), EHJ52063 (S. 마카카), EID26207 (S. 오랄리스), EID33364, EIG27013 (S. 파라상귀니스), EJF37476, EJO19166 (스트렙토코커스 종 BS35b), EJU16049, EJU32481, YP_006298249, ERF61304, ERK04546, ETJ95568 (S. 아갈락티아), TS89875, ETS90967 (스트렙토코커스 종 SR4), ETS92439, EUB27844 (스트렙토코커스 종 BS21), AFJ08616, EUC82735 (스트렙토코커스 종 CM6), EWC92088, EWC94390, EJP25691, YP_008027038, YP_008868573, AGM26527, AHK22391, AHB36273, Q927P4, G3ECR1 또는 Q99ZW2 (S. 피오제네스),)에 개시된 Cas9 아미노산 서열 중 어느 하나를 포함할 수 있다. 이러한 Cas9 단백질 서열의 임의의 변이체가 사용될 수 있지만, 본 명세서의 RNA 성분과 결합할 경우 DNA에 대한 특이적 결합 활성 및 임의로 엔도뉴클레오리틱 활성을 가져야 한다. 이러한 변이체는 참고 Cas9의 아미노산 서열과 적어도 약 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% 또는 99% 동일한 아미노산 서열을 포함할 수 있다.

[0112] 대안적으로, 본 명세서의 Cas9 단백질은, 예를 들어, 본 명세서에 참고로 포함되는 미국 특허 출원 공개 제2010/0093617호에 개시된 바와 같은 서열 번호:462 (S. 써모필러스), 474 (S. 써모필러스), 489 (S. 아갈락티아), 494 (S. 아갈락티아), 499 (S. 뮤탄스), 505 (S. 피오제네스) 또는 518 (S. 피오제네스) 중 어느 하나에 의해 암호화될 수 있다. 또한 대안적으로, 본 명세서의 Cas9 단백질은, 예를 들어 서열 번호:11 또는 서열 번호:11의 잔기 1 내지 1368의 아미노산 서열을 포함할 수 있다. 또한 대안적으로, Cas9 단백질은, 예를 들어, 전술한 아미노산 서열 중 어느 하나와 적어도 약 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% 또는 99% 동일한 아미노산 서열을 포함할 수 있다. 이러한 변이체인 Cas9 단백질은 본 명세서의 RNA 성분과 결합할 경우 DNA에 대한 특이적 결합 활성 및 임의로 절단 또는 닉킹 활성을 가져야 한다.

[0113] 본 명세서에서 사용된 Cas 단백질 (예를 들어, Cas9)의 기원은 RNA 성분(들)이 유래된 것과 동일한 종 또는 상이한 종으로부터 유래될 수 있다. 예를 들어, 스트렙토코커스 종 (예를 들어, S. 피오제네스 또는 S. 써모필러스)으로부터 유래된 Cas9 단백질을 포함하는 RGEN은 동일한 스트렙토코커스 종으로부터 유래된 서열 (예를 들어, crRNA 반복 서열, tracrRNA 서열)을 갖는 적어도 하나의 RNA 성분과 복합체를 형성할 수 있다. 대안적으로, 본 명세서에서 사용된 Cas 단백질 (예를 들어, Cas9)의 기원은 RNA 성분(들)이 유래된 종과 상이한 종으로부터 유래될 수 있다 (Cas 단백질 및 RNA 성분(들)은 서로 이종성일 수 있음); 이러한 이종 Cas/RNA 성분 RGEN은 DNA 표적화 활성을 가져야 한다.

[0114] 특정 표적 DNA 서열에 대한 본 명세서의 Cas 단백질의 결합 활성 및/또는 엔도뉴클레오리틱 활성을 결정하는 것은 본 명세서에 참고로 개시되는 미국 특허 제8697359호에 개시된 바와 같은 본 기술 분야에 알려진 임의의 적합한 검정에 의해 평가될 수 있다. 예를 들어, 통상적이지 않은 효모에서 Cas 단백질과 적합한 RNA 성분을 발현한 후, 삽입-결실의 존재가 예측되는 DNA 표적 부위를 조사함으로써 측정을 행할 수 있다 (이러한 특정 검정의 Cas 단백질은 완전한 엔도뉴클레오리틱 활성 [이중 가닥 절단 활성]을 가질 것이다). 예측되는 표적 부위에서 삽입-결실의 존재를 조사하는 것은, 예를 들어, DNA 서열 분석법을 통해 또는 표적 서열의 기능 상실을 분석하여 삽입-결실 형성을 추정함으로써 수행될 수 있다. 또 다른 예에서, Cas 단백질 활성은 표적 부위 또는 그 부근의 서열에 상동성인 서열을 포함하는 공여 DNA를 제공하는 통상적이지 않은 효모에서 Cas 단백질과 적합한 RNA 성분을 발현시킴으로써 측정될 수 있다. 표적 부위에서의 공여 DNA 서열의 존재는 (공여자와 표적 서열 사이의 성공적인 HR에 의해 예측되는 바와 같은) 표적화가 일어났음을 나타낼 것이다.

[0115] 본 명세서의 Cas 단백질, 예컨대 Cas9는 전형적으로 이종 핵 국재화 서열 (NLS)을 추가로 포함한다. 본 명세서에서, 이종 NLS 아미노산 서열은, 예를 들어, 본 명세서의 효모 세포의 핵에 검출가능한 양으로 Cas 단백질의 축적을 유도하기 충분한 강도일 수 있다. NLS는 염기성의 양으로 하전된 잔기 (예를 들어, 라이신 및/또는 아르기닌)의 하나 (1부분(monopartite)) 또는 그 이상 (예를 들어, 2부분(bipartite))의 짧은 서열 (예를 들어, 2 내지 20개의 잔기)을 포함할 수 있으며, Cas 아미노산 서열 중 어디에도 위치할 수 있지만 단백질 표면 상에 노출되어야 한다. 본 명세서에서, NLS는, 예를 들어, Cas 단백질의 N-말단 또는 C-말단에 작동가능하게 연결될 수 있다. 둘 이상의 NLS 서열은 Cas 단백질 연결될 수 있는데, 예를 들어 Cas 단백질의 N- 및 C-말단 모두에 연결될 수 있다. 본 명세서에서 적합한 NLS 서열의 비제한적인 예에는 모두 본 명세서에 참고로 포함되는 미국 특허 제6660830호 및 제7309576호 (예를 들어, 그 안의 표 1)에 개시된 것들이 포함된다. 본 명세서에서 유용한 NLS의 또 다른 예에는 서열 번호:11의 아미노산 잔기 1373-1379가 포함된다.

[0116] 소정 실시 형태에서, Cas 단백질에 의한 DNA-특이적 표적화를 유도하는 Cas 단백질 및 이의 각각의 RNA 성분 (예를 들어, crRNA)은 개시된 통상적이지 않은 효모에 이종성이다. 이러한 RGEN 성분의 이종성은 Cas 단백질 및 이들의 각각의 RNA 성분이 알려진 원핵생물 (박테리아 및 고세균)에만 존재한다는 사실에 기인한다.

[0117] 본 명세서에서, Cas 단백질은 효모 세포에서의 발현을 위해 코돈-최적화된 오픈 리딩 프레임 (ORF)을 사용하여 통상적이지 않은 효모 세포에서 임의로 발현될 수 있다. 본 명세서에서, "코돈-최적화된" 서열은 이의 코돈 사용 빈도가 숙주 세포의 바람직한 코돈 사용 빈도를 모방하도록 설계된 유전자이다. 야로위아 리폴리티카가 통상적이지 않은 효모 세포인 측면에서, ORF의 코돈 최적화는 본 명세서에 참고로 포함되는 미국 특허 제7125672호에 제공된 바와 같은 야로위아 리폴리티카 코돈 사용 프로파일에 따라 수행될 수 있다.

[0118] 일부 실시 형태에서, Cas 단백질은 하나 이상의 이종 단백질 도메인 (예를 들어, Cas 단백질에 더하여 1, 2, 3 또는 그 이상의 도메인)을 포함하는 융합 단백질의 일부이다. 이러한 융합 단백질은 임의의 추가의 단백질 서열 및 임의로 임의의 2개의 도메인 사이, 예컨대 Cas와 제1 이종 도메인 사이의 링커 서열을 포함할 수 있다. 본 명세서에서 Cas 단백질에 융합될 수 있는 단백질 도메인의 예에는 에피토프 태그 (예를 들어, 히스티딘 [His], V5, FLAG, 인플루엔자 혈구응집소 [HA], myc, VSV-G, 티오레독신 [Trx]), 리포터 (예를 들어, 글루타티온-5-트랜스퍼라제 [GST], 고추냉이 퍼옥시다제 [HRP], 클로람페니콜 아세틸트랜스퍼라제 [CAT], 베타-갈락토시다제, 베타-글루쿠로니다제 [GUS], 루시퍼라제, 녹색 형광 단백질 [GFP], HcRed, DsRed, 시안 형광 단백질 [CFP], 황색 형광 단백질 [YFP], 청색 형광 단백질 [BFP]) 및: 메틸라제 활성, 데메틸라제 활성, 전사 활성화 활성 (예를 들어, VP16 또는 VP64), 전사 억제 활성, 전사 방출 인자 활성, 히스톤 변형 활성, RNA 절단 활성 및 핵산 결합 활성 중 하나 이상의 활성을 갖는 도메인을 제한 없이 포함된다. 다른 실시 형태에서, Cas 단백질은 DNA 분자 또는 다른 분자, 예컨대 말토스 결합 단백질 (MBP), S-태그, Lex A DNA 결합 도메인 (DBD), GAL4A DNA 결합 도메인 및 단순 헤르페스 바이러스 (HSV) VP16에 결합하는 단백질과 융합될 수 있다. 본 명세서에서, Cas 단백질을 포함하는 융합 단백질의 일부가 될 수 있는 추가의 도메인은 본 명세서에 참고로 포함되는 미국 특허 출원 공개 제2011/0059502호에 개시되어 있다. Cas 단백질이 이종 단백질 (예를 들어, 전사 인자)에 융합되는 소정 실시 형태에서, Cas 단백질은 (본 명세서의 적합한 RNA 성분과 복합체인 경우) DNA 인식 및 결합 활성을 가지나, DNA 닉킹 또는 절단 활성은 없다.

[0119] 본 명세서에서, RGEN은 DNA 표적 서열에서 DNA 가닥에 결합할 수 있고, 임의로 절단할 수 있다. 소정 실시 형태에서, RGEN은 DNA 표적 서열의 하나 또는 두 가닥을 절단할 수 있다. RGEN은, 예를 들어, DNA 표적 서열의 두 가닥을 절단할 수 있다.

[0120] 본 명세서에서, DNA 표적 서열의 두 가닥을 절단할 수 있는 RGEN은 전형적으로 이의 엔도뉴클레아제 도메인 모

두를 기능적 상태(functional state)로 포함하는 Cas 단백질을 포함한다 (예를 들어, 야생형 엔도뉴클레아제 도메인 또는 각각의 엔도뉴클레아제 도메인에서 일부 또는 모든 활성을 보유하는 이들의 변이체). 따라서, 야생형 Cas 단백질 (예를 들어, 본 명세서에 개시된 Cas9 단백질) 또는 Cas 단백질 각각의 엔도뉴클레아제 도메인에서 일부 또는 모든 활성을 보유하는 이들의 변이체가 DNA 표적 서열의 두 가닥을 절단할 수 있는 RGEN의 적합한 예이다. 기능적 RuvC 및 HNH 뉴클레아제 도메인을 포함하는 Cas9 단백질은 DNA 표적 서열의 두 가닥을 절단할 수 있는 Cas 단백질의 예이다. 본 명세서에서 DNA 표적 서열의 두 가닥을 절단할 수 있는 RGEN은 전형적으로 절단 부위에 평활 말단 (즉, 뉴클레오티드 돌출부가 없음)을 형성하도록 동일한 위치에서 두 가닥을 절단한다.

[0121] 본 명세서에서 DNA 표적 서열의 한 가닥을 절단할 수 있는 RGEN은 본 명세서에서 닉카아제 활성 (예를 들어, 부분 절단 능력)을 갖는 것으로 특징지어질 수 있다. 본 명세서에서, Cas 닉카아제 (예를 들어, Cas9 닉카아제)는 전형적으로 Cas가 DNA 표적 서열 중 한 가닥만 절단하도록 하는 (즉, 닉을 형성) 하나의 기능적 엔도뉴클레아제 도메인을 포함한다. 예를 들어, Cas9 닉카아제는 (i) 기능장애 돌연변이인 RuvC 도메인 및 (ii) 기능적 HNH 도메인 (예를 들어, 야생형 HNH 도메인)을 포함할 수 있다. 또 다른 예로서, Cas9 닉카아제는 (i) 기능적 RuvC 도메인 (예를 들어, 야생형 RuvC 도메인) 및 (ii) 기능장애 돌연변이인 HNH 도메인을 포함할 수 있다.

[0122] 본 명세서에 사용하기에 적합한 Cas9 닉카아제의 비제한적인 예가 본 명세서에 참고로 포함되는 문헌[Gasiunas et aL. (*ProC. NatL. AcaD. Sci. U.S.A.* 109:E2579-E2586)], 문헌[Jinek et aL. (*Science* 337:816-821)], 문헌[Sapranauskas et aL. (*Nucleic Acids Res.* 39:9275-9282)] 및 미국 특허 출원 공개 제2014/0189896호에 의해 개시되어 있다. 예를 들어, 본 명세서에서 Cas9 닉카아제는 Asp-31 치환 (예를 들어, Asp-31-Ala) (돌연변이 RuvC 도메인의 예) 또는 His-865 치환 (예를 들어, His-865-Ala), Asn-882 치환 (예를 들어, Asn-882-Ala) 또는 Asn-891 치환 (예를 들어, Asn-891-Ala) (HNH 도메인 돌연변이의 예)을 갖는 S. 써모필러스 Cas9를 포함할 수 있다. 또한 예를 들어, 본 명세서에서 Cas9 닉카아제는 Asp-10 치환 (예를 들어, Asp-10-Ala), Glu-762 치환 (예를 들어, Glu-762-Ala) 또는 Asp-986 치환 (예를 들어, Asp-986-Ala) (돌연변이 RuvC 도메인의 예) 또는 His-840 치환 (예를 들어, His-840-Ala), Asn-854 치환 (예를 들어, Asn-854-Ala) 또는 Asn-863 치환 (예를 들어, Asn-863-Ala) (돌연변이 HNH 도메인의 예)을 갖는 S. 피오제네스 Cas9를 포함할 수 있다. S. 피오제네스 Cas9에 관하여, 3개의 RuvC 서브도메인은 일반적으로 각각 아미노산 잔기 1-59, 718-769 및 909-1098에 위치하고, HNH 도메인은 아미노산 잔기 775-908에 위치한다 (문헌[Nishimasu et al., *Cell* 156:935-949]).

[0123] 본 명세서에서, Cas9 닉카아제는 개시된 발명의 통상적이지 않은 효모에서 다양한 용도로 사용될 수 있다. 예를 들어, Cas9 닉카아제는 적합한 공여 폴리뉴클레오티드를 갖는 DNA 표적 부위 서열에 또는 그 부근에서 HR을 촉진하는 데 사용될 수 있다. 닉이 있는 DNA는 NHEJ 공정을 위한 기질이 아니고, HR 공정에 의해 인식되기 때문에, 특정 표적 부위에서 DNA에 닉을 형성하는 것은 적합한 공여 폴리뉴클레오티드를 사용하는 HR을 더 잘 수용할 수 있게 해야 한다.

[0124] 또 다른 예로서, DNA 표적화의 특이성을 높이기 위해 한 쌍의 Cas9 닉카아제가 사용될 수 있다. 일반적으로, 이는 상이한 가이드 서열을 갖는 RNA 성분과 관련되어 있기 때문에, 원하는 표적화를 위한 영역에서 반대 가닥 상의 가까운 DNA 서열을 표적화하고 닉을 형성하는 2개의 Cas9 닉카아제를 제공함으로써 수행될 수 있다. 각각의 DNA 가닥의 이러한 가까운 절단은 DSB (즉, 단일-가닥 돌출부를 갖는 DSB)을 생성하고, 이는 이어서 NHEJ (삽입-결실 형성을 유도함) 또는 HR (제공된다면, 적합한 공여 폴리뉴클레오티드와의 재조합을 유도함)의 기질로 인식된다. 이러한 실시 형태에서 각각의 닉은, 예를 들어, 서로 적어도 약 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90 또는 100 (또는 5 내지 100 사이의 임의의 정수)개의 염기만큼 이격될 수 있다. 본 명세서에서, 상기한 바와 같이 하나 또는 2개의 Cas9 닉카아제 단백질이 Cas9 닉카아제 쌍에서 사용될 수 있다. 예를 들어, 돌연변이 RuvC 도메인을 가지나 기능하는 HNH 도메인 (즉, Cas9 $HNH^{+}/RuvC^{-}$)을 갖는 Cas9 닉카아제 (예를 들어, S. 피오제네스 Cas9 $HNH^{+}/RuvC^{-}$)를 사용할 수 있다. 각각의 닉카아제를 각각의 특정 DNA 부위로 표적화하는 가이드 RNA 서열을 갖는 본 명세서의 적합한 RNA 성분을 사용하여, 서로 가까운 (100개의 염기쌍 이하로 이격된) 특정 DNA 부위로 각각의 Cas9 닉카아제 (예를 들어, Cas9 $HNH^{+}/RuvC^{-}$)를 유도할 것이다.

[0125] 소정 실시 형태에서 RGEN은 DNA 표적 부위는 서열에 결합할 수 있지만, 표적 부위 서열에서 임의의 가닥을 절단하지 않는다. 이러한 RGEN은 이의 모든 뉴클레아제 도메인이 기능장애 돌연변이인 Cas 단백질을 포함할 수 있다. 예를 들어, 본 명세서에서 DNA 표적 부위 서열에 결합할 수 있지만, 표적 부위 서열에서 임의의 가닥을 절단하지 않는 Cas9 단백질은 기능장애 돌연변이인 RuvC 도메인 및 기능장애 돌연변이인 HNH 도메인을 포함할 수 있다. 이러한 Cas9 단백질의 비제한적인 예에는 상기 개시된 RuvC 및 HNH 뉴클레아제 도메인 돌연변이 (예를 들어, Asp-10 치환, 예컨대 Asp-10-Ala 및 His-840 치환, 예컨대 His-840-Ala을 갖는 S. 피오제네스 Cas9) 중

어느 하나가 포함된다. 본 명세서에서, 표적 DNA 서열에 결합하지만 절단하지 않는 Cas 단백질은 유전자 발현을 조절하는 데 사용될 수 있으며, 예를 들어, 이 경우 Cas 단백질은 전사 인자 (또는 이의 일부) (예를 들어, 억제 인자 또는 활성 인자, 예컨대 본 명세서에 개시된 것 중 어느 하나)와 융합될 수 있다. 예를 들어, Asp-10 치환 (예를 들어, Asp-10-Ala) 및 His-840 치환 (예를 들어, His-840-Ala)을 갖는 S. 피오제네스 Cas9를 포함하는 Cas9는 VP16 또는 VP64 전사 활성 인자 도메인에 융합될 수 있다. 이러한 RGEN의 RNA 성분에서 사용되는 가이드 서열은, 예를 들어 유전자 프로모터 또는 다른 조절 요소 (예를 들어, 인트론) 내의 DNA 서열에 상보적일 것이다.

[0126] 소정 측면의 효모는 (i) DNA 표적 서열의 DNA 가닥 중 하나 또는 둘 모두를 절단할 수 있는 RGEN 및 (ii) DNA 표적 부위 또는 그 부근의 서열에 상동성인 적어도 하나의 서열을 포함하는 공여 폴리뉴클레오티드 (본 명세서의 Cas 단백질에 의해 특이적으로 표적화된 서열)을 포함할 수 있다. 적합한 공여 폴리뉴클레오티드는 표적 부위가 (예를 들어, 본 명세서의 Cas 단백질을 사용하여 도입될 수 있는) SSB 또는 DSB를 함유하는 경우, DNA 표적 부위 또는 그 부근의 서열과 HR을 수행할 수 있다. 본 명세서의 공여 폴리뉴클레오티드 내의 "상동성 서열"은, 예를 들어, 서열 표적 부위 또는 그 부근의 서열과 100% 동일성 또는 서열 표적 부위 또는 그 부근의 서열과 적어도 약 95%, 96%, 97%, 98% 또는 99% 동일성을 갖는 적어도 약 25, 50, 75, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 또는 10000개의 뉴클레오티드 또는 약 50 내지 500, 50 내지 550, 50 내지 600, 50 내지 650 또는 50 내지 700개의 뉴클레오티드의 서열을 포함할 수 있거나 이로 이루어질 수 있다.

[0127] 본 명세서의 공여 폴리뉴클레오티드는, 예를 들어, 표적 부위 서열 또는 그 부근의 서열에 이종성인 서열에 의해 이격된 2개의 상동성 서열 (상동성 암)을 가질 수 있다. 이러한 공여 폴리뉴클레오티드와 표적 부위 서열 사이의 HR은 전형적으로 표적 부위에서 서열을 공여 폴리뉴클레오티드의 이종 서열로 대체시킨다 (공여 폴리뉴클레오티드의 상동성 암에 상동성인 표적 부위 서열 사이에 위치한 표적 부위 서열은 공여 폴리뉴클레오티드의 이종 서열에 의해 대체된다). 2개의 상동성 암을 갖는 공여 폴리뉴클레오티드에서, 암은, 예를 들어, 적어도 약 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 75, 100, 250, 500, 1000, 2500, 5000, 10000, 15000, 20000, 25000 또는 30000개의 뉴클레오티드 만큼 이격될 수 있다 (즉, 공여 폴리뉴클레오티드의 이종 서열은 길이가 적어도 약 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 75, 100, 250, 500, 1000, 2500, 5000, 10000, 15000, 20000, 25000 또는 30000개 뉴클레오티드이다). 각각의 상동성 암의 길이 (예를 들어, 상동성 서열에 대해 상기 개시된 임의의 길이)는 동일하거나 상이할 수 있다. 각각의 표적 부위 또는 그 부근의 상동성 서열을 갖는 각각의 암의 % 동일성 (예를 들어, 상동성 서열에 대해 상기 개시된 임의의 % 동일성)은 동일하거나 상이할 수 있다.

[0128] 공여 폴리뉴클레오티드의 상응하는 상동성 서열에 상동성인 표적 부위 서열 또는 그 부근 (대안적으로, 인근 또는 이의 근접에 있는)의 DNA 서열은, 예를 들어, 표적 서열 내의 예측되는 Cas 단백질 절단 부위 (DSB 또는 닉)로부터 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 450, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000 또는 60000 (또는 1 내지 60000 사이의 임의의 정수)개의 뉴클레오티드 (예를 들어, 약 1 내지 1000, 100 내지 1000, 500 내지 1000, 1 내지 500 또는 100 내지 500 뉴클레오티드) 내에 있을 수 있다. 이러한 뉴클레오티드 거리는 절단 부위로부터 상동성 서열의 첫번째 뉴클레오티드까지로 표시될 수 있으며, 절단 부위로부터 상류 또는 하류 방향으로 진행될 수 있다. 예를 들어, 공여 폴리뉴클레오티드의 상응하는 서열에 상동성인 표적 서열 근처의 서열은 표적 서열에서 예측되는 Cas 단백질 절단 부위의 500개의 뉴클레오티드 염기쌍 하류에서 시작할 수 있다. 2개의 상동성 암 (예를 들어, 이종 서열에 의해 이격된 제1 및 제2 상동성 암)을 갖는 공여 폴리뉴클레오티드를 사용하는 본 명세서의 실시 형태에서, 예를 들어, (공여자의 제1 상동성 암과 상동성에 해당하는) 상동성 서열은 예측되는 Cas 절단 부위의 상류에 있을 수 있고 (공여자의 제2 상동성 암과 상동성에 해당하는) 상동성 서열은 예측되는 Cas 절단 부위의 상류에 있을 수 있다. 예측되는 절단 부위로부터 이러한 상류 및 하류 상동성 서열 각각의 뉴클레오티드 거리는 동일하거나 상이할 수 있으며, 예를 들어, 상기 개시된 뉴클레오티드 거리 중 어느 하나일 수 있다. 예를 들어, (공여자의 제1 상동성 암과 상동성에 해당하는) 상동성 서열의 3' 말단은 예측되는 Cas 절단 부위의 600개의 뉴클레오티드 염기쌍 상류에 위치할 수 있고, (공여자의 제2 상동성 암과 상동성에 해당하는) 상동성 서열의 5' 말단은 예측되는 Cas 절단 부위의 400개의 뉴클레오티드 염기쌍 하류에 위치할 수 있다.

[0129] 본 명세서에서, RGEN은 통상적이지 않은 효모의 게놈 내의 염색체, 에피솜 또는 임의의 다른 DNA 분자의 표적 부위 서열에서 DNA 가닥에 결합하고, 임의로 절단할 수 있다. RGEN의 RNA 성분이 표적 서열의 가닥에 상보적인 서열 (가이드 서열)을 포함하는 경우, 이러한 표적 서열의 인식 및 결합은 특이적이다. 소정 실시 형태에서,

표적 부위는 유일할 수 있다 (즉, 대상의 게놈에서 표적 부위 서열은 한번만 발생한다).

[0130] 본 명세서에서, 표적 서열의 길이는, 예를 들어, 적어도 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 또는 30개의 뉴클레오티드; 13 내지 30개의 뉴클레오티드; 17 내지 25개의 뉴클레오티드; 또는 17 내지 20개의 뉴클레오티드일 수 있다. 이러한 길이는 PAM 서열을 포함하거나 제외할 수 있다. 또한, 본 명세서에서 표적 서열의 가닥은 가이드 서열과 혼성화되고, (적합한 PAM이 표적 서열에 인접한 경우, 아래 참조) 표적 서열에 Cas 단백질 또는 Cas 단백질 복합체를 직접 서열-특이적으로 결합시키기 위한 (crRNA 또는 gRNA의) 가이드 서열과의 충분한 상보성을 갖는다. 가이드 서열과 이의 상응하는 DNA 표적 서열의 가닥 사이의 상보성 정도는, 예를 들어, 적어도 약 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100%이다. 본 명세서에서, 표적 부위는, 예를 들어, 유전자 산물 (예를 들어, 단백질 또는 RNA)을 암호화하는 서열 또는 비암호화 서열 (예를 들어, 조절 서열 또는 "정크(junk)" 서열) 내에 위치할 수 있다.

[0131] PAM (프로토스페이서-인접 모티프) 서열은 표적 부위 서열에 인접할 수 있다. PAM 서열은 본 명세서의 RGEN에 의해 인식되는 짧은 DNA 서열이다. 관련된 PAM 및 DNA 표적 서열의 처음 11개의 뉴클레오티드는 Cas9/gRNA 표적화 및 절단에 중요할 수 있다 (문헌[Jiang et al., NaT. Biotech. 31:233-239]). 본 명세서에서, PAM 서열의 길이는 사용되는 Cas 단백질 또는 Cas 단백질 복합체에 따라 다를 수 있지만, 전형적으로는, 예를 들어 2, 3, 4, 5, 6, 7 또는 8개의 뉴클레오티드 길이이다. PAM 서열은, 예를 들어, 결과적으로 RNA 성분 가이드 서열에 상보적인 표적 부위의 가닥에 상보적인 표적 부위 서열로부터 바로 하류에, 또는 이의 하류에 있는 2 또는 3개의 뉴클레오티드 내에 있다. RGEN이 RNA 성분과 복합체를 형성한 핵산 내부 분해 방식으로 (endonucleolytically) 활성인 Cas9 단백질인 본 명세서의 실시 형태에서, Cas9는 RNA 성분에 의해 유도된 바와 같이 표적 서열에 결합하고, PAM 서열의 상류 세 번째 뉴클레오티드 위치의 5' 부근에서 두 가닥을 절단한다. 표적 부위:PAM 서열의 다음의 예를 고려하라:

[0132] 5'-NNNNNNNNNNNNNNNNNNNNNNNXGG-3' (서열 번호:46).

[0133] 이러한 예의 서열에서, N은 A, C, T 또는 G일 수 있고, X는 A, C, T 또는 G일 수 있다 (X는 또한 N_{PAM}로도 지칭됨). 이러한 예에서, PAM 서열은 (밑줄친) XGG이다. 적합한 Cas9/RNA 성분 복합체는 이러한 표적을 이중 밑줄로된 N의 5' 부근에서 절단할 것이다. 서열 번호:46에서 N으로 이루어진 열(string of N's)은 본 명세서의 RNA 성분의 가이드 서열과, 예를 들어, 적어도 약 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 동일한 표적 서열을 나타낸다 (여기서, DNA 표적 서열의 임의의 T는 RNA 가이드 서열의 임의의 U와 정렬될 것임). (본 명세서에서 표적 부위를 나타내는) 이러한 표적 서열을 인식하고 결합하는데 있어서, Cas9 복합체의 RNA 성분의 가이드 서열은, N으로 이루어진 열의 보체 서열과 어닐링될 것이며; 가이드 서열과 표적 부위 보체 사이의 % 상보성은, 예를 들어, 적어도 약 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100%이다. Cas9 닉카아제가 게놈에서 서열 번호:46을 표적으로 하는데 사용되는 경우, 닉카아제는 닉카아제의 어느 엔도뉴클레아제 도메인가 기능장애인지에 따라 이중 밑줄로된 N의 5' 부근에서 또는 상보성 가닥의 동일한 위치에서 닉을 형성할 것이다. (RuvC와 HNH 도메인이 모두 기능장애) 핵산 분해 활성이 없는 Cas9가 게놈에서 서열 번호:46을 표적으로 하는데 사용되는 경우, 표적 서열을 인식하고 결합하지만, 어떤 서열도 절단하지 않는다.

[0134] 본 명세서에서, PAM은 전형적으로 사용되는 RGEN의 타입을 고려하여 선택된다. 본 명세서에서, PAM 서열은, 예를 들어, Cas가 유래될 수 있는 본 명세서에 개시된 임의의 종으로부터 유래된 Cas, 예컨대 Cas9를 포함하는 RGEN에 의해 인식되는 것일 수 있다. 소정 실시 형태에서, PAM 서열은 S. 피오제네스, S. 써모필러스, S. 아갈락티아, N. 메닝기티디스, T. 덴티콜라 또는 F. 노비시다로부터 유래된 Cas9를 포함하는 RGEN에 의해 인식되는 것일 수 있다. 예를 들어, S. 피오제네스로부터 유래된 적합한 Cas9는 NGG의 PAM 서열 (서열 번호:47; N은 A, C, T 또는 G일 수 있음)을 갖는 표적 게놈 서열을 표적화하는 데 사용될 수 있다. 다른 예로서, 적합한 Cas9는 다음과 같은 PAM 서열을 갖는 DNA 서열을 표적화하는 경우 하기의 종 중 어느 하나로부터 유래될 수 있다: S. 써모필러스 (NNAGAA [서열 번호:48]), S. 아갈락티아 (NGG [서열 번호:47]), NNAGAAW [서열 번호:49, W는 A 또는 T임], NGGNG [서열 번호:50]), N. 메닝기티디스 (NNNNGATT [서열 번호:51]), T. 덴티콜라 (NAAAAC [서열 번호:52]) 또는 F. 노비시다 (NG [서열 번호:53]) (여기서, 이러한 모든 특정 PAM 서열에서 N으로 이루어진 서열은 A, C, T 또는 G임). 본 명세서에서 유용한 Cas9/PAM의 다른 예에는, 본 명세서에 참고로 포함되는 문헌[Shah et aL. (*RNA Biology* 10:891-899)] 및 문헌[Esvelt et aL. (*Nature Methods* 10:1116-1121)]에 개시된 것들이 포함된다. 본 명세서에서, 표적 서열의 예는 서열 번호:46을 따르지만, 'XGG' PAM은 전술한 PAM 중 어느 하나에 의해 대체된다.

[0135] 본 명세서의 실시 형태에서, 5'-캡을 갖지 않는 적어도 하나의 RNA 성분은 RGEN에 포함된다. 이러한 캡핑되지 않은 RNA 성분은 통상적이지 않은 효모 내의 염색체 또는 에피솜의 표적 부위 서열에 상보적인 서열을 포함한다. RGEN은 이러한 서열 상보성을 기초하여 표적 부위에서 DNA 가닥에 특이적으로 결합하고, 임의로 절단한다. 따라서, 개시된 발명의 실시 형태에서 RNA 성분의 상보적인 서열은 가이드 서열 또는 가변 표적화 도메인로도 지칭될 수 있다.

[0136] 본 명세서에서, RNA 성분의 가이드 서열 (예를 들어, crRNA 또는 gRNA)은 길이가, 예를 들어, 적어도 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 또는 30개의 리보뉴클레오티드; 13 내지 30개의 리보뉴클레오티드; 17 내지 25개의 리보뉴클레오티드; 또는 17 내지 20개의 리보뉴클레오티드일 수 있다. 일반적으로, 본 명세서에서 가이드 서열은 표적 서열과 혼성화되고, (적합한 PAM이 표적 서열에 인접한 경우) 표적 서열에 Cas 단백질 또는 Cas 단백질 복합체를 직접 서열-특이적으로 결합시키기 위한 (crRNA 또는 gRNA의) 가이드 서열과의 충분한 상보성을 갖는다. 가이드 서열과 이의 상응하는 DNA 표적 서열 사이의 상보성 정도는, 예를 들어, 적어도 약 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100%이다. 가이드 서열은 효모 세포의 DNA 표적 서열로 RGEN을 표적화하도록 적절하게 유전자 조작될 수 있다.

[0137] 본 명세서에서, RNA 성분은, 예를 들어, 가이드 서열을 포함하는 crRNA 및 반복 (tracrRNA 메이트) 서열을 포함할 수 있다. 가이드 서열은 전형적으로 crRNA의 5' 말단에 또는 그 부근 (1, 2, 3, 4, 5, 6, 7, 8, 9, 10 또는 그 이상의 염기 내에)에 위치한다. crRNA의 가이드 서열의 하류는 tracrRNA의 5' 말단에서 서열에 상보적이고 이와 혼성화될 수 있는 "반복" 또는 "tracrRNA 메이트" 서열이다. 가이드 및 tracrRNA 메이트 서열은, 예를 들어, 바로 인접하거나, 1, 2, 3, 4 또는 그 이상의 염기에 의해 이격될 수 있다. tracrRNA 메이트 서열은 tracrRNA의 5' 말단에, 예를 들어, 적어도 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98% 또는 99% 서열 상보성을 갖는다. 일반적으로, 상보성 정도는 tracrRNA 메이트 서열과 tracrRNA 서열의 최적 정렬을 참고하여, 두 서열 중 더 짧은 길이에 따를 수 있다. 본 명세서에서, tracrRNA 메이트 서열의 길이는, 예를 들어 길이가 적어도 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 또는 18개의 리보뉴클레오티드일 수 있고, tracrRNA의 5' 말단에서 동일하거나 유사한 길이 (예를 들어, +/- 1, 2, 3, 4 또는 5개의 염기)의 서열과 혼성화된다. 본 명세서에서, tracrRNA 메이트 서열의 적합한 예에는 서열 번호:54 (guuuuuguacucucaagauuua), 서열 번호:55 (guuuuuguacucuca), 서열 번호:56 (guuuuagagcua, 실시예 참조) 또는 서열 번호:57 (guuuuagagcuag) 또는 (i) 적어도 약 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% 또는 99%의 서열 동일성을 갖고 (ii) tracrRNA의 5'-말단 서열과 어닐링할 수 있는 이들의 변이체가 포함된다. 본 명세서에서, crRNA의 길이는, 예를 들어, 적어도 약 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46 또는 48개의 리보뉴클레오티드; 또는 약 18 내지 48개의 리보뉴클레오티드; 또는 약 25 내지 50개의 리보뉴클레오티드일 수 있다.

[0138] tracrRNA는 타입 II 크리스퍼 시스템의 Cas9 단백질이 RGEN에 포함되는 실시 형태에서 crRNA와 함께 포함되어야 한다. 본 명세서에서, tracrRNA는 5'에서 3' 방향으로 (i) crRNA의 반복 영역 (tracrRNA 메이트 서열)과 어닐링하는 서열 및 (ii) 스템 루프-함유 부분을 포함한다. (i)의 서열의 길이는, 예를 들어, 상기 개시된 임의의 tracrRNA 메이트 서열의 길이와 동일하거나 유사할 수 있다 (예를 들어, +/- 1, 2, 3, 4 또는 5개의 염기). 본 명세서에서, tracrRNA의 총 길이 (즉, 서열 성분 [i] 및 [ii])는, 예를 들어, 적어도 약 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85 또는 90개의 (또는 30 내지 90 사이의 임의의 정수) 리보뉴클레오티드일 수 있다. tracrRNA는 3'-말단에 1, 2, 3, 4, 5 또는 그 이상의 우라실 잔기를 추가로 포함할 수 있으며, 이는 전사 종결 서열을 갖는 tracrRNA를 발현함으로써 존재할 수 있다.

[0139] 본 명세서에서, tracrRNA는, 예를 들어, Cas9 서열이 유래될 수 있는, 상기 열거된 임의의 박테리아 종으로부터 유래될 수 있다. 적합한 tracrRNA 서열의 예에는 본 명세서에 참고로 포함되는 미국 특허 제8697359호 및 문헌 [Chylinski et aL. (*RNA Biology* 10:726-737)]에 개시된 것들이 포함된다. 본 명세서에서 바람직한 tracrRNA는 스트렙토코커스 종 tracrRNA (예를 들어, S. 피오제네스, S. 써모필러스)로부터 유래될 수 있다. 본 명세서의 tracrRNA의 다른 적합한 예는 하기를 포함할 수 있다:

[0140] 서열 번호:58:

[0141] uagcaaguuaaaauaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugc (실시예 참조),

[0142] 서열 번호:59:

[0143] uagcaaguuaaaauaaggcuaguccguuaucaacuugaaaaagug 또는

[0144] 서열 번호:60:

[0145] uagcaaguuaaaauaaggcuaguccguuauca,

[0146] (S. 피오제네스 tracrRNA로부터 유래됨). 본 명세서의 tracrRNA의 다른 적합한 예는 하기를 포함할 수 있다:

[0147] 서열 번호:61:

[0148] uaaaucuugcagaagcuacaaagauaaggcuucaugccgaaaucaacacccugucauuuuauggcaggguguuuucguuauuuaa, 서열 번호:62:

[0149] ugcagaagcuacaaagauaaggcuucaugccgaaaucaacacccugucauuuuauggcagggugu uuucguuauuua

[0150] 또는

[0151] 서열 번호:63:

[0152] ugcagaagcuacaaagauaaggcuucaugccgaaaucaacacccugucauuuuauggcagggugu, (S. 써모필러스 tracrRNA로부터 유래됨).

[0153] 본 명세서의 tracrRNA의 또 다른 예는 (i) 이와 적어도 약 80%, 85%, 90, 91, 92, 93, 94, 95, 96, 97, 98 또는 99%의 서열 동일성을 갖고 (ii) tracrRNA로서 작용할 수 있는 이들 tracrRNA 서열 번호의 변이체이다 (예를 들어, 5'-말단 서열은 crRNA의 tracrRNA 메이트 서열에 어닐링할 수 있고, 5'-말단 서열의 하류 서열은 하나 이상의 헤어핀을 형성할 수 있으며, 변이체 tracrRNA는 Cas9 단백질과 복합체를 형성할 수 있다).

[0154] 본 명세서에 개시된 RGEN의 RNA 성분은, 예를 들어, tracrRNA에 작동가능하게 연결되거나 융합된 crRNA를 포함하는 가이드 RNA (gRNA)를 포함할 수 있다. 특정 바람직한 실시 형태에서, gRNA의 crRNA 성분은 tracrRNA 성분의 상류에 있다 (즉, 이러한 gRNA는 5'에서 3' 방향으로 tracrRNA에 작동가능하게 연결된 crRNA를 포함한다). 본 명세서에 개시된 바와 같은 (예를 들어, 상기 실시 형태) 임의의 crRNA 및/또는 tracrRNA (및/또는 이들의 부분, 예컨대 crRNA 반복 서열, tracrRNA 메이트 서열 또는 tracrRNA 5'-말단 서열)는, 예를 들어, gRNA에 포함될 수 있다.

[0155] 본 명세서에서, gRNA의 crRNA 성분의 tracrRNA 메이트 서열은 tracrRNA 성분의 5'-말단과 어닐링함으로써, 헤어핀 구조를 형성할 수 있어야 한다. (crRNA 성분의) tracrRNA 메이트 서열과 (tracrRNA 성분의) 5'-말단 서열 사이의 길이 및 % 상보성에 관한 상기 개시 중 임의의 것은, 예를 들어 gRNA의 crRNA 및 tracrRNA 성분을 특징지을 수 있다. 이러한 어닐링을 용이하게 하기 위해, crRNA와 tracrRNA 성분의 작동가능한 결합 또는 융합은 바람직하게는 적합한 루프 형성 리보뉴클레오티드 서열을 포함한다 (즉, 루프 형성 서열은 crRNA와 tracrRNA 성분을 함께 연결하여 gRNA를 형성할 수 있음). RNA 루프 형성 서열의 적합한 예에는 GAAA (서열 번호:43, 실시예 참조), CAAA (서열 번호:44) 및 AAAG (서열 번호:45)가 포함된다. 그러나, 대체 루프 서열이 사용되는 바와 같이 더 길거나 더 짧은 루프 서열이 사용될 수 있다. 루프 서열은 바람직하게는 리보뉴클레오티드 트리플렛 (예를 들어, AAA) 및 트리플렛의 어느 하나의 말단에 추가의 리보뉴클레오티드 (예를 들어, C 또는 G)를 포함한다.

[0156] 본 명세서에서, gRNA는 (crRNA 성분의) 이의 tracrRNA 메이트 서열과 tracrRNA 5'-말단 서열 부분을 어닐링하여 헤어핀 ("제1 헤어핀")을 형성한다. gRNA의 tracrRNA 성분의 서열에 따라, 하나 이상의 (예를 들어, 1, 2, 3 또는 4개) 추가의 헤어핀 구조가 이러한 제1 헤어핀의 하류를 형성할 수 있다. 따라서, gRNA는, 예를 들어 최대 5개의 헤어핀 구조를 가질 수 있다. gRNA는 gRNA 서열의 말단에 이어지는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 또는 그 이상의 잔기를 추가로 포함할 수 있는데, 이는, 예를 들어 전사 종결 서열을 갖는 gRNA를 발현함으로써 존재할 수 있다. 이러한 추가의 잔기는, 예를 들어, 종결 서열의 선택에 따라 모두 U 잔기 또는 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% 또는 99%의 U 잔기일 수 있다.

[0157] 개시된 발명에 유용한 적합한 gRNA의 비제한적인 예는 다음을 포함할 수 있다:

[0158] 서열 번호:64:

[0159] NNNNNNNNNNNNNNNNNNNguuuuuguacucucaagauuuaGAAAuaaaucuugcagaagcuacaaagauaaggcuucaugccgaaaucaacacccugucauuuuauggcaggguguuuucguuauuuaa, 서열 번호:65:

[0160] NNNNNNNNNNNNNNNNNNNguuuuuguacucucaGAAAugcagaagcuacaaagauaaggcuucaugccgaaaucaacacccugucauuuuauggcaggguguuuucguuauuuaa, 서열 번호:66:

[0161] NNNNNNNNNNNNNNNNNNNguuuuuguacucucaGAAAugcagaagcuacaaagauaaggcuucaugccgaaaucaacacccugucauuuuauggcagggugu, 서열 번호:67:

[0162] NNNNNNNNNNNNNNNNNNNguuuuuguacucucaGAAAuagcaaguuaaaauaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugc, 서열 번호:68:

[0163] NNNNNNNNNNNNNNNNNNNguuuuagagcuaGAAAuagcaaguuaaaauaaggcuaguccguuaucaacuugaaaaagug, 서열 번호:69:

[0164] NNNNNNNNNNNNNNNNNNNguuuuagagcuaGAAAuagcaaguuaaaauaaggcuaguccguuauca 또는

[0165] 서열 번호:70:

[0166] NNNNNNNNNNNNNNNNNNNguuuuagagcuaGAAAuagcaaguuaaaauaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcuuuu (실시예 참조).

[0167] 서열 번호:64 내지 70의 각각에서, 단일-밑줄로된 서열은 gRNA의 crRN 부분을 나타낸다. 각각의 "N"은 적합한 가이드 서열의 리보뉴클레오티드 염기 (A, U, G 또는 C)를 나타낸다. 소문자의 첫 번째 블록은 tracrRNA 메이트 서열을 나타낸다. 소문자의 두 번째 블록은 gRNA의 tracrRNA 부분을 나타낸다. 이중 밑줄로된 서열은 tracrRNA 메이트 서열과 어닐링하는 tracrRNA 서열 부분에 근접하여 제1 헤어핀을 형성한다. 루프 서열 (GAAA, 서열 번호:43)은 대문자로 표시되며, 이는 각각의 gRNA의 crRNA와 tracrRNA 부분을 작동가능하게 연결한다. 본 명세서의 gRNA의 다른 예는 (i) 이러한 서열과 적어도 약 80%, 85%, 90, 91, 92, 93, 94, 95, 96, 97, 98 또는 99%의 서열 동일성 (이 계산에서 가이드 서열은 제외함)을 갖고 (ii) Cas9 단백질을 특이적으로 표적화하는 gRNA로서 작용하여 표적 DNA 서열에 결합하고, 임의로 닉을 형성하거나 절단할 수 있는 전술한 gRNA의 변이체를 포함한다.

[0168] 본 명세서에서, gRNA는 또한 가이드 서열 (VT 도메인)에 이어서 Cas 엔도뉴클레아제 인식 (CER) 도메인을 갖는 것으로 특징지어질 수 있다. CER 도메인은 tracrRNA 메이트 서열 다음에 tracrRNA 서열을 포함한다. 본 명세서에서 유용한 CER 도메인의 예는 상기 서열 번호:64 내지 70에 포함되는 것들을 포함한다 (각각의 CER 도메인은 VT 도메인의 N으로 이루어진 서열에 이어지는 서열이다). CER 도메인의 또 다른 적합한 예는 서열 번호:1 (실시예 참조)이며, 이는 5'에서 3' 방향으로 서열 번호:56의 tracrRNA 메이트 서열, 서열 번호:43의 루프 형성 서열 (GAAA) 및 서열 번호:58의 tracrRNA 서열을 포함한다.

[0169] 개시된 발명의 RGEN의 RNA 성분은 5'-캡 (7-메틸구아닐레이트 [m^7G] cap)을 갖지 않는다. 따라서, 본 명세서에서 RNA 성분은 이의 5'-말단에 7-메틸구아닐레이트 (m^7G) 캡을 갖지 않는다. 본 명세서에서, RNA 성분은, 예를 들어, 5'-캡 대신에 5'-하이드록실기를 가질 수 있다. 대안적으로, 본 명세서에서 RNA 성분은, 예를 들어, 5'-캡 대신에 5' 포스페이트를 가질 수 있다. 5'-캡핑된 RNA (즉, 5' m^7G 캡을 갖는 RNA)가 핵외수송의 대상이기 때문에 RNA 성분이 전사 후 핵에 더 잘 축적될 수 있다고 여겨진다. 본 명세서에서, 캡핑되지 않은 RNA 성분의 바람직한 예는 적합한 gRNA, crRNA 및/또는 tracrRNA를 포함한다. 소정 실시 형태에서, 본 명세서의 RNA 성분은 RNA 성분의 전구체의 5'-말단에서 리보자임 서열에 의한 RNA 자가처리에 의해 5'-캡이 없고 대신에 임의로

5'-하이드록실기를 갖는다 (즉, gRNA와 같은 RNA 성분의 상류에 리보자임 서열을 포함하는 전구체 RNA는 리보자임 서열을 제거하기 위해 리보자임-매개 자가처리를 거침으로써, 5'-캡이 없는 RNA 성분의 하류를 남긴다). 소정 다른 실시 형태에서, 본 명세서의 RNA 성분은 RNA 폴리머라제 III (Pol III) 프로모터로부터의 전사에 의해 생성되지 않는다.

[0170] 소정 실시 형태에서, 효모는 (ii) RNA 성분을 암호화하는 뉴클레오티드 서열에 (i) 작동가능하게 연결된 프로모터를 포함하는 DNA 폴리뉴클레오티드 서열을 추가로 포함한다. 이러한 폴리뉴클레오티드 서열은 효모에 의해 Cas 단백질과 복합체를 형성하여 RGEN을 형성하는 RNA 성분을 발현하는 데 사용된다. 이러한 폴리뉴클레오티드 서열은, 예를 들어, 플라스미드, 효모 인공 염색체 (YAC), 코스미드, 파스미드, 박테리아 인공 염색체 (BAC), 바이러스 또는 선형 DNA (예를 들어, 선형 PCR 생성물) 또는 폴리뉴클레오티드 서열을 통상적이지 않은 효모 세포로 전달하는 데 유용한 임의의 다른 타입의 벡터 또는 구조물의 형태일 수 있다. 이러한 폴리뉴클레오티드 서열은 본 명세서의 효모 세포 에서 일시적으로 존재할 수 있거나 (즉, 게놈으로 통합되지 않음) 안정적으로 존재할 수 있다 (즉, 게놈으로 통합됨). 또한, 이러한 폴리뉴클레오티드 서열은 하나 이상의 적합한 마커 서열 (예를 들어, 선택 또는 표현형 마커)를 포함하거나 포함하지 않을 수 있다.

[0171] 본 명세서에서, RNA 성분을 발현하기 위해 폴리뉴클레오티드 서열에 포함되는 적합한 프로모터는 통상적이지 않은 효모 세포에서 작동가능하며, 예를 들어, 구성적일 수 있거나 유도성일 수 있다. 소정 측면에서, 프로모터는 단위 시간당 상대적으로 많은 수의 생산 개시를 이끌 수 있는 프로모터 및/또는 효모를 포함하는 효모에서 유전자의 평균 전사 수준보다 높은 전사 수준을 유도하는 프로모터인 강력한 프로모터를 포함할 수 있다.

[0172] 본 명세서에서 유용한 강력한 프로모터의 예에는 본 명세서에 참고로 포함되는 미국 특허 출원 공개 제2012/0252079호 (DGAT2), 제2012/0252093호 (EL1), 2013/0089910호 (ALK2), 제2013/0089911호 (SPS19), 제2006/0019297호 (GPD 및 GPM), 제2011/0059496호 (GPD 및 GPM), 제2005/0130280호 (FBA, FBAIN, FBAINm), 제2006/0057690호 (GPAT) 및 제2010/0068789호 (YAT1)에 개시된 것들이 포함된다. 강력한 프로모터의 다른 적합한 예에는 표 2에 열거된 것들이 포함된다.

[0173] [표 2]

강력한 프로모터

프로모터 명	고유 유전자	참고 문헌 [a]
XPR2	알칼리성 세포외 프로테아제	미국 특허 제 4937189 호; EP220864
TEF	번역 신장 인자 EF1-α (*tef*)	미국 특허 제 6265185 호
GPD, GPM	글리세르알데하이드-3-포스페이트-데하이드로게나제 (*gpd*), 포스포글리세레이트 뮤타아제 (*gpm*)	미국 특허 제 7259255 호 및 제 7459546 호
GPDIN	글리세르알데하이드-3-포스페이트-데하이드로게나제 (*gpd*)	미국 특허 제 7459546 호
GPM/FBAIN	키메라 포스포글리세레이트 뮤타아제 (*gpm*)/ 프룩토스-비스스페이트 알돌라아제 (*fba1*)	미국 특허 제 7202356 호
FBA, FBAIN, FBAINm	프룩토스-비스스페이트 알돌라아제 (*fba1*)	미국 특허 제 7202356 호
GPAT	글리세롤-3-포스페이트 O-아실트랜스퍼라제 (*gpat*)	미국 특허 제 7264949 호
YAT1	암모늄 전달 효소 (*yat1*)	미국 특허 출원 공개 제 2006/0094102 호
EXP1	수송 단백질	미국 특허 제 7932077 호

[0174] [a] 이 표의 각각의 참조 문헌은 본 명세서에 참고로 포함된다.

[0175] 상기 열거된 강력한 프로모터는 야로위아 리폴리티카로부터 유래되지만, 예를 들어, 본 명세서에 개시된 임의의 통상적이지 않은 효모로부터의 이들의 상응하는 프로모터 (예를 들어, 상동체)가 강력한 프로모터로 작용할 수 있다고 여겨진다. 따라서, 강력한 프로모터는, 예를 들어, XPR2, TEF, GPD, GPM, GPDIN, FBA, FBAIN, FBAINm, GPAT, YAT1, EXP1, DGAT2, EL1, ALK2 또는 SPS19 프로모터를 포함할 수 있다. 대안적으로, 전술한 것 중 어느 하나에 상응하는 것과 같은 강력한 프로모터는 다른 유형의 효모 (예를 들어, S. 세레비시아, S. 폼베)로부터 유래될 수 있다 (예를 들어, 본 명세서에 참고로 포함되는 미국 특허 출원 공개 제2010/0150871호에 개시된 임의의 강력한 프로모터). 본 명세서에서 유용한 강력한 프로모터의 다른 예는 PGK1, ADH1, TDH3, TEF1, PHO5,

LEU2 및 GAL1 프로모터 및 본 명세서에 참고로 포함되는 문헌[Velculescu et aL. (*Cell* 88:243-251)]에 개시된 강력한 효모 프로모터를 포함한다. 본 명세서에서 유용한 강력한 프로모터의 또 다른 예는 서열 번호:12 (야로위아 FBA1 프로모터 서열)을 포함할 수 있다.

[0176] 소정 실시 형태에서, 본 명세서의 프로모터는 RNA 폴리머라제 II (Pol II) 프로모터를 포함할 수 있다. 상기 열거된 모든 강력한 프로모터는 적합한 예 Pol II 프로모터의 예라고 여겨진다. Pol II 프로모터로부터의 전사는, 예를 들어, 적어도 약 12개의 단백질 (예를 들어, RPB1-RPN12 단백질)의 RNA 폴리머라제 II 복합체를 포함할 수 있다. 본 명세서에서, Pol II 프로모터로부터 전사된 RNA는 전형적으로 5'-캡핑된다 (예를 들어, 5'-말단에 m^7G 기를 함유함). 본 명세서의 RNA 성분은 5'-캡을 갖지 않기 때문에, RNA 성분으로부터 5'-캡을 제거하는 방법은 본 명세서의 Pol II 프로모터로부터 발현되는 경우 사용되어야 한다. 본 명세서에서, Pol II-전사된 RNA 성분으로부터 5'-캡을 효과적으로 제거하는 적합한 방법은, 예를 들어, 하나 이상의 리보자임 (하기 참조), 그룹 1 자가-스플라이싱 인트론 및 그룹 2 자가-스플라이싱 인트론의 적절한 사용을 포함한다.

[0177] 본 명세서에서 RNA 성분을 암호화하는 뉴클레오티드 서열은, 예를 들어, RNA 성분을 암호화하는 서열의 상류에 있는 리보자임을 추가로 암호화할 수 있다. 따라서, 소정 실시 형태에서, 효모는 (ii) 5'에서 3' 방향으로 리보자임과 RNA 성분을 암호화하는 뉴클레오티드 서열에 (i) 작동가능하게 연결된 프로모터를 포함하는 DNA 폴리뉴클레오티드 서열을 추가로 포함한다. 이러한 폴리뉴클레오티드 서열로부터 발현된 전사물은 자체 촉매 작용으로 리보자임 서열을 제거하여, 5'-캡을 갖지 않지만 RNA 성분 서열을 포함하는 RNA를 생성한다. 이러한 "자가처리된" RNA는 crRNA 또는 gRNA를 포함할 수 있으며, 예를 들어 Cas 단백질, 예컨대 Cas9와 복합체를 형성하여 RGEN을 형성할 수 있다.

[0178] 본 명세서에서, 리보자임은, 예를 들어, 망치머리 (HH) 리보자임, 델타 간염 바이러스 (HDV) 리보자임, 그룹 I 인트론 리보자임, RnaseP 리보자임 또는 헤어핀 리보자임일 수 있다. 본 명세서에서, 리보자임의 다른 비제한적인 예는 VS(Varkud satellite) 리보자임, 글루코사민-6-포스페이트 활성화 리보자임 (glmS) 및 CPEB3 리보자임을 포함한다. 문헌[Lilley (*BiocheM. SoC. TranS.* 39:641-646)]은 리보자임 구조와 활성에 관한 정보를 개시하고 있다. 본 명세서에 사용하기에 적합한 리보자임의 예에는 본 명세서에 참고로 포함되는 EP0707638 및 미국 특허 제6063566호, 제5580967호, 제5616459호 및 제5688670호에 개시된 리보자임이 포함된다.

[0179] 망치머리 리보자임이 특정 바람직한 실시 형태에서 사용된다. 이러한 유형의 리보자임은, 예를 들어, 본 명세서에 참고로 포함되는 문헌[Hammann et aL. (*RNA* 18:871-885)]에 개시된 바와 같은 타입 I, 타입 II 또는 타입 III 망치머리 리보자임일 수 있다. 본 명세서에서 적절하게 사용될 수 있는, 망치머리 리보자임을 암호화하는 DNA를 동정하는 다수의 방법이 문헌[Hammann et al.]에 개시되어 있다. 본 명세서에서, 망치머리 리보자임은, 예를 들어, 바이러스, 바이로이드, 식물 바이러스 위성 RNA, 원핵생물 (예를 들어, 고세균, 시아노박테리아, 아시도박테리아) 또는 진핵생물, 예컨대 식물 (예를 들어, 아라비돕시스 탈리아나(*Arabidopsis thaliana*), 카네이션), 원생생물 (예를 들어, 아메바, 유글레나류), 진균류 (예를 들어, 아스퍼질러스, 야로위아 리폴리티카), 양서류 (예를 들어, 영원(newt), 개구리), 주혈흡충류(schistosome), 곤충 (예를 들어, 귀뚜라미), 연체 동물, 포유류 (예를 들어, 마우스, 인간) 또는 선충류로부터 유래될 수 있다.

[0180] 본 명세서에서, 망치머리 리보자임은 전형적으로 보존된 서열의 짧은 링커에 의해 이격된, 각각 나선구조 I, II 및 III으로 지칭되는 3가지의 염기쌍 형성된 나선 구조를 포함한다. 망치머리 리보자임의 3가지 타입 (I 내지 III)은 일반적으로 리보자임의 5' 및 3' 말단이 어느 나선 구조에 포함되는 지에 기초한다. 예를 들어, 망치머리 리보자임 서열의 5' 및 3' 말단에 의해 스템 I이 형성된다면, 이는 타입 I 망치머리 리보자임으로 지칭될 수 있다. 3가지 가능한 위상학적(topological) 타입 중에서, 타입 I은 원핵생물, 진핵생물 및 RNA 식물 병원균의 게놈에서 발견될 수 있는 반면, 타입 II 망치머리 리보자임은 원핵생물에서만 기재되어 왔으며, 타입 III 망치머리 리보자임은 주로 식물, 식물 병원균 및 원핵생물에서 발견된다. 소정 실시 형태에서, 망치머리 리보자임은 타입 I 망치머리 리보자임이다.

[0181] 소정 실시 형태에서, 망치머리 리보자임을 암호화하는 서열은 적어도 약 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140 또는 150 (또는 40 내지 150 사이의 임의의 정수)개의 뉴클레오티드, 40 내지 100개의 뉴클레오티드 또는 40 내지 60개의 뉴클레오티드를 포함할 수 있다.

[0182] 망치머리 리보자임을 암호화하는 서열은 RNA 성분을 암호화하는 서열의 상류에 있다. 본 명세서에서, 망치머리 리보자임을 암호화하는 서열은, 예를 들어, RNA 성분의 가이드 서열을 암호화하는 서열의 5' 부근이거나, 이의 5'의 적어도 2, 3, 4, 5, 6, 7, 8, 9 또는 10개의 뉴클레오티드일 수 있다 (예를 들어, 가이드 서열은 crRNA 또는 gRNA의 것일 수 있다). 망치머리 리보자임의 처음 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 또는 15번째의 리

보뉴클레오티드는 전형적으로 망치머리 리보자임 서열의 바로 하류에 있는 서열의 리보뉴클레오티드의 처음 동일한 번호에 각각 상보적이어야 한다. 예를 들어, 본 명세서에서, 폴리뉴클레오티드 서열이 RNA 성분의 가이드 서열의 바로 상류에서 망치머리 리보자임 서열을 포함하는 RNA를 암호화하는 경우, 리보자임의 처음 6개의 리보뉴클레오티드는, 예를 들어, 가이드 서열의 처음 6개의 리보뉴클레오티드에 상보적일 수 있다. 이러한 예에서, 망치머리 리보자임은 가이드 서열의 첫 번째 위치의 바로 상류에서 RNA 전사물을 절단할 것이다 (또는 또 다른 방식으로 말하면, 망치머리 리보자임은 리보자임 서열의 바로 하류에서 RNA 전사물을 절단할 것이다). 이러한 논리는 다른 전술한 예시적인 실시 형태에 유사하게 적용된다. 예를 들어, 본 명세서에서, 폴리뉴클레오티드 서열이 RNA 성분의 가이드 서열의 상류에 있는 8개의 잔기인 망치머리 리보자임 서열을 포함하는 RNA를 암호화하는 경우 (예를 들어, 8-잔기 스페이서 서열이 있음), 예를 들어, 리보자임의 처음 6개의 리보뉴클레오티드는 리보자임 서열의 3' 부근의 6개의 리보뉴클레오티드에 상보적일 수 있다. 이러한 예에서, 망치머리 리보자임은 리보자임 서열의 바로 하류에서 RNA 전사물을 절단할 것이다. 또 다른 예로서, 본 명세서에서, 폴리뉴클레오티드 서열이 RNA 성분의 가이드 서열의 바로 상류에서 망치머리 리보자임 서열을 포함하는 RNA를 암호화하는 경우, 리보자임의 처음 10개의 리보뉴클레오티드는, 예를 들어, 가이드 서열의 처음 10개의 리보뉴클레오티드에 상보적일 수 있다. 이러한 예에서, 망치머리 리보자임은 가이드 서열의 첫 번째 위치의 바로 상류에서 RNA 전사물을 절단할 것이다 (또는 또 다른 방식으로 말하면, 망치머리 리보자임은 리보자임 서열의 바로 하류에서 RNA 전사물을 절단할 것이다).

[0183] 망치머리 리보자임 서열의 예를 다음과 같이 나타낼 수 있다:

[0184] NNNNNNcugaugaguccgugaggacgaaacgaguaagcucguc (서열 번호:15, N은 A, U, C 또는 G일 수 있다; 실시예 참조). 서열 번호:15의 처음 6개의 잔기는 본 명세서의 DNA 폴리뉴클레오티드로부터 발현된 RNA 전사물에서 서열 번호:15의 바로 다음에 오는 (예를 들어, 본 명세서에 개시된 crRNA 또는 gRNA의 가이드 서열의) 처음 6개의 잔기에 상보적이도록 (에 어닐링되도록) 설계될 수 있다. 리보자임은 서열 번호:15 바로 다음의 전사물을 절단할 것이다. 서열 번호:15를 바로 뒤따르는 서열 잔기와 어닐링하기 위해 서열 번호:15가 6개의 잔기 ("N")를 갖는 것으로 도시되어 있지만, 이러한 목적을 위한 이러한 리보자임의 시작에는 5 내지 15개의 "N" 잔기가 있을 수 있다. 서열 번호:15를 포함하는 RNA 전사물에 있어서, (i) 망치머리 리보자임의 나선 구조 I은 전사물에서 서열 번호:15 바로 다음의 처음 6개의 잔기와 N 잔기의 어닐링에 의해 형성될 것이고, (ii) 나선 구조 II는 단일-밑줄로 나타낸 상보적인 서열의 어닐링에 의해 형성될 것이며, (iii) 나선 구조 III은 이중-밑줄로 나타낸 상보적인 서열의 어닐링에 의해 형성될 것이라는 점에 유의해야 한다. 따라서, 소정 실시 형태에서, 망치머리 리보자임은 (i) 서열 번호:15와 적어도 약 80%, 85%, 90, 91, 92, 93, 94, 95, 96, 97, 98 또는 99%의 서열 동일성 (이 계산에서 "N" 서열은 제외함)을 갖고 (ii) 서로 어닐링되어 나선 구조 II와 III (나선 구조 I은 "N" 잔기의 적절한 선택으로 형성됨)을 형성하는 서열 번호:15의 단일 밑줄로된 영역 및 이중 밑줄로된 영역과 정렬되는 영역을 갖는 서열 번호:15의 변이체일 수 있다.

[0185] 서열 번호:15 및 (상기) 이들의 다양한 실시 형태에 연결될 수 있는 서열의 예에는 서열 번호:64 내지 70 중 하나를 포함하는 gRNA가 포함된다.

[0186] 본 명세서에서, RNA 성분 (본 명세서에서 "리보자임-RNA 성분 카세트")에 연결된 5' 망치머리 리보자임을 포함하는 RNA 서열을 암호화하는 DNA 폴리뉴클레오티드는 망치머리 리보자임 서열로 바로 시작되는 5'-말단을 갖는 전사물의 전사를 유도하도록 설계될 수 있다 (즉, 전사 시작 부위에서 리보자임 서열이 시작된다). 대안적으로, DNA 폴리뉴클레오티드는 리보자임-RNA 성분 카세트의 상류에 비-리보자임(non-ribozyme) 서열을 갖는 전사물의 전사를 유도하도록 설계될 수 있다. 이러한 5' 비-리보자임 전사물 서열은, 예를 들어, 단지 수 개의 뉴클레오티드 (1 내지 10)만큼 짧거나, 최대 5000 내지 20000개의 뉴클레오티드 만큼 길 수도 있다 (이러한 리보자임의 5' 서열은 리보자임이 그 자체를 RNA 성분으로부터 절단하는 경우, RNA 성분으로부터 제거된다).

[0187] 소정 실시 형태에서, 리보자임-RNA 성분 카세트를 포함하는 DNA 폴리뉴클레오티드는 RNA 성분 서열의 하류에 적합한 전사 종결 서열을 포함할 수 있다. 본 명세서에서 유용한 전사 종결 서열의 예가 본 명세서에 참고로 포함되는 미국 특허 출원 공개 제2014/0186906호에 개시되어 있다. 예를 들어, S. 세레비시아 Sup4 유전자 전사 종결 서열 (예를 들어, 서열 번호:8)이 사용될 수 있다. 이러한 실시 형태는 전형적으로 리보자임-RNA 성분 카세트의 하류에 위치한 리보자임 서열을 포함하지 않는다. 또한, 이러한 실시 형태는 전형적으로 종결 서열의 선택에 따라 RNA 성분 서열의 말단에 이어지는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30개 또는 그 이상의 잔기를 포함한다. 이러한 추가의 잔기는, 예를 들어, 종결 서열의 선택에 따라 모두 U 잔기 또는 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%,

97%, 98% 또는 99%의 U 잔기일 수 있다. 대안적으로, 리보자임 서열 (예를 들어, 망치머리 또는 HDV 리보자임)은 (예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 그 이상의 뉴클레오티드) RNA 성분 서열의 3'일 수 있으며; 이러한 실시 형태에서 RNA 성분 서열은 상류 및 하류 리보자임에 의해 플랭킹된다. 3' 리보자임 서열은 RNA 성분 서열로부터 그 자체를 절단하도록 적절하게 위치될 수 있고; 이러한 절단은, 예를 들어, RNA 성분 서열의 말단에서 정확하게, 또는 RNA 성분 서열의 말단에 이어지는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15개 또는 그 이상의 잔기를 갖도록 전사물을 종료시킬 것이다.

[0188] 소정 실시 형태에서, DNA 폴리뉴클레오티드는 (ii) 하나 초과의 리보자임-RNA 성분 카세트 (즉, 탠덤 카세트)를 포함하는 서열에 (i) 작동가능하게 연결된 프로모터를 포함할 수 있다. 이러한 DNA 폴리뉴클레오티드로부터 발현된 전사물은, 예를 들어, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 그 이상의 리보자임-RNA 성분 카세트를 가질 수 있다. 3' 리보자임 서열은 전사물 서열의 하류로부터 RNA 성분의 절단 및 분리를 허용하기 위해 각각의 RNA 성분 서열 다음에 임의로 (예를 들어, 상기와 같이) 포함될 수 있다. 이러한 실시 형태에서 각각의 RNA 성분은 전형적으로 본 명세서의 RGEN을 특유의 DNA 표적 부위 유도하도록 설계된다. 따라서, 이러한 DNA 폴리뉴클레오티드는, 예를 들어 통상적이지 않은 효모에서 동시에 다수의 상이한 표적 부위를 표적화하는데 적절하게 사용될 수 있고; 이러한 사용은 임의로 다중화 방법으로 특징지어질 수 있다. 3' 리보자임에 연결된 RNA 성분에 연결된 5' 망치머리 리보자임은 본 명세서에서 "리보자임-RNA 성분-리보자임 카세트"로 지칭될 수 있다. 본 명세서에서, 탠덤 리보자임-RNA 성분-리보자임 카세트를 포함하는 전사물을 발현하기 위한 DNA 폴리뉴클레오티드는 각각의 카세트 (예를 들어, 비암호화 스페이서 서열) 사이에 약 10, 20, 30, 40, 50, 60, 70, 80, 90, 100개 또는 그 이상의 뉴클레오티드가 존재하도록 설계될 수 있다. 각각의 카세트 사이의 거리는 동일하거나 상이할 수 있다.

[0189] 상기 실시 형태 중 일부는 망치머리 리보자임 서열에 관하여 기술하고 있지만, 이러한 실시 형태는 또한 이에 따라 망치머리 리보자임 서열 대신에 본 명세서의 임의의 다른 리보자임 서열 (예를 들어, HDV 리보자임)의 관점에서 특징지어질 수 있다. 당업자는 이러한 다른 리보자임 서열을 특정 부위에서 절단하도록 배치하는 방법을 이해할 것이다.

[0190] 소정 실시 형태에서, 효모는 (ii) Cas 단백질 (예를 들어, Cas9)을 암호화하는 뉴클레오티드 서열에 (i) 작동가능하게 연결된 프로모터를 포함하는 DNA 폴리뉴클레오티드 서열을 추가로 포함한다. 이러한 폴리뉴클레오티드 서열은 효모에 의해 RNA 성분과 복합체를 형성하여 RGEN을 형성하는 Cas 단백질을 발현하는 데 사용된다. 이러한 폴리뉴클레오티드 서열은, 예를 들어, 플라스미드, YAC, 코스미드, 파스미드, BAC, 바이러스 또는 선형 DNA (예를 들어, 선형 PCR 생성물), 또는 폴리뉴클레오티드 서열을 통상적이지 않은 효모 세포로 전달하는 데 유용한 임의의 다른 타입의 벡터 또는 구조물의 형태일 수 있다. 본 명세서에 개시된 임의의 Pol II 프로모터가, 예를 들어 사용될 수 있다. 따라서, RNA 성분을 발현하기 위한 DNA 폴리뉴클레오티드 서열에 관련하여 상기 개시된 임의의 특징은 Cas 단백질을 발현하기 위한 DNA 폴리뉴클레오티드 서열에 적용될 수 있다. 이러한 폴리뉴클레오티드 서열은 본 명세서의 효모 세포 에서 일시적으로 존재할 수 있거나 (즉, 게놈으로 통합되지 않음) 안정적으로 존재할 수 있다 (즉, 게놈으로 통합됨). 다른 측면에서, 효모는 Cas 단백질을 발현하기 위한 DNA 폴리뉴클레오티드에 더하여, RNA 성분을 발현하기 위한 DNA 폴리뉴클레오티드를 가질 수 있다 (예를 들어, 상기한 바와 같음). 이러한 DNA 폴리뉴클레오티드는 모두 효모에 안정적이거나 일시적일 수 있거나; Cas 단백질을 발현하기 위한 DNA 폴리뉴클레오티드는 안정적이고 RNA 성분을 발현하기 위한 DNA 폴리뉴클레오티드는 일시적일 수 있다 (또는 그 반대).

[0191] 대안적으로, DNA 폴리뉴클레오티드 서열은 효모 세포에서 RGEN을 제공하기 위한 Cas 단백질과 적합한 RNA 성분 모두를 발현하는 것일 수 있다. 이러한 DNA 폴리뉴클레오티드는, 예를 들어, (i) (RGEN의) RNA 성분 (RNA 성분 카세트)을 암호화하는 뉴클레오티드 서열에 작동가능하게 연결된 프로모터 및 (ii) Cas 단백질 (예를 들어, Cas9) (Cas 카세트)을 암호화하는 뉴클레오티드 서열에 작동가능하게 연결된 프로모터를 포함할 수 있다. Cas 단백질 또는 RNA 성분을 발현하기 위한 DNA 폴리뉴클레오티드에 관련하여 상기 기재된 임의의 특징은, 예를 들어, 통상적이지 않은 효모 세포에서 Cas 단백질과 적합한 RNA 성분을 모두 발현하기 위한 DNA 폴리뉴클레오티드 서열에 적용될 수 있다. 또한, 본 명세서에 개시된 임의의 Cas 단백질과 RNA 성분 (예를 들어, crRNA 또는 gRNA)은 이러한 DNA 폴리뉴클레오티드 서열로부터 발현될 수 있다. 소정 실시 형태에서, 하나 이상의 RNA 성분 및/또는 Cas 카세트는 DNA 폴리뉴클레오티드 서열 내에 포함될 수 있다. 다른 측면에서, 하나 이상의 RNA 성분은 상기한 바와 같이 탠덤으로 발현될 수 있다. Cas 카세트 및 RNA 카세트에서 사용된 프로모터는 동일하거나 상이할 수 있다. 이러한 DNA 폴리뉴클레오티드 서열이 통상적이지 않은 효모와 통상적인 효모 모두에서 RGEN을 발현하는 데 유용할 것으로 생각된다.

[0192] 개시된 발명은 또한 통상적이지 않은 효모 내의 염색체 또는 에피솜의 표적 부위 서열로 RNA-유도 엔도뉴클레아제 (RGEN)를 표적화하는 방법에 관한 것이다. 이러한 방법은 효모의 핵에 5'-캡을 갖지 않는 적어도 하나의 RNA 성분을 포함하는 RGEN을 제공하는 단계를 포함하며, 여기서 RNA 성분은 표적 부위 서열에 상보적인 서열을 포함하고, RGEN은 표적 부위 서열의 전부 또는 일부에 결합하고, 임의로 절단한다.

[0193] 이러한 표적화 방법은, 예를 들어, 방법의 특징 (예를 들어, 효모 타입, RGEN, RNA 성분, 등)각각에 관한 상기 개시된 실시 형태 또는 하기 실시예 중 어느 하나를 사용하여 실시될 수 있다. 따라서, 상기 또는 실시예에서 개시된 특징 중 어느 하나 또는 이들 특징의 임의의 조합이 적절하게 사용되어, 본 명세서의 표적화 방법의 실시 형태를 특징지을 수 있다. 다음의 표적화 방법 특징은 예이다.

[0194] 본 명세서의 표적화 방법의 소정 실시 형태에서, 통상적이지 않은 효모는 다음과 같은 속의 임의의 구성원일 수 있다: 야로위아, 피키아, 쉬반니오마이세스, 클루이베로마이세스, 아르술라, 트리코스포론, 칸디다, 유스틸라고, 토룰롭시스, 자이고사카로마이세스, 트리고놉시스, 크립토코커스, 로도토룰라, 파피아, 스포로볼로마이세스 및 파치솔렌. 야로위아 리폴리티카는 본 명세서에 적합한 야로위아 효모이다. 표적화 방법에서 유용한 통상적이지 않은 효모의 다른 비제한적인 예가 본 명세서에 개시되어 있다.

[0195] 본 명세서의 표적화 방법에 사용하기에 적합한 RGEN은 타입 I, II 또는 III 크리스퍼 시스템의 Cas 단백질을 포함할 수 있다. 소정 실시 형태에서, Cas9 단백질, 예컨대 스트렙토코커스 Cas9가 사용될 수 있다. 표적화 방법에 사용하기에 적합한 스트렙토코커스 Cas9 단백질의 예에는 S. 피오제네스, S. 써모필러스, S. 뉴모니애, S. 아갈락티아, S. 파라상귀니스, S. 오랄리스, S. 살리바리우스, S. 마카카, S. 디스갈락티애, S. 안지노서스, S. 콘스텔라투스, S. 슈도포르시누스 또는 S. 뮤탄스의 Cas9 단백질로부터 유래된 아미노산 서열을 포함하는 Cas9 단백질이 포함된다. 본 명세서의 표적화 방법에 유용한 RGEN 및 Cas9 단백질의 비제한적인 예가 본 명세서에 개시되어 있다. 예를 들어, DNA 표적 서열에서 하나 또는 두 가닥을 절단할 수 있는 RGEN이 사용될 수 있다.

[0196] 본 명세서의 표적화 방법에 사용하기 위한 RGEN의 RNA 성분은, 예를 들어, tracrRNA에 작동가능하게 연결되거나 융합된 crRNA를 포함하는 gRNA를 포함할 수 있다. 본 명세서에 개시된 바와 같은 임의의 crRNA 및/또는 tracrRNA (및/또는 이들의 부분, 예컨대 tracrRNA 메이트 서열 또는 tracrRNA 5'-말단 서열)는, 예를 들어, gRNA에 포함될 수 있다. 또한, 본 명세서에 개시된 임의의 gRNA가, 예를 들어, 표적화 방법에서 사용될 수 있다.

[0197] PAM (프로토스페이서-인접 모티프) 서열은, 예를 들어, 표적 부위 서열에 인접할 수 있다. 본 명세서의 표적화 방법의 소정 실시 형태에서, PAM 서열은 결과적으로 RNA 성분 가이드 서열에 상보적인 표적 부위의 가닥에 상보적인 표적 부위 서열로부터 바로 하류에, 또는 이의 2 또는 3개의 뉴클레오티드 하류에 있다. RGEN이 RNA 성분과 복합체를 형성한 핵산 내부 분해 방식으로 활성인 Cas9 단백질인 본 명세서의 실시 형태에서, Cas9는 RNA 성분에 의해 유도된 바와 같이 표적 서열에 결합하고, PAM 서열의 상류 세 번째 뉴클레오티드 위치의 5' 부근에서 두 가닥을 절단한다. 적합한 PAM 서열의 예에는 S. 피오제네스 (NGG [서열 번호:47]) 및 S. 써모필러스 (NNAGAA [서열 번호:48]) PAM 서열이 포함되며, 이는 각각 각각의 종으로부터 유래된 Cas9 단백질로 표적으로 하는데 사용될 수 있다. 또한, 본 명세서에 개시된 임의의 PAM 서열이, 예를 들어, 표적화 방법에서 사용될 수 있다.

[0198] 본 명세서의 표적화하는 방법의 소정 실시 형태에서, 효모는 (ii) RNA 성분을 암호화하는 뉴클레오티드 서열에 (i) 작동가능하게 연결된 프로모터를 포함하는 DNA 폴리뉴클레오티드 서열을 추가로 포함한다. RNA 성분이 DNA 폴리뉴클레오티드로부터 전사되기 때문에, 이러한 DNA 폴리뉴클레오티드를 사용하여 RGEN의 RNA 성분이 효모의 핵에 제공될 수 있다. 효모 핵에서 (RGEN의) RNA 성분을 발현하기 위한 적합한 DNA 폴리뉴클레오티드 서열의 예가 본 명세서에 개시되어 있다. 본 명세서에 개시된 바와 같은 임의의 프로모터가, 예를 들어, 강력한 프로모터 및/또는 Pol II 프로모터 서열을 포함하는 것과 같은 이러한 DNA 폴리뉴클레오티드 서열에서 사용될 수 있다. 소정 실시 형태에서, RNA 성분을 암호화하는 DNA 폴리뉴클레오티드가 Cas 단백질을 발현하도록 이미 조작된 (예를 들어, 안정한 Cas 발현) 효모에서 RNA 성분을 제공하는 데 사용될 수 있다.

[0199] 본 명세서에서 RNA 성분을 암호화하는 뉴클레오티드 서열은, 예를 들어, RNA 성분을 암호화하는 서열의 상류에 있는 리보자임을 추가로 암호화할 수 있다. 따라서, 본 명세서의 표적화 방법의 소정 실시 형태에서, 효모는 (ii) 5'에서 3' 방향으로 리보자임과 RNA 성분을 암호화하는 뉴클레오티드 서열에 (i) 작동가능하게 연결된 프로모터를 포함하는 DNA 폴리뉴클레오티드 서열을 포함할 수 있다. RNA 성분이 DNA 폴리뉴클레오티드로부터 전사되기 때문에, 이러한 DNA 폴리뉴클레오티드를 사용하여 RGEN의 RNA 성분을 효모의 핵에 제공할 수 있다. 본 명세서에서, 리보자임은, 예를 들어, 망치머리 리보자임, 델타 간염 바이러스 (HDV) 리보자임, 그룹 I 인트론

리보자임, RnaseP 리보자임 또는 헤어핀 리보자임일 수 있다. 본 명세서에 개시된 바와 같은 임의의 리보자임 및 본 명세서에 개시된 바와 같은 임의의 RNA 성분에 연결된 리보자임을 암호화하는 폴리뉴클레오티드 서열이, 예를 들어, 표적화 방법에서 사용될 수 있다.

[0200] 본 명세서의 표적화 방법의 소정 실시 형태에서, 효모는 (ii) Cas 단백질 (예를 들어, Cas9)을 암호화하는 뉴클레오티드 서열에 (i) 작동가능하게 연결된 프로모터를 포함하는 DNA 폴리뉴클레오티드 서열을 추가로 포함한다. 이러한 DNA 폴리뉴클레오티드을 사용하여 효모에서 RGEN의 Cas 단백질 성분을 제공할 수 있다. 효모에서 (RGEN의) Cas 단백질 성분을 발현하기 위한 적합한 DNA 폴리뉴클레오티드 서열의 예가 본 명세서에 개시되어 있다. 본 명세서에 개시된 바와 같은 임의의 프로모터가, 예를 들어, 강력한 프로모터을 포함하는 것과 같은 이러한 DNA 폴리뉴클레오티드 서열에서 사용될 수 있다.

[0201] DNA 표적 부위 서열에 또는 그 부근의 서열에 상동성인 적어도 하나의 서열을 포함하는 공여 폴리뉴클레오티드는 또한 표적화 방법의 소정 실시 형태에서 (표적 부위에 서열에서 닉을 형성하거나 절단하는 RGEN으르 제공하는 것과 함께) 효모에 제공될 수 있다. 적합한 예에는 상동성 암을 갖는 공여 폴리뉴클레오티드가 포함된다. 본 명세서에 개시된 바와 같은 임의의 공여 폴리뉴클레오티드가 표적화 방법에서 사용될 수 있다. 이러한 방법의 이러한 실시 형태는 전형적으로 공여 폴리뉴클레오티드와 표적 서열 (표적 서열의 RGEN-매개 닉킹 또는 절단 후) 사이의 HR을 포함하며; 따라서, 이러한 방법은 임의로 통상적이지 않은 효모에서 HR을 수행하는 방법으로도 지칭될 수 있다. 이러한 방법으로 수행될 수 있는 HR 전략의 예가 본 명세서에 개시되어 있다. 효모 세포에서 표적화하기 위한 공여 DNA 폴리뉴클레오티드의 적합한 양은 효모 세포 당 적어도 약 300, 400, 500, 600, 700 또는 800개의 공여 DNA 분자일 수 있다.

[0202] RGEN 성분을 발현하기 위해 본 명세서에 기재된 DNA 폴리뉴클레오티드를 포함하는 임의의 구조물 또는 벡터는 임의의 표준 기술에 의해 통상적이지 않은 효모 세포로 도입될 수 있다. 이러한 기술에는, 예를 들어 형질전환 (예를 들어, 아세트산리튬 형질전환 (문헌[*Methods in Enzymology*, 194:186-187]), 바이올리스틱 충격 (biolistic impact), 전기천공 및 미세주입이 포함된다. 예로서, 본 명세서에 참고로 포함되는 미국 특허 제4880741호 및 제5071764호 및 문헌[Chen et aL. (*AppL. MicrobioL. BiotechnoL.* 48:232-235)]은 야로위아 리폴리티카에 대한 DNA 전달 기술을 기재한다.

[0203] 본 명세서의 표적화 방법은 통상적이지 않은 효모에서 삽입-결실을 생성하기 위한 목적으로 수행될 수 있다. 이러한 방법은 상기 개시된 바와 같이 수행될 수 있지만, 표적 DNA 부위 또는 그 부근에서 HR을 행할 수 있는 공여 DNA 폴리뉴클레오티드를 추가로 제공하지 않는다. (즉, 이러한 방법에서는 NHEJ가 유도된다). 생성될 수 있는 삽입-결실의 예가 본 명세서에 개시되어 있다. 삽입-결실의 크기는, 예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 그 이상의 염기일 수 있다. 소정 실시 형태에서 삽입-결실은 적어도 약 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140 또는 150개의 염기와 같이 심지어 더 클 수 있다. 또 다른 실시 형태에서, 삽입 또는 결실은 적어도 약 500, 750, 1000 또는 1500개의 염기일 수 있다. 소정 실시 형태에서 삽입-결실을 생성하고자 시도하는 경우, 표적 부위 서열에 단일 염기 치환이 대신 형성될 수 있다. 따라서, 본 명세서의 표적화 방법은, 예를 들어, 단일 염기 치환을 생성하기 위한 목적으로 수행될 수 있다.

[0204] 본 명세서의 삽입-결실 형성을 목적으로 하는 표적화 방법의 소정 실시 형태에서, 통상적이지 않은 효모 (예를 들어, 야로위아 리폴리티카)에서 삽입-결실의 형성의 빈도는 통상적인 효모, 예컨대 S. 세레비시아에서 동일하거나 유사한 표적화 전략을 사용하여 관찰되는 것보다 상당히 더 높다. 예를 들어, 삽입-결실의 형성의 빈도는 통상적인 효모에서 약 0.0001 내지 0.001일 수 있는 반면 (문헌[DiCarlo et al., *Nucleic Acids Res.* 41:4336-4343]), 본 명세서의 통상적이지 않은 효모에서의 빈도는 적어도 약 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75 또는 0.80일 수 있다. 따라서, 본 명세서의 통상적이지 않은 효모에서 삽입-결실의 형성의 빈도는 통상적인 효모에서 동일하거나 유사한 Cas-매개 표적화 전략을 사용하여 발견된 것보다, 예를 들어, 적어도 약 50, 100, 250, 500, 750, 1000, 2000, 4000 또는 8000배 더 높을 수 있다. 이러한 실시 형태의 소정 측면은 공여 DNA를 포함하지 않고/않거나 RGEN 성분 (Cas 및 적합한 RNA 성분)이 동일한 벡터/구조물로부터 발현되는 표적화 방법에 관한 것일 수 있다.

[0205] 본 명세서의 표적화 방법은, 예를 들어, 방법에서 둘 이상의 DNA 표적 부위가 표적화되는 방식으로 수행될 수 있다. 이러한 방법은 본 명세서에 개시된 바와 같은 탠덤 리보자임-RNA 성분 카세트 (예를 들어, 탠덤 리보자임-RNA 성분-리보자임 카세트)를 포함하는 전사물을 발현하는 DNA 폴리뉴클레오티드를 효모에 제공하는 단계를 포함할 수 있다. 이러한 방법은 동일 서열 (예를 들어, 프로모터 또는 오픈 리딩 프레임)에서 매우 가까운 DNA 부위 및/또는 서로 멀리 있는 (예를 들어, 상이한 유전자 및/또는 염색체) 부위를 표적화할 수 있다. 이러한

방법은 표적화의 원하는 결과에 따라 적합한 공여 DNA 폴리뉴클레오티드를 사용하거나 (HR의 경우) 사용하지 않고 (삽입-결실 및/또는 염기 치환으로 유도되는 NHEJ의 경우) 수행될 수 있다

[0206] 소정 실시 형태에서 표적화 방법이 수행되어 단백질 또는 비암호화 RNA를 암호화하는 하나 이상의 DNA 폴리뉴클레오티드 서열을 파괴할 수 있다. 파괴의 표적이 될 수 있는 이러한 서열의 예는 마커 (즉, 마커 유전자)를 암호화하는 것이다. 본 명세서의 마커의 비제한적인 예에는 스크리닝가능한 마커와 선택가능한 마커가 포함된다. 본 명세서에서 스크리닝가능한 마커는 적절한 조건 하에서 효모를 시각적으로 상이하게 만드는 것일 수 있다. 스크리닝가능한 마커의 예에는 베타-글루쿠로니다제 (GUS), 베타-갈락토시다제 (lacZ) 및 형광 단백질 (예를 들어, GFP, RFP, YFP, BFP)을 암호화하는 폴리뉴클레오티드가 포함된다. 본 명세서에서 선택가능한 마커는 효모를 선택적 제제 또는 선택적 환경에 내성으로 만드는 것일 수 있다. 선택가능한 마커의 예는 효모가 각각 외인성으로 공급된 히스티딘, 류신, 트립토판, 메티오닌 또는 우라실의 부재 하에 생존할 수 있게 해주는 영양요구성 마커, 예컨대 HIS3, LEU2, TRP1, MET15 또는 URA3이다. 선택가능한 마커의 다른 예는 효모에 하이그로마이신 B, 누오르세오트리신(nourseothricin), 플레오마이신, 퓨로마인신 또는 네오마이신 (예를 들어, G418)에 대한 내성을 부여하는 것들과 같은 항생제 (항진균제)-저항 마커이다.

[0207] 소정 실시 형태에서 마커를 파괴하는 적어도 하나의 목적은 마커 재활용(marker recycling)일 수 있다. 마커 재활용은, 예를 들어, (i) 효모를 마커 및 이종 DNA 서열로 형질전환시키는 단계, (ii) 마커 및 이종 DNA 서열으로 형질전환된 효모 (여기서, 마커-선택가능한 효모는 전형적으로 이종 DNA 서열을 함유할 확률이 더 높음)를 선택하는 단계, (iii) 마커를 파괴한 후 단계 (i) 내지 (iii)을 필요한 만큼 많이 반복하여 (동일한 마커를 사용하지만, 각각의 사이클은 상이한 이종 DNA 서열을 사용함) 효모를 다수의 이종 DNA 서열로 형질전환시키는 단계를 포함하는 공정이다. 이러한 공정에서 하나 이상의 이종 서열은 공여 폴리뉴클레오티드의 형태로 마커 그 자체를 포함할 수 있다 (예를 들어, 특정 유전자좌를 표적화기 위한 상동성 암에 의해 플랭킹된 마커). 본 명세서의 마커 재활용 공정의 예에는 통상적이지 않은 효모, 예컨대 야로위아 리폴리티카에서 마커로서 URA3을 사용하는 것들이 포함된다.

[0208] 본 명세서에 개시된 조성물 및 방법의 비제한적인 예는 다음과 같다:

[0209] 1. 5'-캡을 갖지 않는 적어도 하나의 RNA 성분을 포함하는 적어도 하나의 RNA-유도 엔도뉴클레아제 (RGEN)를 포함하는 통상적이지 않은 효모로서, 여기서 RNA 성분은 효모 내의 염색체 또는 에피솜 상에 표적 부위 서열에 상보적인 서열을 포함하고, RGEN은 표적 부위 서열의 전부 또는 일부에 결합할 수 있는, 통상적이지 않은 효모.

[0210] 2. RGEN이 표적 부위 서열의 전부 또는 일부에 결합하여 절단할 수 있는, 실시 형태 1의 통상적이지 않은 효모.

[0211] 3. 상기 효모가 야로위아, 피키아, 쉬반니오마이세스, 클루이베로마이세스, 아르술라, 트리코스포론, 칸디다, 유스틸라고, 토룰롭시스, 자이고사카로마이세스, 트리고놉시스, 크립토코커스, 로도토룰라, 파피아, 스포로볼로마이세스 및 파치솔렌으로 이루어진 군으로부터 선택되는 속의 구성원인, 실시 형태 1의 효모.

[0212] 4. RGEN이 크리스퍼(CRISPR; 규칙적으로 사이 간격을 두고 분포하는 짧은 회문구조 반복 서열)-관련 (Cas) 단백질-9 (Cas9) 아미노산 서열을 포함하는, 실시 형태 1의 효모.

[0213] 5. Cas9 단백질이 스트렙토코커스 Cas9 단백질인, 실시 형태 4의 효모.

[0214] 6. RNA 성분이 트랜스-활성화 크리스퍼 RNA (tracrRNA)에 작동가능하게 연결된 크리스퍼 RNA (crRNA)를 포함하는 가이드 RNA (gRNA)를 포함하는, 실시 형태 4의 효모.

[0215] 7. PAM (프로토스페이서-인접 모티프) 서열이 표적 부위 서열에 인접한, 실시 형태 4의 효모.

[0216] 8. 적어도 하나의 뉴클레오티드 서열에 작동가능하게 연결된 프로모터를 포함하는 폴리뉴클레오티드 서열을 포함하는 통상적이지 않은 효모로서, 여기서 상기 뉴클레오티드 서열은 RNA 성분을 암호화하는 DNA 서열의 상류에 리보자임을 암호화하는 DNA 서열을 포함하고, 상기 RNA 성분은 효모 내의 염색체 또는 에피솜 상의 표적 부위 서열에 상보적인 가변 표적화 도메인을 포함하며, RNA 성분은 표적 부위 서열의 전부 또는 일부에 결합할 수 있는 RNA-유도 엔도뉴클레아제 (RGEN)를 형성할 수 있는, 통상적이지 않은 효모.

[0217] 9. RGEN이 표적 부위 서열의 전부 또는 일부에 결합하여 절단할 수 있는, 실시 형태 8의 통상적이지 않은 효모.

[0218] 10. 뉴클레오티드 서열로부터 전사된 RNA가 자체 촉매 작용으로 리보자임을 제거하여 상기 RNA 성분을 생성하며, 상기 RNA 성분이 5' 캡을 포함하지 않는, 실시 형태 8의 통상적이지 않은 효모.

[0219] 11. 리보자임이 망치머리 리보자임, 델타 간염 바이러스 리보자임, 그룹 I 인트론 리보자임, RnaseP 리보자임

또는 헤어핀 리보자임인, 실시 형태 10의 통상적이지 않은 효모.

[0220] 12. 뉴클레오티드 서열로부터 전사된 RNA가 5' 캡이 없는 리보자임-RNA 성분 융합 분자를 생성하기 위해 자체 촉매 작용으로 리보자임을 제거하지 않는, 실시 형태 8의 통상적이지 않은 효모.

[0221] 13. 리보자임이 HDV 리보자임인, 실시 형태 12의 통상적이지 않은 효모.

[0222] 14. 프로모터가 강력한 프로모터인, 실시 형태 8의 통상적이지 않은 효모.

[0223] 15. 프로모터가 Pol II 프로모터 서열을 포함하는, 실시 형태 8의 통상적이지 않은 효모.

[0224] 16. 통상적이지 않은 효모 내의 염색체 또는 에피솜의 표적 부위 서열로 RNA-유도 엔도뉴클레아제 (RGEN)를 표적화하는 방법으로서, 상기 방법은 상기 효모에 5'-캡을 갖지 않는 적어도 하나의 RNA 성분을 포함하는 RGEN을 제공하는 단계를 포함하며, 여기서 RNA 성분은 표적 부위 서열에 상보적인 서열을 포함하고, RGEN은 표적 부위 서열의 전부 또는 일부에 결합하는, 방법.

[0225] 17. RGEN이 표적 부위 서열의 전부 또는 일부에 결합하여 절단할 수 있는, 실시 형태 16의 방법.

[0226] 18. 통상적이지 않은 효모 내의 염색체 또는 에피솜의 표적 부위 서열로 RNA-유도 엔도뉴클레아제 (RGEN)를 표적화하는 방법으로서, 상기 방법은 상기 효모에 적어도 하나의 리보자임-RNA 성분 융합 분자를 포함하는 RGEN을 제공하는 단계를 포함하며, 여기서 RNA 성분은 표적 부위 서열에 상보적인 서열을 포함하고, RGEN은 표적 부위 서열의 전부 또는 일부에 결합하는, 방법.

[0227] 19. RGEN이 표적 부위 서열의 전부 또는 일부에 결합하여 절단할 수 있는, 실시 형태 18의 방법.

[0228] 20. 통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 표적 부위 서열로 RNA-유도 엔도뉴클레아제 (RGEN)를 표적화하는 방법으로서, 상기 방법은 상기 효모에 Cas 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 제1 재조합 DNA 구조물 및 RNA 성분의 상류에 리보자임을 암호화하는 DNA 서열을 포함하는 제2 재조합 DNA 구조물을 제공하는 단계를 포함하며, 여기서 제2 재조합 DNA 구조물로부터 전사된 RNA는 자체 촉매 작용으로 리보자임을 제거하여 상기 RNA 성분을 생성하고, RNA 성분과 Cas9 엔도뉴클레아제는 표적 부위 서열의 전부 또는 일부에 결합할 수 있는 RGEN을 형성할 수 있는, 방법.

[0229] 21. RGEN이 표적 부위 서열의 전부 또는 일부에 결합하여 절단할 수 있는, 실시 형태 20의 방법.

[0230] 22. 통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 표적 부위 서열로 RNA-유도 엔도뉴클레아제 (RGEN)를 표적화하는 방법으로서, 상기 방법은 상기 효모에 Cas 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 제1 재조합 DNA 구조물 및 리보자임-RNA 성분 융합 분자를 암호화하는 DNA 서열을 포함하는 적어도 하나의 제2 재조합 DNA 구조물을 제공하는 단계를 포함하며, 여기서 상기 리보자임-RNA 성분 융합 분자와 Cas 엔도뉴클레아제는 표적 부위 서열의 전부 또는 일부에 결합하고, 임의로 절단할 수 있는 RGEN을 형성할 수 있는, 방법.

[0231] 23. RGEN이 표적 부위 서열의 전부 또는 일부에 결합하여 절단할 수 있는, 실시 형태 22의 방법.

[0232] 24. 통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 표적 부위를 변형시키는 방법으로서, 상기 방법은 통상적이지 않은 효모에 Cas 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 제1 재조합 DNA 구조물 및 RNA 성분의 상류에 리보자임을 암호화하는 DNA 서열을 포함하는 제2 재조합 DNA 구조물을 제공하는 단계를 포함하며, 여기서 제2 재조합 DNA 구조물로부터 전사된 RNA는 자체 촉매 작용으로 리보자임을 제거하여 5' 캡을 포함하지 않는 상기 RNA 성분을 생성하고, Cas9 엔도뉴클레아제는 상기 표적 부위에서 단일 또는 이중 가닥 절단을 도입하는, 방법.

[0233] 25. 통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 표적 부위를 변형시키는 방법으로서, 상기 방법은 통상적이지 않은 효모에 Cas 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 제1 재조합 DNA 구조물 및 5' 캡을 포함하지 않는 리보자임-RNA 성분 융합 분자를 암호화하는 DNA 서열을 포함하는 제2 재조합 DNA 구조물을 제공하는 단계를 포함하며, 여기서 상기 리보자임-RNA 성분 융합 분자와 Cas9 엔도뉴클레아제는 상기 표적 부위에 단일 또는 이중 가닥 절단을 도입하는 RGEN을 형성할 수 있는, 방법.

[0234] 26. 통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 다수의 표적 부위를 변형시키는 방법으로서, 상기 방법은 통상적이지 않은 효모에 Cas 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 적어도 하나의 제1 재조합 DNA 구조물 및 적어도 하나의 폴리뉴클레오티드에 작동가능하게 연결된 프로모터를 포함하는 적어도 하나의 제2 재조합 DNA 구조물을 제공하는 단계를 포함하며, 여기서 상기 적어도 하나의 폴리뉴클레오티드는 RNA 성분의 상

류에 리보자임을 포함하는 RNA 분자를 암호화하고, 상기 RNA 분자는 자체 촉매 작용으로 리보자임을 제거하여 상기 RNA 성분을 생성하며, Cas9 엔도뉴클레아제는 상기 표적 부위에서 단일 또는 이중 가닥 절단을 도입하는, 방법.

[0235] 27. 통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 다수의 표적 부위를 변형시키는 방법으로서, 상기 방법은 통상적이지 않은 효모에 Cas9 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 적어도 하나의 제1 재조합 DNA 구조물 및 적어도 하나의 폴리뉴클레오티드에 작동가능하게 연결된 프로모터를 포함하는 적어도 하나의 제2 재조합 DNA 구조물을 제공하는 단계를 포함하며, 여기서 상기 적어도 하나의 폴리뉴클레오티드는 리보자임-RNA 성분 융합 분자를 암호화하고, 상기 리보자임-RNA 성분 융합 분자와 Cas9 엔도뉴클레아제는 상기 표적 부위에 단일 또는 이중 가닥 절단을 도입하는 RGEN을 형성할 수 있는, 방법.

[0236] 28. 상기 표적에서 변형을 갖는 적어도 하나의 통상적이지 않은 효모 세포를 동정하는 단계를 추가로 포함하며, 여기서 변형은 상기 표적 부위에서 하나 이상의 뉴클레오티드의 적어도 하나의 결실, 첨가 또는 치환을 포함하는, 실시 형태 22 내지 실시 형태 25 중 어느 하나의 방법.

[0237] 29. 상기 효모에 공여 DNA를 제공하는 단계를 추가로 포함하며, 여기서 상기 공여 DNA는 관심 대상의 폴리뉴클레오티드를 포함하는, 실시 형태 24 내지 실시 형태 28 중 어느 하나의 방법.

[0238] 30. 상기 표적 부위에 통합된 관심 대상의 폴리뉴클레오티드를 이의 염색체 또는 에피솜에 포함하는 적어도 하나의 효모 세포를 동정하는 단계를 추가로 포함하는, 실시 형태 29의 방법.

[0239] 31. 통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 뉴클레오티드 서열을 편집하는 방법으로서, 상기 방법은 통상적이지 않은 효모에 폴리뉴클레오티드 변형 주형 DNA, Cas 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 제1 재조합 DNA 구조물 및 RNA 성분의 상류에 리보자임을 암호화하는 DNA 서열을 포함하는 제2 재조합 DNA 구조물을 제공하는 단계를 포함하며, 여기서 제2 재조합 DNA 구조물로부터 전사된 RNA는 자체 촉매 작용으로 리보자임을 제거하여 5' 캡을 포함하지 않는 상기 RNA 성분을 생성하고, Cas9 엔도뉴클레아제는 상기 효모의 염색체 또는 에피솜의 표적 부위에 단일 또는 이중 가닥 절단을 도입하며, 상기 폴리뉴클레오티드 변형 주형 DNA는 상기 뉴클레오티드 서열의 적어도 하나의 뉴클레오티드 변형을 포함하는, 방법.

[0240] 32. 통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 뉴클레오티드 서열을 편집하는 방법으로서, 상기 방법은 통상적이지 않은 효모에 폴리뉴클레오티드 변형 주형 DNA, Cas 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 제1 재조합 DNA 구조물 및 5' 캡을 포함하지 않는 리보자임-RNA 성분 융합 분자를 암호화하는 DNA 서열을 포함하는 제2 재조합 DNA 구조물을 제공하는 단계를 포함하며, 여기서 상기 리보자임-RNA 성분 융합 분자와 Cas9 엔도뉴클레아제는 상기 효모의 염색체 또는 에피솜의 표적 부위에 단일 또는 이중 가닥 절단을 도입하는 RGEN을 형성할 수 있고, 상기 폴리뉴클레오티드 변형 주형 DNA는 상기 뉴클레오티드 서열의 적어도 하나의 뉴클레오티드 변형을 포함하는, 방법.

[0241] 33. 통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 뉴클레오티드 서열을 편집하는 방법으로서, 상기 방법은 통상적이지 않은 효모에 Cas9 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 적어도 하나의 제1 재조합 DNA 구조물 및 적어도 하나의 폴리뉴클레오티드에 작동가능하게 연결된 프로모터를 포함하는 적어도 하나의 제2 재조합 DNA 구조물을 제공하는 단계를 포함하며, 여기서 상기 적어도 하나의 폴리뉴클레오티드는 RNA 성분의 상류에 리보자임을 포함하는 RNA 분자를 암호화하고, 상기 RNA 분자는 자체 촉매 작용으로 리보자임을 제거하여 5' 캡을 포함하지 않는 상기 RNA 성분을 생성하며, Cas9 엔도뉴클레아제는 상기 효모의 염색체 또는 에피솜의 표적 부위에 단일 또는 이중 가닥 절단을 도입하고, 상기 폴리뉴클레오티드 변형 주형 DNA는 상기 뉴클레오티드 서열의 적어도 하나의 뉴클레오티드 변형을 포함하는, 방법.

[0242] 34. 통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 뉴클레오티드 서열을 편집하는 방법으로서, 상기 방법은 통상적이지 않은 효모에 적어도 하나의 폴리뉴클레오티드 변형 주형 DNA, Cas 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 적어도 하나의 제1 재조합 DNA 구조물 및 적어도 하나의 폴리뉴클레오티드에 작동가능하게 연결된 프로모터를 포함하는 적어도 하나의 제2 재조합 DNA 구조물을 제공하는 단계를 포함하며, 여기서 상기 적어도 하나의 폴리뉴클레오티드는 5' 캡을 포함하지 않는 리보자임-RNA 성분 융합 분자를 암호화하고, 상기 리보자임-RNA 성분 융합 분자와 Cas9 엔도뉴클레아제는 상기 효모의 염색체 또는 에피솜의 표적 부위에 단일 또는 이중 가닥 절단을 도입하는 RGEN을 형성할 수 있으며, 상기 폴리뉴클레오티드 변형 주형 DNA는 상기 뉴클레오티드 서열의 적어도 하나의 뉴클레오티드 변형을 포함하는, 방법.

[0243] 35. 제1 재조합 DNA와 제2 재조합 DNA가 동일한 플라스미드 상에 위치하는, 실시 형태 24 내지 실시 형태 34 중

어느 하나의 방법.

[0244] 36. 제1 재조합 DNA와 제2 재조합 DNA가 별도의 플라스미드 상에 위치하는, 실시 형태 24 내지 실시 형태 34 중 어느 하나의 방법.

[0245] 37. 통상적이지 않은 효모 내의 염색체 또는 에피솜 상의 뉴클레오티드 서열을 침묵시키는 방법으로서, 상기 방법은 통상적이지 않은 효모에 불활성화된 Cas9 엔도뉴클레아제를 암호화하는 DNA 서열을 포함하는 적어도 하나의 제1 재조합 DNA 구조물 및 적어도 하나의 폴리뉴클레오티드에 작동가능하게 연결된 프로모터를 포함하는 적어도 하나의 제2 재조합 DNA 구조물을 제공하는 단계를 포함하며, 여기서 상기 적어도 하나의 폴리뉴클레오티드는 5' 캡을 포함하지 않는 리보자임-RNA 성분 융합 분자를 암호화하고, 상기 리보자임-RNA 성분 융합 분자와 불활성화된 Cas9 엔도뉴클레아제는 상기 효모의 염색체 또는 에피솜 내의 상기 뉴클레오티드 서열에 결합하는 RGEN을 형성함으로써 상기 뉴클레오티드 서열의 전사를 차단할 수 있는, 방법.

[0246] 38. 통상적이지 않은 효모에서 유전자 변형을 위한 다수의 가이드 RNA를 고효율로 생성하는 방법으로서, 상기 방법은 하기 단계를 포함하는, 방법:

[0247] a) 5' 에서 3' 순서로 리보자임을 암호화하는 제1 DNA 서열, 역선별제를 암호화하는 제2 DNA 서열, 가이드 RNA의 CER 도메인을 암호화하는 제3 DNA 서열 및 종결 서열에 작동가능하게 연결된 프로모터를 포함하는 재조합 DNA 구조물을 제공하는 단계;

[0248] b) (a)의 재조합 DNA 구조물에 적어도 하나의 올리고뉴클레오티드 듀플렉스를 제공하는 단계로서, 상기 올리고뉴클레오티드 듀플렉스는 가이드 RNA 표적 서열의 가변 표적화 도메인 (VT)을 암호화할 수 있는 DNA 서열을 포함하는 제1 단일 가닥 올리고뉴클레오티드와 가변 표적화 도메인을 암호화하는 DNA 서열에 상보적인 서열을 포함하는 제2 단일 가닥 올리고뉴클레오티드를 조합하는 것으로부터 유래되는 단계;

[0249] c) (a)의 역선별제를 (b)의 적어도 하나의 올리고듀플렉스와 교환함으로써, 각각이 가이드 RNA의 가변 표적화 도메인을 암호화할 수 있는 DNA 서열을 포함하는 재조합 DNA 구조물의 라이브러리를 생성하는 단계; 및

[0250] d) (c)의 재조합 DNA 구조물의 라이브러리를 전사함으로써, 리보자임-가이드 RNA 분자의 라이브러리를 생성하는 단계.

[0251] 39. 리보자임-가이드 RNA 분자가 리보자임과 리보자임 상류에 임의의 RNA 서열을 자체 촉매 작용으로 제거하도록 상기 분자의 라이브러리를 유도하여, 5' 캡을 함유하지 않는 가이드 RNA 분자의 라이브러리를 생성하는 단계를 추가로 포함하는, 실시 형태 38의 방법.

[0252] 40. 리보자임-가이드 RNA 분자가 리보자임 상류에 임의의 RNA 서열을 절단하도록 상기 분자의 라이브러리를 유도하여, 5' 캡을 함유하지 않는 리보자임-gRNA 융합 분자를 생성하는 단계를 추가로 포함하는, 실시 형태 38의 방법.

[0253] 41. (ii) 리보자임과 가이드 RNA를 암호화하는 뉴클레오티드 서열에 작동가능하게 연결된 (i) 폴리머라제-II 프로모터를 포함하는 재조합 DNA 서열로서, 여기서 상기 리보자임은 상기 가이드 RNA의 상류이고, (ii)의 뉴클레오티드 서열로부터 전사된 RNA는 자체 촉매 작용으로 리보자임을 제거하여 상기 가이드 RNA를 생성하며, 상기 가이드 RNA는 통상적이지 않은 효모의 게놈 내의 표적 부위를 인식하고, 결합하며, 임의로 절단할 수 있는 RGEN을 형성할 수 있는, 재조합 DNA 서열.

[0254] 42. 리보자임과 가이드 RNA를 포함하는 재조합 RNA 서열로서, 여기서 상기 리보자임은 상기 가이드 RNA의 상류이고, 자체 촉매 작용으로 상기 리보자임을 제거하여 상기 가이드 RNA를 생성하며, 상기 가이드 RNA는 통상적이지 않은 효모의 게놈 내의 표적 부위를 인식하고, 결합하며, 임의로 절단할 수 있는 RGEN을 형성할 수 있는, 재조합 RNA 서열.

[0255] 43. (ii) 리보자임과 가이드 RNA를 암호화하는 뉴클레오티드 서열에 작동가능하게 연결된 (i) 폴리머라제-II 프로모터를 포함하는 재조합 DNA 서열로서, 여기서 상기 리보자임은 상기 가이드 RNA의 상류이고, (ii)의 뉴클레오티드 서열로부터 전사된 RNA는 리보자임-가이드 RNA 융합 분자를 생성하며, 상기 리보자임-가이드 RNA 융합 분자는 통상적이지 않은 효모의 게놈 내의 표적 부위를 인식하고, 결합하며, 임의로 절단할 수 있는 RGEN을 형성할 수 있는, 재조합 DNA 서열.

[0256] 44. 리보자임-가이드 RNA 융합 분자를 포함하는 재조합 RNA 서열로서, 여기서 상기 리보자임-가이드 RNA 융합 분자는 통상적이지 않은 효모의 게놈 내의 표적 부위를 인식하고, 결합하며, 임의로 절단할 수 있는 RGEN을 형

성할 수 있는, 재조합 RNA 서열.

[0257] 실시예

[0258] 개시된 발명은 하기 실시예에서 추가로 정의된다. 이들 실시예는 본 발명의 소정의 바람직한 측면을 나타내지만 단지 예로서만 주어짐을 이해해야 한다. 상기 논의 및 이들 실시예로부터, 당업자는 본 발명의 본질적인 특징을 확인할 수 있으며, 본 발명의 사상 및 범주로부터 벗어나지 않고서, 본 발명을 다양하게 변경 및 변형하여 본 발명이 다양한 용도와 조건에 적합하도록 할 수 있다.

[0259] 실시예 1

[0260] 야로위아의 Pol III 프로모터로부터 발현된 sgRNA는 Cas9를 표적 부위로 유도하지 않으며 DNA 절단을 매개하지 않음

[0261] 이러한 실시예는 Leu2 유전자좌를 표적으로 하는 야로위아 리폴리티카에서 sgRNA와 Cas9 단백질을 발현하도록 설계된 벡터 및 카세트를 개시한다. 이러한 효모에서 생성된 sgRNA와 Cas9가 상호작용하고, 표적 부위를 찾아 절단할 수 있는 경우, 오류가 발생하기 쉬운 비상동 말단 결합 (NHEJ)을 통해 표적 부위에 돌연변이가 발생될 것이다.

[0262] 도 1은 가변 표적화 도메인 (VT) (가이드 서열)과 Cas 엔도뉴클레아제 인식 도메인 (CER) 두 영역을 함유하는 단일 RNA 분자인 sgRNA 분자를 도시한다. VT 영역은 표적화된 핵산 분자와 동일성을 갖는 RNA 폴리뉴클레오티드의 20mer일 수 있다. VT 도메인은 PAM 모티프 (예를 들어, NGG, 서열 번호:47)의 5'에 위치하는 표적 부위에서 절단을 위한 표적 부위를 특정한다. CER 도메인은 Cas9 단백질과 상호작용하여, VT 도메인이 상호작용하도록 하여 Cas9 단백질 절단을 유도한다 (문헌[Jinek et al., *Science* 337:816-821]). VT 도메인과 CER 도메인 모두가 sgRNA의 기능을 위해 필요하다.

[0263] 야로위아의 LEU2 유전자좌의 암호화 영역에서 3개의 개별 표적 부위 (Leu2-1, Leu2-2, Leu2-3)로 Cas9를 표적화하는 VT 도메인을 암호화하는 DNA 서열을 표 3에 열거한다. 표 3은 또한 야로위아 CAN1 유전자좌의 암호화 영역을 표적으로 하는 VT 도메인을 암호화하는 DNA 서열을 열거한다.

[0264] [표 3]

Cas9 로 야로위아의 LEU2 또는 CAN1 유전자좌를 표적화하기 위한 sgRNA VT 도메인을 암호화하는 DNA 서열

Leu2-1[a] (서열 번호:2)	TCCAAGAAGATTGTTCTTCT
Leu2-2[a] (서열 번호:3)	CTCCGTCATCCCCGGTTCTC
Leu2-3[a] (서열 번호:4)	CGGCGACTTCTGTGGCCCCG
Can1-1[b] (서열 번호:17)	TCAAACGATTACCCACCCTC

[a] Leu2-1, Leu2-2 및 Leu2-3 에 의해 표적화된 LEU2 유전자 부위는 PAM 부위로서 각각 CGG, TGG 또는 AGG 를 갖는다.

[b] Can1-1 에 의해 표적화된 CAN1 유전자 부위는 PAM 부위로서 CGG 를 갖는다.

[0265]

[0266] 표 3의 LEU2-표적화 DNA 서열 각각을 스트렙토코커스 피오제네스 Cas9 단백질과 상호작용하는 CER 도메인 (서열 번호:1)을 암호화하는 DNA 서열에 개별적으로 융합시켜, CER 도메인과 VT 도메인을 모두 갖는 완전 sgRNA를 암호화하는 DNA 서열을 생성하였다 (서열 번호:1은 5'에서 3' 방향으로 서열 번호:56의 tracrRNA 메이트 서열, 서열 번호:43의 루프 형성 서열 (GAAA) 및 서열 번호:58의 tracrRNA 서열을 포함함에 유의한다). 세포의 핵에서 이러한 sgRNA을 발현시키고, 핵외수송과 5' 변형 시스템을 회피하기 위하여, sgRNA를 암호화하는 DNA 서열을 사카로마이세스 세레비시아 (Snr52 [서열 번호:5] 또는 Rpr1 [서열 번호:6]) 또는 야로위아 리폴리티카 (Snr52 [서열 번호:7])로부터의 RNA Pol III 프로모터의 제어 하에 두었다. 구체적으로, Sc Snr52를 Leu2-1에 융합시키고, Sc Rpr1을 Leu2-2에 융합시키며, Yl Snr52를 Leu2-3에 융합시켰다. 각각의 sgRNA를 암호화하는 DNA 서열의 3' 말단을 사카로마이세스 세레비시아 (서열 번호:8)의 Sup4 유전자로부터의 강력한 종결자에 융합시켰다. 따라서, 3개의 상이한 Pol III-유도된 sgRNA 카세트를 제조하였다.

[0267] 표준 기술에 따라 스트렙토코커스 피오제네스 M1 GAS (SF370)로부터의 Cas9 유전자의 오픈 리딩 프레임을 야로위아에서의 발현을 위해 코돈-최적화하여, 서열 번호:9를 생성하였다. 유인원 바이러스 40 (SV40) 1부분 핵 국재화 신호 (NLS)와 짧은 링커 (4개의 아미노산)를 암호화하는 DNA 서열을 서열 번호:9의 마지막 센스 코돈 다음에 혼입하여 서열 번호:10을 만들었다. 서열 번호:10은 서열 번호:11에 나타낸 아미노산 서열을 암호화한다. 서열 번호:11의 마지막 7개의 아미노산은 첨가된 NLS를 암호화하면 반면, 서열 번호:11의 1369 내지 1372 위치

의 잔기는 첨가된 링커를 암호화한다. 야로위아 코돈-최적화된 Cas9-NLS 서열 (서열 번호:10)을 표준 분자생물학 기술에 의해 야로위아 항시성 프로모터인 FBA1 (서열 번호:12)로 융합시켰다. 항시성 FBA1 프로모터, 야로위아 코돈-최적화된 Cas9 및 SV40 NLS를 함유하는 야로위아 코돈-최적화된 Cas9 발현 카세트 (서열 번호:13)의 예가 도 2a에 예시되어 있다. 이러한 Cas9 발현 카세트 (서열 번호:13)를 플라스미드 pZUF로 클로닝하여, 구조물 pZUFCas9 (도 3a, 서열 번호:14)를 만들었다.

[0268] (상기) 각각의 sgRNA 발현 카세트를 pZUFCas9 (서열 번호:14)의 PacI/ClaI 부위로 개별적으로 클로닝하여, 효모 세포를 야로위아 코돈-최적화된 Cas9 발현 카세트 및 Pol III-유도된 sgRNA 발현 카세트로 공동형질전환(co-transform)하는 데 사용될 수 있는 pZUFCas9/sgRNA 구조물을 만들었다. 이러한 구조물의 예는 pZUFCas9/PolIII-sgRNA (도 3b)이며, 이는 야로위아에서 Leu2-3을 표적화하기 위한 Y1 Snr52 - sgRNA 발현 카세트를 함유한다.

[0269] 우라실 영양요구성 야로위아 리폴리티카 세포를 200 ng의 플라스미드 pZUFCas9 (서열 번호:14) 또는 특정 pZUFCas9/sgRNA (예를 들어, pZUFCas9/PolIII-sgRNA, 도 3b)로 형질전환시키고, 우라실이 없는 완전 최소 플레이트 (CM-ura) 상에서 우라실 원영양성(prototroph)을 선별하였다. CM-ura 플레이트 상에 발생하는 콜로니를 류신이 없는 완전 최소 플레이트 (CM-leu) 상에서 류신 영양요구성에 대하여 스크리닝하였다. 우라실 원영양성 형질전환체 중 어느 것도 류신 영양요구성을 나타내지 않았다. 이러한 결과는 야로위아 코돈-최적화된 Cas9 및 Pol III 프로모터-유도된 sgRNA가 발현되지 않았고/거나, 생성되지 않았으며/거나, 상호작용하지 않았고/거나, DNA를 표적하지 않았으며/거나 DNA를 절단하지 않았음을 제시한다. 이러한 실험이 류신 영양요구성을 생성하는 경우, 이러한 결과는 Cas9/sgRNA 복합체가 Leu2 암호화 영역을 표적화하고 절단하여 오류가 발생하기 쉬운 NHEJ를 유도하고, 결과적으로 삽입-결실을 형성하여, 프레임시프트 돌연변이를 일으킨다는 것을 나타낼 가능성이 있다.

[0270] 따라서, sgRNA의 Pol III-유도된 발현은 야로위아에서 기능적 Cas9-sgRNA 복합체를 제공하는 데 유용하지 않을 수 있다.

[0271] 실시예 2

[0272] DNA 폴리머라제 II 프로모터에 의해 유도된 5'- 및 3'-리보자임을 포함하는 야로위아-최적화된 sgRNA 발현 카세트

[0273] 이러한 실시예는 야로위아에서 발현 및 Cas9-매개 표적화를 위해 최적화된 sgRNA를 개시한다. 이러한 발현을 위해 사용된 각각의 카세트는 5'-리보자임 및 3'-리보자임 (리보자임-sgRNA-리보자임 또는 RGR)에 융합된 sgRNA의 전사를 유도하기 위한 Pol II 프로모터를 포함하였다. 5' 및 3' 리보자임을 제공하여 sgRNA로부터 Pol II 프로모터-관련 전사물 변형, 예컨대 5' 캡 구조를 제거하여, sgRNA 서열만을 남겼다. 이러한 발현 카세트는 sgRNA 발현을 위한 더 넓은 프로모터 선택을 허용한다. 또한, 이러한 카세트로부터 전사된 sgRNA는 5'-캡 구조가 없기 때문에 핵외수송의 대상이 아니다. 이러한 특징은 야로위아 세포에서 sgRNA의 강한 발현을 허용하여, Cas9 엔도뉴클레아제를 생체 내 게놈의 표적화 영역으로 유도할 수 있다.

[0274] sgRNA 서열에 5' 망치머리 (HH) 및 3' 델타 간염 바이러스 (HDV) 리보자임을 첨가함으로써 일부 RNA 폴리머라제 (예를 들어 Pol II)에 의해 전사된 프로모터에서 발생하는 전사후 변형을 고려하지 않고 임의의 프로모터로부터 sgRNA의 발현을 허용하고, sgRNA 발현을 위한 현재의 프로모터의 제한된 선택을 피한다. 이러한 sgRNA가 발현되는 경우, 예비-sgRNA 전사물에 존재하는 리보자임은 자가절단(autocleave)됨으로써, 전사물로부터 분리되어, 변형되지 않은 sgRNA를 남긴다.

[0275] 시험된 각각의 sgRNA의 경우, sgRNA를 암호화하는 DNA 서열을 (i) 이의 5'-말단에서 5' HH 리보자임을 암호화하는 서열 (서열 번호:15)로 (ii) 이의 3'-말단에서 3' HDV 리보자임을 암호화하는 서열 (서열 번호:16)로 융합시켰다. HH 리보자임의 5'-결합은 HH 리보자임의 처음 6개의 뉴클레오티드가 sgRNA의 VT 영역 (가이드 서열)의 처음 6개의 뉴클레오티드의 역 상보서열이 되도록 하였다. 각각의 리보자임-플랭킹된 예비-sgRNA (RGR)를 표준 분자생물학 기술을 사용하여 FBA1 프로모터 (서열 번호:12)에 융합시켜, 야로위아-최적화된 sgRNA 발현 카세트 (도 2b에 도시된 최종 카세트)를 얻었다. 이러한 카세트의 예시적인 서열이 서열 번호:18에 도시되어 있으며, 이는 RGR을 암호화하는 서열 (HH-sgRNA-HDV)에 작동가능하게 연결된 FBA1 프로모터 (서열 번호:12)를 포함하고, 여기서 sgRNA는 이의 CER 도메인으로서 서열 번호:17 (Can1-1) 및 서열 번호:1에 의해 암호화되는 VT 도메인을 포함한다 (서열 번호:18, pRF38 (서열 번호:19) 및 pRF84 (서열 번호:41)의 CER 도메인-암호화 영역 각각이 서열 번호:1의 위치 73-74 (CER 도메인)에 상응하는 잔기 위치 사이에 있는, 첨가된 'TGG'를 가짐에 유의한다).

이러한 VT 도메인은 야로위아 CAN1 유전자 오픈 리딩 프레임의 암호화 영역에 있는 부위를 표적으로 한다 (진뱅크 수탁 번호 NC_006068, YALI0B19338g, 염색체 B의 ~bp 2557513-2559231). 암호화된 HH 리보자임의 처음 6개의 잔기는 sgRNA의 처음 6개의 잔기 (즉, VT 도메인의 처음 6개의 잔기)에 상보적이다. 서열 번호:18의 서열 번호:12 (FBA1 프로모터) 바로 다음에 3개의 잔기 (ATG)가 존재하며, 이는 예비-sgRNA의 발현과 리보자임-매개 자체촉매 작용에 영향을 주지 않는 것으로 여겨짐에 유의한다. 서열 번호:18을 pRF38 (도 3c, 서열 번호:19)로 명명된 구조물로 클로닝하였다.

[0276] 따라서, 5' 및 3' pol II 프로모터-관련 전사물 변형없이 sgRNA를 발현시키기 위한 DNA 카세트를 제조하였다. 이러한 유형의 카세트를 야로위아에서의 Cas9 유전자 표적화를 위해 실시예 3에서 사용하였다.

[0277] 실시예 3

[0278] 야로위아-최적화된 sgRNA가 sgRNA/Cas9 엔도뉴클레아제 시스템에서 사용되어 염색체 DNA를 절단할 수 있음

[0279] 이러한 실시예는 실시예 2에 기재된 바와 같은 야로위아-최적화된 sgRNA 발현 카세트를 사용하여 야로위아에서 염색체 DNA를 인식하여 절단하기 위해 Cas9와 함께 작용할 수 있는 sgRNA를 발현하는 것을 개시한다. 이러한 절단은 절단 부위에서 오류가 발생하기 쉬운 NHEJ DNA 수복으로 인해 예측되는 DNA 절단 부위의 영역에서 돌연변이가 발생한 것에 의해 나타났다.

[0280] 야로위아 리폴리티카의 CAN1 유전자는 절단의 표적이었다. 야로위아 형질전환체에서의 CAN1의 성공적인 표적화를 각각 표현형 (카나바닌 저항) 및 돌연변이 빈도 및 스펙트럼에 대한 서열 분석으로 조사하였다.

[0281] Ura^- 야로위아 리폴리티카 세포 (균주 ATCC 20362로부터 직접 유래된 우라실 영양요구성인 균주 Y2224가 본 명세서에 참고로 포함되는 미국 특허 출원 공개 제2010/0062502호에 개시되어 있음)를 리튬 이온-매개 형질전환 (문헌[Ito et al., J. *Bacteriology* 153:163-168])에 의해 pZUFCas9 (도 3a, 서열 번호:14) 및 CAN1 유전자좌를 표적화하기 위한 야로위아-최적화된 RGR 예비-sgRNA 카세트 (서열 번호:18에 포함됨)를 함유하는 pRF38 (도 3c, 서열 번호:19)로부터 증폭된 선형 PCR 생성물로 공동형질전환시켰다. 이러한 PCR 증폭에 사용된 프라이머는 서열 번호:20 (정방향) 및 서열 번호:21 (역방향)이었다. pZUFCas9 (서열 번호:14)로만 형질전환된 Ura^- 야로위아 리폴리티카 세포 (Y2224)는 음성 대조군으로 역할을 하였다. pZUFCas9 (서열 번호:14) 및 RGR 예비-sgRNA 발현 카세트로 형질전환된 세포를 CM-ura 배지 상에서 우라실 원영양성으로서 선별하였다. CM-ura 플레이트를 우라실과 아르기닌이 없고, 60 μg/ml의 독성 아르기닌 유사체인 카나바닌 (Cm+can)이 보충된 완전 최소 배지 상으로 복제 평판(replica-plating)하여 CAN1 유전자에 기능 상실 돌연변이를 함유하는 세포를 스크리닝하였다. 기능적 CAN1 유전자를 가진 세포는 카나바닌을 세포로 수송하여 세포 사멸을 일으킬 수 있다. CAN1 유전자에 기능 상실 대립유전자를 가진 세포는 카나바닌을 수송하지 못하며 CM+can 플레이트 상에서 성장할 수 있다.

[0282] 카나바닌 내성의 표현형 스크리닝에 의한 회복된 기능 상실 돌연변이체의 빈도는 Cas9 단독으로 형질전환된 세포에서는 0이었다 (도 4). 그러나, Cas9가 RGR 예비-sgRNA 발현 카세트로 공동형질전환된 경우, 카나바닌-내성 형질전환체의 빈도가 10%로 증가하였다 (도 4).

[0283] 카나바닌-내성 콜로니의 CAN1 유전자좌를 정방향 (서열 번호:22) 및 역방향 (서열 번호:23) PCR 프라이머를 사용하여 증폭시켰다. PCR 생성물을 자이모클린(Zymoclean)™ 및 농축기 컬럼 (캘리포니아주 얼바인 소재의 자이모 리서치(Zymo Research))을 사용하여 정제하였다. PCR 생성물을 서열 분석 프라이머 서열 번호:24를 사용하여 서열 분석하였다 (생어법). 서열을 표적 부위를 함유하는 야생형 (WT) 야로위아 CAN1 암호화 서열과 정렬시켰다 (도 5). Cas9와 sgRNA를 모두 발현하는 세포의 CAN1 유전자좌에서 주된 기능 상실 돌연변이 (서열분석된 분리주의 73%)는 Cas9 절단 부위에서의 -1 프레임시프트 돌연변이었다 (도 5). 보다 적은 수의 다른 결실 및 삽입이 CAN1 유전자좌에서의 돌연변이의 나머지를 구성하였다. 총, 90%의 돌연변이가 작은 결실 또는 삽입이었다 (도 5). 드물게, 다른 염색체로부터의 소량의 서열의 삽입 (4%), 절단 부위에서 야로위아-최적화된 sgRNA 발현 카세트의 삽입 (1.5%) 또는 더 큰 결실 (1%)과 같은 다른 이벤트가 발생하였다. 스크리닝한 3.5%의 카나바닌-내성 콜로니가 CAN1 유전자좌에서 서열 분석으로 결정되지 않은 복잡한 재배열을 가졌다. 전체적으로, CAN1 표적 부위에서 관찰된 돌연변이는 Cas9/sgRNA 복합체에 의해 생성된 절단을 수복하기 위해 오류가 발생하기 쉬운 NHEJ가 세포에서 사용되었음을 나타낸다.

[0284] (i) CAN1-특이적 Cas9 엔도뉴클레아제가 발현되도록 형질전환된 세포에서 카나바닌-내성 콜로니의 증가된 빈도 및 (ii) 카나바닌-내성 돌연변이는 예측되는 Cas9 절단 부위에서의 오류가 발생하기 쉬운 NHEJ 이벤트에 의한

것임을 나타내는 서열 분석 데이터 모두에 의해 실시예 2에 기재된 야로위아-최적화된 Cas9 및 RGR 예비-sgRNA 발현 카세트가 야로위아 염색체 DNA를 절단하여 돌연변이를 생성함을 확인하였다.

[0285] 따라서, RNA 성분의 5' 캡이 리보자임의 자체 촉매 작용으로 제거되는, 5'-캡이 없는 RGEN (예를 들어, Cas9)의 RNA 성분 (예를 들어, sgRNA)을 발현하는 것은 통상적이지 않은 효모에서 DNA 서열의 RGEN-매개 표적화를 허용한다.

[0286] 실시예 4

[0287] (3' 리보자임없이) 5'-리보자임을 사용하여 발현된 야로위아-최적화된 sgRNA는 염색체 DNA를 절단하기 위한 sgRNA/Cas9 엔도뉴클레아제 시스템에 유용함

[0288] 이러한 예에서, 3' 리보자임없이 5' HH 리보자임만을 함유하는 야로위아-최적화된 카세트로부터 생성된 sgRNA의 기능을 평가하여, sgRNA가 Cas9와 상호작용하고, DNA 표적 서열을 인식하며, Cas9에 의한 DNA 절단을 유도하고, 오류가 발생하기 쉬운 NHEJ에 의한 돌연변이를 유도할 수 있는지를 결정하였다.

[0289] Pol II 프로모터로부터 전사된 RNA는 이들의 5' 및 3' 말단 모두에서 심하게 가공되고 변형되어, Pol II 프로모터로부터 기능적 sgRNA를 생성하기 위해 5' 및 3' 말단을 모두 절단해야 한다는 것을 시사한다. 플랭킹 영역을 갖는 시험관 내에서 생성된 sgRNA가 (i) 5'-플랭킹 영역이 존재하는 경우 비기능적이고 (ii) 3' 플랭킹 영역이 존재하는 경우 기능적으로 상당히 손상된다는 것이 이전부터 밝혀져 왔다 (문헌[Gao et al., J. *Integr. Plant BioL.* 56:343-349]). 5' 리보자임과 3' 플랭킹 영역도 함유하는 예비-sgRNA가 Cas9와 함께 사카로마이세스 세레비시아로 발현되는 경우, sgRNA는 절단을 위한 표적 부위로 Cas9를 유도하도록 기능하지 않았다 (문헌[Gao et al., *ibid*]).

[0290] (3'-위치한 리보자임이 없는) 5' 리보자임-플랭킹된 sgRNA가 통상적이지 않은 효모에서 기능할 수 있는지 시험하기 위해, 5'에서 3' 방향으로, S. 세레비시아 Sup4 유전자 (서열 번호:8)로부터의 강력한 전사 종결자에 융합된 Can1-1 표적 부위 (서열 번호:17)를 표적화하는 sgRNA (서열 번호:70의 예)를 암호화하는 서열에 융합된 HH 리보자임 (서열 번호:15)에 융합된 FBA1 프로모터 (서열 번호:12)를 함유하는 야로위아-최적화된 sgRNA 발현 카세트 (서열 번호:25)를 구축하였다 (이러한 카세트는 RG [리보자임-sgRNA] RNA를 발현하는 것으로 특징지어질 수 있다). RG 발현 카세트에서 암호화된 sgRNA는 CER 도메인 (서열 번호:1)에 연결된 서열 번호:17에 상응하는 VT 도메인을 포함한다. 암호화된 HH 리보자임의 처음 6개의 잔기는 sgRNA의 처음 6개의 잔기 (즉, VT 도메인의 처음 6개의 잔기)에 상보적이다. 서열 번호:25의 서열 번호:12 (FBA1 프로모터) 바로 다음에 3개의 잔기 (ATG)가 존재하며, 이는 예비-sgRNA의 발현과 리보자임-매개 자체촉매 작용에 영향을 주지 않는 것으로 여겨짐에 유의한다. 이러한 야로위아-최적화된 RG 발현 카세트 (서열 번호:25)가 도 2c에 예시되어 있다.

[0291] Cas9와 상호작용하고, Cas9를 Cas9에 의한 절단을 위한 DNA 표적 서열로 유도할 수 있는 sgRNA를 발현하는 야로위아-최적화된 RG 카세트의 능력을 시험하기 위해, RG 구조물 (서열 번호:25) 또는 RGR 구조물 (서열 번호:18, 실시예 2)을 함유하는 PCR 생성물을 pZUFCas9 (서열 번호:14)과 함께 Ura^- 야로위아 리폴리티카 세포 (Y2224)로 리튬 이온-매개 형질전환에 의해 공동형질전환시켰다 (문헌[Ito et al., *ibid*]). Ura^+ 형질전환체를 CM+can 플레이트 상으로 복제 평판하여 카나바닌-내성 세포 (실시예 3에서와 같이)에 대해 스크리닝하였는데, 여기서 RG 또는 RGR 예비-sgRNA로부터 생성된 sgRNA는 Cas9를 유도하는 데 작용하여 CAN1 표적 서열을 절단하여 NHEJ를 통한 오류가 발생하기 쉬운 수복을 초래하였다. 야로위아-최적화된 RG 또는 RGR 카세트가 Cas9 매개 절단을 표적 부위로 유도하는 빈도는 동일하며 (도 6), 이는 S. 세레비시아를 사용한 문헌[J. *Integr. Plant BioL.* 56:343-349]의 결과와 대조적으로 야로위아에서 효율적인 Cas9/sgRNA 표적 절단 및 돌연변이를 위해 3' 리보자임이 필요하지 않음을 나타내었다.

[0292] 이러한 예는 통상적이지 않은 효모, 예컨대 야로위아에서 리보자임 전략을 사용하는 경우, Pol II 프로모터로부터 기능적 sgRNA를 생성하기 위해 5'-플랭킹 리보자임만이 필요할 수 있음을 입증한다. 이러한 결과는 Cas9에 의한 표적 서열의 효율적인 절단 및 돌연변이를 위해 5' 및 3' 리보자임이 모두 필요한 통상적인 효모인 S. 세레비시아에서 관찰된 것과 대조적이다 (문헌[Gao et al., *ibid*]).

[0293] 따라서, 이러한 예는 RNA 성분의 5' 캡이 리보자임의 자체 촉매 작용으로 제거되는, 5'-캡이 없는 RGEN (예를 들어, Cas9)의 RNA 성분 (예를 들어, sgRNA)을 발현하는 것은 통상적이지 않은 효모에서 DNA 서열의 RGEN-매개 표적화를 허용함을 추가로 입증한다.

[0294] 실시예 5

[0295] Cas9/sgRNA-유도된 DNA 이중 가닥 절단의 상동 재조합 (HR) 수복을 촉진하기 위한 선형 폴리뉴클레오티드 변형 주형의 사용

[0296] 이러한 실시예는 야로위아에서 야로위아-최적화된 Cas9 및 예비-sgRNA 발현 카세트를 발현시킴으로써 생성된 이중 가닥 절단 (DSB)을 수복하기 위해 선형 폴리뉴클레오티드 변형 주형 DNA 서열을 사용하는 HR 시스템(machinery)의 능력을 시험하는 방법을 개시한다. 각각 염색체 DNA에서 Cas9/sgRNA 표적화 부위 외부의 영역에 상동성인 5'- 및 3'-암 서열을 갖는 3개의 상이한 선형 주형 서열을 제조하였다.

[0297] 처음 2개의 유형의 폴리뉴클레오티드 변형 주형 서열을 상보적인 합성 올리고뉴클레오티드로부터 제조하였다. 상보적인 올리고뉴클레오티드를 어닐링한 후 에탄올 침전에 의해 정제하였다.

[0298] 상보성 올리고뉴클레오티드 (서열 번호:28 및 29)를 사용하여 제1 폴리뉴클레오티드 변형 주형을 제조하였고, 20-뉴클레오티드 Can1-1 표적 부위 (서열 번호:17), 3-뉴클레오티드 PAM 도메인 및 Can1-1 표적 부위의 바로 상류에 있는 2개의 뉴클레오티드를 결실시킴으로써 CAN1 유전자에서 -1bp 프레임시프트를 일으키는 8개의 코돈 및 1개의 염기쌍을 결실시키도록 설계하였다. 서열 번호:28과 이의 역 상보서열인 서열 번호:29를 어닐링함으로써 제1 폴리뉴클레오티드 변형 주형을 조립하였다. 제1 공여 DNA의 상동성 암 (각각 약 50-bp)은 서로 바로 옆에 위치하며; 이들 사이에는 이종 서열이 없다.

[0299] 상보성 올리고뉴클레오티드 (서열 번호:30 및 31)을 사용하여 제2 폴리뉴클레오티드 변형 주형을 제조하였고, CAN1 오픈 리딩 프레임에 2개의 인-프레임(in-frame) 번역 종결 코돈 (즉, 넌센스 돌연변이)을 생성하도록 설계하였다. 또한, Can1-1 표적 부위 하류의 PAM 서열 (CGG를 ATG로 대체함) 및 시드(seed) 서열의 제1 뉴클레오티드 (즉, 서열 번호:17의 Can1-1 표적 서열의 마지막 잔기) (C를 G로 대체함)를 파괴하도록 설계되었다. 서열 번호:30과 이의 역 상보서열인 서열 번호:31을 어닐링함으로써 이러한 폴리뉴클레오티드 변형 주형을 생성하였다. 상기로부터 얻을 수 있듯이, 제2 공여 DNA의 상동성 암 (각각 약 50-bp)은 이종 서열의 단지 수개의 염기쌍에 의해 이격된다.

[0300] 2개의 PCR 생성물을 생성함으로써 부분적으로 제3 폴리뉴클레오티드 변형 주형을 제조하였다. PCR 생성물 (서열 번호:32, 서열 번호:33 [정방향] 및 서열 번호:34 [역방향]의 프라이머를 사용하여 야로위아 리폴리티카 ATCC 20362 게놈 DNA로부터 증폭됨) 중 하나에서, 서열 번호:32의 위치 638은 CAN1 오픈 리딩 프레임 시작 코돈 상류의 뉴클레오티드 3 bp에 상응한다. 역방향 프라이머 (서열 번호:34)는 CAN1 오픈 리딩 프레임 하류의 37 bp에 위치하는 서열에 상보적인 17개의 뉴클레오티드를 첨가한다. 제2 PCR 생성물 (서열 번호:35, 서열 번호:36 [정방향] 및 서열 번호:37 [역방향]의 프라이머를 사용하여 야로위아 리폴리티카 ATCC 20362 게놈 DNA로부터 증폭됨)은 CAN1 오픈 리딩 프레임의 종결 코돈 하류의 14번째 염기쌍에서 시작하는 637개의 염기쌍(637 base pairs starting 14 base pairs)을 포함한다. 정방향 프라이머 (서열 번호:36)는 CAN1 오픈 리딩 프레임 상류의 종결되는 2개의 염기쌍(ending 2 base pairs) 영역에 상보적인 20개의 뉴클레오티드를 첨가한다. 상류 (서열 번호:32) PCR 생성물 및 하류 PCR 생성물 (서열 번호:35) 모두를 자이모클린™ 및 농축기 컬럼을 사용하여 정제하였다. 이러한 PCR 생성물 (각각 10 ng)을 새로운 PCR 반응에서 혼합하였다. 생성물 상류의 3'-최외각(most) 37개의 뉴클레오티드는 생성물 하류의 5'-최외각 37개의 뉴클레오티드와 동일하다. 상류 및 하류 단편을 사용하여 서로를 프라이밍하여, 상류 및 하류 서열 모두를 함유하는 중첩 말단으로부터의 합성에 의해 단일 생성물 (서열 번호:38)을 생성하였다 (문헌[by Horton et al., *Biotechniques* 54:129-133]에 기재된 기술). 서열 번호:38 공여 DNA의 상동성 암 (각각 600-bp 이상)은 서로 바로 옆에 위치하며; 이들 사이에는 이종 서열이 없다. 이러한 폴리뉴클레오티드 변형 주형은 Can1-1 표적 부위의 Cas9/sgRNA-매개 이중 가닥 절단의 영역에서 전체 CAN1 오픈 리딩 프레임을 포함하는 큰 결실을 가능하게 할 수 있다.

[0301] Ura$^-$ 야로위아 리폴리티카 세포 (Y2224)를 상기 리튬 이온 형질전환 방법을 사용하여 (i) pZUFCas9 (서열 번호:14), (ii) 1 ㎍의 야로위아-최적화된 RGR 예비-sgRNA 발현 카세트 (서열 번호:18) 및 (iii) 1 nmol의 "프레임시프트 주형" DNA (서열 번호:28), 1 nmol의 "점 돌연변이 주형" DNA (서열 번호:30) 또는 1 ㎍의 "큰 결실 주형" DNA (서열 번호:38)로 형질전환하였다. 형질전환된 세포를 CM-ura 플레이트 상에서 우라실에 대한 원영양성으로서 회수하였다. 원영양성 콜로니를 CM+can에 복제 평판으로 스크리닝하여 CAN1 돌연변이를 가진 카나바닌-내성 세포를 동정하였다. 각각의 형질전환으로부터의 Can^R 콜로니의 CAN1 유전자좌를 정방향 (서열 번호:22) 및 역방향 프라이머 (서열 번호:23)를 사용하는 PCR 증폭을 통해 스크리닝하였다. 각각의 PCR 생성물을 ExoSAP-IT® (캘리포니아주 산타 클라라 소재의 아피메트릭스(Affymetrix))를 사용하여 정제하고, 서열 분석 프

라이머 서열 번호:24를 사용하여 서열분석 (생어법)하였다. Can^R 콜로니의 총 개수에서 (특정 주형 DNA가 형질전환에 사용되었다는 관점에서) 예측되는 상동 재조합 이벤트를 나타내는 콜로니의 빈도는 약 15%이었다 (도 7).

[0302] 3개의 상이한 폴리뉴클레오티드 변형 주형 DNA 서열의 HR 수복의 효율성이 약간 상이하다 (도 8). 특히, 각각의 이러한 주형에 대한 HR 빈도는 약 11% (큰 결실 및 프레임시프트 공여자) 내지 22% (점 돌연변이 주형) (도 8)이었으며, 이는 Can1-1 표적 부위에서의 Cas9/sgRNA-생성된 절단 이벤트의 일부가 폴리뉴클레오티드 변형 주형 DNA가 제공되는 경우 충실도가 높은(high-fidelity) 방식으로 HR 경로를 사용하여 수복되었음을 나타낸다.

[0303] DNA 수복의 2가지 주요 경로인 NHEJ 또는 HR의 사용은 야로위아에서 NHEJ에 대한 분명한 편향을 나타내며 (도 7), 이는 통상적인 효모의 Cas9/sgRNA-매개 절단 이벤트에서의 수복의 연구에서 관찰된 것과는 상이하다. 예를 들어, 문헌[DiCarlo et aL. (*Nucleic Acids Res.* 41:4336-4343)]은 Cas9/sgRNA-매개 DNA 절단의 수복을 위해 공여 DNA가 제공되는 경우 얻어지는 거의 모든 S. 세레비시아 돌연변이체가 HR을 통해 생성되었지만, 공여 DNA가 제공되지 않는 경우 빈도는 4 내지 5배만큼(4 to 5 orders of magnitude) 떨어졌고, 이는 HR에 대한 분명한 편향을 나타내었다. 대조적으로, Cas9/sgRNA (RGR 카세트로부터 발현된 sgRNA) 절단 부위에서 야로위아의 총 돌연변이 빈도는 폴리뉴클레오티드 변형 주형 DNA를 수용하거나 수용하지 않는 형질전환체 (도 9, 두 유형의 형질전환체에 대한 ~15% 돌연변이율을 나타냄) 사이에서 다르지 않았고, HR은 공여 DNA가 제공되는 경우 생성되는 돌연변이 형질전환체의 약 15%만을 차지한다 (도 7). 따라서, 상기 관찰된 바와 같은 야로위아의 폴리뉴클레오티드 변형 주형 DNA 서열을 갖는 HR의 빈도는 단지 약 2.25%이었고, 이는 통상적인 효모에서 공여 DNA를 사용하여 관찰된 100%에 가까운 HR-매개 돌연변이율과 현저한 대조를 이룬다 (문헌[DiCarlo et al., *ibid*]).

[0304] 따라서, 이러한 예는 RNA 성분의 5' 캡이 리보자임의 자체 촉매 작용으로 제거되는, 5'-캡이 없는 RGEN (예를 들어, Cas9)의 RNA 성분 (예를 들어, sgRNA)을 발현하는 것은 통상적이지 않은 효모에서 DNA 서열의 RGEN-매개 표적화를 허용함을 추가로 입증한다. 이러한 예는 또한 통상적이지 않은 효모에서 RGEN-매개 절단이 적합한 공여 DNA (폴리뉴클레오티드 변형 주형)가 제공되는 경우 소정 비율로 HR에 의해 수복될 수 있음을 입증한다.

[0305] 실시예 6

[0306] 단일 안정한 벡터로부터의 Cas9 및 야로위아-최적화된 RGR 또는 RG 예비-sgRNA의 발현은 Cas9/sgRNA-매개 표적 DNA 절단을 제공함

[0307] 이러한 예에서, 야로위아-최적화된 RGR 또는 RG 예비-sgRNA 발현 카세트를 각각 야로위아-최적화된 Cas9 발현 카세트로서 동일한 안정한 발현 플라스미드로 개별적으로 옮겼다. 특히, 서열 번호:18 (RGR 발현에 대한) 또는 서열 번호:25 (RG 발현에 대한)를 각각 개별적으로 pZUFCas9 (도 3a, 서열 번호:14)로 클로닝하였다. 이는 단일-성분 형질전환이 세포에서 Cas9 엔도뉴클레아제, 및 RG 또는 RGR 예비-sgRNA를 발현하도록 허용함으로써, Cas9/sgRNA-매개 표적 부위 절단, 이어서 오류가 발생하기 쉬운 NHEJ 수복을 제공하였다.

[0308] 야로위아-최적화된 RGR (서열 번호:18) 또는 RG (서열 번호:25) sgRNA 발현 카세트를 정방향 (서열 번호:39) 및 역방향 (서열 번호:40) 프라이머를 사용하여 PCR에 의해 증폭시켰다. 각각의 산물을 PacI/ClaI 제한 부위에서 플라스미드 pZUFCas9 (서열 번호:14)로 개별적으로 클로닝하여, 각각 Cas9 발현 및 최적화된 RGR 예비-sgRNA (pRF84, 서열 번호:41, 도 10a) 또는 최적화된 RG 예비-sgRNA (pRF85, 서열 번호:42, 도 10b)의 발현을 위한 각각의 카세트를 갖는 2개의 새로운 플라스미드를 생성하였다.

[0309] Cas9/sgRNA-매개 표적 부위 (Can1-1) 절단을 제공하기 위해 각각 Cas9 및 sgRNA를 효율적으로 발현하는 pRF84 (서열 번호:41) 및 pRF85 (서열 번호:42) 플라스미드 구조물의 능력을 시험하기 위해, Ura^- 야로위아 리폴리티카 세포 (Y2224)를 상기 리튬 이온 형질전환 방법을 사용하여 200 ng의 pRF84 (서열 번호:41), pRF85 (서열 번호:42) 또는 pZUFCas9 (서열 번호:14)로 형질전환하였다. 각각의 플라스미드로 형질전환된 세포를 CM-ura 배지 상에서 우라실 원영양성으로 선별하였다. 각각의 형질전환으로부터의 우라실 원영양성을 CM+can 상의 복제 평판에 의해 CAN1 돌연변이체에 대해 스크리닝하였다. CM+can 플레이트 상에서 성장한 콜로니의 수를 사용하여 pZUFCas9 (Cas9 단독 발현), pRF84 (Cas9 및 RGR 예비-sgRNA 발현) 또는 pRF85 (Cas9 및 RG 예비-sgRNA 발현)로 형질전환된 세포의 CAN1 돌연변이 빈도 (도 11)를 생성하였다. CAN1 유전자좌에서 pZUFCas9 (서열 번호:14)로 형질전환된 야로위아 세포의 Cas9/sgRNA-매개 돌연변이의 빈도가 0인 반면, (i) Cas9 및 (ii) RGR 예비-sgRNA (pRF84) 또는 RG sgRNA (pRF85)를 발현하는 세포의 CAN1 돌연변이 빈도 (~69%)는 카나바닌-내성 (도 11)에 의해 나타낸 바와 유사하다.

[0310] 이러한 결과는 동일한 벡터로부터 Cas9 및 예비-sgRNA를 발현하는 것이 상당히 더 높은 비율의 Cas9/sgRNA-매개 절단을 유도하고 결과적으로 예측되는 절단 부위에서 NHEJ-매개 돌연변이를 유도함을 나타낸다. Cas9 및 예비-sgRNA (RGR 또는 RG 예비-sgRNA)를 암호화하는 별개의 서열로 형질전환된 야로위아 세포는 약 5%의 표적화된 돌연변이 빈도 (실시예 4, 도 6)를 나타내는 반면, Cas9 및 sgRNA 암호화 서열을 형질전환에 사용되는 동일한 벡터 상에 위치시키면 표적화된 돌연변이 빈도는 약 69%이었다 (도 11).

[0311] 따라서, 통상적이지 않은 효모를 형질전환하는데 사용되는 동일한 구조물로부터 Cas 단백질 및 이의 상응하는 RNA 성분을 발현하면 RGEN 단백질 및 RNA 성분을 발현하기 위해 별개의 구조물을 사용하는 것과 비교하여 효모에서 Cas-매개 DNA 표적화의 비율이 더 높아진다.

[0312] 실시예 7

[0313] 야로위아 리폴리티카에서 HDV 리보자임-sgRNA 융합을 사용하는 고-효율 유전자 표적화

[0314] 이러한 실시예는 단일 가이드 RNA (HDV 리보자임에 의해 5' 말단 상에 플랭킹된 sgRNA) (리보자임-단일 가이드 RNA 융합)의 사용을 논의한다. 발현되는 경우, HDV 리보자임은 그 자체의 서열의 5'을 절단하여, 임의의 선행되는 전사물을 제거하지만, sgRNA의 5' 말단에 융합된 HDV 서열을 남겨둔다.

[0315] 플라스미드 pZuf-Cas9 (서열 번호: 14)를 애질런트(Agilent) 퀵체인지(QuickChange) 및 다음과 같은 프라이머 AarI-제거-1 (AGAAGTATCCTACCATCTACcatctccGAAAGAAACTCGTCGATTCC, 서열 번호: 90) 및 AarI-제거-2 (GGAATCGACGAGTTTCTTTCggagatgGTAGATGGTAGGATACTTCT, 서열 번호:91)를 사용하여 돌연변이화하여, pZuf-Cas9 (서열 번호: 14) 상의 Cas9 유전자 (서열 번호: 10)에 존재하는 내인성 AarI 부위를 제거하고, pRF109 (서열 번호: 92)를 생성하였다. 변형된 Aar1- Cas9 유전자 (서열 번호: 93)를 pRF109로부터의 NcoI/NotI 단편으로서 pZufCas9의 NcoI/NotI 부위에 클로닝하여, 기존 Cas9 유전자 (서열 번호: 10)를 Aar1- Cas9 유전자로 대체하여 pRF141 (서열 번호: 94)을 생성하였다.

[0316] 고효율 클로닝 카세트 (도 12a, 서열 번호: 95)는 y152 프로모터 (서열 번호: 96), HDV 리보자임 (서열 번호: 16), 대장균 역선별 카세트 *rpsL* (서열 번호: 97), 가이드 RNA CER 도메인을 암호화하는 DNA (서열 번호: 1) 및 S. 세레비시아 Sup4 종결자 (서열 번호: 8)로 이루어진다. 고효율 클로닝 카세트 (서열 번호: 95)의 말단을 플랭킹하는 것은 PacI 및 ClaI 제한 효소 인식 부위이다. 고효율 클로닝 카세트를 pRF141 (서열 번호: 94)의 PacI/ClaI 부위로 클로닝하여 pRF291 (서열 번호: 98)을 생성하였다. *rpsL* 역선별 카세트 (서열 번호: 97)는 S12 리보솜 단백질 서브유닛(subunit)을 암호화하는 대장균 유전자 *rpsL*의 WT 카피를 함유한다 (문헌[*Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, 1987, First eD. American Society of Microbiology, Washington, DC]). S12 서브유닛의 일부 돌연변이는 열성 방식으로(recessive manner) (문헌[Lederberg J. 1951. Streptomycin Resistance; a genetically recessive mutatioN. Journal of bacteriology 61:549-550]) 항생제 스트렙토마이신에 대한 내성을 일으키므로 (문헌[Ozaki M, Mizushima S, Nomura M. 1969. Identification and functional characterization of the protein controlled by the streptomycin-Resistant locus in E. *coli*. Nature 222:333-339]), *rpsL* 유전자의 야생형 카피가 존재하는 경우 균주는 표현형적으로 스트렙토마이신에 민감하다. Top10 (라이프 테크놀로지스(Life technologies))과 같은 일반적인 클로닝 균주는 세포가 스트렙토마이신에 내성을 갖도록 그들의 염색체 상에 rpsL 카피를 가지고 있다.

[0317] 가이드 RNA의 가변 표적화 도메인을 암호화하는 DNA 단편을 플라스미드 (예컨대 pRF291)로 클로닝하는 것은 어닐링되는 경우 가변 표적화 도메인을 암호화하는 DNA 단편 및 고효율 클로닝 카세트에 존재하는 2개의 AarI 부위로의 클로닝을 위한 정확한 돌출부를 함유하는 2개의 부분적으로 상보적인 올리고뉴클레오티드를 필요로 한다. 2개의 올리고뉴클레오티드 Can1-1F (AATGGGACtcaaacgattacccaccctcGTTT, 서열 번호: 99) 및 Can1-1R (TCTAAAACgagggtgggtaatcgtttgaGTCC, 서열 번호: 100)를 100μM에서 듀플렉스 완충액 (30mM HEPES pH 7.5, 100mM 아세트산나트륨)에서 재현탁시켰다. Can1-1F (서열 번호 99) 및 Can1-1R (서열 번호 100)을 단일 튜브 내에서 각각 50μM의 최종 농도로 혼합하고, 5분 동안 95℃로 가열하여, 0.1℃/min으로 25℃로 냉각시켜, 2개의 올리고뉴클레오티드를 어닐링하여, Can1-1 표적 부위를 표적화할 수 있는 가이드 RNA의 가변 표적 도메인을 암호화하는 DNA 단편을 함유하는 작은 듀플렉스 DNA 분자 (도 12b)를 형성하였다 (PAM 서열 CGG를 포함하는 서열 번호: 101로 나타냄). 최종 부피 20μl 중에 50ng의 pRF291, Can1-1F 및 Can1-1R로 구성된 2.5μM의 소형 듀플렉스 DNA, 1x T4 리가제 완충액 (50mM Tris-HCl, 10mM $MgCl_2$, 1mM

ATP, 10mM DTT pH 7.5), 0.5 μM의 AarI 올리고뉴클레오티드, 2 단위 AarI, 40 단위 T4 DNA 리가제를 함유하는 단일 튜브 분해(digestion)/라이게이션 반응을 생성하였다. 이중가닥 Can1-1F 및 Can1-1R 듀플렉스가 없는 제2 대조 반응물을 또한 조립하였다. 반응물을 30분 동안 37℃에서 인큐베이션하였다. 10μl의 각각의 반응물을 이전에 기재된 바와 같은 Top10 대장균 세포로 형질전환하였다 (문헌[Green MR, Sambrook J. 2012. Molecular Cloning: A Laboratory Manual, Fourth Edition eD. Cold Spring Harbor Laboratory PRess, Cold Spring Harbor, NY]). Can1-1F 및 Can1-1R의 듀플렉스가 AarI 제한 부위 (도 12a)에 의해 플랭킹된 *rpsL* 역선별 마커를 대체한 pRF291의 존재를 선별하기 위해, 세포를 100 μg/ml의 앰피실린 및 50 μg/ml의 스트렙토마이신을 함유하는 1.5% (w/v) 박토(Bacto) 한천으로 고형화된 용원성 브로쓰 상에 플레이팅하였다. 고효율 클로닝 카세트를 함유하는 pRF291의 존재는 플라스미드 상의 역선별 카세트의 존재로 인해 항생제 암피실린에 대해 표현형적으로 내성이지만 항생제 스트렙토마이신에 민감성인 콜로니를 생성하였다. 그러나, 역선별 카세트가 AarI 효소를 통해 제거되고 Can1-1 이중 DNA가 부위로 라이게이션되는 경우 (AarI에 대한 인식 서열을 제거함), 플라스미드로 형질전환된 세포는 암피실린 내성, 스트렙토마이신 내성 표현형을 가졌다 (도 12a). Can1-1 가변 표적화 도메인 표적화 (역선별 카세트를 대체함)를 암호화하는 DNA 단편을 함유하는 pRF291은 sup4 종결자 (서열 번호: 8)에 융합된 가이드 CER 도메인 (서열 번호: 1)을 암호화하는 DNA에 융합된 Can1-1 가변 표적화 도메인 (서열 번호: 17)을 암호화하는 DNA에 융합된 HDV 리보자임 (서열 번호: 16)을 암호화하는 DNA에 융합된 y152 프로모터를 함유하는 재조합 HDV- sgRNA 발현 카세트 (서열 번호: 102)를 생성하였다. 이러한 구조물 pRF303 (서열 번호: 103)을 함유하는 플라스미드를 사용하여 (Cas9 엔도뉴클레아제와 복합체를 형성한 경우) HDV 리보자임-가이드 RNA (서열 번호: 104)를 암호화하여, 돌연변이 유발을 위해 야로위아 리폴리티카의 Can1 유전자 (서열 번호: 21)를 표적화하였다.

[0318] (문헌[Richard M, Quijano RR, Bezzate S, Bordon-Pallier F, Gaillardin C. 2001. Journal of bacteriology 183:3098-3107]에 기재된 바와 같이) 플라스미드가 없거나 sgRNA 발현 카세트 (pRF291, 서열 번호: 98)가 없는 100 ng의 플라스미드, RGR 발현 카세트 (서열 번호: 41)를 갖는 pRF84 플라스미드, 5' 리보자임이 sgRNA (서열 번호: 42)로부터 그 자체를 제거하는 RG 카세트를 갖는 pRF85 플라스미드 또는 야로위아의 Can1-1 표적 부위를 표적화하는 HDV-sgRNA 융합 발현 카세트 (서열 번호: 102)를 갖는 pRF303(서열 번호 103)로 야로위아 리폴리티카를 형질전환하였다. 형질전환체를 우라실 원영양성에 대해 선별하고, Can1 유전자에서 아르기닌 유사체 카나바닌에 대한 표현형 내성에 의해 돌연변이에 대해 기록하였다. HDV-sgRNA 융합을 발현하는 플라스미드는 리보자임으로부터 유리된 sgRNA 중 하나를 발현하는 플라스미드와 동일한 빈도로 Can1 유전자에서 기능 상실 돌연변이를 유발하였으며, 이는 HDV 리보자임과 sgRNA 표적화 Can1-1의 5' 융합이 sgRNA 기능에 영향을 주지 않음을 시사하였다 (표 4).

[0319] [표 4]

상이한 sgRNA 변이체를 통한 Can1-1 표적 서열의 돌연변이 빈도.

플라스미드	sgRNA 변이체	Can^R 빈도 ± SD
pRF291	sgRNA 없음	0±0
pRF84	sgRNA를 생성하는 RGR	0.70±0.04
pRF85	sgRNA를 생성하는 RG	0.73±0.11
pRF303	HDV-sgRNA 융합체	0.81±0.15

[0320]

[0321] 다수의 추가의 표적 부위 (표 5)를 표적화하는 가변 표적 도메인을 암호화하는 다수의 추가의 DNA 단편을 상기 기재한 바와 동일한 전략을 사용하여 도 12a에 예시된 바와 같이 pRF291 (서열 번호: 98) 플라스미드로 클로닝하였다. Can1 유전자 (서열 번호: 105) 내에서, can1-2 표적 부위 (서열 번호: 106) 및 다른 표적 부위, 예컨대 sou2-1 (서열 번호: 107), Sou2-2 (서열 번호: 108), Tg11-1 (서열 번호: 112), Acos10-1 (서열 번호: 113), Fat1-1 (서열 번호: 114) 및 Ura3-1 (서열 번호: 116)을 표적화하는 제2 표적 부위를 표적화하는 가변 표적화 도메인을 암호화하는 DNA 단편을 포함함.

[0322] [표 5]

Cas9 로 야로위아에서 상이한 유전자좌를 표적화하기 위해 sgRNA VT 도메인을 암호화하는 DNA 서열

	sgRNA 의 가변 표적화 도메인을 암호화하는 DNA	야로위아 표적 부위 + PAM 서열 (볼드)
Can1-2	서열 번호: 106 의 염기 1-20	GGCCCACTCGGATGACTCAG**AGG** (서열 번호: 106)
Sou2-1	서열 번호: 107 의 염기 1-24	GTCTGGACCTTCCACCCTCGCCAC**GGG** (서열 번호: 107)
Sou2-2	서열 번호: 108 의 염기 1-22	GCAGTCCCGTGGCGAGGGTGGA**AGG** (서열 번호: 108)
TGL1-1	서열 번호: 112 의 염기 1-20	CAGCTCGAGACGTCCTAGAA**CGG** (서열 번호: 112)
Acos10-1	서열 번호: 113 의 염기 1-20	TTCCTCTGTCACAGACGTTT**CGG** (서열 번호: 113)
Fat1-1	서열 번호: 114 의 염기 1-20	GAAAAGTGCGTTTTGATTCT**CGG** (서열 번호: 114)
ura3-1	서열 번호: 116 의 염기 1-20	GCCGCTCGAGTGCTCAAGC**TCG** (서열 번호: 116)

[0323]

[0324] 표적 부위의 돌연변이 빈도는 모든 HDV-sgRNA 융합체가 Cas9 엔도뉴클레아제와 복합체를 형성할 수 있음을 나타내었으며, 이는 NHEJ를 통한 돌연변이를 유도하는 각각의 표적 부위에서 절단을 일으켰다 (표 6).

[0325] [표 6]

HDV-sgRNA 융합을 사용한 야로위아 리폴리티카의 다양한 표적 부위에서의 돌연변이 빈도.

표적 부위	돌연변이 빈도 ± SD
Can1-2	0.76±0.15
Sou2-1	0.19
Sou2-2	0.30
TGL1-1	0.88
Acos10-1	0.36
Fat1-1	0.50
ura3-1	0.92

[0326]

[0327] 실시예 8

[0328] 불활성화된-Cas9 및 HDV-sgRNA 융합을 사용한 유전자 침묵.

[0329] HNH 및 RuvC 뉴클레아제 도메인 (서열 번호: 117)에 돌연변이를 함유하는 촉매적으로 불활성화된 Cas9 변이체는 sgRNA와 상호작용하고 생체 내 표적 부위에 결합할 수 있지만, 표적 DNA의 어느 한 가닥을 절단할 수는 없다. DNA에 결합하지만 절단하지 않는 이러한 작용 방식은 영구적인 유전적 변화를 일으키지 않으면서, 염색체의 특정 유전자좌의 발현을 일시적으로 감소시키는 데 사용될 수 있다.

[0330] 야로위아 리폴리티카를 위한 촉매적으로 불활성화된 Cas9 발현 카세트를 생성하기 위해, 프라이머 D10AF (GAAATACTCCATCGGCCTGGCCATTGGAACCAACTCTGTCG, 서열 번호: 118) 및 D10AR (CGACAGAGTTGGTTCCAATGGCCAGGCCGATGGAGTATTTC, 서열 번호: 119)로 기재된 바와 같은 퀵체인지 부위-지정 돌연변이 유발 (스트라타진(Stratagene))을 사용하여 D10A 돌연변이를 플라스미드 pZufCas9 (서열 번호: 14)에 도입하였다. 이는 RuvC 뉴클레아제 (서열 번호 120) 및 구조물 pRF111 (서열 번호 121)을 함유하는 상응하는 플라스미드를 불활성화시키는 D10A 돌연변이를 갖는 야로위아 코돈 최적화된 Cas9 유전자를 생성시켰다. 제2 뉴클레아제 도메인 (HNH)을 불활성화시키기 위해, H840A 돌연변이를 야로위아 코돈 최적화된 D10A 유전자로 도입하여 야로위아 코돈 최적화된 Cas9 불활성화된 유전자 (서열 번호: 124) 및 야로위아에서의 발현을 위한 유전자를 포함하는 플라스미드인 pRF143 (서열 번호: 125)를 생성하는 프라이머 H840A1 (TCAGCGACTACGATGTGGACGCCATTGTCCCTCAATCCTTTCT, 서열 번호: 122) 및 H840A2 (AGAAAGGATTGAGGGACAATGGCGTCCACATCGTAGTCGCTGA, 서열

번호: 123)를 사용하여 추가의 1회의 퀵체인지 돌연변이 유발 (스트라타진)을 수행하였다.

[0331] 야로위아 리폴리티카에서 유전자 침묵을 평가하기 위해, 야로위아 코돈 최적화된 dsRED익스프레스 오픈 리딩 프레임 (서열 번호: 126)을 5' NcoI 제한 부위 및 3' NotI 제한 부위 (서열 번호: 127)를 갖는 클로닝 단편으로서 제조하였다. 클로닝 단편 (서열 번호: 127)을 pZufCas9의 NcoI/NotI 부위로 클로닝하여 플라스미드 pRF165 (서열 번호: 129) 상에 함유된 FAB1-dsRED 융합 카세트 (서열 번호: 128)를 생성하는 야로위아 최적화된 dsRED익스프레스 클로닝 단편 (서열 번호: 127)에 융합된 FBA1 프로모터 (서열 번호: 12)를 생성하였다. FBA1-dsRED익스프레스 카세트 (서열 번호: 128)를 염색체로 통합하기 위해, 카세트 (서열 번호: 130)를 함유하는 PmeI-NotI 단편을 통합 플라스미드 p2P069 (서열 번호: 131)의 PmeI/NotI 부위로 라이게이션하여 FBA1-dsRED익스프레스 발현 카세트인 pRF201 (서열 번호: 132)을 포함하는 통합 벡터를 생성하였다. FBA1-dsRED익스프레스 융합체 및 Leu2 유전자 (서열 번호: 133)의 카피를 포함하는 pRF201의 SphI/AscI 단편을 표준 기술을 사용하여 류신 원영양성에 대해 선별함으로써 야로위아의 염색체로 통합하였다 (문헌[Richard M, Quijano RR, Bezzate S, Bordon-Pallier F, Gaillardin C. 2001. Tagging morphogenetic genes by insertional mutagenesis in the yeast Yarrowia lipolyticA. Journal of bacteriology 183:3098-3107]). FBA1-dsRED익스프레스 발현 카세트의 존재를 표준 PCR 기술 및 프라이머 HY026 (GCGCGTTTAAACCATCATCTAAGGGCCTCAAAACTACC, 서열 번호: 134) 및 HY027 (GAGAGCGGCCGCTTAAAGAAACAGATGGTGTCTTCCCT, 서열 번호: 135)을 사용하여 야로위아 게놈에서 확인하였다. FBA1-dsRED익스프레스 카세트 (서열 번호: 128)를 함유하는 2개의 독립 균주 YRF41 및 YRF42를 추가의 사용을 위해 선택하였다.

[0332] 야로위아 최적화된 dsRED익스프레스 발현 카세트 (서열 번호: 128)를 표적화하기 위한 sgRNA를 생성하기 위해 실시예 12와 유사한 전략을 사용하였다. 도 13a에 도시된 바와 같이, 플라스미드 구조물인 pRF169 (서열 번호: 136)는 야로위아로부터의 GPD 프로모터 (서열 번호: 137), 역선별가능한 마커, 가이드 RNA CER 도메인 (서열 번호: 1)을 암호화하는 DNA 및 Sup4 종결자 (서열 번호: 8) 카세트 (서열 번호: 138)를 함유하였다. HH 리보자임을 암호화하는 DNA 단편에 연결된 야로위아의 표적 부위를 표적화하는 sgRNA의 가변 표적화 도메인을 암호화하는 DNA를, 도 13b에 도시된 바와 같이 망치머리 리보자임의 처음 6개의 뉴클레오티드가 가변 표적 도메인의 처음 6개의 뉴클레오티드의 역 상보서열이되도록 하는 HH 리보자임을 암호화하는 DNA 단편을 제외하고는 실시예 12에 기재된 바와 같이 pRF169 (서열 번호 136)로 클로닝하였다. 정확한 돌출부를 갖는 이중 가닥 올리고뉴클레오티드가 AarI 부위 사이의 역선별 카세트를 대체하는 경우, 리보자임-가이드 RNA (RG) 발현 카세트가 생성되었다 (도 13-a). 전사되는 경우, HH 리보자임은 리보자임-가이드 RNA 분자로부터 5' 전사물 및 그 자체를 제거하여, 세포에 온전한 sgRNA를 남긴다. dsRED익스프레스 오픈 리딩 프레임 (서열 번호: 126)을 표적으로 하는 3개의 가이드 RNA를 생성하였는데; 2개는 주형 가닥, ds-주형-1 (서열 번호: 139), ds-주형-2 (서열 번호: 140)를 표적화하고 하나는 비주형 가닥 ds-비주형-1 (서열 번호: 141)을 표적화하였다.

[0333] 각각의 표적 부위에 대해, 표적 특이적인 망치머리 리보자임, 가변 표적 도메인 (VTD) 및 pRF169의 AarI 부위로 클로닝하기 위한 정확한 중첩 말단을 암호화하는 DNA 서열을 함유하는 2개의 올리고뉴클레오티드를 설계하였다. 각 부위의 올리고뉴클레오티드; ds-주형-1F (서열 번호 144) ds-주형-1R (서열 번호 145), ds-주형-2F (서열 번호 146), ds-주형-2R (서열 번호 147) ds-비주형-1F (서열 번호 148) 및 ds-비주형-1R (서열 번호 149)을 이중 가닥화하여(duplexed) pRF169의 고효율 카세트 (도 13a 및 도 13b)에 남아있는 AarI 돌출부로 클로닝하기 위한 정확한 돌출부를 갖는 이중 가닥 DNA 분자를 형성하였고, pRF291로 클로닝하기 위해 실시예 12에 기재된 바와 같이 수행하였다. 역선별 카세트를 대체하는 sgRNA의 가변 표적 도메인을 암호화하는 DNA 단편의 삽입은 Sup4 종결자에 융합된 가이드 RNA CER 도메인을 암호화하는 DNA에 융합된 망치머리 리보자임-표적 부위 듀플렉스 DNA에 융합된 GPD 프로모터를 갖는 각각의 표적 부위에 대한 새로운 플라스미드를 생성하였다 (도 13a). 이러한 듀플렉스를 함유하는 플라스미드는 pRF296 (ds-주형-1, 서열 번호: 150), pRF298 (ds-주형-2, 서열 번호: 151), pRF300 (ds-비주형-1, 서열 번호: 152)이다.

[0334] 유전자 침묵을 위한 구조물을 생성하기 위해, pRF143으로부터의 불활성화된 Cas9 (서열 번호: 125)를 표준 기술을 사용하여 NcoI/NotI 단편으로서 pRF293, pRF298 및 pRF300에 클로닝하고, 이들 플라스미드의 NcoI/NotI 부위에 존재하는 기능적 Cas9 (서열 번호: 93)를 대체하여 각각 플라스미드 pRF339 (서열 번호: 153), pRF341 (서열 번호: 154) 및 pRF342 (서열 번호: 155)를 생성하였다.

[0335] 균주 YRF41 및 YRF42는 표준 기술을 사용하여 pRF339, pRF341 및 pRF343으로 우라실 원영양성으로 형질전환되었

다 (문헌[Richard M, Quijano RR, Bezzate S, Bordon-Pallier F, Gaillardin C. 2001. Tagging morphogenetic genes by insertional mutagenesis in the yeast Yarrowia lipolyticA. Journal of bacteriology 183:3098-3107]). 각각의 형질전환에 있어서, 12개의 형질전환체를 우라실이 없는 플레이트 상에 도말 정제하여, 플라스미드를 유지하였다. 각각의 분리주를 사용하여 2ml의 CM-ura 브로쓰 (테크노바(Teknova))를 접종하고, 30℃, 250 RPM에서 하룻밤 동안 성장시켰다. 각각 2 내지 5㎕를 하룻밤 동안 200㎕의 ddH_2O로 희석하고, 아큐리(Accuri) 유세포 분석기의 dsRED익스프레스 채널에서 분석하여, 각 세포 내의 dsRED익스프레스 단백질의 양을 평가하였다. 7,151 내지 10,000개의 세포를 각 배양물로부터 분석하였다. dsRED익스프레스 발현 카세트가 없는 야로위아 세포의 평균 형광도를 분석한 각 배양물의 평균 형광도에서 제하여, 이들을 평균낸 각 균주/플라스미드 조합 내에서 보정된 평균 형광도를 얻었고, 표준 편차를 측정하였다 (표 7). 목적 유전자를 표적화하는 발현 벡터를 통해 발현된 리보자임-sgRBA (RG)와 조합된 불활성화된 Cas9는 유전자의 발현을 2 내지 10배로 침묵시켰다. 배수 침묵화(fold silencing)는 표적 부위의 위치 및 쇄의 수(strandedness) 및/또는 야로위아 세포에서 기능적 형태의 DNA 폴리머라제 프로모터로부터 발현되는 리보자임 플랭킹된 sgRNA의 능력에 따라 달라졌다 (표 7).

[0336] [표 7]

2 개의 FBA-dsRED 익스프레스 통합된 균주에서의 3 개의 표적 부위에 의한 유전자 침묵.

균주	플라스미드	표적 부위	평균 형광도 ± SD	표적수의 배수(Fold of No Target)
YRF41	없음	없음	540.6±2.9	1
YRF41	pRF339 (서열 번호: 69)	ds-주형-1	299.2±138.7	0.55±0.26
YRF41	pRF341 (서열 번호: 70)	ds-주형-2	257.9±139.3	0.48±0.26
YRF41	pRF343 (서열 번호: 71)	ds-비주형-1	169.4±45.3	0.31±0.08
YRF42	없음	없음	871.2±36.9	1
YRF42	pRF339 (서열 번호: 69)	ds-주형-1	194.3±121.1	0.22±0.14
YRF42	pRF341 (서열 번호: 70)	ds-주형-2	168.7±191.6	0.19±0.22
YRF42	pRF343 (서열 번호: 71)	ds-비주형-1	94.9±109.6	0.11±0.13

[0337]

[0338] 실시예 9

[0339] 단일 플라스미드로부터 발현된 Cas9 및 HDV 리보자임-sgRNA 융합체 (RG)를 사용한 정확한 유전자 편집.

[0340] 이러한 예에서 우리는 동일한 안정한 벡터로부터 발현된 Cas9와 HDV-sgRNA 융합체의 안정한 발현이 상동성 재조합을 통한 정확한 유전자 편집을 위한 기질이될 수 있는 야로위아의 표적 부위에 DNA 이중 가닥 절단을 생성할 수 있음을 보여준다.

[0341] 실시예 4에 기재된 Can1 결실 폴리뉴클레오티드 변형 주형 DNA (서열 번호: 38)를 HinDIII로 분해하고 표준 기법을 사용하여 pUC18의 HinDIII 부위로 클로닝하여 pRF80 (서열 번호: 156)을 생성하였다. 더 짧은 Can1 결실 편집 주형 (서열 번호: 157)을 표준 PCR 기술 및 프라이머 80F (AGCTTGCTACGTTAGGAGAA, 서열 번호: 158) 및 80R (TATGAGCTTATCCTGTATCG, 서열 번호: 159)을 사용하여 pRF80으로부터 증폭시켜, 더 많은 양의 편집 주형을 생성하였다.

[0342] Ura 영양요구성 야로위아 세포를 표준 기술을 사용하여 (문헌[Richard M, Quijano RR, Bezzate S, Bordon-Pallier F, Gaillardin C. 2001. Tagging morphogenetic genes by insertional mutagenesis in the yeast Yarrowia lipolyticA. Journal of bacteriology 183:3098-3107]) Cas9 유전자의 카피를 포함하지만 sgRNA가 없는 100ng의 플라스미드 pRF291, 및 Cas9 유전자의 카피 및 편집 주형 DNA가 없거나 1000 ng의 짧은 Can1 결실

편집 주형 (서열 번호: 157)을 갖는 Can1-1 표적 부위 HDV-sgRNA 발현 카세트를 함께 포함하는 pRF303으로 형질전환하였다. 형질전환체를 CM-ura 배지 (테크노바) 상에서 선별하였다. 각각의 형질전환에 대해, 20개의 개별 콜로니를 CM-ura 배지 (테크노바) 상에서 도말 정제하였다. 도말 정제된 각각의 콜로니로부터 4개의 개별 콜로니 (형질전환 당 총 80개)를 60 ㎍/ml의 L- 카나바닌을 함유하는 CM-arg 플레이트 상으로 패치하여, Can-1 유전자에서 기능적 대립유전자의 결실을 포함하는 콜로니에 대해 스크리닝하였다. 카나바닌에 대한 내성을 보이는 패치를 기록하고 유전자 불활성화의 빈도를 기록하였다 (표 8). 상동성 재조합으로 인해 Can1 기능을 상실한 콜로니 및 NHEJ로 인해 Can1 기능을 상실한 콜로니를 결정하기 위해, Can1 유전자좌 (서열 번호: 160)를 Can1-PCRF (GGAAGGCACATATGGCAAGG, 서열 번호: 22) 및 Can1-PCRR (GTAAGAGTGGTTTGCTCCAGG, 서열 번호: 23)을 사용하여 증폭시켰다. 이전 실시예에 기재된 바와 같이 작은 삽입-결실을 갖는 세포에서, PCR 생성물은 Can1 결실 편집 주형과 상동성 재조합에 의한 결실을 함유하는 균주에서 WT Can1 좌위 (서열 번호: 160)와 크기 (2125bp)가 매우 유사해야 하며, Can1-PCRF (서열 번호: 22) 및 Can1-PCRR (서열 번호: 23)을 갖는 PCR 단편 (서열 번호: 161)은 더 작을 것이다 (392bp). 2 μl의 PCR 생성물을 전기영동으로 분해하고, 표준 기술을 사용하여 영상을 얻었다 (그림 14). 편집 주형 (서열 번호: 161)과의 재조합에 해당하는 짧은 밴드를 갖는 도말 정제 시에, 1개 이상의 콜로니를 생성하는 원래의 20개의 도말된 콜로니의 백분율을 사용하여 HR의 빈도를 측정하였다 (표 8). pRF303 (서열 번호: 103)을 수용한 세포에서, 카나바닌 내성 콜로니의 빈도는 세포가 편집 주형을 받았는지 여부와 유사하였다 (표 8). 형질전환된 세포의 총 개체군 중 pRF303 (서열 번호: 103) 및 Can1 짧은 편집 형질 (서열 번호: 157) 모두를 수용한 세포에서, 약 1/10이 편집 주형(서열 번호: 157)으로부터의 Can1 유전자좌의 정확한 편집을 포함하였다 (표 8).

[0343] [표 8]

카나바닌 저항 빈도 및 정확한 편집의 빈도.

플라스미드	sgRNA	편집 주형	Can^R 빈도 ± SD	HR 빈도 ± SD
pRF291 (서열 번호: 98)	없음	없음	0 ± 0	측정되지 않음
pRF291 (서열 번호: 98)	없음	Can1 짧은 서열 (서열 번호: 157)	0 ± 0	측정되지 않음
pRF303 (서열 번호: 103)	HDV-Can1-1 sgRNA	없음	0.80 ± 0.10	측정되지 않음
pRF303 (서열 번호: 103)	HDV-Can1-1 sgRNA	Can1 짧은 서열 (서열 번호: 157)	0.72 ± 0.12	0.09±0.05

[0344]

[0345] 실시예 10

[0346] 야로위아에서 URA3 유전자 불활성화

[0347] 본 실시예는 야로위아에서 URA3 유전자 불활성화를 위해 단일 가이드 RNA (sgRNA)와 Cas9 엔도뉴클레아제를 개별적으로 또는 함께 발현하는 플라스미드의 구성 및 사용을 기술한다.

[0348] pYRH235 및 pYRH236은 각각 URA3.1 표적 서열 (5'-ctgttcagagacagtttcct-3; 서열 번호:165)을 표적화하는 리보자임 플랭킹된 예비-sgRNA (RGR-URA3.1; 서열 번호: 164) 및 URA3.2 표적 서열 (5'-taacatccagagaagcacac-3'; 서열 번호:167)을 표적화하는 리보자임 플랭킹된 예비-sgRNA (RGR-URA3.2; 서열 번호: 166)를 발현하였다. RGR-URA3.1의 DNA 단편을 암호화하는 NcoI-NotI 제한 분해 단편 및 RGR-URA3.2를 암호화하는 BspHI-NotI 제한 분해 단편을 FBA1L 프로모터 (서열 번호: 168)에 융합시켜 각각 pYRH235 및 pYRH236을 생성하였다. pYRH235 및 pYRH236 플라스미드는 설포닐 우레아 내성을 부여하는 단일 아미노산 변화 (W497L)를 갖는 천연 아세토하이드록시산 합성효소 (AHAS 또는 아세토락테이트 합성효소; E.C.4.1.3.18; 서열 번호: 169)의 마커 유전자를 함유하였다.

[0349] 야로위아 균주 ATCC20362의 Ura^- 유도체 (Y2224)를 먼저 SphI-BsiWI 제한 효소 분해에 의해 선형화된 pZufCas9 (서열 번호: 14)로 형질전환시키고, 형질전환체를 우라실이 없는 완전 최소 (CM) 플레이트에 상에서 선별하였다. 선형화된 Cas9 발현 카세트는 야로위아 게놈에 무작위로 통합되었으므로, 형질전환체는 URA3 유전자의 카피를 적어도 2개를 함유하였다. 이어서, sgRNA를 발현하는 pYRH235 또는 pYRH236을 Cas9를 발현하는 야로위아 균주로 형질전환시키고, 형질전환체를 600 mg/L의 설포닐 우레아를 함유하는 CM 플레이트 상에서 선별하였다. 50개의 형질전환체를 CM-ura 플레이트 및 5-FOA를 갖는 SC 플레이트에 패치하여, URA3에 대한 Cas9 및 sgRNA에 의한 URA3 유전자 불활성화의 빈도를 찾았다. 각각 94% 및 100%의 pYRH235 및 pYRH236 형질전환체가

우라실 영양요구성이 되었다.

[0350] 표적 부위 URA3.1 또는 URA3.2에서의 돌연변이의 서열 분석 확인을 수행하였다. ZufCas9 및 pYRH235의 20개의 형질전환체를 서열 분석을 위해 무작위로 선택하고, 각각의 콜로니를 플라스미드 pZufCas9의 URA3 유전자의 돌연변이 및 천연 게놈 URA3에 대해 분석하였다. 플라스미드 pZufCas9로부터의 URA3 유전자를 서열 분석하기 위해, URA3에 대한 RHO705 (서열 번호: 170) 및 FBA1 프로모터 서열에 대한 RHO719 (서열 번호: 171)를 사용하여 영역을 PCR 증폭하였고, 프라이머 RHO733 (서열 번호: 172) 또는 RHO734 (서열 번호: 173)를 사용하여 주형으로서 PCR 증폭 산물을 서열 분석하였다. 천연 게놈 기원의 URA3 유전자를 서열 분석하기 위해, 프라이머 RHO705 (서열 번호: 170) 및 RHO707 (서열 번호: 174)을 사용하여 PCR 증폭을 수행하였고, 프라이머 RHO733 (서열 번호: 172) 및 RHO734 (서열 번호: 173)를 사용하여 주형으로서 PCR 증폭 산물을 서열 분석하였다. 20개의 콜로니 모두는 플라스미드 및 게놈 유래 URA3 유전자 모두에서 돌연변이를 함유하였다 (도 15). 5개의 대표적인 콜로니 (콜로니 1, 2, 3, 5 및 6, 각각 서열 번호: 176, 177, 178, 179 및 180, 및 서열 번호: 181, 182, 183, 184 및 185)의 플라스미드 및 게놈 유래 URA3 유전자, 및 야생형 URA3.1 (서열 번호: 175)에 대한 서열 분석 결과의 단편 정렬을 도 15에 나타낸다. 이러한 결과는 동일한 세포에서 다수의 유전자 카피가 야로위아의 sgRNA/Cas9 엔도뉴클레아제 시스템에 의해 표적화되고 돌연변이되었음을 보여준다.

[0351] 실시예 11

[0352] 야로위아에서의 URA3 유전자 돌연변이 또는 결실.

[0353] 본 실시예는 마커 재활용에 사용하기 위해 야로위아에서의 URA3 유전자 돌연변이 또는 결실을 위한 동일한 벡터 시스템 상에 2개의 sgRNA 및 Cas9 엔도뉴클레아제를 발현하는 플라스미드의 구성 및 사용을 기술한다.

[0354] pYRH222는 FBA1 프로모터 (서열 번호: 12), 및 도 16a에 도시된 URA3.2 표적 서열 (서열 번호: 167)을 표적화하는 리보자임 플랭킹된 예비-sgRNA (RGR-URA3.2; 서열 번호: 166)를 암호화하는 FBA1L 프로모터 유도된 DNA 단편 하에 Cas9 엔도뉴클레아제 (서열 번호: 10)를 발현한다. YRH222 벡터는 플라스미드의 염색체외 복제를 수용하는 자가 복제 서열(autonomously replicating sequence) (ARS18, 서열 번호: 208)뿐만 아니라 TDH1 (GPD라고도 지칭됨) 프로모터 (서열 번호: 187) 하에서 발현된 하이그로마이신 항생제 내성 선택 마커 (서열 번호: 186)를 함유하였다 (문헌[PNAS, Fournier, P. et al., 1993, 90:4912-4916]). ARS18의 존재는 선별 압력이 없는 경우 세포가 플라스미드를 상실하게 만들었다.

[0355] pYRH282는 pYRH222로부터 유래되었다. pYRH235 유래의 RGR-URA3.1 (서열 번호: 164)을 암호화하는 DNA 단편에 융합된 FBA1L 프로모터 (서열 번호: 168)를 프라이머 RHO804 (서열 번호: 188) 및 RHO805 (서열 번호: 189)를 사용하여 PCR 증폭시켰다. 이어서, PCR 생성물을 BsiWI로 분해하고 pYRH222로 클로닝하였다. 클로닝된 유전자의 배향 및 서열 동일성은 서열 분석에 의해 확인되었고, 구조물은 pYRH282로 명명되었다.

[0356] pYRH283은 pYRH222로부터 유래되었다. RGR-URA3.3 (서열 번호: 191)을 암호화하는 DNA에 융합된 TDH1 프로모터 (서열 번호: 187)로 이루어진 BsiWI 부위 (서열 번호: 190)에 의해 플랭킹된 합성 DNA 단편을 IDT (아이오와주 코랄빌 소재)에 의해 합성하고 BsiWI 부위에서 pYRH222로 클로닝하였다. 클로닝된 유전자의 배향 및 서열 동일성은 서열 분석에 의해 확인되었고, 구조물은 pYRH283으로 명명되었다.

[0357] 야로위아 균주 ATCC20362의 자손을 pYRH222, pYRH282 및 pYRH283으로 형질전환시키고, 형질전환체를 300 mg/L의 하이그로마이신을 함유하는 YPD 플레이트 상에서 선별하였다. 상대적으로 높은 배경 성장이 DNA 대조군이 없는 플레이트 상에서 관찰되었다 (표 9). 각 구조물의 30개의 형질전환체를 무작위로 선별하고, 5-FOA가 함유된 SC 플레이트 상으로 도말하여 우라실 영양요구성에 대해 역선별하였다. DNA 대조군이 없는 플레이트로부터의 콜로니에서는 성장이 관찰되지 않았다. 4 내지 11 개의 패치는 pYRH222, pYRH282 및 pYRH283 형질전환체로 성장을 보였다. 콜로니 PCR을 프라이머 RHO610 (서열 번호: 192) 및 RHO611 (서열 번호: 193)을 사용하여 수행하여, sgRNA 표적화 부위를 함유하는 DNA 영역을 증폭시켰고, PCR 증폭된 생성물은 아가로스 겔 상에서 상이한 이동을 나타내었다 (도 17). PCR 생성물을 주형으로 사용하여 서열 분석을 수행하고, 프라이머 RHO704 (서열 번호: 194)를 사용하여 서열 분석을 수행하였다.

[0358] pYRH222 형질전환체의 경우에, 11개의 중 6개의 서열 분석이 성공적으로 수행되었고, 이들 모두는 URA3.2 표적 부위에서 돌연변이되었다 (도 16b; 서열 번호: 195-201). pYRH282의 경우, 성공적으로 서열 분석된 모두가 표적 부위에서 돌연변이를 나타냈으며, 그 중 2개가 2개의 표적 부위 사이에서 결실을 보였다 (도 16c; 서열 번호: 202 내지 204). pYRH283의 경우, 8개 중 7개의 성공적인 서열 분석이 표적 부위에서 돌연변이를 나타냈으며, 그 중 2개가 2개의 표적 부위 사이에서 결실을 보였고 (도 16d; 서열 번호: 205 내지 207), URA3 유전자

의 거의 완벽한 결실을 생성하였다.

[0359] 본 실시예는 야로위아에서 sgRNA/Cas9 엔도뉴클레아제 시스템을 사용하여 2개의 표적 부위 사이에서 표적화된 결실을 만들기 위해 동일한 플라스미드 상에서 2개의 가이드 RNA가 발현되는 것을 보여주는데, 동정을 겔을 사용하거나 서열 분석에 의해 수행하였다. 이러한 플라스미드 상의 ARS18 (서열 번호: 208)의 존재는 선별 압력이 없을 경우 세포가 플라스미드를 상실하게 하여, URA3 마커 재활용을 위해 플라스미드를 반복적으로 사용할 수 있게 하였다.

[0360] [표 9]

pYRH222, pYRH282 및 pYRH283 형질전환체의 분석. DNA 대조군을 포함하지 않는 각각의 형질전환 플레이트에 대한 형질전환체의 수를 기록하였다.

	Hyg 플레이트 상의 콜로니	5-FOA 상에서 패칭됨	5-FOA 상에서 성장	돌연변이를 표적화함/서열분석됨
DNA 대조군이 없음	131	30	0	
pYRH222 (URA3.2)	352	30	11	6/6
pYRH282 (URA3.2+URA3.1)	244	30	4	4/4 (2 개의 결실)
pYRH283 (URA3.2+3.3)	178	30	10	7/8 (2 개의 결실)

[0361]

[0362] 실시예 12

[0363] 유전자 불활성화를 위한 야로위아에서의 Csy4 (Cas6)의 사용

[0364] 본 실시예는 통상적이지 않은 효모에서 DNA 서열 (예를 들어, CAN1이지만 이에 한정되지 않음)을 표적화할 수 있는 RGEN 복합체를 형성할 수 있는 5' 캡이 없는 가이드 RNA를 생성하기 위한 Csy4 (Cas6으로도 지칭됨)의 사용을 기술한다.

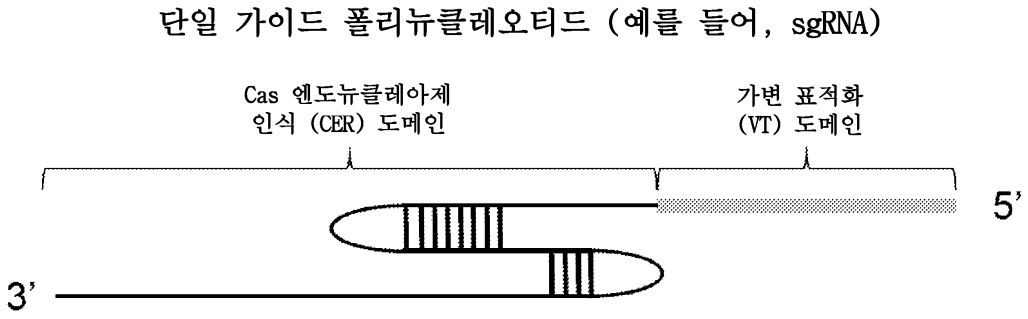
[0365] Csy4를 암호화하는 유전자 (Cas6으로도 알려짐)를 야로위아에서 CAN1 유전자 불활성화를 위해 28bp의 Csy4 인식 부위에 의해 플랭킹된 sgRNA를 표적화하는 CAN1를 암호화하는 DNA와 함께 Cas9 발현 플라스미드 상으로 도입하였다.

[0366] pYRH290은 FBA1 프로모터 (서열 번호: 12) 하에 Cas9 엔도뉴클레아제 (서열 번호: 10)를 FBA1 프로모터 (서열 번호: 210) 하에 Csy4 발현 (서열 번호: 209)을 위한 야로위아 리폴리티카 코돈-최적화된 유전자 하에 발현하였다. pYRH290은 또한 CAN1 표적 서열 (서열 번호: 214)을 표적화하는 28bp의 Csy4 엔도뉴클레아제 인식 서열 (서열 번호: 212) 플랭킹된 예비-sgRNA (서열 번호: 213)를 암호화하는 DNA 단편 (TDH1:28bp-gCAN1-28bp; 서열 번호: 211)을 함유하였다. Csy4에 의한 처리 후, 생성된 sgRNA (서열 번호: 222)는 8-뉴클레오티드 5'-플랭킹 서열 (서열 번호: 223) 및 20-뉴클레오티드 3'-플랭킹 서열 (서열 번호: 224)을 함유하였다.

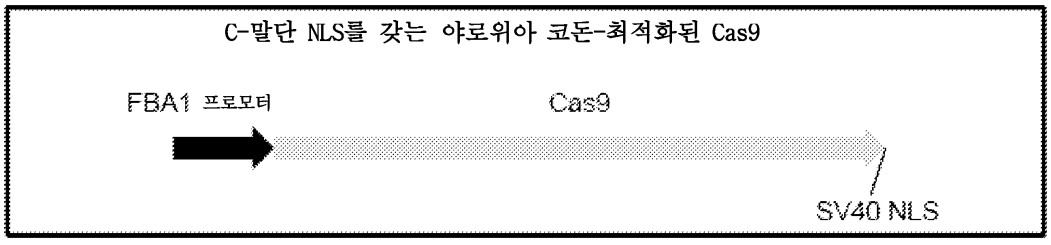
[0367] 야로위아 균주 ATCC20362의 Ura^- 유도체 (Y2224)를 pYRH290으로 형질전환시키고, 형질전환체를 우라실이 없는 CM 플레이트 상에서 선별하였다. 86개의 형질전환체를 카나바닌을 함유하는 CM 플레이트로 복제 평판하여, can1 돌연변이체를 선별하였다. 86개 중 40개의 형질전환체가 카나바닌을 함유하는 CM 플레이트 상에서 성장시켰다. 40개 중 16개의 카나바닌 내성 콜로니를 서열 분석하여, CAN1 표적 부위 (서열 번호: 214)에서 돌연변이를 확인하였고, 14개의 콜로니를 CAN1 표적 부위에서 돌연변이를 확인하였다. 도 18은 CAN1 표적 부위 (서열 번호: 215) 및 콜로니 14, 16, 18, 19, 24 및 25 (각각 서열 번호: 216 내지 221)의 CAN1 표적 서열에서 돌연변이를 포함하는 야생형 CAN1 유전자의 단편의 정렬을 도시한다.

도면

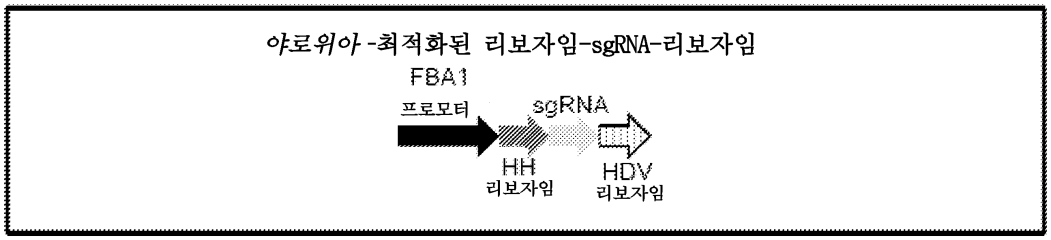
도면1



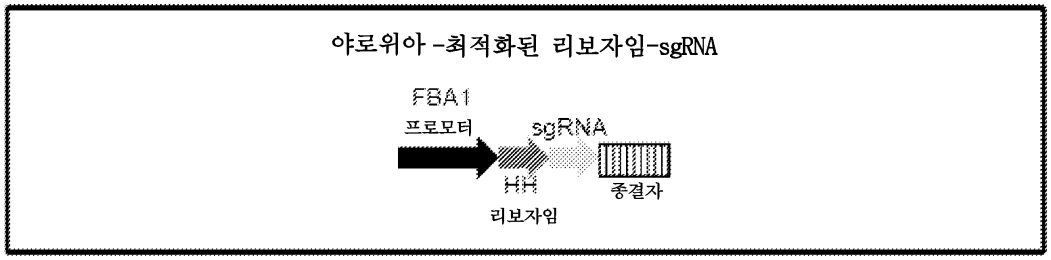
도면2a



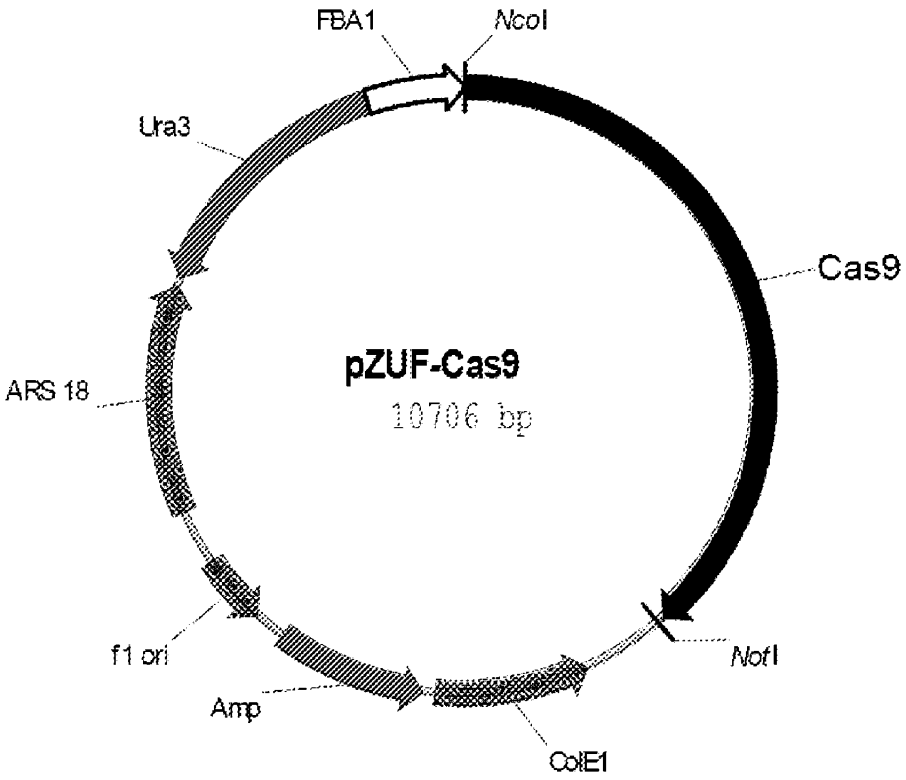
도면2b



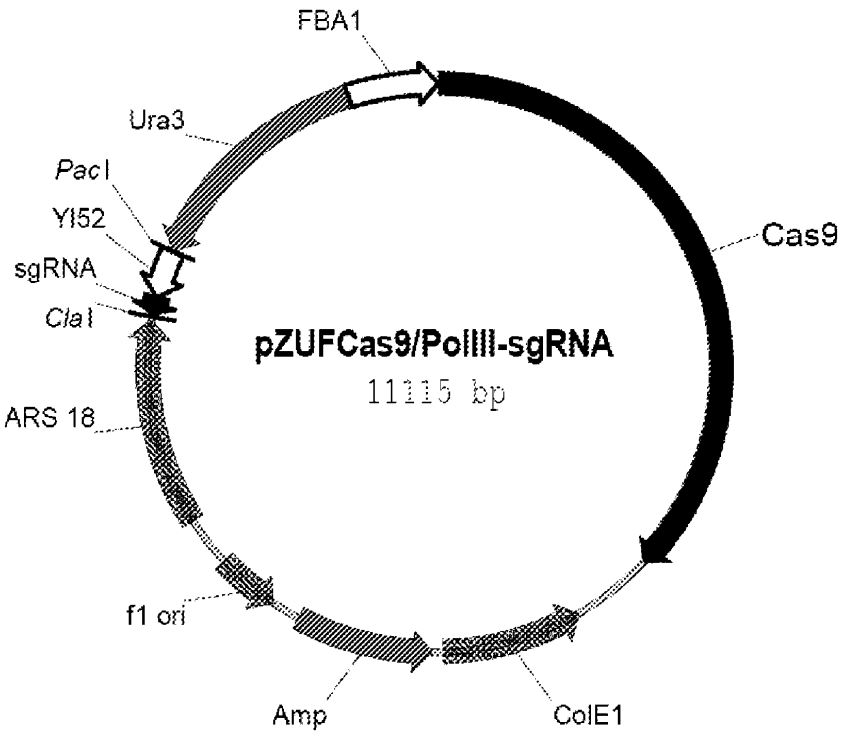
도면2c



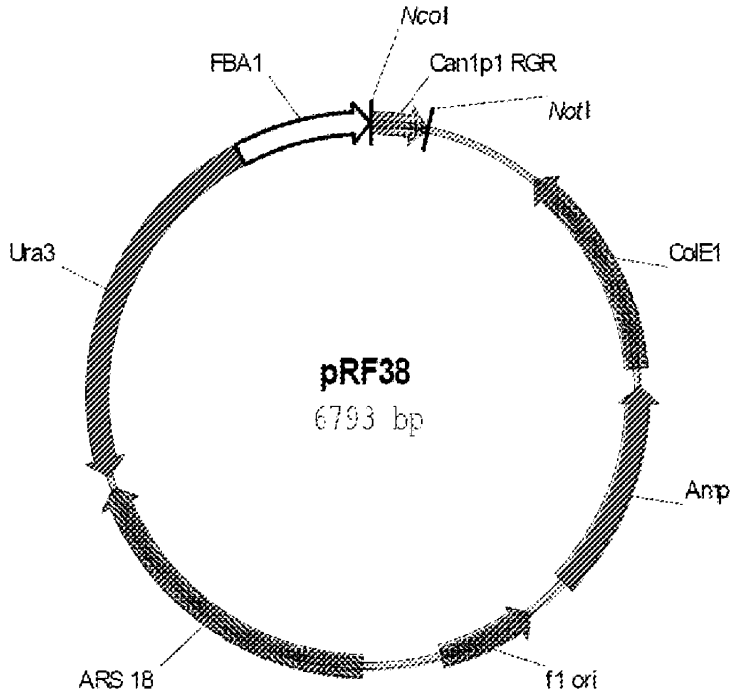
도면3a


FBA1
NcoI
Ura3
Cas9
pZUF-Cas9
10706 bp
ARS 18
f1 ori
NotI
Amp
ColE1

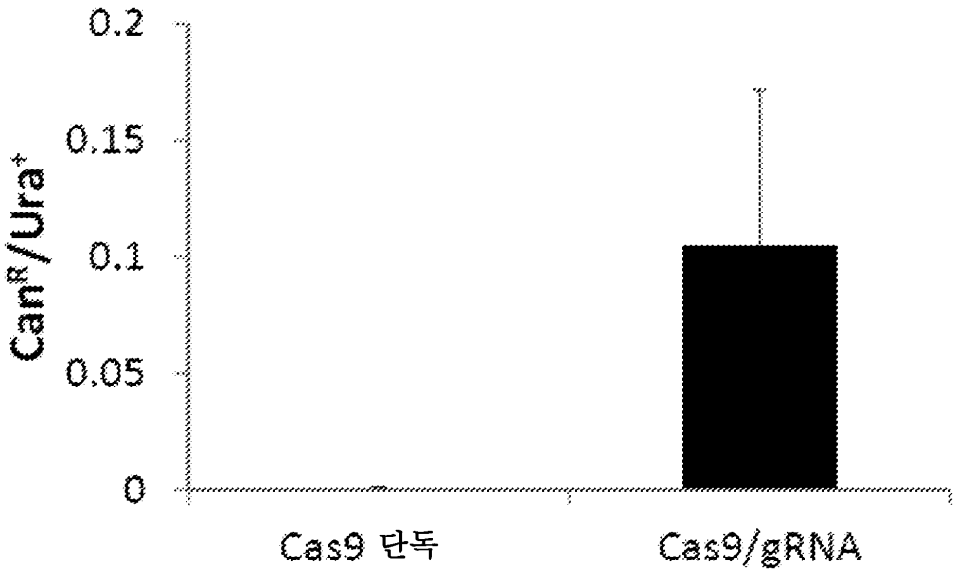
도면3b


FBA1
Ura3
PacI
YI52
sgRNA
ClaI
Cas9
pZUFCas9/PolIII-sgRNA
11115 bp
ARS 18
f1 ori
Amp
ColE1

도면3c


NcoI
FBA1
Can1p1 RGR
NotI
Ura3
ColE1
pRF38
6793 bp
Amp
ARS 18
f1 ori

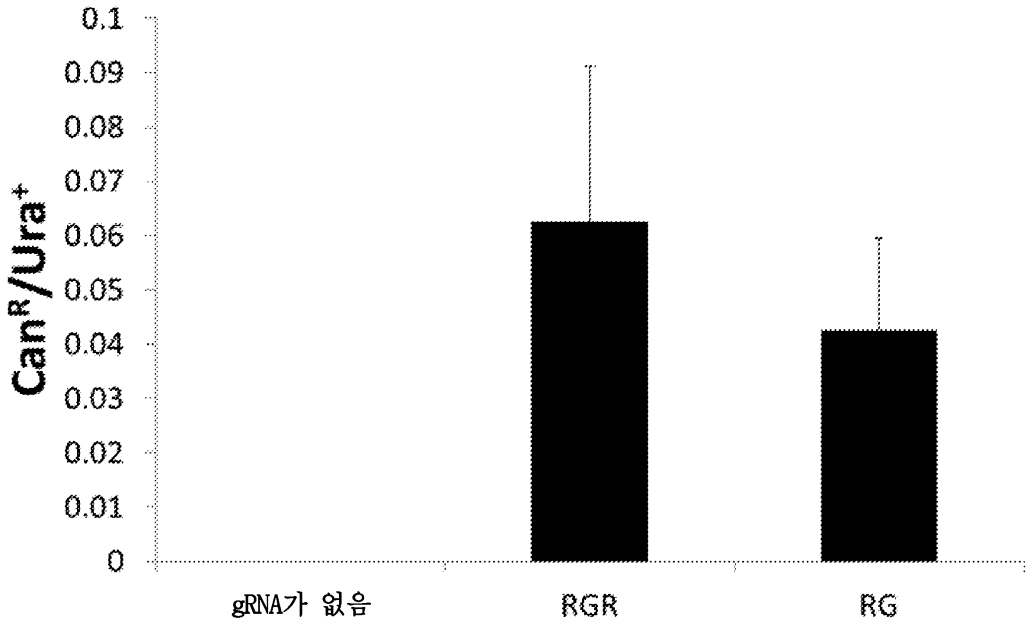
도면4


0.2
0.15
0.1
0.05
0
Can^R/Ura^+
Cas9 단독
Cas9/gRNA

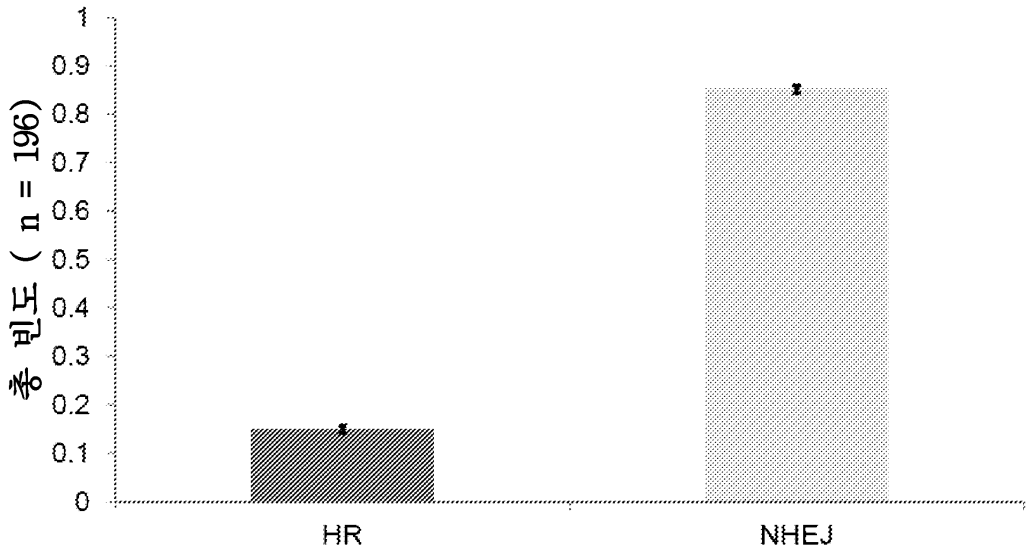
도면5

명칭	서열	수	빈도	서열 번호
WT	CAATGGAAAAGACATTTT**TCAAACGATTACCCACC-CTC**CGGGACTGAGGCCCAC	0	0	71
1	CAATGGAAAAGACATTTTCAAACGATTACCCACC--TCCGGGACTGAGGCCCAC	141	0.73	72
2	CAATGGAAAAGACATTTTCAAACGATTACCCAC---TCCGGGACTGAGGCCCAC	11	0.057	73
3	CAATGGAAAAGACATTTTCAAACGATTACCCACC-----GGGACTGAGGCCCAC	3	0.015	74
4	CAATGGAAAAGACATTTTCAAACGATTACCC--------GGGACTGAGGCCCAC	2	0.010	75
5	C----------ACTGAGGCCCAC	2	0.010	76
6	CAATGGAAAAGACATTTTCAAACGATTACCCACCACTCCGGGACTGAGGCCCAC	2	0.010	77
7	CAATGGAAAAGACATTTTCAAACGATTACC------TCCGGGACTGAGGCCCAC	1	0.005	78
8	CAATGGAAAAGACATTTTCAAACGATTACCC--------GGGACTGAGGCCCAC	1	0.005	79
9	CAATGGAAAAGACATTTTCAAACGATTACCCACC----------TGAGGCCCAC	1	0.005	80
10	CAATGGAAAAGACATTTTCAAACGATTACCCA------CGGGACTGAGGCCCAC	1	0.005	81
11	CAATGGAAAAGACATTTTCAAACGATTACACAC--------------GGCCCAC	1	0.005	82
12	CAATGGAAAAGACATTTTC----------TCCGGGACTGAGGCCCAC	1	0.005	83
13	CAATGGAAAAGACATTTTCAAACG----------CTCCGGGACTGAGGCCCAC	1	0.005	84
14	CAATGGAAAAGACATTTTCAAACGAT----------CCGGGACTGAGGCCCAC	1	0.005	85
15	CAATGGAAAAGACATTTTCAAACGATAC----------CCCAC	1	0.005	86
16	CAATGGAAAAGACATTTTCAAACGATTACCCACCCCTCCGGGACTGAGGCCCAC	1	0.005	87
17	CAATGGAAAAGACATTTTCAAACGATTACCC-----TCCGGGACTGAGGCCCAC	1	0.005	88
18	CAATGGAAAAGACATTTTCAAACGATTACCCACCCCTCCGGGACTGAGGCCCAC	1	0.005	89

도면6

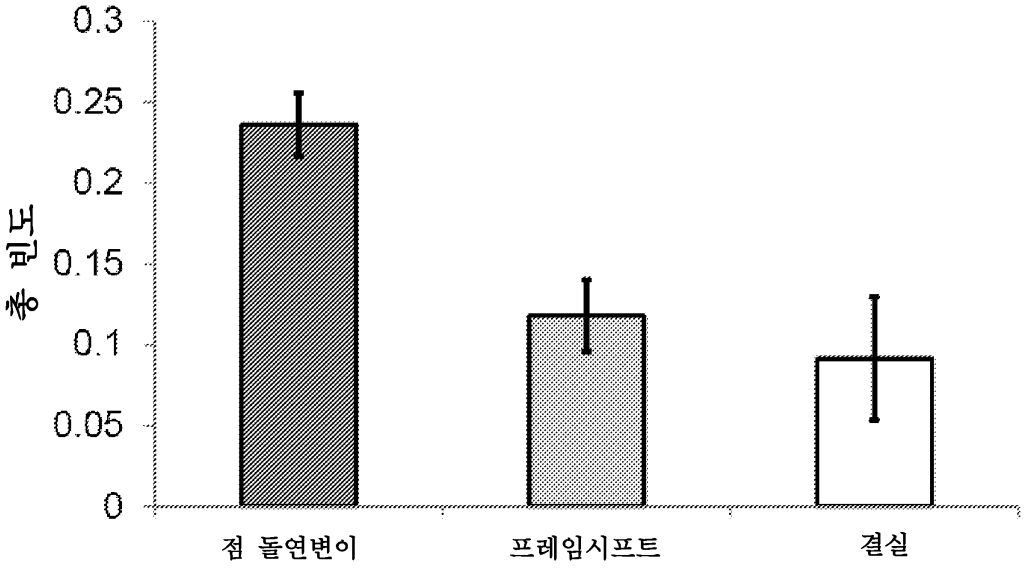


도면7



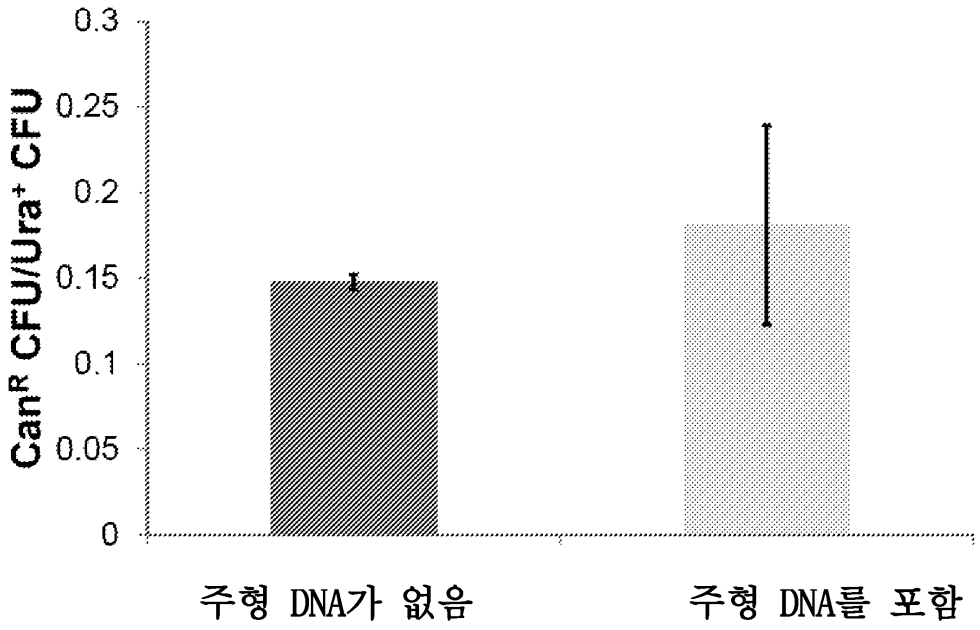
DNA 수복 경로

도면8

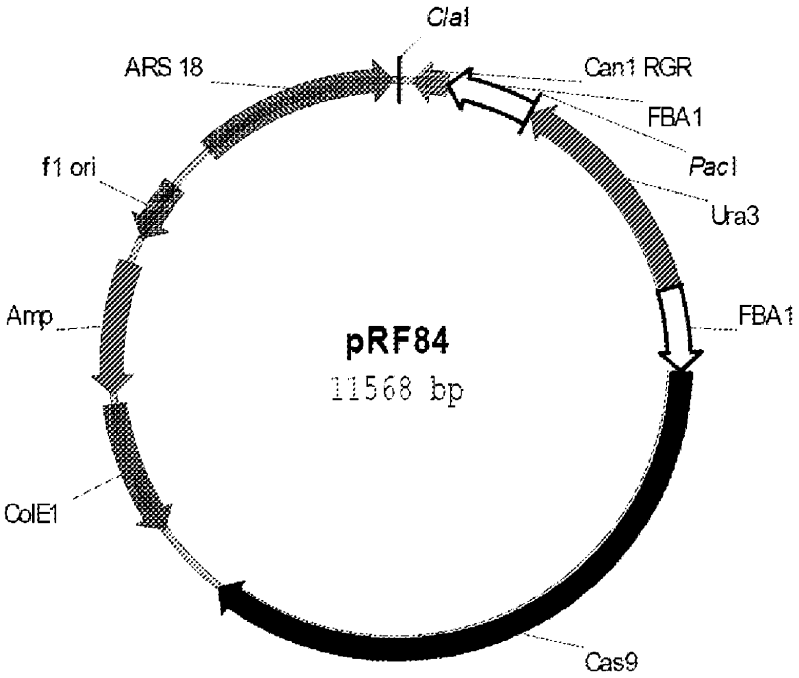


폴리뉴클레오티드 변형 주형의 유형

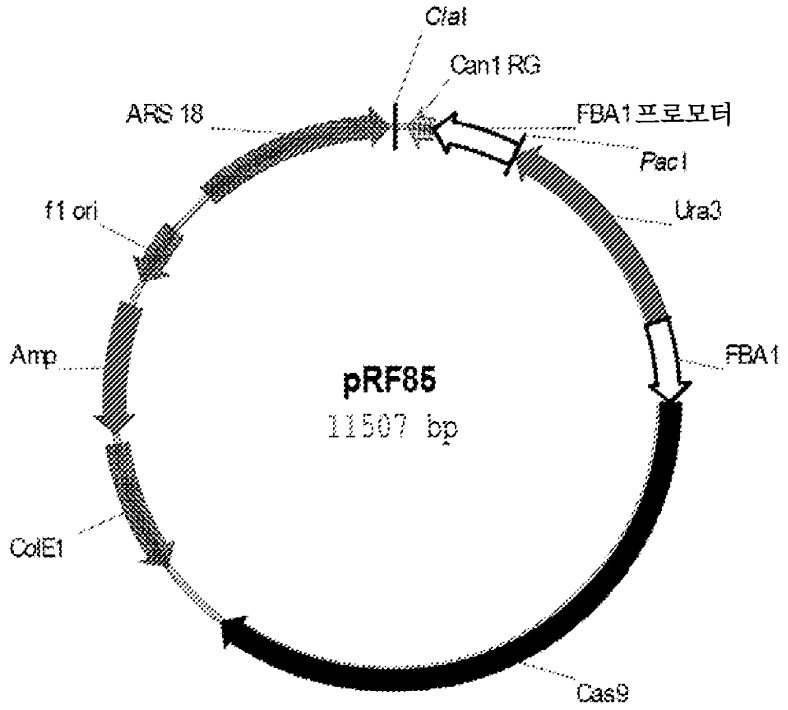
도면9


0.3
0.25
0.2
0.15
0.1
0.05
0
CanR CFU/Ura+ CFU
주형 DNA가 없음
주형 DNA를 포함

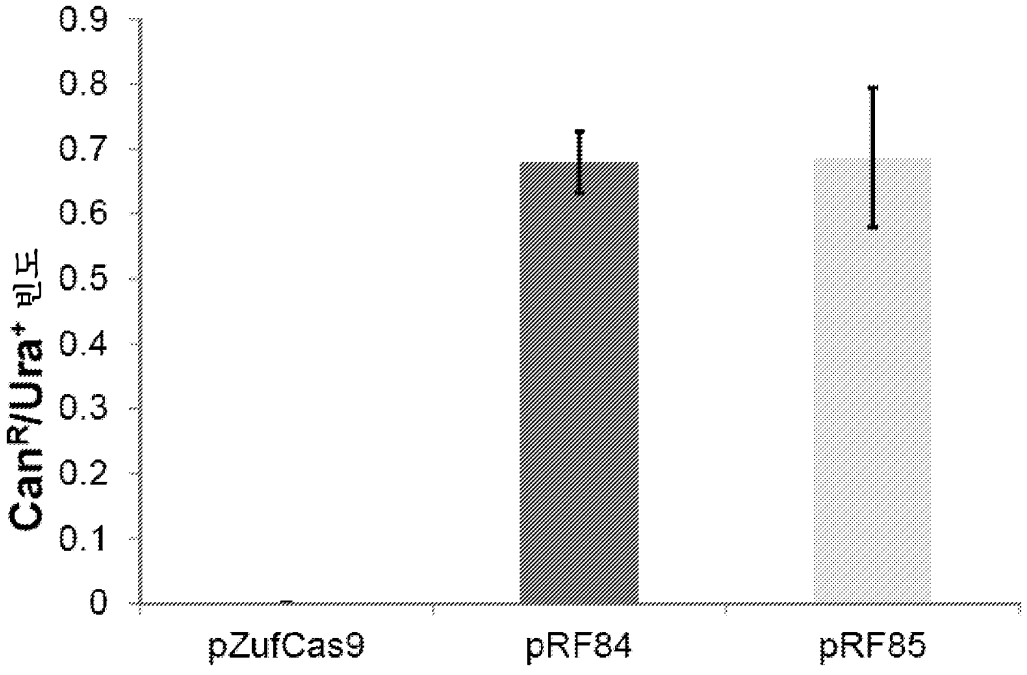
도면10a


ClaI
ARS 18
Can1 RGR
FBA1
PacI
Ura3
FBA1
f1 ori
Amp
pRF84
11568 bp
ColE1
Cas9

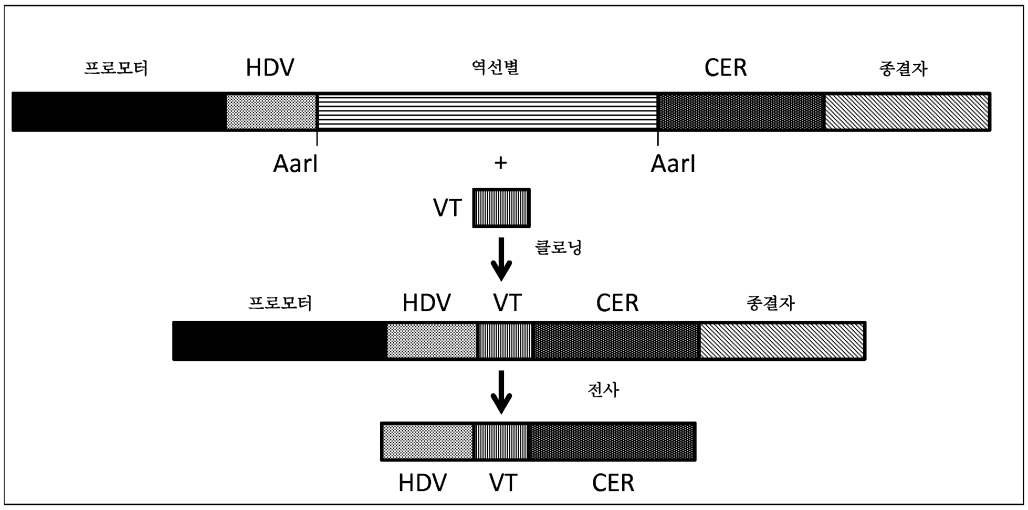
도면10b


ClaI
Can1 RG
ARS 18
FBA1 프로모터
PacI
f1 ori
Ura3
Amp
FBA1
pRF85
11507 bp
ColE1
Cas9

도면11


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
CanR/Ura+ 빈도
pZufCas9
pRF84
pRF85

도면12a

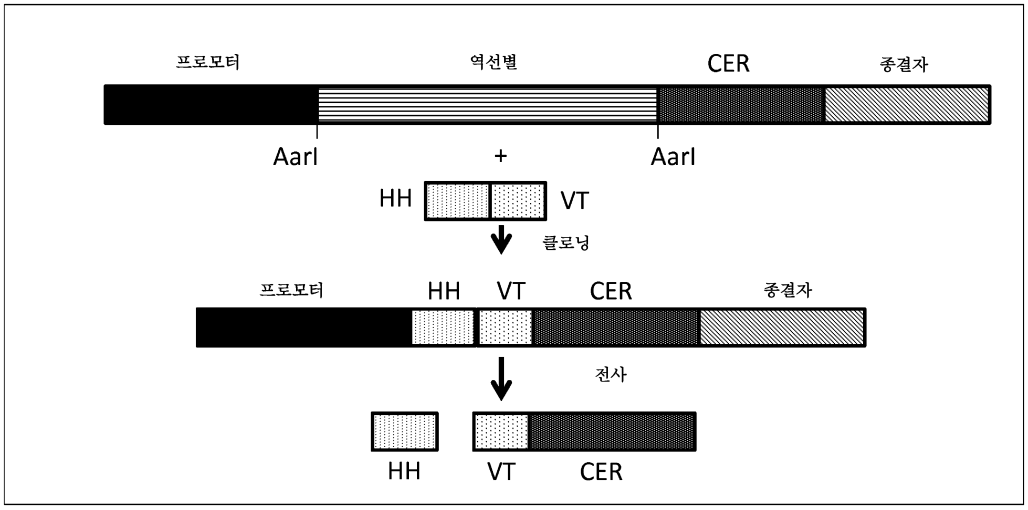


도면12b

AATGGGACtcaaacgattacccaccctcGTTT 서열 번호: 99

CCTGagtttgctaatgggtgggagCAAAATCT 서열 번호: 100

도면13a



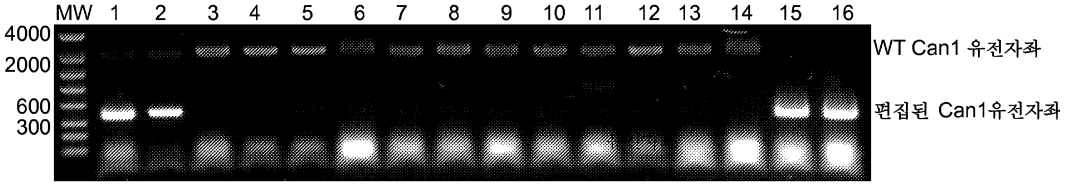
도면13b

CATGCGTAAACTGACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC TCAGTTTCAGTACGGTTCTA**GTTT** 서열 번호:162

GCATTTGACTGACTACTCAGGCACTCCTGCTTTGCTCATTCGAGCAG AGTCAAAGTCATGCCAAGAT**CAAAATCT** 서열 번호:163

←——— HH ———→ ←——— VT ———→

도면14



도면15

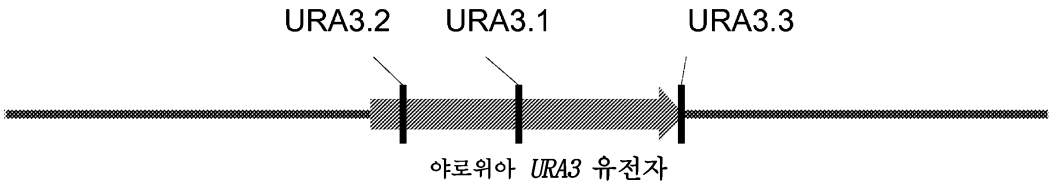
플라스미드 URA3

URA3.1	CGTCCTCCTTCTTCTGTTCAGAGACAGTTTCCTCGGCACCAGCTCGCAGGCCA	서열 번호 : 175
콜로니 1	CGTCCTCCTTCTTCTGTTCAGAGACAGTT-CCTCGGCACCAGCTCGCAGGCCA	서열 번호 : 176
콜로니 2	CGTCCTCCTTCTTCTGTTCAGAGACAGT--CCTCGGCACCAGCTCGCAGGCCA	서열 번호 : 177
콜로니 3	CGTCCTCCTTCTTCTGTTCAGAGACAGTT-CCTCGGCACCAGCTCGCAGGCCA	서열 번호 : 178
콜로니 5	CGTCCTCCTTCTTCTGTTCAGAGACAGTT-CCTCGGCACCAGCTCGCAGGCCA	서열 번호 : 179
콜로니 6	CGTCCTCCTTCTTCTGTTCAGAGACAGTT-CCTCGGCACCAGCTCGCAGGCCA	서열 번호 : 180

게놈 URA3

URA3.1	CGTCCTCCTTCTTCTGTTCAGAGACAGTTTCCTCGGCACCAGCTCGCAGGCCA	서열 번호 : 175
콜로니 1	CGTCCTCCTTCTTCTGTTCAGAGACAGTN-CCNCGGCCCC	서열 번호 : 181
콜로니 2	CGTCCTCCTTCTTCTGTTCAGAGACAGTTT**A**CCTCGGCACCAGCTCGCAGGCCA	서열 번호 : 182
콜로니 3	CGTCCTCCTTCTTCTGTTCAGAGACAGTTT---------CAGCTCGCAGGCCA	서열 번호 : 183
콜로니 5	CGTCCTCCTTCTTCTGTTCAGANACAGTTT----GGCACCANCTCGCAGGCCA	서열 번호 : 184
콜로니 6	CGTCCTCCTTCTTCTGTTCAGAGACAGTTT**C**CCTCGGCACCAGCTCGCAGGCCA	서열 번호 : 185

도면16a



야로위아 *URA3* 유전자

도면16b

pYRH222	URA3.2	
WT	AAACCAACCTGTGTGCTTCTCTGGATGTTACCACCACCA	서열 번호: 195
콜로니 3	AAACCAACCTGTGT**T**GNTNNNNNG	서열 번호: 196
콜로니 4	AAACCAACCTGTGTG**T**TTCTNGGNNTNNNCCCCCCCC	서열 번호: 197
콜로니 5	AAACCAACCTGTGTG**TG**TCTCTGGATGTTACCACCACNN	서열 번호: 198
콜로니 6	AAACCAACCTGTGTG**T**TTTCNNGGNNNT	서열 번호: 199
콜로니 9	AAACCAACCTGTGT**T**GCTTCTCTGGATGT	서열 번호: 200
콜로니 10	AAACCAACCTGTGNNTTCNNNNGNNNNTNCCNCCCCCNA	서열 번호: 201

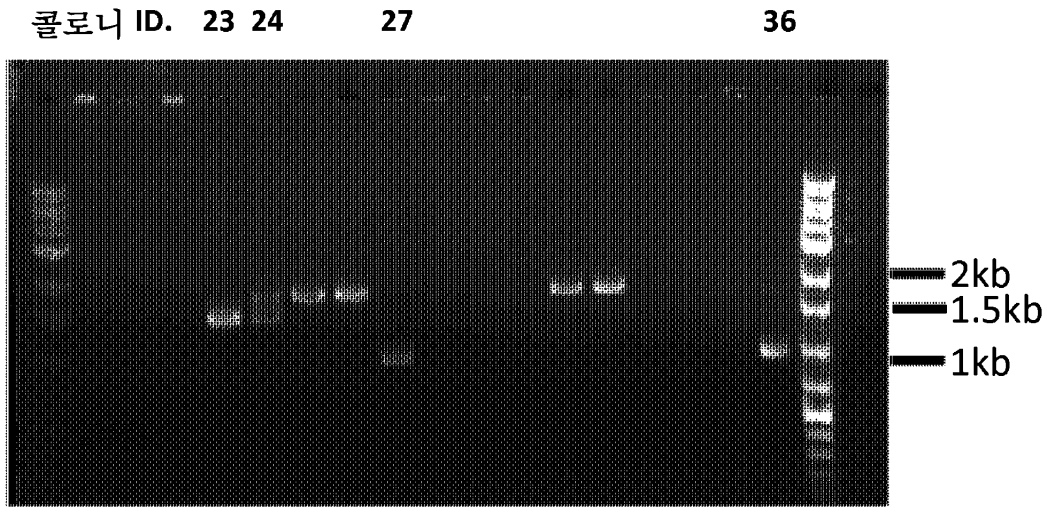
도면16c

pYRH282	URA3.2		URA3.1	서열 번호
WT	AAACCAA**CCT**GTGTGCTTCTCTGGATGTT		GCTGGTG**CCG**AGGAAACTGTCTCTGAACAGAAGA	202
23	AAACCAA**CCT**GTG----------	결실	----------AAACTGTCTCTGAACAGAAGA	203
24	AAACCAA**CCT**GTG----------	결실	----------AAACTGTCTCTGANCNNAANA	204

도면16d

pYRH283	URA3.2		URA3.3	서열 번호
WT	AAACCAA**CCT**GTGTGCTTCTCTGGATGTT		ATTAACTGTTAGAGGTTAGACTA**TGG**ATATGTA	205
27	AAACCAA**CCT**GTGT----------	결실	----------CTA**TGG**ATATGTA	206
36	AAACCAA**CCT**GTG----------	결실	----------**G**ATATGTA	207

도면17



도면18

CAN1-WT	GAAAAGACATTTTCAAACGATTACCCACCCTC**CGG**GACTGAGGCC	서열 번호: 215
콜로니14	GAAAAGACATTTTCAAACGATTACCCACC-TC**CGG**GACTGAGGCC	서열 번호: 216
콜로니16	GAAAAGACATTTTCAAACGATTACCCACCCTC**CGG**GACTGAGGCC	서열 번호: 217
콜로니18	GAAAAGACATTTTCAAACGATTACCCACC-TC**CGG**GACTGAGGCC	서열 번호: 218
콜로니19	GAAAANACNTTTTCAAACGATTACCCACC-TC**CGG**GACTGAGGCC	서열 번호: 219
콜로니24	GAAAAGACATTTTCAAACGATTACCCACC-TC**CGG**GACTGAGGCC	서열 번호: 220
콜로니25	GAAAAGACATTTTCAAACGATTACCCACC-TC**CGG**GACTGAGGCC	서열 번호: 221

서 열 목 록

SEQUENCE LISTING

<110> E. I. du Pont de Nemours and Company

Fan, Xiaochun

Frisch, Ryan

Hong, Seung-Pyo

<120> GENETIC TARGETING IN NON-CONVENTIONAL YEAST USING AN RNA-GUIDED ENDONUCLEASE

<130> CL6272

<150> 62/036652

<151> 2014-08-13

<160> 224

<170> PatentIn version 3.5

<210> 1

<211> 80

<212> RNA

<213> artificial sequence

<220><223> Cas9 endonuclease recognition (CER) domain

<400> 1

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60

ggcaccgagu cggugcuuuu 80

<210> 2

<211> 20

<212> DNA

<213> Yarrowia lipolytica

<400> 2

tccaagaaga ttgttcttct 20

<210> 3

<211> 20

<212> DNA

<213> Yarrowia lipolytica

<400> 3

ctccgtcatc cccggttctc 20

<210> 4

<211> 20

<212> DNA

<213> Yarrowia lipolytica

<400> 4

cggcgacttc tgtggccccg 20

<210> 5

<211> 300

<212> DNA

<213> Saccharomyces cerevisiae

<400> 5

aacaattatc tcaaaattca cccactcttc atctttgaaa agataatgta tgattatgct 60

ttcactcata tttatacaga aacttgatgt tttctttcga gtatatacaa ggtgattaca 120

tgtacgtttg aagtacaact ctagattttg tagtgccctc ttgggctagc ggtaaaggtg 180

cgcatttttt cacaccctac aatgttctgt tcaaaagatt ttggtcaaac gctgtagaag 240

tgaaagttgg tgcgcatgtt tcggcgttcg aaacttctcc gcagtgaaag ataaatgatc 300

<210> 6

<211> 300

<212> DNA

<213> Saccharomyces cerevisiae

<400> 6

taaaaatcaa tcaatcatcg tgtgttttat atgtctctta tctaagtata agaatatcca 60

tagttaatat tcacttacgc tacctttaa cctgtaatca ttgtcaacag gatatgttaa 120

cgacccacat tgataaacgc tagtatttct ttttcctctt cttattggcc ggctgtctct 180

atactcccct atagtctgtt tcttttcgtt tcgattgttt tacgtttgag gcctcgtggc 240

gcacatggta cgctgtggtg ctcgcggctg ggaacgaaac tctgggagct gcgattggca 300

<210> 7

<211> 300

<212> DNA

<213> Yarrowia lipolytica

<400> 7

attttttttg attttctttt ttgaccccgt cttcaattac acttcccaac tgggaacacc 60

cctctttatc gacccatttt aggtaattta ccctagccca ttgtctccat aaggaatatt 120

accctaaccc acagtccagg gtgcccaggt ccttctttgg ccaaatttta acttcggtcc 180

tatggcacag cggtagcgcg tgagattgca aatcttaagg tcccgagttc gaatctcggt 240

gggacctagt tatttttgat agataatttc gtgatgatta gaaacttaac gcaaaataat 300

<210> 8

<211> 20

<212> DNA

<213> Saccharomyces cerevisiae

<400> 8

ttttttttgtt ttttatgtct 20

<210> 9

<211> 4107

<212> DNA

<213> artificial sequence

<220><223> S. pyogenes Cas9

<400> 9

atggacaaga aatactccat cggcctggac attggaacca actctgtcgg ctgggctgtc 60

atcaccgacg agtacaaggt gccctccaag aaattcaagg tcctcggaaa caccgatcga 120

cactccatca agaaaaacct cattggtgcc ctgttgttcg attctggcga gactgccgaa 180

gctaccagac tcaagcgaac tgctcggcga cgttacaccc gacggaagaa ccgaatctgc 240

tacctgcagg agatcttttc caacgagatg gccaaggtgg acgattcgtt ctttcatcga 300

ctggaggaat ccttcctcgt cgaggaagac aagaaacacg agcgtcatcc catctttggc 360

aacattgtgg acgaggttgc ttaccacgag aagtatccta ccatctacca cctgcgaaag 420

aaactcgtcg attccaccga caaggcggat ctcagactta tctacctcgc tctggcacac 480

atgatcaagt ttcgaggtca tttcctcatc gagggcgatc tcaatcccga caacagcgat 540

gtggacaagc tgttcattca gctcgttcag acctacaacc agctgttcga ggaaaacccc 600

atcaatgcct ccggagtcga tgcaaaggcc atcttgtctg ctcgactctc gaagagcaga 660

cgactggaga acctcattgc ccaacttcct ggcgagaaaa agaacggact gtttggcaac 720

ctcattgccc tttctcttgg tctcacaccc aacttcaagt ccaacttcga tctggcggag 780

gacgccaagc tccagctgtc caaggacacc tacgacgatg acctcgacaa cctgcttgca 840

cagattggcg atcagtacgc cgacctgttt ctcgctgcca agaacctttc ggatgctatt 900

ctcttgtctg acattctgcg agtcaacacc gagatcacaa aggctcccct ttctgcctcc 960

atgatcaagc gatacgacga gcaccatcag gatctcacac tgctcaaggc tcttgtccga 1020

cagcaactgc ccgagaagta caaggagatc tttttcgatc agtcgaagaa cggctacgct 1080

ggatacatcg acggcggagc ctctcaggaa gagttctaca agttcatcaa gccaattctc 1140

gagaagatgg acggaaccga ggaactgctt gtcaagctca atcgagagga tctgcttcgg 1200

aagcaacgaa ccttcgacaa cggcagcatt cctcatcaga tccacctcgg tgagctgcac 1260

gccattcttc gacgtcagga agacttctac ccctttctca aggacaaccg agagaagatc 1320

gagaagattc ttacctttcg aatcccctac tatgttggtc ctcttgccag aggaaactct 1380

cgatttgctt ggatgactcg aaagtccgag gaaaccatca ctccctggaa cttcgaggaa 1440

gtcgtggaca agggtgcctc tgcacagtcc ttcatcgagc gaatgaccaa cttcgacaag 1500

aatctgccca acgagaaggt tcttcccaag cattcgctgc tctacgagta ctttacagtc 1560

tacaacgaac tcaccaaagt caagtacgtt accgagggaa tgcgaaagcc tgccttcttg 1620

tctggcgaac agaagaaagc cattgtcgat ctcctgttca agaccaaccg aaaggtcact 1680

gttaagcagc tcaaggagga ctacttcaag aaaatcgagt gtttcgacag cgtcgagatt 1740

tccggagttg aggaccgatt caacgcctct ttgggcacct atcacgatct gctcaagatt 1800

atcaaggaca aggattttct cgacaacgag gaaaacgagg acattctgga ggacatcgtg 1860

ctcactctta ccctgttcga agatcgggag atgatcgagg aacgactcaa gacatacgct 1920

cacctgttcg acgacaaggt catgaaacaa ctcaagcgac gtagatacac cggctgggga 1980

agactttcgc gaaagctcat caacggcatc agagacaagc agtccggaaa gaccattctg 2040

gactttctca agtccgatgg ctttgccaac cgaaacttca tgcagctcat tcacgacgat 2100

tctcttacct tcaaggagga catccagaag gcacaagtgt ccggtcaggg cgacagcttg 2160

cacgaacata ttgccaacct ggctggttcg ccagccatca agaaaggcat tctccagact 2220

gtcaaggttg tcgacgagct ggtgaaggtc atgggacgtc acaagcccga gaacattgtg 2280

atcgagatgg ccagagagaa ccagacaact caaaagggtc agaaaaactc gcgagagcgg 2340

atgaagcgaa tcgaggaagg catcaaggag ctgggatccc agattctcaa ggagcatccc 2400

gtcgagaaca ctcaactgca gaacgagaag ctgtatctct actatctgca gaatggtcga 2460

gacatgtacg tggatcagga actggacatc aatcgtctca gcgactacga tgtggaccac 2520

attgtccctc aatcctttct caaggacgat tctatcgaca acaaggtcct tacacgatcc 2580

gacaagaaca gaggcaagtc ggacaacgtt cccagcgaag aggtggtcaa aaagatgaag 2640

aactactggc gacagctgct caacgccaag ctcattaccc agcgaaagtt cgacaatctt 2700

accaaggccg agcgaggcgg tctgtccgag ctcgacaagg ctggcttcat caagcgtcaa 2760

ctcgtcgaga ccagacagat cacaaagcac gtcgcacaga ttctcgattc tcggatgaac 2820

accaagtacg acgagaacga caagctcatc cgagaggtca aggtgattac tctcaagtcc 2880

aaactggtct ccgatttccg aaaggacttt cagttctaca aggtgcgaga gatcaacaat 2940

taccaccatg cccacgatgc ttacctcaac gccgtcgttg gcactgcgct catcaagaaa 3000

taccccaagc tcgaaagcga gttcgtttac ggcgattaca aggtctacga cgttcgaaag 3060

atgattgcca agtccgaaca ggagattggc aaggctactg ccaagtactt cttttactcc 3120

aacatcatga acttttcaa gaccgagatc accttggcca acggagagat tcgaaagaga 3180

ccacttatcg agaccaacgg cgaaactgga gagatcgtgt gggacaaggg tcgagacttt 3240

gcaaccgtgc gaaaggttct gtcgatgcct caggtcaaca tcgtcaagaa aaccgaggtt 3300

cagactggcg gattctccaa ggagtcgatt ctgcccaagc gaaactccga caagctcatc 3360

gctcgaaaga aagactggga tcccaagaaa tacggtggct tcgattctcc taccgtcgcc 3420

tattccgtgc ttgtcgttgc gaaggtcgag aagggcaagt ccaaaaagct caagtccgtc 3480

aaggagctgc tcggaattac catcatggag cgatcgagct tcgagaagaa tcccatcgac 3540

ttcttggaag ccaagggtta caaggaggtc aagaaagacc tcattatcaa gctgcccaag 3600

tactctctgt tcgaactgga gaacggtcga aagcgtatgc tcgcctccgc tggcgagctg 3660

cagaagggaa acgagcttgc cttgccttcg aagtacgtca actttctcta tctggcttct 3720

cactacgaga agctcaaggg ttctcccgag gacaacgaac agaagcaact cttcgttgag 3780

cagcacaaac attacctcga cgagattatc gagcagattt ccgagttttc gaagcgagtc 3840

atcctggctg atgccaactt ggacaaggtg ctctctgcct acaacaagca tcgggacaaa 3900

cccattcgag aacaggcgga gaacatcatt cacctgttta ctcttaccaa cctgggtgct 3960

cctgcagctt tcaagtactt cgataccact atcgaccgaa agcggtacac atccaccaag 4020

gaggttctcg atgccaccct gattcaccag tccatcactg gcctgtacga gacccgaatc 4080

gacctgtctc agcttggtgg cgactaa 4107

<210> 10

<211> 4140

<212> DNA

<213> artificial sequence

<220><223> S. pyogenes Cas9 with NLS

<400> 10

atggacaaga aatactccat cggcctggac attggaacca actctgtcgg ctgggctgtc 60

atcaccgacg agtacaaggt gccctccaag aaattcaagg tcctcggaaa caccgatcga 120

cactccatca agaaaaacct cattggtgcc ctgttgttcg attctggcga gactgccgaa 180

gctaccagac tcaagcgaac tgctcggcga cgttacaccc gacggaagaa ccgaatctgc 240

tacctgcagg agatcttttc caacgagatg gccaaggtgg acgattcgtt ctttcatcga 300

ctggaggaat ccttcctcgt cgaggaagac aagaaacacg agcgtcatcc catctttggc 360

aacattgtgg acgaggttgc ttaccacgag aagtatccta ccatctacca cctgcgaaag 420

aaactcgtcg attccaccga caaggcggat ctcagactta tctacctcgc tctggcacac 480

atgatcaagt ttcgaggtca tttcctcatc gagggcgatc tcaatcccga caacagcgat 540

gtggacaagc tgttcattca gctcgttcag acctacaacc agctgttcga ggaaaacccc 600

atcaatgcct ccggagtcga tgcaaaggcc atcttgtctg ctcgactctc gaagagcaga 660

cgactggaga acctcattgc ccaacttcct ggcgagaaaa agaacggact gtttggcaac 720

ctcattgccc tttctcttgg tctcacaccc aacttcaagt ccaacttcga tctggcggag 780

gacgccaagc tccagctgtc caaggacacc tacgacgatg acctcgacaa cctgcttgca 840

cagattggcg atcagtacgc cgacctgttt ctcgctgcca agaacctttc ggatgctatt 900

ctcttgtctg acattctgcg agtcaacacc gagatcacaa aggctcccct ttctgcctcc 960

atgatcaagc gatacgacga gcaccatcag gatctcacac tgctcaaggc tcttgtccga 1020

cagcaactgc ccgagaagta caaggagatc tttttcgatc agtcgaagaa cggctacgct 1080

ggatacatcg acggcggagc ctctcaggaa gagttctaca agttcatcaa gccaattctc 1140

gagaagatgg acggaaccga ggaactgctt gtcaagctca atcgagagga tctgcttcgg 1200

aagcaacgaa ccttcgacaa cggcagcatt cctcatcaga tccacctcgg tgagctgcac 1260

gccattcttc gacgtcagga agacttctac ccctttctca aggacaaccg agagaagatc 1320

gagaagattc ttacctttcg aatcccctac tatgttggtc ctcttgccag aggaaactct 1380

cgatttgctt ggatgactcg aaagtccgag gaaaccatca ctccctggaa cttcgaggaa 1440

gtcgtggaca agggtgcctc tgcacagtcc ttcatcgagc gaatgaccaa cttcgacaag 1500

aatctgccca acgagaaggt tcttcccaag cattcgctgc tctacgagta ctttacagtc 1560

tacaacgaac tcaccaaagt caagtacgtt accgagggaa tgcgaaagcc tgccttcttg 1620

tctggcgaac agaagaaagc cattgtcgat ctcctgttca agaccaaccg aaaggtcact 1680

gttaagcagc tcaaggagga ctacttcaag aaaatcgagt gtttcgacag cgtcgagatt 1740

tccggagttg aggaccgatt caacgcctct ttgggcacct atcacgatct gctcaagatt 1800

atcaaggaca aggattttct cgacaacgag gaaaacgagg acattctgga ggacatcgtg 1860

ctcactctta ccctgttcga agatcgggag atgatcgagg aacgactcaa gacatacgct 1920

cacctgttcg acgacaaggt catgaaacaa ctcaagcgac gtagatacac cggctgggga 1980

agactttcgc gaaagctcat caacggcatc agagacaagc agtccggaaa gaccattctg 2040

gactttctca agtccgatgg ctttgccaac cgaaacttca tgcagctcat tcacgacgat 2100

tctcttacct tcaaggagga catccagaag gcacaagtgt ccggtcaggg cgacagcttg 2160

cacgaacata ttgccaacct ggctggttcg ccagccatca agaaaggcat tctccagact 2220

gtcaaggttg tcgacgagct ggtgaaggtc atgggacgtc acaagcccga gaacattgtg 2280

atcgagatgg ccagagagaa ccagacaact caaaagggtc agaaaaactc gcgagagcgg 2340

atgaagcgaa tcgaggaagg catcaaggag ctgggatccc agattctcaa ggagcatccc 2400

gtcgagaaca ctcaactgca gaacgagaag ctgtatctct actatctgca gaatggtcga 2460

gacatgtacg tggatcagga actggacatc aatcgtctca gcgactacga tgtggaccac 2520

attgtccctc aatcctttct caaggacgat tctatcgaca acaaggtcct tacacgatcc 2580

gacaagaaca gaggcaagtc ggacaacgtt cccagcgaag aggtggtcaa aaagatgaag 2640

aactactggc gacagctgct caacgccaag ctcattaccc agcgaaagtt cgacaatctt 2700

accaaggccg agcgaggcgg tctgtccgag ctcgacaagg ctggcttcat caagcgtcaa 2760

ctcgtcgaga ccagacagat cacaaagcac gtcgcacaga ttctcgattc tcggatgaac 2820

accaagtacg acgagaacga caagctcatc cgagaggtca aggtgattac tctcaagtcc 2880

aaactggtct ccgatttccg aaaggacttt cagttctaca aggtgcgaga gatcaacaat 2940

taccaccatg cccacgatgc ttacctcaac gccgtcgttg gcactgcgct catcaagaaa 3000

taccccaagc tcgaaagcga gttcgtttac ggcgattaca aggtctacga cgttcgaaag 3060

atgattgcca agtccgaaca ggagattggc aaggctactg ccaagtactt cttttactcc 3120

aacatcatga acttttcaa gaccgagatc accttggcca acggagagat tcgaaagaga 3180

ccacttatcg agaccaacgg cgaaactgga gagatcgtgt gggacaaggg tcgagacttt 3240

gcaaccgtgc gaaaggttct gtcgatgcct caggtcaaca tcgtcaagaa aaccgaggtt 3300

cagactggcg gattctccaa ggagtcgatt ctgcccaagc gaaactccga caagctcatc 3360

gctcgaaaga aagactggga tcccaagaaa tacggtggct tcgattctcc taccgtcgcc 3420

tattccgtgc ttgtcgttgc gaaggtcgag aagggcaagt ccaaaaagct caagtccgtc 3480

aaggagctgc tcggaattac catcatggag cgatcgagct tcgagaagaa tcccatcgac 3540

ttcttggaag ccaagggtta caaggaggtc aagaaagacc tcattatcaa gctgcccaag 3600

tactctctgt tcgaactgga gaacggtcga aagcgtatgc tcgcctccgc tggcgagctg 3660

cagaagggaa acgagcttgc cttgccttcg aagtacgtca actttctcta tctggcttct 3720

cactacgaga agctcaaggg ttctcccgag gacaacgaac agaagcaact cttcgttgag 3780

cagcacaaac attacctcga cgagattatc gagcagattt ccgagttttc gaagcgagtc 3840

atcctggctg atgccaactt ggacaaggtg ctctctgcct acaacaagca tcgggacaaa 3900

cccattcgag aacaggcgga gaacatcatt cacctgttta ctcttaccaa cctgggtgct 3960

cctgcagctt tcaagtactt cgataccact atcgaccgaa agcggtacac atccaccaag 4020

gaggttctcg atgccaccct gattcaccag tccatcactg gcctgtacga gacccgaatc 4080

gacctgtctc agcttggtgg cgactccaga gccgatccca agaaaaagcg aaaggtctaa 4140

<210> 11

<211> 1379

<212> PRT

<213> artificial sequence

<220><223> S. pyogenes Cas9 with NLS

<400> 11

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

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

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

65 70 75 80

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

85 90 95

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

100 105 110

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

115 120 125

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

130 135 140

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

145 150 155 160

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

165 170 175

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

180 185 190

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

195 200 205

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

210 215 220

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

225 230 235 240

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

245 250 255

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

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

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

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

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

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

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

530 535 540

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

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

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

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

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

675 680 685

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

690 695 700

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

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

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

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

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

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

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

1010 1015 1020

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

1025 1030 1035

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

1040 1045 1050

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

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

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

1085 1090 1095

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

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

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

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

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

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

Ser Arg Ala Asp Pro Lys Lys Lys Arg Lys Val
1370 1375

<210> 12

<211> 543

<212> DNA

<213> Yarrowia lipolytica

<400> 12

tcgacgttta aaccatcatc taagggcctc aaaactacct cggaactgct gcgctgatct 60

ggacaccaca gaggttccga gcactttagg ttgcaccaaa tgtcccacca ggtgcaggca 120

gaaaacgctg gaacagcgtg tacagtttgt cttaacaaaa agtgagggcg ctgaggtcga 180

gcagggtggt gtgacttgtt atagccttta gagctgcgaa agcgcgtatg gatttggctc 240

atcaggccag attgagggtc tgtggacaca tgtcatgtta gtgtacttca atcgcccccc 300

ggatatagcc ccgacaatag gccgtggcct catttttttg ccttccgcac atttccattg 360

ctcggtaccc acaccttgct tctcctgcac ttgccaacct taatactggt ttacattgac 420

caacatctta caagcggggg gcttgtctag ggtatatata aacagtggct ctcccaatcg 480

gttgccagtc tctttttttcc tttctttccc cacagattcg aaatctaaac tacacatcac 540

acc 543

<210> 13

<211> 4683

<212> DNA

<213> artificial sequence

<220><223> Cas9-NLS expression cassette

<400> 13

tcgacgttta aaccatcatc taagggcctc aaaactacct cggaactgct gcgctgatct 60

ggacaccaca gaggttccga gcactttagg ttgcaccaaa tgtcccacca ggtgcaggca 120

gaaaacgctg gaacagcgtg tacagtttgt cttaacaaaa agtgagggcg ctgaggtcga 180

gcagggtggt gtgacttgtt atagccttta gagctgcgaa agcgcgtatg gatttggctc 240

atcaggccag attgagggtc tgtggacaca tgtcatgtta gtgtacttca atcgcccccct 300

ggatatagcc ccgacaatag gccgtggcct catttttttg ccttccgcac atttccattg 360

ctcggtaccc acaccttgct tctcctgcac ttgccaacct taatactggt ttacattgac 420

caacatctta caagcggggg gcttgtctag ggtatatata aacagtggct ctcccaatcg 480

gttgccagtc tctttttttcc tttctttccc cacagattcg aaatctaaac tacacatcac 540

accatggaca agaaatactc catcggcctg gacattggaa ccaactctgt cggctgggct 600

gtcatcaccg acgagtacaa ggtgccctcc aagaaattca aggtcctcgg aaacaccgat 660

cgacactcca tcaagaaaaa cctcattggt gccctgttgt tcgattctgg cgagactgcc 720

gaagctacca gactcaagcg aactgctcgg cgacgttaca cccgacggaa gaaccgaatc 780

tgctacctgc aggagatctt ttccaacgag atggccaagg tggacgattc gttctttcat 840

cgactggagg aatccttcct cgtcgaggaa gacaagaaac acgagcgtca tcccatcttt 900

ggcaacattg tggacgaggt tgcttaccac gagaagtatc ctaccatcta ccacctgcga 960

aagaaactcg tcgattccac cgacaaggcg gatctcagac ttatctacct cgctctggca 1020

cacatgatca agtttcgagg tcatttcctc atcgagggcg atctcaatcc cgacaacagc 1080

gatgtggaca agctgttcat tcagctcgtt cagacctaca accagctgtt cgaggaaaac 1140

cccatcaatg cctccggagt cgatgcaaag gccatcttgt ctgctcgact ctcgaagagc 1200

agacgactgg agaacctcat tgcccaactt cctggcgaga aaaagaacgg actgtttggc 1260

aacctcattg ccctttctct tggtctcaca cccaacttca agtccaactt cgatctggcg 1320

gaggacgcca agctccagct gtccaaggac acctacgacg atgacctcga caacctgctt 1380

gcacagattg gcgatcagta cgccgacctg tttctcgctg ccaagaacct ttcggatgct 1440

attctcttgt ctgacattct gcgagtcaac accgagatca caaaggctcc cctttctgcc 1500

tccatgatca agcgatacga cgagcaccat caggatctca cactgctcaa ggctcttgtc 1560

cgacagcaac tgcccgagaa gtacaaggag atctttttcg atcagtcgaa gaacggctac 1620

gctggataca tcgacggcgg agcctctcag gaagagttct acaagttcat caagccaatt 1680

ctcgagaaga tggacggaac cgaggaactg cttgtcaagc tcaatcgaga ggatctgctt 1740

cggaagcaac gaaccttcga caacggcagc attcctcatc agatccacct cggtgagctg 1800

cacgccattc ttcgacgtca ggaagacttc tacccctttc tcaaggacaa ccgagagaag 1860

atcgagaaga ttcttacctt tcgaatcccc tactatgttg gtcctcttgc cagaggaaac 1920

tctcgatttg cttggatgac tcgaaagtcc gaggaaacca tcactccctg gaacttcgag 1980

gaagtcgtgg acaagggtgc ctctgcacag tccttcatcg agcgaatgac caacttcgac 2040

aagaatctgc ccaacgagaa ggttcttccc aagcattcgc tgctctacga gtactttaca 2100

gtctacaacg aactcaccaa agtcaagtac gttaccgagg gaatgcgaaa gcctgccttc 2160

ttgtctggcg aacagaagaa agccattgtc gatctcctgt tcaagaccaa ccgaaaggtc 2220

actgttaagc agctcaagga ggactacttc aagaaaatcg agtgtttcga cagcgtcgag 2280

atttccggag ttgaggaccg attcaacgcc tctttgggca cctatcacga tctgctcaag 2340

attatcaagg acaaggattt tctcgacaac gaggaaaacg aggacattct ggaggacatc 2400

gtgctcactc ttaccctgtt cgaagatcgg gagatgatcg aggaacgact caagacatac 2460

gctcacctgt tcgacgacaa ggtcatgaaa caactcaagc gacgtagata caccggctgg 2520

ggaagacttt cgcgaaagct catcaacggc atcagagaca agcagtccgg aaagaccatt 2580

ctggactttc tcaagtccga tggctttgcc aaccgaaact tcatgcagct cattcacgac 2640

gattctctta ccttcaagga ggacatccag aaggcacaag tgtccggtca gggcgacagc 2700

ttgcacgaac atattgccaa cctggctggt tcgccagcca tcaagaaagg cattctccag 2760

actgtcaagg ttgtcgacga gctggtgaag gtcatgggac gtcacaagcc cgagaacatt 2820

gtgatcgaga tggccagaga gaaccagaca actcaaaagg gtcagaaaaa ctcgcgagag 2880

cggatgaagc gaatcgagga aggcatcaag gagctgggat cccagattct caaggagcat 2940

cccgtcgaga acactcaact gcagaacgag aagctgtatc tctactatct gcagaatggt 3000

cgagacatgt acgtggatca ggaactggac atcaatcgtc tcagcgacta cgatgtggac 3060

cacattgtcc ctcaatcctt tctcaaggac gattctatcg acaacaaggt ccttacacga 3120

tccgacaaga acagaggcaa gtcggacaac gttcccagcg aagaggtggt caaaaagatg 3180

aagaactact ggcgacagct gctcaacgcc aagctcatta cccagcgaaa gttcgacaat 3240

cttaccaagg ccgagcgagg cggtctgtcc gagctcgaca aggctggctt catcaagcgt 3300

caactcgtcg agaccagaca gatcacaaag cacgtcgcac agattctcga ttctcggatg 3360

aacaccaagt acgacgagaa cgacaagctc atccgagagg tcaaggtgat tactctcaag 3420

tccaaactgg tctccgattt ccgaaaggac tttcagttct acaaggtgcg agagatcaac 3480

aattaccacc atgcccacga tgcttacctc aacgccgtcg ttggcactgc gctcatcaag 3540

aaatacccca agctcgaaag cgagttcgtt tacggcgatt acaaggtcta cgacgttcga 3600

aagatgattg ccaagtccga acaggagatt ggcaaggcta ctgccaagta cttcttttac 3660

tccaacatca tgaacttttt caagaccgag atcaccttgg ccaacggaga gattcgaaag 3720

agaccactta tcgagaccaa cggcgaaact ggagagatcg tgtgggacaa gggtcgagac 3780

tttgcaaccg tgcgaaaggt tctgtcgatg cctcaggtca acatcgtcaa gaaaaccgag 3840

gttcagactg gcggattctc caaggagtcg attctgccca agcgaaactc cgacaagctc 3900

atcgctcgaa agaaagactg ggatcccaag aaatacggtg gcttcgattc tcctaccgtc 3960

gcctattccg tgcttgtcgt tgcgaaggtc gagaagggca agtccaaaaa gctcaagtcc 4020

gtcaaggagc tgctcggaat taccatcatg gagcgatcga gcttcgagaa gaatcccatc 4080

gacttcttgg aagccaaggg ttacaaggag gtcaagaaag acctcattat caagctgccc 4140

aagtactctc tgttcgaact ggagaacggt cgaaagcgta tgctcgcctc cgctggcgag 4200

ctgcagaagg gaaacgagct tgccttgcct tcgaagtacg tcaactttct ctatctggct 4260

tctcactacg agaagctcaa gggttctccc gaggacaacg aacagaagca actcttcgtt 4320

gagcagcaca aacattacct cgacgagatt atcgagcaga tttccgagtt ttcgaagcga 4380

gtcatcctgg ctgatgccaa cttggacaag gtgctctctg cctacaacaa gcatcgggac 4440

aaacccattc gagaacaggc ggagaacatc attcacctgt ttactcttac caacctgggt 4500

gctcctgcag ctttcaagta cttcgatacc actatcgacc gaaagcggta cacatccacc 4560

aaggaggttc tcgatgccac cctgattcac cagtccatca ctggcctgta cgagacccga 4620

atcgacctgt ctcagcttgg tggcgactcc agagccgatc ccaagaaaaa gcgaaaggtc 4680

taa 4683

<210> 14

<211> 10706

<212> DNA

<213> artificial sequence

<220><223> pZUFCas9 plasmid

<400> 14

catggacaag aaatactcca tcggcctgga cattggaacc aactctgtcg gctgggctgt 60

catcaccgac gagtacaagg tgccctccaa gaaattcaag gtcctcggaa acaccgatcg 120

acactccatc aagaaaaacc tcattggtgc cctgttgttc gattctggcg agactgccga 180

agctaccaga ctcaagcgaa ctgctcggcg acgttacacc cgacggaaga accgaatctg 240

ctacctgcag gagatctttt ccaacgagat ggccaaggtg gacgattcgt tctttcatcg 300

actggaggaa tccttcctcg tcgaggaaga caagaaacac gagcgtcatc ccatctttgg 360

caacattgtg gacgaggttg cttaccacga gaagtatcct accatctacc acctgcgaaa 420

gaaactcgtc gattccaccg acaaggcgga tctcagactt atctacctcg ctctggcaca 480

catgatcaag tttcgaggtc atttcctcat cgagggcgat ctcaatcccg acaacagcga 540

tgtggacaag ctgttcattc agctcgttca gacctacaac cagctgttcg aggaaaaccc 600

catcaatgcc tccggagtcg atgcaaaggc catcttgtct gctcgactct cgaagagcag 660

acgactggag aacctcattg cccaacttcc tggcgagaaa aagaacggac tgtttggcaa 720

cctcattgcc ctttctcttg gtctcacacc caacttcaag tccaacttcg atctggcgga 780

ggacgccaag ctccagctgt ccaaggacac ctacgacgat gacctcgaca acctgcttgc 840

acagattggc gatcagtacg ccgacctgtt tctcgctgcc aagaaccttt cggatgctat 900

tctcttgtct gacattctgc gagtcaacac cgagatcaca aaggctcccc tttctgcctc 960

catgatcaag cgatacgacg agcaccatca ggatctcaca ctgctcaagg ctcttgtccg 1020

acagcaactg cccgagaagt acaaggagat ctttttcgat cagtcgaaga acggctacgc 1080

tggatacatc gacggcggag cctctcagga agagttctac aagttcatca agccaattct 1140

cgagaagatg gacggaaccg aggaactgct tgtcaagctc aatcgagagg atctgcttcg 1200

gaagcaacga accttcgaca acggcagcat tcctcatcag atccacctcg gtgagctgca 1260

cgccattctt cgacgtcagg aagacttcta ccccttttctc aaggacaacc gagagaagat 1320

cgagaagatt cttacctttc gaatccccta ctatgttggt cctcttgcca gaggaaactc 1380

tcgatttgct tggatgactc gaaagtccga ggaaaccatc actccctgga acttcgagga 1440

agtcgtggac aagggtgcct ctgcacagtc cttcatcgag cgaatgacca acttcgacaa 1500

gaatctgccc aacgagaagg ttcttcccaa gcattcgctg ctctacgagt actttacagt 1560

ctacaacgaa ctcaccaaag tcaagtacgt taccgaggga atgcgaaagc ctgccttctt 1620

gtctggcgaa cagaagaaag ccattgtcga tctcctgttc aagaccaacc gaaaggtcac 1680

tgttaagcag ctcaaggagg actacttcaa gaaaatcgag tgtttcgaca gcgtcgagat 1740

ttccggagtt gaggaccgat tcaacgcctc tttgggcacc tatcacgatc tgctcaagat 1800

tatcaaggac aaggattttc tcgacaacga ggaaaacgag gacattctgg aggacatcgt 1860

gctcactctt accctgttcg aagatcggga gatgatcgag gaacgactca agacatacgc 1920

tcacctgttc gacgacaagg tcatgaaaca actcaagcga cgtagataca ccggctgggg 1980

aagactttcg cgaaagctca tcaacggcat cagagacaag cagtccggaa agaccattct 2040

ggactttctc aagtccgatg gctttgccaa ccgaaacttc atgcagctca ttcacgacga 2100

ttctcttacc ttcaaggagg acatccagaa ggcacaagtg tccggtcagg gcgacagctt 2160

gcacgaacat attgccaacc tggctggttc gccagccatc aagaaaggca ttctccagac 2220

tgtcaaggtt gtcgacgagc tggtgaaggt catgggacgt cacaagcccg agaacattgt 2280

gatcgagatg gccagagaga accagacaac tcaaaagggt cagaaaaact cgcgagagcg 2340

gatgaagcga atcgaggaag gcatcaagga gctgggatcc cagattctca aggagcatcc 2400

cgtcgagaac actcaactgc agaacgagaa gctgtatctc tactatctgc agaatggtcg 2460

agacatgtac gtggatcagg aactggacat caatcgtctc agcgactacg atgtggacca 2520

cattgtccct caatcctttc tcaaggacga ttctatcgac aacaaggtcc ttacacgatc 2580

cgacaagaac agaggcaagt cggacaacgt tcccagcgaa gaggtggtca aaaagatgaa 2640

gaactactgg cgacagctgc tcaacgccaa gctcattacc cagcgaaagt tcgacaatct 2700

taccaaggcc gagcgaggcg gtctgtccga gctcgacaag gctggcttca tcaagcgtca 2760

actcgtcgag accagacaga tcacaaagca cgtcgcacag attctcgatt ctcggatgaa 2820

caccaagtac gacgagaacg acaagctcat ccgagaggtc aaggtgatta ctctcaagtc 2880

caaactggtc tccgatttcc gaaaggactt tcagttctac aaggtgcgag agatcaacaa 2940

ttaccaccat gcccacgatg cttacctcaa cgccgtcgtt ggcactgcgc tcatcaagaa 3000

ataccccaag ctcgaaagcg agttcgttta cggcgattac aaggtctacg acgttcgaaa 3060

gatgattgcc aagtccgaac aggagattgg caaggctact gccaagtact tcttttactc 3120

caacatcatg aacttttca agaccgagat caccttggcc aacggagaga ttcgaaagag 3180

accacttatc gagaccaacg gcgaaactgg agagatcgtg tgggacaagg gtcgagactt 3240

tgcaaccgtg cgaaaggttc tgtcgatgcc tcaggtcaac atcgtcaaga aaaccgaggt 3300

tcagactggc ggattctcca aggagtcgat tctgcccaag cgaaactccg acaagctcat 3360

cgctcgaaag aaagactggg atcccaagaa atacggtggc ttcgattctc ctaccgtcgc 3420

ctattccgtg cttgtcgttg cgaaggtcga gaagggcaag tccaaaaagc tcaagtccgt 3480

caaggagctg ctcggaatta ccatcatgga gcgatcgagc ttcgagaaga atcccatcga 3540

cttcttggaa gccaagggtt acaaggaggt caagaaagac ctcattatca agctgcccaa 3600

gtactctctg ttcgaactgg agaacggtcg aaagcgtatg ctcgcctccg ctggcgagct 3660

gcagaaggga aacgagcttg ccttgccttc gaagtacgtc aactttctct atctggcttc 3720

tcactacgag aagctcaagg gttctcccga ggacaacgaa cagaagcaac tcttcgttga 3780

gcagcacaaa cattacctcg acgagattat cgagcagatt tccgagtttt cgaagcgagt 3840

catcctggct gatgccaact tggacaaggt gctctctgcc tacaacaagc atcgggacaa 3900

acccattcga gaacaggcgg agaacatcat tcacctgttt actcttacca acctgggtgc 3960

tcctgcagct ttcaagtact tcgataccac tatcgaccga aagcggtaca catccaccaa 4020

ggaggttctc gatgccaccc tgattcacca gtccatcact ggcctgtacg agacccgaat 4080

cgacctgtct cagcttggtg gcgactccag agccgatccc aagaaaaagc gaaaggtcta 4140

agcggccgca agtgtggatg gggaagtgag tgcccggttc tgtgtgcaca attggcaatc 4200

caagatggat ggattcaaca cagggatata gcgagctacg tggtggtgcg aggatatagc 4260

aacggatatt tatgtttgac acttgagaat gtacgataca agcactgtcc aagtacaata 4320

ctaaacatac tgtacatact catactcgta cccgggcaac ggtttcactt gagtgcagtg 4380

gctagtgctc ttactcgtac agtgtgcaat actgcgtatc atagtctttg atgtatatcg 4440

tattcattca tgttagttgc gtacgagccg gaagcataaa gtgtaaagcc tggggtgcct 4500

aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa 4560

acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta 4620

ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc 4680

gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg 4740

caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt 4800

tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa 4860

gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct 4920

ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc 4980

cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg 5040

tcgttcgctc caagctgggc tgtgtgcacg aacccccccgt tcagcccgac cgctgcgcct 5100

tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag 5160

cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga 5220

agtggtggcc taactacggc tacactagaa ggacagtatt tggtatctgc gctctgctga 5280

agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg 5340

gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag 5400

aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag 5460

ggattttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat 5520

gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct 5580

taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac 5640

tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa 5700

tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg 5760

gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt 5820

gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca 5880

ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt 5940

cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct 6000

tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg 6060

cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg 6120

agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg 6180

cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa 6240

aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt 6300

aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt 6360

gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt 6420

gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca 6480

tgagcggata catatttgaa tgtatttaga aaaataaaca aatagggggtt ccgcgcacat 6540

ttccccgaaa agtgccacct gacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg 6600

tggttacgcg cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt 6660

tcttcccttc ctttctcgcc acgttcgccg gctttccccg tcaagctcta aatcgggggc 6720

tcccttttagg gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg 6780

gtgatggttc acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg 6840

agtccacgtt ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct 6900

cggtctattc ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg 6960

agctgattta acaaaaattt aacgcgaatt ttaacaaaat attaacgctt acaatttcca 7020

ttcgccattc aggctgcgca actgttggga agggcgatcg gtgcgggcct cttcgctatt 7080

acgccagctg gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa cgccagggtt 7140

ttcccagtca cgacgttgta aaacgacggc cagtgaattg taatacgact cactataggg 7200

cgaattgggt accgggcccc ccctcgaggt cgatggtgtc gataagcttg atatcgaatt 7260

catgtcacac aaaccgatct tcgcctcaag gaaacctaat tctacatccg agagactgcc 7320

gagatccagt ctacactgat taattttcgg gccaataatt taaaaaaatc gtgttatata 7380

atattatatg tattatatat atacatcatg atgatactga cagtcatgtc ccattgctaa 7440

atagacagac tccatctgcc gcctccaact gatgttctca atatttaagg ggtcatctcg 7500
cattgtttaa taataaacag actccatcta ccgcctccaa atgatgttct caaaatatat 7560
tgtatgaact tatttttatt acttagtatt attagacaac ttacttgctt tatgaaaaac 7620
acttcctatt taggaaacaa tttataatgg cagttcgttc atttaacaat ttatgtagaa 7680
taaatgttat aaatgcgtat gggaaatctt aaatatggat agcataaatg atatctgcat 7740
tgcctaattc gaaatcaaca gcaacgaaaa aaatcccttg tacaacataa atagtcatcg 7800
agaaatatca actatcaaag aacagctatt cacacgttac tattgagatt attattggac 7860

gagaatcaca cactcaactg tctttctctc ttctagaaat acaggtacaa gtatgtacta 7920
ttctcattgt tcatacttct agtcatttca tcccacatat tccttggatt tctctccaat 7980
gaatgacatt ctatcttgca aattcaacaa ttataataag atataccaaa gtagcggtat 8040
agtggcaatc aaaaagcttc tctggtgtgc ttctcgtatt tatttttatt ctaatgatcc 8100
attaaaggta tatatttatt tcttgttata taatcctttt gtttattaca tgggctggat 8160
acataaaggt attttgattt aatttttgc ttaaattcaa tccccctcg ttcagtgtca 8220
actgtaatgg taggaaatta ccatactttt gaagaagcaa aaaaaatgaa agaaaaaaaa 8280

aatcgtattt ccaggttaga cgttccgcag aatctagaat gcggtatgcg gtacattgtt 8340
cttcgaacgt aaaagttgcg ctccctgaga tattgtacat ttttgctttt acaagtacaa 8400
gtacatcgta caactatgta ctactgttga tgcatccaca acagtttgtt ttgttttttt 8460
ttgttttttt tttttctaat gattcattac cgctatgtat acctacttgt acttgtagta 8520
agccgggtta ttggcgttca attaatcata gacttatgaa tctgcacggt gtgcgctgcg 8580
agttactttt agcttatgca tgctacttgg gtgtaatatt gggatctgtt cggaaatcaa 8640
cggatgctca atcgatttcg acagtaatta attaagtcat acacaagtca gctttcttcg 8700

agcctcatat aagtataagt agttcaacgt attagcactg tacccagcat ctccgtatcg 8760
agaaacacaa caacatgccc cattggacag atcatgcgga tacacaggtt gtgcagtatc 8820
atacatactc gatcagacag gtcgtctgac catcatacaa gctgaacaag cgctccatac 8880
ttgcacgctc tctatataca cagttaaatt acatatccat agtctaacct ctaacagtta 8940
atcttctggt aagcctccca gccagccttc tggtatcgct tggcctcctc aataggatct 9000
cggttctggc cgtacagacc tcggccgaca attatgatat ccgttccggt agacatgaca 9060
tcctcaacag ttcggtactg ctgtccgaga gcgtctccct tgtcgtcaag acccaccccg 9120

ggggtcagaa taagccagtc ctcagagtcg cccttaggtc ggttctgggc aatgaagcca 9180

accacaaact cggggtcgga tcgggcaagc tcaatggtct gcttggagta ctcgccagtg 9240

gccagagagc ccttgcaaga cagctcggcc agcatgagca gacctctggc cagcttctcg 9300

ttgggagagg ggactaggaa ctccttgtac tgggagttct cgtagtcaga gacgtcctcc 9360

ttcttctgtt cagagacagt ttcctcggca ccagctcgca ggccagcaat gattccggtt 9420

ccgggtacac cgtgggcgtt ggtgatatcg gaccactcgg cgattcggtg acaccggtac 9480

tggtgcttga cagtgttgcc aatatctgcg aactttctgt cctcgaacag gaagaaaccg 9540

tgcttaagag caagttcctt gaggggggagc acagtgccgg cgtaggtgaa gtcgtcaatg 9600

atgtcgatat gggttttgat catgcacaca taaggtccga ccttatcggc aagctcaatg 9660

agctccttgg tggtggtaac atccagagaa gcacacaggt tggttttctt ggctgccacg 9720

agcttgagca ctcgagcggc aaaggcggac ttgtggacgt tagctcgagc ttcgtaggag 9780

ggcattttgg tggtgaagag gagactgaaa taaatttagt ctgcagaact ttttatcgga 9840

accttatctg gggcagtgaa gtatatgtta tggtaatagt tacgagttag ttgaacttat 9900

agatagactg gactatacgg ctatcggtcc aaattagaaa gaacgtcaat ggctctctgg 9960

gcgtcgcctt tgccgacaaa aatgtgatca tgatgaaagc cagcaatgac gttgcagctg 10020

atattgttgt cggccaaccg cgccgaaaac gcagctgtca gacccacagc ctccaacgaa 10080

gaatgtatcg tcaaagtgat ccaagcacac tcatagttgg agtcgtactc caaaggcggc 10140

aatgacgagt cagacagata ctcgtcgacg tttaaaccat catctaaggg cctcaaaact 10200

acctcggaac tgctgcgctg atctggacac cacagaggtt ccgagcactt taggttgcac 10260

caaatgtccc accaggtgca ggcagaaaac gctggaacag cgtgtacagt ttgtcttaac 10320

aaaaagtgag ggcgctgagg tcgagcaggg tggtgtgact tgttatagcc tttagagctg 10380

cgaaagcgcg tatggatttg gctcatcagg ccagattgag ggtctgtgga cacatgtcat 10440

gttagtgtac ttcaatcgcc ccctggatat agccccgaca ataggccgtg gcctcatttt 10500

tttgccttcc gcacatttcc attgctcggt acccacacct tgcttctcct gcacttgcca 10560

accttaatac tggtttacat tgaccaacat cttacaagcg gggggcttgt ctagggtata 10620

tataaacagt ggctctccca atcggttgcc agtctctttt ttcctttctt tccccacaga 10680

ttcgaaatct aaactacaca tcacac 10706

<210> 15

<211> 43

<212> RNA

<213> artificial sequence

<220><223> Hammerhead (HH) ribozyme

<220><221> misc_feature

<222> (1)..(6)

<223> n = A, C, U, or G

<400> 15

nnnnnncuga ugaguccgug aggacgaaac gaguaagcuc guc 43

<210> 16

<211> 68

<212> RNA

<213> hepatitis delta virus

<400> 16

ggccggcaug gucccagccu ccucgcuggc gccggcuggg caacaugcuu cggcauggcg 60

aaugggac 68

<210> 17

<211> 20

<212> DNA

<213> Yarrowia lipolytica

<400> 17

tcaaacgatt acccaccctc 20

<210> 18

<211> 760

<212> DNA

<213> artificial sequence

<220><223> RGR expression cassette

<400> 18

tcgacgttta aaccatcatc taagggcctc aaaactacct cggaactgct gcgctgatct 60

ggacaccaca gaggttccga gcactttagg ttgcaccaaa tgtcccacca ggtgcaggca 120

gaaaacgctg gaacagcgtg tacagtttgt cttaacaaaa agtgagggcg ctgaggtcga 180

gcagggtggt gtgacttgtt atagccttta gagctgcgaa agcgcgtatg gatttggctc 240

atcaggccag attgagggtc tgtggacaca tgtcatgtta gtgtacttca atcgcccccct 300

ggatatagcc ccgacaatag gccgtggcct catttttttg ccttccgcac atttccattg 360

ctcggtaccc acaccttgct tctcctgcac ttgccaacct taatactggt ttacattgac 420

caacatctta caagcggggg gcttgtctag ggtatatata aacagtggct ctcccaatcg 480

gttgccagtc tctttttcc tttctttccc cacagattcg aaatctaaac tacacatcac 540

accatggttt gactgatgag tccgtgagga cgaaacgagt aagctcgtct caaacgatta 600

cccaccctcg ttttagagct agaaatagca agttaaaata aggctagtcc gttatcaact 660

tgaaaaagtg gcaccgagtc ggtggtgctt ttggccggca tggtcccagc ctcctcgctg 720

gcgccggctg ggcaacatgc ttcggcatgg cgaatgggac 760

<210> 19

<211> 6793

<212> DNA

<213> artificial sequence

<220><223> pRF38 plasmid

<400> 19

catggtttga ctgatgagtc cgtgaggacg aaacgagtaa gctcgtctca aacgattacc 60

caccctcgtt ttagagctag aaatagcaag ttaaaataag gctagtccgt tatcaacttg 120

aaaaagtggc accgagtcgg tggtgctttt ggccggcatg gtcccagcct cctcgctggc 180

gccggctggg caacatgctt cggcatggcg aatgggacaa gcttgggggc ggccgcaagt 240

gtggatgggg aagtgagtgc ccggttctgt gtgcacaatt ggcaatccaa gatggatgga 300

ttcaacacag ggatatagcg agctacgtgg tggtgcgagg atatagcaac ggatatttat 360

gtttgacact tgagaatgta cgatacaagc actgtccaag tacaatacta aacatactgt 420

acatactcat actcgtaccc gggcaacggt ttcacttgag tgcagtggct agtgctctta 480

ctcgtacagt gtgcaatact gcgtatcata gtctttgatg tatatcgtat tcattcatgt 540

tagttgcgta cgagccggaa gcataaagtg taaagcctgg ggtgcctaat gagtgagcta 600

actcacatta attgcgttgc gctcactgcc cgctttccag tcgggaaacc tgtcgtgcca 660

gctgcattaa tgaatcggcc aacgcgcggg gagaggcggt ttgcgtattg ggcgctcttc 720

cgcttcctcg ctcactgact cgctgcgctc ggtcgttcgg ctgcggcgag cggtatcagc 780

tcactcaaag gcggtaatac ggttatccac agaatcaggg gataacgcag gaaagaacat 840

gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc tggcgttttt 900

ccataggctc cgccccccctg acgagcatca caaaaatcga cgctcaagtc agaggtggcg 960

aaacccgaca ggactataaa gataccaggc gtttcccccct ggaagctccc tcgtgcgctc 1020

tcctgttccg accctgccgc ttaccggata cctgtccgcc tttctccctt cgggaagcgt 1080

ggcgctttct catagctcac gctgtaggta tctcagttcg gtgtaggtcg ttcgctccaa 1140

gctgggctgt gtgcacgaac cccccgttca gcccgaccgc tgcgccttat ccggtaacta 1200

tcgtcttgag tccaacccgg taagacacga cttatcgcca ctggcagcag ccactggtaa 1260

caggattagc agagcgaggt atgtaggcgg tgctacagag ttcttgaagt ggtggcctaa 1320

ctacggctac actagaagga cagtatttgg tatctgcgct ctgctgaagc cagttacctt 1380

cggaaaaaga gttggtagct cttgatccgg caaacaaacc accgctggta gcggtggttt 1440

ttttgtttgc aagcagcaga ttacgcgcag aaaaaaagga tctcaagaag atcctttgat 1500

cttttctacg gggtctgacg ctcagtggaa cgaaaactca cgttaaggga ttttggtcat 1560

gagattatca aaaaggatct tcacctagat ccttttaaat taaaaatgaa gttttaaatc 1620

aatctaaagt atatatgagt aaacttggtc tgacagttac caatgcttaa tcagtgaggc 1680

acctatctca gcgatctgtc tatttcgttc atccatagtt gcctgactcc ccgtcgtgta 1740

gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga taccgcgaga 1800

cccacgctca ccggctccag atttatcagc aataaaccag ccagccggaa gggccgagcg 1860

cagaagtggt cctgcaactt tatccgcctc catccagtct attaattgtt gccgggaagc 1920

tagagtaagt agttcgccag ttaatagttt gcgcaacgtt gttgccattg ctacaggcat 1980

cgtggtgtca cgctcgtcgt ttggtatggc ttcattcagc tccggttccc aacgatcaag 2040

gcgagttaca tgatccccca tgttgtgcaa aaaagcggtt agctccttcg gtcctccgat 2100

cgttgtcaga agtaagttgg ccgcagtgtt atcactcatg gttatggcag cactgcataa 2160

ttctcttact gtcatgccat ccgtaagatg cttttctgtg actggtgagt actcaaccaa 2220

gtcattctga gaatagtgta tgcggcgacc gagttgctct tgcccggcgt caatacggga 2280

taataccgcg ccacatagca gaactttaaa agtgctcatc attggaaaac gttcttcggg 2340

gcgaaaactc tcaaggatct taccgctgtt gagatccagt tcgatgtaac ccactcgtgc 2400

acccaactga tcttcagcat cttttacttt caccagcgtt tctgggtgag caaaaacagg 2460

aaggcaaaat gccgcaaaaa agggaataag ggcgacacgg aaatgttgaa tactcatact 2520

cttccttttt caatattatt gaagcattta tcagggttat tgtctcatga gcggatacat 2580

atttgaatgt atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt 2640

gccacctgac gcgccctgta gcggcgcatt aagcgcggcg ggtgtggtgg ttacgcgcag 2700

cgtgaccgct acacttgcca gcgccctagc gcccgctcct ttcgctttct tcccttcctt 2760

tctcgccacg ttcgccggct ttccccgtca agctctaaat cgggggctcc ctttagggtt 2820

ccgatttagt gctttacggc acctcgaccc caaaaaactt gattagggtg atggttcacg 2880

tagtgggcca tcgccctgat agacggtttt tcgcccttg acgttggagt ccacgttctt 2940

taatagtgga ctcttgttcc aaactggaac aacactcaac cctatctcgg tctattcttt 3000

tgatttataa gggattttgc cgatttcggc ctattggtta aaaaatgagc tgatttaaca 3060

aaaatttaac gcgaatttta acaaaatatt aacgcttaca atttccattc gccattcagg 3120

ctgcgcaact gttgggaagg gcgatcggtg cgggcctctt cgctattacg ccagctggcg 3180

aaaggggggat gtgctgcaag gcgattaagt tgggtaacgc cagggttttc ccagtcacga 3240

cgttgtaaaa cgacggccag tgaattgtaa tacgactcac tatagggcga attgggtacc 3300

gggccccccc tcgaggtcga tggtgtcgat aagcttgata tcgaattcat gtcacacaaa 3360

ccgatcttcg cctcaaggaa acctaattct acatccgaga gactgccgag atccagtcta 3420

cactgattaa ttttcgggcc aataatttaa aaaaatcgtg ttatataata ttatatgtat 3480

tatatatata catcatgatg atactgacag tcatgtccca ttgctaaata gacagactcc 3540

atctgccgcc tccaactgat gttctcaata tttaaggggt catctcgcat tgtttaataa 3600

taaacagact ccatctaccg cctccaaatg atgttctcaa aatatattgt atgaacttat 3660

ttttattact tagtattatt agacaactta cttgctttat gaaaaacact tcctatttag 3720

gaaacaattt ataatggcag ttcgttcatt taacaattta tgtagaataa atgttataaa 3780

tgcgtatggg aaatcttaaa tatggatagc ataaatgata tctgcattgc ctaattcgaa 3840

atcaacagca acgaaaaaaa tcccttgtac aacataaata gtcatcgaga aatatcaact 3900

atcaaagaac agctattcac acgttactat tgagattatt attggacgag aatcacacac 3960

tcaactgtct ttctctcttc tagaaataca ggtacaagta tgtactattc tcattgttca 4020

tacttctagt catttcatcc cacatattcc ttggatttct ctccaatgaa tgacattcta 4080

tcttgcaaat tcaacaatta taataagata taccaaagta gcggtatagt ggcaatcaaa 4140

aagcttctct ggtgtgcttc tcgtatttat ttttattcta atgatccatt aaaggtatat 4200

atttatttct tgttatataa tccttttgtt tattacatgg gctggataca taaaggtatt 4260

ttgatttaat tttttgctta aattcaatcc cccctcgttc agtgtcaact gtaatggtag 4320

gaaattacca tacttttgaa gaagcaaaaa aaatgaaaga aaaaaaaaat cgtatttcca 4380

ggttagacgt tccgcagaat ctagaatgcg gtatgcggta cattgttctt cgaacgtaaa 4440

agttgcgctc cctgagatat tgtacatttt tgcttttaca agtacaagta catcgtacaa 4500

ctatgtacta ctgttgatgc atccacaaca gtttgttttg ttttttttg ttttttttt 4560

ttctaatgat tcattaccgc tatgtatacc tacttgtact tgtagtaagc cgggttattg 4620

gcgttcaatt aatcatagac ttatgaatct gcacggtgtg cgctgcgagt tacttttagc 4680

ttatgcatgc tacttgggtg taatattggg atctgttcgg aaatcaacgg atgctcaatc 4740

gatttcgaca gtaattaatt aagtcataca caagtcagct ttcttcgagc ctcatataag 4800

tataagtagt tcaacgtatt agcactgtac ccagcatctc cgtatcgaga aacacaacaa 4860

catgccccat tggacagatc atgcggatac acaggttgtg cagtatcata catactcgat 4920

cagacaggtc gtctgaccat catacaagct gaacaagcgc tccatacttg cacgctctct 4980

atatacacag ttaaattaca tatccatagt ctaacctcta acagttaatc ttctggtaag 5040
cctcccagcc agccttctgg tatcgcttgg cctcctcaat aggatctcgg ttctggccgt 5100
acagacctcg gccgacaatt atgatatccg ttccggtaga catgacatcc tcaacagttc 5160
ggtactgctg tccgagagcg tctcccttgt cgtcaagacc caccccgggg gtcagaataa 5220
gccagtcctc agagtcgccc ttaggtcggt tctgggcaat gaagccaacc acaaactcgg 5280
ggtcggatcg ggcaagctca atggtctgct tggagtactc gccagtggcc agagagccct 5340
tgcaagacag ctcggccagc atgagcagac ctctggccag cttctcgttg ggagagggga 5400

ctaggaactc cttgtactgg gagttctcgt agtcagagac gtcctccttc ttctgttcag 5460
agacagtttc ctcggcacca gctcgcaggc cagcaatgat tccggttccg ggtacaccgt 5520
gggcgttggt gatatcggac cactcggcga ttcggtgaca ccggtactgg tgcttgacag 5580
tgttgccaat atctgcgaac tttctgtcct cgaacaggaa gaaaccgtgc ttaagagcaa 5640
gttccttgag ggggagcaca gtgccggcgt aggtgaagtc gtcaatgatg tcgatatggg 5700
ttttgatcat gcacacataa ggtccgacct tatcggcaag ctcaatgagc tccttggtgg 5760
tggtaacatc cagagaagca cacaggttgg ttttcttggc tgccacgagc ttgagcactc 5820

gagcggcaaa ggcggacttg tggacgttag ctcgagcttc gtaggagggc attttggtgg 5880
tgaagaggag actgaaataa atttagtctg cagaactttt tatcggaacc ttatctgggg 5940
cagtgaagta tatgttatgg taatagttac gagttagttg aacttataga tagactggac 6000
tatacggcta tcggtccaaa ttagaaagaa cgtcaatggc tctctgggcg tcgcctttgc 6060
cgacaaaaat gtgatcatga tgaaagccag caatgacgtt gcagctgata ttgttgtcgg 6120
ccaaccgcgc cgaaaacgca gctgtcagac ccacagcctc caacgaagaa tgtatcgtca 6180
aagtgatcca agcacactca tagttggagt cgtactccaa aggcggcaat gacgagtcag 6240

acagatactc gtcgacgttt aaaccatcat ctaagggcct caaaactacc tcggaactgc 6300
tgcgctgatc tggacaccac agaggttccg agcactttag gttgcaccaa atgtcccacc 6360
aggtgcaggc agaaaacgct ggaacagcgt gtacagtttg tcttaacaaa aagtgagggc 6420
gctgaggtcg agcagggtgg tgtgacttgt tatagccttt agagctgcga aagcgcgtat 6480
ggatttggct catcaggcca gattgagggt ctgtggacac atgtcatgtt agtgtacttc 6540
aatcgccccc tggatatagc cccgacaata ggccgtggcc tcatttttt gccttccgca 6600
catttccatt gctcggtacc cacaccttgc ttctcctgca cttgccaacc ttaatactgg 6660

tttacattga ccaacatctt acaagcgggg ggcttgtcta gggtatatat aaacagtggc 6720

tctcccaatc ggttgccagt ctctttttc ctttctttcc ccacagattc gaaatctaaa 6780

ctacacatca cac 6793

<210> 20

<211> 19

<212> DNA

<213> artificial sequence

<220><223> RGR forward PCR primer

<400> 20

cgagtcagac agatactcg 19

<210> 21

<211> 19

<212> DNA

<213> artificial sequence

<220><223> RGR reverse PCR primer

<400> 21

ccctgtgttg aatccatcc 19

<210> 22

<211> 20

<212> DNA

<213> Yarrowia lipolytica

<400> 22

ggaaggcaca tatggcaagg 20

<210> 23

<211> 21

<212> DNA

<213> Yarrowia lipolytica

<400> 23

gtaagagtgg tttgctccag g 21

<210> 24

<211> 21

<212> DNA

<213> Yarrowia lipolytica

<400> 24

gcacaggtat ttctgccctt c 21

<210> 25

<211> 709

<212> DNA

<213> artificial sequence

<220><223> RG expression cassette

<400> 25

tcgacgttta aaccatcatc taagggcctc aaaactacct cggaactgct gcgctgatct 60

ggacaccaca gaggttccga gcactttagg ttgcaccaaa tgtcccacca ggtgcaggca 120

gaaaacgctg gaacagcgtg tacagtttgt cttaacaaaa agtgagggcg ctgaggtcga 180

gcagggtggt gtgacttgtt atagccttta gagctgcgaa agcgcgtatg gatttggctc 240

atcaggccag attgagggtc tgtggacaca tgtcatgtta gtgtacttca atcgcccccc 300

ggatatagcc ccgacaatag gccgtggcct catttttttg ccttccgcac atttccattg 360

ctcggtaccc acaccttgct tctcctgcac ttgccaacct taatactggt ttacattgac 420

caacatctta caagcggggg gcttgtctag ggtatatata aacagtggct ctcccaatcg 480

gttgccagtc tctttttttcc tttctttccc cacagattcg aaatctaaac tacacatcac 540

accatggttt gactgatgag tccgtgagga cgaaacgagt aagctcgtct caaacgatta 600

cccaccctcg ttttagagct agaaatagca agttaaaata aggctagtcc gttatcaact 660

tgaaaaagtg gcaccgagtc ggtgcttttt ttttttgttt tttatgtct 709

<210> 26

<211> 10

<212> DNA

<213> unknown

<220><223> poly-A

<400> 26

aaaaaaaaaa 10

<210> 27

<211> 10

<212> DNA

<213> unknown

<220><223> poly-T

<400> 27

tttttttttt 10

<210> 28

<211> 100

<212> DNA

<213> artificial sequence

<220><223> CAN1 frameshift donor

<400> 28

ccttaacgac cctgccgtct ccatccatcc gaccacaatg gaaaagacat gactgaggcc 60

cacatccaca tcaaccacac ggcccactcg gatgactcag 100

<210> 29

<211> 100

<212> DNA

<213> artificial sequence

<220><223> CAN1 frameshift donor complement

<400> 29

ctgagtcatc cgagtgggcc gtgtggttga tgtggatgtg ggcctcagtc atgtcttttc 60

cattgtggtc ggatggatgg agacggcagg gtcgttaagg 100

<210> 30

<211> 106

<212> DNA

<213> artificial sequence

<220><223> CAN1 point mutation donor

<400> 30

ccatccatcc gaccacaatg gaaaagacat tttcaaacga ttacccaccc tgatgaactg 60

aggcccacat ccacatcaac cacacggccc actcggatga ctcaga 106

<210> 31

<211> 106

<212> DNA

<213> artificial sequence

<220><223> CAN1 point mutation donor complement

<400> 31

tctgagtcat ccgagtgggc cgtgtggttg atgtggatgt gggcctcagt tcatcagggt 60

gggtaatcgt ttgaaaatgt cttttccatt gtggtcggat ggatgg 106

<210> 32

<211> 655

<212> DNA

<213> artificial sequence

<220><223> CAN1 upstream donor arm

<400> 32

gggaagcctt gctacgttag gagaagacgc acggcgatga tacgggtacc cctcatgaca 60

tcaatatccg ctgcccctct tgccagcaag gcgtcagcag gtgctttttt cgctattttc 120

accagaccac agccttttc cttgtgtctc atcttggatt ccttcaaagg caactcaccg 180

cacctccgag tcgtgtgaac aatgtaataa taggctattg actttttcc cacctgttta 240

gcgccaaacc caaagcgctt ttcgccccca ctgcagcccg atggaaggca catatggcaa 300

gggaaaagtc ttcaggtaat acatgcctgc tgcaactata tgtactctga ctcattccct 360

cagacgtggg tcatagacag ctgttttaaa ccgggcaaat caatctctgt cgcacaggta 420

tttctgccct tcaaaaccag gttgccacat cagattccat caaagttttt cagactaact 480

tcaatcttaa acggcatctc acaacaagcg aattggacgg aaaaaaagcg tctatcatta 540

ccggcaccta tccacactaa gacagtacta aaggacgacg ctccccacga aacgacgttt 600

cgaccttaac gaccctgccg tctccatcca tccgaccact cccgacgctc tctcc 655

<210> 33

<211> 29

<212> DNA

<213> artificial sequence

<220><223> Forward PCR primer for amplifying CAN1 upstream donor arm

<400> 33

gggaagcttg ctacgttagg agaagacgc 29

<210> 34

<211> 37

<212> DNA

<213> artificial sequence

<220><223> Reverse PCR primer for amplifying CAN1 upstream donor arm

<400> 34

ggagagagcg tcgggagtgg tcggatggat ggagacg 37

<210> 35

<211> 658

<212> DNA

<213> artificial sequence

<220><223> CAN1 downstream donor arm

<400> 35

cgtctccatc catccgacca ctcccgacgc tctctcctgg agcaaaccac tcttaccaag 60

catatagcat atataataac gtattgaatt tattaactga ttgaattgag agtaaagcca 120

gtagcgttgt acggctgtag ctttttagaa aagtggcaga tgagcgatgg tggatatgaa 180

agtaccttta cggcatgtag cgacacaaga tcgcttccaa gaactcgaca ttcaagccca 240

gctcgtacaa gaaaatgaac tagccaatca tatgaactag cacattgaag tcaccgcatc 300

atctctgttg gaaacgacgc gcatgtactc gtgcgtagta aatccgtatc tgtacactcg 360

aaagattaca gtatgtagta gtagcatgac taacgatgta acgtccaaat aacgctctgt 420

gcctactcct gtagatgcat tagaccacct gctaacgtct acacgttatg tccgttagct 480

ccaagattgc acttttccct caaagactct gctgggttac gtcatggtct ctttcgggtc 540

tctggtccgt tctctgcccg cccatatccg cccaggctgc tacgatacag gataagctca 600

taagcttaga ttatttttcc ggaatgacat cacgatgcag tggtggaagg atgtatgg 658

<210> 36

<211> 37

<212> DNA

<213> artificial sequence

<220><223> Forward PCR primer for amplifying CAN1 downstream donor arm

<400> 36

cgtctccatc catccgacca ctcccgacgc tctctcc 37

<210> 37

<211> 22

<212> DNA

<213> artificial sequence

<220><223> Reverse PCR primer for amplifying CAN1 downstream donor arm

<400> 37

ccatacatcc ttccaccact gc 22

<210> 38

<211> 1276

<212> DNA

<213> artificial sequence

<220><223> CAN1 large deletion donor

<400> 38

gggaagcctt gctacgttag gagaagacgc acggcgatga tacgggtacc cctcatgaca 60

tcaatatccg ctgcccctct tgccagcaag gcgtcagcag gtgctttttt cgctattttc 120

accagaccac agccttttc cttgtgtctc atcttggatt ccttcaaagg caactcaccg 180

cacctccgag tcgtgtgaac aatgtaataa taggctattg actttttcc cacctgttta 240

gcgccaaacc caaagcgctt ttcgccccca ctgcagcccg atggaaggca catatggcaa 300

gggaaaagtc ttcaggtaat acatgcctgc tgcaactata tgtactctga ctcattccct 360

cagacgtggg tcatagacag ctgttttaaa ccgggcaaat caatctctgt cgcacaggta 420

tttctgccct tcaaaaccag gttgccacat cagattccat caaagttttt cagactaact 480

tcaatcttaa acggcatctc acaacaagcg aattggacgg aaaaaaagcg tctatcatta 540

ccggcaccta tccacactaa gacagtacta aaggacgacg ctccccacga aacgacgttt 600

cgaccttaac gaccctgccg tctccatcca tccgaccact cccgacgctc tctcctggag 660

caaaccactc ttaccaagca tatagcatat ataataacgt attgaattta ttaactgatt 720

gaattgagag taaagccagt agcgttgtac ggctgtagct ttttagaaaa gtggcagatg 780

agcgatggtg gatatgaaag tacctttacg gcatgtagcg acacaagatc gcttccaaga 840

actcgacatt caagcccagc tcgtacaaga aaatgaacta gccaatcata tgaactagca 900

cattgaagtc accgcatcat ctctgttgga aacgacgcgc atgtactcgt gcgtagtaaa 960

tccgtatctg tacactcgaa agattacagt atgtagtagt agcatgacta acgatgtaac 1020

gtccaaataa cgctctgtgc ctactcctgt agatgcatta gaccacctgc taacgtctac 1080

acgttatgtc cgttagctcc aagattgcac ttttccctca aagactctgc tgggttacgt 1140

catggtctct ttcgggtctc tggtccgttc tctgcccgcc catatccgcc caggctgcta 1200

cgatacagga taagctcata agcttagatt atttttccgg aatgacatca cgatgcagtg 1260

gtggaaggat gtatgg 1276

<210> 39

<211> 31

<212> DNA

<213> artificial sequence

<220><223> RG/RGR forward PCR primer

<400> 39

ggggttaatt aacgagtcag acagatactc g 31

<210> 40

<211> 29

<212> DNA

<213> artificial sequence

<220><223> RG/RGR reverse PCR primer

<400> 40

ggggatcgat ccctgtgttg aatccatcc 29

<210> 41

<211> 11568

<212> DNA

<213> artificial sequence

<220><223> pRF84 plasmid

<400> 41

cgatccctgt gttgaatcca tccatcttgg attgccaatt gtgcacacag aaccgggcac 60

tcacttcccc atccacactt gcggccgccc ccaagcttgt cccattcgcc atgccgaagc 120

atgttgccca gccggcgcca gcgaggaggc tgggaccatg ccggccaaaa gcaccaccga 180

ctcggtgcca ctttttcaag ttgataacgg actagcctta ttttaacttg ctatttctag 240

ctctaaaacg agggtgggta atcgtttgag acgagcttac tcgtttcgtc ctcacggact 300

catcagtcaa accatggtgt gatgtgtagt ttagatttcg aatctgtggg gaaagaaagg 360

aaaaaagaga ctggcaaccg attgggagag ccactgttta tatataccct agacaagccc 420

cccgcttgta agatgttggt caatgtaaac cagtattaag gttggcaagt gcaggagaag 480

caaggtgtgg gtaccgagca atggaaatgt gcggaaggca aaaaaatgag gccacggcct 540

attgtcgggg ctatatccag ggggcgattg aagtacacta acatgacatg tgtccacaga 600

ccctcaatct ggcctgatga gccaaatcca tacgcgcttt cgcagctcta aaggctataa 660

caagtcacac caccctgctc gacctcagcg ccctcacttt ttgttaagac aaactgtaca 720

cgctgttcca gcgttttctg cctgcacctg gtgggacatt tggtgcaacc taaagtgctc 780

ggaacctctg tggtgtccag atcagcgcag cagttccgag gtagttttga ggcccttaga 840

tgatggttta aacgtcgacg agtatctgtc tgactcgtta attaagtcat acacaagtca 900

gctttcttcg agcctcatat aagtataagt agttcaacgt attagcactg tacccagcat 960

ctccgtatcg agaaacacaa caacatgccc cattggacag atcatgcgga tacacaggtt 1020

gtgcagtatc atacatactc gatcagacag gtcgtctgac catcatacaa gctgaacaag 1080

cgctccatac ttgcacgctc tctatataca cagttaaatt acatatccat agtctaacct 1140

ctaacagtta atcttctggt aagcctccca gccagccttc tggtatcgct tggcctcctc 1200

aataggatct cggttctggc cgtacagacc tcggccgaca attatgatat ccgttccggt 1260

agacatgaca tcctcaacag ttcggtactg ctgtccgaga gcgtctccct tgtcgtcaag 1320

acccaccccg ggggtcagaa taagccagtc ctcagagtcg cccttaggtc ggttctgggc 1380

aatgaagcca accacaaact cggggtcgga tcgggcaagc tcaatggtct gcttggagta 1440

ctcgccagtg gccagagagc ccttgcaaga cagctcggcc agcatgagca gacctctggc 1500

cagcttctcg ttgggagagg ggactaggaa ctccttgtac tgggagttct cgtagtcaga 1560

gacgtcctcc ttcttctgtt cagagacagt ttcctcggca ccagctcgca ggccagcaat 1620

gattccggtt ccgggtacac cgtgggcgtt ggtgatatcg gaccactcgg cgattcggtg 1680

acaccggtac tggtgcttga cagtgttgcc aatatctgcg aactttctgt cctcgaacag 1740

gaagaaaccg tgcttaagag caagttcctt gagggggagc acagtgccgg cgtaggtgaa 1800

gtcgtcaatg atgtcgatat gggttttgat catgcacaca taaggtccga ccttatcggc 1860

aagctcaatg agctccttgg tggtggtaac atccagagaa gcacacaggt tggttttctt 1920

ggctgccacg agcttgagca ctcgagcggc aaaggcggac ttgtggacgt tagctcgagc 1980

ttcgtaggag ggcattttgg tggtgaagag gagactgaaa taaatttagt ctgcagaact 2040

ttttatcgga accttatctg gggcagtgaa gtatatgtta tggtaatagt tacgagttag 2100

ttgaacttat agatagactg gactatacgg ctatcggtcc aaattagaaa gaacgtcaat 2160

ggctctctgg gcgtcgcctt tgccgacaaa aatgtgatca tgatgaaagc cagcaatgac 2220

gttgcagctg atattgttgt cggccaaccg cgccgaaaac gcagctgtca gacccacagc 2280

ctccaacgaa gaatgtatcg tcaaagtgat ccaagcacac tcatagttgg agtcgtactc 2340

caaaggcggc aatgacgagt cagacagata ctcgtcgacg tttaaaccat catctaaggg 2400

cctcaaaact acctcggaac tgctgcgctg atctggacac cacagaggtt ccgagcactt 2460

taggttgcac caaatgtccc accaggtgca ggcagaaaac gctggaacag cgtgtacagt 2520

ttgtcttaac aaaaagtgag ggcgctgagg tcgagcaggg tggtgtgact tgttatagcc 2580

tttagagctg cgaaagcgcg tatggatttg gctcatcagg ccagattgag ggtctgtgga 2640

cacatgtcat gttagtgtac ttcaatcgcc ccctggatat agccccgaca ataggccgtg 2700

gcctcatttt tttgccttcc gcacatttcc attgctcggt acccacacct tgcttctcct 2760

gcacttgcca accttaatac tggtttacat tgaccaacat cttacaagcg gggggcttgt 2820

ctagggtata tataaacagt ggctctccca atcggttgcc agtctctttt ttcctttctt 2880

tccccacaga ttcgaaatct aaactacaca tcacaccatg gacaagaaat actccatcgg 2940
cctggacatt ggaaccaact ctgtcggctg ggctgtcatc accgacgagt acaaggtgcc 3000
ctccaagaaa ttcaaggtcc tcggaaacac cgatcgacac tccatcaaga aaaacctcat 3060
tggtgccctg ttgttcgatt ctggcgagac tgccgaagct accagactca agcgaactgc 3120
tcggcgacgt tacacccgac ggaagaaccg aatctgctac ctgcaggaga tcttttccaa 3180
cgagatggcc aaggtggacg attcgttctt tcatcgactg gaggaatcct tcctcgtcga 3240
ggaagacaag aaacacgagc gtcatcccat ctttggcaac attgtggacg aggttgctta 3300

ccacgagaag tatcctacca tctaccacct gcgaaagaaa ctcgtcgatt ccaccgacaa 3360
ggcggatctc agacttatct acctcgctct ggcacacatg atcaagtttc gaggtcattt 3420
cctcatcgag ggcgatctca atcccgacaa cagcgatgtg gacaagctgt tcattcagct 3480
cgttcagacc tacaaccagc tgttcgagga aaaccccatc aatgcctccg gagtcgatgc 3540
aaaggccatc ttgtctgctc gactctcgaa gagcagacga ctggagaacc tcattgccca 3600
acttcctggc gagaaaaaga acggactgtt tggcaacctc attgcccttt ctcttggtct 3660
cacacccaac ttcaagtcca acttcgatct ggcggaggac gccaagctcc agctgtccaa 3720

ggacacctac gacgatgacc tcgacaacct gcttgcacag attggcgatc agtacgccga 3780
cctgtttctc gctgccaaga acctttcgga tgctattctc ttgtctgaca ttctgcgagt 3840
caacaccgag atcacaaagg ctccccctttc tgcctccatg atcaagcgat acgacgagca 3900
ccatcaggat ctcacactgc tcaaggctct tgtccgacag caactgcccg agaagtacaa 3960
ggagatcttt ttcgatcagt cgaagaacgg ctacgctgga tacatcgacg gcggagcctc 4020
tcaggaagag ttctacaagt tcatcaagcc aattctcgag aagatggacg gaaccgagga 4080
actgcttgtc aagctcaatc gagaggatct gcttcggaag caacgaacct tcgacaacgg 4140

cagcattcct catcagatcc acctcggtga gctgcacgcc attcttcgac gtcaggaaga 4200
cttctacccc tttctcaagg acaaccgaga gaagatcgag aagattctta cctttcgaat 4260
cccctactat gttggtcctc ttgccagagg aaactctcga tttgcttgga tgactcgaaa 4320
gtccgaggaa accatcactc cctggaactt cgaggaagtc gtggacaagg gtgcctctgc 4380
acagtccttc atcgagcgaa tgaccaactt cgacaagaat ctgcccaacg agaaggttct 4440
tcccaagcat tcgctgctct acgagtactt tacagtctac aacgaactca ccaaagtcaa 4500
gtacgttacc gagggaatgc gaaagcctgc cttcttgtct ggcgaacaga agaaagccat 4560

tgtcgatctc ctgttcaaga ccaaccgaaa ggtcactgtt aagcagctca aggaggacta 4620

cttcaagaaa atcgagtgtt tcgacagcgt cgagatttcc ggagttgagg accgattcaa 4680

cgcctctttg ggcacctatc acgatctgct caagattatc aaggacaagg attttctcga 4740

caacgaggaa aacgaggaca ttctggagga catcgtgctc actcttaccc tgttcgaaga 4800

tcgggagatg atcgaggaac gactcaagac atacgctcac ctgttcgacg acaaggtcat 4860

gaaacaactc aagcgacgta gatacaccgg ctggggaaga ctttcgcgaa agctcatcaa 4920

cggcatcaga gacaagcagt ccggaaagac cattctggac tttctcaagt ccgatggctt 4980

tgccaaccga aacttcatgc agctcattca cgacgattct cttaccttca aggaggacat 5040

ccagaaggca caagtgtccg gtcagggcga cagcttgcac gaacatattg ccaacctggc 5100

tggttcgcca gccatcaaga aaggcattct ccagactgtc aaggttgtcg acgagctggt 5160

gaaggtcatg ggacgtcaca agcccgagaa cattgtgatc gagatggcca gagagaacca 5220

gacaactcaa aagggtcaga aaaactcgcg agagcggatg aagcgaatcg aggaaggcat 5280

caaggagctg ggatcccaga ttctcaagga gcatcccgtc gagaacactc aactgcagaa 5340

cgagaagctg tatctctact atctgcagaa tggtcgagac atgtacgtgg atcaggaact 5400

ggacatcaat cgtctcagcg actacgatgt ggaccacatt gtccctcaat cctttctcaa 5460

ggacgattct atcgacaaca aggtccttac acgatccgac aagaacagag gcaagtcgga 5520

caacgttccc agcgaagagg tggtcaaaaa gatgaagaac tactggcgac agctgctcaa 5580

cgccaagctc attacccagc gaaagttcga caatcttacc aaggccgagc gaggcggtct 5640

gtccgagctc gacaaggctg gcttcatcaa gcgtcaactc gtcgagacca gacagatcac 5700

aaagcacgtc gcacagattc tcgattctcg gatgaacacc aagtacgacg agaacgacaa 5760

gctcatccga gaggtcaagg tgattactct caagtccaaa ctggtctccg atttccgaaa 5820

ggactttcag ttctacaagg tgcgagagat caacaattac caccatgccc acgatgctta 5880

cctcaacgcc gtcgttggca ctgcgctcat caagaaatac cccaagctcg aaagcgagtt 5940

cgtttacggc gattacaagg tctacgacgt tcgaaagatg attgccaagt ccgaacagga 6000

gattggcaag gctactgcca agtacttctt ttactccaac atcatgaact ttttcaagac 6060

cgagatcacc ttggccaacg gagagattcg aaagagacca cttatcgaga ccaacggcga 6120

aactggagag atcgtgtggg acaagggtcg agactttgca accgtgcgaa aggttctgtc 6180

gatgcctcag gtcaacatcg tcaagaaaac cgaggttcag actggcggat tctccaagga 6240

gtcgattctg cccaagcgaa actccgacaa gctcatcgct cgaaagaaag actgggatcc 6300

caagaaatac ggtggcttcg attctcctac cgtcgcctat tccgtgcttg tcgttgcgaa 6360

ggtcgagaag ggcaagtcca aaaagctcaa gtccgtcaag gagctgctcg gaattaccat 6420

catggagcga tcgagcttcg agaagaatcc catcgacttc ttggaagcca agggttacaa 6480

ggaggtcaag aaagacctca ttatcaagct gcccaagtac tctctgttcg aactggagaa 6540

cggtcgaaag cgtatgctcg cctccgctgg cgagctgcag aagggaaacg agcttgcctt 6600

gccttcgaag tacgtcaact ttctctatct ggcttctcac tacgagaagc tcaagggttc 6660

tcccgaggac aacgaacaga agcaactctt cgttgagcag cacaaacatt acctcgacga 6720

gattatcgag cagatttccg agttttcgaa gcgagtcatc ctggctgatg ccaacttgga 6780

caaggtgctc tctgcctaca acaagcatcg ggacaaaccc attcgagaac aggcggagaa 6840

catcattcac ctgtttactc ttaccaacct gggtgctcct gcagctttca agtacttcga 6900

taccactatc gaccgaaagc ggtacacatc caccaaggag gttctcgatg ccaccctgat 6960

tcaccagtcc atcactggcc tgtacgagac ccgaatcgac ctgtctcagc ttggtggcga 7020

ctccagagcc gatcccaaga aaaagcgaaa ggtctaagcg gccgcaagtg tggatgggga 7080

agtgagtgcc cggttctgtg tgcacaattg gcaatccaag atggatggat tcaacacagg 7140

gatatagcga gctacgtggt ggtgcgagga tatagcaacg gatatttatg tttgacactt 7200

gagaatgtac gatacaagca ctgtccaagt acaatactaa acatactgta catactcata 7260

ctcgtacccg ggcaacggtt tcacttgagt gcagtggcta gtgctcttac tcgtacagtg 7320

tgcaatactg cgtatcatag tctttgatgt atatcgtatt cattcatgtt agttgcgtac 7380

gagccggaag cataaagtgt aaagcctggg gtgcctaatg agtgagctaa ctcacattaa 7440

ttgcgttgcg ctcactgccc gctttccagt cgggaaacct gtcgtgccag ctgcattaat 7500

gaatcggcca acgcgcgggg agaggcggtt tgcgtattgg gcgctcttcc gcttcctcgc 7560

tcactgactc gctgcgctcg gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg 7620

cggtaatacg gttatccaca gaatcagggg ataacgcagg aaagaacatg tgagcaaaag 7680

gccagcaaaa ggccaggaac cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc 7740

gccccccctga cgagcatcac aaaaatcgac gctcaagtca gaggtggcga aacccgacag 7800

gactataaag ataccaggcg tttccccctg gaagctccct cgtgcgctct cctgttccga 7860

ccctgccgct taccggatac ctgtccgcct ttctcccttc gggaagcgtg gcgctttctc 7920

atagctcacg ctgtaggtat ctcagttcgg tgtaggtcgt tcgctccaag ctgggctgtg 7980

tgcacgaacc ccccgttcag cccgaccgct gcgccttatc cggtaactat cgtcttgagt 8040

ccaacccggt aagacacgac ttatcgccac tggcagcagc cactggtaac aggattagca 8100

gagcgaggta tgtaggcggt gctacagagt tcttgaagtg gtggcctaac tacggctaca 8160

ctagaaggac agtatttggt atctgcgctc tgctgaagcc agttaccttc ggaaaaagag 8220

ttggtagctc ttgatccggc aaacaaacca ccgctggtag cggtggtttt tttgtttgca 8280

agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg 8340

ggtctgacgc tcagtggaac gaaaactcac gttaagggat tttggtcatg agattatcaa 8400
aaaggatctt cacctagatc cttttaaatt aaaaatgaag ttttaaatca atctaaagta 8460
tatatgagta aacttggtct gacagttacc aatgcttaat cagtgaggca cctatctcag 8520
cgatctgtct atttcgttca tccatagttg cctgactccc cgtcgtgtag ataactacga 8580
tacgggaggg cttaccatct ggccccagtg ctgcaatgat accgcgagac ccacgctcac 8640
cggctccaga tttatcagca ataaaccagc cagccggaag ggccgagcgc agaagtggtc 8700
ctgcaacttt atccgcctcc atccagtcta ttaattgttg ccgggaagct agagtaagta 8760

gttcgccagt taatagtttg cgcaacgttg ttgccattgc tacaggcatc gtggtgtcac 8820
gctcgtcgtt tggtatggct tcattcagct ccggttccca acgatcaagg cgagttacat 8880
gatcccccat gttgtgcaaa aaagcggtta gctccttcgg tcctccgatc gttgtcagaa 8940
gtaagttggc cgcagtgtta tcactcatgg ttatggcagc actgcataat tctcttactg 9000
tcatgccatc cgtaagatgc ttttctgtga ctggtgagta ctcaaccaag tcattctgag 9060
aatagtgtat gcggcgaccg agttgctctt gcccggcgtc aatacgggat aataccgcgc 9120
cacatagcag aactttaaaa gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct 9180

caaggatctt accgctgttg agatccagtt cgatgtaacc cactcgtgca cccaactgat 9240
cttcagcatc ttttactttc accagcgttt ctgggtgagc aaaaacagga aggcaaaatg 9300
ccgcaaaaaa gggaataagg gcgacacgga aatgttgaat actcatactc ttccttttc 9360
aatattattg aagcatttat cagggttatt gtctcatgag cggatacata tttgaatgta 9420
tttagaaaaa taaacaaata ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg 9480
cgccctgtag cggcgcatta agcgcggcgg gtgtggtggt tacgcgcagc gtgaccgcta 9540
cacttgccag cgccctagcg cccgctcctt tcgctttctt cccttccttt ctcgccacgt 9600

tcgccggctt tccccgtcaa gctctaaatc gggggctccc tttagggttc cgatttagtg 9660
ctttacggca cctcgacccc aaaaaacttg attagggtga tggttcacgt agtgggccat 9720
cgccctgata gacggttttt cgcccctttga cgttggagtc cacgttcttt aatagtggac 9780
tcttgttcca aactggaaca acactcaacc ctatctcggt ctattctttt gatttataag 9840
ggattttgcc gatttcggcc tattggttaa aaaatgagct gatttaacaa aaatttaacg 9900
cgaattttaa caaaatatta acgcttacaa tttccattcg ccattcaggc tgcgcaactg 9960
ttgggaaggg cgatcggtgc gggcctcttc gctattacgc cagctggcga aaggggatg 10020

tgctgcaagg cgattaagtt gggtaacgcc agggttttcc cagtcacgac gttgtaaaac 10080

gacggccagt gaattgtaat acgactcact atagggcgaa ttgggtaccg ggcccccccct 10140

cgaggtcgat ggtgtcgata agcttgatat cgaattcatg tcacacaaac cgatcttcgc 10200

ctcaaggaaa cctaattcta catccgagag actgccgaga tccagtctac actgattaat 10260

tttcgggcca ataatttaaa aaaatcgtgt tatataatat tatatgtatt atatatatac 10320

atcatgatga tactgacagt catgtcccat tgctaaatag acagactcca tctgccgcct 10380

ccaactgatg ttctcaatat ttaaggggtc atctcgcatt gtttaataat aaacagactc 10440

catctaccgc ctccaaatga tgttctcaaa atatattgta tgaacttatt tttattactt 10500

agtattatta gacaacttac ttgctttatg aaaaacactt cctatttagg aaacaattta 10560

taatggcagt tcgttcattt aacaatttat gtagaataaa tgttataaat gcgtatggga 10620

aatcttaaat atggatagca taaatgatat ctgcattgcc taattcgaaa tcaacagcaa 10680

cgaaaaaaat cccttgtaca acataaatag tcatcgagaa atatcaacta tcaaagaaca 10740

gctattcaca cgttactatt gagattatta ttggacgaga atcacacact caactgtctt 10800

tctctcttct agaaatacag gtacaagtat gtactattct cattgttcat acttctagtc 10860

atttcatccc acatattcct tggatttctc tccaatgaat gacattctat cttgcaaatt 10920

caacaattat aataagatat accaaagtag cggtatagtg gcaatcaaaa agcttctctg 10980

gtgtgcttct cgtatttatt tttattctaa tgatccatta aaggtatata tttatttctt 11040

gttatataat ccttttgttt attacatggg ctggatacat aaaggtattt tgatttaatt 11100

ttttgcttaa attcaatccc ccctcgttca gtgtcaactg taatggtagg aaattaccat 11160

acttttgaag aagcaaaaaa aatgaaagaa aaaaaaaatc gtatttccag gttagacgtt 11220

ccgcagaatc tagaatgcgg tatgcggtac attgttcttc gaacgtaaaa gttgcgctcc 11280

ctgagatatt gtacattttt gcttttacaa gtacaagtac atcgtacaac tatgtactac 11340

tgttgatgca tccacaacag tttgttttgt tttttttttgt ttttttttttt tctaatgatt 11400

cattaccgct atgtatacct acttgtactt gtagtaagcc gggttattgg cgttcaatta 11460

atcatagact tatgaatctg cacggtgtgc gctgcgagtt acttttagct tatgcatgct 11520

acttgggtgt aatattggga tctgttcgga aatcaacgga tgctcaat 11568

<210> 42

<211> 11507

<212> DNA

<213> artificial sequence

<220><223> pRF85 plasmid

<400> 42

cgatccctgt gttgaatcca tccatcttgg attgccaatt gtgcacacag aaccgggcac 60

tcacttcccc atccacactt gcggccgcag acataaaaaa caaaaaaaaa aagcaccgac 120
tcggtgccac tttttcaagt tgataacgga ctagccttat tttaacttgc tatttctagc 180
tctaaaacga gggtgggtaa tcgtttgaga cgagcttact cgtttcgtcc tcacggactc 240
atcagtcaaa ccatggtgtg atgtgtagtt tagatttcga atctgtgggg aaagaaagga 300
aaaaagagac tggcaaccga ttgggagagc cactgtttat atatacccta gacaagcccc 360
ccgcttgtaa gatgttggtc aatgtaaacc agtattaagg ttggcaagtg caggagaagc 420
aaggtgtggg taccgagcaa tggaaatgtg cggaaggcaa aaaaatgagg ccacggccta 480

ttgtcggggc tatatccagg gggcgattga agtacactaa catgacatgt gtccacagac 540
cctcaatctg gcctgatgag ccaaatccat acgcgctttc gcagctctaa aggctataac 600
aagtcacacc accctgctcg acctcagcgc cctcactttt tgttaagaca aactgtacac 660
gctgttccag cgttttctgc ctgcacctgg tgggacattt ggtgcaacct aaagtgctcg 720
gaacctctgt ggtgtccaga tcagcgcagc agttccgagg tagttttgag gcccttagat 780
gatggtttaa acgtcgacga gtatctgtct gactcgttaa ttaagtcata cacaagtcag 840
ctttcttcga gcctcatata agtataagta gttcaacgta ttagcactgt acccagcatc 900

tccgtatcga gaaacacaac aacatgcccc attggacaga tcatgcggat acacaggttg 960
tgcagtatca tacatactcg atcagacagg tcgtctgacc atcatacaag ctgaacaagc 1020
gctccatact tgcacgctct ctatatacac agttaaatta catatccata gtctaacctc 1080
taacagttaa tcttctggta agcctcccag ccagccttct ggtatcgctt ggcctcctca 1140
ataggatctc ggttctggcc gtacagacct cggccgacaa ttatgatatc cgttccggta 1200
gacatgacat cctcaacagt tcggtactgc tgtccgagag cgtctccctt gtcgtcaaga 1260
cccaccccgg gggtcagaat aagccagtcc tcagagtcgc ccttaggtcg gttctgggca 1320

atgaagccaa ccacaaactc ggggtcggat cgggcaagct caatggtctg cttggagtac 1380
tcgccagtgg ccagagagcc cttgcaagac agctcggcca gcatgagcag acctctggcc 1440
agcttctcgt tgggagaggg gactaggaac tccttgtact gggagttctc gtagtcagag 1500
acgtcctcct tcttctgttc agagacagtt tcctcggcac cagctcgcag gccagcaatg 1560
attccggttc cgggtacacc gtgggcgttg gtgatatcgg accactcggc gattcggtga 1620
caccggtact ggtgcttgac agtgttgcca atatctgcga actttctgtc ctcgaacagg 1680
aagaaaccgt gcttaagagc aagttccttg agggggagca cagtgccggc gtaggtgaag 1740

tcgtcaatga tgtcgatatg ggttttgatc atgcacacat aaggtccgac cttatcggca 1800

agctcaatga gctccttggt ggtggtaaca tccagagaag cacacaggtt ggttttcttg 1860

gctgccacga gcttgagcac tcgagcggca aaggcggact tgtggacgtt agctcgagct 1920

tcgtaggagg gcattttggt ggtgaagagg agactgaaat aaatttagtc tgcagaactt 1980

tttatcggaa ccttatctgg ggcagtgaag tatatgttat ggtaatagtt acgagttagt 2040

tgaacttata gatagactgg actatacggc tatcggtcca aattagaaag aacgtcaatg 2100

gctctctggg cgtcgccttt gccgacaaaa atgtgatcat gatgaaagcc agcaatgacg 2160

ttgcagctga tattgttgtc ggccaaccgc gccgaaaacg cagctgtcag acccacagcc 2220

tccaacgaag aatgtatcgt caaagtgatc caagcacact catagttgga gtcgtactcc 2280

aaaggcggca atgacgagtc agacagatac tcgtcgacgt ttaaaccatc atctaagggc 2340

ctcaaaacta cctcggaact gctgcgctga tctggacacc acagaggttc cgagcacttt 2400

aggttgcacc aaatgtccca ccaggtgcag gcagaaaacg ctggaacagc gtgtacagtt 2460

tgtcttaaca aaaagtgagg gcgctgaggt cgagcagggt ggtgtgactt gttatagcct 2520

ttagagctgc gaaagcgcgt atggatttgg ctcatcaggc cagattgagg gtctgtggac 2580

acatgtcatg ttagtgtact tcaatcgccc cctggatata gccccgacaa taggccgtgg 2640

cctcattttt ttgccttccg cacatttcca ttgctcggta cccacacctt gcttctcctg 2700

cacttgccaa ccttaatact ggtttacatt gaccaacatc ttacaagcgg ggggcttgtc 2760

tagggtatat ataaacagtg gctctcccaa tcggttgcca gtctcttttt tcctttcttt 2820

ccccacagat tcgaaatcta aactacacat cacaccatgg acaagaaata ctccatcggc 2880

ctggacattg gaaccaactc tgtcggctgg gctgtcatca ccgacgagta caaggtgccc 2940

tccaagaaat tcaaggtcct cggaaacacc gatcgacact ccatcaagaa aaacctcatt 3000

ggtgccctgt tgttcgattc tggcgagact gccgaagcta ccagactcaa gcgaactgct 3060

cggcgacgtt acacccgacg gaagaaccga atctgctacc tgcaggagat cttttccaac 3120

gagatggcca aggtggacga ttcgttcttt catcgactgg aggaatcctt cctcgtcgag 3180

gaagacaaga aacacgagcg tcatcccatc tttggcaaca ttgtggacga ggttgcttac 3240

cacgagaagt atcctaccat ctaccacctg cgaaagaaac tcgtcgattc caccgacaag 3300

gcggatctca gacttatcta cctcgctctg gcacacatga tcaagtttcg aggtcatttc 3360

ctcatcgagg gcgatctcaa tcccgacaac agcgatgtgg acaagctgtt cattcagctc 3420

gttcagacct acaaccagct gttcgaggaa aaccccatca atgcctccgg agtcgatgca 3480

aaggccatct tgtctgctcg actctcgaag agcagacgac tggagaacct cattgcccaa 3540

cttcctggcg agaaaaagaa cggactgttt ggcaacctca ttgccctttc tcttggtctc 3600

acacccaact tcaagtccaa cttcgatctg gcggaggacg ccaagctcca gctgtccaag 3660

gacacctacg acgatgacct cgacaacctg cttgcacaga ttggcgatca gtacgccgac 3720

ctgtttctcg ctgccaagaa cctttcggat gctattctct tgtctgacat tctgcgagtc 3780

aacaccgaga tcacaaaggc tcccctttct gcctccatga tcaagcgata cgacgagcac 3840

catcaggatc tcacactgct caaggctctt gtccgacagc aactgcccga gaagtacaag 3900

gagatctttt tcgatcagtc gaagaacggc tacgctggat acatcgacgg cggagcctct 3960

caggaagagt tctacaagtt catcaagcca attctcgaga agatggacgg aaccgaggaa 4020

ctgcttgtca agctcaatcg agaggatctg cttcggaagc aacgaacctt cgacaacggc 4080

agcattcctc atcagatcca cctcggtgag ctgcacgcca ttcttcgacg tcaggaagac 4140

ttctacccct ttctcaagga caaccgagag aagatcgaga agattcttac ctttcgaatc 4200

ccctactatg ttggtcctct tgccagagga aactctcgat ttgcttggat gactcgaaag 4260

tccgaggaaa ccatcactcc ctggaacttc gaggaagtcg tggacaaggg tgcctctgca 4320

cagtccttca tcgagcgaat gaccaacttc gacaagaatc tgcccaacga gaaggttctt 4380

cccaagcatt cgctgctcta cgagtacttt acagtctaca acgaactcac caaagtcaag 4440

tacgttaccg agggaatgcg aaagcctgcc ttcttgtctg gcgaacagaa gaaagccatt 4500

gtcgatctcc tgttcaagac caaccgaaag gtcactgtta agcagctcaa ggaggactac 4560

ttcaagaaaa tcgagtgttt cgacagcgtc gagatttccg gagttgagga ccgattcaac 4620

gcctctttgg gcacctatca cgatctgctc aagattatca aggacaagga ttttctcgac 4680

aacgaggaaa acgaggacat tctggaggac atcgtgctca ctcttaccct gttcgaagat 4740

cgggagatga tcgaggaacg actcaagaca tacgctcacc tgttcgacga caaggtcatg 4800

aaacaactca agcgacgtag atacaccggc tggggaagac tttcgcgaaa gctcatcaac 4860

ggcatcagag acaagcagtc cggaaagacc attctggact ttctcaagtc cgatggcttt 4920

gccaaccgaa acttcatgca gctcattcac gacgattctc ttaccttcaa ggaggacatc 4980

cagaaggcac aagtgtccgg tcagggcgac agcttgcacg aacatattgc caacctggct 5040

ggttcgccag ccatcaagaa aggcattctc cagactgtca aggttgtcga cgagctggtg 5100

aaggtcatgg gacgtcacaa gcccgagaac attgtgatcg agatggccag agagaaccag 5160

acaactcaaa agggtcagaa aaactcgcga gagcggatga agcgaatcga ggaaggcatc 5220

aaggagctgg gatcccagat tctcaaggag catcccgtcg agaacactca actgcagaac 5280

gagaagctgt atctctacta tctgcagaat ggtcgagaca tgtacgtgga tcaggaactg 5340

gacatcaatc gtctcagcga ctacgatgtg gaccacattg tccctcaatc ctttctcaag 5400

gacgattcta tcgacaacaa ggtccttaca cgatccgaca agaacagagg caagtcggac 5460

aacgttccca gcgaagaggt ggtcaaaaag atgaagaact actggcgaca gctgctcaac 5520

gccaagctca ttacccagcg aaagttcgac aatcttacca aggccgagcg aggcggtctg 5580
tccgagctcg acaaggctgg cttcatcaag cgtcaactcg tcgagaccag acagatcaca 5640
aagcacgtcg cacagattct cgattctcgg atgaacacca agtacgacga gaacgacaag 5700
ctcatccgag aggtcaaggt gattactctc aagtccaaac tggtctccga tttccgaaag 5760
gactttcagt tctacaaggt gcgagagatc aacaattacc accatgccca cgatgcttac 5820
ctcaacgccg tcgttggcac tgcgctcatc aagaaatacc ccaagctcga aagcgagttc 5880
gtttacggcg attacaaggt ctacgacgtt cgaaagatga ttgccaagtc cgaacaggag 5940

attggcaagg ctactgccaa gtacttcttt tactccaaca tcatgaactt tttcaagacc 6000
gagatcacct tggccaacgg agagattcga aagagaccac ttatcgagac caacggcgaa 6060
actggagaga tcgtgtggga caagggtcga gactttgcaa ccgtgcgaaa ggttctgtcg 6120
atgcctcagg tcaacatcgt caagaaaacc gaggttcaga ctggcggatt ctccaaggag 6180
tcgattctgc ccaagcgaaa ctccgacaag ctcatcgctc gaaagaaaga ctgggatccc 6240
aagaaatacg gtggcttcga ttctcctacc gtcgcctatt ccgtgcttgt cgttgcgaag 6300
gtcgagaagg gcaagtccaa aaagctcaag tccgtcaagg agctgctcgg aattaccatc 6360

atggagcgat cgagcttcga gaagaatccc atcgacttct tggaagccaa gggttacaag 6420
gaggtcaaga aagacctcat tatcaagctg cccaagtact ctctgttcga actggagaac 6480
ggtcgaaagc gtatgctcgc ctccgctggc gagctgcaga agggaaacga gcttgccttg 6540
ccttcgaagt acgtcaactt tctctatctg gcttctcact acgagaagct caagggttct 6600
cccgaggaca acgaacagaa gcaactcttc gttgagcagc acaaacatta cctcgacgag 6660
attatcgagc agatttccga gttttcgaag cgagtcatcc tggctgatgc caacttggac 6720
aaggtgctct ctgcctacaa caagcatcgg gacaaaccca ttcgagaaca ggcggagaac 6780

atcattcacc tgtttactct taccaacctg ggtgctcctg cagctttcaa gtacttcgat 6840
accactatcg accgaaagcg gtacacatcc accaaggagg ttctcgatgc caccctgatt 6900
caccagtcca tcactggcct gtacgagacc cgaatcgacc tgtctcagct tggtggcgac 6960
tccagagccg atcccaagaa aaagcgaaag gtctaagcgg ccgcaagtgt ggatggggaa 7020
gtgagtgccc ggttctgtgt gcacaattgg caatccaaga tggatggatt caacacaggg 7080
atatagcgag ctacgtggtg gtgcgaggat atagcaacgg atatttatgt ttgacacttg 7140
agaatgtacg atacaagcac tgtccaagta caatactaaa catactgtac atactcatac 7200

tcgtacccgg gcaacggttt cacttgagtg cagtggctag tgctcttact cgtacagtgt 7260

gcaatactgc gtatcatagt ctttgatgta tatcgtattc attcatgtta gttgcgtacg 7320

agccggaagc ataaagtgta aagcctgggg tgcctaatga gtgagctaac tcacattaat 7380

tgcgttgcgc tcactgcccg ctttccagtc gggaaacctg tcgtgccagc tgcattaatg 7440

aatcggccaa cgcgcgggga gaggcggttt gcgtattggg cgctcttccg cttcctcgct 7500

cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc actcaaaggc 7560

ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg 7620

ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg 7680

ccccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg 7740

actataaaga taccaggcgt ttcccccctgg aagctccctc gtgcgctctc ctgttccgac 7800

cctgccgctt accggatacc tgtccgcctt tctcccttcg ggaagcgtgg cgctttctca 7860

tagctcacgc tgtaggtatc tcagttcggt gtaggtcgtt cgctccaagc tgggctgtgt 7920

gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc ggtaactatc gtcttgagtc 7980

caacccggta agacacgact tatcgccact ggcagcagcc actggtaaca ggattagcag 8040

agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact acggctacac 8100

tagaaggaca gtatttggta tctgcgctct gctgaagcca gttaccttcg gaaaaagagt 8160

tggtagctct tgatccggca aacaaaccac cgctggtagc ggtggttttt ttgtttgcaa 8220

gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cctttgatct tttctacggg 8280

gtctgacgct cagtggaacg aaaactcacg ttaagggatt ttggtcatga gattatcaaa 8340

aaggatcttc acctagatcc ttttaaatta aaaatgaagt tttaaatcaa tctaaagtat 8400

atatgagtaa acttggtctg acagttacca atgcttaatc agtgaggcac ctatctcagc 8460

gatctgtcta tttcgttcat ccatagttgc ctgactcccc gtcgtgtaga taactacgat 8520

acgggagggc ttaccatctg gccccagtgc tgcaatgata ccgcgagacc cacgctcacc 8580

ggctccagat ttatcagcaa taaaccagcc agccggaagg gccgagcgca gaagtggtcc 8640

tgcaacttta tccgcctcca tccagtctat taattgttgc cgggaagcta gagtaagtag 8700

ttcgccagtt aatagtttgc gcaacgttgt tgccattgct acaggcatcg tggtgtcacg 8760

ctcgtcgttt ggtatggctt cattcagctc cggttcccaa cgatcaaggc gagttacatg 8820

atcccccatg ttgtgcaaaa aagcggttag ctccttcggt cctccgatcg ttgtcagaag 8880

taagttggcc gcagtgttat cactcatggt tatggcagca ctgcataatt ctcttactgt 8940

catgccatcc gtaagatgct tttctgtgac tggtgagtac tcaaccaagt cattctgaga 9000

atagtgtatg cggcgaccga gttgctcttg cccggcgtca atacgggata ataccgcgcc 9060

acatagcaga actttaaaag tgctcatcat tggaaaacgt tcttcggggc gaaaactctc 9120

aaggatctta ccgctgttga gatccagttc gatgtaaccc actcgtgcac ccaactgatc 9180

ttcagcatct tttactttca ccagcgtttc tgggtgagca aaaacaggaa ggcaaaatgc 9240

cgcaaaaaag ggaataaggg cgacacggaa atgttgaata ctcatactct tccttttca 9300

atattattga agcatttatc agggttattg tctcatgagc ggatacatat ttgaatgtat 9360

ttagaaaaat aaacaaatag gggttccgcg cacatttccc cgaaaagtgc cacctgacgc 9420

gccctgtagc ggcgcattaa gcgcggcggg tgtggtggtt acgcgcagcg tgaccgctac 9480

acttgccagc gccctagcgc ccgctccttt cgctttcttc ccttcctttc tcgccacgtt 9540

cgccggcttt ccccgtcaag ctctaaatcg ggggctccct ttagggttcc gatttagtgc 9600

tttacggcac ctcgacccca aaaaacttga ttagggtgat ggttcacgta gtgggccatc 9660

gccctgatag acggtttttc gccctttgac gttggagtcc acgttcttta atagtggact 9720

cttgttccaa actggaacaa cactcaaccc tatctcggtc tattcttttg atttataagg 9780

gattttgccg atttcggcct attggttaaa aaatgagctg atttaacaaa aatttaacgc 9840

gaattttaac aaaatattaa cgcttacaat ttccattcgc cattcaggct gcgcaactgt 9900

tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa agggggatgt 9960

gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg 10020

acggccagtg aattgtaata cgactcacta tagggcgaat tgggtaccgg gccccccctc 10080

gaggtcgatg gtgtcgataa gcttgatatc gaattcatgt cacacaaacc gatcttcgcc 10140

tcaaggaaac ctaattctac atccgagaga ctgccgagat ccagtctaca ctgattaatt 10200

ttcgggccaa taatttaaaa aaatcgtgtt atataatatt atatgtatta tatatataca 10260

tcatgatgat actgacagtc atgtcccatt gctaaataga cagactccat ctgccgcctc 10320

caactgatgt tctcaatatt taaggggtca tctcgcattg tttaataata aacagactcc 10380

atctaccgcc tccaaatgat gttctcaaaa tatattgtat gaacttattt ttattactta 10440

gtattattag acaacttact tgctttatga aaaacacttc ctatttagga aacaatttat 10500

aatggcagtt cgttcattta acaatttatg tagaataaat gttataaatg cgtatgggaa 10560

atcttaaata tggatagcat aaatgatatc tgcattgcct aattcgaaat caacagcaac 10620

gaaaaaaatc ccttgtacaa cataaatagt catcgagaaa tatcaactat caaagaacag 10680

ctattcacac gttactattg agattattat tggacgagaa tcacacactc aactgtcttt 10740

ctctcttcta gaaatacagg tacaagtatg tactattctc attgttcata cttctagtca 10800

tttcatccca catattcctt ggatttctct ccaatgaatg acattctatc ttgcaaattc 10860

aacaattata ataagatata ccaaagtagc ggtatagtgg caatcaaaaa gcttctctgg 10920

tgtgcttctc gtatttattt ttattctaat gatccattaa aggtatatat ttatttcttg 10980

ttatataatc cttttgttta ttacatgggc tggatacata aaggtatttt gatttaattt 11040

tttgcttaaa ttcaatcccc cctcgttcag tgtcaactgt aatggtagga aattaccata 11100

cttttgaaga agcaaaaaaa atgaaagaaa aaaaaaatcg tatttccagg ttagacgttc 11160

cgcagaatct agaatgcggt atgcggtaca ttgttcttcg aacgtaaaag ttgcgctccc 11220

tgagatattg tacatttttg cttttacaag tacaagtaca tcgtacaact atgtactact 11280

gttgatgcat ccacaacagt ttgttttgtt tttttttgtt ttttttttt ctaatgattc 11340

attaccgcta tgtataccta cttgtacttg tagtaagccg ggttattggc gttcaattaa 11400

tcatagactt atgaatctgc acggtgtgcg ctgcgagtta cttttagctt atgcatgcta 11460

cttgggtgta atattgggat ctgttcggaa atcaacggat gctcaat 11507

<210> 43

<211> 4

<212> RNA

<213> unknown

<220><223> RNA loop-forming sequence (GAAA)

<400> 43

gaaa 4

<210> 44

<211> 4

<212> RNA

<213> unknown

<220><223> RNA loop-forming sequence (CAAA)

<400> 44

caaa 4

<210> 45

<211> 4

<212> RNA

<213> unknown

<220><223> RNA loop-forming sequence (AAAG)

<400> 45

aaag 4

<210> 46

<211> 23

<212> DNA

<213> unknown

<220><223> Example of a Cas9 target site:PAM sequence

<220><221> misc_feature

<222> (1)..(20)

<223> n = A, C, T, or G

<220><221> misc_feature

<222> (21)..(21)

<223> n = A, C, T, or G (indicated as an "X" in Specification)

<400> 46

nnnnnnnnnn nnnnnnnnnn ngg 23

<210> 47

<211> 3

<212> DNA

<213> unknown

<220><223> PAM sequence NGG

<220><221> misc_feature

<222> (1)..(1)

<223> n = A, C, T, or G

<400> 47

ngg 3

<210> 48

<211> 6

<212> DNA

<213> unknown

<220><223> PAM sequence NNAGAA

<220><221> misc_feature

<222> (1)..(2)

<223> n = A, C, T, or G

<400> 48

nnagaa 6

<210> 49

<211> 7

<212> DNA

<213> unknown

<220><223> PAM sequence NNAGAAW

<220><221> misc_feature

<222> (1)..(2)

<223> n = A, C, T, or G

<220><221> misc_feature

<222> (7)..(7)

<223> w = A or T

<400> 49

nnagaaw 7

<210> 50

<211> 5

<212> DNA

<213> unknown

<220><223> PAM sequence NGGNG

<220><221> misc_feature

<222> (1)..(1)

<223> n = A, C, T, or G

<220><221> misc_feature

<222> (4)..(4)

<223> n = A, C, T, or G

<400> 50

nggng 5

<210> 51

<211> 8

<212> DNA

<213> unknown

<220><223> PAM sequence NNNNGATT

<220><221> misc_feature

<222> (1)..(4)

<223> n = A, C, T, or G

<400> 51

nnnngatt 8

<210> 52

<211> 6

<212> DNA

<213> unknown

<220><223> PAM sequence NAAAAC

<220><221> misc_feature

<222> (1)..(1)

<223> n = A, C, T, or G

<400> 52

naaaac 6

<210> 53

<211> 2

<212> DNA

<213> unknown

<220><223> PAM sequence NG

<220><221> misc_feature

<222> (1)..(1)

<223> n = A, C, T, or G

<400> 53

ng 2

<210> 54

<211> 22

<212> RNA

<213> unknown

<220><223> TracrRNA mate sequence example 1

<400> 54

guuuuuguac ucucaagauu ua 22

<210> 55

<211> 15

<212> RNA

<213> unknown

<220><223> TracrRNA mate sequence example 2

<400> 55

guuuuuguac ucuca 15

<210> 56

<211> 12

<212> RNA

<213> unknown

<220><223> TracrRNA mate sequence example 3

<400> 56

guuuuagagc ua 12

<210> 57

<211> 13

<212> RNA

<213> unknown

<220><223> TracrRNA mate sequence example 4

<400> 57

guuuuagagc uag 13

<210> 58

<211> 60

<212> RNA

<213> Streptococcus pyogenes

<400> 58

uagcaaguua aaauaaggcu aguccguuau caacuugaaa aaguggcacc gagucggugc 60

<210> 59

<211> 45

<212> RNA

<213> Streptococcus pyogenes

<400> 59

uagcaaguua aaauaaggcu aguccguuau caacuugaaa aagug 45

<210> 60

<211> 32

<212> RNA

<213> Streptococcus pyogenes

<400> 60

uagcaaguua aaauaaggcu aguccguuau ca 32

<210> 61

<211> 85

<212> RNA

<213> Streptococcus thermophilus

<400> 61

uaaaucuugc agaagcuaca aagauaaggc uucaugccga aaucaacacc cugucauuuu 60

auggcagggu guuuucguua uuuaa 85

<210> 62

<211> 77

<212> RNA

<213> Streptococcus thermophilus

<400> 62

ugcagaagcu acaaagauaa ggcuucaugc cgaaaucaac acccugucau uuuauggcag 60

gguguuuucg uuauuua 77

<210> 63

<211> 65

<212> RNA

<213> Streptococcus thermophilus

<400> 63

ugcagaagcu acaaagauaa ggcuucaugc cgaaaucaac acccugucau uuuauggcag 60

ggugu 65

<210> 64

<211> 131

<212> RNA

<213> artificial sequence

<220><223> gRNA example 1

<220><221> misc_feature

<222> (1)..(20)

<223> n = A, C, U, or G

<400> 64

nnnnnnnnnn nnnnnnnnnn guuuuuguac ucucaagauu uagaaauaaa ucuugcagaa 60

gcuacaaaga uaaggcuuca ugccgaaauc aacacccugu cauuuuaugg caggguguuu 120

ucguuauuua a 131

<210> 65

<211> 117

<212> RNA

<213> artificial sequence

<220><223> gRNA example 2

<220>

<221> misc_feature

<222> (1)..(20)

<223> n = A, C, U, or G

<400> 65

nnnnnnnnnn nnnnnnnnnn guuuuuguac ucucagaaau gcagaagcua caaagauaag 60

gcuucaugcc gaaaucaaca cccugucauu uuauggcagg guguuuucgu uauuuaa 117

<210> 66

<211> 104

<212> RNA

<213> artificial sequence

<220><223> gRNA example 3

<220><221> misc_feature

<222> (1)..(20)

<223> n = A, C, U, or G

<400> 66

nnnnnnnnnn nnnnnnnnnn guuuuuguac ucucagaaau gcagaagcua caaagauaag 60

gcuucaugcc gaaaucaaca cccugucauu uuauggcagg gugu 104

<210> 67

<211> 99

<212> RNA

<213> artificial sequence

<220><223> gRNA example 4

<220><221> misc_feature

<222> (1)..(20)

<223> n = A, C, U, or G

<400> 67

nnnnnnnnnn nnnnnnnnnn guuuuuguac ucucagaaau agcaaguuaa aauaaggcua 60

guccguuauc aacuugaaaa aguggcaccg agucggugc 99

<210> 68

<211> 81

<212> RNA

<213> artificial sequence

<220><223> gRNA example 5

<220><221> misc_feature

<222> (1)..(20)

<223> n = A, C, U, or G

<400> 68

nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu g 81

<210> 69

<211> 68

<212> RNA

<213> artificial sequence

<220><223> gRNA example 6

<220><221> misc_feature

<222> (1)..(20)

<223> n = A, C, U, or G

<400> 69

nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuauca 68

<210> 70

<211> 100

<212> RNA

<213> artificial sequence

<220><223> gRNA example 7

<220><221> misc_feature

<222> (1)..(20)

<223> n = A, C, U, or G

<400> 70

nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 71

<211> 53

<212> DNA

<213> Yarrowia lipolytica

<400> 71

caatggaaaa gacattttca aacgattacc caccctccgg gactgaggcc cac 53

<210> 72

<211> 52

<212> DNA

<213> artificial sequence

<220><223> Sequence 1 shown in Figure 5

<400> 72

caatggaaaa gacattttca aacgattacc cacctccggg actgaggccc ac 52

<210> 73

<211> 51

<212> DNA

<213> artificial sequence

<220><223> Sequence 2 shown in Figure 5

<400> 73

caatggaaaa gacattttca aacgattacc cactccggga ctgaggccca c 51

<210> 74

<211> 49

<212> DNA

<213> artificial sequence

<220><223> Sequence 3 shown in Figure 5

<400> 74

caatggaaaa gacattttca aacgattacc caccgggact gaggcccac 49

<210> 75

<211> 46

<212> DNA

<213> artificial sequence

<220><223> Sequence 4 shown in Figure 5

<400> 75

caatggaaaa gacattttca aacgattacc cgggactgag gcccac 46

<210> 76

<211> 13

<212> DNA

<213> artificial sequence

<220><223> Sequence 5 shown in Figure 5

<400> 76

cactgaggcc cac 13

<210> 77

<211> 54

<212> DNA

<213> artificial sequence

<220><223> Sequence 6 shown in Figure 5

<400> 77

caatggaaaa gacattttca aacgattacc caccactccg ggactgaggc ccac 54

<210> 78

<211> 48

<212> DNA

<213> artificial sequence

<220><223> Sequence 7 shown in Figure 5

<400> 78

caatggaaaa gacattttca aacgattacc tccgggactg aggcccac 48

<210> 79

<211> 46

<212> DNA

<213> artificial sequence

<220><223> Sequence 8 shown in Figure 5

<400> 79

caatggaaaa gacattttca aacgattacc cgggactgag gcccac 46

<210> 80

<211> 44

<212> DNA

<213

> artificial sequence

<220><223> Sequence 9 shown in Figure 5

<400> 80

caatggaaaa gacattttca aacgattacc cacctgaggc ccac 44

<210> 81

<211> 48

<212> DNA

<213> artificial sequence

<220><223> Sequence 10 shown in Figure 5

<400> 81

caatggaaaa gacattttca aacgattacc cacgggactg aggcccac 48

<210> 82

<211> 40

<212> DNA

<213> artificial sequence

<220><223> Sequence 11 shown in Figure 5

<400> 82

caatggaaaa gacattttca aacgattaca cacggcccac 40

<210> 83

<211> 37

<212> DNA

<213> artificial sequence

<220><223> Sequence 12 shown in Figure 5

<400> 83

caatggaaaa gacattttct ccgggactga ggcccac 37

<210> 84

<211> 43

<212> DNA

<213> artificial sequence

<220><223> Sequence 13 shown in Figure 5

<400> 84

caatggaaaa gacattttca aacgctccgg gactgaggcc cac 43

<210> 85

<211> 43

<212> DNA

<213> artificial sequence

<220><223> Sequence 14 shown in Figure 5

<400>

85

caatggaaaa gacattttca aacgatccgg gactgaggcc cac 43

<210> 86

<211> 33

<212> DNA

<213> artificial sequence

<220><223> Sequence 15 shown in Figure 5

<400> 86

caatggaaaa gacattttca aacgataccc cac 33

<210> 87

<211> 54

<212> DNA

<213> artificial sequence

<220><223> Sequence 16 shown in Figure 5

<400> 87

caatggaaaa gacattttca aacgattacc cacccctccg ggactgaggc ccac 54

<210> 88

<211> 49

<212> DNA

<213> artificial sequence

<220><223> Sequence 17 shown in Figure 5

<400> 88

caatggaaaa gacattttca aacgattacc ctccgggact gaggcccac 49

<210> 89

<211> 54

<212> DNA

<213> artificial sequence

<220><223> Sequence 18 shown in Figure 5

<400> 89

caatggaaaa gacattttca aacgattacc cacccctccg ggactgaggc ccac 54

<210> 90

<211> 47

<212> DNA

<213> Artificial sequence

<220><223> AarI-removal-1

<400> 90

agaagtatcc taccatctac catctccgaa agaaactcgt cgattcc 47

<210> 91

<211> 47

<212> DNA

<213> Artificial sequence

<220><223> AarI-removal-2

<400> 91

ggaatcgacg agtttctttc ggagatggta gatggtagga tacttct 47

<210> 92

<211> 10706

<212> DNA

<213> Artificial sequence

<220><223> pRF109

<400> 92

catggacaag aaatactcca tcggcctgga cattggaacc aactctgtcg gctgggctgt 60

catcaccgac gagtacaagg tgccctccaa gaaattcaag gtcctcggaa acaccgatcg 120

acactccatc aagaaaaacc tcattggtgc cctgttgttc gattctggcg agactgccga 180

agctaccaga ctcaagcgaa ctgctcggcg acgttacacc cgacggaaga accgaatctg 240

ctacctgcag gagatctttt ccaacgagat ggccaaggtg gacgattcgt tctttcatcg 300

actggaggaa tccttcctcg tcgaggaaga caagaaacac gagcgtcatc ccatctttgg 360

caacattgtg gacgaggttg cttaccacga gaagtatcct accatctacc atctccgaaa 420

gaaactcgtc gattccaccg acaaggcgga tctcagactt atctacctcg ctctggcaca 480

catgatcaag tttcgaggtc atttcctcat cgagggcgat ctcaatcccg acaacagcga 540

tgtggacaag ctgttcattc agctcgttca gacctacaac cagctgttcg aggaaaaccc 600

catcaatgcc tccggagtcg atgcaaaggc catcttgtct gctcgactct cgaagagcag 660

acgactggag aacctcattg cccaacttcc tggcgagaaa aagaacggac tgtttggcaa 720

cctcattgcc ctttctcttg gtctcacacc caacttcaag tccaacttcg atctggcgga 780

ggacgccaag ctccagctgt ccaaggacac ctacgacgat gacctcgaca acctgcttgc 840

acagattggc gatcagtacg ccgacctgtt tctcgctgcc aagaaccttt cggatgctat 900

tctcttgtct gacattctgc gagtcaacac cgagatcaca aaggctcccc tttctgcctc 960

catgatcaag cgatacgacg agcaccatca ggatctcaca ctgctcaagg ctcttgtccg 1020

acagcaactg cccgagaagt acaaggagat ctttttcgat cagtcgaaga acggctacgc 1080

tggatacatc gacggcggag cctctcagga agagttctac aagttcatca agccaattct 1140

cgagaagatg gacggaaccg aggaactgct tgtcaagctc aatcgagagg atctgcttcg 1200

gaagcaacga accttcgaca acggcagcat tcctcatcag atccacctcg gtgagctgca 1260

cgccattctt cgacgtcagg aagacttcta ccccttttctc aaggacaacc gagagaagat 1320

cgagaagatt cttacctttc gaatccccta ctatgttggt cctcttgcca gaggaaactc 1380

tcgatttgct tggatgactc gaaagtccga ggaaaccatc actccctgga acttcgagga 1440

agtcgtggac aagggtgcct ctgcacagtc cttcatcgag cgaatgacca acttcgacaa 1500

gaatctgccc aacgagaagg ttcttcccaa gcattcgctg ctctacgagt actttacagt 1560

ctacaacgaa ctcaccaaag tcaagtacgt taccgaggga atgcgaaagc ctgccttctt 1620

gtctggcgaa cagaagaaag ccattgtcga tctcctgttc aagaccaacc gaaaggtcac 1680

tgttaagcag ctcaaggagg actacttcaa gaaaatcgag tgtttcgaca gcgtcgagat 1740

ttccggagtt gaggaccgat tcaacgcctc tttgggcacc tatcacgatc tgctcaagat 1800

tatcaaggac aaggattttc tcgacaacga ggaaaacgag gacattctgg aggacatcgt 1860

gctcactctt accctgttcg aagatcggga gatgatcgag gaacgactca agacatacgc 1920

tcacctgttc gacgacaagg tcatgaaaca actcaagcga cgtagataca ccggctgggg 1980

aagactttcg cgaaagctca tcaacggcat cagagacaag cagtccggaa agaccattct 2040

ggactttctc aagtccgatg gctttgccaa ccgaaacttc atgcagctca ttcacgacga 2100

ttctcttacc ttcaaggagg acatccagaa ggcacaagtg tccggtcagg gcgacagctt 2160

gcacgaacat attgccaacc tggctggttc gccagccatc aagaaaggca ttctccagac 2220

tgtcaaggtt gtcgacgagc tggtgaaggt catgggacgt cacaagcccg agaacattgt 2280

gatcgagatg gccagagaga accagacaac tcaaaagggt cagaaaaact cgcgagagcg 2340

gatgaagcga atcgaggaag gcatcaagga gctgggatcc cagattctca aggagcatcc 2400

cgtcgagaac actcaactgc agaacgagaa gctgtatctc tactatctgc agaatggtcg 2460

agacatgtac gtggatcagg aactggacat caatcgtctc agcgactacg atgtggacca 2520

cattgtccct caatcctttc tcaaggacga ttctatcgac aacaaggtcc ttacacgatc 2580

cgacaagaac agaggcaagt cggacaacgt tcccagcgaa gaggtggtca aaaagatgaa 2640

gaactactgg cgacagctgc tcaacgccaa gctcattacc cagcgaaagt tcgacaatct 2700

taccaaggcc gagcgaggcg gtctgtccga gctcgacaag gctggcttca tcaagcgtca 2760

actcgtcgag accagacaga tcacaaagca cgtcgcacag attctcgatt ctcggatgaa 2820

caccaagtac gacgagaacg acaagctcat ccgagaggtc aaggtgatta ctctcaagtc 2880

caaactggtc tccgatttcc gaaaggactt tcagttctac aaggtgcgag agatcaacaa 2940

ttaccaccat gcccacgatg cttacctcaa cgccgtcgtt ggcactgcgc tcatcaagaa 3000

ataccccaag ctcgaaagcg agttcgttta cggcgattac aaggtctacg acgttcgaaa 3060

gatgattgcc aagtccgaac aggagattgg caaggctact gccaagtact tcttttactc 3120

caacatcatg aacttttca agaccgagat caccttggcc aacggagaga ttcgaaagag 3180

accacttatc gagaccaacg gcgaaactgg agagatcgtg tgggacaagg gtcgagactt 3240

tgcaaccgtg cgaaaggttc tgtcgatgcc tcaggtcaac atcgtcaaga aaaccgaggt 3300

tcagactggc ggattctcca aggagtcgat tctgcccaag cgaaactccg acaagctcat 3360

cgctcgaaag aaagactggg atcccaagaa atacggtggc ttcgattctc ctaccgtcgc 3420

ctattccgtg cttgtcgttg cgaaggtcga gaagggcaag tccaaaaagc tcaagtccgt 3480

caaggagctg ctcggaatta ccatcatgga gcgatcgagc ttcgagaaga atcccatcga 3540

cttcttggaa gccaagggtt acaaggaggt caagaaagac ctcattatca agctgcccaa 3600

gtactctctg ttcgaactgg agaacggtcg aaagcgtatg ctcgcctccg ctggcgagct 3660

gcagaaggga aacgagcttg ccttgccttc gaagtacgtc aactttctct atctggcttc 3720

tcactacgag aagctcaagg gttctcccga ggacaacgaa cagaagcaac tcttcgttga 3780

gcagcacaaa cattacctcg acgagattat cgagcagatt tccgagtttt cgaagcgagt 3840

catcctggct gatgccaact tggacaaggt gctctctgcc tacaacaagc atcgggacaa 3900

acccattcga gaacaggcgg agaacatcat tcacctgttt actcttacca acctgggtgc 3960

tcctgcagct ttcaagtact tcgataccac tatcgaccga aagcggtaca catccaccaa 4020

ggaggttctc gatgccaccc tgattcacca gtccatcact ggcctgtacg agacccgaat 4080

cgacctgtct cagcttggtg gcgactccag agccgatccc aagaaaaagc gaaaggtcta 4140

agcggccgca agtgtggatg gggaagtgag tgcccggttc tgtgtgcaca attggcaatc 4200

caagatggat ggattcaaca cagggatata gcgagctacg tggtggtgcg aggatatagc 4260

aacggatatt tatgtttgac acttgagaat gtacgataca agcactgtcc aagtacaata 4320

ctaaacatac tgtacatact catactcgta cccgggcaac ggtttcactt gagtgcagtg 4380

gctagtgctc ttactcgtac agtgtgcaat actgcgtatc atagtctttg atgtatatcg 4440
tattcattca tgttagttgc gtacgagccg gaagcataaa gtgtaaagcc tggggtgcct 4500
aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa 4560
acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta 4620
ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc 4680
gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg 4740
caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt 4800

tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa 4860
gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct 4920
ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc 4980
cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg 5040
tcgttcgctc caagctgggc tgtgtgcacg aacccccccgt tcagcccgac cgctgcgcct 5100
tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag 5160
cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga 5220

agtggtggcc taactacggc tacactagaa ggacagtatt tggtatctgc gctctgctga 5280
agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg 5340
gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag 5400
aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag 5460
ggattttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat 5520
gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct 5580
taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac 5640

tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa 5700
tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg 5760
gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt 5820
gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca 5880
ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt 5940
cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct 6000
tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg 6060

cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg 6120

agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg 6180

cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa 6240

aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt 6300

aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt 6360

gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt 6420

gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca 6480

tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat 6540

ttccccgaaa agtgccacct gacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg 6600

tggttacgcg cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt 6660

tcttcccttc ctttctcgcc acgttcgccg gctttccccg tcaagctcta aatcgggggc 6720

tcccttttagg gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg 6780

gtgatggttc acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg 6840

agtccacgtt ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct 6900

cggtctattc ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg 6960

agctgattta acaaaaattt aacgcgaatt ttaacaaaat attaacgctt acaatttcca 7020

ttcgccattc aggctgcgca actgttggga agggcgatcg gtgcgggcct cttcgctatt 7080

acgccagctg gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa cgccagggtt 7140

ttcccagtca cgacgttgta aaacgacggc cagtgaattg taatacgact cactataggg 7200

cgaattgggt accgggcccc ccctcgaggt cgatggtgtc gataagcttg atatcgaatt 7260

catgtcacac aaaccgatct tcgcctcaag gaaacctaat tctacatccg agagactgcc 7320

gagatccagt ctacactgat taattttcgg gccaataatt taaaaaaatc gtgttatata 7380

atattatatg tattatatat atacatcatg atgatactga cagtcatgtc ccattgctaa 7440

atagacagac tccatctgcc gcctccaact gatgttctca atatttaagg ggtcatctcg 7500

cattgtttaa taataaacag actccatcta ccgcctccaa atgatgttct caaaatatat 7560

tgtatgaact tatttttatt acttagtatt attagacaac ttacttgctt tatgaaaaac 7620

acttcctatt taggaaacaa tttataatgg cagttcgttc atttaacaat ttatgtagaa 7680

taaatgttat aaatgcgtat gggaaatctt aaatatggat agcataaatg atatctgcat 7740

tgcctaattc gaaatcaaca gcaacgaaaa aaatcccttg tacaacataa atagtcatcg 7800

agaaatatca actatcaaag aacagctatt cacacgttac tattgagatt attattggac 7860

gagaatcaca cactcaactg tctttctctc ttctagaaat acaggtacaa gtatgtacta 7920

ttctcattgt tcatacttct agtcatttca tcccacatat tccttggatt tctctccaat 7980

gaatgacatt ctatcttgca aattcaacaa ttataataag atataccaaa gtagcggtat 8040

agtggcaatc aaaaagcttc tctggtgtgc ttctcgtatt tatttttatt ctaatgatcc 8100

attaaaggta tatatttatt tcttgttata taatcctttt gtttattaca tgggctggat 8160

acataaaggt attttgattt aatttttgc ttaaattcaa tccccctcg ttcagtgtca 8220

actgtaatgg taggaaatta ccatactttt gaagaagcaa aaaaaatgaa agaaaaaaaa 8280

aatcgtattt ccaggttaga cgttccgcag aatctagaat gcggtatgcg gtacattgtt 8340

cttcgaacgt aaaagttgcg ctccctgaga tattgtacat ttttgctttt acaagtacaa 8400

gtacatcgta caactatgta ctactgttga tgcatccaca acagtttgtt ttgttttttt 8460

ttgttttttt tttttctaat gattcattac cgctatgtat acctacttgt acttgtagta 8520

agccgggtta ttggcgttca attaatcata gacttatgaa tctgcacggt gtgcgctgcg 8580

agttactttt agcttatgca tgctacttgg gtgtaatatt gggatctgtt cggaaatcaa 8640

cggatgctca atcgatttcg acagtaatta attaagtcat acacaagtca gctttcttcg 8700

agcctcatat aagtataagt agttcaacgt attagcactg tacccagcat ctccgtatcg 8760

agaaacacaa caacatgccc cattggacag atcatgcgga tacacaggtt gtgcagtatc 8820

atacatactc gatcagacag gtcgtctgac catcatacaa gctgaacaag cgctccatac 8880

ttgcacgctc tctatataca cagttaaatt acatatccat agtctaacct ctaacagtta 8940

atcttctggt aagcctccca gccagccttc tggtatcgct tggcctcctc aataggatct 9000

cggttctggc cgtacagacc tcggccgaca attatgatat ccgttccggt agacatgaca 9060

tcctcaacag ttcggtactg ctgtccgaga gcgtctccct tgtcgtcaag acccaccccg 9120

ggggtcagaa taagccagtc ctcagagtcg cccttaggtc ggttctgggc aatgaagcca 9180

accacaaact cggggtcgga tcgggcaagc tcaatggtct gcttggagta ctcgccagtg 9240

gccagagagc ccttgcaaga cagctcggcc agcatgagca gacctctggc cagcttctcg 9300

ttgggagagg ggactaggaa ctccttgtac tgggagttct cgtagtcaga gacgtcctcc 9360

ttcttctgtt cagagacagt ttcctcggca ccagctcgca ggccagcaat gattccggtt 9420

ccgggtacac cgtgggcgtt ggtgatatcg gaccactcgg cgattcggtg acaccggtac 9480

tggtgcttga cagtgttgcc aatatctgcg aactttctgt cctcgaacag gaagaaaccg 9540

tgcttaagag caagttcctt gaggggggagc acagtgccgg cgtaggtgaa gtcgtcaatg 9600

atgtcgatat gggttttgat catgcacaca taaggtccga ccttatcggc aagctcaatg 9660

agctccttgg tggtggtaac atccagagaa gcacacaggt tggttttctt ggctgccacg 9720

agcttgagca ctcgagcggc aaaggcggac ttgtggacgt tagctcgagc ttcgtaggag 9780

ggcattttgg tggtgaagag gagactgaaa taaatttagt ctgcagaact ttttatcgga 9840

accttatctg gggcagtgaa gtatatgtta tggtaatagt tacgagttag ttgaacttat 9900
agatagactg gactatacgg ctatcggtcc aaattagaaa gaacgtcaat ggctctctgg 9960
gcgtcgcctt tgccgacaaa aatgtgatca tgatgaaagc cagcaatgac gttgcagctg 10020
atattgttgt cggccaaccg cgccgaaaac gcagctgtca gacccacagc ctccaacgaa 10080
gaatgtatcg tcaaagtgat ccaagcacac tcatagttgg agtcgtactc caaaggcggc 10140
aatgacgagt cagacagata ctcgtcgacg tttaaaccat catctaaggg cctcaaaact 10200
acctcggaac tgctgcgctg atctggacac cacagaggtt ccgagcactt taggttgcac 10260

caaatgtccc accaggtgca ggcagaaaac gctggaacag cgtgtacagt ttgtcttaac 10320
aaaaagtgag ggcgctgagg tcgagcaggg tggtgtgact tgttatagcc tttagagctg 10380
cgaaagcgcg tatggatttg gctcatcagg ccagattgag ggtctgtgga cacatgtcat 10440
gttagtgtac ttcaatcgcc ccctggatat agccccgaca ataggccgtg gcctcatttt 10500
tttgccttcc gcacatttcc attgctcggt acccacacct tgcttctcct gcacttgcca 10560
accttaatac tggtttacat tgaccaacat cttacaagcg gggggcttgt ctagggtata 10620
tataaacagt ggctctccca atcggttgcc agtctctttt ttcctttctt tccccacaga 10680

ttcgaaatct aaactacaca tcacac 10706

<210> 93
<211> 4140
<212> DNA
<213> Artificial sequence
<220><223> Aar1- Cas9 gene
<400> 93

atggacaaga aatactccat cggcctggac attggaacca actctgtcgg ctgggctgtc 60
atcaccgacg agtacaaggt gccctccaag aaattcaagg tcctcggaaa caccgatcga 120
cactccatca agaaaaacct cattggtgcc ctgttgttcg attctggcga gactgccgaa 180
gctaccagac tcaagcgaac tgctcggcga cgttacaccc gacggaagaa ccgaatctgc 240
tacctgcagg agatcttttc caacgagatg gccaaggtgg acgattcgtt ctttcatcga 300

ctggaggaat ccttcctcgt cgaggaagac aagaaacacg agcgtcatcc catctttggc 360
aacattgtgg acgaggttgc ttaccacgag aagtatccta ccatctacca tctccgaaag 420
aaactcgtcg attccaccga caaggcggat ctcagactta tctacctcgc tctggcacac 480
atgatcaagt ttcgaggtca tttcctcatc gagggcgatc tcaatcccga caacagcgat 540

gtggacaagc tgttcattca gctcgttcag acctacaacc agctgttcga ggaaaacccc 600

atcaatgcct ccggagtcga tgcaaaggcc atcttgtctg ctcgactctc gaagagcaga 660

cgactggaga acctcattgc ccaacttcct ggcgagaaaa agaacggact gtttggcaac 720

ctcattgccc tttctcttgg tctcacaccc aacttcaagt ccaacttcga tctggcggag 780

gacgccaagc tccagctgtc caaggacacc tacgacgatg acctcgacaa cctgcttgca 840

cagattggcg atcagtacgc cgacctgttt ctcgctgcca agaacctttc ggatgctatt 900

ctcttgtctg acattctgcg agtcaacacc gagatcacaa aggctcccct ttctgcctcc 960

atgatcaagc gatacgacga gcaccatcag gatctcacac tgctcaaggc tcttgtccga 1020

cagcaactgc ccgagaagta caaggagatc tttttcgatc agtcgaagaa cggctacgct 1080

ggatacatcg acggcggagc ctctcaggaa gagttctaca agttcatcaa gccaattctc 1140

gagaagatgg acggaaccga ggaactgctt gtcaagctca atcgagagga tctgcttcgg 1200

aagcaacgaa ccttcgacaa cggcagcatt cctcatcaga tccacctcgg tgagctgcac 1260

gccattcttc gacgtcagga agacttctac ccctttctca aggacaaccg agagaagatc 1320

gagaagattc ttacctttcg aatcccctac tatgttggtc ctcttgccag aggaaactct 1380

cgatttgctt ggatgactcg aaagtccgag gaaaccatca ctccctggaa cttcgaggaa 1440

gtcgtggaca agggtgcctc tgcacagtcc ttcatcgagc gaatgaccaa cttcgacaag 1500

aatctgccca acgagaaggt tcttcccaag cattcgctgc tctacgagta ctttacagtc 1560

tacaacgaac tcaccaaagt caagtacgtt accgagggaa tgcgaaagcc tgccttcttg 1620

tctggcgaac agaagaaagc cattgtcgat ctcctgttca agaccaaccg aaaggtcact 1680

gttaagcagc tcaaggagga ctacttcaag aaaatcgagt gtttcgacag cgtcgagatt 1740

tccggagttg aggaccgatt caacgcctct ttgggcacct atcacgatct gctcaagatt 1800

atcaaggaca aggattttct cgacaacgag gaaaacgagg acattctgga ggacatcgtg 1860

ctcactctta ccctgttcga agatcgggag atgatcgagg aacgactcaa gacatacgct 1920

cacctgttcg acgacaaggt catgaaacaa ctcaagcgac gtagatacac cggctgggga 1980

agactttcgc gaaagctcat caacggcatc agagacaagc agtccggaaa gaccattctg 2040

gactttctca agtccgatgg ctttgccaac cgaaacttca tgcagctcat tcacgacgat 2100

tctcttacct tcaaggagga catccagaag gcacaagtgt ccggtcaggg cgacagcttg 2160

cacgaacata ttgccaacct ggctggttcg ccagccatca agaaaggcat tctccagact 2220

gtcaaggttg tcgacgagct ggtgaaggtc atgggacgtc acaagcccga gaacattgtg 2280

atcgagatgg ccagagagaa ccagacaact caaaagggtc agaaaaactc gcgagagcgg 2340

atgaagcgaa tcgaggaagg catcaaggag ctgggatccc agattctcaa ggagcatccc 2400

gtcgagaaca ctcaactgca gaacgagaag ctgtatctct actatctgca gaatggtcga 2460
gacatgtacg tggatcagga actggacatc aatcgtctca gcgactacga tgtggaccac 2520
attgtccctc aatcctttct caaggacgat tctatcgaca acaaggtcct tacacgatcc 2580
gacaagaaca gaggcaagtc ggacaacgtt cccagcgaag aggtggtcaa aaagatgaag 2640
aactactggc gacagctgct caacgccaag ctcattaccc agcgaaagtt cgacaatctt 2700
accaaggccg agcgaggcgg tctgtccgag ctcgacaagg ctggcttcat caagcgtcaa 2760
ctcgtcgaga ccagacagat cacaaagcac gtcgcacaga ttctcgattc tcggatgaac 2820

accaagtacg acgagaacga caagctcatc cgagaggtca aggtgattac tctcaagtcc 2880
aaactggtct ccgatttccg aaaggacttt cagttctaca aggtgcgaga gatcaacaat 2940
taccaccatg cccacgatgc ttacctcaac gccgtcgttg gcactgcgct catcaagaaa 3000
taccccaagc tcgaaagcga gttcgtttac ggcgattaca aggtctacga cgttcgaaag 3060
atgattgcca agtccgaaca ggagattggc aaggctactg ccaagtactt cttttactcc 3120
aacatcatga acttttcaa gaccgagatc accttggcca acggagagat tcgaaagaga 3180
ccacttatcg agaccaacgg cgaaactgga gagatcgtgt gggacaaggg tcgagacttt 3240

gcaaccgtgc gaaaggttct gtcgatgcct caggtcaaca tcgtcaagaa aaccgaggtt 3300
cagactggcg gattctccaa ggagtcgatt ctgcccaagc gaaactccga caagctcatc 3360
gctcgaaaga aagactggga tcccaagaaa tacggtggct tcgattctcc taccgtcgcc 3420
tattccgtgc ttgtcgttgc gaaggtcgag aagggcaagt ccaaaaagct caagtccgtc 3480
aaggagctgc tcggaattac catcatggag cgatcgagct tcgagaagaa tcccatcgac 3540
ttcttggaag ccaagggtta caaggaggtc aagaaagacc tcattatcaa gctgcccaag 3600
tactctctgt tcgaactgga gaacggtcga aagcgtatgc tcgcctccgc tggcgagctg 3660

cagaagggaa acgagcttgc cttgccttcg aagtacgtca actttctcta tctggcttct 3720
cactacgaga agctcaaggg ttctcccgag gacaacgaac agaagcaact cttcgttgag 3780
cagcacaaac attacctcga cgagattatc gagcagattt ccgagttttc gaagcgagtc 3840
atcctggctg atgccaactt ggacaaggtg ctctctgcct acaacaagca tcgggacaaa 3900
cccattcgag aacaggcgga gaacatcatt cacctgttta ctcttaccaa cctgggtgct 3960
cctgcagctt tcaagtactt cgataccact atcgaccgaa agcggtacac atccaccaag 4020
gaggttctcg atgccaccct gattcaccag tccatcactg gcctgtacga gacccgaatc 4080

gacctgtctc agcttggtgg cgactccaga gccgatccca agaaaaagcg aaaggtctaa 4140

<210> 94

<211> 10706

<212> DNA

<213> ARtificial sequence

<220><223> pRF141

<400> 94

catggacaag aaatactcca tcggcctgga cattggaacc aactctgtcg gctgggctgt 60

catcaccgac gagtacaagg tgccctccaa gaaattcaag gtcctcggaa acaccgatcg 120

acactccatc aagaaaaacc tcattggtgc cctgttgttc gattctggcg agactgccga 180

agctaccaga ctcaagcgaa ctgctcggcg acgttacacc cgacggaaga accgaatctg 240

ctacctgcag gagatctttt ccaacgagat ggccaaggtg gacgattcgt tctttcatcg 300

actggaggaa tccttcctcg tcgaggaaga caagaaacac gagcgtcatc ccatctttgg 360

caacattgtg gacgaggttg cttaccacga gaagtatcct accatctacc atctccgaaa 420

gaaactcgtc gattccaccg acaaggcgga tctcagactt atctacctcg ctctggcaca 480

catgatcaag tttcgaggtc atttcctcat cgagggcgat ctcaatcccg acaacagcga 540

tgtggacaag ctgttcattc agctcgttca gacctacaac cagctgttcg aggaaaaccc 600

catcaatgcc tccggagtcg atgcaaaggc catcttgtct gctcgactct cgaagagcag 660

acgactggag aacctcattg cccaacttcc tggcgagaaa aagaacggac tgtttggcaa 720

cctcattgcc ctttctcttg gtctcacacc caacttcaag tccaacttcg atctggcgga 780

ggacgccaag ctccagctgt ccaaggacac ctacgacgat gacctcgaca acctgcttgc 840

acagattggc gatcagtacg ccgacctgtt tctcgctgcc aagaaccttt cggatgctat 900

tctcttgtct gacattctgc gagtcaacac cgagatcaca aaggctcccc tttctgcctc 960

catgatcaag cgatacgacg agcaccatca ggatctcaca ctgctcaagg ctcttgtccg 1020

acagcaactg cccgagaagt acaaggagat ctttttcgat cagtcgaaga acggctacgc 1080

tggatacatc gacggcggag cctctcagga agagttctac aagttcatca agccaattct 1140

cgagaagatg gacggaaccg aggaactgct tgtcaagctc aatcgagagg atctgcttcg 1200

gaagcaacga accttcgaca acggcagcat tcctcatcag atccacctcg gtgagctgca 1260

cgccattctt cgacgtcagg aagacttcta ccccttttctc aaggacaacc gagagaagat 1320

cgagaagatt cttacctttc gaatccccta ctatgttggt cctcttgcca gaggaaactc 1380

tcgatttgct tggatgactc gaaagtccga ggaaaccatc actccctgga acttcgagga 1440

agtcgtggac aagggtgcct ctgcacagtc cttcatcgag cgaatgacca acttcgacaa 1500

gaatctgccc aacgagaagg ttcttcccaa gcattcgctg ctctacgagt actttacagt 1560

ctacaacgaa ctcaccaaag tcaagtacgt taccgaggga atgcgaaagc ctgccttctt 1620

gtctggcgaa cagaagaaag ccattgtcga tctcctgttc aagaccaacc gaaaggtcac 1680

tgttaagcag ctcaaggagg actacttcaa gaaaatcgag tgtttcgaca gcgtcgagat 1740

ttccggagtt gaggaccgat tcaacgcctc tttgggcacc tatcacgatc tgctcaagat 1800

tatcaaggac aaggattttc tcgacaacga ggaaaacgag gacattctgg aggacatcgt 1860

gctcactctt accctgttcg aagatcggga gatgatcgag gaacgactca agacatacgc 1920

tcacctgttc gacgacaagg tcatgaaaca actcaagcga cgtagataca ccggctgggg 1980

aagactttcg cgaaagctca tcaacggcat cagagacaag cagtccggaa agaccattct 2040

ggactttctc aagtccgatg gctttgccaa ccgaaacttc atgcagctca ttcacgacga 2100

ttctcttacc ttcaaggagg acatccagaa ggcacaagtg tccggtcagg gcgacagctt 2160

gcacgaacat attgccaacc tggctggttc gccagccatc aagaaaggca ttctccagac 2220

tgtcaaggtt gtcgacgagc tggtgaaggt catgggacgt cacaagcccg agaacattgt 2280

gatcgagatg gccagagaga accagacaac tcaaaagggt cagaaaaact cgcgagagcg 2340

gatgaagcga atcgaggaag gcatcaagga gctgggatcc cagattctca aggagcatcc 2400

cgtcgagaac actcaactgc agaacgagaa gctgtatctc tactatctgc agaatggtcg 2460

agacatgtac gtggatcagg aactggacat caatcgtctc agcgactacg atgtggacca 2520

cattgtccct caatcctttc tcaaggacga ttctatcgac aacaaggtcc ttacacgatc 2580

cgacaagaac agaggcaagt cggacaacgt tcccagcgaa gaggtggtca aaaagatgaa 2640

gaactactgg cgacagctgc tcaacgccaa gctcattacc cagcgaaagt tcgacaatct 2700

taccaaggcc gagcgaggcg gtctgtccga gctcgacaag gctggcttca tcaagcgtca 2760

actcgtcgag accagacaga tcacaaagca cgtcgcacag attctcgatt ctcggatgaa 2820

caccaagtac gacgagaacg acaagctcat ccgagaggtc aaggtgatta ctctcaagtc 2880

caaactggtc tccgatttcc gaaaggactt tcagttctac aaggtgcgag agatcaacaa 2940

ttaccaccat gcccacgatg cttacctcaa cgccgtcgtt ggcactgcgc tcatcaagaa 3000

ataccccaag ctcgaaagcg agttcgttta cggcgattac aaggtctacg acgttcgaaa 3060

gatgattgcc aagtccgaac aggagattgg caaggctact gccaagtact tcttttactc 3120

caacatcatg aacttttca agaccgagat caccttggcc aacggagaga ttcgaaagag 3180

accacttatc gagaccaacg gcgaaactgg agagatcgtg tgggacaagg gtcgagactt 3240

tgcaaccgtg cgaaaggttc tgtcgatgcc tcaggtcaac atcgtcaaga aaaccgaggt 3300

tcagactggc ggattctcca aggagtcgat tctgcccaag cgaaactccg acaagctcat 3360

cgctcgaaag aaagactggg atcccaagaa atacggtggc ttcgattctc ctaccgtcgc 3420

ctattccgtg cttgtcgttg cgaaggtcga gaagggcaag tccaaaaagc tcaagtccgt 3480

caaggagctg ctcggaatta ccatcatgga gcgatcgagc ttcgagaaga atcccatcga 3540

cttcttggaa gccaagggtt acaaggaggt caagaaagac ctcattatca agctgcccaa 3600

gtactctctg ttcgaactgg agaacggtcg aaagcgtatg ctcgcctccg ctggcgagct 3660

gcagaaggga aacgagcttg ccttgccttc gaagtacgtc aactttctct atctggcttc 3720

tcactacgag aagctcaagg gttctcccga ggacaacgaa cagaagcaac tcttcgttga 3780

gcagcacaaa cattacctcg acgagattat cgagcagatt tccgagtttt cgaagcgagt 3840

catcctggct gatgccaact tggacaaggt gctctctgcc tacaacaagc atcgggacaa 3900

acccattcga gaacaggcgg agaacatcat tcacctgttt actcttacca acctgggtgc 3960

tcctgcagct ttcaagtact tcgataccac tatcgaccga aagcggtaca catccaccaa 4020

ggaggttctc gatgccaccc tgattcacca gtccatcact ggcctgtacg agacccgaat 4080

cgacctgtct cagcttggtg gcgactccag agccgatccc aagaaaaagc gaaaggtcta 4140

agcggccgca agtgtggatg gggaagtgag tgcccggttc tgtgtgcaca attggcaatc 4200

caagatggat ggattcaaca cagggatata gcgagctacg tggtggtgcg aggatatagc 4260

aacggatatt tatgtttgac acttgagaat gtacgataca agcactgtcc aagtacaata 4320

ctaaacatac tgtacatact catactcgta cccgggcaac ggtttcactt gagtgcagtg 4380

gctagtgctc ttactcgtac agtgtgcaat actgcgtatc atagtctttg atgtatatcg 4440

tattcattca tgttagttgc gtacgagccg gaagcataaa gtgtaaagcc tggggtgcct 4500

aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa 4560

acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta 4620

ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc 4680

gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg 4740

caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt 4800

tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa 4860

gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct 4920

ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc 4980

cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg 5040

tcgttcgctc caagctgggc tgtgtgcacg aaccccccgt tcagcccgac cgctgcgcct 5100

tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag 5160

cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga 5220

agtggtggcc taactacggc tacactagaa ggacagtatt tggtatctgc gctctgctga 5280

agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg 5340

gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag 5400

aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag 5460

ggattttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat 5520

gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct 5580

taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac 5640

tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa 5700

tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg 5760

gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt 5820

gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca 5880

ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt 5940

cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct 6000

tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg 6060

cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg 6120

agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg 6180

cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa 6240

aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt 6300

aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt 6360

gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt 6420

gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca 6480

tgagcggata catatttgaa tgtatttaga aaaataaaca aatagggggtt ccgcgcacat 6540

ttccccgaaa agtgccacct gacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg 6600

tggttacgcg cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt 6660

tcttcccttc ctttctcgcc acgttcgccg gctttccccg tcaagctcta aatcgggggc 6720

tcccttttagg gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg 6780

gtgatggttc acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg 6840

agtccacgtt ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct 6900

cggtctattc ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg 6960

agctgattta acaaaaattt aacgcgaatt ttaacaaaat attaacgctt acaatttcca 7020

ttcgccattc aggctgcgca actgttggga agggcgatcg gtgcgggcct cttcgctatt 7080
acgccagctg gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa cgccagggtt 7140
ttcccagtca cgacgttgta aaacgacggc cagtgaattg taatacgact cactataggg 7200
cgaattgggt accgggcccc ccctcgaggt cgatggtgtc gataagcttg atatcgaatt 7260
catgtcacac aaaccgatct tcgcctcaag gaaacctaat tctacatccg agagactgcc 7320
gagatccagt ctacactgat taattttcgg gccaataatt taaaaaaatc gtgttatata 7380
atattatatg tattatatat atacatcatg atgatactga cagtcatgtc ccattgctaa 7440

atagacagac tccatctgcc gcctccaact gatgttctca atatttaagg ggtcatctcg 7500
cattgtttaa taataaacag actccatcta ccgcctccaa atgatgttct caaaatatat 7560
tgtatgaact tatttttatt acttagtatt attagacaac ttacttgctt tatgaaaaac 7620
acttcctatt taggaaacaa tttataatgg cagttcgttc atttaacaat ttatgtagaa 7680
taaatgttat aaatgcgtat gggaaatctt aaatatggat agcataaatg atatctgcat 7740
tgcctaattc gaaatcaaca gcaacgaaaa aaatcccttg tacaacataa atagtcatcg 7800
agaaatatca actatcaaag aacagctatt cacacgttac tattgagatt attattggac 7860

gagaatcaca cactcaactg tctttctctc ttctagaaat acaggtacaa gtatgtacta 7920
ttctcattgt tcatacttct agtcatttca tcccacatat tccttggatt tctctccaat 7980
gaatgacatt ctatcttgca aattcaacaa ttataataag atataccaaa gtagcggtat 8040
agtggcaatc aaaaagcttc tctggtgtgc ttctcgtatt tatttttatt ctaatgatcc 8100
attaaaggta tatatttatt tcttgttata taatcctttt gtttattaca tgggctggat 8160
acataaaggt attttgattt aatttttttgc ttaaattcaa tccccccctcg ttcagtgtca 8220
actgtaatgg taggaaatta ccatactttt gaagaagcaa aaaaaatgaa agaaaaaaaa 8280

aatcgtattt ccaggttaga cgttccgcag aatctagaat gcggtatgcg gtacattgtt 8340
cttcgaacgt aaaagttgcg ctccctgaga tattgtacat ttttgctttt acaagtacaa 8400
gtacatcgta caactatgta ctactgttga tgcatccaca acagtttgtt ttgttttttt 8460
ttgttttttt ttttttctaat gattcattac cgctatgtat acctacttgt acttgtagta 8520
agccgggtta ttggcgttca attaatcata gacttatgaa tctgcacggt gtgcgctgcg 8580
agttactttt agcttatgca tgctacttgg gtgtaatatt gggatctgtt cggaaatcaa 8640
cggatgctca atcgatttcg acagtaatta attaagtcat acacaagtca gctttcttcg 8700

agcctcatat aagtataagt agttcaacgt attagcactg tacccagcat ctccgtatcg 8760

agaaacacaa caacatgccc cattggacag atcatgcgga tacacaggtt gtgcagtatc 8820

atacatactc gatcagacag gtcgtctgac catcatacaa gctgaacaag cgctccatac 8880

ttgcacgctc tctatataca cagttaaatt acatatccat agtctaacct ctaacagtta 8940

atcttctggt aagcctccca gccagccttc tggtatcgct tggcctcctc aataggatct 9000

cggttctggc cgtacagacc tcggccgaca attatgatat ccgttccggt agacatgaca 9060

tcctcaacag ttcggtactg ctgtccgaga gcgtctccct tgtcgtcaag acccaccccg 9120

ggggtcagaa taagccagtc ctcagagtcg cccttaggtc ggttctgggc aatgaagcca 9180

accacaaact cggggtcgga tcgggcaagc tcaatggtct gcttggagta ctcgccagtg 9240

gccagagagc ccttgcaaga cagctcggcc agcatgagca gacctctggc cagcttctcg 9300

ttgggagagg ggactaggaa ctccttgtac tgggagttct cgtagtcaga gacgtcctcc 9360

ttcttctgtt cagagacagt ttcctcggca ccagctcgca ggccagcaat gattccggtt 9420

ccgggtacac cgtgggcgtt ggtgatatcg gaccactcgg cgattcggtg acaccggtac 9480

tggtgcttga cagtgttgcc aatatctgcg aactttctgt cctcgaacag gaagaaaccg 9540

tgcttaagag caagttcctt gagggggagc acagtgccgg cgtaggtgaa gtcgtcaatg 9600

atgtcgatat gggttttgat catgcacaca taaggtccga ccttatcggc aagctcaatg 9660

agctccttgg tggtggtaac atccagagaa gcacacaggt tggttttctt ggctgccacg 9720

agcttgagca ctcgagcggc aaaggcggac ttgtggacgt tagctcgagc ttcgtaggag 9780

ggcattttgg tggtgaagag gagactgaaa taaatttagt ctgcagaact ttttatcgga 9840

accttatctg gggcagtgaa gtatatgtta tggtaatagt tacgagttag ttgaacttat 9900

agatagactg gactatacgg ctatcggtcc aaattagaaa gaacgtcaat ggctctctgg 9960

gcgtcgcctt tgccgacaaa aatgtgatca tgatgaaagc cagcaatgac gttgcagctg 10020

atattgttgt cggccaaccg cgccgaaaac gcagctgtca gacccacagc ctccaacgaa 10080

gaatgtatcg tcaaagtgat ccaagcacac tcatagttgg agtcgtactc caaaggcggc 10140

aatgacgagt cagacagata ctcgtcgacg tttaaaccat catctaaggg cctcaaaact 10200

acctcggaac tgctgcgctg atctggacac cacagaggtt ccgagcactt taggttgcac 10260

caaatgtccc accaggtgca ggcagaaaac gctggaacag cgtgtacagt ttgtcttaac 10320

aaaaagtgag ggcgctgagg tcgagcaggg tggtgtgact tgttatagcc tttagagctg 10380

cgaaagcgcg tatggatttg gctcatcagg ccagattgag ggtctgtgga cacatgtcat 10440

gttagtgtac ttcaatcgcc ccctggatat agccccgaca ataggccgtg gcctcatttt 10500

tttgccttcc gcacatttcc attgctcggt acccacacct tgcttctcct gcacttgcca 10560

accttaatac tggtttacat tgaccaacat cttacaagcg gggggcttgt ctagggtata 10620

tataaacagt ggctctccca atcggttgcc agtctctttt ttcctttctt tccccacaga 10680

ttcgaaatct aaactacaca tcacac 10706

<210> 95

<211> 1048

<212> DNA

<213> Artificial sequence

<220><223> High-throughput cloning cassette

<400> 95

gcgcacgtta attaaatttt ttttgatttt ctttttgac cccgtcttca attacacttc 60

ccaactggga acacccctct ttatcgaccc attttaggta atttacccta gcccattgtc 120

tccataagga atattaccct aacccacagt ccagggtgcc caggtccttc tttggccaaa 180

ttttaacttc ggtcctatgg cacagcggta gcgcgtgaga ttgcaaatct taaggtcccg 240

agttcgaatc tcggtgggac ctagttattt ttgatagata atttcgtgat gattagaaac 300

ttaacgcaaa ataatggccg gcatggtccc agcctcctcg ctggcgccgg ctgggcaaca 360

tgcttcggca tggcgaatgg gacgcaggtg atggcgggat cgttgtatat ttcttgacac 420

cttttcggca tcgccctaaa ttcggcgtcc tcatattgtg tgaggacgtt ttattacgtg 480

tttacgaagc aaaagctaaa accaggagct atttaatggc aacagttaac cagctggtac 540

gcaaaccacg tgctcgcaaa gttgcgaaaa gcaacgtgcc tgcgctggaa gcatgcccgc 600

aaaaacgtgg cgtatgtact cgtgtatata ctaccactcc taaaaaaccg aactccgcgc 660

tgcgtaaagt atgccgtgtt cgtctgacta acggtttcga agtgacttcc tacatcggtg 720

gtgaaggtca caacctgcag gagcactccg tgatcctgat ccgtggcggt cgtgttaaag 780

acctcccggg tgttcgttac cacaccgtac gtggtgcgct tgactgctcc ggcgttaaag 840

accgtaagca ggctcgttcc aagtatggcg tgaagcgtcc taaggcttag gttaataaca 900

ggcctgctgg taatcgcagg ccttttttatt tttacacctg cgttttagag ctagaaatag 960

caagttaaaa taaggctagt ccgttatcaa cttgaaaaag tggcaccgag tcggtgcttt 1020

tttttttgtt ttttatcgat gcgcgcac 1048

<210> 96

<211> 300

<212> DNA

<213> Yarrowia lipolytica

<400> 96

attttttttg attttctttt ttgaccccgt cttcaattac acttcccaac tgggaacacc 60

cctctttatc gacccatttt aggtaattta ccctagccca ttgtctccat aaggaatatt 120

accctaaccc acagtccagg gtgcccaggt ccttctttgg ccaaatttta acttcggtcc 180

tatggcacag cggtagcgcg tgagattgca aatcttaagg tcccgagttc gaatctcggt 240

gggacctagt tatttttgat agataatttc gtgatgatta gaaacttaac gcaaaataat 300

<210> 97

<211> 544

<212> DNA

<213> Escherichia coli

<220><221> misc_feature

<222> (1)..(544)

<223> rpsL counterselection cassette

<400> 97

atggcgggat cgttgtatat ttcttgacac cttttcggca tcgccctaaa ttcggcgtcc 60

tcatattgtg tgaggacgtt ttattacgtg tttacgaagc aaaagctaaa accaggagct 120

atttaatggc aacagttaac cagctggtac gcaaaccacg tgctcgcaaa gttgcgaaaa 180

gcaacgtgcc tgcgctggaa gcatgcccgc aaaaacgtgg cgtatgtact cgtgtatata 240

ctaccactcc taaaaaaccg aactccgcgc tgcgtaaagt atgccgtgtt cgtctgacta 300

acggtttcga agtgacttcc tacatcggtg gtgaaggtca caacctgcag gagcactccg 360

tgatcctgat ccgtggcggt cgtgttaaag acctcccggg tgttcgttac cacaccgtac 420

gtggtgcgct tgactgctcc ggcgttaaag accgtaagca ggctcgttcc aagtatggcg 480

tgaagcgtcc taaggcttag gttaataaca ggcctgctgg taatcgcagg ccttttttatt 540

ttta 544

<210> 98

<211> 11714

<212> DNA

<213> Artificial Sequence

<220><223> pRF291

<400> 98

cgataaaaaa caaaaaaaaa agcaccgact cggtgccact ttttcaagtt gataacggac 60

tagccttatt ttaacttgct atttctagct ctaaaacgca ggtgtaaaaa taaaaaggcc 120

tgcgattacc agcaggcctg ttattaacct aagccttagg acgcttcacg ccatacttgg 180

aacgagcctg cttacggtct ttaacgccgg agcagtcaag cgcaccacgt acggtgtggt 240

aacgaacacc cgggaggtct ttaacacgac cgccacggat caggatcacg gagtgctcct 300

gcaggttgtg accttcacca ccgatgtagg aagtcacttc gaaaccgtta gtcagacgaa 360

cacggcatac tttacgcagc gcggagttcg gttttttagg agtggtagta tatacacgag 420

tacatacgcc acgtttttgc gggcatgctt ccagcgcagg cacgttgctt ttcgcaactt 480

tgcgagcacg tggtttgcgt accagctggt taactgttgc cattaaatag ctcctggttt 540

tagcttttgc ttcgtaaaca cgtaataaaa cgtcctcaca caatatgagg acgccgaatt 600

tagggcgatg ccgaaaaggt gtcaagaaat atacaacgat cccgccatca cctgcgtccc 660

attcgccatg ccgaagcatg ttgcccagcc ggcgccagcg aggaggctgg gaccatgccg 720

gccattattt tgcgttaagt ttctaatcat cacgaaatta tctatcaaaa ataactaggt 780

cccaccgaga ttcgaactcg ggaccttaag atttgcaatc tcacgcgcta ccgctgtgcc 840

ataggaccga agttaaaatt tggccaaaga aggacctggg caccctggac tgtgggttag 900

ggtaatattc cttatggaga caatgggcta gggtaaatta cctaaaatgg gtcgataaag 960

aggggtgttc ccagttggga agtgtaattg aagacggggt caaaaaagaa aatcaaaaaa 1020

aatttaatta agtcatacac aagtcagctt tcttcgagcc tcatataagt ataagtagtt 1080

caacgtatta gcactgtacc cagcatctcc gtatcgagaa acacaacaac atgccccatt 1140

ggacagatca tgcggataca caggttgtgc agtatcatac atactcgatc agacaggtcg 1200

tctgaccatc atacaagctg aacaagcgct ccatacttgc acgctctcta tatacacagt 1260

taaattacat atccatagtc taacctctaa cagttaatct tctggtaagc ctcccagcca 1320

gccttctggt atcgcttggc ctcctcaata ggatctcggt tctggccgta cagacctcgg 1380

ccgacaatta tgatatccgt tccggtagac atgacatcct caacagttcg gtactgctgt 1440

ccgagagcgt ctcccttgtc gtcaagaccc accccggggg tcagaataag ccagtcctca 1500

gagtcgccct taggtcggtt ctgggcaatg aagccaacca caaactcggg gtcggatcgg 1560

gcaagctcaa tggtctgctt ggagtactcg ccagtggcca gagagccctt gcaagacagc 1620

tcggccagca tgagcagacc tctggccagc ttctcgttgg gagaggggac taggaactcc 1680

ttgtactggg agttctcgta gtcagagacg tcctccttct tctgttcaga gacagtttcc 1740

tcggcaccag ctcgcaggcc agcaatgatt ccggttccgg gtacaccgtg ggcgttggtg 1800

atatcggacc actcggcgat tcggtgacac cggtactggt gcttgacagt gttgccaata 1860

tctgcgaact ttctgtcctc gaacaggaag aaaccgtgct taagagcaag ttccttgagg 1920

gggagcacag tgccggcgta ggtgaagtcg tcaatgatgt cgatatgggt tttgatcatg 1980

cacacataag gtccgacctt atcggcaagc tcaatgagct ccttggtggt ggtaacatcc 2040

agagaagcac acaggttggt tttcttggct gccacgagct tgagcactcg agcggcaaag 2100

gcggacttgt ggacgttagc tcgagcttcg taggagggca ttttggtggt gaagaggaga 2160

ctgaaataaa tttagtctgc agaacttttt atcggaacct tatctggggc agtgaagtat 2220

atgttatggt aatagttacg agttagttga acttatagat agactggact atacggctat 2280

cggtccaaat tagaaagaac gtcaatggct ctctgggcgt cgcctttgcc gacaaaaatg 2340

tgatcatgat gaaagccagc aatgacgttg cagctgatat tgttgtcggc caaccgcgcc 2400

gaaaacgcag ctgtcagacc cacagcctcc aacgaagaat gtatcgtcaa agtgatccaa 2460

gcacactcat agttggagtc gtactccaaa ggcggcaatg acgagtcaga cagatactcg 2520

tcgacgttta aaccatcatc taagggcctc aaaactacct cggaactgct gcgctgatct 2580

ggacaccaca gaggttccga gcactttagg ttgcaccaaa tgtcccacca ggtgcaggca 2640

gaaaacgctg gaacagcgtg tacagtttgt cttaacaaaa agtgagggcg ctgaggtcga 2700

gcagggtggt gtgacttgtt atagccttta gagctgcgaa agcgcgtatg gatttggctc 2760

atcaggccag attgagggtc tgtggacaca tgtcatgtta gtgtacttca atcgcccccc 2820

ggatatagcc ccgacaatag gccgtggcct catttttttg ccttccgcac atttccattg 2880

ctcggtaccc acaccttgct tctcctgcac ttgccaacct taatactggt ttacattgac 2940

caacatctta caagcggggg gcttgtctag ggtatatata aacagtggct ctcccaatcg 3000

gttgccagtc tctttttcc tttctttccc cacagattcg aaatctaaac tacacatcac 3060

accatggaca agaaatactc catcggcctg gacattggaa ccaactctgt cggctgggct 3120

gtcatcaccg acgagtacaa ggtgccctcc aagaaattca aggtcctcgg aaacaccgat 3180

cgacactcca tcaagaaaaa cctcattggt gccctgttgt tcgattctgg cgagactgcc 3240

gaagctacca gactcaagcg aactgctcgg cgacgttaca cccgacggaa gaaccgaatc 3300

tgctacctgc aggagatctt ttccaacgag atggccaagg tggacgattc gttctttcat 3360

cgactggagg aatccttcct cgtcgaggaa gacaagaaac acgagcgtca tcccatcttt 3420

ggcaacattg tggacgaggt tgcttaccac gagaagtatc ctaccatcta ccatctccga 3480

aagaaactcg tcgattccac cgacaaggcg gatctcagac ttatctacct cgctctggca 3540

cacatgatca agtttcgagg tcatttcctc atcgagggcg atctcaatcc cgacaacagc 3600

gatgtggaca agctgttcat tcagctcgtt cagacctaca accagctgtt cgaggaaaac 3660

cccatcaatg cctccggagt cgatgcaaag gccatcttgt ctgctcgact ctcgaagagc 3720

agacgactgg agaacctcat tgcccaactt cctggcgaga aaaagaacgg actgtttggc 3780

aacctcattg ccctttctct tggtctcaca cccaacttca agtccaactt cgatctggcg 3840

gaggacgcca agctccagct gtccaaggac acctacgacg atgacctcga caacctgctt 3900

gcacagattg gcgatcagta cgccgacctg tttctcgctg ccaagaacct ttcggatgct 3960

attctcttgt ctgacattct gcgagtcaac accgagatca caaaggctcc cctttctgcc 4020

tccatgatca agcgatacga cgagcaccat caggatctca cactgctcaa ggctcttgtc 4080

cgacagcaac tgcccgagaa gtacaaggag atctttttcg atcagtcgaa gaacggctac 4140

gctggataca tcgacggcgg agcctctcag gaagagttct acaagttcat caagccaatt 4200

ctcgagaaga tggacggaac cgaggaactg cttgtcaagc tcaatcgaga ggatctgctt 4260

cggaagcaac gaaccttcga caacggcagc attcctcatc agatccacct cggtgagctg 4320

cacgccattc ttcgacgtca ggaagacttc tacccctttc tcaaggacaa ccgagagaag 4380

atcgagaaga ttcttacctt tcgaatcccc tactatgttg gtcctcttgc cagaggaaac 4440

tctcgatttg cttggatgac tcgaaagtcc gaggaaacca tcactccctg gaacttcgag 4500

gaagtcgtgg acaagggtgc ctctgcacag tccttcatcg agcgaatgac caacttcgac 4560

aagaatctgc ccaacgagaa ggttcttccc aagcattcgc tgctctacga gtactttaca 4620

gtctacaacg aactcaccaa agtcaagtac gttaccgagg gaatgcgaaa gcctgccttc 4680

ttgtctggcg aacagaagaa agccattgtc gatctcctgt tcaagaccaa ccgaaaggtc 4740

actgttaagc agctcaagga ggactacttc aagaaaatcg agtgtttcga cagcgtcgag 4800

atttccggag ttgaggaccg attcaacgcc tctttgggca cctatcacga tctgctcaag 4860

attatcaagg acaaggattt tctcgacaac gaggaaaacg aggacattct ggaggacatc 4920

gtgctcactc ttaccctgtt cgaagatcgg gagatgatcg aggaacgact caagacatac 4980

gctcacctgt tcgacgacaa ggtcatgaaa caactcaagc gacgtagata caccggctgg 5040

ggaagacttt cgcgaaagct catcaacggc atcagagaca agcagtccgg aaagaccatt 5100

ctggactttc tcaagtccga tggctttgcc aaccgaaact tcatgcagct cattcacgac 5160

gattctctta ccttcaagga ggacatccag aaggcacaag tgtccggtca gggcgacagc 5220

ttgcacgaac atattgccaa cctggctggt tcgccagcca tcaagaaagg cattctccag 5280

actgtcaagg ttgtcgacga gctggtgaag gtcatgggac gtcacaagcc cgagaacatt 5340

gtgatcgaga tggccagaga gaaccagaca actcaaaagg gtcagaaaaa ctcgcgagag 5400

cggatgaagc gaatcgagga aggcatcaag gagctgggat cccagattct caaggagcat 5460

cccgtcgaga acactcaact gcagaacgag aagctgtatc tctactatct gcagaatggt 5520

cgagacatgt acgtggatca ggaactggac atcaatcgtc tcagcgacta cgatgtggac 5580

cacattgtcc ctcaatcctt tctcaaggac gattctatcg acaacaaggt ccttacacga 5640

tccgacaaga acagaggcaa gtcggacaac gttcccagcg aagaggtggt caaaaagatg 5700

aagaactact ggcgacagct gctcaacgcc aagctcatta cccagcgaaa gttcgacaat 5760
cttaccaagg ccgagcgagg cggtctgtcc gagctcgaca aggctggctt catcaagcgt 5820
caactcgtcg agaccagaca gatcacaaag cacgtcgcac agattctcga ttctcggatg 5880
aacaccaagt acgacgagaa cgacaagctc atccgagagg tcaaggtgat tactctcaag 5940
tccaaactgg tctccgattt ccgaaaggac tttcagttct acaaggtgcg agagatcaac 6000
aattaccacc atgcccacga tgcttacctc aacgccgtcg ttggcactgc gctcatcaag 6060
aaatacccca agctcgaaag cgagttcgtt tacggcgatt acaaggtcta cgacgttcga 6120

aagatgattg ccaagtccga acaggagatt ggcaaggcta ctgccaagta cttcttttac 6180
tccaacatca tgaacttttt caagaccgag atcaccttgg ccaacggaga gattcgaaag 6240
agaccactta tcgagaccaa cggcgaaact ggagagatcg tgtgggacaa gggtcgagac 6300
tttgcaaccg tgcgaaaggt tctgtcgatg cctcaggtca acatcgtcaa gaaaaccgag 6360
gttcagactg gcggattctc caaggagtcg attctgccca agcgaaactc cgacaagctc 6420
atcgctcgaa agaaagactg ggatcccaag aaatacggtg gcttcgattc tcctaccgtc 6480
gcctattccg tgcttgtcgt tgcgaaggtc gagaagggca agtccaaaaa gctcaagtcc 6540

gtcaaggagc tgctcggaat taccatcatg gagcgatcga gcttcgagaa gaatcccatc 6600
gacttcttgg aagccaaggg ttacaaggag gtcaagaaag acctcattat caagctgccc 6660
aagtactctc tgttcgaact ggagaacggt cgaaagcgta tgctcgcctc cgctggcgag 6720
ctgcagaagg gaaacgagct tgccttgcct tcgaagtacg tcaactttct ctatctggct 6780
tctcactacg agaagctcaa gggttctccc gaggacaacg aacagaagca actcttcgtt 6840
gagcagcaca aacattacct cgacgagatt atcgagcaga tttccgagtt ttcgaagcga 6900
gtcatcctgg ctgatgccaa cttggacaag gtgctctctg cctacaacaa gcatcgggac 6960

aaacccattc gagaacaggc ggagaacatc attcacctgt ttactcttac caacctgggt 7020
gctcctgcag ctttcaagta cttcgatacc actatcgacc gaaagcggta cacatccacc 7080
aaggaggttc tcgatgccac cctgattcac cagtccatca ctggcctgta cgagacccga 7140
atcgacctgt ctcagcttgg tggcgactcc agagccgatc ccaagaaaaa gcgaaaggtc 7200
taagcggccg caagtgtgga tggggaagtg agtgcccggt tctgtgtgca caattggcaa 7260
tccaagatgg atggattcaa cacagggata tagcgagcta cgtggtggtg cgaggatata 7320
gcaacggata tttatgtttg acacttgaga atgtacgata caagcactgt ccaagtacaa 7380

tactaaacat actgtacata ctcatactcg tacccgggca acggtttcac ttgagtgcag 7440

tggctagtgc tcttactcgt acagtgtgca atactgcgta tcatagtctt tgatgtatat 7500

cgtattcatt catgttagtt gcgtacgagc cggaagcata aagtgtaaag cctggggtgc 7560

ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgctt tccagtcggg 7620

aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag gcggtttgcg 7680

tattgggcgc tcttccgctt cctcgctcac tgactcgctg cgctcggtcg ttcggctgcg 7740

gcgagcggta tcagctcact caaaggcggt aatacggtta tccacagaat caggggataa 7800

cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc 7860

gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa atcgacgctc 7920

aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag 7980

ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt ccgcctttct 8040

cccttcggga agcgtggcgc tttctcatag ctcacgctgt aggtatctca gttcggtgta 8100

ggtcgttcgc tccaagctgg gctgtgtgca cgaacccccc gttcagcccg accgctgcgc 8160

cttatccggt aactatcgtc ttgagtccaa cccggtaaga cacgacttat cgccactggc 8220

agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagttctt 8280

gaagtggtgg cctaactacg gctacactag aaggacagta tttggtatct gcgctctgct 8340

gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc 8400

tggtagcggt ggtttttttg tttgcaagca gcagattacg cgcagaaaaa aaggatctca 8460

agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa actcacgtta 8520

agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa 8580

atgaagtttt aaatcaatct aaagtatata tgagtaaact tggtctgaca gttaccaatg 8640

cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca tagttgcctg 8700

actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc ccagtgctgc 8760

aatgataccg cgagacccac gctcaccggc tccagattta tcagcaataa accagccagc 8820

cggaagggcc gagcgcagaa gtggtcctgc aactttatcc gcctccatcc agtctattaa 8880

ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca acgttgttgc 8940

cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat tcagctccgg 9000

ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag cggttagctc 9060

cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac tcatggttat 9120

ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt ctgtgactgg 9180

tgagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt gctcttgccc 9240

ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc tcatcattgg 9300

aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat ccagttcgat 9360

gtaacccact cgtgcaccca actgatcttc agcatctttt actttcacca gcgtttctgg 9420

gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataagggcga cacggaaatg 9480

ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg gttattgtct 9540

catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg ttccgcgcac 9600

atttccccga aaagtgccac ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt 9660

ggtggttacg cgcagcgtga ccgctacact tgccagcgcc ctagcgcccg ctcctttcgc 9720

tttcttccct tcctttctcg ccacgttcgc cggctttccc cgtcaagctc taaatcgggg 9780

gctcccttta gggttccgat ttagtgcttt acggcacctc gaccccaaaa aacttgatta 9840

gggtgatggt tcacgtagtg ggccatcgcc ctgatagacg gtttttcgcc ctttgacgtt 9900

ggagtccacg ttctttaata gtggactctt gttccaaact ggaacaacac tcaaccctat 9960

ctcggtctat tcttttgatt tataagggat tttgccgatt tcggcctatt ggttaaaaaa 10020

tgagctgatt taacaaaaat ttaacgcgaa ttttaacaaa atattaacgc ttacaatttc 10080

cattcgccat tcaggctgcg caactgttgg gaagggcgat cggtgcgggc ctcttcgcta 10140

ttacgccagc tggcgaaagg gggatgtgct gcaaggcgat taagttgggt aacgccaggg 10200

ttttcccagt cacgacgttg taaaacgacg gccagtgaat tgtaatacga ctcactatag 10260

ggcgaattgg gtaccgggcc ccccctcgag gtcgatggtg tcgataagct tgatatcgaa 10320

ttcatgtcac acaaaccgat cttcgcctca aggaaaccta attctacatc cgagagactg 10380

ccgagatcca gtctacactg attaattttc gggccaataa tttaaaaaaa tcgtgttata 10440

taatattata tgtattatat atatacatca tgatgatact gacagtcatg tcccattgct 10500

aaatagacag actccatctg ccgcctccaa ctgatgttct caatatttaa ggggtcatct 10560

cgcattgttt aataataaac agactccatc taccgcctcc aaatgatgtt ctcaaaatat 10620

attgtatgaa cttattttta ttacttagta ttattagaca acttacttgc tttatgaaaa 10680

acacttccta tttaggaaac aatttataat ggcagttcgt tcatttaaca atttatgtag 10740

aataaatgtt ataaatgcgt atgggaaatc ttaaatatgg atagcataaa tgatatctgc 10800

attgcctaat tcgaaatcaa cagcaacgaa aaaaatccct tgtacaacat aaatagtcat 10860

cgagaaatat caactatcaa agaacagcta ttcacacgtt actattgaga ttattattgg 10920

acgagaatca cacactcaac tgtctttctc tcttctagaa atacaggtac aagtatgtac 10980

tattctcatt gttcatactt ctagtcattt catcccacat attccttgga tttctctcca 11040

atgaatgaca ttctatcttg caaattcaac aattataata agatatacca aagtagcggt 11100

atagtggcaa tcaaaaagct tctctggtgt gcttctcgta tttatttta ttctaatgat 11160

ccattaaagg tatatattta tttcttgtta tataatcctt ttgtttatta catgggctgg 11220

atacataaag gtattttgat ttaatttttt gcttaaattc aatccccct cgttcagtgt 11280

caactgtaat ggtaggaaat taccatactt ttgaagaagc aaaaaaaatg aaagaaaaaa 11340

aaaatcgtat ttccaggtta gacgttccgc agaatctaga atgcggtatg cggtacattg 11400

ttcttcgaac gtaaaagttg cgctccctga gatattgtac atttttgctt ttacaagtac 11460

aagtacatcg tacaactatg tactactgtt gatgcatcca caacagtttg ttttgttttt 11520

ttttgttttt ttttttttcta atgattcatt accgctatgt atacctactt gtacttgtag 11580

taagccgggt tattggcgtt caattaatca tagacttatg aatctgcacg gtgtgcgctg 11640

cgagttactt ttagcttatg catgctactt gggtgtaata ttgggatctg ttcggaaatc 11700

aacggatgct caat 11714

<210> 99

<211> 32

<212> DNA

<213> Artificial sequence

<220><223> Can1-1F

<400> 99

aatgggactc aaacgattac ccaccctcgt tt 32

<210> 100

<211> 32

<212> DNA

<213> Artificial sequence

<220><223> Can1-1R

<400> 100

tctaaaacga gggtgggtaa tcgtttgagt cc 32

<210> 101

<211> 23

<212> DNA

<213> Yarrowia lipolytica

<220><221> misc_feature

<222> (1)..(23)

<223> Can1-1 target site and PAM

<400> 101

tcaaacgatt acccaccctc cgg 23

<210> 102

<211> 480

<212> DNA

<213> Artificial sequence

<220><223> Can1-1 gRNA expression cassette

<400> 102

attttttttg attttctttt ttgaccccgt cttcaattac acttcccaac tgggaacacc 60

cctctttatc gacccatttt aggtaattta ccctagccca ttgtctccat aaggaatatt 120

accctaaccc acagtccagg gtgcccaggt ccttctttgg ccaaatttta acttcggtcc 180

tatggcacag cggtagcgcg tgagattgca aatcttaagg tcccgagttc gaatctcggt 240

gggacctagt tatttttgat agataatttc gtgatgatta gaaacttaac gcaaaataat 300

ggccggcatg gtcccagcct cctcgctggc gccggctggg caacatgctt cggcatggcg 360

aatgggactc aaacgattac ccaccctcgt tttagagcta gaaatagcaa ttaaaataag 420

gctagtccgt tatcaacttg aaaaagtggc accgagtcgg tgctttttttt tttgtttttt 480

<210> 103

<211> 11176

<212> DNA

<213> Artificial sequence

<220><223> pRF303

<400> 103

tctaaaacga gggtgggtaa tcgtttgagt cccattcgcc atgccgaagc atgttgccca 60

gccggcgcca gcgaggaggc tgggaccatg ccggccatta ttttgcgtta agtttctaat 120

catcacgaaa ttatctatca aaaataacta ggtcccaccg agattcgaac tcgggacctt 180

aagatttgca atctcacgcg ctaccgctgt gccataggac cgaagttaaa atttggccaa 240

agaaggacct gggcaccctg gactgtgggt tagggtaata ttccttatgg agacaatggg 300

ctagggtaaa ttacctaaaa tgggtcgata aagaggggtg ttcccagttg ggaagtgtaa 360

ttgaagacgg ggtcaaaaaa gaaaatcaaa aaaaatttaa ttaagtcata cacaagtcag 420

ctttcttcga gcctcatata agtataagta gttcaacgta ttagcactgt acccagcatc 480

tccgtatcga gaaacacaac aacatgcccc attggacaga tcatgcggat acacaggttg 540

tgcagtatca tacatactcg atcagacagg tcgtctgacc atcatacaag ctgaacaagc 600

gctccatact tgcacgctct ctatatacac agttaaatta catatccata gtctaacctc 660

taacagttaa tcttctggta agcctcccag ccagccttct ggtatcgctt ggcctcctca 720

ataggatctc ggttctggcc gtacagacct cggccgacaa ttatgatatc cgttccggta 780

gacatgacat cctcaacagt tcggtactgc tgtccgagag cgtctccctt gtcgtcaaga 840

cccaccccgg gggtcagaat aagccagtcc tcagagtcgc ccttaggtcg gttctgggca 900

atgaagccaa ccacaaactc ggggtcggat cgggcaagct caatggtctg cttggagtac 960

tcgccagtgg ccagagagcc cttgcaagac agctcggcca gcatgagcag acctctggcc 1020

agcttctcgt tgggagaggg gactaggaac tccttgtact gggagttctc gtagtcagag 1080

acgtcctcct tcttctgttc agagacagtt tcctcggcac cagctcgcag gccagcaatg 1140

attccggttc cgggtacacc gtgggcgttg gtgatatcgg accactcggc gattcggtga 1200

caccggtact ggtgcttgac agtgttgcca atatctgcga actttctgtc ctcgaacagg 1260

aagaaaccgt gcttaagagc aagttccttg agggggagca cagtgccggc gtaggtgaag 1320

tcgtcaatga tgtcgatatg ggttttgatc atgcacacat aaggtccgac cttatcggca 1380

agctcaatga gctccttggt ggtggtaaca tccagagaag cacacaggtt ggttttcttg 1440

gctgccacga gcttgagcac tcgagcggca aaggcggact tgtggacgtt agctcgagct 1500

tcgtaggagg gcattttggt ggtgaagagg agactgaaat aaatttagtc tgcagaactt 1560

tttatcggaa ccttatctgg ggcagtgaag tatatgttat ggtaatagtt acgagttagt 1620

tgaacttata gatagactgg actatacggc tatcggtcca aattagaaag aacgtcaatg 1680

gctctctggg cgtcgccttt gccgacaaaa atgtgatcat gatgaaagcc agcaatgacg 1740

ttgcagctga tattgttgtc ggccaaccgc gccgaaaacg cagctgtcag acccacagcc 1800

tccaacgaag aatgtatcgt caaagtgatc caagcacact catagttgga gtcgtactcc 1860

aaaggcggca atgacgagtc agacagatac tcgtcgacgt ttaaaccatc atctaagggc 1920

ctcaaaacta cctcggaact gctgcgctga tctggacacc acagaggttc cgagcacttt 1980

aggttgcacc aaatgtccca ccaggtgcag gcagaaaacg ctggaacagc gtgtacagtt 2040

tgtcttaaca aaaagtgagg gcgctgaggt cgagcagggt ggtgtgactt gttatagcct 2100

ttagagctgc gaaagcgcgt atggatttgg ctcatcaggc cagattgagg gtctgtggac 2160

acatgtcatg ttagtgtact tcaatcgccc cctggatata gccccgacaa taggccgtgg 2220

cctcattttt ttgccttccg cacatttcca ttgctcggta cccacacctt gcttctcctg 2280

cacttgccaa ccttaatact ggtttacatt gaccaacatc ttacaagcgg ggggcttgtc 2340

tagggtatat ataaacagtg gctctcccaa tcggttgcca gtctcttttt tcctttcttt 2400

ccccacagat tcgaaatcta aactacacat cacaccatgg acaagaaata ctccatcggc 2460

ctggacattg gaaccaactc tgtcggctgg gctgtcatca ccgacgagta caaggtgccc 2520

tccaagaaat tcaaggtcct cggaaacacc gatcgacact ccatcaagaa aaacctcatt 2580
ggtgccctgt tgttcgattc tggcgagact gccgaagcta ccagactcaa gcgaactgct 2640
cggcgacgtt acacccgacg gaagaaccga atctgctacc tgcaggagat cttttccaac 2700
gagatggcca aggtggacga ttcgttcttt catcgactgg aggaatcctt cctcgtcgag 2760
gaagacaaga aacacgagcg tcatcccatc tttggcaaca ttgtggacga ggttgcttac 2820
cacgagaagt atcctaccat ctaccatctc cgaaagaaac tcgtcgattc caccgacaag 2880
gcggatctca gacttatcta cctcgctctg gcacacatga tcaagtttcg aggtcatttc 2940

ctcatcgagg gcgatctcaa tcccgacaac agcgatgtgg acaagctgtt cattcagctc 3000
gttcagacct acaaccagct gttcgaggaa aaccccatca atgcctccgg agtcgatgca 3060
aaggccatct tgtctgctcg actctcgaag agcagacgac tggagaacct cattgcccaa 3120
cttcctggcg agaaaaagaa cggactgttt ggcaacctca ttgccctttc tcttggtctc 3180
acacccaact tcaagtccaa cttcgatctg gcggaggacg ccaagctcca gctgtccaag 3240
gacacctacg acgatgacct cgacaacctg cttgcacaga ttggcgatca gtacgccgac 3300
ctgtttctcg ctgccaagaa cctttcggat gctattctct tgtctgacat tctgcgagtc 3360

aacaccgaga tcacaaaggc tcccctttct gcctccatga tcaagcgata cgacgagcac 3420
catcaggatc tcacactgct caaggctctt gtccgacagc aactgcccga gaagtacaag 3480
gagatctttt tcgatcagtc gaagaacggc tacgctggat acatcgacgg cggagcctct 3540
caggaagagt tctacaagtt catcaagcca attctcgaga agatggacgg aaccgaggaa 3600
ctgcttgtca agctcaatcg agaggatctg cttcggaagc aacgaacctt cgacaacggc 3660
agcattcctc atcagatcca cctcggtgag ctgcacgcca ttcttcgacg tcaggaagac 3720
ttctacccct ttctcaagga caaccgagag aagatcgaga agattcttac ctttcgaatc 3780

ccctactatg ttggtcctct tgccagagga aactctcgat ttgcttggat gactcgaaag 3840
tccgaggaaa ccatcactcc ctggaacttc gaggaagtcg tggacaaggg tgcctctgca 3900
cagtccttca tcgagcgaat gaccaacttc gacaagaatc tgcccaacga gaaggttctt 3960
cccaagcatt cgctgctcta cgagtacttt acagtctaca acgaactcac caaagtcaag 4020
tacgttaccg agggaatgcg aaagcctgcc ttcttgtctg gcgaacagaa gaaagccatt 4080
gtcgatctcc tgttcaagac caaccgaaag gtcactgtta agcagctcaa ggaggactac 4140
ttcaagaaaa tcgagtgttt cgacagcgtc gagatttccg gagttgagga ccgattcaac 4200

gcctctttgg gcacctatca cgatctgctc aagattatca aggacaagga ttttctcgac 4260

aacgaggaaa acgaggacat tctggaggac atcgtgctca ctcttaccct gttcgaagat 4320

cgggagatga tcgaggaacg actcaagaca tacgctcacc tgttcgacga caaggtcatg 4380

aaacaactca agcgacgtag atacaccggc tggggaagac tttcgcgaaa gctcatcaac 4440

ggcatcagag acaagcagtc cggaaagacc attctggact ttctcaagtc cgatggcttt 4500

gccaaccgaa acttcatgca gctcattcac gacgattctc ttaccttcaa ggaggacatc 4560

cagaaggcac aagtgtccgg tcagggcgac agcttgcacg aacatattgc caacctggct 4620

ggttcgccag ccatcaagaa aggcattctc cagactgtca aggttgtcga cgagctggtg 4680

aaggtcatgg gacgtcacaa gcccgagaac attgtgatcg agatggccag agagaaccag 4740

acaactcaaa agggtcagaa aaactcgcga gagcggatga agcgaatcga ggaaggcatc 4800

aaggagctgg gatcccagat tctcaaggag catcccgtcg agaacactca actgcagaac 4860

gagaagctgt atctctacta tctgcagaat ggtcgagaca tgtacgtgga tcaggaactg 4920

gacatcaatc gtctcagcga ctacgatgtg gaccacattg tccctcaatc ctttctcaag 4980

gacgattcta tcgacaacaa ggtccttaca cgatccgaca agaacagagg caagtcggac 5040

aacgttccca gcgaagaggt ggtcaaaaag atgaagaact actggcgaca gctgctcaac 5100

gccaagctca ttacccagcg aaagttcgac aatcttacca aggccgagcg aggcggtctg 5160

tccgagctcg acaaggctgg cttcatcaag cgtcaactcg tcgagaccag acagatcaca 5220

aagcacgtcg cacagattct cgattctcgg atgaacacca agtacgacga gaacgacaag 5280

ctcatccgag aggtcaaggt gattactctc aagtccaaac tggtctccga tttccgaaag 5340

gactttcagt tctacaaggt gcgagagatc aacaattacc accatgccca cgatgcttac 5400

ctcaacgccg tcgttggcac tgcgctcatc aagaaatacc ccaagctcga aagcgagttc 5460

gtttacggcg attacaaggt ctacgacgtt cgaaagatga ttgccaagtc cgaacaggag 5520

attggcaagg ctactgccaa gtacttcttt tactccaaca tcatgaactt tttcaagacc 5580

gagatcacct tggccaacgg agagattcga aagagaccac ttatcgagac caacggcgaa 5640

actggagaga tcgtgtggga caagggtcga gactttgcaa ccgtgcgaaa ggttctgtcg 5700

atgcctcagg tcaacatcgt caagaaaacc gaggttcaga ctggcggatt ctccaaggag 5760

tcgattctgc ccaagcgaaa ctccgacaag ctcatcgctc gaaagaaaga ctgggatccc 5820

aagaaatacg gtggcttcga ttctcctacc gtcgcctatt ccgtgcttgt cgttgcgaag 5880

gtcgagaagg gcaagtccaa aaagctcaag tccgtcaagg agctgctcgg aattaccatc 5940

atggagcgat cgagcttcga gaagaatccc atcgacttct tggaagccaa gggttacaag 6000

gaggtcaaga aagacctcat tatcaagctg cccaagtact ctctgttcga actggagaac 6060

ggtcgaaagc gtatgctcgc ctccgctggc gagctgcaga agggaaacga gcttgccttg 6120

ccttcgaagt acgtcaactt tctctatctg gcttctcact acgagaagct caagggttct 6180

cccgaggaca acgaacagaa gcaactcttc gttgagcagc acaaacatta cctcgacgag 6240

attatcgagc agatttccga gttttcgaag cgagtcatcc tggctgatgc caacttggac 6300

aaggtgctct ctgcctacaa caagcatcgg gacaaaccca ttcgagaaca ggcggagaac 6360

atcattcacc tgtttactct taccaacctg ggtgctcctg cagctttcaa gtacttcgat 6420

accactatcg accgaaagcg gtacacatcc accaaggagg ttctcgatgc caccctgatt 6480

caccagtcca tcactggcct gtacgagacc cgaatcgacc tgtctcagct tggtggcgac 6540

tccagagccg atcccaagaa aaagcgaaag gtctaagcgg ccgcaagtgt ggatggggaa 6600

gtgagtgccc ggttctgtgt gcacaattgg caatccaaga tggatggatt caacacaggg 6660

atatagcgag ctacgtggtg gtgcgaggat atagcaacgg atatttatgt ttgacacttg 6720

agaatgtacg atacaagcac tgtccaagta caatactaaa catactgtac atactcatac 6780

tcgtacccgg gcaacggttt cacttgagtg cagtggctag tgctcttact cgtacagtgt 6840

gcaatactgc gtatcatagt ctttgatgta tatcgtattc attcatgtta gttgcgtacg 6900

agccggaagc ataaagtgta aagcctgggg tgcctaatga gtgagctaac tcacattaat 6960

tgcgttgcgc tcactgcccg ctttccagtc gggaaacctg tcgtgccagc tgcattaatg 7020

aatcggccaa cgcgcgggga gaggcggttt gcgtattggg cgctcttccg cttcctcgct 7080

cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc actcaaaggc 7140

ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg 7200

ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg 7260

ccccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg 7320

actataaaga taccaggcgt ttcccccctgg aagctccctc gtgcgctctc ctgttccgac 7380

cctgccgctt accggatacc tgtccgcctt tctcccttcg ggaagcgtgg cgctttctca 7440

tagctcacgc tgtaggtatc tcagttcggt gtaggtcgtt cgctccaagc tgggctgtgt 7500

gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc ggtaactatc gtcttgagtc 7560

caacccggta agacacgact tatcgccact ggcagcagcc actggtaaca ggattagcag 7620

agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact acggctacac 7680

tagaaggaca gtatttggta tctgcgctct gctgaagcca gttaccttcg gaaaaagagt 7740

tggtagctct tgatccggca aacaaaccac cgctggtagc ggtggttttt ttgtttgcaa 7800

gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cctttgatct tttctacggg 7860

gtctgacgct cagtggaacg aaaactcacg ttaagggatt ttggtcatga gattatcaaa 7920

aaggatcttc acctagatcc ttttaaatta aaaatgaagt tttaaatcaa tctaaagtat 7980

atatgagtaa acttggtctg acagttacca atgcttaatc agtgaggcac ctatctcagc 8040

gatctgtcta tttcgttcat ccatagttgc ctgactcccc gtcgtgtaga taactacgat 8100

acgggagggc ttaccatctg gccccagtgc tgcaatgata ccgcgagacc cacgctcacc 8160

ggctccagat ttatcagcaa taaaccagcc agccggaagg gccgagcgca gaagtggtcc 8220

tgcaacttta tccgcctcca tccagtctat taattgttgc cgggaagcta gagtaagtag 8280

ttcgccagtt aatagtttgc gcaacgttgt tgccattgct acaggcatcg tggtgtcacg 8340

ctcgtcgttt ggtatggctt cattcagctc cggttcccaa cgatcaaggc gagttacatg 8400

atcccccatg ttgtgcaaaa aagcggttag ctccttcggt cctccgatcg ttgtcagaag 8460

taagttggcc gcagtgttat cactcatggt tatggcagca ctgcataatt ctcttactgt 8520

catgccatcc gtaagatgct tttctgtgac tggtgagtac tcaaccaagt cattctgaga 8580

atagtgtatg cggcgaccga gttgctcttg cccggcgtca atacgggata ataccgcgcc 8640

acatagcaga actttaaaag tgctcatcat tggaaaacgt tcttcggggc gaaaactctc 8700

aaggatctta ccgctgttga gatccagttc gatgtaaccc actcgtgcac ccaactgatc 8760

ttcagcatct tttactttca ccagcgtttc tgggtgagca aaaacaggaa ggcaaaatgc 8820

cgcaaaaaag ggaataaggg cgacacggaa atgttgaata ctcatactct tccttttca 8880

atattattga agcatttatc agggttattg tctcatgagc ggatacatat ttgaatgtat 8940

ttagaaaaat aaacaaatag gggttccgcg cacatttccc cgaaaagtgc cacctgacgc 9000

gccctgtagc ggcgcattaa gcgcggcggg tgtggtggtt acgcgcagcg tgaccgctac 9060

acttgccagc gccctagcgc ccgctccttt cgctttcttc ccttcctttc tcgccacgtt 9120

cgccggcttt ccccgtcaag ctctaaatcg ggggctccct ttagggttcc gatttagtgc 9180

tttacggcac ctcgacccca aaaaacttga ttagggtgat ggttcacgta gtgggccatc 9240

gccctgatag acggtttttc gccctttgac gttggagtcc acgttcttta atagtggact 9300

cttgttccaa actggaacaa cactcaaccc tatctcggtc tattcttttg atttataagg 9360

gattttgccg atttcggcct attggttaaa aaatgagctg atttaacaaa aatttaacgc 9420

gaattttaac aaaatattaa cgcttacaat ttccattcgc cattcaggct gcgcaactgt 9480

tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa aggggatgt 9540

gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg 9600

acggccagtg aattgtaata cgactcacta tagggcgaat tgggtaccgg gccccccctc 9660

gaggtcgatg gtgtcgataa gcttgatatc gaattcatgt cacacaaacc gatcttcgcc 9720

tcaaggaaac ctaattctac atccgagaga ctgccgagat ccagtctaca ctgattaatt 9780

ttcgggccaa taatttaaaa aaatcgtgtt atataatatt atatgtatta tatatataca 9840

tcatgatgat actgacagtc atgtcccatt gctaaataga cagactccat ctgccgcctc 9900

caactgatgt tctcaatatt taaggggtca tctcgcattg tttaataata aacagactcc 9960

atctaccgcc tccaaatgat gttctcaaaa tatattgtat gaacttattt ttattactta 10020

gtattattag acaacttact tgctttatga aaaacacttc ctatttagga aacaatttat 10080

aatggcagtt cgttcattta acaatttatg tagaataaat gttataaatg cgtatgggaa 10140

atcttaaata tggatagcat aaatgatatc tgcattgcct aattcgaaat caacagcaac 10200

gaaaaaaatc ccttgtacaa cataaatagt catcgagaaa tatcaactat caaagaacag 10260

ctattcacac gttactattg agattattat tggacgagaa tcacacactc aactgtcttt 10320

ctctcttcta gaaatacagg tacaagtatg tactattctc attgttcata cttctagtca 10380

tttcatccca catattcctt ggatttctct ccaatgaatg acattctatc ttgcaaattc 10440

aacaattata ataagatata ccaaagtagc ggtatagtgg caatcaaaaa gcttctctgg 10500

tgtgcttctc gtatttattt ttattctaat gatccattaa aggtatatat ttatttcttg 10560

ttatataatc cttttgttta ttacatgggc tggatacata aaggtatttt gatttaattt 10620

tttgcttaaa ttcaatcccc cctcgttcag tgtcaactgt aatggtagga aattaccata 10680

cttttgaaga agcaaaaaaa atgaaagaaa aaaaaaatcg tatttccagg ttagacgttc 10740

cgcagaatct agaatgcggt atgcggtaca ttgttcttcg aacgtaaaag ttgcgctccc 10800

tgagatattg tacatttttg cttttacaag tacaagtaca tcgtacaact atgtactact 10860

gttgatgcat ccacaacagt ttgttttgtt ttttttttgt ttttttttct ctaatgattc 10920

attaccgcta tgtataccta cttgtacttg tagtaagccg ggttattggc gttcaattaa 10980

tcatagactt atgaatctgc acggtgtgcg ctgcgagtta cttttagctt atgcatgcta 11040

cttgggtgta atattgggat ctgttcggaa atcaacggat gctcaatcga taaaaaacaa 11100

aaaaaaaagc accgactcgg tgccactttt tcaagttgat aacggactag ccttatttta 11160

acttgctatt tctagc 11176

<210> 104

<211> 168

<212> RNA

<213> Artificial sequence

<220><223> ribozyme-guide RNA of Cas1-1

<400> 104

ggccggcaug gucccagccu ccucgcuggc gccggcuggg caacaugcuu cggcauggcg 60

aaugggacuc aaacgauuac ccacccucgu uuuagagcua gaaauagcaa guuaaaauaa 120

ggcuaguccg uuaucaacuu gaaaaagugg caccgagucg gugcuuuu 168

<210> 105

<211> 1719

<212> DNA

<213> Yarrowia lipolytica

<220><221> misc_feature

<222> (1)..(1719)

<223> Can1 gene

<400> 105

atggaaaaga cattttcaaa cgattaccca ccctccggga ctgaggccca catccacatc 60

aaccacacgg cccactcgga tgactcagag gaggtgccct cgcacaagga aaattacaac 120

accagtggcc acgacctgga ggagtccgac ccggataacc atgtcggtga gaccctcgag 180

gtcaagcgag gtctcaagat gcgacacatc tccatgatct cgcttggagg aaccattggt 240

accggtctct tcattggtac cggaggagct ctccagcagg ccggtccctg tggcgccctc 300

gtcgcctacg tgttcatggc caccattgtc tactctgttg ccgagtctct tggagaactg 360

gctacgtaca ttcccatcac cggctccttt gccgtcttta ctacccgata tctgtcacag 420

tcgtttggtg cctccatggg ctggctatac tggttctcgt gggcgatcac cttcgccatc 480

gagctcaaca ccattggtcc cgtgattgag tactggactg acgccgttcc tactgctgcc 540

tggattgcca tcttcttcgt catcctcact accatcaact tcttccccgt gggcttctat 600

ggcgaagtcg agttctgggt ggcctccgtg aaggtcattg ccatcattgg atggctcatc 660

tacgcgctct gcatgacgtg tggagcaggt gtaacaggtc ctgtgggatt cagatactgg 720

aaccaccccg gacccatggg agacggaatc tggaccgacg gcgtgcccat tgtgcgaaac 780

gcgcccggtc gacgattcat gggatggctc aattcgctcg ttaacgccgc cttcacctac 840

cagggctgtg agctggtcgg agtcactgcc ggtgaggccc agaaccccag aaagtccgtc 900

cctcgagcca tcaaccgagt ctttgctcga atttgcatct tctacattgg ctctatcttc 960

ttcatgggca tgctcgtgcc ctttaacgac cccaagctga ccgatgactc ctccgtcatc 1020

gcctcctctc cttttgttat tgccattatc aactctggca ccaaggtgct ccctcacatt 1080

ttcaacgccg tcattctcat caccctgatt tcggcaggaa actccaacgt ctacattggc 1140

tcgcgagtgg tctacgccct ggctgactcc ggaaccgcac caaagttctt caagcgaacc 1200

accaagaagg gagtgccgta cgtggcagtc tgcttcacct cggcgtttgg tctgctggcc 1260

ttcatgtctg tgtccgagtc gtcgtccact gtcttcgact ggttcatcaa catctccgct 1320

gtggccggcc tcatctgttg ggccttcatc tctgcctccc acatccgatt catgcaagtg 1380

cttaagcaca gagggatctc cagagatacg ctgcccttca aggcacgatg gcagccattc 1440

tactcatggt acgcgctcgt ctccatcatc ttcatcactc tcatccaggg cttcacgtcc 1500

ttctggcact ttaccgccgc caagttcatg actgcataca tctccgtcat tgtctgggtc 1560

ggtttgtaca ttatcttcca gtgtctgttc cgatgcaagt tccttatccc tattgaggat 1620

gtggacattg acaccggccg acgagagatt gacgacgatg tgtgggagga gaagatcccc 1680

acaaagtggt acgagaagtt ttggaatatt attgcataa 1719

<210> 106

<211> 23

<212> DNA

<213> Yarrowia lipolytica

<220><221> misc_feature

<222> (1)..(23)

<223> Can1-2 target site and PAM

<400> 106

ggcccactcg gatgactcag agg 23

<210> 107

<211> 27

<212> DNA

<213> Yarrowia lipolytica

<220><221> misc_feature

<222> (1)..(17)

<223> Sou2-1 target site and PAM

<400> 107

gtctggacct tccaccctcg ccacggg 27

<210> 108

<211> 25

<212> DNA

<213> Yarrowia lipolytica

<220><221> misc_feature

<222> (1)..(25)

<223> Sou2-2 target site and PAM

<400> 108

gcagtcccgt ggcgagggtg gaagg 25

<210> 109

<211> 20

<212> RNA

<213> Artificial sequence

<220><223> Variable targeting domain of Can1-2

<400> 109

ggcccacucg gaugacucag 20

<210> 110

<211> 24

<212> RNA

<213> artificial sequence

<220><223> Variable targeting domain of Sou2-1

<400> 110

gucuggaccu uccacccucg ccac 24

<210> 111

<211> 22

<212> RNA

<213> artificial sequence

<220><223> Variable targeting domain of Sou2-2

<400> 111

gcagucccgu ggcgagggug ga 22

<210> 112

<211> 23

<212> DNA

<213> Yarrowia lipolytica

<400> 112

cagctcgaga cgtcctagaa cgg 23

<210> 113

<211> 23

<212> DNA

<213> Yarrowia lipolytica

<400> 113

ttcctctgtc acagacgttt cgg 23

<210> 114

<211> 23

<212> DNA

<213> Yarrowia lipolytica

<400> 114

gaaaagtgcg ttttgattct cgg 23

<210> 115

<211> 19

<212> RNA

<213> artificial sequence

<220><223> Variable targeting domain of ura3-1

<400> 115

gccgcucgag ugcucaagc 19

<210> 116

<211> 22

<212> DNA

<213> Yarrowia lipolytica

<220><221> misc_feature

<222> (1)..(22)

<223> URa3-1 target site and PAM

<400> 116

gccgctcgag tgctcaagct cg 22

<210> 117

<211> 1379

<212> PRT

<213> Artificial sequence

<220><223> Cas9-SV40 NLS D10A H840A

<400> 117

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

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

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

245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

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

275 280 285

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

290 295 300

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

305 310 315 320

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

325 330 335

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

340 345 350

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

355 360 365

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

370 375 380

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

385 390 395 400

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

405 410 415

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

420 425 430

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

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

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

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

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

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

755 760 765

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

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

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

805 810 815

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

820 825 830

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

835 840 845

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

850 855 860

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

865 870 875 880

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

885 890 895

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

900 905 910

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

915 920 925

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

930 935 940

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

945 950 955 960

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

965 970 975

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

980 985 990

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

995 1000 1005

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

1010 1015 1020

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

1025 1030 1035

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

1040 1045 1050

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

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

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

1085 1090 1095

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

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

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

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

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

1235 1240 1245

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

1250 1255 1260

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

1265 1270 1275

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

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

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

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1325 1330 1335

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

1340 1345 1350

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

1355 1360 1365

Ser Arg Ala Asp Pro Lys Lys Lys Arg Lys Val

1370 1375

<210> 118

<211> 41

<212> DNA

<213> Artificial sequence

<220><223> D10AF

<400> 118

gaaatactcc atcggcctgg ccattggaac caactctgtc g 41

<210> 119

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> D10AR

<400> 119

cgacagagtt ggttccaatg gccaggccga tggagtattt c 41

<210> 120

<211> 4140

<212> DNA

<213> Artificial Sequence

<220><223> Yarrowia optimized Cas9 D10A

<400> 120

atggacaaga aatactccat cggcctggcc attggaacca actctgtcgg ctgggctgtc 60

atcaccgacg agtacaaggt gccctccaag aaattcaagg tcctcggaaa caccgatcga 120

cactccatca agaaaaacct cattggtgcc ctgttgttcg attctggcga gactgccgaa 180

gctaccagac tcaagcgaac tgctcggcga cgttacaccc gacggaagaa ccgaatctgc 240

tacctgcagg agatcttttc caacgagatg gccaaggtgg acgattcgtt ctttcatcga 300

ctggaggaat ccttcctcgt cgaggaagac aagaaacacg agcgtcatcc catctttggc 360

aacattgtgg acgaggttgc ttaccacgag aagtatccta ccatctacca cctgcgaaag 420

aaactcgtcg attccaccga caaggcggat ctcagactta tctacctcgc tctggcacac 480

atgatcaagt ttcgaggtca tttcctcatc gagggcgatc tcaatcccga caacagcgat 540

gtggacaagc tgttcattca gctcgttcag acctacaacc agctgttcga ggaaaacccc 600

atcaatgcct ccggagtcga tgcaaaggcc atcttgtctg ctcgactctc gaagagcaga 660

cgactggaga acctcattgc ccaacttcct ggcgagaaaa agaacggact gtttggcaac 720

ctcattgccc tttctcttgg tctcacaccc aacttcaagt ccaacttcga tctggcggag 780

gacgccaagc tccagctgtc caaggacacc tacgacgatg acctcgacaa cctgcttgca 840

cagattggcg atcagtacgc cgacctgttt ctcgctgcca agaacctttc ggatgctatt 900

ctcttgtctg acattctgcg agtcaacacc gagatcacaa aggctcccct ttctgcctcc 960

atgatcaagc gatacgacga gcaccatcag gatctcacac tgctcaaggc tcttgtccga 1020

cagcaactgc ccgagaagta caaggagatc tttttcgatc agtcgaagaa cggctacgct 1080

ggatacatcg acggcggagc ctctcaggaa gagttctaca agttcatcaa gccaattctc 1140

gagaagatgg acggaaccga ggaactgctt gtcaagctca atcgagagga tctgcttcgg 1200

aagcaacgaa ccttcgacaa cggcagcatt cctcatcaga tccacctcgg tgagctgcac 1260

gccattcttc gacgtcagga agacttctac ccctttctca aggacaaccg agagaagatc 1320

gagaagattc ttacctttcg aatcccctac tatgttggtc ctcttgccag aggaaactct 1380

cgatttgctt ggatgactcg aaagtccgag gaaaccatca ctccctggaa cttcgaggaa 1440

gtcgtggaca agggtgcctc tgcacagtcc ttcatcgagc gaatgaccaa cttcgacaag 1500

aatctgccca acgagaaggt tcttcccaag cattcgctgc tctacgagta ctttacagtc 1560

tacaacgaac tcaccaaagt caagtacgtt accgagggaa tgcgaaagcc tgccttcttg 1620

tctggcgaac agaagaaagc cattgtcgat ctcctgttca agaccaaccg aaaggtcact 1680

gttaagcagc tcaaggagga ctacttcaag aaaatcgagt gtttcgacag cgtcgagatt 1740

tccggagttg aggaccgatt caacgcctct ttgggcacct atcacgatct gctcaagatt 1800

atcaaggaca aggattttct cgacaacgag gaaaacgagg acattctgga ggacatcgtg 1860

ctcactctta ccctgttcga agatcgggag atgatcgagg aacgactcaa gacatacgct 1920

cacctgttcg acgacaaggt catgaaacaa ctcaagcgac gtagatacac cggctgggga 1980

agactttcgc gaaagctcat caacggcatc agagacaagc agtccggaaa gaccattctg 2040

gactttctca agtccgatgg ctttgccaac cgaaacttca tgcagctcat tcacgacgat 2100

tctcttacct tcaaggagga catccagaag gcacaagtgt ccggtcaggg cgacagcttg 2160

cacgaacata ttgccaacct ggctggttcg ccagccatca agaaaggcat tctccagact 2220

gtcaaggttg tcgacgagct ggtgaaggtc atgggacgtc acaagcccga gaacattgtg 2280

atcgagatgg ccagagagaa ccagacaact caaaagggtc agaaaaactc gcgagagcgg 2340

atgaagcgaa tcgaggaagg catcaaggag ctgggatccc agattctcaa ggagcatccc 2400

gtcgagaaca ctcaactgca gaacgagaag ctgtatctct actatctgca gaatggtcga 2460

gacatgtacg tggatcagga actggacatc aatcgtctca gcgactacga tgtggaccac 2520

attgtccctc aatcctttct caaggacgat tctatcgaca acaaggtcct tacacgatcc 2580

gacaagaaca gaggcaagtc ggacaacgtt cccagcgaag aggtggtcaa aaagatgaag 2640

aactactggc gacagctgct caacgccaag ctcattaccc agcgaaagtt cgacaatctt 2700

accaaggccg agcgaggcgg tctgtccgag ctcgacaagg ctggcttcat caagcgtcaa 2760

ctcgtcgaga ccagacagat cacaaagcac gtcgcacaga ttctcgattc tcggatgaac 2820

accaagtacg acgagaacga caagctcatc cgagaggtca aggtgattac tctcaagtcc 2880

aaactggtct ccgatttccg aaaggacttt cagttctaca aggtgcgaga gatcaacaat 2940

taccaccatg cccacgatgc ttacctcaac gccgtcgttg gcactgcgct catcaagaaa 3000

taccccaagc tcgaaagcga gttcgtttac ggcgattaca aggtctacga cgttcgaaag 3060

atgattgcca agtccgaaca ggagattggc aaggctactg ccaagtactt cttttactcc 3120

aacatcatga acttttcaa gaccgagatc accttggcca acggagagat tcgaaagaga 3180

ccacttatcg agaccaacgg cgaaactgga gagatcgtgt gggacaaggg tcgagacttt 3240

gcaaccgtgc gaaaggttct gtcgatgcct caggtcaaca tcgtcaagaa aaccgaggtt 3300

cagactggcg gattctccaa ggagtcgatt ctgcccaagc gaaactccga caagctcatc 3360

gctcgaaaga aagactggga tcccaagaaa tacggtggct tcgattctcc taccgtcgcc 3420

tattccgtgc ttgtcgttgc gaaggtcgag aagggcaagt ccaaaaagct caagtccgtc 3480

aaggagctgc tcggaattac catcatggag cgatcgagct tcgagaagaa tcccatcgac 3540

ttcttggaag ccaagggtta caaggaggtc aagaaagacc tcattatcaa gctgcccaag 3600

tactctctgt tcgaactgga gaacggtcga aagcgtatgc tcgcctccgc tggcgagctg 3660

cagaagggaa acgagcttgc cttgccttcg aagtacgtca actttctcta tctggcttct 3720

cactacgaga agctcaaggg ttctcccgag gacaacgaac agaagcaact cttcgttgag 3780

cagcacaaac attacctcga cgagattatc gagcagattt ccgagttttc gaagcgagtc 3840

atcctggctg atgccaactt ggacaaggtg ctctctgcct acaacaagca tcgggacaaa 3900

cccattcgag aacaggcgga gaacatcatt cacctgttta ctcttaccaa cctgggtgct 3960

cctgcagctt tcaagtactt cgataccact atcgaccgaa agcggtacac atccaccaag 4020

gaggttctcg atgccaccct gattcaccag tccatcactg gcctgtacga gacccgaatc 4080

gacctgtctc agcttggtgg cgactccaga gccgatccca agaaaaagcg aaaggtctaa 4140

<210> 121

<211> 10706

<212> DNA

<213> Artificial sequence

<220><223> pRF111

<400> 121

catggacaag aaatactcca tcggcctggc cattggaacc aactctgtcg gctgggctgt 60

catcaccgac gagtacaagg tgccctccaa gaaattcaag gtcctcggaa acaccgatcg 120

acactccatc aagaaaaacc tcattggtgc cctgttgttc gattctggcg agactgccga 180

agctaccaga ctcaagcgaa ctgctcggcg acgttacacc cgacggaaga accgaatctg 240

ctacctgcag gagatctttt ccaacgagat ggccaaggtg gacgattcgt tctttcatcg 300

actggaggaa tccttcctcg tcgaggaaga caagaaacac gagcgtcatc ccatctttgg 360

caacattgtg gacgaggttg cttaccacga gaagtatcct accatctacc acctgcgaaa 420

gaaactcgtc gattccaccg acaaggcgga tctcagactt atctacctcg ctctggcaca 480

catgatcaag tttcgaggtc atttcctcat cgagggcgat ctcaatcccg acaacagcga 540

tgtggacaag ctgttcattc agctcgttca gacctacaac cagctgttcg aggaaaaccc 600

catcaatgcc tccggagtcg atgcaaaggc catcttgtct gctcgactct cgaagagcag 660

acgactggag aacctcattg cccaacttcc tggcgagaaa aagaacggac tgtttggcaa 720

cctcattgcc ctttctcttg gtctcacacc caacttcaag tccaacttcg atctggcgga 780

ggacgccaag ctccagctgt ccaaggacac ctacgacgat gacctcgaca acctgcttgc 840

acagattggc gatcagtacg ccgacctgtt tctcgctgcc aagaaccttt cggatgctat 900

tctcttgtct gacattctgc gagtcaacac cgagatcaca aaggctcccc tttctgcctc 960

catgatcaag cgatacgacg agcaccatca ggatctcaca ctgctcaagg ctcttgtccg 1020

acagcaactg cccgagaagt acaaggagat ctttttcgat cagtcgaaga acggctacgc 1080

tggatacatc gacggcggag cctctcagga agagttctac aagttcatca agccaattct 1140

cgagaagatg gacggaaccg aggaactgct tgtcaagctc aatcgagagg atctgcttcg 1200

gaagcaacga accttcgaca acggcagcat tcctcatcag atccacctcg gtgagctgca 1260

cgccattctt cgacgtcagg aagacttcta ccccttttctc aaggacaacc gagagaagat 1320

cgagaagatt cttacctttc gaatccccta ctatgttggt cctcttgcca gaggaaactc 1380

tcgatttgct tggatgactc gaaagtccga ggaaaccatc actccctgga acttcgagga 1440

agtcgtggac aagggtgcct ctgcacagtc cttcatcgag cgaatgacca acttcgacaa 1500

gaatctgccc aacgagaagg ttcttcccaa gcattcgctg ctctacgagt actttacagt 1560

ctacaacgaa ctcaccaaag tcaagtacgt taccgaggga atgcgaaagc ctgccttctt 1620

gtctggcgaa cagaagaaag ccattgtcga tctcctgttc aagaccaacc gaaaggtcac 1680

tgttaagcag ctcaaggagg actacttcaa gaaaatcgag tgtttcgaca gcgtcgagat 1740

ttccggagtt gaggaccgat tcaacgcctc tttgggcacc tatcacgatc tgctcaagat 1800

tatcaaggac aaggattttc tcgacaacga ggaaaacgag gacattctgg aggacatcgt 1860

gctcactctt accctgttcg aagatcggga gatgatcgag gaacgactca agacatacgc 1920

tcacctgttc gacgacaagg tcatgaaaca actcaagcga cgtagataca ccggctgggg 1980

aagactttcg cgaaagctca tcaacggcat cagagacaag cagtccggaa agaccattct 2040

ggactttctc aagtccgatg gctttgccaa ccgaaacttc atgcagctca ttcacgacga 2100

ttctcttacc ttcaaggagg acatccagaa ggcacaagtg tccggtcagg gcgacagctt 2160

gcacgaacat attgccaacc tggctggttc gccagccatc aagaaaggca ttctccagac 2220

tgtcaaggtt gtcgacgagc tggtgaaggt catgggacgt cacaagcccg agaacattgt 2280

gatcgagatg gccagagaga accagacaac tcaaaagggt cagaaaaact cgcgagagcg 2340

gatgaagcga atcgaggaag gcatcaagga gctgggatcc cagattctca aggagcatcc 2400

cgtcgagaac actcaactgc agaacgagaa gctgtatctc tactatctgc agaatggtcg 2460

agacatgtac gtggatcagg aactggacat caatcgtctc agcgactacg atgtggacca 2520

cattgtccct caatcctttc tcaaggacga ttctatcgac aacaaggtcc ttacacgatc 2580

cgacaagaac agaggcaagt cggacaacgt tcccagcgaa gaggtggtca aaaagatgaa 2640
gaactactgg cgacagctgc tcaacgccaa gctcattacc cagcgaaagt tcgacaatct 2700
taccaaggcc gagcgaggcg gtctgtccga gctcgacaag gctggcttca tcaagcgtca 2760
actcgtcgag accagacaga tcacaaagca cgtcgcacag attctcgatt ctcggatgaa 2820
caccaagtac gacgagaacg acaagctcat ccgagaggtc aaggtgatta ctctcaagtc 2880
caaactggtc tccgatttcc gaaaggactt tcagttctac aaggtgcgag agatcaacaa 2940
ttaccaccat gcccacgatg cttacctcaa cgccgtcgtt ggcactgcgc tcatcaagaa 3000

ataccccaag ctcgaaagcg agttcgttta cggcgattac aaggtctacg acgttcgaaa 3060
gatgattgcc aagtccgaac aggagattgg caaggctact gccaagtact tcttttactc 3120
caacatcatg aacttttca agaccgagat caccttggcc aacggagaga ttcgaaagag 3180
accacttatc gagaccaacg gcgaaactgg agagatcgtg tgggacaagg gtcgagactt 3240
tgcaaccgtg cgaaaggttc tgtcgatgcc tcaggtcaac atcgtcaaga aaaccgaggt 3300
tcagactggc ggattctcca aggagtcgat tctgcccaag cgaaactccg acaagctcat 3360
cgctcgaaag aaagactggg atcccaagaa atacggtggc ttcgattctc ctaccgtcgc 3420

ctattccgtg cttgtcgttg cgaaggtcga gaagggcaag tccaaaaagc tcaagtccgt 3480
caaggagctg ctcggaatta ccatcatgga gcgatcgagc ttcgagaaga atcccatcga 3540
cttcttggaa gccaagggtt acaaggaggt caagaaagac ctcattatca agctgcccaa 3600
gtactctctg ttcgaactgg agaacggtcg aaagcgtatg ctcgcctccg ctggcgagct 3660
gcagaaggga aacgagcttg ccttgccttc gaagtacgtc aactttctct atctggcttc 3720
tcactacgag aagctcaagg gttctcccga ggacaacgaa cagaagcaac tcttcgttga 3780
gcagcacaaa cattacctcg acgagattat cgagcagatt tccgagtttt cgaagcgagt 3840

catcctggct gatgccaact tggacaaggt gctctctgcc tacaacaagc atcgggacaa 3900
acccattcga gaacaggcgg agaacatcat tcacctgttt actcttacca acctgggtgc 3960
tcctgcagct ttcaagtact tcgataccac tatcgaccga aagcggtaca catccaccaa 4020
ggaggttctc gatgccaccc tgattcacca gtccatcact ggcctgtacg agacccgaat 4080
cgacctgtct cagcttggtg gcgactccag agccgatccc aagaaaaagc gaaaggtcta 4140
agcggccgca agtgtggatg gggaagtgag tgcccggttc tgtgtgcaca attggcaatc 4200
caagatggat ggattcaaca cagggatata gcgagctacg tggtggtgcg aggatatagc 4260

aacggatatt tatgtttgac acttgagaat gtacgataca agcactgtcc aagtacaata 4320

ctaaacatac tgtacatact catactcgta cccgggcaac ggtttcactt gagtgcagtg 4380

gctagtgctc ttactcgtac agtgtgcaat actgcgtatc atagtctttg atgtatatcg 4440

tattcattca tgttagttgc gtacgagccg gaagcataaa gtgtaaagcc tggggtgcct 4500

aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa 4560

acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta 4620

ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc 4680

gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg 4740

caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt 4800

tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa 4860

gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct 4920

ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc 4980

cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg 5040

tcgttcgctc caagctgggc tgtgtgcacg aaccccccgt tcagcccgac cgctgcgcct 5100

tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag 5160

cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga 5220

agtggtggcc taactacggc tacactagaa ggacagtatt tggtatctgc gctctgctga 5280

agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg 5340

gtagcggtgg tttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag 5400

aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag 5460

ggattttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat 5520

gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct 5580

taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac 5640

tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa 5700

tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg 5760

gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt 5820

gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca 5880

ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt 5940

cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct 6000

tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg 6060

cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg 6120

agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg 6180

cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa 6240

aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt 6300

aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt 6360

gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt 6420

gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca 6480

tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat 6540

ttccccgaaa agtgccacct gacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg 6600

tggttacgcg cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt 6660

tcttcccttc ctttctcgcc acgttcgccg gctttccccg tcaagctcta aatcgggggc 6720

tccctttagg gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg 6780

gtgatggttc acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg 6840

agtccacgtt ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct 6900

cggtctattc ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg 6960

agctgattta acaaaaattt aacgcgaatt ttaacaaaat attaacgctt acaatttcca 7020

ttcgccattc aggctgcgca actgttggga agggcgatcg gtgcgggcct cttcgctatt 7080

acgccagctg gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa cgccagggtt 7140

ttcccagtca cgacgttgta aaacgacggc cagtgaattg taatacgact cactataggg 7200

cgaattgggt accgggcccc ccctcgaggt cgatggtgtc gataagcttg atatcgaatt 7260

catgtcacac aaaccgatct tcgcctcaag gaaacctaat tctacatccg agagactgcc 7320

gagatccagt ctacactgat taattttcgg gccaataatt taaaaaaatc gtgttatata 7380

atattatatg tattatatat atacatcatg atgatactga cagtcatgtc ccattgctaa 7440

atagacagac tccatctgcc gcctccaact gatgttctca atatttaagg ggtcatctcg 7500

cattgtttaa taataaacag actccatcta ccgcctccaa atgatgttct caaaatatat 7560

tgtatgaact tatttttatt acttagtatt attagacaac ttacttgctt tatgaaaaac 7620

acttcctatt taggaaacaa tttataatgg cagttcgttc atttaacaat ttatgtagaa 7680

taaatgttat aaatgcgtat gggaaatctt aaatatggat agcataaatg atatctgcat 7740

tgcctaattc gaaatcaaca gcaacgaaaa aaatcccttg tacaacataa atagtcatcg 7800

agaaatatca actatcaaag aacagctatt cacacgttac tattgagatt attattggac 7860

gagaatcaca cactcaactg tctttctctc ttctagaaat acaggtacaa gtatgtacta 7920

ttctcattgt tcatacttct agtcatttca tcccacatat tccttggatt tctctccaat 7980

gaatgacatt ctatcttgca aattcaacaa ttataataag atataccaaa gtagcggtat 8040

agtggcaatc aaaaagcttc tctggtgtgc ttctcgtatt tatttttatt ctaatgatcc 8100
attaaaggta tatatttatt tcttgttata taatcctttt gtttattaca tgggctggat 8160
acataaaggt attttgattt aatttttttgc ttaaattcaa tccccccctcg ttcagtgtca 8220
actgtaatgg taggaaatta ccatactttt gaagaagcaa aaaaaatgaa agaaaaaaaa 8280
aatcgtattt ccaggttaga cgttccgcag aatctagaat gcggtatgcg gtacattgtt 8340
cttcgaacgt aaaagttgcg ctccctgaga tattgtacat ttttgctttt acaagtacaa 8400
gtacatcgta caactatgta ctactgttga tgcatccaca acagtttgtt ttgttttttt 8460

ttgttttttt tttttctaat gattcattac cgctatgtat acctacttgt acttgtagta 8520
agccgggtta ttggcgttca attaatcata gacttatgaa tctgcacggt gtgcgctgcg 8580
agttactttt agcttatgca tgctacttgg gtgtaatatt gggatctgtt cggaaatcaa 8640
cggatgctca atcgatttcg acagtaatta attaagtcat acacaagtca gctttcttcg 8700
agcctcatat aagtataagt agttcaacgt attagcactg tacccagcat ctccgtatcg 8760
agaaacacaa caacatgccc cattggacag atcatgcgga tacacaggtt gtgcagtatc 8820
atacatactc gatcagacag gtcgtctgac catcatacaa gctgaacaag cgctccatac 8880

ttgcacgctc tctatataca cagttaaatt acatatccat agtctaacct ctaacagtta 8940
atcttctggt aagcctccca gccagccttc tggtatcgct tggcctcctc aataggatct 9000
cggttctggc cgtacagacc tcggccgaca attatgatat ccgttccggt agacatgaca 9060
tcctcaacag ttcggtactg ctgtccgaga gcgtctccct tgtcgtcaag acccaccccg 9120
ggggtcagaa taagccagtc ctcagagtcg cccttaggtc ggttctgggc aatgaagcca 9180
accacaaact cggggtcgga tcgggcaagc tcaatggtct gcttggagta ctcgccagtg 9240
gccagagagc ccttgcaaga cagctcggcc agcatgagca gacctctggc cagcttctcg 9300

ttgggagagg ggactaggaa ctccttgtac tgggagttct cgtagtcaga gacgtcctcc 9360
ttcttctgtt cagagacagt ttcctcggca ccagctcgca ggccagcaat gattccggtt 9420
ccgggtacac cgtgggcgtt ggtgatatcg gaccactcgg cgattcggtg acaccggtac 9480
tggtgcttga cagtgttgcc aatatctgcg aactttctgt cctcgaacag gaagaaaccg 9540
tgcttaagag caagttcctt gaggggggagc acagtgccgg cgtaggtgaa gtcgtcaatg 9600
atgtcgatat gggttttgat catgcacaca taaggtccga ccttatcggc aagctcaatg 9660
agctccttgg tggtggtaac atccagagaa gcacacaggt tggttttctt ggctgccacg 9720

agcttgagca ctcgagcggc aaaggcggac ttgtggacgt tagctcgagc ttcgtaggag 9780

ggcattttgg tggtgaagag gagactgaaa taaatttagt ctgcagaact ttttatcgga 9840

accttatctg gggcagtgaa gtatatgtta tggtaatagt tacgagttag ttgaacttat 9900

agatagactg gactatacgg ctatcggtcc aaattagaaa gaacgtcaat ggctctctgg 9960

gcgtcgcctt tgccgacaaa aatgtgatca tgatgaaagc cagcaatgac gttgcagctg 10020

atattgttgt cggccaaccg cgccgaaaac gcagctgtca gacccacagc ctccaacgaa 10080

gaatgtatcg tcaaagtgat ccaagcacac tcatagttgg agtcgtactc caaaggcggc 10140

aatgacgagt cagacagata ctcgtcgacg tttaaaccat catctaaggg cctcaaaact 10200

acctcggaac tgctgcgctg atctggacac cacagaggtt ccgagcactt taggttgcac 10260

caaatgtccc accaggtgca ggcagaaaac gctggaacag cgtgtacagt ttgtcttaac 10320

aaaaagtgag ggcgctgagg tcgagcaggg tggtgtgact tgttatagcc tttagagctg 10380

cgaaagcgcg tatggatttg gctcatcagg ccagattgag ggtctgtgga cacatgtcat 10440

gttagtgtac ttcaatcgcc ccctggatat agccccgaca ataggccgtg gcctcatttt 10500

tttgccttcc gcacatttcc attgctcggt acccacacct tgcttctcct gcacttgcca 10560

accttaatac tggtttacat tgaccaacat cttacaagcg gggggcttgt ctagggtata 10620

tataaacagt ggctctccca atcggttgcc agtctctttt ttcctttctt tccccacaga 10680

ttcgaaatct aaactacaca tcacac 10706

<210> 122

<211> 43

<212> DNA

<213> Artificial sequence

<220><223> H840A1

<400> 122

tcagcgacta cgatgtggac gccattgtcc ctcaatcctt tct 43

<210> 123

<211> 43

<212> DNA

<213> Artificial sequence

<220><223> H840A2

<400> 123

agaaaggatt gagggacaat ggcgtccaca tcgtagtcgc tga 43

<210> 124

<211> 4140

<212> DNA

<213> Artificial Sequence

<220><223> YArrowia codon optimized inactivated Cas9

<400> 124

atggacaaga aatactccat cggcctggcc attggaacca actctgtcgg ctgggctgtc 60

atcaccgacg agtacaaggt gccctccaag aaattcaagg tcctcggaaa caccgatcga 120

cactccatca agaaaaacct cattggtgcc ctgttgttcg attctggcga gactgccgaa 180

gctaccagac tcaagcgaac tgctcggcga cgttacaccc gacggaagaa ccgaatctgc 240

tacctgcagg agatcttttc caacgagatg gccaaggtgg acgattcgtt ctttcatcga 300

ctggaggaat ccttcctcgt cgaggaagac aagaaacacg agcgtcatcc catctttggc 360

aacattgtgg acgaggttgc ttaccacgag aagtatccta ccatctacca cctgcgaaag 420

aaactcgtcg attccaccga caaggcggat ctcagactta tctacctcgc tctggcacac 480

atgatcaagt ttcgaggtca tttcctcatc gagggcgatc tcaatcccga caacagcgat 540

gtggacaagc tgttcattca gctcgttcag acctacaacc agctgttcga ggaaaacccc 600

atcaatgcct ccggagtcga tgcaaaggcc atcttgtctg ctcgactctc gaagagcaga 660

cgactggaga acctcattgc ccaacttcct ggcgagaaaa agaacggact gtttggcaac 720

ctcattgccc tttctcttgg tctcacaccc aacttcaagt ccaacttcga tctggcggag 780

gacgccaagc tccagctgtc caaggacacc tacgacgatg acctcgacaa cctgcttgca 840

cagattggcg atcagtacgc cgacctgttt ctcgctgcca agaacctttc ggatgctatt 900

ctcttgtctg acattctgcg agtcaacacc gagatcacaa aggctcccct ttctgcctcc 960

atgatcaagc gatacgacga gcaccatcag gatctcacac tgctcaaggc tcttgtccga 1020

cagcaactgc ccgagaagta caaggagatc tttttcgatc agtcgaagaa cggctacgct 1080

ggatacatcg acggcggagc ctctcaggaa gagttctaca agttcatcaa gccaattctc 1140

gagaagatgg acggaaccga ggaactgctt gtcaagctca atcgagagga tctgcttcgg 1200

aagcaacgaa ccttcgacaa cggcagcatt cctcatcaga tccacctcgg tgagctgcac 1260

gccattcttc gacgtcagga agacttctac ccctttctca aggacaaccg agagaagatc 1320

gagaagattc ttacctttcg aatcccctac tatgttggtc ctcttgccag aggaaactct 1380

cgatttgctt ggatgactcg aaagtccgag gaaaccatca ctccctggaa cttcgaggaa 1440

gtcgtggaca agggtgcctc tgcacagtcc ttcatcgagc gaatgaccaa cttcgacaag 1500

aatctgccca acgagaaggt tcttcccaag cattcgctgc tctacgagta ctttacagtc 1560

tacaacgaac tcaccaaagt caagtacgtt accgagggaa tgcgaaagcc tgccttcttg 1620

tctggcgaac agaagaaagc cattgtcgat ctcctgttca agaccaaccg aaaggtcact 1680

gttaagcagc tcaaggagga ctacttcaag aaaatcgagt gtttcgacag cgtcgagatt 1740

tccggagttg aggaccgatt caacgcctct ttgggcacct atcacgatct gctcaagatt 1800

atcaaggaca aggattttct cgacaacgag gaaaacgagg acattctgga ggacatcgtg 1860

ctcactctta ccctgttcga agatcgggag atgatcgagg aacgactcaa gacatacgct 1920

cacctgttcg acgacaaggt catgaaacaa ctcaagcgac gtagatacac cggctgggga 1980

agactttcgc gaaagctcat caacggcatc agagacaagc agtccggaaa gaccattctg 2040

gactttctca agtccgatgg ctttgccaac cgaaacttca tgcagctcat tcacgacgat 2100

tctcttacct tcaaggagga catccagaag gcacaagtgt ccggtcaggg cgacagcttg 2160

cacgaacata ttgccaacct ggctggttcg ccagccatca agaaaggcat tctccagact 2220

gtcaaggttg tcgacgagct ggtgaaggtc atgggacgtc acaagcccga gaacattgtg 2280

atcgagatgg ccagagagaa ccagacaact caaaagggtc agaaaaactc gcgagagcgg 2340

atgaagcgaa tcgaggaagg catcaaggag ctgggatccc agattctcaa ggagcatccc 2400

gtcgagaaca ctcaactgca gaacgagaag ctgtatctct actatctgca gaatggtcga 2460

gacatgtacg tggatcagga actggacatc aatcgtctca gcgactacga tgtggacgcc 2520

attgtccctc aatcctttct caaggacgat tctatcgaca acaaggtcct tacacgatcc 2580

gacaagaaca gaggcaagtc ggacaacgtt cccagcgaag aggtggtcaa aaagatgaag 2640

aactactggc gacagctgct caacgccaag ctcattaccc agcgaaagtt cgacaatctt 2700

accaaggccg agcgaggcgg tctgtccgag ctcgacaagg ctggcttcat caagcgtcaa 2760

ctcgtcgaga ccagacagat cacaaagcac gtcgcacaga ttctcgattc tcggatgaac 2820

accaagtacg acgagaacga caagctcatc cgagaggtca aggtgattac tctcaagtcc 2880

aaactggtct ccgatttccg aaaggacttt cagttctaca aggtgcgaga gatcaacaat 2940

taccaccatg cccacgatgc ttacctcaac gccgtcgttg gcactgcgct catcaagaaa 3000

taccccaagc tcgaaagcga gttcgtttac ggcgattaca aggtctacga cgttcgaaag 3060

atgattgcca agtccgaaca ggagattggc aaggctactg ccaagtactt cttttactcc 3120

aacatcatga acttttcaa gaccgagatc accttggcca acggagagat tcgaaagaga 3180

ccacttatcg agaccaacgg cgaaactgga gagatcgtgt gggacaaggg tcgagacttt 3240

gcaaccgtgc gaaaggttct gtcgatgcct caggtcaaca tcgtcaagaa aaccgaggtt 3300

cagactggcg gattctccaa ggagtcgatt ctgcccaagc gaaactccga caagctcatc 3360

gctcgaaaga aagactggga tcccaagaaa tacggtggct tcgattctcc taccgtcgcc 3420

tattccgtgc ttgtcgttgc gaaggtcgag aagggcaagt ccaaaaagct caagtccgtc 3480

aaggagctgc tcggaattac catcatggag cgatcgagct tcgagaagaa tcccatcgac 3540

ttcttggaag ccaagggtta caaggaggtc aagaaagacc tcattatcaa gctgcccaag 3600

tactctctgt tcgaactgga gaacggtcga aagcgtatgc tcgcctccgc tggcgagctg 3660

cagaagggaa acgagcttgc cttgccttcg aagtacgtca actttctcta tctggcttct 3720

cactacgaga agctcaaggg ttctcccgag gacaacgaac agaagcaact cttcgttgag 3780

cagcacaaac attacctcga cgagattatc gagcagattt ccgagttttc gaagcgagtc 3840

atcctggctg atgccaactt ggacaaggtg ctctctgcct acaacaagca tcgggacaaa 3900

cccattcgag aacaggcgga gaacatcatt cacctgttta ctcttaccaa cctgggtgct 3960

cctgcagctt tcaagtactt cgataccact atcgaccgaa agcggtacac atccaccaag 4020

gaggttctcg atgccaccct gattcaccag tccatcactg gcctgtacga gacccgaatc 4080

gacctgtctc agcttggtgg cgactccaga gccgatccca agaaaaagcg aaaggtctaa 4140

<210> 125

<211> 10706

<212> DNA

<213> Artificial sequence

<220><223> pRF143

<400> 125

catggacaag aaatactcca tcggcctggc cattggaacc aactctgtcg gctgggctgt 60

catcaccgac gagtacaagg tgccctccaa gaaattcaag gtcctcggaa acaccgatcg 120

acactccatc aagaaaaacc tcattggtgc cctgttgttc gattctggcg agactgccga 180

agctaccaga ctcaagcgaa ctgctcggcg acgttacacc cgacggaaga accgaatctg 240

ctacctgcag gagatctttt ccaacgagat ggccaaggtg gacgattcgt tctttcatcg 300

actggaggaa tccttcctcg tcgaggaaga caagaaacac gagcgtcatc ccatctttgg 360

caacattgtg gacgaggttg cttaccacga gaagtatcct accatctacc acctgcgaaa 420

gaaactcgtc gattccaccg acaaggcgga tctcagactt atctacctcg ctctggcaca 480

catgatcaag tttcgaggtc atttcctcat cgagggcgat ctcaatcccg acaacagcga 540

tgtggacaag ctgttcattc agctcgttca gacctacaac cagctgttcg aggaaaaccc 600

catcaatgcc tccggagtcg atgcaaaggc catcttgtct gctcgactct cgaagagcag 660

acgactggag aacctcattg cccaacttcc tggcgagaaa aagaacggac tgtttggcaa 720

cctcattgcc ctttctcttg gtctcacacc caacttcaag tccaacttcg atctggcgga 780

ggacgccaag ctccagctgt ccaaggacac ctacgacgat gacctcgaca acctgcttgc 840

acagattggc gatcagtacg ccgacctgtt tctcgctgcc aagaaccttt cggatgctat 900

tctcttgtct gacattctgc gagtcaacac cgagatcaca aaggctcccc tttctgcctc 960

catgatcaag cgatacgacg agcaccatca ggatctcaca ctgctcaagg ctcttgtccg 1020

acagcaactg cccgagaagt acaaggagat ctttttcgat cagtcgaaga acggctacgc 1080

tggatacatc gacggcggag cctctcagga agagttctac aagttcatca agccaattct 1140

cgagaagatg gacggaaccg aggaactgct tgtcaagctc aatcgagagg atctgcttcg 1200

gaagcaacga accttcgaca acggcagcat tcctcatcag atccacctcg gtgagctgca 1260

cgccattctt cgacgtcagg aagacttcta ccccttttctc aaggacaacc gagagaagat 1320

cgagaagatt cttacctttc gaatccccta ctatgttggt cctcttgcca gaggaaactc 1380

tcgatttgct tggatgactc gaaagtccga ggaaaccatc actccctgga acttcgagga 1440

agtcgtggac aagggtgcct ctgcacagtc cttcatcgag cgaatgacca acttcgacaa 1500

gaatctgccc aacgagaagg ttcttcccaa gcattcgctg ctctacgagt actttacagt 1560

ctacaacgaa ctcaccaaag tcaagtacgt taccgaggga atgcgaaagc ctgccttctt 1620

gtctggcgaa cagaagaaag ccattgtcga tctcctgttc aagaccaacc gaaaggtcac 1680

tgttaagcag ctcaaggagg actacttcaa gaaaatcgag tgtttcgaca gcgtcgagat 1740

ttccggagtt gaggaccgat tcaacgcctc tttgggcacc tatcacgatc tgctcaagat 1800

tatcaaggac aaggattttc tcgacaacga ggaaaacgag gacattctgg aggacatcgt 1860

gctcactctt accctgttcg aagatcggga gatgatcgag gaacgactca agacatacgc 1920

tcacctgttc gacgacaagg tcatgaaaca actcaagcga cgtagataca ccggctgggg 1980

aagactttcg cgaaagctca tcaacggcat cagagacaag cagtccggaa agaccattct 2040

ggactttctc aagtccgatg gctttgccaa ccgaaacttc atgcagctca ttcacgacga 2100

ttctcttacc ttcaaggagg acatccagaa ggcacaagtg tccggtcagg gcgacagctt 2160

gcacgaacat attgccaacc tggctggttc gccagccatc aagaaaggca ttctccagac 2220

tgtcaaggtt gtcgacgagc tggtgaaggt catgggacgt cacaagcccg agaacattgt 2280

gatcgagatg gccagagaga accagacaac tcaaaagggt cagaaaaact cgcgagagcg 2340

gatgaagcga atcgaggaag gcatcaagga gctgggatcc cagattctca aggagcatcc 2400

cgtcgagaac actcaactgc agaacgagaa gctgtatctc tactatctgc agaatggtcg 2460

agacatgtac gtggatcagg aactggacat caatcgtctc agcgactacg atgtggacgc 2520

cattgtccct caatcctttc tcaaggacga ttctatcgac aacaaggtcc ttacacgatc 2580

cgacaagaac agaggcaagt cggacaacgt tcccagcgaa gaggtggtca aaaagatgaa 2640

gaactactgg cgacagctgc tcaacgccaa gctcattacc cagcgaaagt tcgacaatct 2700

taccaaggcc gagcgaggcg gtctgtccga gctcgacaag gctggcttca tcaagcgtca 2760

actcgtcgag accagacaga tcacaaagca cgtcgcacag attctcgatt ctcggatgaa 2820

caccaagtac gacgagaacg acaagctcat ccgagaggtc aaggtgatta ctctcaagtc 2880

caaactggtc tccgatttcc gaaaggactt tcagttctac aaggtgcgag agatcaacaa 2940

ttaccaccat gcccacgatg cttacctcaa cgccgtcgtt ggcactgcgc tcatcaagaa 3000

ataccccaag ctcgaaagcg agttcgttta cggcgattac aaggtctacg acgttcgaaa 3060

gatgattgcc aagtccgaac aggagattgg caaggctact gccaagtact tcttttactc 3120

caacatcatg aacttttca agaccgagat caccttggcc aacggagaga ttcgaaagag 3180

accacttatc gagaccaacg gcgaaactgg agagatcgtg tgggacaagg gtcgagactt 3240

tgcaaccgtg cgaaaggttc tgtcgatgcc tcaggtcaac atcgtcaaga aaaccgaggt 3300

tcagactggc ggattctcca aggagtcgat tctgcccaag cgaaactccg acaagctcat 3360

cgctcgaaag aaagactggg atcccaagaa atacggtggc ttcgattctc ctaccgtcgc 3420

ctattccgtg cttgtcgttg cgaaggtcga gaagggcaag tccaaaaagc tcaagtccgt 3480

caaggagctg ctcggaatta ccatcatgga gcgatcgagc ttcgagaaga atcccatcga 3540

cttcttggaa gccaagggtt acaaggaggt caagaaagac ctcattatca agctgcccaa 3600

gtactctctg ttcgaactgg agaacggtcg aaagcgtatg ctcgcctccg ctggcgagct 3660

gcagaaggga aacgagcttg ccttgccttc gaagtacgtc aactttctct atctggcttc 3720

tcactacgag aagctcaagg gttctcccga ggacaacgaa cagaagcaac tcttcgttga 3780

gcagcacaaa cattacctcg acgagattat cgagcagatt tccgagtttt cgaagcgagt 3840

catcctggct gatgccaact tggacaaggt gctctctgcc tacaacaagc atcgggacaa 3900

acccattcga gaacaggcgg agaacatcat tcacctgttt actcttacca acctgggtgc 3960

tcctgcagct ttcaagtact tcgataccac tatcgaccga aagcggtaca catccaccaa 4020

ggaggttctc gatgccaccc tgattcacca gtccatcact ggcctgtacg agacccgaat 4080

cgacctgtct cagcttggtg gcgactccag agccgatccc aagaaaaagc gaaaggtcta 4140

agcggccgca agtgtggatg gggaagtgag tgcccggttc tgtgtgcaca attggcaatc 4200

caagatggat ggattcaaca cagggatata gcgagctacg tggtggtgcg aggatatagc 4260

aacggatatt tatgtttgac acttgagaat gtacgataca agcactgtcc aagtacaata 4320

ctaaacatac tgtacatact catactcgta cccgggcaac ggtttcactt gagtgcagtg 4380

gctagtgctc ttactcgtac agtgtgcaat actgcgtatc atagtctttg atgtatatcg 4440

tattcattca tgttagttgc gtacgagccg gaagcataaa gtgtaaagcc tggggtgcct 4500

aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa 4560
acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta 4620
ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc 4680
gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg 4740
caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt 4800
tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa 4860
gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct 4920

ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc 4980
cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg 5040
tcgttcgctc caagctgggc tgtgtgcacg aaccccccgt tcagcccgac cgctgcgcct 5100
tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag 5160
cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga 5220
agtggtggcc taactacggc tacactagaa ggacagtatt tggtatctgc gctctgctga 5280
agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg 5340

gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag 5400
aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag 5460
ggattttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat 5520
gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct 5580
taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac 5640
tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa 5700
tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg 5760

gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt 5820
gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca 5880
ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt 5940
cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct 6000
tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg 6060
cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg 6120
agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg 6180

cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa 6240

aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt 6300

aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt 6360

gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt 6420

gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca 6480

tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat 6540

ttccccgaaa agtgccacct gacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg 6600

tggttacgcg cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt 6660

tcttcccttc ctttctcgcc acgttcgccg gctttccccg tcaagctcta aatcgggggc 6720

tccctttagg gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg 6780

gtgatggttc acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg 6840

agtccacgtt ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct 6900

cggtctattc ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg 6960

agctgattta acaaaaattt aacgcgaatt ttaacaaaat attaacgctt acaatttcca 7020

ttcgccattc aggctgcgca actgttggga agggcgatcg gtgcgggcct cttcgctatt 7080

acgccagctg gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa cgccagggtt 7140

ttcccagtca cgacgttgta aaacgacggc cagtgaattg taatacgact cactataggg 7200

cgaattgggt accgggcccc ccctcgaggt cgatggtgtc gataagcttg atatcgaatt 7260

catgtcacac aaaccgatct tcgcctcaag gaaacctaat tctacatccg agagactgcc 7320

gagatccagt ctacactgat taattttcgg gccaataatt taaaaaaatc gtgttatata 7380

atattatatg tattatatat atacatcatg atgatactga cagtcatgtc ccattgctaa 7440

atagacagac tccatctgcc gcctccaact gatgttctca atatttaagg ggtcatctcg 7500

cattgtttaa taataaacag actccatcta ccgcctccaa atgatgttct caaaatatat 7560

tgtatgaact tatttttatt acttagtatt attagacaac ttacttgctt tatgaaaaac 7620

acttcctatt taggaaacaa tttataatgg cagttcgttc atttaacaat ttatgtagaa 7680

taaatgttat aaatgcgtat gggaaatctt aaatatggat agcataaatg atatctgcat 7740

tgcctaattc gaaatcaaca gcaacgaaaa aaatcccttg tacaacataa atagtcatcg 7800

agaaatatca actatcaaag aacagctatt cacacgttac tattgagatt attattggac 7860

gagaatcaca cactcaactg tctttctctc ttctagaaat acaggtacaa gtatgtacta 7920

ttctcattgt tcatacttct agtcatttca tcccacatat tccttggatt tctctccaat 7980

gaatgacatt ctatcttgca aattcaacaa ttataataag atataccaaa gtagcggtat 8040

agtggcaatc aaaaagcttc tctggtgtgc ttctcgtatt tatttttatt ctaatgatcc 8100

attaaaggta tatatttatt tcttgttata taatcctttt gtttattaca tgggctggat 8160

acataaaggt attttgattt aatttttgc ttaaattcaa tccccctcg ttcagtgtca 8220

actgtaatgg taggaaatta ccatactttt gaagaagcaa aaaaaatgaa agaaaaaaaa 8280

aatcgtattt ccaggttaga cgttccgcag aatctagaat gcggtatgcg gtacattgtt 8340

cttcgaacgt aaaagttgcg ctccctgaga tattgtacat ttttgctttt acaagtacaa 8400

gtacatcgta caactatgta ctactgttga tgcatccaca acagtttgtt ttgttttttt 8460

ttgttttttt tttttctaat gattcattac cgctatgtat acctacttgt acttgtagta 8520

agccgggtta ttggcgttca attaatcata gacttatgaa tctgcacggt gtgcgctgcg 8580

agttactttt agcttatgca tgctacttgg gtgtaatatt gggatctgtt cggaaatcaa 8640

cggatgctca atcgatttcg acagtaatta attaagtcat acacaagtca gctttcttcg 8700

agcctcatat aagtataagt agttcaacgt attagcactg tacccagcat ctccgtatcg 8760

agaaacacaa caacatgccc cattggacag atcatgcgga tacacaggtt gtgcagtatc 8820

atacatactc gatcagacag gtcgtctgac catcatacaa gctgaacaag cgctccatac 8880

ttgcacgctc tctatataca cagttaaatt acatatccat agtctaacct ctaacagtta 8940

atcttctggt aagcctccca gccagccttc tggtatcgct tggcctcctc aataggatct 9000

cggttctggc cgtacagacc tcggccgaca attatgatat ccgttccggt agacatgaca 9060

tcctcaacag ttcggtactg ctgtccgaga gcgtctccct tgtcgtcaag acccaccccg 9120

ggggtcagaa taagccagtc ctcagagtcg cccttaggtc ggttctgggc aatgaagcca 9180

accacaaact cggggtcgga tcgggcaagc tcaatggtct gcttggagta ctcgccagtg 9240

gccagagagc ccttgcaaga cagctcggcc agcatgagca gacctctggc cagcttctcg 9300

ttgggagagg ggactaggaa ctccttgtac tgggagttct cgtagtcaga gacgtcctcc 9360

ttcttctgtt cagagacagt ttcctcggca ccagctcgca ggccagcaat gattccggtt 9420

ccgggtacac cgtgggcgtt ggtgatatcg gaccactcgg cgattcggtg acaccggtac 9480

tggtgcttga cagtgttgcc aatatctgcg aactttctgt cctcgaacag gaagaaaccg 9540

tgcttaagag caagttcctt gaggggggagc acagtgccgg cgtaggtgaa gtcgtcaatg 9600

atgtcgatat gggttttgat catgcacaca taaggtccga ccttatcggc aagctcaatg 9660

agctccttgg tggtggtaac atccagagaa gcacacaggt tggttttctt ggctgccacg 9720

agcttgagca ctcgagcggc aaaggcggac ttgtggacgt tagctcgagc ttcgtaggag 9780

ggcattttgg tggtgaagag gagactgaaa taaatttagt ctgcagaact ttttatcgga 9840

accttatctg gggcagtgaa gtatatgtta tggtaatagt tacgagttag ttgaacttat 9900

agatagactg gactatacgg ctatcggtcc aaattagaaa gaacgtcaat ggctctctgg 9960

gcgtcgcctt tgccgacaaa aatgtgatca tgatgaaagc cagcaatgac gttgcagctg 10020
atattgttgt cggccaaccg cgccgaaaac gcagctgtca gacccacagc ctccaacgaa 10080
gaatgtatcg tcaaagtgat ccaagcacac tcatagttgg agtcgtactc caaaggcggc 10140
aatgacgagt cagacagata ctcgtcgacg tttaaaccat catctaaggg cctcaaaact 10200
acctcggaac tgctgcgctg atctggacac cacagaggtt ccgagcactt taggttgcac 10260
caaatgtccc accaggtgca ggcagaaaac gctggaacag cgtgtacagt ttgtcttaac 10320
aaaaagtgag ggcgctgagg tcgagcaggg tggtgtgact tgttatagcc tttagagctg 10380

cgaaagcgcg tatggatttg gctcatcagg ccagattgag ggtctgtgga cacatgtcat 10440
gttagtgtac ttcaatcgcc ccctggatat agccccgaca ataggccgtg gcctcatttt 10500
tttgccttcc gcacatttcc attgctcggt acccacacct tgcttctcct gcacttgcca 10560
accttaatac tggtttacat tgaccaacat cttacaagcg gggggcttgt ctagggtata 10620
tataaacagt ggctctccca atcggttgcc agtctctttt ttcctttctt tccccacaga 10680
ttcgaaatct aaactacaca tcacac 10706

<210> 126
<211> 678
<212> DNA
<213> Artificial sequence

<220><223> Yarrowia optimized dsREDexpress ORF

<400> 126

atggcctcct cggaggacgt catcaaggag ttcatgcgat tcaaggtccg aatggaaggc 60
tccgtgaacg gtcacgagtt tgagattgag ggagagggtg aaggccgacc ctacgaaggc 120
acccagaccg cgaagctgaa ggtgaccaag ggtggacccc tgcccttcgc ctgggacatt 180
ctgtctcctc agtttcagta cggttctaag gtgtacgtga agcaccctgc tgacattccc 240
gactacaaga aactttcctt tcccgagggc ttcaagtggg agcgagttat gaacttcgag 300
gatggcggtg tcgttaccgt tactcaggac tcctcgctcc aggacggctc gttcatctac 360

aaggttaagt tcatcggtgt caacttccct agcgatggac ccgtcatgca aaagaaaact 420
atgggatggg aagcctctac agagcggctg taccctcgag acggagtgtt gaagggcgag 480
attcacaagg ccctgaagct caaggacggt ggacactatc tcgttgagtt taagtctatc 540
tacatggcaa agaaacccgt gcagcttcca ggctactatt acgtcgattc caagctcgat 600
atcaccagcc ataatgagga ctacactatt gtcgaacagt acgagcgtgc tgagggaaga 660

caccatctgt ttctttaa 678

<210> 127

<211> 696

<212> DNA

<213> Artificial sequence

<220><223> Yarrowia optimized dsREDexpress cloning fragment

<400> 127

ggggccatgg cctcctcgga ggacgtcatc aaggagttca tgcgattcaa ggtccgaatg 60

gaaggctccg tgaacggtca cgagtttgag attgagggag agggtgaagg ccgaccctac 120

gaaggcaccc agaccgcgaa gctgaaggtg accaagggtg gacccctgcc cttcgcctgg 180

gacattctgt ctcctcagtt tcagtacggt tctaaggtgt acgtgaagca ccctgctgac 240

attcccgact acaagaaact ttcctttccc gagggcttca agtgggagcg agttatgaac 300

ttcgaggatg gcggtgtcgt taccgttact caggactcct cgctccagga cggctcgttc 360

atctacaagg ttaagttcat cggtgtcaac ttccctagcg atggacccgt catgcaaaag 420

aaaactatgg gatgggaagc ctctacagag cggctgtacc ctcgagacgg agtgttgaag 480

ggcgagattc acaaggccct gaagctcaag gacggtggac actatctcgt tgagtttaag 540

tctatctaca tggcaaagaa acccgtgcag cttccaggct actattacgt cgattccaag 600

ctcgatatca ccagccataa tgaggactac actattgtcg aacagtacga gcgtgctgag 660

ggaagacacc atctgtttct ttaagcggcc gcgggg 696

<210> 128

<211> 1212

<212> DNA

<213> Artificial Sequence

<220><223> FBA1-dsREDexpress expression cassette

<400> 128

aaaccatcat ctaagggcct caaaactacc tcggaactgc tgcgctgatc tggacaccac 60

agaggttccg agcactttag gttgcaccaa atgtcccacc aggtgcaggc agaaaacgct 120

ggaacagcgt gtacagtttg tcttaacaaa aagtgagggc gctgaggtcg agcagggtgg 180

tgtgacttgt tatagccttt agagctgcga aagcgcgtat ggatttggct catcaggcca 240

gattgagggt ctgtggacac atgtcatgtt agtgtacttc aatcgccccc tggatatagc 300

cccgacaata ggccgtggcc tcatttttt gccttccgca catttccatt gctcggtacc 360

cacaccttgc ttctcctgca cttgccaacc ttaatactgg tttacattga ccaacatctt 420

acaagcgggg ggcttgtcta gggtatatat aaacagtggc tctcccaatc ggttgccagt 480

ctctttttc ctttctttcc ccacagattc gaaatctaaa ctacacatca caccatggcc 540

tcctcggagg acgtcatcaa ggagttcatg cgattcaagg tccgaatgga aggctccgtg 600

aacggtcacg agtttgagat tgagggagag ggtgaaggcc gaccctacga aggcacccag 660

accgcgaagc tgaaggtgac caagggtgga cccctgccct tcgcctggga cattctgtct 720

cctcagtttc agtacggttc taaggtgtac gtgaagcacc ctgctgacat tcccgactac 780

aagaaacttt cctttcccga gggcttcaag tgggagcgag ttatgaactt cgaggatggc 840

ggtgtcgtta ccgttactca ggactcctcg ctccaggacg gctcgttcat ctacaaggtt 900

aagttcatcg gtgtcaactt ccctagcgat ggacccgtca tgcaaaagaa aactatggga 960

tgggaagcct ctacagagcg gctgtaccct cgagacggag tgttgaaggg cgagattcac 1020

aaggccctga agctcaagga cggtggacac tatctcgttg agtttaagtc tatctacatg 1080

gcaaagaaac ccgtgcagct tccaggctac tattacgtcg attccaagct cgatatcacc 1140

agccataatg aggactacac tattgtcgaa cagtacgagc gtgctgaggg aagacaccat 1200

ctgtttcttt aa 1212

<210> 129

<211> 7244

<212> DNA

<213> Artificial Sequence

<220><223> pRF165

<400> 129

ggccgcaagt gtggatgggg aagtgagtgc ccggttctgt gtgcacaatt ggcaatccaa 60

gatggatgga ttcaacacag ggatatagcg agctacgtgg tggtgcgagg atatagcaac 120

ggatatttat gtttgacact tgagaatgta cgatacaagc actgtccaag tacaatacta 180

aacatactgt acatactcat actcgtaccc gggcaacggt ttcacttgag tgcagtggct 240

agtgctctta ctcgtacagt gtgcaatact gcgtatcata gtctttgatg tatatcgtat 300

tcattcatgt tagttgcgta cgagccggaa gcataaagtg taaagcctgg ggtgcctaat 360

gagtgagcta actcacatta attgcgttgc gctcactgcc cgctttccag tcgggaaacc 420

tgtcgtgcca gctgcattaa tgaatcggcc aacgcgcggg gagaggcggt ttgcgtattg 480

ggcgctcttc cgcttcctcg ctcactgact cgctgcgctc ggtcgttcgg ctgcggcgag 540

cggtatcagc tcactcaaag gcggtaatac ggttatccac agaatcaggg gataacgcag 600

gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc 660

tggcgttttt ccataggctc cgccccctg acgagcatca caaaaatcga cgctcaagtc 720

agaggtggcg aaacccgaca ggactataaa gataccaggc gtttccccct ggaagctccc 780

tcgtgcgctc tcctgttccg accctgccgc ttaccggata cctgtccgcc tttctccctt 840

cgggaagcgt ggcgctttct catagctcac gctgtaggta tctcagttcg gtgtaggtcg 900

ttcgctccaa gctgggctgt gtgcacgaac cccccgttca gcccgaccgc tgcgccttat 960

ccggtaacta tcgtcttgag tccaacccgg taagacacga cttatcgcca ctggcagcag 1020

ccactggtaa caggattagc agagcgaggt atgtaggcgg tgctacagag ttcttgaagt 1080

ggtggcctaa ctacggctac actagaagga cagtatttgg tatctgcgct ctgctgaagc 1140

cagttacctt cggaaaaaga gttggtagct cttgatccgg caaacaaacc accgctggta 1200

gcggtggttt ttttgtttgc aagcagcaga ttacgcgcag aaaaaaagga tctcaagaag 1260

atcctttgat cttttctacg gggtctgacg ctcagtggaa cgaaaactca cgttaaggga 1320

ttttggtcat gagattatca aaaaggatct tcacctagat ccttttaaat taaaaatgaa 1380

gttttaaatc aatctaaagt atatatgagt aaacttggtc tgacagttac caatgcttaa 1440

tcagtgaggc acctatctca gcgatctgtc tatttcgttc atccatagtt gcctgactcc 1500

ccgtcgtgta gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga 1560

taccgcgaga cccacgctca ccggctccag atttatcagc aataaaccag ccagccggaa 1620

gggccgagcg cagaagtggt cctgcaactt tatccgcctc catccagtct attaattgtt 1680

gccgggaagc tagagtaagt agttcgccag ttaatagttt gcgcaacgtt gttgccattg 1740

ctacaggcat cgtggtgtca cgctcgtcgt ttggtatggc ttcattcagc tccggttccc 1800

aacgatcaag gcgagttaca tgatccccca tgttgtgcaa aaaagcggtt agctccttcg 1860

gtcctccgat cgttgtcaga agtaagttgg ccgcagtgtt atcactcatg gttatggcag 1920

cactgcataa ttctcttact gtcatgccat ccgtaagatg cttttctgtg actggtgagt 1980

actcaaccaa gtcattctga gaatagtgta tgcggcgacc gagttgctct tgcccggcgt 2040

caatacggga taataccgcg ccacatagca gaactttaaa agtgctcatc attggaaaac 2100

gttcttcggg gcgaaaactc tcaaggatct taccgctgtt gagatccagt tcgatgtaac 2160

ccactcgtgc acccaactga tcttcagcat cttttacttt caccagcgtt tctgggtgag 2220

caaaaacagg aaggcaaaat gccgcaaaaa agggaataag ggcgacacgg aaatgttgaa 2280

tactcatact cttccttttt caatattatt gaagcattta tcagggttat tgtctcatga 2340

gcggatacat atttgaatgt atttagaaaa ataaacaaat aggggttccg cgcacatttc 2400

cccgaaaagt gccacctgac gcgccctgta gcggcgcatt aagcgcggcg ggtgtggtgg 2460

ttacgcgcag cgtgaccgct acacttgcca gcgccctagc gcccgctcct ttcgctttct 2520

tcccttcctt tctcgccacg ttcgccggct ttccccgtca agctctaaat cgggggctcc 2580

ctttagggtt ccgatttagt gctttacggc acctcgaccc caaaaaactt gattagggtg 2640

atggttcacg tagtgggcca tcgccctgat agacggtttt tcgccctttg acgttggagt 2700

ccacgttctt taatagtgga ctcttgttcc aaactggaac aacactcaac cctatctcgg 2760

tctattcttt tgatttataa gggattttgc cgatttcggc ctattggtta aaaaatgagc 2820

tgatttaaca aaaatttaac gcgaatttta acaaaatatt aacgcttaca atttccattc 2880

gccattcagg ctgcgcaact gttgggaagg gcgatcggtg cgggcctctt cgctattacg 2940

ccagctggcg aaaggggatg tgctgcaag gcgattaagt tgggtaacgc cagggttttc 3000

ccagtcacga cgttgtaaaa cgacggccag tgaattgtaa tacgactcac tatagggcga 3060

attgggtacc gggcccccccc tcgaggtcga tggtgtcgat aagcttgata tcgaattcat 3120

gtcacacaaa ccgatcttcg cctcaaggaa acctaattct acatccgaga gactgccgag 3180

atccagtcta cactgattaa ttttcgggcc aataatttaa aaaaatcgtg ttatataata 3240

ttatatgtat tatatatata catcatgatg atactgacag tcatgtccca ttgctaaata 3300

gacagactcc atctgccgcc tccaactgat gttctcaata tttaaggggt catctcgcat 3360

tgtttaataa taaacagact ccatctaccg cctccaaatg atgttctcaa aatatattgt 3420

atgaacttat ttttattact tagtattatt agacaactta cttgctttat gaaaaacact 3480

tcctatttag gaaacaattt ataatggcag ttcgttcatt taacaattta tgtagaataa 3540

atgttataaa tgcgtatggg aaatcttaaa tatggatagc ataaatgata tctgcattgc 3600

ctaattcgaa atcaacagca acgaaaaaaa tcccttgtac aacataaata gtcatcgaga 3660

aatatcaact atcaaagaac agctattcac acgttactat tgagattatt attggacgag 3720

aatcacacac tcaactgtct ttctctcttc tagaaataca ggtacaagta tgtactattc 3780

tcattgttca tacttctagt catttcatcc cacatattcc ttggatttct ctccaatgaa 3840

tgacattcta tcttgcaaat tcaacaatta taataagata taccaaagta gcggtatagt 3900

ggcaatcaaa aagcttctct ggtgtgcttc tcgtatttat ttttattcta atgatccatt 3960

aaaggtatat atttatttct tgttatataa tccttttgtt tattacatgg gctggataca 4020

taaaggtatt ttgatttaat tttttgctta aattcaatcc cccctcgttc agtgtcaact 4080

gtaatggtag gaaattacca tacttttgaa gaagcaaaaa aaatgaaaga aaaaaaaaat 4140

cgtatttcca ggttagacgt tccgcagaat ctagaatgcg gtatgcggta cattgttctt 4200

cgaacgtaaa agttgcgctc cctgagatat tgtacatttt tgcttttaca agtacaagta 4260

catcgtacaa ctatgtacta ctgttgatgc atccacaaca gtttgttttg ttttttttg 4320

ttttttttt ttctaatgat tcattaccgc tatgtatacc tacttgtact tgtagtaagc 4380

cgggttattg gcgttcaatt aatcatagac ttatgaatct gcacggtgtg cgctgcgagt 4440

tacttttagc ttatgcatgc tacttgggtg taatattggg atctgttcgg aaatcaacgg 4500

atgctcaatc gatttcgaca gtaattaatt aagtcataca caagtcagct ttcttcgagc 4560

ctcatataag tataagtagt tcaacgtatt agcactgtac ccagcatctc cgtatcgaga 4620

aacacaacaa catgccccat tggacagatc atgcggatac acaggttgtg cagtatcata 4680

catactcgat cagacaggtc gtctgaccat catacaagct gaacaagcgc tccatacttg 4740

cacgctctct atatacacag ttaaattaca tatccatagt ctaacctcta acagttaatc 4800

ttctggtaag cctcccagcc agccttctgg tatcgcttgg cctcctcaat aggatctcgg 4860

ttctggccgt acagacctcg gccgacaatt atgatatccg ttccggtaga catgacatcc 4920

tcaacagttc ggtactgctg tccgagagcg tctcccttgt cgtcaagacc caccccgggg 4980

gtcagaataa gccagtcctc agagtcgccc ttaggtcggt tctgggcaat gaagccaacc 5040

acaaactcgg ggtcggatcg ggcaagctca atggtctgct tggagtactc gccagtggcc 5100

agagagccct tgcaagacag ctcggccagc atgagcagac ctctggccag cttctcgttg 5160

ggagagggga ctaggaactc cttgtactgg gagttctcgt agtcagagac gtcctccttc 5220

ttctgttcag agacagtttc ctcggcacca gctcgcaggc cagcaatgat tccggttccg 5280

ggtacaccgt gggcgttggt gatatcggac cactcggcga ttcggtgaca ccggtactgg 5340

tgcttgacag tgttgccaat atctgcgaac tttctgtcct cgaacaggaa gaaaccgtgc 5400

ttaagagcaa gttccttgag ggggagcaca gtgccggcgt aggtgaagtc gtcaatgatg 5460

tcgatatggg ttttgatcat gcacacataa ggtccgacct tatcggcaag ctcaatgagc 5520

tccttggtgg tggtaacatc cagagaagca cacaggttgg ttttcttggc tgccacgagc 5580

ttgagcactc gagcggcaaa ggcggacttg tggacgttag ctcgagcttc gtaggagggc 5640

attttggtgg tgaagaggag actgaaataa atttagtctg cagaactttt tatcggaacc 5700

ttatctgggg cagtgaagta tatgttatgg taatagttac gagttagttg aacttataga 5760

tagactggac tatacggcta tcggtccaaa ttagaaagaa cgtcaatggc tctctgggcg 5820

tcgcctttgc cgacaaaaat gtgatcatga tgaaagccag caatgacgtt gcagctgata 5880

ttgttgtcgg ccaaccgcgc cgaaaacgca gctgtcagac ccacagcctc caacgaagaa 5940

tgtatcgtca aagtgatcca agcacactca tagttggagt cgtactccaa aggcggcaat 6000

gacgagtcag acagatactc gtcgacgttt aaaccatcat ctaagggcct caaaactacc 6060

tcggaactgc tgcgctgatc tggacaccac agaggttccg agcactttag gttgcaccaa 6120

atgtcccacc aggtgcaggc agaaaacgct ggaacagcgt gtacagtttg tcttaacaaa 6180

aagtgagggc gctgaggtcg agcagggtgg tgtgacttgt tatagccttt agagctgcga 6240
aagcgcgtat ggatttggct catcaggcca gattgagggt ctgtggacac atgtcatgtt 6300
agtgtacttc aatcgccccc tggatatagc cccgacaata ggccgtggcc tcatttttt 6360
gccttccgca catttccatt gctcggtacc cacaccttgc ttctcctgca cttgccaacc 6420
ttaatactgg tttacattga ccaacatctt acaagcgggg ggcttgtcta gggtatatat 6480
aaacagtggc tctcccaatc ggttgccagt ctctttttc ctttctttcc ccacagattc 6540
gaaatctaaa ctacacatca caccatggcc tcctcggagg acgtcatcaa ggagttcatg 6600

cgattcaagg tccgaatgga aggctccgtg aacggtcacg agtttgagat tgagggagag 6660
ggtgaaggcc gaccctacga aggcacccag accgcgaagc tgaaggtgac caagggtgga 6720
cccctgccct tcgcctggga cattctgtct cctcagtttc agtacggttc taaggtgtac 6780
gtgaagcacc ctgctgacat tcccgactac aagaaacttt cctttcccga gggcttcaag 6840
tgggagcgag ttatgaactt cgaggatggc ggtgtcgtta ccgttactca ggactcctcg 6900
ctccaggacg gctcgttcat ctacaaggtt aagttcatcg gtgtcaactt ccctagcgat 6960
ggacccgtca tgcaaaagaa aactatggga tgggaagcct ctacagagcg gctgtaccct 7020

cgagacggag tgttgaaggg cgagattcac aaggccctga agctcaagga cggtggacac 7080
tatctcgttg agtttaagtc tatctacatg gcaaagaaac ccgtgcagct tccaggctac 7140
tattacgtcg attccaagct cgatatcacc agccataatg aggactacac tattgtcgaa 7200
cagtacgagc gtgctgaggg aagacaccat ctgtttcttt aagc 7244

<210> 130
<211> 1214
<212> DNA
<213> Artificial sequence
<220><223> FBA1 Yarrowia dsREDexpress cassette from pRF165 on PmeI NotI fragment
<400> 130
aaaccatcat ctaagggcct caaaactacc tcggaactgc tgcgctgatc tggacaccac 60

agaggttccg agcactttag gttgcaccaa atgtcccacc aggtgcaggc agaaaacgct 120
ggaacagcgt gtacagtttg tcttaacaaa aagtgagggc gctgaggtcg agcagggtgg 180
tgtgacttgt tatagccttt agagctgcga aagcgcgtat ggatttggct catcaggcca 240
gattgagggt ctgtggacac atgtcatgtt agtgtacttc aatcgccccc tggatatagc 300

cccgacaata ggccgtggcc tcatttttt gccttccgca catttccatt gctcggtacc 360

cacaccttgc ttctcctgca cttgccaacc ttaatactgg tttacattga ccaacatctt 420

acaagcgggg ggcttgtcta gggtatatat aaacagtggc tctcccaatc ggttgccagt 480

ctctttttc ctttctttcc ccacagattc gaaatctaaa ctacacatca caccatggcc 540

tcctcggagg acgtcatcaa ggagttcatg cgattcaagg tccgaatgga aggctccgtg 600

aacggtcacg agtttgagat tgagggagag ggtgaaggcc gaccctacga aggcacccag 660

accgcgaagc tgaaggtgac caagggtgga cccctgccct tcgcctggga cattctgtct 720

cctcagtttc agtacggttc taaggtgtac gtgaagcacc ctgctgacat tcccgactac 780

aagaaacttt cctttcccga gggcttcaag tgggagcgag ttatgaactt cgaggatggc 840

ggtgtcgtta ccgttactca ggactcctcg ctccaggacg gctcgttcat ctacaaggtt 900

aagttcatcg gtgtcaactt ccctagcgat ggacccgtca tgcaaaagaa aactatggga 960

tgggaagcct ctacagagcg gctgtaccct cgagacggag tgttgaaggg cgagattcac 1020

aaggccctga agctcaagga cggtggacac tatctcgttg agtttaagtc tatctacatg 1080

gcaaagaaac ccgtgcagct tccaggctac tattacgtcg attccaagct cgatatcacc 1140

agccataatg aggactacac tattgtcgaa cagtacgagc gtgctgaggg aagacaccat 1200

ctgtttcttt aagc 1214

<210> 131

<211> 11526

<212> DNA

<213> Artificial sequence

<220><223> p2P069 integration vector

<400> 131

ggccgcaagt gtggatgggg aagtgagtgc ccggttctgt gtgcacaatt ggcaatccaa 60

gatggatgga ttcaacacag ggatatagcg agctacgtgg tggtgcgagg atatagcaac 120

ggatatttat gtttgacact tgagaatgta cgatacaagc actgtccaag tacaatacta 180

aacatactgt acatactcat actcgtaccc gggcaacggt ttcacttgag tgcagtggct 240

agtgctctta ctcgtacagt gtgcaatact gcgtatcata gtctttgatg tatatcgtat 300

tcattcatgt tagttgcgta cgggcgtcgt tgcttgtgtg atttttgagg acccatccct 360

ttggtatata agtatactct ggggttaagg ttgcccgtgt agtctaggtt atagttttca 420

tgtgaaatac cgagagccga gggagaataa acgggggtat ttggacttgt ttttttcgcg 480

gaaaagcgtc gaatcaaccc tgcgggcctt gcaccatgtc cacgacgtgt ttctcgcccc 540

aattcgcccc ttgcacgtca aaattaggcc tccatctaga ccccctccata acatgtgact 600

gtggggaaaa gtataaggga aaccatgcaa ccatagacga cgtgaaagac ggggaggaac 660

caatggaggc caaagaaatg gggtagcaac agtccaggag acagacaagg agacaaggag 720

agggcgcccg aaagatcgga aaaacaaaca tgtccaattg gggcagtgac ggaaacgaca 780

cggacacttc agtacaatgg accgaccatc tccaagccag ggttattccg gtatcacctt 840

ggccgtaacc tcccgctggt acctgatatt gtacacgttc acattcaata tactttcagc 900

tacaataaga gaggctgttt gtcgggcatg tgtgtccgtc gtatggggtg atgtccgagg 960

gcgaaattcg ctacaagctt aactctggcg cttgtccagt atgaatagac aagtcaagac 1020

cagtggtgcc atgattgaca gggaggtaca agacttcgat actcgagcat tactcggact 1080

tgtggcgatt gaacagacgg gcgatcgctt ctcccccgta ttgccggcgc gccagctgca 1140

ttaatgaatc ggccaacgcg cggggagagg cggtttgcgt attgggcgct cttccgcttc 1200

ctcgctcact gactcgctgc gctcggtcgt tcggctgcgg cgagcggtat cagctcactc 1260

aaaggcggta atacggttat ccacagaatc aggggataac gcaggaaaga acatgtgagc 1320

aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag 1380

gctccgcccc cctgacgagc atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc 1440

gacaggacta taaagatacc aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt 1500

tccgaccctg ccgcttaccg gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct 1560

ttctcatagc tcacgctgta ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg 1620

ctgtgtgcac gaacccccccg ttcagcccga ccgctgcgcc ttatccggta actatcgtct 1680

tgagtccaac ccggtaagac acgacttatc gccactggca gcagccactg gtaacaggat 1740

tagcagagcg aggtatgtag gcggtgctac agagttcttg aagtggtggc ctaactacgg 1800

ctacactaga agaacagtat ttggtatctg cgctctgctg aagccagtta ccttcggaaa 1860

aagagttggt agctcttgat ccggcaaaca aaccaccgct ggtagcggtg gtttttttgt 1920

ttgcaagcag cagattacgc gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc 1980

tacggggtct gacgctcagt ggaacgaaaa ctcacgttaa gggattttgg tcatgagatt 2040

atcaaaaagg atcttcacct agatcctttt aaattaaaaa tgaagtttta aatcaatcta 2100

aagtatatat gagtaaactt ggtctgacag ttaccaatgc ttaatcagtg aggcacctat 2160

ctcagcgatc tgtctatttc gttcatccat agttgcctga ctccccgtcg tgtagataac 2220

tacgatacgg gagggcttac catctggccc cagtgctgca atgataccgc gagacccacg 2280

ctcaccggct ccagatttat cagcaataaa ccagccagcc ggaagggccg agcgcagaag 2340

tggtcctgca actttatccg cctccatcca gtctattaat tgttgccggg aagctagagt 2400

aagtagttcg ccagttaata gtttgcgcaa cgttgttgcc attgctacag gcatcgtggt 2460

gtcacgctcg tcgtttggta tggcttcatt cagctccggt tcccaacgat caaggcgagt 2520
tacatgatcc cccatgttgt gcaaaaaagc ggttagctcc ttcggtcctc cgatcgttgt 2580
cagaagtaag ttggccgcag tgttatcact catggttatg gcagcactgc ataattctct 2640
tactgtcatg ccatccgtaa gatgcttttc tgtgactggt gagtactcaa ccaagtcatt 2700
ctgagaatag tgtatgcggc gaccgagttg ctcttgcccg gcgtcaatac gggataatac 2760
cgcgccacat agcagaactt taaaagtgct catcattgga aaacgttctt cggggcgaaa 2820
actctcaagg atcttaccgc tgttgagatc cagttcgatg taacccactc gtgcacccaa 2880

ctgatcttca gcatctttta ctttcaccag cgtttctggg tgagcaaaaa caggaaggca 2940
aaatgccgca aaaaagggaa taagggcgac acggaaatgt tgaatactca tactcttcct 3000
ttttcaatat tattgaagca tttatcaggg ttattgtctc atgagcggat acatatttga 3060
atgtatttag aaaaataaac aaataggggt tccgcgcaca tttccccgaa aagtgccacc 3120
tgatgcggtg tgaaataccg cacagatgcg taaggagaaa ataccgcatc aggaaattgt 3180
aagcgttaat attttgttaa aattcgcgtt aaatttttgt taaatcagct catttttaa 3240
ccaataggcc gaaatcggca aaatccctta taaatcaaaa gaatagaccg agatagggtt 3300

gagtgttgtt ccagtttgga acaagagtcc actattaaag aacgtggact ccaacgtcaa 3360
agggcgaaaa accgtctatc agggcgatgg cccactacgt gaaccatcac cctaatcaag 3420
ttttttgggg tcgaggtgcc gtaaagcact aaatcggaac cctaaaggga gcccccgatt 3480
tagagcttga cggggaaagc cggcgaacgt ggcgagaaag gaagggaaga aagcgaaagg 3540
agcgggcgct agggcgctgg caagtgtagc ggtcacgctg cgcgtaacca ccacacccgc 3600
cgcgcttaat gcgccgctac agggcgcgtc cattcgccat tcaggctgcg caactgttgg 3660
gaagggcgat cggtgcgggc ctcttcgcta ttacgccagc tggcgaaagg gggatgtgct 3720

gcaaggcgat taagttgggt aacgccaggg ttttcccagt cacgacgttg taaaacgacg 3780
gccagtgaat tgtaatacga ctcactatag ggcgaattgg gcccgacgtc gcatgcgctg 3840
atgacacttt ggtctgaaag agatgcattt tgaatcccaa acttgcagtg cccaagtgac 3900
atacatctcc gcgttttgga aaatgttcag aaacagttga ttgtgttgga atggggaatg 3960
gggaatggaa aaatgactca agtatcaatt ccaaaaactt ctctggctgg cagtacctac 4020
tgtccatact actgcatttt ctccagtcag gccactctat actcgacgac acagtagtaa 4080
aacccagata atttcgacat aaacaagaaa acagacccaa taatatttat atatagtcag 4140

ccgtttgtcc agttcagact gtaatagccg aaaaaaaatc caaagtttct attctaggaa 4200

aatatattcc aatattttta attcttaatc tcatttattt tattctagcg aaatacattt 4260

cagctacttg agacatgtga tacccacaaa tcggattcgg actcggttgt tcagaagagc 4320

atatggcatt cgtgctcgct tgttcacgta ttcttcctgt tccatctctt ggccgacaat 4380

cacacaaaaa tggggttttt tttttaattc taatgattca ttacagcaaa attgagatat 4440

agcagaccac gtattccata atcaccaagg aagttcttgg gcgtcttaat taagttgcga 4500

cacatgtctt gatagtatct tggcttctct ctcttgagct tttccataac aagttcttct 4560

gcctccagga agtccatggt gaatgattct tatactcaga aggaaatgct taacgatttc 4620

gggtgtgagt tgacaaggag agagagaaaa gaagaggaaa ggtaattcgg ggacggtggt 4680

cttttatacc cttggctaaa gtcccaacca caaagcaaaa aaattttcag tagtctattt 4740

tgcgtccggc atgggttacc cggatggcca gacaaagaaa ctagtacaaa gtctgaacaa 4800

gcgtagattc cagactgcag taccctacgc ccttaacggc aagtgtggga accgggggag 4860

gtttgatatg tggggtgaag ggggctctcg ccggggttgg gcccgctact gggtcaattt 4920

ggggtcaatt ggggcaattg gggctgtttt ttgggacaca aatacgccgc caacccggtc 4980

tctcctgaat tctgcagatg ggctgcagga attccgtcgt cgcctgagtc gacatcattt 5040

atttaccagt tggccacaaa cccttgacga tctcgtatgt cccctccgac atactcccgg 5100

ccggctgggg tacgttcgat agcgctatcg gcatcgacaa ggtttgggtc cctagccgat 5160

accgcactac ctgagtcaca atcttcggag gtttagtctt ccacatagca cgggcaaaag 5220

tgcgtatata tacaagagcg tttgccagcc acagattttc actccacaca ccacatcaca 5280

catacaacca cacacatcca caatggaacc cgaaactaag aagaccaaga ctgactccaa 5340

gaagattgtt cttctcggcg gcgacttctg tggccccgag gtgattgccg aggccgtcaa 5400

ggtgctcaag tctgttgctg aggcctccgg caccgagttt gtgtttgagg accgactcat 5460

tggaggagct gccattgaga aggagggcga gcccatcacc gacgctactc tcgacatctg 5520

ccgaaaggct gactctatta tgctcggtgc tgtcggaggc gctgccaaca ccgtatggac 5580

cactcccgac ggacgaaccg acgtgcgacc cgagcagggt ctcctcaagc tgcgaaagga 5640

cctgaacctg tacgccaacc tgcgaccctg ccagctgctg tcgcccaagc tcgccgatct 5700

ctcccccatc cgaaacgttg agggcaccga cttcatcatt gtccgagagc tcgtcggagg 5760

tatctacttt ggagagcgaa aggaggatga cggatctggc gtcgcttccg acaccgagac 5820

ctactccgtt cctgaggttg agcgaattgc ccgaatggcc gccttcctgg cccttcagca 5880

caacccccct cttcccgtgt ggtctcttga caaggccaac gtgctggcct cctctcgact 5940

ttggcgaaag actgtcactc gagtcctcaa ggacgaattc ccccagctcg agctcaacca 6000

ccagctgatc gactcggccg ccatgatcct catcaagcag ccctccaaga tgaatggtat 6060

catcatcacc accaacatgt ttggcgatat catctccgac gaggcctccg tcatccccgg 6120

ttctctgggt ctgctgccct ccgcctctct ggcttctctg cccgacacca acgaggcgtt 6180

cggtctgtac gagccctgtc acggatctgc ccccgatctc ggcaagcaga aggtcaaccc 6240

cattgccacc attctgtctg ccgccatgat gctcaagttc tctcttaaca tgaagcccgc 6300

cggtgacgct gttgaggctg ccgtcaagga gtccgtcgag gctggtatca ctaccgccga 6360

tatcggaggc tcttcctcca cctccgaggt cggagacttg ttgccaacaa ggtcaaggag 6420

ctgctcaaga aggagtaagt cgtttctacg acgcattgat ggaaggagca aactgacgcg 6480

cctgcgggtt ggtctaccgg cagggtccgc tagtgtataa gactctataa aaagggccct 6540

gccctgctaa tgaaatgatg atttataatt taccggtgta gcaaccttga ctagaagaag 6600

cagattgggt gtgtttgtag tggaggacag tggtacgttt tggaaacagt cttcttgaaa 6660

gtgtcttgtc tacagtatat tcactcataa cctcaatagc caagggtgta gtcggtttat 6720

taaaggaagg gagttgtggc tgatgtggat agatatcttt aagctggcga ctgcacccaa 6780

cgagtgtggt ggtagcttgt tagatctgta tattcggtaa gatatatttt gtggggtttt 6840

agtggtgttt aaaccatcat ctaagggcct caaaactacc tcggaactgc tgcgctgatc 6900

tggacaccac agaggttccg agcactttag gttgcaccaa atgtcccacc aggtgcaggc 6960

agaaaacgct ggaacagcgt gtacagtttg tcttaacaaa aagtgagggc gctgaggtcg 7020

agcagggtgg tgtgacttgt tatagccttt agagctgcga aagcgcgtat ggatttggct 7080

catcaggcca gattgagggt ctgtggacac atgtcatgtt agtgtacttc aatcgccccc 7140

tggatatagc cccgacaata ggccgtggcc tcatttttt gccttccgca catttccatt 7200

gctcggtacc cacaccttgc ttctcctgca cttgccaacc ttaatactgg tttacattga 7260

ccaacatctt acaagcgggg ggcttgtcta gggtatatat aaacagtggc tctcccaatc 7320

ggttgccagt ctctttttc ctttctttcc ccacagattc gaaatctaaa ctacacatca 7380

caccatggac aagaaatact ccatcggcct ggacattgga accaactctg tcggctgggc 7440

tgtcatcacc gacgagtaca aggtgccctc caagaaattc aaggtcctcg gaaacaccga 7500

tcgacactcc atcaagaaaa acctcattgg tgccctgttg ttcgattctg gcgagactgc 7560

cgaagctacc agactcaagc gaactgctcg gcgacgttac acccgacgga agaaccgaat 7620

ctgctacctg caggagatct tttccaacga gatggccaag gtggacgatt cgttctttca 7680

tcgactggag gaatccttcc tcgtcgagga agacaagaaa cacgagcgtc atcccatctt 7740

tggcaacatt gtggacgagg ttgcttacca cgagaagtat cctaccatct accacctgcg 7800

aaagaaactc gtcgattcca ccgacaaggc ggatctcaga cttatctacc tcgctctggc 7860

acacatgatc aagtttcgag gtcatttcct catcgagggc gatctcaatc ccgacaacag 7920

cgatgtggac aagctgttca ttcagctcgt tcagacctac aaccagctgt tcgaggaaaa 7980
ccccatcaat gcctccggag tcgatgcaaa ggccatcttg tctgctcgac tctcgaagag 8040
cagacgactg gagaacctca ttgcccaact tcctggcgag aaaaagaacg gactgtttgg 8100
caacctcatt gccctttctc ttggtctcac acccaacttc aagtccaact tcgatctggc 8160
ggaggacgcc aagctccagc tgtccaagga cacctacgac gatgacctcg acaacctgct 8220
tgcacagatt ggcgatcagt acgccgacct gtttctcgct gccaagaacc tttcggatgc 8280
tattctcttg tctgacattc tgcgagtcaa caccgagatc acaaaggctc ccctttctgc 8340

ctccatgatc aagcgatacg acgagcacca tcaggatctc acactgctca aggctcttgt 8400
ccgacagcaa ctgcccgaga agtacaagga gatctttttc gatcagtcga agaacggcta 8460
cgctggatac atcgacggcg gagcctctca ggaagagttc tacaagttca tcaagccaat 8520
tctcgagaag atggacggaa ccgaggaact gcttgtcaag ctcaatcgag aggatctgct 8580
tcggaagcaa cgaaccttcg acaacggcag cattcctcat cagatccacc tcggtgagct 8640
gcacgccatt cttcgacgtc aggaagactt ctaccccttt ctcaaggaca accgagagaa 8700
gatcgagaag attcttacct ttcgaatccc ctactatgtt ggtcctcttg ccagaggaaa 8760

ctctcgattt gcttggatga ctcgaaagtc cgaggaaacc atcactccct ggaacttcga 8820
ggaagtcgtg gacaagggtg cctctgcaca gtccttcatc gagcgaatga ccaacttcga 8880
caagaatctg cccaacgaga aggttcttcc caagcattcg ctgctctacg agtactttac 8940
agtctacaac gaactcacca aagtcaagta cgttaccgag ggaatgcgaa agcctgcctt 9000
cttgtctggc gaacagaaga aagccattgt cgatctcctg ttcaagacca accgaaaggt 9060
cactgttaag cagctcaagg aggactactt caagaaaatc gagtgtttcg acagcgtcga 9120
gatttccgga gttgaggacc gattcaacgc ctctttgggc acctatcacg atctgctcaa 9180

gattatcaag gacaaggatt ttctcgacaa cgaggaaaac gaggacattc tggaggacat 9240
cgtgctcact cttaccctgt tcgaagatcg ggagatgatc gaggaacgac tcaagacata 9300
cgctcacctg ttcgacgaca aggtcatgaa acaactcaag cgacgtagat acaccggctg 9360
gggaagactt tcgcgaaagc tcatcaacgg catcagagac aagcagtccg gaaagaccat 9420
tctggacttt ctcaagtccg atggctttgc caaccgaaac ttcatgcagc tcattcacga 9480
cgattctctt accttcaagg aggacatcca gaaggcacaa gtgtccggtc agggcgacag 9540
cttgcacgaa catattgcca acctggctgg ttcgccagcc atcaagaaag gcattctcca 9600

gactgtcaag gttgtcgacg agctggtgaa ggtcatggga cgtcacaagc ccgagaacat 9660

tgtgatcgag atggccagag agaaccagac aactcaaaag ggtcagaaaa actcgcgaga 9720

gcggatgaag cgaatcgagg aaggcatcaa ggagctggga tcccagattc tcaaggagca 9780

tcccgtcgag aacactcaac tgcagaacga gaagctgtat ctctactatc tgcagaatgg 9840

tcgagacatg tacgtggatc aggaactgga catcaatcgt ctcagcgact acgatgtgga 9900

ccacattgtc cctcaatcct ttctcaagga cgattctatc gacaacaagg tccttacacg 9960

atccgacaag aacagaggca agtcggacaa cgttcccagc gaagaggtgg tcaaaaagat 10020

gaagaactac tggcgacagc tgctcaacgc caagctcatt acccagcgaa agttcgacaa 10080

tcttaccaag gccgagcgag gcggtctgtc cgagctcgac aaggctggct tcatcaagcg 10140

tcaactcgtc gagaccagac agatcacaaa gcacgtcgca cagattctcg attctcggat 10200

gaacaccaag tacgacgaga acgacaagct catccgagag gtcaaggtga ttactctcaa 10260

gtccaaactg gtctccgatt tccgaaagga ctttcagttc tacaaggtgc gagagatcaa 10320

caattaccac catgcccacg atgcttacct caacgccgtc gttggcactg cgctcatcaa 10380

gaaatacccc aagctcgaaa gcgagttcgt ttacggcgat tacaaggtct acgacgttcg 10440

aaagatgatt gccaagtccg aacaggagat tggcaaggct actgccaagt acttctttta 10500

ctccaacatc atgaactttt tcaagaccga gatcaccttg gccaacggag agattcgaaa 10560

gagaccactt atcgagacca acggcgaaac tggagagatc gtgtgggaca agggtcgaga 10620

ctttgcaacc gtgcgaaagg ttctgtcgat gcctcaggtc aacatcgtca agaaaaccga 10680

ggttcagact ggcggattct ccaaggagtc gattctgccc aagcgaaact ccgacaagct 10740

catcgctcga aagaaagact gggatcccaa gaaatacggt ggcttcgatt ctcctaccgt 10800

cgcctattcc gtgcttgtcg ttgcgaaggt cgagaagggc aagtccaaaa agctcaagtc 10860

cgtcaaggag ctgctcggaa ttaccatcat ggagcgatcg agcttcgaga agaatcccat 10920

cgacttcttg gaagccaagg gttacaagga ggtcaagaaa gacctcatta tcaagctgcc 10980

caagtactct ctgttcgaac tggagaacgg tcgaaagcgt atgctcgcct ccgctggcga 11040

gctgcagaag ggaaacgagc ttgccttgcc ttcgaagtac gtcaactttc tctatctggc 11100

ttctcactac gagaagctca agggttctcc cgaggacaac gaacagaagc aactcttcgt 11160

tgagcagcac aaacattacc tcgacgagat tatcgagcag atttccgagt tttcgaagcg 11220

agtcatcctg gctgatgcca acttggacaa ggtgctctct gcctacaaca agcatcggga 11280

caaacccatt cgagaacagg cggagaacat cattcacctg tttactctta ccaacctggg 11340

tgctcctgca gctttcaagt acttcgatac cactatcgac cgaaagcggt acacatccac 11400

caaggaggtt ctcgatgcca ccctgattca ccagtccatc actggcctgt acgagacccg 11460

aatcgacctg tctcagcttg gtggcgactc cagagccgat cccaagaaaa agcgaaaggt 11520

ctaagc 11526

<210> 132

<211> 8064

<212> DNA

<213> artificial sequence

<220><223> pRF201

<400> 132

aaaccatcat ctaagggcct caaaactacc tcggaactgc tgcgctgatc tggacaccac 60

agaggttccg agcactttag gttgcaccaa atgtcccacc aggtgcaggc agaaaacgct 120

ggaacagcgt gtacagtttg tcttaacaaa aagtgagggc gctgaggtcg agcagggtgg 180

tgtgacttgt tatagccttt agagctgcga aagcgcgtat ggatttggct catcaggcca 240

gattgagggt ctgtggacac atgtcatgtt agtgtacttc aatcgccccc tggatatagc 300

cccgacaata ggccgtggcc tcatttttt gccttccgca catttccatt gctcggtacc 360

cacaccttgc ttctcctgca cttgccaacc ttaatactgg tttacattga ccaacatctt 420

acaagcgggg ggcttgtcta gggtatatat aaacagtggc tctcccaatc ggttgccagt 480

ctctttttc ctttctttcc ccacagattc gaaatctaaa ctacacatca caccatggcc 540

tcctcggagg acgtcatcaa ggagttcatg cgattcaagg tccgaatgga aggctccgtg 600

aacggtcacg agtttgagat tgagggagag ggtgaaggcc gaccctacga aggcacccag 660

accgcgaagc tgaaggtgac caagggtgga cccctgccct tcgcctggga cattctgtct 720

cctcagtttc agtacggttc taaggtgtac gtgaagcacc ctgctgacat tcccgactac 780

aagaaacttt cctttcccga gggcttcaag tgggagcgag ttatgaactt cgaggatggc 840

ggtgtcgtta ccgttactca ggactcctcg ctccaggacg gctcgttcat ctacaaggtt 900

aagttcatcg gtgtcaactt ccctagcgat ggacccgtca tgcaaaagaa aactatggga 960

tgggaagcct ctacagagcg gctgtaccct cgagacggag tgttgaaggg cgagattcac 1020

aaggccctga agctcaagga cggtggacac tatctcgttg agtttaagtc tatctacatg 1080

gcaaagaaac ccgtgcagct tccaggctac tattacgtcg attccaagct cgatatcacc 1140

agccataatg aggactacac tattgtcgaa cagtacgagc gtgctgaggg aagacaccat 1200

ctgtttcttt aagcggccgc aagtgtggat ggggaagtga gtgcccggtt ctgtgtgcac 1260

aattggcaat ccaagatgga tggattcaac acagggatat agcgagctac gtggtggtgc 1320

gaggatatag caacggatat ttatgtttga cacttgagaa tgtacgatac aagcactgtc 1380

caagtacaat actaaacata ctgtacatac tcatactcgt acccgggcaa cggtttcact 1440

tgagtgcagt ggctagtgct cttactcgta cagtgtgcaa tactgcgtat catagtcttt 1500

gatgtatatc gtattcattc atgttagttg cgtacgggcg tcgttgcttg tgtgattttt 1560

gaggacccat ccctttggta tataagtata ctctggggtt aaggttgccc gtgtagtcta 1620

ggttatagtt ttcatgtgaa ataccgagag ccgagggaga ataaacgggg gtatttggac 1680

ttgtttttttt cgcggaaaag cgtcgaatca accctgcggg ccttgcacca tgtccacgac 1740

gtgtttctcg ccccaattcg ccccttgcac gtcaaaatta ggcctccatc tagacccctc 1800

cataacatgt gactgtgggg aaaagtataa gggaaaccat gcaaccatag acgacgtgaa 1860

agacggggag gaaccaatgg aggccaaaga aatggggtag caacagtcca ggagacagac 1920

aaggagacaa ggagagggcg cccgaaagat cggaaaaaca aacatgtcca attggggcag 1980

tgacggaaac gacacggaca cttcagtaca atggaccgac catctccaag ccagggttat 2040

tccggtatca ccttggccgt aacctcccgc tggtacctga tattgtacac gttcacattc 2100

aatatacttt cagctacaat aagagaggct gtttgtcggg catgtgtgtc cgtcgtatgg 2160

ggtgatgtcc gagggcgaaa ttcgctacaa gcttaactct ggcgcttgtc cagtatgaat 2220

agacaagtca agaccagtgg tgccatgatt gacagggagg tacaagactt cgatactcga 2280

gcattactcg gacttgtggc gattgaacag acgggcgatc gcttctcccc cgtattgccg 2340

gcgcgccagc tgcattaatg aatcggccaa cgcgcgggga gaggcggttt gcgtattggg 2400

cgctcttccg cttcctcgct cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg 2460

gtatcagctc actcaaaggc ggtaatacgg ttatccacag aatcagggga taacgcagga 2520

aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg 2580

gcgtttttcc ataggctccg ccccccctgac gagcatcaca aaaatcgacg ctcaagtcag 2640

aggtggcgaa acccgacagg actataaaga taccaggcgt ttcccccctgg aagctccctc 2700

gtgcgctctc ctgttccgac cctgccgctt accggatacc tgtccgcctt tctcccttcg 2760

ggaagcgtgg cgctttctca tagctcacgc tgtaggtatc tcagttcggt gtaggtcgtt 2820

cgctccaagc tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc 2880

ggtaactatc gtcttgagtc caacccggta agacacgact tatcgccact ggcagcagcc 2940

actggtaaca ggattagcag agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg 3000

tggcctaact acggctacac tagaagaaca gtatttggta tctgcgctct gctgaagcca 3060

gttaccttcg gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc 3120

ggtggttttt ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat 3180

cctttgatct tttctacggg gtctgacgct cagtggaacg aaaactcacg ttaagggatt 3240

ttggtcatga gattatcaaa aaggatcttc acctagatcc ttttaaatta aaaatgaagt 3300

tttaaatcaa tctaaagtat atatgagtaa acttggtctg acagttacca atgcttaatc 3360

agtgaggcac ctatctcagc gatctgtcta tttcgttcat ccatagttgc ctgactcccc 3420

gtcgtgtaga taactacgat acgggagggc ttaccatctg gccccagtgc tgcaatgata 3480

ccgcgagacc cacgctcacc ggctccagat ttatcagcaa taaaccagcc agccggaagg 3540

gccgagcgca gaagtggtcc tgcaacttta tccgcctcca tccagtctat taattgttgc 3600

cgggaagcta gagtaagtag ttcgccagtt aatagtttgc gcaacgttgt tgccattgct 3660

acaggcatcg tggtgtcacg ctcgtcgttt ggtatggctt cattcagctc cggttcccaa 3720

cgatcaaggc gagttacatg atcccccatg ttgtgcaaaa aagcggttag ctccttcggt 3780

cctccgatcg ttgtcagaag taagttggcc gcagtgttat cactcatggt tatggcagca 3840

ctgcataatt ctcttactgt catgccatcc gtaagatgct tttctgtgac tggtgagtac 3900

tcaaccaagt cattctgaga atagtgtatg cggcgaccga gttgctcttg cccggcgtca 3960

atacgggata ataccgcgcc acatagcaga actttaaaag tgctcatcat tggaaaacgt 4020

tcttcggggc gaaaactctc aaggatctta ccgctgttga gatccagttc gatgtaaccc 4080

actcgtgcac ccaactgatc ttcagcatct tttactttca ccagcgtttc tgggtgagca 4140

aaaacaggaa ggcaaaatgc cgcaaaaaag ggaataaggg cgacacggaa atgttgaata 4200

ctcatactct tccttttca atattattga agcatttatc agggttattg tctcatgagc 4260

ggatacatat ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc 4320

cgaaaagtgc cacctgatgc ggtgtgaaat accgcacaga tgcgtaagga gaaaataccg 4380

catcaggaaa ttgtaagcgt taatattttg ttaaaattcg cgttaaattt ttgttaaatc 4440

agctcatttt ttaaccaata ggccgaaatc ggcaaaatcc cttataaatc aaaagaatag 4500

accgagatag ggttgagtgt tgttccagtt tggaacaaga gtccactatt aaagaacgtg 4560

gactccaacg tcaaagggcg aaaaaccgtc tatcagggcg atggcccact acgtgaacca 4620

tcaccctaat caagtttttt ggggtcgagg tgccgtaaag cactaaatcg gaaccctaaa 4680

gggagccccc gatttagagc ttgacgggga aagccggcga acgtggcgag aaaggaaggg 4740

aagaaagcga aaggagcggg cgctagggcg ctggcaagtg tagcggtcac gctgcgcgta 4800

accaccacac ccgccgcgct taatgcgccg ctacagggcg cgtccattcg ccattcaggc 4860

tgcgcaactg ttgggaaggg cgatcggtgc gggcctcttc gctattacgc cagctggcga 4920

aaggggatg tgctgcaagg cgattaagtt gggtaacgcc agggttttcc cagtcacgac 4980

gttgtaaaac gacggccagt gaattgtaat acgactcact atagggcgaa ttgggcccga 5040

cgtcgcatgc gctgatgaca ctttggtctg aaagagatgc attttgaatc ccaaacttgc 5100

agtgcccaag tgacatacat ctccgcgttt tggaaaatgt tcagaaacag ttgattgtgt 5160
tggaatgggg aatggggaat ggaaaaatga ctcaagtatc aattccaaaa acttctctgg 5220
ctggcagtac ctactgtcca tactactgca ttttctccag tcaggccact ctatactcga 5280
cgacacagta gtaaaaccca gataatttcg acataaacaa gaaaacagac ccaataatat 5340
ttatatatag tcagccgttt gtccagttca gactgtaata gccgaaaaaa aatccaaagt 5400
ttctattcta ggaaaatata ttccaatatt tttaattctt aatctcattt attttattct 5460
agcgaaatac atttcagcta cttgagacat gtgataccca caaatcggat tcggactcgg 5520

ttgttcagaa gagcatatgg cattcgtgct cgcttgttca cgtattcttc ctgttccatc 5580
tcttggccga caatcacaca aaaatggggt ttttttttta attctaatga ttcattacag 5640
caaaattgag atatagcaga ccacgtattc cataatcacc aaggaagttc ttgggcgtct 5700
taattaagtt gcgacacatg tcttgatagt atcttggctt ctctctcttg agcttttcca 5760
taacaagttc ttctgcctcc aggaagtcca tggtgaatga ttcttatact cagaaggaaa 5820
tgcttaacga tttcgggtgt gagttgacaa ggagagagag aaaagaagag gaaaggtaat 5880
tcggggacgg tggtctttta taccccttggc taaagtccca accacaaagc aaaaaaattt 5940

tcagtagtct attttgcgtc cggcatgggt tacccggatg gccagacaaa gaaactagta 6000
caaagtctga acaagcgtag attccagact gcagtaccct acgcccttaa cggcaagtgt 6060
gggaaccggg ggaggtttga tatgtggggt gaaggggggct ctcgccgggg ttgggcccgc 6120
tactgggtca atttggggtc aattggggca attggggctg ttttttggga cacaaatacg 6180
ccgccaaccc ggtctctcct gaattctgca gatgggctgc aggaattccg tcgtcgcctg 6240
agtcgacatc atttatttac cagttggcca caaacccttg acgatctcgt atgtcccctc 6300
cgacatactc ccggccggct ggggtacgtt cgatagcgct atcggcatcg acaaggtttg 6360

ggtccctagc cgataccgca ctacctgagt cacaatcttc ggaggtttag tcttccacat 6420
agcacgggca aaagtgcgta tatatacaag agcgtttgcc agccacagat tttcactcca 6480
cacaccacat cacacataca accacacaca tccacaatgg aacccgaaac taagaagacc 6540
aagactgact ccaagaagat tgttcttctc ggcggcgact tctgtggccc cgaggtgatt 6600
gccgaggccg tcaaggtgct caagtctgtt gctgaggcct ccggcaccga gtttgtgttt 6660
gaggaccgac tcattggagg agctgccatt gagaaggagg gcgagcccat caccgacgct 6720
actctcgaca tctgccgaaa ggctgactct attatgctcg gtgctgtcgg aggcgctgcc 6780

aacaccgtat ggaccactcc cgacggacga accgacgtgc gacccgagca gggtctcctc 6840

aagctgcgaa aggacctgaa cctgtacgcc aacctgcgac cctgccagct gctgtcgccc 6900

aagctcgccg atctctcccc catccgaaac gttgagggca ccgacttcat cattgtccga 6960

gagctcgtcg gaggtatcta ctttggagag cgaaaggagg atgacggatc tggcgtcgct 7020

tccgacaccg agacctactc cgttcctgag gttgagcgaa ttgcccgaat ggccgccttc 7080

ctggcccttc agcacaaccc ccctcttccc gtgtggtctc ttgacaaggc caacgtgctg 7140

gcctcctctc gactttggcg aaagactgtc actcgagtcc tcaaggacga attcccccag 7200

ctcgagctca accaccagct gatcgactcg gccgccatga tcctcatcaa gcagccctcc 7260

aagatgaatg gtatcatcat caccaccaac atgtttggcg atatcatctc cgacgaggcc 7320

tccgtcatcc ccggttctct gggtctgctg ccctccgcct ctctggcttc tctgcccgac 7380

accaacgagg cgttcggtct gtacgagccc tgtcacggat ctgcccccga tctcggcaag 7440

cagaaggtca accccattgc caccattctg tctgccgcca tgatgctcaa gttctctctt 7500

aacatgaagc ccgccggtga cgctgttgag gctgccgtca aggagtccgt cgaggctggt 7560

atcactaccg ccgatatcgg aggctcttcc tccacctccg aggtcggaga cttgttgcca 7620

acaaggtcaa ggagctgctc aagaaggagt aagtcgtttc tacgacgcat tgatggaagg 7680

agcaaactga cgcgcctgcg ggttggtcta ccggcagggt ccgctagtgt ataagactct 7740

ataaaaaggg ccctgccctg ctaatgaaat gatgatttat aatttaccgg tgtagcaacc 7800

ttgactagaa gaagcagatt gggtgtgttt gtagtggagg acagtggtac gttttggaaa 7860

cagtcttctt gaaagtgtct tgtctacagt atattcactc ataacctcaa tagccaaggg 7920

tgtagtcggt ttattaaagg aagggagttg tggctgatgt ggatagatat ctttaagctg 7980

gcgactgcac ccaacgagtg tggtggtagc ttgttagatc tgtatattcg gtaagatata 8040

ttttgtgggg ttttagtggt gttt 8064

<210> 133

<211> 5356

<212> DNA

<213> Artificial seuqnece

<220><223> AscI/SphI integration fragment from pRF201

<400> 133

cgctgatgac actttggtct gaaagagatg cattttgaat cccaaacttg cagtgcccaa 60

gtgacataca tctccgcgtt ttggaaaatg ttcagaaaca gttgattgtg ttggaatggg 120

gaatggggaa tggaaaaatg actcaagtat caattccaaa aacttctctg gctggcagta 180

cctactgtcc atactactgc attttctcca gtcaggccac tctatactcg acgacacagt 240

agtaaaaccc agataatttc gacataaaca agaaaacaga cccaataata tttatatata 300

gtcagccgtt tgtccagttc agactgtaat agccgaaaaa aaatccaaag tttctattct 360

aggaaaatat attccaatat ttttaattct taatctcatt tattttattc tagcgaaata 420

catttcagct acttgagaca tgtgataccc acaaatcgga ttcggactcg gttgttcaga 480

agagcatatg gcattcgtgc tcgcttgttc acgtattctt cctgttccat ctcttggccg 540

acaatcacac aaaaatgggg ttttttttttt aattctaatg attcattaca gcaaaattga 600

gatatagcag accacgtatt ccataatcac caaggaagtt cttgggcgtc ttaattaagt 660

tgcgacacat gtcttgatag tatcttggct tctctctctt gagcttttcc ataacaagtt 720

cttctgcctc caggaagtcc atggtgaatg attcttatac tcagaaggaa atgcttaacg 780

atttcgggtg tgagttgaca aggagagaga gaaaagaaga ggaaaggtaa ttcggggacg 840

gtggtctttt atacccttgg ctaaagtccc aaccacaaag caaaaaaatt ttcagtagtc 900

tattttgcgt ccggcatggg ttacccggat ggccagacaa agaaactagt acaaagtctg 960

aacaagcgta gattccagac tgcagtaccc tacgccctta acggcaagtg tgggaaccgg 1020

gggaggtttg atatgtgggg tgaaaggggc tctcgccggg gttgggcccg ctactgggtc 1080

aatttggggt caattggggc aattggggct gttttttggg acacaaatac gccgccaacc 1140

cggtctctcc tgaattctgc agatgggctg caggaattcc gtcgtcgcct gagtcgacat 1200

catttattta ccagttggcc acaaaccctt gacgatctcg tatgtcccct ccgacatact 1260

cccggccggc tggggtacgt tcgatagcgc tatcggcatc gacaaggttt gggtccctag 1320

ccgataccgc actacctgag tcacaatctt cggaggttta gtcttccaca tagcacgggc 1380

aaaagtgcgt atatatacaa gagcgtttgc cagccacaga ttttcactcc acacaccaca 1440

tcacacatac aaccacacac atccacaatg gaacccgaaa ctaagaagac caagactgac 1500

tccaagaaga ttgttcttct cggcggcgac ttctgtggcc ccgaggtgat tgccgaggcc 1560

gtcaaggtgc tcaagtctgt tgctgaggcc tccggcaccg agtttgtgtt tgaggaccga 1620

ctcattggag gagctgccat tgagaaggag ggcgagccca tcaccgacgc tactctcgac 1680

atctgccgaa aggctgactc tattatgctc ggtgctgtcg gaggcgctgc caacaccgta 1740

tggaccactc ccgacggacg aaccgacgtg cgacccgagc agggtctcct caagctgcga 1800

aaggacctga acctgtacgc caacctgcga ccctgccagc tgctgtcgcc caagctcgcc 1860

gatctctccc ccatccgaaa cgttgagggc accgacttca tcattgtccg agagctcgtc 1920

ggaggtatct actttggaga gcgaaaggag gatgacggat ctggcgtcgc ttccgacacc 1980

gagacctact ccgttcctga ggttgagcga attgcccgaa tggccgcctt cctggccctt 2040

cagcacaacc cccctcttcc cgtgtggtct cttgacaagg ccaacgtgct ggcctcctct 2100

cgactttggc gaaagactgt cactcgagtc ctcaaggacg aattccccca gctcgagctc 2160

aaccaccagc tgatcgactc ggccgccatg atcctcatca agcagccctc caagatgaat 2220

ggtatcatca tcaccaccaa catgtttggc gatatcatct ccgacgaggc ctccgtcatc 2280

cccggttctc tgggtctgct gccctccgcc tctctggctt ctctgcccga caccaacgag 2340

gcgttcggtc tgtacgagcc ctgtcacgga tctgcccccg atctcggcaa gcagaaggtc 2400

aaccccattg ccaccattct gtctgccgcc atgatgctca agttctctct taacatgaag 2460

cccgccggtg acgctgttga ggctgccgtc aaggagtccg tcgaggctgg tatcactacc 2520

gccgatatcg gaggctcttc ctccacctcc gaggtcggag acttgttgcc aacaaggtca 2580

aggagctgct caagaaggag taagtcgttt ctacgacgca ttgatggaag gagcaaactg 2640

acgcgcctgc gggttggtct accggcaggg tccgctagtg tataagactc tataaaaagg 2700

gccctgccct gctaatgaaa tgatgattta taatttaccg gtgtagcaac cttgactaga 2760

agaagcagat tgggtgtgtt tgtagtggag gacagtggta cgttttggaa acagtcttct 2820

tgaaagtgtc ttgtctacag tatattcact cataacctca atagccaagg gtgtagtcgg 2880

tttattaaag gaagggagtt gtggctgatg tggatagata tctttaagct ggcgactgca 2940

cccaacgagt gtggtggtag cttgttagat ctgtatattc ggtaagatat attttgtggg 3000

gttttagtgg tgtttaaacc atcatctaag ggcctcaaaa ctacctcgga actgctgcgc 3060

tgatctggac accacagagg ttccgagcac tttaggttgc accaaatgtc ccaccaggtg 3120

caggcagaaa acgctggaac agcgtgtaca gtttgtctta acaaaaagtg agggcgctga 3180

ggtcgagcag ggtggtgtga cttgttatag cctttagagc tgcgaaagcg cgtatggatt 3240

tggctcatca ggccagattg agggtctgtg gacacatgtc atgttagtgt acttcaatcg 3300

ccccctggat atagccccga caataggccg tggcctcatt tttttgcctt ccgcacattt 3360

ccattgctcg gtacccacac cttgcttctc ctgcacttgc caaccttaat actggtttac 3420

attgaccaac atcttacaag cgggggggctt gtctagggta tatataaaca gtggctctcc 3480

caatcggttg ccagtctctt ttttcctttc tttccccaca gattcgaaat ctaaactaca 3540

catcacacca tggcctcctc ggaggacgtc atcaaggagt tcatgcgatt caaggtccga 3600

atggaaggct ccgtgaacgg tcacgagttt gagattgagg gagagggtga aggccgaccc 3660

tacgaaggca cccagaccgc gaagctgaag gtgaccaagg gtggaccccct gcccttcgcc 3720

tgggacattc tgtctcctca gtttcagtac ggttctaagg tgtacgtgaa gcaccctgct 3780

gacattcccg actacaagaa actttccttt cccgagggct tcaagtggga gcgagttatg 3840

aacttcgagg atggcggtgt cgttaccgtt actcaggact cctcgctcca ggacggctcg 3900

ttcatctaca aggttaagtt catcggtgtc aacttcccta gcgatggacc cgtcatgcaa 3960

aagaaaacta tgggatggga agcctctaca gagcggctgt accctcgaga cggagtgttg 4020

aagggcgaga ttcacaaggc cctgaagctc aaggacggtg gacactatct cgttgagttt 4080

aagtctatct acatggcaaa gaaacccgtg cagcttccag gctactatta cgtcgattcc 4140

aagctcgata tcaccagcca taatgaggac tacactattg tcgaacagta cgagcgtgct 4200

gagggaagac accatctgtt tctttaagcg gccgcaagtg tggatgggga agtgagtgcc 4260

cggttctgtg tgcacaattg gcaatccaag atggatggat tcaacacagg gatatagcga 4320

gctacgtggt ggtgcgagga tatagcaacg gatatttatg tttgacactt gagaatgtac 4380

gatacaagca ctgtccaagt acaatactaa acatactgta catactcata ctcgtacccg 4440

ggcaacggtt tcacttgagt gcagtggcta gtgctcttac tcgtacagtg tgcaatactg 4500

cgtatcatag tctttgatgt atatcgtatt cattcatgtt agttgcgtac gggcgtcgtt 4560

gcttgtgtga tttttgagga cccatccctt tggtatataa gtatactctg gggttaaggt 4620

tgcccgtgta gtctaggtta tagttttcat gtgaaatacc gagagccgag ggagaataaa 4680

cgggggtatt tggacttgtt tttttcgcgg aaaagcgtcg aatcaaccct gcgggccttg 4740

caccatgtcc acgacgtgtt tctcgcccca attcgcccct tgcacgtcaa aattaggcct 4800

ccatctagac ccctccataa catgtgactg tggggaaaag tataagggaa accatgcaac 4860

catagacgac gtgaaagacg gggaggaacc aatggaggcc aaagaaatgg ggtagcaaca 4920

gtccaggaga cagacaagga gacaaggaga gggcgcccga aagatcggaa aaacaaacat 4980

gtccaattgg ggcagtgacg gaaacgacac ggacacttca gtacaatgga ccgaccatct 5040

ccaagccagg gttattccgg tatcaccttg gccgtaacct cccgctggta cctgatattg 5100

tacacgttca cattcaatat actttcagct acaataagag aggctgtttg tcgggcatgt 5160

gtgtccgtcg tatggggtga tgtccgaggg cgaaattcgc tacaagctta actctggcgc 5220

ttgtccagta tgaatagaca agtcaagacc agtggtgcca tgattgacag ggaggtacaa 5280

gacttcgata ctcgagcatt actcggactt gtggcgattg aacagacggg cgatcgcttc 5340

tcccccgtat tgccgg 5356

<210> 134

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> HY026

<400> 134

gcgcgtttaa accatcatct aagggcctca aaactacc 38

<210> 135

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> HY027

<400> 135

gagagcggcc gcttaaagaa acagatggtg tcttccct 38

<210>

136

<211> 11952

<212> DNA

<213> Artificial Sequence

<220><223> pRF169

<400> 136

catggacaag aaatactcca tcggcctgga cattggaacc aactctgtcg gctgggctgt 60

catcaccgac gagtacaagg tgccctccaa gaaattcaag gtcctcggaa acaccgatcg 120

acactccatc aagaaaaacc tcattggtgc cctgttgttc gattctggcg agactgccga 180

agctaccaga ctcaagcgaa ctgctcggcg acgttacacc cgacggaaga accgaatctg 240

ctacctgcag gagatctttt ccaacgagat ggccaaggtg gacgattcgt tctttcatcg 300

actggaggaa tccttcctcg tcgaggaaga caagaaacac gagcgtcatc ccatctttgg 360

caacattgtg gacgaggttg cttaccacga gaagtatcct accatctacc atctccgaaa 420

gaaactcgtc gattccaccg acaaggcgga tctcagactt atctacctcg ctctggcaca 480

catgatcaag tttcgaggtc atttcctcat cgagggcgat ctcaatcccg acaacagcga 540

tgtggacaag ctgttcattc agctcgttca gacctacaac cagctgttcg aggaaaaccc 600

catcaatgcc tccggagtcg atgcaaaggc catcttgtct gctcgactct cgaagagcag 660

acgactggag aacctcattg cccaacttcc tggcgagaaa aagaacggac tgtttggcaa 720

cctcattgcc ctttctcttg gtctcacacc caacttcaag tccaacttcg atctggcgga 780

ggacgccaag ctccagctgt ccaaggacac ctacgacgat gacctcgaca acctgcttgc 840

acagattggc gatcagtacg ccgacctgtt tctcgctgcc aagaaccttt cggatgctat 900

tctcttgtct gacattctgc gagtcaacac cgagatcaca aaggctcccc tttctgcctc 960

catgatcaag cgatacgacg agcaccatca ggatctcaca ctgctcaagg ctcttgtccg 1020

acagcaactg cccgagaagt acaaggagat ctttttcgat cagtcgaaga acggctacgc 1080

tggatacatc gacggcggag cctctcagga agagttctac aagttcatca agccaattct 1140

cgagaagatg gacggaaccg aggaactgct tgtcaagctc aatcgagagg atctgcttcg 1200

gaagcaacga accttcgaca acggcagcat tcctcatcag atccacctcg gtgagctgca 1260

cgccattctt cgacgtcagg aagacttcta ccccttttctc aaggacaacc gagagaagat 1320

cgagaagatt cttacctttc gaatccccta ctatgttggt cctcttgcca gaggaaactc 1380

tcgatttgct tggatgactc gaaagtccga ggaaaccatc actccctgga acttcgagga 1440

agtcgtggac aagggtgcct ctgcacagtc cttcatcgag cgaatgacca acttcgacaa 1500

gaatctgccc aacgagaagg ttcttcccaa gcattcgctg ctctacgagt actttacagt 1560

ctacaacgaa ctcaccaaag tcaagtacgt taccgaggga atgcgaaagc ctgccttctt 1620

gtctggcgaa cagaagaaag ccattgtcga tctcctgttc aagaccaacc gaaaggtcac 1680

tgttaagcag ctcaaggagg actacttcaa gaaaatcgag tgtttcgaca gcgtcgagat 1740

ttccggagtt gaggaccgat tcaacgcctc tttgggcacc tatcacgatc tgctcaagat 1800

tatcaaggac aaggattttc tcgacaacga ggaaaacgag gacattctgg aggacatcgt 1860

gctcactctt accctgttcg aagatcggga gatgatcgag gaacgactca agacatacgc 1920

tcacctgttc gacgacaagg tcatgaaaca actcaagcga cgtagataca ccggctgggg 1980

aagactttcg cgaaagctca tcaacggcat cagagacaag cagtccggaa agaccattct 2040

ggactttctc aagtccgatg gctttgccaa ccgaaacttc atgcagctca ttcacgacga 2100

ttctcttacc ttcaaggagg acatccagaa ggcacaagtg tccggtcagg gcgacagctt 2160

gcacgaacat attgccaacc tggctggttc gccagccatc aagaaaggca ttctccagac 2220

tgtcaaggtt gtcgacgagc tggtgaaggt catgggacgt cacaagcccg agaacattgt 2280

gatcgagatg gccagagaga accagacaac tcaaaagggt cagaaaaact cgcgagagcg 2340

gatgaagcga atcgaggaag gcatcaagga gctgggatcc cagattctca aggagcatcc 2400

cgtcgagaac actcaactgc agaacgagaa gctgtatctc tactatctgc agaatggtcg 2460

agacatgtac gtggatcagg aactggacat caatcgtctc agcgactacg atgtggacca 2520

cattgtccct caatcctttc tcaaggacga ttctatcgac aacaaggtcc ttacacgatc 2580

cgacaagaac agaggcaagt cggacaacgt tcccagcgaa gaggtggtca aaaagatgaa 2640

gaactactgg cgacagctgc tcaacgccaa gctcattacc cagcgaaagt tcgacaatct 2700

taccaaggcc gagcgaggcg gtctgtccga gctcgacaag gctggcttca tcaagcgtca 2760

actcgtcgag accagacaga tcacaaagca cgtcgcacag attctcgatt ctcggatgaa 2820

caccaagtac gacgagaacg acaagctcat ccgagaggtc aaggtgatta ctctcaagtc 2880

caaactggtc tccgatttcc gaaaggactt tcagttctac aaggtgcgag agatcaacaa 2940
ttaccaccat gcccacgatg cttacctcaa cgccgtcgtt ggcactgcgc tcatcaagaa 3000
ataccccaag ctcgaaagcg agttcgttta cggcgattac aaggtctacg acgttcgaaa 3060
gatgattgcc aagtccgaac aggagattgg caaggctact gccaagtact tcttttactc 3120
caacatcatg aacttttca agaccgagat caccttggcc aacggagaga ttcgaaagag 3180
accacttatc gagaccaacg gcgaaactgg agagatcgtg tgggacaagg gtcgagactt 3240
tgcaaccgtg cgaaaggttc tgtcgatgcc tcaggtcaac atcgtcaaga aaaccgaggt 3300

tcagactggc ggattctcca aggagtcgat tctgcccaag cgaaactccg acaagctcat 3360
cgctcgaaag aaagactggg atcccaagaa atacggtggc ttcgattctc ctaccgtcgc 3420
ctattccgtg cttgtcgttg cgaaggtcga gaagggcaag tccaaaaagc tcaagtccgt 3480
caaggagctg ctcggaatta ccatcatgga gcgatcgagc ttcgagaaga atcccatcga 3540
cttcttggaa gccaagggtt acaaggaggt caagaaagac ctcattatca agctgcccaa 3600
gtactctctg ttcgaactgg agaacggtcg aaagcgtatg ctcgcctccg ctggcgagct 3660
gcagaaggga aacgagcttg ccttgccttc gaagtacgtc aactttctct atctggcttc 3720

tcactacgag aagctcaagg gttctcccga ggacaacgaa cagaagcaac tcttcgttga 3780
gcagcacaaa cattacctcg acgagattat cgagcagatt tccgagtttt cgaagcgagt 3840
catcctggct gatgccaact tggacaaggt gctctctgcc tacaacaagc atcgggacaa 3900
acccattcga gaacaggcgg agaacatcat tcacctgttt actcttacca acctgggtgc 3960
tcctgcagct ttcaagtact tcgataccac tatcgaccga aagcggtaca catccaccaa 4020
ggaggttctc gatgccaccc tgattcacca gtccatcact ggcctgtacg agacccgaat 4080
cgacctgtct cagcttggtg gcgactccag agccgatccc aagaaaaagc gaaaggtcta 4140

agcggccgca agtgtggatg gggaagtgag tgcccggttc tgtgtgcaca attggcaatc 4200
caagatggat ggattcaaca cagggatata gcgagctacg tggtggtgcg aggatatagc 4260
aacggatatt tatgtttgac acttgagaat gtacgataca agcactgtcc aagtacaata 4320
ctaaacatac tgtacatact catactcgta cccgggcaac ggtttcactt gagtgcagtg 4380
gctagtgctc ttactcgtac agtgtgcaat actgcgtatc atagtctttg atgtatatcg 4440
tattcattca tgttagttgc gtacgagccg gaagcataaa gtgtaaagcc tggggtgcct 4500
aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa 4560

acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta 4620

ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc 4680

gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg 4740

caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt 4800

tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa 4860

gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct 4920

ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc 4980

cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg 5040

tcgttcgctc caagctgggc tgtgtgcacg aacccccccgt tcagcccgac cgctgcgcct 5100

tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag 5160

cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga 5220

agtggtggcc taactacggc tacactagaa ggacagtatt tggtatctgc gctctgctga 5280

agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg 5340

gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag 5400

aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag 5460

ggattttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat 5520

gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct 5580

taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac 5640

tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa 5700

tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg 5760

gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt 5820

gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca 5880

ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt 5940

cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct 6000

tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg 6060

cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg 6120

agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg 6180

cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa 6240

aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt 6300

aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt 6360

gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt 6420

gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca 6480

tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat 6540

ttccccgaaa agtgccacct gacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg 6600

tggttacgcg cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt 6660

tcttcccttc ctttctcgcc acgttcgccg gctttccccg tcaagctcta aatcgggggc 6720

tccctttagg gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg 6780

gtgatggttc acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg 6840

agtccacgtt ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct 6900

cggtctattc ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg 6960

agctgattta acaaaaattt aacgcgaatt ttaacaaaat attaacgctt acaatttcca 7020

ttcgccattc aggctgcgca actgttggga agggcgatcg gtgcgggcct cttcgctatt 7080

acgccagctg gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa cgccagggtt 7140

ttcccagtca cgacgttgta aaacgacggc cagtgaattg taatacgact cactataggg 7200

cgaattgggt accgggcccc ccctcgaggt cgatggtgtc gataagcttg atatcgaatt 7260

catgtcacac aaaccgatct tcgcctcaag gaaacctaat tctacatccg agagactgcc 7320

gagatccagt ctacactgat taattttcgg gccaataatt taaaaaaatc gtgttatata 7380

atattatatg tattatatat atacatcatg atgatactga cagtcatgtc ccattgctaa 7440

atagacagac tccatctgcc gcctccaact gatgttctca atatttaagg ggtcatctcg 7500

cattgtttaa taataaacag actccatcta ccgcctccaa atgatgttct caaaatatat 7560

tgtatgaact tatttttatt acttagtatt attagacaac ttacttgctt tatgaaaaac 7620

acttcctatt taggaaacaa tttataatgg cagttcgttc atttaacaat ttatgtagaa 7680

taaatgttat aaatgcgtat gggaaatctt aaatatggat agcataaatg atatctgcat 7740

tgcctaattc gaaatcaaca gcaacgaaaa aaatcccttg tacaacataa atagtcatcg 7800

agaaatatca actatcaaag aacagctatt cacacgttac tattgagatt attattggac 7860

gagaatcaca cactcaactg tctttctctc ttctagaaat acaggtacaa gtatgtacta 7920

ttctcattgt tcatacttct agtcatttca tcccacatat tccttggatt tctctccaat 7980

gaatgacatt ctatcttgca aattcaacaa ttataataag atataccaaa gtagcggtat 8040

agtggcaatc aaaaagcttc tctggtgtgc ttctcgtatt tatttttatt ctaatgatcc 8100

attaaaggta tatatttatt tcttgttata taatcctttt gtttattaca tgggctggat 8160

acataaaggt attttgattt aatttttttgc ttaaattcaa tccccccctcg ttcagtgtca 8220

actgtaatgg taggaaatta ccatactttt gaagaagcaa aaaaaatgaa agaaaaaaaa 8280

aatcgtattt ccaggttaga cgttccgcag aatctagaat gcggtatgcg gtacattgtt 8340

cttcgaacgt aaaagttgcg ctccctgaga tattgtacat ttttgctttt acaagtacaa 8400
gtacatcgta caactatgta ctactgttga tgcatccaca acagtttgtt ttgttttttt 8460
ttgttttttt tttttctaat gattcattac cgctatgtat acctacttgt acttgtagta 8520
agccgggtta ttggcgttca attaatcata gacttatgaa tctgcacggt gtgcgctgcg 8580
agttactttt agcttatgca tgctacttgg gtgtaatatt gggatctgtt cggaaatcaa 8640
cggatgctca atcgatcgcc agcaacgcgg ccttagacat aaaaaacaaa aaaaaaaagc 8700
accgactcgg tgccactttt tcaagttgat aacggactag ccttatttta acttgctatt 8760

tctagctcta aaacgcaggt gtaaaaataa aaaggcctgc gattaccagc aggcctgtta 8820
ttaacctaag ccttaggacg cttcacgcca tacttggaac gagcctgctt acggtcttta 8880
acgccggagc agtcaagcgc accacgtacg gtgtggtaac gaacacccgg gaggtcttta 8940
acacgaccgc cacggatcag gatcacggag tgctcctgca ggttgtgacc ttcaccaccg 9000
atgtaggaag tcacttcgaa accgttagtc agacgaacac ggcatacttt acgcagcgcg 9060
gagttcggtt ttttaggagt ggtagtatat acacgagtac atacgccacg tttttgcggg 9120
catgcttcca gcgcaggcac gttgcttttc gcaactttgc gagcacgtgg tttgcgtacc 9180

agctggttaa ctgttgccat taaatagctc ctggttttag cttttgcttc gtaaacacgt 9240
aataaaacgt cctcacacaa tatgaggacg ccgaatttag ggcgatgccg aaaaggtgtc 9300
aagaaatata caacgatccc gccatcggcg cgcccacctg ctacgcatgg ttgatgtgtg 9360
tttaattcaa gaatgaatat agagaagaga agaagaaaaa agattcaatt gagccggcga 9420
tgcagaccct tatataaatg ttgccttgga cagacggagc aagcccgccc aaacctacgt 9480
tcggtataat atgttaagct ttttaacaca aaggtttggc ttggggtaac ctgatgtggt 9540
gcaaaagacc gggcgttggc gagccattgc gcgggcgaat ggggtcgtga ctcgtctcaa 9600

attcgagggc gtgcctcaat tcgtgccccc gtggcttttt cccgccgttt ccgccccgtt 9660
tgcaccactg cagccgcttc tttggttcgg acaccttgct gcgagctagg tgccttgtgc 9720
tacttaaaaa gtggcctccc aacaccaaca tgacatgagt gcgtgggcca agacacgttg 9780
gcggggtcgc agtcggctca atggcccgga aaaaacgctg ctggagctgg ttcggacgca 9840
gtccgccgcg gcgtatggat atccgcaagg ttccatagcg ccattgccct ccgtcggcgt 9900
ctatcccgca accttaatta agtcatacac aagtcagctt tcttcgagcc tcatataagt 9960
ataagtagtt caacgtatta gcactgtacc cagcatctcc gtatcgagaa acacaacaac 10020

atgccccatt ggacagatca tgcggataca caggttgtgc agtatcatac atactcgatc 10080

agacaggtcg tctgaccatc atacaagctg aacaagcgct ccatacttgc acgctctcta 10140

tatacacagt taaattacat atccatagtc taacctctaa cagttaatct tctggtaagc 10200

ctcccagcca gccttctggt atcgcttggc ctcctcaata ggatctcggt tctggccgta 10260

cagacctcgg ccgacaatta tgatatccgt tccggtagac atgacatcct caacagttcg 10320

gtactgctgt ccgagagcgt ctcccttgtc gtcaagaccc accccggggg tcagaataag 10380

ccagtcctca gagtcgccct taggtcggtt ctgggcaatg aagccaacca caaactcggg 10440

gtcggatcgg gcaagctcaa tggtctgctt ggagtactcg ccagtggcca gagagccctt 10500

gcaagacagc tcggccagca tgagcagacc tctggccagc ttctcgttgg gagaggggac 10560

taggaactcc ttgtactggg agttctcgta gtcagagacg tcctccttct tctgttcaga 10620

gacagtttcc tcggcaccag ctcgcaggcc agcaatgatt ccggttccgg gtacaccgtg 10680

ggcgttggtg atatcggacc actcggcgat tcggtgacac cggtactggt gcttgacagt 10740

gttgccaata tctgcgaact ttctgtcctc gaacaggaag aaaccgtgct taagagcaag 10800

ttccttgagg gggagcacag tgccggcgta ggtgaagtcg tcaatgatgt cgatatgggt 10860

tttgatcatg cacacataag gtccgacctt atcggcaagc tcaatgagct ccttggtggt 10920

ggtaacatcc agagaagcac acaggttggt tttcttggct gccacgagct tgagcactcg 10980

agcggcaaag gcggacttgt ggacgttagc tcgagcttcg taggagggca ttttggtggt 11040

gaagaggaga ctgaaataaa tttagtctgc agaacttttt atcggaacct tatctggggc 11100

agtgaagtat atgttatggt aatagttacg agttagttga acttatagat agactggact 11160

atacggctat cggtccaaat tagaaagaac gtcaatggct ctctgggcgt cgcctttgcc 11220

gacaaaaatg tgatcatgat gaaagccagc aatgacgttg cagctgatat tgttgtcggc 11280

caaccgcgcc gaaaacgcag ctgtcagacc cacagcctcc aacgaagaat gtatcgtcaa 11340

agtgatccaa gcacactcat agttggagtc gtactccaaa ggcggcaatg acgagtcaga 11400

cagatactcg tcgacgttta aaccatcatc taagggcctc aaaactacct cggaactgct 11460

gcgctgatct ggacaccaca gaggttccga gcactttagg ttgcaccaaa tgtcccacca 11520

ggtgcaggca gaaaacgctg gaacagcgtg tacagtttgt cttaacaaaa agtgagggcg 11580

ctgaggtcga gcagggtggt gtgacttgtt atagccttta gagctgcgaa agcgcgtatg 11640

gatttggctc atcaggccag attgagggtc tgtggacaca tgtcatgtta gtgtacttca 11700

atcgcccccct ggatatagcc ccgacaatag gccgtggcct catttttttg ccttccgcac 11760

atttccattg ctcggtaccc acaccttgct tctcctgcac ttgccaacct taatactggt 11820

ttacattgac caacatctta caagcggggg gcttgtctag ggtatatata aacagtggct 11880

ctcccaatcg gttgccagtc tctttttcc tttctttccc cacagattcg aaatctaaac 11940

tacacatcac ac 11952

<210> 137

<211> 564

<212> DNA

<213> Yarrowia lipolytica

<220><221> misc_feature

<222> (1)..(564)

<223> GPD Promoter

<400> 137

ggttgcggga tagacgccga cggagggcaa tggcgctatg gaaccttgcg gatatccata 60

cgccgcggcg gactgcgtcc gaaccagctc cagcagcgtt ttttccgggc cattgagccg 120

actgcgaccc cgccaacgtg tcttggccca cgcactcatg tcatgttggt gttgggaggc 180

cacttttaa gtagcacaag gcacctagct cgcagcaagg tgtccgaacc aaagaagcgg 240

ctgcagtggt gcaaacgggg cggaaacggc gggaaaaagc cacgggggca cgaattgagg 300

cacgccctcg aatttgagac gagtcacgac cccattcgcc cgcgcaatgg ctcgccaacg 360

cccggtcttt tgcaccacat caggttaccc caagccaaac ctttgtgtta aaaagcttaa 420

catattatac cgaacgtagg tttgggcggg cttgctccgt ctgtccaagg caacatttat 480

ataagggtct gcatcgccgg ctcaattgaa tctttttct tcttctcttc tctatattca 540

ttcttgaatt aaacacacat caac 564

<210> 138

<211> 1271

<212> DNA

<213> Artificial sequence

<220><223> GPD promoter-counterselectable marker-CER-terminator

<400> 138

ttaattaagg ttgcgggata gacgccgacg gagggcaatg gcgctatgga accttgcgga 60

tatccatacg ccgcggcgga ctgcgtccga accagctcca gcagcgtttt ttccgggcca 120

ttgagccgac tgcgaccccg ccaacgtgtc ttggcccacg cactcatgtc atgttggtgt 180

tgggaggcca ctttttaagt agcacaaggc acctagctcg cagcaaggtg tccgaaccaa 240

agaagcggct gcagtggtgc aaacggggcg gaaacggcgg gaaaaagcca cgggggcacg 300

aattgaggca cgccctcgaa tttgagacga gtcacgaccc cattcgcccg cgcaatggct 360

cgccaacgcc cggtcttttg caccacatca ggttacccca agccaaacct ttgtgttaaa 420

aagcttaaca tattataccg aacgtaggtt tgggcgggct tgctccgtct gtccaaggca 480

acatttatat aagggtctgc atcgccggct caattgaatc ttttttcttc ttctcttctc 540

tatattcatt cttgaattaa acacacatca accatgcgta gcaggtgggc gcgccgatgg 600

cgggatcgtt gtatatttct tgacaccttt tcggcatcgc cctaaattcg gcgtcctcat 660

attgtgtgag gacgttttat tacgtgttta cgaagcaaaa gctaaaacca ggagctattt 720

aatggcaaca gttaaccagc tggtacgcaa accacgtgct cgcaaagttg cgaaaagcaa 780

cgtgcctgcg ctggaagcat gcccgcaaaa acgtggcgta tgtactcgtg tatatactac 840

cactcctaaa aaaccgaact ccgcgctgcg taaagtatgc cgtgttcgtc tgactaacgg 900

tttcgaagtg acttcctaca tcggtggtga aggtcacaac ctgcaggagc actccgtgat 960

cctgatccgt ggcggtcgtg ttaaagacct cccgggtgtt cgttaccaca ccgtacgtgg 1020

tgcgcttgac tgctccggcg ttaaagaccg taagcaggct cgttccaagt atggcgtgaa 1080

gcgtcctaag gcttaggtta ataacaggcc tgctggtaat cgcaggcctt tttattttta 1140

cacctgcgtt ttagagctag aaatagcaag ttaaaataag gctagtccgt tatcaacttg 1200

aaaaagtggc accgagtcgg tgcttttttt ttttgttttt tatgtctaag gccgcgttgc 1260

tggcgatcga t 1271

<210> 139

<211> 23

<212> DNA

<213> Artificial sequence

<220><223> ds-temp-1 target site

<400> 139

tcagtttcag tacggttcta agg 23

<210> 140

<211> 23

<212> DNA

<213> Artificial sequence

<220><223> ds-temp-2 target site

<400> 140

cggtgtcgtt accgttactc agg 23

<210> 141

<211> 23

<212> DNA

<213> Artificial sequence

<220><223> ds-nontemp-3 target site

<400> 141

cttggtcacc ttcagcttcg cgg 23

<210> 142

<211> 63

<212> DNA

<213> Artificial sequence

<220><223> Hammerhead ribozyme-VT fusion

<220><221> misc_feature

<222> (1)..(6)

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

<220><221> misc_feature

<222> (44)..(63)

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

<400> 142

nnnnnnctga tgagtccgtg aggacgaaac gagtaagctc gtcnnnnnnn nnnnnnnnnn 60

nnn 63

<210> 143

<211> 63

<212> RNA

<213> Artificial sequence

<220><223> Hammerhead ribozyme-VT fusion

<220><221> misc_feature

<222> (1)..(6)

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

<220><221> misc_feature

<222> (44)..(63)

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

<400> 143

nnnnnncuga ugaguccgug aggacgaaac gaguaagcuc gucnnnnnnn nnnnnnnnnn 60

nnn 63

<210> 144

<211> 75

<212> DNA

<213> Artificial sequence

<220><223> ds-temp-1F

<400> 144

catgcgtaaa ctgactgatg agtccgtgag gacgaaacga gtaagctcgt ctcagtttca 60

gtacggttct agttt 75

<210> 145

<211> 75

<212> DNA

<213> Artificial sequence

<220><223> 145 ds-temp-1R

<400> 145

tctaaaacta gaaccgtact gaaactgaga cgagcttact cgtttcgtcc tcacggactc 60

atcagtcagt ttacg 75

<210> 146

<211> 75

<212> DNA

<213> Artificial sequence

<220><223> 146 ds-temp-2F

<400> 146

catgcgtaac accgctgatg agtccgtgag gacgaaacga gtaagctcgt ccggtgtcgt 60

taccgttact cgttt 75

<210> 147

<211> 75

<212> DNA

<213> Artificial sequence

<220><223> ds-temp-2R

<400> 147

tctaaaacga gtaacggtaa cgacaccgga cgagcttact cgtttcgtcc tcacggactc 60

atcagcggtg ttacg 75

<210> 148

<211> 75

<212> DNA

<213> Artificial sequence

<220><223> ds-nontemp-1F

<400> 148

catgcgtaac caggctgatg agtccgtgag gacgaaacga gtaagctcgt ccttggtcac 60

cttcagcttc ggttt 75

<210> 149

<211> 75

<212> DNA

<213> Artificial sequence

<220><223> ds-nontemp-1R

<400> 149

tctaaaaccg aagctgaagg tgaccaagga cgagcttact cgtttcgtcc tcacggactc 60

atcagcctgg ttacg 75

<210> 150

<211> 11448

<212> DNA

<213> Artificial sequence

<220><223> pRF296

<400> 150

catggacaag aaatactcca tcggcctgga cattggaacc aactctgtcg gctgggctgt 60

catcaccgac gagtacaagg tgccctccaa gaaattcaag gtcctcggaa acaccgatcg 120

acactccatc aagaaaaacc tcattggtgc cctgttgttc gattctggcg agactgccga 180

agctaccaga ctcaagcgaa ctgctcggcg acgttacacc cgacggaaga accgaatctg 240

ctacctgcag gagatctttt ccaacgagat ggccaaggtg gacgattcgt tctttcatcg 300

actggaggaa tccttcctcg tcgaggaaga caagaaacac gagcgtcatc ccatctttgg 360

caacattgtg gacgaggttg cttaccacga gaagtatcct accatctacc atctccgaaa 420

gaaactcgtc gattccaccg acaaggcgga tctcagactt atctacctcg ctctggcaca 480

catgatcaag tttcgaggtc atttcctcat cgagggcgat ctcaatcccg acaacagcga 540

tgtggacaag ctgttcattc agctcgttca gacctacaac cagctgttcg aggaaaaccc 600

catcaatgcc tccggagtcg atgcaaaggc catcttgtct gctcgactct cgaagagcag 660

acgactggag aacctcattg cccaacttcc tggcgagaaa aagaacggac tgtttggcaa 720

cctcattgcc ctttctcttg gtctcacacc caacttcaag tccaacttcg atctggcgga 780

ggacgccaag ctccagctgt ccaaggacac ctacgacgat gacctcgaca acctgcttgc 840

acagattggc gatcagtacg ccgacctgtt tctcgctgcc aagaaccttt cggatgctat 900

tctcttgtct gacattctgc gagtcaacac cgagatcaca aaggctcccc tttctgcctc 960

catgatcaag cgatacgacg agcaccatca ggatctcaca ctgctcaagg ctcttgtccg 1020

acagcaactg cccgagaagt acaaggagat ctttttcgat cagtcgaaga acggctacgc 1080

tggatacatc gacggcggag cctctcagga agagttctac aagttcatca agccaattct 1140

cgagaagatg gacggaaccg aggaactgct tgtcaagctc aatcgagagg atctgcttcg 1200

gaagcaacga accttcgaca acggcagcat tcctcatcag atccacctcg gtgagctgca 1260

cgccattctt cgacgtcagg aagacttcta ccccttttctc aaggacaacc gagagaagat 1320

cgagaagatt cttacctttc gaatccccta ctatgttggt cctcttgcca gaggaaactc 1380

tcgatttgct tggatgactc gaaagtccga ggaaaccatc actccctgga acttcgagga 1440

agtcgtggac aagggtgcct ctgcacagtc cttcatcgag cgaatgacca acttcgacaa 1500

gaatctgccc aacgagaagg ttcttcccaa gcattcgctg ctctacgagt actttacagt 1560

ctacaacgaa ctcaccaaag tcaagtacgt taccgaggga atgcgaaagc ctgccttctt 1620

gtctggcgaa cagaagaaag ccattgtcga tctcctgttc aagaccaacc gaaaggtcac 1680

tgttaagcag ctcaaggagg actacttcaa gaaaatcgag tgtttcgaca gcgtcgagat 1740

ttccggagtt gaggaccgat tcaacgcctc tttgggcacc tatcacgatc tgctcaagat 1800

tatcaaggac aaggattttc tcgacaacga ggaaaacgag gacattctgg aggacatcgt 1860

gctcactctt accctgttcg aagatcggga gatgatcgag gaacgactca agacatacgc 1920

tcacctgttc gacgacaagg tcatgaaaca actcaagcga cgtagataca ccggctgggg 1980

aagactttcg cgaaagctca tcaacggcat cagagacaag cagtccggaa agaccattct 2040

ggactttctc aagtccgatg gctttgccaa ccgaaacttc atgcagctca ttcacgacga 2100

ttctcttacc ttcaaggagg acatccagaa ggcacaagtg tccggtcagg gcgacagctt 2160

gcacgaacat attgccaacc tggctggttc gccagccatc aagaaaggca ttctccagac 2220

tgtcaaggtt gtcgacgagc tggtgaaggt catgggacgt cacaagcccg agaacattgt 2280

gatcgagatg gccagagaga accagacaac tcaaaagggt cagaaaaact cgcgagagcg 2340

gatgaagcga atcgaggaag gcatcaagga gctgggatcc cagattctca aggagcatcc 2400

cgtcgagaac actcaactgc agaacgagaa gctgtatctc tactatctgc agaatggtcg 2460

agacatgtac gtggatcagg aactggacat caatcgtctc agcgactacg atgtggacca 2520

cattgtccct caatcctttc tcaaggacga ttctatcgac aacaaggtcc ttacacgatc 2580

cgacaagaac agaggcaagt cggacaacgt tcccagcgaa gaggtggtca aaaagatgaa 2640

gaactactgg cgacagctgc tcaacgccaa gctcattacc cagcgaaagt tcgacaatct 2700

taccaaggcc gagcgaggcg gtctgtccga gctcgacaag gctggcttca tcaagcgtca 2760

actcgtcgag accagacaga tcacaaagca cgtcgcacag attctcgatt ctcggatgaa 2820

caccaagtac gacgagaacg acaagctcat ccgagaggtc aaggtgatta ctctcaagtc 2880

caaactggtc tccgatttcc gaaaggactt tcagttctac aaggtgcgag agatcaacaa 2940

ttaccaccat gcccacgatg cttacctcaa cgccgtcgtt ggcactgcgc tcatcaagaa 3000

ataccccaag ctcgaaagcg agttcgttta cggcgattac aaggtctacg acgttcgaaa 3060

gatgattgcc aagtccgaac aggagattgg caaggctact gccaagtact tcttttactc 3120

caacatcatg aacttttca agaccgagat caccttggcc aacggagaga ttcgaaagag 3180

accacttatc gagaccaacg gcgaaactgg agagatcgtg tgggacaagg gtcgagactt 3240

tgcaaccgtg cgaaaggttc tgtcgatgcc tcaggtcaac atcgtcaaga aaaccgaggt 3300

tcagactggc ggattctcca aggagtcgat tctgcccaag cgaaactccg acaagctcat 3360

cgctcgaaag aaagactggg atcccaagaa atacggtggc ttcgattctc ctaccgtcgc 3420

ctattccgtg cttgtcgttg cgaaggtcga gaagggcaag tccaaaaagc tcaagtccgt 3480

caaggagctg ctcggaatta ccatcatgga gcgatcgagc ttcgagaaga atcccatcga 3540

cttcttggaa gccaagggtt acaaggaggt caagaaagac ctcattatca agctgcccaa 3600

gtactctctg ttcgaactgg agaacggtcg aaagcgtatg ctcgcctccg ctggcgagct 3660

gcagaaggga aacgagcttg ccttgccttc gaagtacgtc aactttctct atctggcttc 3720

tcactacgag aagctcaagg gttctcccga ggacaacgaa cagaagcaac tcttcgttga 3780

gcagcacaaa cattacctcg acgagattat cgagcagatt tccgagtttt cgaagcgagt 3840

catcctggct gatgccaact tggacaaggt gctctctgcc tacaacaagc atcgggacaa 3900

acccattcga gaacaggcgg agaacatcat tcacctgttt actcttacca acctgggtgc 3960

tcctgcagct ttcaagtact tcgataccac tatcgaccga aagcggtaca catccaccaa 4020

ggaggttctc gatgccaccc tgattcacca gtccatcact ggcctgtacg agacccgaat 4080

cgacctgtct cagcttggtg gcgactccag agccgatccc aagaaaaagc gaaaggtcta 4140

agcggccgca agtgtggatg gggaagtgag tgcccggttc tgtgtgcaca attggcaatc 4200

caagatggat ggattcaaca cagggatata gcgagctacg tggtggtgcg aggatatagc 4260

aacggatatt tatgtttgac acttgagaat gtacgataca agcactgtcc aagtacaata 4320

ctaaacatac tgtacatact catactcgta cccgggcaac ggtttcactt gagtgcagtg 4380
gctagtgctc ttactcgtac agtgtgcaat actgcgtatc atagtctttg atgtatatcg 4440
tattcattca tgttagttgc gtacgagccg gaagcataaa gtgtaaagcc tggggtgcct 4500
aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa 4560
acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta 4620
ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc 4680
gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg 4740

caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt 4800
tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa 4860
gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct 4920
ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc 4980
cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg 5040
tcgttcgctc caagctgggc tgtgtgcacg aaccccccgt tcagcccgac cgctgcgcct 5100
tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag 5160

cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga 5220
agtggtggcc taactacggc tacactagaa ggacagtatt tggtatctgc gctctgctga 5280
agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg 5340
gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag 5400
aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag 5460
ggattttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat 5520
gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct 5580

taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac 5640
tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa 5700
tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg 5760
gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt 5820
gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca 5880
ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt 5940
cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct 6000

tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg 6060

cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg 6120

agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg 6180

cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa 6240

aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt 6300

aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt 6360

gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt 6420

gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca 6480

tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat 6540

ttccccgaaa agtgccacct gacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg 6600

tggttacgcg cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt 6660

tcttcccttc ctttctcgcc acgttcgccg gctttccccg tcaagctcta aatcgggggc 6720

tcccttttagg gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg 6780

gtgatggttc acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg 6840

agtccacgtt ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct 6900

cggtctattc ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg 6960

agctgattta acaaaaattt aacgcgaatt ttaacaaaat attaacgctt acaatttcca 7020

ttcgccattc aggctgcgca actgttggga agggcgatcg gtgcgggcct cttcgctatt 7080

acgccagctg gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa cgccagggtt 7140

ttcccagtca cgacgttgta aaacgacggc cagtgaattg taatacgact cactataggg 7200

cgaattgggt accgggcccc ccctcgaggt cgatggtgtc gataagcttg atatcgaatt 7260

catgtcacac aaaccgatct tcgcctcaag gaaacctaat tctacatccg agagactgcc 7320

gagatccagt ctacactgat taattttcgg gccaataatt taaaaaaatc gtgttatata 7380

atattatatg tattatatat atacatcatg atgatactga cagtcatgtc ccattgctaa 7440

atagacagac tccatctgcc gcctccaact gatgttctca atatttaagg ggtcatctcg 7500

cattgtttaa taataaacag actccatcta ccgcctccaa atgatgttct caaaatatat 7560

tgtatgaact tatttttatt acttagtatt attagacaac ttacttgctt tatgaaaaac 7620

acttcctatt taggaaacaa tttataatgg cagttcgttc atttaacaat ttatgtagaa 7680

taaatgttat aaatgcgtat gggaaatctt aaatatggat agcataaatg atatctgcat 7740

tgcctaattc gaaatcaaca gcaacgaaaa aaatcccttg tacaacataa atagtcatcg 7800

agaaatatca actatcaaag aacagctatt cacacgttac tattgagatt attattggac 7860

gagaatcaca cactcaactg tctttctctc ttctagaaat acaggtacaa gtatgtacta 7920

ttctcattgt tcatacttct agtcatttca tcccacatat tccttggatt tctctccaat 7980

gaatgacatt ctatcttgca aattcaacaa ttataataag atataccaaa gtagcggtat 8040

agtggcaatc aaaaagcttc tctggtgtgc ttctcgtatt tatttttatt ctaatgatcc 8100

attaaaggta tatatttatt tcttgttata taatcctttt gtttattaca tgggctggat 8160

acataaaggt attttgattt aatttttgc ttaaattcaa tccccctcg ttcagtgtca 8220

actgtaatgg taggaaatta ccatactttt gaagaagcaa aaaaaatgaa agaaaaaaaa 8280

aatcgtattt ccaggttaga cgttccgcag aatctagaat gcggtatgcg gtacattgtt 8340

cttcgaacgt aaaagttgcg ctccctgaga tattgtacat ttttgctttt acaagtacaa 8400

gtacatcgta caactatgta ctactgttga tgcatccaca acagtttgtt ttgttttttt 8460

ttgttttttt ttttttctaat gattcattac cgctatgtat acctacttgt acttgtagta 8520

agccgggtta ttggcgttca attaatcata gacttatgaa tctgcacggt gtgcgctgcg 8580

agttactttt agcttatgca tgctacttgg gtgtaatatt gggatctgtt cggaaatcaa 8640

cggatgctca atcgatcgcc agcaacgcgg ccttagacat aaaaaacaaa aaaaaaaagc 8700

accgactcgg tgccactttt tcaagttgat aacggactag ccttatttta acttgctatt 8760

tctagctcta aaactagaac cgtactgaaa ctgagacgag cttactcgtt tcgtcctcac 8820

ggactcatca gtcagtttac gcatggttga tgtgtgttta attcaagaat gaatatagag 8880

aagagaagaa gaaaaaagat tcaattgagc cggcgatgca gacccttata taaatgttgc 8940

cttggacaga cggagcaagc ccgcccaaac ctacgttcgg tataatatgt taagcttttt 9000

aacacaaagg tttggcttgg ggtaacctga tgtggtgcaa aagaccgggc gttggcgagc 9060

cattgcgcgg gcgaatgggg tcgtgactcg tctcaaattc gagggcgtgc ctcaattcgt 9120

gcccccgtgg ctttttcccg ccgtttccgc cccgtttgca ccactgcagc cgcttctttg 9180

gttcggacac cttgctgcga gctaggtgcc ttgtgctact taaaaagtgg cctcccaaca 9240

ccaacatgac atgagtgcgt gggccaagac acgttggcgg ggtcgcagtc ggctcaatgg 9300

cccggaaaaa acgctgctgg agctggttcg gacgcagtcc gccgcggcgt atggatatcc 9360

gcaaggttcc atagcgccat tgccctccgt cggcgtctat cccgcaacct taattaagtc 9420

atacacaagt cagctttctt cgagcctcat ataagtataa gtagttcaac gtattagcac 9480

tgtacccagc atctccgtat cgagaaacac aacaacatgc cccattggac agatcatgcg 9540

gatacacagg ttgtgcagta tcatacatac tcgatcagac aggtcgtctg accatcatac 9600

aagctgaaca agcgctccat acttgcacgc tctctatata cacagttaaa ttacatatcc 9660

atagtctaac ctctaacagt taatcttctg gtaagcctcc cagccagcct tctggtatcg 9720

cttggcctcc tcaataggat ctcggttctg gccgtacaga cctcggccga caattatgat 9780

atccgttccg gtagacatga catcctcaac agttcggtac tgctgtccga gagcgtctcc 9840
cttgtcgtca agacccaccc cgggggtcag aataagccag tcctcagagt cgcccttagg 9900
tcggttctgg gcaatgaagc caaccacaaa ctcggggtcg gatcgggcaa gctcaatggt 9960
ctgcttggag tactcgccag tggccagaga gcccttgcaa gacagctcgg ccagcatgag 10020
cagacctctg gccagcttct cgttgggaga ggggactagg aactccttgt actgggagtt 10080
ctcgtagtca gagacgtcct ccttcttctg ttcagagaca gtttcctcgg caccagctcg 10140
caggccagca atgattccgg ttccgggtac accgtgggcg ttggtgatat cggaccactc 10200

ggcgattcgg tgacaccggt actggtgctt gacagtgttg ccaatatctg cgaactttct 10260
gtcctcgaac aggaagaaac cgtgcttaag agcaagttcc ttgaggggga gcacagtgcc 10320
ggcgtaggtg aagtcgtcaa tgatgtcgat atgggttttg atcatgcaca cataaggtcc 10380
gaccttatcg gcaagctcaa tgagctcctt ggtggtggta acatccagag aagcacacag 10440
gttggttttc ttggctgcca cgagcttgag cactcgagcg gcaaaggcgg acttgtggac 10500
gttagctcga gcttcgtagg agggcatttt ggtggtgaag aggagactga aataaattta 10560
gtctgcagaa ctttttatcg gaaccttatc tggggcagtg aagtatatgt tatggtaata 10620

gttacgagtt agttgaactt atagatagac tggactatac ggctatcggt ccaaattaga 10680
aagaacgtca atggctctct gggcgtcgcc tttgccgaca aaaatgtgat catgatgaaa 10740
gccagcaatg acgttgcagc tgatattgtt gtcggccaac cgcgccgaaa acgcagctgt 10800
cagacccaca gcctccaacg aagaatgtat cgtcaaagtg atccaagcac actcatagtt 10860
ggagtcgtac tccaaaggcg gcaatgacga gtcagacaga tactcgtcga cgtttaaacc 10920
atcatctaag ggcctcaaaa ctacctcgga actgctgcgc tgatctggac accacagagg 10980
ttccgagcac tttaggttgc accaaatgtc ccaccaggtg caggcagaaa acgctggaac 11040

agcgtgtaca gtttgtctta acaaaaagtg agggcgctga ggtcgagcag ggtggtgtga 11100
cttgttatag cctttagagc tgcgaaagcg cgtatggatt tggctcatca ggccagattg 11160
agggtctgtg gacacatgtc atgttagtgt acttcaatcg ccccctggat atagccccga 11220
caataggccg tggcctcatt tttttgcctt ccgcacattt ccattgctcg gtacccacac 11280
cttgcttctc ctgcacttgc caaccttaat actggtttac attgaccaac atcttacaag 11340
cgggggggctt gtctagggta tatataaaca gtggctctcc caatcggttg ccagtctctt 11400
ttttcctttc tttccccaca gattcgaaat ctaaactaca catcacac 11448

<210> 151

<211> 11448

<212> DNA

<213> Artificial Sequence

<220><223> pRF298

<400> 151

catggacaag aaatactcca tcggcctgga cattggaacc aactctgtcg gctgggctgt 60

catcaccgac gagtacaagg tgccctccaa gaaattcaag gtcctcggaa acaccgatcg 120

acactccatc aagaaaaacc tcattggtgc cctgttgttc gattctggcg agactgccga 180

agctaccaga ctcaagcgaa ctgctcggcg acgttacacc cgacggaaga accgaatctg 240

ctacctgcag gagatctttt ccaacgagat ggccaaggtg gacgattcgt tctttcatcg 300

actggaggaa tccttcctcg tcgaggaaga caagaaacac gagcgtcatc ccatctttgg 360

caacattgtg gacgaggttg cttaccacga gaagtatcct accatctacc atctccgaaa 420

gaaactcgtc gattccaccg acaaggcgga tctcagactt atctacctcg ctctggcaca 480

catgatcaag tttcgaggtc atttcctcat cgagggcgat ctcaatcccg acaacagcga 540

tgtggacaag ctgttcattc agctcgttca gacctacaac cagctgttcg aggaaaaccc 600

catcaatgcc tccggagtcg atgcaaaggc catcttgtct gctcgactct cgaagagcag 660

acgactggag aacctcattg cccaacttcc tggcgagaaa aagaacggac tgtttggcaa 720

cctcattgcc ctttctcttg gtctcacacc caacttcaag tccaacttcg atctggcgga 780

ggacgccaag ctccagctgt ccaaggacac ctacgacgat gacctcgaca acctgcttgc 840

acagattggc gatcagtacg ccgacctgtt tctcgctgcc aagaaccttt cggatgctat 900

tctcttgtct gacattctgc gagtcaacac cgagatcaca aaggctcccc tttctgcctc 960

catgatcaag cgatacgacg agcaccatca ggatctcaca ctgctcaagg ctcttgtccg 1020

acagcaactg cccgagaagt acaaggagat ctttttcgat cagtcgaaga acggctacgc 1080

tggatacatc gacggcggag cctctcagga agagttctac aagttcatca agccaattct 1140

cgagaagatg gacggaaccg aggaactgct tgtcaagctc aatcgagagg atctgcttcg 1200

gaagcaacga accttcgaca acggcagcat tcctcatcag atccacctcg gtgagctgca 1260

cgccattctt cgacgtcagg aagacttcta ccccttttctc aaggacaacc gagagaagat 1320

cgagaagatt cttacctttc gaatccccta ctatgttggt cctcttgcca gaggaaactc 1380

tcgatttgct tggatgactc gaaagtccga ggaaaccatc actccctgga acttcgagga 1440

agtcgtggac aagggtgcct ctgcacagtc cttcatcgag cgaatgacca acttcgacaa 1500

gaatctgccc aacgagaagg ttcttcccaa gcattcgctg ctctacgagt actttacagt 1560

ctacaacgaa ctcaccaaag tcaagtacgt taccgaggga atgcgaaagc ctgccttctt 1620

gtctggcgaa cagaagaaag ccattgtcga tctcctgttc aagaccaacc gaaaggtcac 1680
tgttaagcag ctcaaggagg actacttcaa gaaaatcgag tgtttcgaca gcgtcgagat 1740
ttccggagtt gaggaccgat tcaacgcctc tttgggcacc tatcacgatc tgctcaagat 1800
tatcaaggac aaggattttc tcgacaacga ggaaaacgag gacattctgg aggacatcgt 1860
gctcactctt accctgttcg aagatcggga gatgatcgag gaacgactca agacatacgc 1920
tcacctgttc gacgacaagg tcatgaaaca actcaagcga cgtagataca ccggctgggg 1980
aagactttcg cgaaagctca tcaacggcat cagagacaag cagtccggaa agaccattct 2040

ggactttctc aagtccgatg gctttgccaa ccgaaacttc atgcagctca ttcacgacga 2100
ttctcttacc ttcaaggagg acatccagaa ggcacaagtg tccggtcagg gcgacagctt 2160
gcacgaacat attgccaacc tggctggttc gccagccatc aagaaaggca ttctccagac 2220
tgtcaaggtt gtcgacgagc tggtgaaggt catgggacgt cacaagcccg agaacattgt 2280
gatcgagatg gccagagaga accagacaac tcaaaagggt cagaaaaact cgcgagagcg 2340
gatgaagcga atcgaggaag gcatcaagga gctgggatcc cagattctca aggagcatcc 2400
cgtcgagaac actcaactgc agaacgagaa gctgtatctc tactatctgc agaatggtcg 2460

agacatgtac gtggatcagg aactggacat caatcgtctc agcgactacg atgtggacca 2520
cattgtccct caatcctttc tcaaggacga ttctatcgac aacaaggtcc ttacacgatc 2580
cgacaagaac agaggcaagt cggacaacgt tcccagcgaa gaggtggtca aaaagatgaa 2640
gaactactgg cgacagctgc tcaacgccaa gctcattacc cagcgaaagt tcgacaatct 2700
taccaaggcc gagcgaggcg gtctgtccga gctcgacaag gctggcttca tcaagcgtca 2760
actcgtcgag accagacaga tcacaaagca cgtcgcacag attctcgatt ctcggatgaa 2820
caccaagtac gacgagaacg acaagctcat ccgagaggtc aaggtgatta ctctcaagtc 2880

caaactggtc tccgatttcc gaaaggactt tcagttctac aaggtgcgag agatcaacaa 2940
ttaccaccat gcccacgatg cttacctcaa cgccgtcgtt ggcactgcgc tcatcaagaa 3000
ataccccaag ctcgaaagcg agttcgttta cggcgattac aaggtctacg acgttcgaaa 3060
gatgattgcc aagtccgaac aggagattgg caaggctact gccaagtact tcttttactc 3120
caacatcatg aacttttca agaccgagat caccttggcc aacggagaga ttcgaaagag 3180
accacttatc gagaccaacg gcgaaactgg agagatcgtg tgggacaagg gtcgagactt 3240
tgcaaccgtg cgaaaggttc tgtcgatgcc tcaggtcaac atcgtcaaga aaaccgaggt 3300

tcagactggc ggattctcca aggagtcgat tctgcccaag cgaaactccg acaagctcat 3360

cgctcgaaag aaagactggg atcccaagaa atacggtggc ttcgattctc ctaccgtcgc 3420

ctattccgtg cttgtcgttg cgaaggtcga gaagggcaag tccaaaaagc tcaagtccgt 3480

caaggagctg ctcggaatta ccatcatgga gcgatcgagc ttcgagaaga atcccatcga 3540

cttcttggaa gccaagggtt acaaggaggt caagaaagac ctcattatca agctgcccaa 3600

gtactctctg ttcgaactgg agaacggtcg aaagcgtatg ctcgcctccg ctggcgagct 3660

gcagaaggga aacgagcttg ccttgccttc gaagtacgtc aactttctct atctggcttc 3720

tcactacgag aagctcaagg gttctcccga ggacaacgaa cagaagcaac tcttcgttga 3780

gcagcacaaa cattacctcg acgagattat cgagcagatt tccgagtttt cgaagcgagt 3840

catcctggct gatgccaact tggacaaggt gctctctgcc tacaacaagc atcgggacaa 3900

acccattcga gaacaggcgg agaacatcat tcacctgttt actcttacca acctgggtgc 3960

tcctgcagct ttcaagtact tcgataccac tatcgaccga aagcggtaca catccaccaa 4020

ggaggttctc gatgccaccc tgattcacca gtccatcact ggcctgtacg agacccgaat 4080

cgacctgtct cagcttggtg gcgactccag agccgatccc aagaaaaagc gaaaggtcta 4140

agcggccgca agtgtggatg gggaagtgag tgcccggttc tgtgtgcaca attggcaatc 4200

caagatggat ggattcaaca cagggatata gcgagctacg tggtggtgcg aggatatagc 4260

aacggatatt tatgtttgac acttgagaat gtacgataca agcactgtcc aagtacaata 4320

ctaaacatac tgtacatact catactcgta cccgggcaac ggtttcactt gagtgcagtg 4380

gctagtgctc ttactcgtac agtgtgcaat actgcgtatc atagtctttg atgtatatcg 4440

tattcattca tgttagttgc gtacgagccg gaagcataaa gtgtaaagcc tggggtgcct 4500

aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa 4560

acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta 4620

ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc 4680

gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg 4740

caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt 4800

tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa 4860

gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct 4920

ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc 4980

cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg 5040

tcgttcgctc caagctgggc tgtgtgcacg aaccccccgt tcagcccgac cgctgcgcct 5100

tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag 5160

cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga 5220

agtggtggcc taactacggc tacactagaa ggacagtatt tggtatctgc gctctgctga 5280

agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg 5340

gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag 5400

aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag 5460

ggattttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat 5520

gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct 5580

taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac 5640

tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa 5700

tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg 5760

gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt 5820

gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca 5880

ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt 5940

cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct 6000

tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg 6060

cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg 6120

agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg 6180

cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa 6240

aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt 6300

aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt 6360

gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt 6420

gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca 6480

tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat 6540

ttccccgaaa agtgccacct gacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg 6600

tggttacgcg cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt 6660

tcttcccttc ctttctcgcc acgttcgccg gctttccccg tcaagctcta aatcgggggc 6720

tcccttttagg gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg 6780

gtgatggttc acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg 6840

agtccacgtt ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct 6900

cggtctattc ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg 6960

agctgattta acaaaaattt aacgcgaatt ttaacaaaat attaacgctt acaatttcca 7020

ttcgccattc aggctgcgca actgttggga agggcgatcg gtgcgggcct cttcgctatt 7080

acgccagctg gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa cgccagggtt 7140
ttcccagtca cgacgttgta aaacgacggc cagtgaattg taatacgact cactataggg 7200
cgaattgggt accgggcccc ccctcgaggt cgatggtgtc gataagcttg atatcgaatt 7260
catgtcacac aaaccgatct tcgcctcaag gaaacctaat tctacatccg agagactgcc 7320
gagatccagt ctacactgat taattttcgg gccaataatt taaaaaaatc gtgttatata 7380
atattatatg tattatatat atacatcatg atgatactga cagtcatgtc ccattgctaa 7440
atagacagac tccatctgcc gcctccaact gatgttctca atatttaagg ggtcatctcg 7500

cattgtttaa taataaacag actccatcta ccgcctccaa atgatgttct caaaatatat 7560
tgtatgaact tatttttatt acttagtatt attagacaac ttacttgctt tatgaaaaac 7620
acttcctatt taggaaacaa tttataatgg cagttcgttc atttaacaat ttatgtagaa 7680
taaatgttat aaatgcgtat gggaaatctt aaatatggat agcataaatg atatctgcat 7740
tgcctaattc gaaatcaaca gcaacgaaaa aaatcccttg tacaacataa atagtcatcg 7800
agaaatatca actatcaaag aacagctatt cacacgttac tattgagatt attattggac 7860
gagaatcaca cactcaactg tctttctctc ttctagaaat acaggtacaa gtatgtacta 7920

ttctcattgt tcatacttct agtcatttca tcccacatat tccttggatt tctctccaat 7980
gaatgacatt ctatcttgca aattcaacaa ttataataag atataccaaa gtagcggtat 8040
agtggcaatc aaaaagcttc tctggtgtgc ttctcgtatt tatttttatt ctaatgatcc 8100
attaaaggta tatatttatt tcttgttata taatcctttt gtttattaca tgggctggat 8160
acataaaggt attttgattt aatttttttgc ttaaattcaa tccccccctcg ttcagtgtca 8220
actgtaatgg taggaaatta ccatactttt gaagaagcaa aaaaaatgaa agaaaaaaaa 8280
aatcgtattt ccaggttaga cgttccgcag aatctagaat gcggtatgcg gtacattgtt 8340

cttcgaacgt aaaagttgcg ctccctgaga tattgtacat ttttgctttt acaagtacaa 8400
gtacatcgta caactatgta ctactgttga tgcatccaca acagtttgtt ttgttttttt 8460
ttgttttttt ttttttctaat gattcattac cgctatgtat acctacttgt acttgtagta 8520
agccgggtta ttggcgttca attaatcata gacttatgaa tctgcacggt gtgcgctgcg 8580
agttactttt agcttatgca tgctacttgg gtgtaatatt gggatctgtt cggaaatcaa 8640
cggatgctca atcgatcgcc agcaacgcgg ccttagacat aaaaaacaaa aaaaaaaagc 8700
accgactcgg tgccactttt tcaagttgat aacggactag ccttatttta acttgctatt 8760

tctagctcta aaacgagtaa cggtaacgac accggacgag cttactcgtt tcgtcctcac 8820

ggactcatca gcggtgttac gcatggttga tgtgtgttta attcaagaat gaatatagag 8880

aagagaagaa gaaaaaagat tcaattgagc cggcgatgca gacccttata taaatgttgc 8940

cttggacaga cggagcaagc ccgcccaaac ctacgttcgg tataatatgt taagcttttt 9000

aacacaaagg tttggcttgg ggtaacctga tgtggtgcaa aagaccgggc gttggcgagc 9060

cattgcgcgg gcgaatgggg tcgtgactcg tctcaaattc gagggcgtgc ctcaattcgt 9120

gcccccgtgg ctttttcccg ccgtttccgc cccgtttgca ccactgcagc cgcttctttg 9180

gttcggacac cttgctgcga gctaggtgcc ttgtgctact taaaaagtgg cctcccaaca 9240

ccaacatgac atgagtgcgt gggccaagac acgttggcgg ggtcgcagtc ggctcaatgg 9300

cccggaaaaa acgctgctgg agctggttcg gacgcagtcc gccgcggcgt atggatatcc 9360

gcaaggttcc atagcgccat tgccctccgt cggcgtctat cccgcaacct taattaagtc 9420

atacacaagt cagctttctt cgagcctcat ataagtataa gtagttcaac gtattagcac 9480

tgtacccagc atctccgtat cgagaaacac aacaacatgc cccattggac agatcatgcg 9540

gatacacagg ttgtgcagta tcatacatac tcgatcagac aggtcgtctg accatcatac 9600

aagctgaaca agcgctccat acttgcacgc tctctatata cacagttaaa ttacatatcc 9660

atagtctaac ctctaacagt taatcttctg gtaagcctcc cagccagcct tctggtatcg 9720

cttggcctcc tcaataggat ctcggttctg gccgtacaga cctcggccga caattatgat 9780

atccgttccg gtagacatga catcctcaac agttcggtac tgctgtccga gagcgtctcc 9840

cttgtcgtca agacccaccc cgggggtcag aataagccag tcctcagagt cgcccttagg 9900

tcggttctgg gcaatgaagc caaccacaaa ctcggggtcg gatcgggcaa gctcaatggt 9960

ctgcttggag tactcgccag tggccagaga gcccttgcaa gacagctcgg ccagcatgag 10020

cagacctctg gccagcttct cgttgggaga ggggactagg aactccttgt actgggagtt 10080

ctcgtagtca gagacgtcct ccttcttctg ttcagagaca gtttcctcgg caccagctcg 10140

caggccagca atgattccgg ttccgggtac accgtgggcg ttggtgatat cggaccactc 10200

ggcgattcgg tgacaccggt actggtgctt gacagtgttg ccaatatctg cgaactttct 10260

gtcctcgaac aggaagaaac cgtgcttaag agcaagttcc ttgaggggga gcacagtgcc 10320

ggcgtaggtg aagtcgtcaa tgatgtcgat atgggttttg atcatgcaca cataaggtcc 10380

gaccttatcg gcaagctcaa tgagctcctt ggtggtggta acatccagag aagcacacag 10440

gttggttttc ttggctgcca cgagcttgag cactcgagcg gcaaaggcgg acttgtggac 10500

gttagctcga gcttcgtagg agggcatttt ggtggtgaag aggagactga aataaattta 10560

gtctgcagaa ctttttatcg gaaccttatc tggggcagtg aagtatatgt tatggtaata 10620

gttacgagtt agttgaactt atagatagac tggactatac ggctatcggt ccaaattaga 10680

aagaacgtca atggctctct gggcgtcgcc tttgccgaca aaaatgtgat catgatgaaa 10740

gccagcaatg acgttgcagc tgatattgtt gtcggccaac cgcgccgaaa acgcagctgt 10800

cagacccaca gcctccaacg aagaatgtat cgtcaaagtg atccaagcac actcatagtt 10860

ggagtcgtac tccaaaggcg gcaatgacga gtcagacaga tactcgtcga cgtttaaacc 10920

atcatctaag ggcctcaaaa ctacctcgga actgctgcgc tgatctggac accacagagg 10980

ttccgagcac tttaggttgc accaaatgtc ccaccaggtg caggcagaaa acgctggaac 11040

agcgtgtaca gtttgtctta acaaaaagtg agggcgctga ggtcgagcag ggtggtgtga 11100

cttgttatag cctttagagc tgcgaaagcg cgtatggatt tggctcatca ggccagattg 11160

agggtctgtg gacacatgtc atgttagtgt acttcaatcg ccccctggat atagccccga 11220

caataggccg tggcctcatt tttttgcctt ccgcacattt ccattgctcg gtacccacac 11280

cttgcttctc ctgcacttgc caaccttaat actggtttac attgaccaac atcttacaag 11340

cgggggggctt gtctagggta tatataaaca gtggctctcc caatcggttg ccagtctctt 11400

ttttcctttc tttccccaca gattcgaaat ctaaactaca catcacac 11448

<210> 152

<211> 11448

<212> DNA

<213> Artificial Sequence

<220><223> pRF300

<400> 152

catggacaag aaatactcca tcggcctgga cattggaacc aactctgtcg gctgggctgt 60

catcaccgac gagtacaagg tgccctccaa gaaattcaag gtcctcggaa acaccgatcg 120

acactccatc aagaaaaacc tcattggtgc cctgttgttc gattctggcg agactgccga 180

agctaccaga ctcaagcgaa ctgctcggcg acgttacacc cgacggaaga accgaatctg 240

ctacctgcag gagatctttt ccaacgagat ggccaaggtg gacgattcgt tctttcatcg 300

actggaggaa tccttcctcg tcgaggaaga caagaaacac gagcgtcatc ccatctttgg 360

caacattgtg gacgaggttg cttaccacga gaagtatcct accatctacc atctccgaaa 420

gaaactcgtc gattccaccg acaaggcgga tctcagactt atctacctcg ctctggcaca 480

catgatcaag tttcgaggtc atttcctcat cgagggcgat ctcaatcccg acaacagcga 540

tgtggacaag ctgttcattc agctcgttca gacctacaac cagctgttcg aggaaaaccc 600

catcaatgcc tccggagtcg atgcaaaggc catcttgtct gctcgactct cgaagagcag 660

acgactggag aacctcattg cccaacttcc tggcgagaaa aagaacggac tgtttggcaa 720

cctcattgcc ctttctcttg gtctcacacc caacttcaag tccaacttcg atctggcgga 780

ggacgccaag ctccagctgt ccaaggacac ctacgacgat gacctcgaca acctgcttgc 840

acagattggc gatcagtacg ccgacctgtt tctcgctgcc aagaaccttt cggatgctat 900

tctcttgtct gacattctgc gagtcaacac cgagatcaca aaggctcccc tttctgcctc 960

catgatcaag cgatacgacg agcaccatca ggatctcaca ctgctcaagg ctcttgtccg 1020

acagcaactg cccgagaagt acaaggagat ctttttcgat cagtcgaaga acggctacgc 1080

tggatacatc gacggcggag cctctcagga agagttctac aagttcatca agccaattct 1140

cgagaagatg gacggaaccg aggaactgct tgtcaagctc aatcgagagg atctgcttcg 1200

gaagcaacga accttcgaca acggcagcat tcctcatcag atccacctcg gtgagctgca 1260

cgccattctt cgacgtcagg aagacttcta ccccttttctc aaggacaacc gagagaagat 1320

cgagaagatt cttacctttc gaatccccta ctatgttggt cctcttgcca gaggaaactc 1380

tcgatttgct tggatgactc gaaagtccga ggaaaccatc actccctgga acttcgagga 1440

agtcgtggac aagggtgcct ctgcacagtc cttcatcgag cgaatgacca acttcgacaa 1500

gaatctgccc aacgagaagg ttcttcccaa gcattcgctg ctctacgagt actttacagt 1560

ctacaacgaa ctcaccaaag tcaagtacgt taccgaggga atgcgaaagc ctgccttctt 1620

gtctggcgaa cagaagaaag ccattgtcga tctcctgttc aagaccaacc gaaaggtcac 1680

tgttaagcag ctcaaggagg actacttcaa gaaaatcgag tgtttcgaca gcgtcgagat 1740

ttccggagtt gaggaccgat tcaacgcctc tttgggcacc tatcacgatc tgctcaagat 1800

tatcaaggac aaggattttc tcgacaacga ggaaaacgag gacattctgg aggacatcgt 1860

gctcactctt accctgttcg aagatcggga gatgatcgag gaacgactca agacatacgc 1920

tcacctgttc gacgacaagg tcatgaaaca actcaagcga cgtagataca ccggctgggg 1980

aagactttcg cgaaagctca tcaacggcat cagagacaag cagtccggaa agaccattct 2040

ggactttctc aagtccgatg gctttgccaa ccgaaacttc atgcagctca ttcacgacga 2100

ttctcttacc ttcaaggagg acatccagaa ggcacaagtg tccggtcagg gcgacagctt 2160

gcacgaacat attgccaacc tggctggttc gccagccatc aagaaaggca ttctccagac 2220

tgtcaaggtt gtcgacgagc tggtgaaggt catgggacgt cacaagcccg agaacattgt 2280

gatcgagatg gccagagaga accagacaac tcaaaagggt cagaaaaact cgcgagagcg 2340

gatgaagcga atcgaggaag gcatcaagga gctgggatcc cagattctca aggagcatcc 2400

cgtcgagaac actcaactgc agaacgagaa gctgtatctc tactatctgc agaatggtcg 2460

agacatgtac gtggatcagg aactggacat caatcgtctc agcgactacg atgtggacca 2520

cattgtccct caatcctttc tcaaggacga ttctatcgac aacaaggtcc ttacacgatc 2580

cgacaagaac agaggcaagt cggacaacgt tcccagcgaa gaggtggtca aaaagatgaa 2640

gaactactgg cgacagctgc tcaacgccaa gctcattacc cagcgaaagt tcgacaatct 2700

taccaaggcc gagcgaggcg gtctgtccga gctcgacaag gctggcttca tcaagcgtca 2760

actcgtcgag accagacaga tcacaaagca cgtcgcacag attctcgatt ctcggatgaa 2820

caccaagtac gacgagaacg acaagctcat ccgagaggtc aaggtgatta ctctcaagtc 2880

caaactggtc tccgatttcc gaaaggactt tcagttctac aaggtgcgag agatcaacaa 2940

ttaccaccat gcccacgatg cttacctcaa cgccgtcgtt ggcactgcgc tcatcaagaa 3000

ataccccaag ctcgaaagcg agttcgttta cggcgattac aaggtctacg acgttcgaaa 3060

gatgattgcc aagtccgaac aggagattgg caaggctact gccaagtact tcttttactc 3120

caacatcatg aacttttca agaccgagat caccttggcc aacggagaga ttcgaaagag 3180

accacttatc gagaccaacg gcgaaactgg agagatcgtg tgggacaagg gtcgagactt 3240

tgcaaccgtg cgaaaggttc tgtcgatgcc tcaggtcaac atcgtcaaga aaaccgaggt 3300

tcagactggc ggattctcca aggagtcgat tctgcccaag cgaaactccg acaagctcat 3360

cgctcgaaag aaagactggg atcccaagaa atacggtggc ttcgattctc ctaccgtcgc 3420

ctattccgtg cttgtcgttg cgaaggtcga gaagggcaag tccaaaaagc tcaagtccgt 3480

caaggagctg ctcggaatta ccatcatgga gcgatcgagc ttcgagaaga atcccatcga 3540

cttcttggaa gccaagggtt acaaggaggt caagaaagac ctcattatca agctgcccaa 3600

gtactctctg ttcgaactgg agaacggtcg aaagcgtatg ctcgcctccg ctggcgagct 3660

gcagaaggga aacgagcttg ccttgccttc gaagtacgtc aactttctct atctggcttc 3720

tcactacgag aagctcaagg gttctcccga ggacaacgaa cagaagcaac tcttcgttga 3780

gcagcacaaa cattacctcg acgagattat cgagcagatt tccgagtttt cgaagcgagt 3840

catcctggct gatgccaact tggacaaggt gctctctgcc tacaacaagc atcgggacaa 3900

acccattcga gaacaggcgg agaacatcat tcacctgttt actcttacca acctgggtgc 3960

tcctgcagct ttcaagtact tcgataccac tatcgaccga aagcggtaca catccaccaa 4020

ggaggttctc gatgccaccc tgattcacca gtccatcact ggcctgtacg agacccgaat 4080

cgacctgtct cagcttggtg gcgactccag agccgatccc aagaaaaagc gaaaggtcta 4140

agcggccgca agtgtggatg gggaagtgag tgcccggttc tgtgtgcaca attggcaatc 4200

caagatggat ggattcaaca cagggatata gcgagctacg tggtggtgcg aggatatagc 4260

aacggatatt tatgtttgac acttgagaat gtacgataca agcactgtcc aagtacaata 4320

ctaaacatac tgtacatact catactcgta cccgggcaac ggtttcactt gagtgcagtg 4380

gctagtgctc ttactcgtac agtgtgcaat actgcgtatc atagtctttg atgtatatcg 4440
tattcattca tgttagttgc gtacgagccg gaagcataaa gtgtaaagcc tggggtgcct 4500
aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa 4560
acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta 4620
ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc 4680
gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg 4740
caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt 4800

tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa 4860
gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct 4920
ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc 4980
cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg 5040
tcgttcgctc caagctgggc tgtgtgcacg aacccccccgt tcagcccgac cgctgcgcct 5100
tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag 5160
cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga 5220

agtggtggcc taactacggc tacactagaa ggacagtatt tggtatctgc gctctgctga 5280
agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg 5340
gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag 5400
aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag 5460
ggattttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat 5520
gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct 5580
taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac 5640

tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa 5700
tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg 5760
gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt 5820
gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca 5880
ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt 5940
cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct 6000
tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg 6060

cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg 6120

agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg 6180

cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa 6240

aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt 6300

aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt 6360

gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt 6420

gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca 6480

tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat 6540

ttccccgaaa agtgccacct gacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg 6600

tggttacgcg cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt 6660

tcttcccttc ctttctcgcc acgttcgccg gctttccccg tcaagctcta aatcgggggc 6720

tcccttttagg gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg 6780

gtgatggttc acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg 6840

agtccacgtt ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct 6900

cggtctattc ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg 6960

agctgattta acaaaaattt aacgcgaatt ttaacaaaat attaacgctt acaatttcca 7020

ttcgccattc aggctgcgca actgttggga agggcgatcg gtgcgggcct cttcgctatt 7080

acgccagctg gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa cgccagggtt 7140

ttcccagtca cgacgttgta aaacgacggc cagtgaattg taatacgact cactataggg 7200

cgaattgggt accgggcccc ccctcgaggt cgatggtgtc gataagcttg atatcgaatt 7260

catgtcacac aaaccgatct tcgcctcaag gaaacctaat tctacatccg agagactgcc 7320

gagatccagt ctacactgat taattttcgg gccaataatt taaaaaaatc gtgttatata 7380

atattatatg tattatatat atacatcatg atgatactga cagtcatgtc ccattgctaa 7440

atagacagac tccatctgcc gcctccaact gatgttctca atatttaagg ggtcatctcg 7500

cattgtttaa taataaacag actccatcta ccgcctccaa atgatgttct caaaatatat 7560

tgtatgaact tatttttatt acttagtatt attagacaac ttacttgctt tatgaaaaac 7620

acttcctatt taggaaacaa tttataatgg cagttcgttc atttaacaat ttatgtagaa 7680

taaatgttat aaatgcgtat gggaaatctt aaatatggat agcataaatg atatctgcat 7740

tgcctaattc gaaatcaaca gcaacgaaaa aaatcccttg tacaacataa atagtcatcg 7800

agaaatatca actatcaaag aacagctatt cacacgttac tattgagatt attattggac 7860

gagaatcaca cactcaactg tctttctctc ttctagaaat acaggtacaa gtatgtacta 7920

ttctcattgt tcatacttct agtcatttca tcccacatat tccttggatt tctctccaat 7980

gaatgacatt ctatcttgca aattcaacaa ttataataag atataccaaa gtagcggtat 8040

agtggcaatc aaaaagcttc tctggtgtgc ttctcgtatt tatttttatt ctaatgatcc 8100

attaaaggta tatatttatt tcttgttata taatcctttt gtttattaca tgggctggat 8160

acataaaggt attttgattt aatttttttgc ttaaattcaa tccccctcg ttcagtgtca 8220

actgtaatgg taggaaatta ccatactttt gaagaagcaa aaaaaatgaa agaaaaaaaa 8280

aatcgtattt ccaggttaga cgttccgcag aatctagaat gcggtatgcg gtacattgtt 8340

cttcgaacgt aaaagttgcg ctccctgaga tattgtacat ttttgctttt acaagtacaa 8400

gtacatcgta caactatgta ctactgttga tgcatccaca acagtttgtt ttgttttttt 8460

ttgttttttt tttttctaat gattcattac cgctatgtat acctacttgt acttgtagta 8520

agccgggtta ttggcgttca attaatcata gacttatgaa tctgcacggt gtgcgctgcg 8580

agttactttt agcttatgca tgctacttgg gtgtaatatt gggatctgtt cggaaatcaa 8640

cggatgctca atcgatcgcc agcaacgcgg ccttagacat aaaaaacaaa aaaaaaaagc 8700

accgactcgg tgccactttt tcaagttgat aacggactag ccttatttta acttgctatt 8760

tctagctcta aaaccgaagc tgaaggtgac caaggacgag cttactcgtt tcgtcctcac 8820

ggactcatca gcctggttac gcatggttga tgtgtgttta attcaagaat gaatatagag 8880

aagagaagaa gaaaaaagat tcaattgagc cggcgatgca gacccttata taaatgttgc 8940

cttggacaga cggagcaagc ccgcccaaac ctacgttcgg tataatatgt taagcttttt 9000

aacacaaagg tttggcttgg ggtaacctga tgtggtgcaa aagaccgggc gttggcgagc 9060

cattgcgcgg gcgaatgggg tcgtgactcg tctcaaattc gagggcgtgc ctcaattcgt 9120

gcccccgtgg ctttttcccg ccgtttccgc cccgtttgca ccactgcagc cgcttctttg 9180

gttcggacac cttgctgcga gctaggtgcc ttgtgctact taaaaagtgg cctcccaaca 9240

ccaacatgac atgagtgcgt gggccaagac acgttggcgg ggtcgcagtc ggctcaatgg 9300

cccggaaaaa acgctgctgg agctggttcg gacgcagtcc gccgcggcgt atggatatcc 9360

gcaaggttcc atagcgccat tgccctccgt cggcgtctat cccgcaacct taattaagtc 9420

atacacaagt cagctttctt cgagcctcat ataagtataa gtagttcaac gtattagcac 9480

tgtacccagc atctccgtat cgagaaacac aacaacatgc cccattggac agatcatgcg 9540

gatacacagg ttgtgcagta tcatacatac tcgatcagac aggtcgtctg accatcatac 9600

aagctgaaca agcgctccat acttgcacgc tctctatata cacagttaaa ttacatatcc 9660

atagtctaac ctctaacagt taatcttctg gtaagcctcc cagccagcct tctggtatcg 9720

cttggcctcc tcaataggat ctcggttctg gccgtacaga cctcggccga caattatgat 9780

atccgttccg gtagacatga catcctcaac agttcggtac tgctgtccga gagcgtctcc 9840

cttgtcgtca agacccaccc cgggggtcag aataagccag tcctcagagt cgcccttagg 9900

tcggttctgg gcaatgaagc caaccacaaa ctcggggtcg gatcgggcaa gctcaatggt 9960

ctgcttggag tactcgccag tggccagaga gcccttgcaa gacagctcgg ccagcatgag 10020

cagacctctg gccagcttct cgttgggaga ggggactagg aactccttgt actgggagtt 10080

ctcgtagtca gagacgtcct ccttcttctg ttcagagaca gtttcctcgg caccagctcg 10140

caggccagca atgattccgg ttccgggtac accgtgggcg ttggtgatat cggaccactc 10200

ggcgattcgg tgacaccggt actggtgctt gacagtgttg ccaatatctg cgaactttct 10260

gtcctcgaac aggaagaaac cgtgcttaag agcaagttcc ttgaggggga gcacagtgcc 10320

ggcgtaggtg aagtcgtcaa tgatgtcgat atgggttttg atcatgcaca cataaggtcc 10380

gaccttatcg gcaagctcaa tgagctcctt ggtggtggta acatccagag aagcacacag 10440

gttggttttc ttggctgcca cgagcttgag cactcgagcg gcaaaggcgg acttgtggac 10500

gttagctcga gcttcgtagg agggcatttt ggtggtgaag aggagactga aataaattta 10560

gtctgcagaa ctttttatcg gaaccttatc tggggcagtg aagtatatgt tatggtaata 10620

gttacgagtt agttgaactt atagatagac tggactatac ggctatcggt ccaaattaga 10680

aagaacgtca atggctctct gggcgtcgcc tttgccgaca aaaatgtgat catgatgaaa 10740

gccagcaatg acgttgcagc tgatattgtt gtcggccaac cgcgccgaaa acgcagctgt 10800

cagacccaca gcctccaacg aagaatgtat cgtcaaagtg atccaagcac actcatagtt 10860

ggagtcgtac tccaaaggcg gcaatgacga gtcagacaga tactcgtcga cgtttaaacc 10920

atcatctaag ggcctcaaaa ctacctcgga actgctgcgc tgatctggac accacagagg 10980

ttccgagcac tttaggttgc accaaatgtc ccaccaggtg caggcagaaa acgctggaac 11040

agcgtgtaca gtttgtctta acaaaaagtg agggcgctga ggtcgagcag ggtggtgtga 11100

cttgttatag cctttagagc tgcgaaagcg cgtatggatt tggctcatca ggccagattg 11160

agggtctgtg gacacatgtc atgttagtgt acttcaatcg ccccctggat atagccccga 11220

caataggccg tggcctcatt tttttgcctt ccgcacattt ccattgctcg gtacccacac 11280

cttgcttctc ctgcacttgc caaccttaat actggtttac attgaccaac atcttacaag 11340

cgggggggctt gtctagggta tatataaaca gtggctctcc caatcggttg ccagtctctt 11400

ttttcctttc tttccccaca gattcgaaat ctaaactaca catcacac 11448

<210> 153

<211> 11448

<212> DNA

<213> Artificial sequence

<220><223> pRF339

<400> 153

catggacaag aaatactcca tcggcctggc cattggaacc aactctgtcg gctgggctgt 60

catcaccgac gagtacaagg tgccctccaa gaaattcaag gtcctcggaa acaccgatcg 120

acactccatc aagaaaaacc tcattggtgc cctgttgttc gattctggcg agactgccga 180

agctaccaga ctcaagcgaa ctgctcggcg acgttacacc cgacggaaga accgaatctg 240

ctacctgcag gagatctttt ccaacgagat ggccaaggtg gacgattcgt tctttcatcg 300

actggaggaa tccttcctcg tcgaggaaga caagaaacac gagcgtcatc ccatctttgg 360

caacattgtg gacgaggttg cttaccacga gaagtatcct accatctacc acctgcgaaa 420

gaaactcgtc gattccaccg acaaggcgga tctcagactt atctacctcg ctctggcaca 480

catgatcaag tttcgaggtc atttcctcat cgagggcgat ctcaatcccg acaacagcga 540

tgtggacaag ctgttcattc agctcgttca gacctacaac cagctgttcg aggaaaaccc 600

catcaatgcc tccggagtcg atgcaaaggc catcttgtct gctcgactct cgaagagcag 660

acgactggag aacctcattg cccaacttcc tggcgagaaa aagaacggac tgtttggcaa 720

cctcattgcc ctttctcttg gtctcacacc caacttcaag tccaacttcg atctggcgga 780

ggacgccaag ctccagctgt ccaaggacac ctacgacgat gacctcgaca acctgcttgc 840

acagattggc gatcagtacg ccgacctgtt tctcgctgcc aagaaccttt cggatgctat 900

tctcttgtct gacattctgc gagtcaacac cgagatcaca aaggctcccc tttctgcctc 960

catgatcaag cgatacgacg agcaccatca ggatctcaca ctgctcaagg ctcttgtccg 1020

acagcaactg cccgagaagt acaaggagat ctttttcgat cagtcgaaga acggctacgc 1080

tggatacatc gacggcggag cctctcagga agagttctac aagttcatca agccaattct 1140

cgagaagatg gacggaaccg aggaactgct tgtcaagctc aatcgagagg atctgcttcg 1200

gaagcaacga accttcgaca acggcagcat tcctcatcag atccacctcg gtgagctgca 1260

cgccattctt cgacgtcagg aagacttcta ccccttttctc aaggacaacc gagagaagat 1320

cgagaagatt cttacctttc gaatccccta ctatgttggt cctcttgcca gaggaaactc 1380

tcgatttgct tggatgactc gaaagtccga ggaaaccatc actccctgga acttcgagga 1440

agtcgtggac aagggtgcct ctgcacagtc cttcatcgag cgaatgacca acttcgacaa 1500

gaatctgccc aacgagaagg ttcttcccaa gcattcgctg ctctacgagt actttacagt 1560

ctacaacgaa ctcaccaaag tcaagtacgt taccgaggga atgcgaaagc ctgccttctt 1620

gtctggcgaa cagaagaaag ccattgtcga tctcctgttc aagaccaacc gaaaggtcac 1680

tgttaagcag ctcaaggagg actacttcaa gaaaatcgag tgtttcgaca gcgtcgagat 1740

ttccggagtt gaggaccgat tcaacgcctc tttgggcacc tatcacgatc tgctcaagat 1800

tatcaaggac aaggattttc tcgacaacga ggaaaacgag gacattctgg aggacatcgt 1860

gctcactctt accctgttcg aagatcggga gatgatcgag gaacgactca agacatacgc 1920

tcacctgttc gacgacaagg tcatgaaaca actcaagcga cgtagataca ccggctgggg 1980

aagactttcg cgaaagctca tcaacggcat cagagacaag cagtccggaa agaccattct 2040

ggactttctc aagtccgatg gctttgccaa ccgaaacttc atgcagctca ttcacgacga 2100

ttctcttacc ttcaaggagg acatccagaa ggcacaagtg tccggtcagg gcgacagctt 2160

gcacgaacat attgccaacc tggctggttc gccagccatc aagaaaggca ttctccagac 2220

tgtcaaggtt gtcgacgagc tggtgaaggt catgggacgt cacaagcccg agaacattgt 2280

gatcgagatg gccagagaga accagacaac tcaaaagggt cagaaaaact cgcgagagcg 2340

gatgaagcga atcgaggaag gcatcaagga gctgggatcc cagattctca aggagcatcc 2400

cgtcgagaac actcaactgc agaacgagaa gctgtatctc tactatctgc agaatggtcg 2460

agacatgtac gtggatcagg aactggacat caatcgtctc agcgactacg atgtggacgc 2520

cattgtccct caatcctttc tcaaggacga ttctatcgac aacaaggtcc ttacacgatc 2580

cgacaagaac agaggcaagt cggacaacgt tcccagcgaa gaggtggtca aaaagatgaa 2640

gaactactgg cgacagctgc tcaacgccaa gctcattacc cagcgaaagt tcgacaatct 2700

taccaaggcc gagcgaggcg gtctgtccga gctcgacaag gctggcttca tcaagcgtca 2760

actcgtcgag accagacaga tcacaaagca cgtcgcacag attctcgatt ctcggatgaa 2820

caccaagtac gacgagaacg acaagctcat ccgagaggtc aaggtgatta ctctcaagtc 2880

caaactggtc tccgatttcc gaaaggactt tcagttctac aaggtgcgag agatcaacaa 2940

ttaccaccat gcccacgatg cttacctcaa cgccgtcgtt ggcactgcgc tcatcaagaa 3000

ataccccaag ctcgaaagcg agttcgttta cggcgattac aaggtctacg acgttcgaaa 3060

gatgattgcc aagtccgaac aggagattgg caaggctact gccaagtact tcttttactc 3120

caacatcatg aacttttca agaccgagat caccttggcc aacggagaga ttcgaaagag 3180

accacttatc gagaccaacg gcgaaactgg agagatcgtg tgggacaagg gtcgagactt 3240

tgcaaccgtg cgaaaggttc tgtcgatgcc tcaggtcaac atcgtcaaga aaaccgaggt 3300

tcagactggc ggattctcca aggagtcgat tctgcccaag cgaaactccg acaagctcat 3360

cgctcgaaag aaagactggg atcccaagaa atacggtggc ttcgattctc ctaccgtcgc 3420

ctattccgtg cttgtcgttg cgaaggtcga gaagggcaag tccaaaaagc tcaagtccgt 3480

caaggagctg ctcggaatta ccatcatgga gcgatcgagc ttcgagaaga atcccatcga 3540

cttcttggaa gccaagggtt acaaggaggt caagaaagac ctcattatca agctgcccaa 3600

gtactctctg ttcgaactgg agaacggtcg aaagcgtatg ctcgcctccg ctggcgagct 3660

gcagaaggga aacgagcttg ccttgccttc gaagtacgtc aactttctct atctggcttc 3720

tcactacgag aagctcaagg gttctcccga ggacaacgaa cagaagcaac tcttcgttga 3780

gcagcacaaa cattacctcg acgagattat cgagcagatt tccgagtttt cgaagcgagt 3840

catcctggct gatgccaact tggacaaggt gctctctgcc tacaacaagc atcgggacaa 3900

acccattcga gaacaggcgg agaacatcat tcacctgttt actcttacca acctgggtgc 3960

tcctgcagct ttcaagtact tcgataccac tatcgaccga aagcggtaca catccaccaa 4020

ggaggttctc gatgccaccc tgattcacca gtccatcact ggcctgtacg agacccgaat 4080

cgacctgtct cagcttggtg gcgactccag agccgatccc aagaaaaagc gaaaggtcta 4140

agcggccgca agtgtggatg gggaagtgag tgcccggttc tgtgtgcaca attggcaatc 4200

caagatggat ggattcaaca cagggatata gcgagctacg tggtggtgcg aggatatagc 4260

aacggatatt tatgtttgac acttgagaat gtacgataca agcactgtcc aagtacaata 4320

ctaaacatac tgtacatact catactcgta cccgggcaac ggtttcactt gagtgcagtg 4380

gctagtgctc ttactcgtac agtgtgcaat actgcgtatc atagtctttg atgtatatcg 4440

tattcattca tgttagttgc gtacgagccg gaagcataaa gtgtaaagcc tggggtgcct 4500

aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa 4560

acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta 4620

ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc 4680

gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg 4740

caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt 4800

tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa 4860

gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct 4920

ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc 4980

cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg 5040

tcgttcgctc caagctgggc tgtgtgcacg aaccccccgt tcagcccgac cgctgcgcct 5100

tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag 5160

cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga 5220

agtggtggcc taactacggc tacactagaa ggacagtatt tggtatctgc gctctgctga 5280

agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg 5340

gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag 5400

aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag 5460

ggattttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat 5520

gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct 5580

taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac 5640

tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa 5700

tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg 5760

gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt 5820

gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca 5880

ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt 5940

cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct 6000

tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg 6060

cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg 6120

agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg 6180

cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa 6240

aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt 6300

aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt 6360

gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt 6420

gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca 6480

tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat 6540

ttccccgaaa agtgccacct gacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg 6600

tggttacgcg cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt 6660

tcttcccttc ctttctcgcc acgttcgccg gctttccccg tcaagctcta aatcgggggc 6720

tcccttttagg gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg 6780

gtgatggttc acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg 6840

agtccacgtt ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct 6900

cggtctattc ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg 6960

agctgattta acaaaaattt aacgcgaatt ttaacaaaat attaacgctt acaatttcca 7020

ttcgccattc aggctgcgca actgttggga agggcgatcg gtgcgggcct cttcgctatt 7080

acgccagctg gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa cgccagggtt 7140

ttcccagtca cgacgttgta aaacgacggc cagtgaattg taatacgact cactataggg 7200
cgaattgggt accgggcccc ccctcgaggt cgatggtgtc gataagcttg atatcgaatt 7260
catgtcacac aaaccgatct tcgcctcaag gaaacctaat tctacatccg agagactgcc 7320
gagatccagt ctacactgat taattttcgg gccaataatt taaaaaaatc gtgttatata 7380
atattatatg tattatatat atacatcatg atgatactga cagtcatgtc ccattgctaa 7440
atagacagac tccatctgcc gcctccaact gatgttctca atatttaagg ggtcatctcg 7500
cattgtttaa taataaacag actccatcta ccgcctccaa atgatgttct caaaatatat 7560

tgtatgaact tatttttatt acttagtatt attagacaac ttacttgctt tatgaaaaac 7620
acttcctatt taggaaacaa tttataatgg cagttcgttc atttaacaat ttatgtagaa 7680
taaatgttat aaatgcgtat gggaaatctt aaatatggat agcataaatg atatctgcat 7740
tgcctaattc gaaatcaaca gcaacgaaaa aaatcccttg tacaacataa atagtcatcg 7800
agaaatatca actatcaaag aacagctatt cacacgttac tattgagatt attattggac 7860
gagaatcaca cactcaactg tctttctctc ttctagaaat acaggtacaa gtatgtacta 7920
ttctcattgt tcatacttct agtcatttca tcccacatat tccttggatt tctctccaat 7980

gaatgacatt ctatcttgca aattcaacaa ttataataag atataccaaa gtagcggtat 8040
agtggcaatc aaaaagcttc tctggtgtgc ttctcgtatt tatttttatt ctaatgatcc 8100
attaaaggta tatatttatt tcttgttata taatcctttt gtttattaca tgggctggat 8160
acataaaggt attttgattt aatttttgc ttaaattcaa tccccctcg ttcagtgtca 8220
actgtaatgg taggaaatta ccatactttt gaagaagcaa aaaaaatgaa agaaaaaaaa 8280
aatcgtattt ccaggttaga cgttccgcag aatctagaat gcggtatgcg gtacattgtt 8340
cttcgaacgt aaaagttgcg ctccctgaga tattgtacat ttttgctttt acaagtacaa 8400

gtacatcgta caactatgta ctactgttga tgcatccaca acagtttgtt ttgttttttt 8460
ttgttttttt tttttctaat gattcattac cgctatgtat acctacttgt acttgtagta 8520
agccgggtta ttggcgttca attaatcata gacttatgaa tctgcacggt gtgcgctgcg 8580
agttactttt agcttatgca tgctacttgg gtgtaatatt gggatctgtt cggaaatcaa 8640
cggatgctca atcgatcgcc agcaacgcgg ccttagacat aaaaaacaaa aaaaaaaagc 8700
accgactcgg tgccactttt tcaagttgat aacggactag ccttatttta acttgctatt 8760
tctagctcta aaactagaac cgtactgaaa ctgagacgag cttactcgtt tcgtcctcac 8820

ggactcatca gtcagtttac gcatggttga tgtgtgttta attcaagaat gaatatagag 8880

aagagaagaa gaaaaaagat tcaattgagc cggcgatgca gacccttata taaatgttgc 8940

cttggacaga cggagcaagc ccgcccaaac ctacgttcgg tataatatgt taagcttttt 9000

aacacaaagg tttggcttgg ggtaacctga tgtggtgcaa aagaccgggc gttggcgagc 9060

cattgcgcgg gcgaatgggg tcgtgactcg tctcaaattc gagggcgtgc ctcaattcgt 9120

gcccccgtgg ctttttcccg ccgtttccgc cccgtttgca ccactgcagc cgcttctttg 9180

gttcggacac cttgctgcga gctaggtgcc ttgtgctact taaaaagtgg cctcccaaca 9240

ccaacatgac atgagtgcgt gggccaagac acgttggcgg ggtcgcagtc ggctcaatgg 9300

cccggaaaaa acgctgctgg agctggttcg gacgcagtcc gccgcggcgt atggatatcc 9360

gcaaggttcc atagcgccat tgccctccgt cggcgtctat cccgcaacct taattaagtc 9420

atacacaagt cagctttctt cgagcctcat ataagtataa gtagttcaac gtattagcac 9480

tgtacccagc atctccgtat cgagaaacac aacaacatgc cccattggac agatcatgcg 9540

gatacacagg ttgtgcagta tcatacatac tcgatcagac aggtcgtctg accatcatac 9600

aagctgaaca agcgctccat acttgcacgc tctctatata cacagttaaa ttacatatcc 9660

atagtctaac ctctaacagt taatcttctg gtaagcctcc cagccagcct tctggtatcg 9720

cttggcctcc tcaataggat ctcggttctg gccgtacaga cctcggccga caattatgat 9780

atccgttccg gtagacatga catcctcaac agttcggtac tgctgtccga gagcgtctcc 9840

cttgtcgtca agacccaccc cgggggtcag aataagccag tcctcagagt cgcccttagg 9900

tcggttctgg gcaatgaagc caaccacaaa ctcggggtcg gatcgggcaa gctcaatggt 9960

ctgcttggag tactcgccag tggccagaga gcccttgcaa gacagctcgg ccagcatgag 10020

cagacctctg gccagcttct cgttgggaga ggggactagg aactccttgt actgggagtt 10080

ctcgtagtca gagacgtcct ccttcttctg ttcagagaca gtttcctcgg caccagctcg 10140

caggccagca atgattccgg ttccgggtac accgtgggcg ttggtgatat cggaccactc 10200

ggcgattcgg tgacaccggt actggtgctt gacagtgttg ccaatatctg cgaactttct 10260

gtcctcgaac aggaagaaac cgtgcttaag agcaagttcc ttgaggggga gcacagtgcc 10320

ggcgtaggtg aagtcgtcaa tgatgtcgat atgggttttg atcatgcaca cataaggtcc 10380

gaccttatcg gcaagctcaa tgagctcctt ggtggtggta acatccagag aagcacacag 10440

gttggttttc ttggctgcca cgagcttgag cactcgagcg gcaaaggcgg acttgtggac 10500

gttagctcga gcttcgtagg agggcatttt ggtggtgaag aggagactga aataaattta 10560

gtctgcagaa ctttttatcg gaaccttatc tggggcagtg aagtatatgt tatggtaata 10620

gttacgagtt agttgaactt atagatagac tggactatac ggctatcggt ccaaattaga 10680

aagaacgtca atggctctct gggcgtcgcc tttgccgaca aaaatgtgat catgatgaaa 10740

gccagcaatg acgttgcagc tgatattgtt gtcggccaac cgcgccgaaa acgcagctgt 10800

cagacccaca gcctccaacg aagaatgtat cgtcaaagtg atccaagcac actcatagtt 10860

ggagtcgtac tccaaaggcg gcaatgacga gtcagacaga tactcgtcga cgtttaaacc 10920

atcatctaag ggcctcaaaa ctacctcgga actgctgcgc tgatctggac accacagagg 10980

ttccgagcac tttaggttgc accaaatgtc ccaccaggtg caggcagaaa acgctggaac 11040

agcgtgtaca gtttgtctta acaaaaagtg agggcgctga ggtcgagcag ggtggtgtga 11100

cttgttatag cctttagagc tgcgaaagcg cgtatggatt tggctcatca ggccagattg 11160

agggtctgtg gacacatgtc atgttagtgt acttcaatcg ccccctggat atagccccga 11220

caataggccg tggcctcatt tttttgcctt ccgcacattt ccattgctcg gtacccacac 11280

cttgcttctc ctgcacttgc caaccttaat actggtttac attgaccaac atcttacaag 11340

cgggggggctt gtctagggta tatataaaca gtggctctcc caatcggttg ccagtctctt 11400

ttttcctttc tttccccaca gattcgaaat ctaaactaca catcacac 11448

<210> 154

<211> 11448

<212> DNA

<213> Artificial Sequence

<220><223> pRF341

<400> 154

catggacaag aaatactcca tcggcctggc cattggaacc aactctgtcg gctgggctgt 60

catcaccgac gagtacaagg tgccctccaa gaaattcaag gtcctcggaa acaccgatcg 120

acactccatc aagaaaaacc tcattggtgc cctgttgttc gattctggcg agactgccga 180

agctaccaga ctcaagcgaa ctgctcggcg acgttacacc cgacggaaga accgaatctg 240

ctacctgcag gagatctttt ccaacgagat ggccaaggtg gacgattcgt tctttcatcg 300

actggaggaa tccttcctcg tcgaggaaga caagaaacac gagcgtcatc ccatctttgg 360

caacattgtg gacgaggttg cttaccacga gaagtatcct accatctacc acctgcgaaa 420

gaaactcgtc gattccaccg acaaggcgga tctcagactt atctacctcg ctctggcaca 480

catgatcaag tttcgaggtc atttcctcat cgagggcgat ctcaatcccg acaacagcga 540

tgtggacaag ctgttcattc agctcgttca gacctacaac cagctgttcg aggaaaaccc 600

catcaatgcc tccggagtcg atgcaaaggc catcttgtct gctcgactct cgaagagcag 660

acgactggag aacctcattg cccaacttcc tggcgagaaa aagaacggac tgtttggcaa 720

cctcattgcc ctttctcttg gtctcacacc caacttcaag tccaacttcg atctggcgga 780

ggacgccaag ctccagctgt ccaaggacac ctacgacgat gacctcgaca acctgcttgc 840

acagattggc gatcagtacg ccgacctgtt tctcgctgcc aagaaccttt cggatgctat 900

tctcttgtct gacattctgc gagtcaacac cgagatcaca aaggctcccc tttctgcctc 960

catgatcaag cgatacgacg agcaccatca ggatctcaca ctgctcaagg ctcttgtccg 1020

acagcaactg cccgagaagt acaaggagat ctttttcgat cagtcgaaga acggctacgc 1080

tggatacatc gacggcggag cctctcagga agagttctac aagttcatca agccaattct 1140

cgagaagatg gacggaaccg aggaactgct tgtcaagctc aatcgagagg atctgcttcg 1200

gaagcaacga accttcgaca acggcagcat tcctcatcag atccacctcg gtgagctgca 1260

cgccattctt cgacgtcagg aagacttcta ccccttttctc aaggacaacc gagagaagat 1320

cgagaagatt cttacctttc gaatccccta ctatgttggt cctcttgcca gaggaaactc 1380

tcgatttgct tggatgactc gaaagtccga ggaaaccatc actccctgga acttcgagga 1440

agtcgtggac aagggtgcct ctgcacagtc cttcatcgag cgaatgacca acttcgacaa 1500

gaatctgccc aacgagaagg ttcttcccaa gcattcgctg ctctacgagt actttacagt 1560

ctacaacgaa ctcaccaaag tcaagtacgt taccgaggga atgcgaaagc ctgccttctt 1620

gtctggcgaa cagaagaaag ccattgtcga tctcctgttc aagaccaacc gaaaggtcac 1680

tgttaagcag ctcaaggagg actacttcaa gaaaatcgag tgtttcgaca gcgtcgagat 1740

ttccggagtt gaggaccgat tcaacgcctc tttgggcacc tatcacgatc tgctcaagat 1800

tatcaaggac aaggattttc tcgacaacga ggaaaacgag gacattctgg aggacatcgt 1860

gctcactctt accctgttcg aagatcggga gatgatcgag gaacgactca agacatacgc 1920

tcacctgttc gacgacaagg tcatgaaaca actcaagcga cgtagataca ccggctgggg 1980

aagactttcg cgaaagctca tcaacggcat cagagacaag cagtccggaa agaccattct 2040

ggactttctc aagtccgatg gctttgccaa ccgaaacttc atgcagctca ttcacgacga 2100

ttctcttacc ttcaaggagg acatccagaa ggcacaagtg tccggtcagg gcgacagctt 2160

gcacgaacat attgccaacc tggctggttc gccagccatc aagaaaggca ttctccagac 2220

tgtcaaggtt gtcgacgagc tggtgaaggt catgggacgt cacaagcccg agaacattgt 2280

gatcgagatg gccagagaga accagacaac tcaaaagggt cagaaaaact cgcgagagcg 2340

gatgaagcga atcgaggaag gcatcaagga gctgggatcc cagattctca aggagcatcc 2400

cgtcgagaac actcaactgc agaacgagaa gctgtatctc tactatctgc agaatggtcg 2460

agacatgtac gtggatcagg aactggacat caatcgtctc agcgactacg atgtggacgc 2520

cattgtccct caatcctttc tcaaggacga ttctatcgac aacaaggtcc ttacacgatc 2580

cgacaagaac agaggcaagt cggacaacgt tcccagcgaa gaggtggtca aaaagatgaa 2640

gaactactgg cgacagctgc tcaacgccaa gctcattacc cagcgaaagt tcgacaatct 2700

taccaaggcc gagcgaggcg gtctgtccga gctcgacaag gctggcttca tcaagcgtca 2760

actcgtcgag accagacaga tcacaaagca cgtcgcacag attctcgatt ctcggatgaa 2820

caccaagtac gacgagaacg acaagctcat ccgagaggtc aaggtgatta ctctcaagtc 2880

caaactggtc tccgatttcc gaaaggactt tcagttctac aaggtgcgag agatcaacaa 2940

ttaccaccat gcccacgatg cttacctcaa cgccgtcgtt ggcactgcgc tcatcaagaa 3000

ataccccaag ctcgaaagcg agttcgttta cggcgattac aaggtctacg acgttcgaaa 3060

gatgattgcc aagtccgaac aggagattgg caaggctact gccaagtact tcttttactc 3120

caacatcatg aacttttca agaccgagat caccttggcc aacggagaga ttcgaaagag 3180

accacttatc gagaccaacg gcgaaactgg agagatcgtg tgggacaagg gtcgagactt 3240

tgcaaccgtg cgaaaggttc tgtcgatgcc tcaggtcaac atcgtcaaga aaaccgaggt 3300

tcagactggc ggattctcca aggagtcgat tctgcccaag cgaaactccg acaagctcat 3360

cgctcgaaag aaagactggg atcccaagaa atacggtggc ttcgattctc ctaccgtcgc 3420

ctattccgtg cttgtcgttg cgaaggtcga gaagggcaag tccaaaaagc tcaagtccgt 3480

caaggagctg ctcggaatta ccatcatgga gcgatcgagc ttcgagaaga atcccatcga 3540

cttcttggaa gccaagggtt acaaggaggt caagaaagac ctcattatca agctgcccaa 3600

gtactctctg ttcgaactgg agaacggtcg aaagcgtatg ctcgcctccg ctggcgagct 3660

gcagaaggga aacgagcttg ccttgccttc gaagtacgtc aactttctct atctggcttc 3720

tcactacgag aagctcaagg gttctcccga ggacaacgaa cagaagcaac tcttcgttga 3780

gcagcacaaa cattacctcg acgagattat cgagcagatt tccgagtttt cgaagcgagt 3840

catcctggct gatgccaact tggacaaggt gctctctgcc tacaacaagc atcgggacaa 3900

acccattcga gaacaggcgg agaacatcat tcacctgttt actcttacca acctgggtgc 3960

tcctgcagct ttcaagtact tcgataccac tatcgaccga aagcggtaca catccaccaa 4020

ggaggttctc gatgccaccc tgattcacca gtccatcact ggcctgtacg agacccgaat 4080

cgacctgtct cagcttggtg gcgactccag agccgatccc aagaaaaagc gaaaggtcta 4140

agcggccgca agtgtggatg gggaagtgag tgcccggttc tgtgtgcaca attggcaatc 4200

caagatggat ggattcaaca cagggatata gcgagctacg tggtggtgcg aggatatagc 4260

aacggatatt tatgtttgac acttgagaat gtacgataca agcactgtcc aagtacaata 4320

ctaaacatac tgtacatact catactcgta cccgggcaac ggtttcactt gagtgcagtg 4380

gctagtgctc ttactcgtac agtgtgcaat actgcgtatc atagtctttg atgtatatcg 4440

tattcattca tgttagttgc gtacgagccg gaagcataaa gtgtaaagcc tggggtgcct 4500
aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa 4560
acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta 4620
ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc 4680
gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg 4740
caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt 4800
tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa 4860

gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct 4920
ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc 4980
cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg 5040
tcgttcgctc caagctgggc tgtgtgcacg aaccccccgt tcagcccgac cgctgcgcct 5100
tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag 5160
cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga 5220
agtggtggcc taactacggc tacactagaa ggacagtatt tggtatctgc gctctgctga 5280

agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg 5340
gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag 5400
aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag 5460
ggattttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat 5520
gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct 5580
taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac 5640
tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa 5700

tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg 5760
gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt 5820
gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca 5880
ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt 5940
cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct 6000
tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg 6060
cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg 6120

agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg 6180

cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa 6240

aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt 6300

aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt 6360

gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt 6420

gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca 6480

tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat 6540

ttccccgaaa agtgccacct gacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg 6600

tggttacgcg cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt 6660

tcttcccttc ctttctcgcc acgttcgccg gctttccccg tcaagctcta aatcgggggc 6720

tcccttttagg gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg 6780

gtgatggttc acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg 6840

agtccacgtt ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct 6900

cggtctattc ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg 6960

agctgattta acaaaaattt aacgcgaatt ttaacaaaat attaacgctt acaatttcca 7020

ttcgccattc aggctgcgca actgttggga agggcgatcg gtgcgggcct cttcgctatt 7080

acgccagctg gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa cgccagggtt 7140

ttcccagtca cgacgttgta aaacgacggc cagtgaattg taatacgact cactataggg 7200

cgaattgggt accgggcccc ccctcgaggt cgatggtgtc gataagcttg atatcgaatt 7260

catgtcacac aaaccgatct tcgcctcaag gaaacctaat tctacatccg agagactgcc 7320

gagatccagt ctacactgat taattttcgg gccaataatt taaaaaaatc gtgttatata 7380

atattatatg tattatatat atacatcatg atgatactga cagtcatgtc ccattgctaa 7440

atagacagac tccatctgcc gcctccaact gatgttctca atatttaagg ggtcatctcg 7500

cattgtttaa taataaacag actccatcta ccgcctccaa atgatgttct caaaatatat 7560

tgtatgaact tatttttatt acttagtatt attagacaac ttacttgctt tatgaaaaac 7620

acttcctatt taggaaacaa tttataatgg cagttcgttc atttaacaat ttatgtagaa 7680

taaatgttat aaatgcgtat gggaaatctt aaatatggat agcataaatg atatctgcat 7740

tgcctaattc gaaatcaaca gcaacgaaaa aaatcccttg tacaacataa atagtcatcg 7800

agaaatatca actatcaaag aacagctatt cacacgttac tattgagatt attattggac 7860

gagaatcaca cactcaactg tctttctctc ttctagaaat acaggtacaa gtatgtacta 7920

ttctcattgt tcatacttct agtcatttca tcccacatat tccttggatt tctctccaat 7980

gaatgacatt ctatcttgca aattcaacaa ttataataag atataccaaa gtagcggtat 8040

agtggcaatc aaaaagcttc tctggtgtgc ttctcgtatt tatttttatt ctaatgatcc 8100

attaaaggta tatatttatt tcttgttata taatcctttt gtttattaca tgggctggat 8160

acataaaggt attttgattt aatttttgc ttaaattcaa tccccctcg ttcagtgtca 8220

actgtaatgg taggaaatta ccatactttt gaagaagcaa aaaaaatgaa agaaaaaaaa 8280

aatcgtattt ccaggttaga cgttccgcag aatctagaat gcggtatgcg gtacattgtt 8340

cttcgaacgt aaaagttgcg ctccctgaga tattgtacat ttttgctttt acaagtacaa 8400

gtacatcgta caactatgta ctactgttga tgcatccaca acagtttgtt ttgttttttt 8460

ttgttttttt tttttctaat gattcattac cgctatgtat acctacttgt acttgtagta 8520

agccgggtta ttggcgttca attaatcata gacttatgaa tctgcacggt gtgcgctgcg 8580

agttactttt agcttatgca tgctacttgg gtgtaatatt gggatctgtt cggaaatcaa 8640

cggatgctca atcgatcgcc agcaacgcgg ccttagacat aaaaaacaaa aaaaaaaagc 8700

accgactcgg tgccactttt tcaagttgat aacggactag ccttatttta acttgctatt 8760

tctagctcta aaacgagtaa cggtaacgac accggacgag cttactcgtt tcgtcctcac 8820

ggactcatca gcggtgttac gcatggttga tgtgtgttta attcaagaat gaatatagag 8880

aagagaagaa gaaaaaagat tcaattgagc cggcgatgca gacccttata taaatgttgc 8940

cttggacaga cggagcaagc ccgcccaaac ctacgttcgg tataatatgt taagcttttt 9000

aacacaaagg tttggcttgg ggtaacctga tgtggtgcaa aagaccgggc gttggcgagc 9060

cattgcgcgg gcgaatgggg tcgtgactcg tctcaaattc gagggcgtgc ctcaattcgt 9120

gcccccgtgg ctttttcccg ccgtttccgc cccgtttgca ccactgcagc cgcttctttg 9180

gttcggacac cttgctgcga gctaggtgcc ttgtgctact taaaaagtgg cctcccaaca 9240

ccaacatgac atgagtgcgt gggccaagac acgttggcgg ggtcgcagtc ggctcaatgg 9300

cccggaaaaa acgctgctgg agctggttcg gacgcagtcc gccgcggcgt atggatatcc 9360

gcaaggttcc atagcgccat tgccctccgt cggcgtctat cccgcaacct taattaagtc 9420

atacacaagt cagctttctt cgagcctcat ataagtataa gtagttcaac gtattagcac 9480

tgtacccagc atctccgtat cgagaaacac aacaacatgc cccattggac agatcatgcg 9540

gatacacagg ttgtgcagta tcatacatac tcgatcagac aggtcgtctg accatcatac 9600

aagctgaaca agcgctccat acttgcacgc tctctatata cacagttaaa ttacatatcc 9660

atagtctaac ctctaacagt taatcttctg gtaagcctcc cagccagcct tctggtatcg 9720

cttggcctcc tcaataggat ctcggttctg gccgtacaga cctcggccga caattatgat 9780

atccgttccg gtagacatga catcctcaac agttcggtac tgctgtccga gagcgtctcc 9840

cttgtcgtca agacccaccc cgggggtcag aataagccag tcctcagagt cgcccttagg 9900

tcggttctgg gcaatgaagc caaccacaaa ctcggggtcg gatcgggcaa gctcaatggt 9960
ctgcttggag tactcgccag tggccagaga gcccttgcaa gacagctcgg ccagcatgag 10020
cagacctctg gccagcttct cgttgggaga ggggactagg aactccttgt actgggagtt 10080
ctcgtagtca gagacgtcct ccttcttctg ttcagagaca gtttcctcgg caccagctcg 10140
caggccagca atgattccgg ttccgggtac accgtgggcg ttggtgatat cggaccactc 10200
ggcgattcgg tgacaccggt actggtgctt gacagtgttg ccaatatctg cgaactttct 10260
gtcctcgaac aggaagaaac cgtgcttaag agcaagttcc ttgaggggga gcacagtgcc 10320

ggcgtaggtg aagtcgtcaa tgatgtcgat atgggttttg atcatgcaca cataaggtcc 10380
gaccttatcg gcaagctcaa tgagctcctt ggtggtggta acatccagag aagcacacag 10440
gttggttttc ttggctgcca cgagcttgag cactcgagcg gcaaaggcgg acttgtggac 10500
gttagctcga gcttcgtagg agggcatttt ggtggtgaag aggagactga aataaattta 10560
gtctgcagaa ctttttatcg gaaccttatc tggggcagtg aagtatatgt tatggtaata 10620
gttacgagtt agttgaactt atagatagac tggactatac ggctatcggt ccaaattaga 10680
aagaacgtca atggctctct gggcgtcgcc tttgccgaca aaaatgtgat catgatgaaa 10740

gccagcaatg acgttgcagc tgatattgtt gtcggccaac cgcgccgaaa acgcagctgt 10800
cagacccaca gcctccaacg aagaatgtat cgtcaaagtg atccaagcac actcatagtt 10860
ggagtcgtac tccaaaggcg gcaatgacga gtcagacaga tactcgtcga cgtttaaacc 10920
atcatctaag ggcctcaaaa ctacctcgga actgctgcgc tgatctggac accacagagg 10980
ttccgagcac tttaggttgc accaaatgtc ccaccaggtg caggcagaaa acgctggaac 11040
agcgtgtaca gtttgtctta acaaaaagtg agggcgctga ggtcgagcag ggtggtgtga 11100
cttgttatag cctttagagc tgcgaaagcg cgtatggatt tggctcatca ggccagattg 11160

agggtctgtg gacacatgtc atgttagtgt acttcaatcg ccccctggat atagccccga 11220
caataggccg tggcctcatt tttttgcctt ccgcacattt ccattgctcg gtacccacac 11280
cttgcttctc ctgcacttgc caaccttaat actggtttac attgaccaac atcttacaag 11340
cgggggggctt gtctagggta tatataaaca gtggctctcc caatcggttg ccagtctctt 11400
ttttcctttc tttccccaca gattcgaaat ctaaactaca catcacac 11448

<210> 155
<211> 11448
<212> DNA
<213> Artificial sequence

<220><223> pRF343

<400> 155

catggacaag aaatactcca tcggcctggc cattggaacc aactctgtcg gctgggctgt 60

catcaccgac gagtacaagg tgccctccaa gaaattcaag gtcctcggaa acaccgatcg 120

acactccatc aagaaaaacc tcattggtgc cctgttgttc gattctggcg agactgccga 180

agctaccaga ctcaagcgaa ctgctcggcg acgttacacc cgacggaaga accgaatctg 240

ctacctgcag gagatctttt ccaacgagat ggccaaggtg gacgattcgt tctttcatcg 300

actggaggaa tccttcctcg tcgaggaaga caagaaacac gagcgtcatc ccatctttgg 360

caacattgtg gacgaggttg cttaccacga gaagtatcct accatctacc acctgcgaaa 420

gaaactcgtc gattccaccg acaaggcgga tctcagactt atctacctcg ctctggcaca 480

catgatcaag tttcgaggtc atttcctcat cgagggcgat ctcaatcccg acaacagcga 540

tgtggacaag ctgttcattc agctcgttca gacctacaac cagctgttcg aggaaaaccc 600

catcaatgcc tccggagtcg atgcaaaggc catcttgtct gctcgactct cgaagagcag 660

acgactggag aacctcattg cccaacttcc tggcgagaaa aagaacggac tgtttggcaa 720

cctcattgcc ctttctcttg gtctcacacc caacttcaag tccaacttcg atctggcgga 780

ggacgccaag ctccagctgt ccaaggacac ctacgacgat gacctcgaca acctgcttgc 840

acagattggc gatcagtacg ccgacctgtt tctcgctgcc aagaaccttt cggatgctat 900

tctcttgtct gacattctgc gagtcaacac cgagatcaca aaggctcccc tttctgcctc 960

catgatcaag cgatacgacg agcaccatca ggatctcaca ctgctcaagg ctcttgtccg 1020

acagcaactg cccgagaagt acaaggagat ctttttcgat cagtcgaaga acggctacgc 1080

tggatacatc gacggcggag cctctcagga agagttctac aagttcatca agccaattct 1140

cgagaagatg gacggaaccg aggaactgct tgtcaagctc aatcgagagg atctgcttcg 1200

gaagcaacga accttcgaca acggcagcat tcctcatcag atccacctcg gtgagctgca 1260

cgccattctt cgacgtcagg aagacttcta ccccttttctc aaggacaacc gagagaagat 1320

cgagaagatt cttacctttc gaatccccta ctatgttggt cctcttgcca gaggaaactc 1380

tcgatttgct tggatgactc gaaagtccga ggaaaccatc actccctgga acttcgagga 1440

agtcgtggac aagggtgcct ctgcacagtc cttcatcgag cgaatgacca acttcgacaa 1500

gaatctgccc aacgagaagg ttcttcccaa gcattcgctg ctctacgagt actttacagt 1560

ctacaacgaa ctcaccaaag tcaagtacgt taccgaggga atgcgaaagc ctgccttctt 1620

gtctggcgaa cagaagaaag ccattgtcga tctcctgttc aagaccaacc gaaaggtcac 1680

tgttaagcag ctcaaggagg actacttcaa gaaaatcgag tgtttcgaca gcgtcgagat 1740

ttccggagtt gaggaccgat tcaacgcctc tttgggcacc tatcacgatc tgctcaagat 1800
tatcaaggac aaggattttc tcgacaacga ggaaaacgag gacattctgg aggacatcgt 1860
gctcactctt accctgttcg aagatcggga gatgatcgag gaacgactca agacatacgc 1920
tcacctgttc gacgacaagg tcatgaaaca actcaagcga cgtagataca ccggctgggg 1980
aagactttcg cgaaagctca tcaacggcat cagagacaag cagtccggaa agaccattct 2040
ggactttctc aagtccgatg gctttgccaa ccgaaacttc atgcagctca ttcacgacga 2100
ttctcttacc ttcaaggagg acatccagaa ggcacaagtg tccggtcagg gcgacagctt 2160

gcacgaacat attgccaacc tggctggttc gccagccatc aagaaaggca ttctccagac 2220
tgtcaaggtt gtcgacgagc tggtgaaggt catgggacgt cacaagcccg agaacattgt 2280
gatcgagatg gccagagaga accagacaac tcaaaagggt cagaaaaact cgcgagagcg 2340
gatgaagcga atcgaggaag gcatcaagga gctgggatcc cagattctca aggagcatcc 2400
cgtcgagaac actcaactgc agaacgagaa gctgtatctc tactatctgc agaatggtcg 2460
agacatgtac gtggatcagg aactggacat caatcgtctc agcgactacg atgtggacgc 2520
cattgtccct caatcctttc tcaaggacga ttctatcgac aacaaggtcc ttacacgatc 2580

cgacaagaac agaggcaagt cggacaacgt tcccagcgaa gaggtggtca aaaagatgaa 2640
gaactactgg cgacagctgc tcaacgccaa gctcattacc cagcgaaagt tcgacaatct 2700
taccaaggcc gagcgaggcg gtctgtccga gctcgacaag gctggcttca tcaagcgtca 2760
actcgtcgag accagacaga tcacaaagca cgtcgcacag attctcgatt ctcggatgaa 2820
caccaagtac gacgagaacg acaagctcat ccgagaggtc aaggtgatta ctctcaagtc 2880
caaactggtc tccgatttcc gaaaggactt tcagttctac aaggtgcgag agatcaacaa 2940
ttaccaccat gcccacgatg cttacctcaa cgccgtcgtt ggcactgcgc tcatcaagaa 3000

ataccccaag ctcgaaagcg agttcgttta cggcgattac aaggtctacg acgttcgaaa 3060
gatgattgcc aagtccgaac aggagattgg caaggctact gccaagtact tcttttactc 3120
caacatcatg aacttttca agaccgagat caccttggcc aacggagaga ttcgaaagag 3180
accacttatc gagaccaacg gcgaaactgg agagatcgtg tgggacaagg gtcgagactt 3240
tgcaaccgtg cgaaaggttc tgtcgatgcc tcaggtcaac atcgtcaaga aaaccgaggt 3300
tcagactggc ggattctcca aggagtcgat tctgcccaag cgaaactccg acaagctcat 3360
cgctcgaaag aaagactggg atcccaagaa atacggtggc ttcgattctc ctaccgtcgc 3420

ctattccgtg cttgtcgttg cgaaggtcga gaagggcaag tccaaaaagc tcaagtccgt 3480

caaggagctg ctcggaatta ccatcatgga gcgatcgagc ttcgagaaga atcccatcga 3540

cttcttggaa gccaagggtt acaaggaggt caagaaagac ctcattatca agctgcccaa 3600

gtactctctg ttcgaactgg agaacggtcg aaagcgtatg ctcgcctccg ctggcgagct 3660

gcagaaggga aacgagcttg ccttgccttc gaagtacgtc aactttctct atctggcttc 3720

tcactacgag aagctcaagg gttctcccga ggacaacgaa cagaagcaac tcttcgttga 3780

gcagcacaaa cattacctcg acgagattat cgagcagatt tccgagtttt cgaagcgagt 3840

catcctggct gatgccaact tggacaaggt gctctctgcc tacaacaagc atcgggacaa 3900

acccattcga gaacaggcgg agaacatcat tcacctgttt actcttacca acctgggtgc 3960

tcctgcagct ttcaagtact tcgataccac tatcgaccga aagcggtaca catccaccaa 4020

ggaggttctc gatgccaccc tgattcacca gtccatcact ggcctgtacg agacccgaat 4080

cgacctgtct cagcttggtg gcgactccag agccgatccc aagaaaaagc gaaaggtcta 4140

agcggccgca agtgtggatg gggaagtgag tgcccggttc tgtgtgcaca attggcaatc 4200

caagatggat ggattcaaca cagggatata gcgagctacg tggtggtgcg aggatatagc 4260

aacggatatt tatgtttgac acttgagaat gtacgataca agcactgtcc aagtacaata 4320

ctaaacatac tgtacatact catactcgta cccgggcaac ggtttcactt gagtgcagtg 4380

gctagtgctc ttactcgtac agtgtgcaat actgcgtatc atagtctttg atgtatatcg 4440

tattcattca tgttagttgc gtacgagccg gaagcataaa gtgtaaagcc tggggtgcct 4500

aatgagtgag ctaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa 4560

acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta 4620

ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc 4680

gagcggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg 4740

caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt 4800

tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa 4860

gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct 4920

ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc 4980

cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg 5040

tcgttcgctc caagctgggc tgtgtgcacg aaccccccgt tcagcccgac cgctgcgcct 5100

tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccactggcag 5160

cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga 5220

agtggtggcc taactacggc tacactagaa ggacagtatt tggtatctgc gctctgctga 5280

agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaacaa accaccgctg 5340

gtagcggtgg ttttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag 5400

aagatccttt gatcttttct acggggtctg acgctcagtg gaacgaaaac tcacgttaag 5460

ggattttggt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat 5520

gaagttttaa atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct 5580

taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcatccata gttgcctgac 5640

tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa 5700

tgataccgcg agacccacgc tcaccggctc cagatttatc agcaataaac cagccagccg 5760

gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt 5820

gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca 5880

ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt 5940

cccaacgatc aaggcgagtt acatgatccc ccatgttgtg caaaaaagcg gttagctcct 6000

tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc atggttatgg 6060

cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg 6120

agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg 6180

cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa 6240

aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt 6300

aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt 6360

gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt 6420

gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca 6480

tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat 6540

ttccccgaaa agtgccacct gacgcgccct gtagcggcgc attaagcgcg gcgggtgtgg 6600

tggttacgcg cagcgtgacc gctacacttg ccagcgccct agcgcccgct cctttcgctt 6660

tcttcccttc ctttctcgcc acgttcgccg gctttccccg tcaagctcta aatcgggggc 6720

tccctttagg gttccgattt agtgctttac ggcacctcga ccccaaaaaa cttgattagg 6780

gtgatggttc acgtagtggg ccatcgccct gatagacggt ttttcgccct ttgacgttgg 6840

agtccacgtt ctttaatagt ggactcttgt tccaaactgg aacaacactc aaccctatct 6900

cggtctattc ttttgattta taagggattt tgccgatttc ggcctattgg ttaaaaaatg 6960

agctgattta acaaaaattt aacgcgaatt ttaacaaaat attaacgctt acaatttcca 7020

ttcgccattc aggctgcgca actgttggga agggcgatcg gtgcgggcct cttcgctatt 7080

acgccagctg gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa cgccagggtt 7140

ttcccagtca cgacgttgta aaacgacggc cagtgaattg taatacgact cactataggg 7200

cgaattgggt accgggcccc ccctcgaggt cgatggtgtc gataagcttg atatcgaatt 7260
catgtcacac aaaccgatct tcgcctcaag gaaacctaat tctacatccg agagactgcc 7320
gagatccagt ctacactgat taattttcgg gccaataatt taaaaaaatc gtgttatata 7380
atattatatg tattatatat atacatcatg atgatactga cagtcatgtc ccattgctaa 7440
atagacagac tccatctgcc gcctccaact gatgttctca atatttaagg ggtcatctcg 7500
cattgtttaa taataaacag actccatcta ccgcctccaa atgatgttct caaaatatat 7560
tgtatgaact tatttttatt acttagtatt attagacaac ttacttgctt tatgaaaaac 7620

acttcctatt taggaaacaa tttataatgg cagttcgttc atttaacaat ttatgtagaa 7680
taaatgttat aaatgcgtat gggaaatctt aaatatggat agcataaatg atatctgcat 7740
tgcctaattc gaaatcaaca gcaacgaaaa aaatcccttg tacaacataa atagtcatcg 7800
agaaatatca actatcaaag aacagctatt cacacgttac tattgagatt attattggac 7860
gagaatcaca cactcaactg tctttctctc ttctagaaat acaggtacaa gtatgtacta 7920
ttctcattgt tcatacttct agtcatttca tcccacatat tccttggatt tctctccaat 7980
gaatgacatt ctatcttgca aattcaacaa ttataataag atataccaaa gtagcggtat 8040

agtggcaatc aaaaagcttc tctggtgtgc ttctcgtatt tatttttatt ctaatgatcc 8100
attaaaggta tatatttatt tcttgttata taatcctttt gtttattaca tgggctggat 8160
acataaaggt attttgattt aatttttgc ttaaattcaa tccccctcg ttcagtgtca 8220
actgtaatgg taggaaatta ccatactttt gaagaagcaa aaaaaatgaa agaaaaaaaa 8280
aatcgtattt ccaggttaga cgttccgcag aatctagaat gcggtatgcg gtacattgtt 8340
cttcgaacgt aaaagttgcg ctccctgaga tattgtacat ttttgctttt acaagtacaa 8400
gtacatcgta caactatgta ctactgttga tgcatccaca acagtttgtt ttgttttttt 8460

ttgttttttt tttttctaat gattcattac cgctatgtat acctacttgt acttgtagta 8520
agccgggtta ttggcgttca attaatcata gacttatgaa tctgcacggt gtgcgctgcg 8580
agttactttt agcttatgca tgctacttgg gtgtaatatt gggatctgtt cggaaatcaa 8640
cggatgctca atcgatcgcc agcaacgcgg ccttagacat aaaaaacaaa aaaaaaaagc 8700
accgactcgg tgccactttt tcaagttgat aacggactag ccttatttta acttgctatt 8760
tctagctcta aaaccgaagc tgaaggtgac caaggacgag cttactcgtt tcgtcctcac 8820
ggactcatca gcctggttac gcatggttga tgtgtgttta attcaagaat gaatatagag 8880

aagagaagaa gaaaaaagat tcaattgagc cggcgatgca gacccttata taaatgttgc 8940

cttggacaga cggagcaagc ccgcccaaac ctacgttcgg tataatatgt taagcttttt 9000

aacacaaagg tttggcttgg ggtaacctga tgtggtgcaa aagaccgggc gttggcgagc 9060

cattgcgcgg gcgaatgggg tcgtgactcg tctcaaattc gagggcgtgc ctcaattcgt 9120

gcccccgtgg ctttttcccg ccgtttccgc cccgtttgca ccactgcagc cgcttctttg 9180

gttcggacac cttgctgcga gctaggtgcc ttgtgctact taaaaagtgg cctcccaaca 9240

ccaacatgac atgagtgcgt gggccaagac acgttggcgg ggtcgcagtc ggctcaatgg 9300

cccggaaaaa acgctgctgg agctggttcg gacgcagtcc gccgcggcgt atggatatcc 9360

gcaaggttcc atagcgccat tgccctccgt cggcgtctat cccgcaacct taattaagtc 9420

atacacaagt cagctttctt cgagcctcat ataagtataa gtagttcaac gtattagcac 9480

tgtacccagc atctccgtat cgagaaacac aacaacatgc cccattggac agatcatgcg 9540

gatacacagg ttgtgcagta tcatacatac tcgatcagac aggtcgtctg accatcatac 9600

aagctgaaca agcgctccat acttgcacgc tctctatata cacagttaaa ttacatatcc 9660

atagtctaac ctctaacagt taatcttctg gtaagcctcc cagccagcct tctggtatcg 9720

cttggcctcc tcaataggat ctcggttctg gccgtacaga cctcggccga caattatgat 9780

atccgttccg gtagacatga catcctcaac agttcggtac tgctgtccga gagcgtctcc 9840

cttgtcgtca agacccaccc cgggggtcag aataagccag tcctcagagt cgcccttagg 9900

tcggttctgg gcaatgaagc caaccacaaa ctcggggtcg gatcgggcaa gctcaatggt 9960

ctgcttggag tactcgccag tggccagaga gcccttgcaa gacagctcgg ccagcatgag 10020

cagacctctg gccagcttct cgttgggaga ggggactagg aactccttgt actgggagtt 10080

ctcgtagtca gagacgtcct ccttcttctg ttcagagaca gtttcctcgg caccagctcg 10140

caggccagca atgattccgg ttccgggtac accgtgggcg ttggtgatat cggaccactc 10200

ggcgattcgg tgacaccggt actggtgctt gacagtgttg ccaatatctg cgaactttct 10260

gtcctcgaac aggaagaaac cgtgcttaag agcaagttcc ttgaggggga gcacagtgcc 10320

ggcgtaggtg aagtcgtcaa tgatgtcgat atgggttttg atcatgcaca cataaggtcc 10380

gaccttatcg gcaagctcaa tgagctcctt ggtggtggta acatccagag aagcacacag 10440

gttggttttc ttggctgcca cgagcttgag cactcgagcg gcaaaggcgg acttgtggac 10500

gttagctcga gcttcgtagg agggcatttt ggtggtgaag aggagactga aataaattta 10560

gtctgcagaa ctttttatcg gaaccttatc tggggcagtg aagtatatgt tatggtaata 10620

gttacgagtt agttgaactt atagatagac tggactatac ggctatcggt ccaaattaga 10680

aagaacgtca atggctctct gggcgtcgcc tttgccgaca aaaatgtgat catgatgaaa 10740

gccagcaatg acgttgcagc tgatattgtt gtcggccaac cgcgccgaaa acgcagctgt 10800

cagacccaca gcctccaacg aagaatgtat cgtcaaagtg atccaagcac actcatagtt 10860

ggagtcgtac tccaaaggcg gcaatgacga gtcagacaga tactcgtcga cgtttaaacc 10920

atcatctaag ggcctcaaaa ctacctcgga actgctgcgc tgatctggac accacagagg 10980

ttccgagcac tttaggttgc accaaatgtc ccaccaggtg caggcagaaa acgctggaac 11040

agcgtgtaca gtttgtctta acaaaaagtg agggcgctga ggtcgagcag ggtggtgtga 11100

cttgttatag cctttagagc tgcgaaagcg cgtatggatt tggctcatca ggccagattg 11160

agggtctgtg gacacatgtc atgttagtgt acttcaatcg ccccctggat atagccccga 11220

caataggccg tggcctcatt tttttgcctt ccgcacattt ccattgctcg gtacccacac 11280

cttgcttctc ctgcacttgc caaccttaat actggtttac attgaccaac atcttacaag 11340

cgggggggctt gtctagggta tatataaaca gtggctctcc caatcggttg ccagtctctt 11400

ttttcctttc tttccccaca gattcgaaat ctaaactaca catcacac 11448

<210> 156

<211> 3901

<212> DNA

<213> Artificial sequence

<220><223> pRF80

<400> 156

agcttgctac gttaggagaa gacgcacggc gatgatacgg gtacccctca tgacatcaat 60

atccgctgcc cctcttgcca gcaaggcgtc agcaggtgct tttttcgcta ttttcaccag 120

accacagcct ttttccttgt gtctcatctt ggattccttc aaaggcaact caccgcacct 180

ccgagtcgtg tgaacaatgt aataataggc tattgacttt tttcccacct gtttagcgcc 240

aaacccaaag cgcttttcgc ccccactgca gcccgatgga aggcacatat ggcaagggaa 300

aagtcttcag gtaatacatg cctgctgcaa ctatatgtac tctgactcat tccctcagac 360

gtgggtcata gacagctgtt ttaaaccggg caaatcaatc tctgtcgcac aggtatttct 420

gcccttcaaa accaggttgc cacatcagat tccatcaaag tttttcagac taacttcaat 480

cttaaacggc atctcacaac aagcgaattg gacggaaaaa aagcgtctat cattaccggc 540

acctatccac actaagacag tactaaagga cgacgctccc cacgaaacga cgtttcgacc 600

ttaacgaccc tgccgtctcc atccatccga ccactcccga cgctctctcc tggagcaaac 660

cactcttacc aagcatatag catatataat aacgtattga atttattaac tgattgaatt 720

gagagtaaag ccagtagcgt tgtacggctg tagcttttta gaaaagtggc agatgagcga 780

tggtggatat gaaagtacct ttacggcatg tagcgacaca agatcgcttc caagaactcg 840

acattcaagc ccagctcgta caagaaaatg aactagccaa tcatatgaac tagcacattg 900

aagtcaccgc atcatctctg ttggaaacga cgcgcatgta ctcgtgcgta gtaaatccgt 960

atctgtacac tcgaaagatt acagtatgta gtagtagcat gactaacgat gtaacgtcca 1020

aataacgctc tgtgcctact cctgtagatg cattagacca cctgctaacg tctacacgtt 1080

atgtccgtta gctccaagat tgcacttttc cctcaaagac tctgctgggt tacgtcatgg 1140

tctctttcgg gtctctggtc cgttctctgc ccgcccatat ccgcccaggc tgctacgata 1200

caggataagc tcataagctt gcatgcctgc aggtcgactc tagaggatcc ccgggtaccg 1260

agctcgaatt cgtaatcatg gtcatagctg tttcctgtgt gaaattgtta tccgctcaca 1320

attccacaca acatacgagc cggaagcata aagtgtaaag cctggggtgc ctaatgagtg 1380

agctaactca cattaattgc gttgcgctca ctgcccgctt tccagtcggg aaacctgtcg 1440

tgccagctgc attaatgaat cggccaacgc gcggggagag gcggtttgcg tattgggcgc 1500

tcttccgctt cctcgctcac tgactcgctg cgctcggtcg ttcggctgcg gcgagcggta 1560

tcagctcact caaaggcggt aatacggtta tccacagaat caggggataa cgcaggaaag 1620

aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc gttgctggcg 1680

tttttccata ggctccgccc ccctgacgag catcacaaaa atcgacgctc aagtcagagg 1740

tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag ctccctcgtg 1800

cgctctcctg ttccgaccct gccgcttacc ggatacctgt ccgcctttct cccttcggga 1860

agcgtggcgc tttctcatag ctcacgctgt aggtatctca gttcggtgta ggtcgttcgc 1920

tccaagctgg gctgtgtgca cgaaccccccc gttcagcccg accgctgcgc cttatccggt 1980

aactatcgtc ttgagtccaa cccggtaaga cacgacttat cgccactggc agcagccact 2040

ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagttctt gaagtggtgg 2100

cctaactacg gctacactag aaggacagta tttggtatct gcgctctgct gaagccagtt 2160

accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc tggtagcggt 2220

ggtttttttg tttgcaagca gcagattacg cgcagaaaaa aaggatctca agaagatcct 2280

ttgatctttt ctacggggtc tgacgctcag tggaacgaaa actcacgtta agggattttg 2340

gtcatgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa atgaagtttt 2400

aaatcaatct aaagtatata tgagtaaact tggtctgaca gttaccaatg cttaatcagt 2460

gaggcaccta tctcagcgat ctgtctattt cgttcatcca tagttgcctg actccccgtc 2520

gtgtagataa ctacgatacg ggagggctta ccatctggcc ccagtgctgc aatgataccg 2580

cgagacccac gctcaccggc tccagattta tcagcaataa accagccagc cggaagggcc 2640

gagcgcagaa gtggtcctgc aactttatcc gcctccatcc agtctattaa ttgttgccgg 2700

gaagctagag taagtagttc gccagttaat agtttgcgca acgttgttgc cattgctaca 2760

ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat tcagctccgg ttcccaacga 2820

tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag cggttagctc cttcggtcct 2880

ccgatcgttg tcagaagtaa gttggccgca gtgttatcac tcatggttat ggcagcactg 2940

cataattctc ttactgtcat gccatccgta agatgctttt ctgtgactgg tgagtactca 3000

accaagtcat tctgagaata gtgtatgcgg cgaccgagtt gctcttgccc ggcgtcaata 3060

cgggataata ccgcgccaca tagcagaact ttaaaagtgc tcatcattgg aaaacgttct 3120

tcggggcgaa aactctcaag gatcttaccg ctgttgagat ccagttcgat gtaacccact 3180

cgtgcaccca actgatcttc agcatctttt actttcacca gcgtttctgg gtgagcaaaa 3240

acaggaaggc aaaatgccgc aaaaaaggga ataagggcga cacggaaatg ttgaatactc 3300

atactcttcc tttttcaata ttattgaagc atttatcagg gttattgtct catgagcgga 3360

tacatatttg aatgtattta gaaaaataaa caaatagggg ttccgcgcac atttccccga 3420

aaagtgccac ctgacgtcta agaaaccatt attatcatga cattaaccta taaaaatagg 3480

cgtatcacga ggccctttcg tctcgcgcgt ttcggtgatg acggtgaaaa cctctgacac 3540

atgcagctcc cggagacggt cacagcttgt ctgtaagcgg atgccgggag cagacaagcc 3600

cgtcagggcg cgtcagcggg tgttggcggg tgtcggggct ggcttaacta tgcggcatca 3660

gagcagattg tactgagagt gcaccatatg cggtgtgaaa taccgcacag atgcgtaagg 3720

agaaaatacc gcatcaggcg ccattcgcca ttcaggctgc gcaactgttg ggaagggcga 3780

tcggtgcggg cctcttcgct attacgccag ctggcgaaag ggggatgtgc tgcaaggcga 3840

ttaagttggg taacgccagg gttttcccag tcacgacgtt gtaaaacgac ggccagtgcc 3900

a 3901

<210> 157

<211> 1215

<212> DNA

<213> Artificial Sequence

<220><223> short Can1 deletion editing template

<400> 157

agcttgctac gttaggagaa gacgcacggc gatgatacgg gtacccctca tgacatcaat 60

atccgctgcc cctcttgcca gcaaggcgtc agcaggtgct tttttcgcta ttttcaccag 120

accacagcct ttttccttgt gtctcatctt ggattccttc aaaggcaact caccgcacct 180

ccgagtcgtg tgaacaatgt aataataggc tattgacttt tttcccacct gtttagcgcc 240

aaacccaaag cgcttttcgc ccccactgca gcccgatgga aggcacatat ggcaagggaa 300

aagtcttcag gtaatacatg cctgctgcaa ctatatgtac tctgactcat tccctcagac 360

gtgggtcata gacagctgtt ttaaaccggg caaatcaatc tctgtcgcac aggtatttct 420

gcccttcaaa accaggttgc cacatcagat tccatcaaag tttttcagac taacttcaat 480

cttaaacggc atctcacaac aagcgaattg gacggaaaaa aagcgtctat cattaccggc 540

acctatccac actaagacag tactaaagga cgacgctccc cacgaaacga cgtttcgacc 600

ttaacgaccc tgccgtctcc atccatccga ccactcccga cgctctctcc tggagcaaac 660

cactcttacc aagcatatag catatataat aacgtattga atttattaac tgattgaatt 720

gagagtaaag ccagtagcgt tgtacggctg tagcttttta gaaaagtggc agatgagcga 780

tggtggatat gaaagtacct ttacggcatg tagcgacaca agatcgcttc caagaactcg 840

acattcaagc ccagctcgta caagaaaatg aactagccaa tcatatgaac tagcacattg 900

aagtcaccgc atcatctctg ttggaaacga cgcgcatgta ctcgtgcgta gtaaatccgt 960

atctgtacac tcgaaagatt acagtatgta gtagtagcat gactaacgat gtaacgtcca 1020

aataacgctc tgtgcctact cctgtagatg cattagacca cctgctaacg tctacacgtt 1080

atgtccgtta gctccaagat tgcacttttc cctcaaagac tctgctgggt tacgtcatgg 1140

tctctttcgg gtctctggtc cgttctctgc ccgcccatat ccgcccaggc tgctacgata 1200

caggataagc tcata 1215

<210> 158

<211> 20

<212> DNA

<213> Artificial sequence

<220><223> primer 80F

<400> 158

agcttgctac gttaggagaa 20

<210> 159

<211> 20

<212> DNA

<213> Artificial sequence

<220><223> primer 80R

<400> 159

tatgagctta tcctgtatcg 20

<210> 160

<211> 2125

<212> DNA

<213> Yarrowia lipolytica

<220><221> misc_feature

<222> (1)..(2125)

<223> Can1 locus WT (wild type)

<400> 160

ggaaggcaca tatggcaagg gaaaagtctt caggtaatac atgcctgctg caactatatg 60

tactctgact cattccctca gacgtgggtc atagacagct gttttaaacc gggcaaatca 120

atctctgtcg cacaggtatt tctgcccttc aaaaccaggt tgccacatca gattccatca 180

aagtttttca gactaacttc aatcttaaac ggcatctcac aacaagcgaa ttggacggaa 240

aaaaagcgtc tatcattacc ggcacctatc cacactaaga cagtactaaa ggacgacgct 300

ccccacgaaa cgacgtttcg accttaacga ccctgccgtc tccatccatc cgaccacaat 360

ggaaaagaca ttttcaaacg attacccacc ctccgggact gaggcccaca tccacatcaa 420

ccacacggcc cactcggatg actcagagga ggtgccctcg cacaaggaaa attacaacac 480

cagtggccac gacctggagg agtccgaccc ggataaccat gtcggtgaga ccctcgaggt 540

caagcgaggt ctcaagatgc gacacatctc catgatctcg cttggaggaa ccattggtac 600

cggtctcttc attggtaccg gaggagctct ccagcaggcc ggtccctgtg gcgccctcgt 660

cgcctacgtg ttcatggcca ccattgtcta ctctgttgcc gagtctcttg gagaactggc 720

tacgtacatt cccatcaccg gctcctttgc cgtctttact acccgatatc tgtcacagtc 780

gtttggtgcc tccatgggct ggctatactg gttctcgtgg gcgatcacct tcgccatcga 840

gctcaacacc attggtcccg tgattgagta ctggactgac gccgttccta ctgctgcctg 900

gattgccatc ttcttcgtca tcctcactac catcaacttc ttccccgtgg gcttctatgg 960

cgaagtcgag ttctgggtgg cctccgtgaa ggtcattgcc atcattggat ggctcatcta 1020

cgcgctctgc atgacgtgtg gagcaggtgt aacaggtcct gtgggattca gatactggaa 1080

ccaccccgga cccatgggag acggaatctg gaccgacggc gtgcccattg tgcgaaacgc 1140

gcccggtcga cgattcatgg gatggctcaa ttcgctcgtt aacgccgcct tcacctacca 1200

gggctgtgag ctggtcggag tcactgccgg tgaggcccag aaccccagaa agtccgtccc 1260

tcgagccatc aaccgagtct ttgctcgaat ttgcatcttc tacattggct ctatcttctt 1320

catgggcatg ctcgtgccct ttaacgaccc caagctgacc gatgactcct ccgtcatcgc 1380

ctcctctcct tttgttattg ccattatcaa ctctggcacc aaggtgctcc ctcacatttt 1440

caacgccgtc attctcatca ccctgatttc ggcaggaaac tccaacgtct acattggctc 1500

gcgagtggtc tacgccctgg ctgactccgg aaccgcacca aagttcttca agcgaaccac 1560

caagaaggga gtgccgtacg tggcagtctg cttcacctcg gcgtttggtc tgctggcctt 1620

catgtctgtg tccgagtcgt cgtccactgt cttcgactgg ttcatcaaca tctccgctgt 1680

ggccggcctc atctgttggg ccttcatctc tgcctcccac atccgattca tgcaagtgct 1740

taagcacaga gggatctcca gagatacgct gcccttcaag gcacgatggc agccattcta 1800

ctcatggtac gcgctcgtct ccatcatctt catcactctc atccagggct tcacgtcctt 1860

ctggcacttt accgccgcca agttcatgac tgcatacatc tccgtcattg tctgggtcgg 1920

tttgtacatt atcttccagt gtctgttccg atgcaagttc cttatcccta ttgaggatgt 1980

ggacattgac accggccgac gagagattga cgacgatgtg tgggaggaga agatccccac 2040

aaagtggtac gagaagtttt ggaatattat tgcataagaa gatcggggat tcccgacgct 2100

ctctcctgga gcaaaccact cttac 2125

<210> 161

<211> 392

<212> DNA

<213> Artificial sequence

<220><223> Can1 Locus deletion strain

<400> 161

ggaaggcaca tatggcaagg gaaaagtctt caggtaatac atgcctgctg caactatatg 60

tactctgact cattccctca gacgtgggtc atagacagct gttttaaacc gggcaaatca 120

atctctgtcg cacaggtatt tctgcccttc aaaaccaggt tgccacatca gattccatca 180

aagtttttca gactaacttc aatcttaaac ggcatctcac aacaagcgaa ttggacggaa 240

aaaaagcgtc tatcattacc ggcacctatc cacactaaga cagtactaaa ggacgacgct 300

ccccacgaaa cgacgtttcg accttaacga ccctgccgtc tccatccatc cgaccactcc 360

cgacgctctc tcctggagca aaccactctt ac 392

<210> 162

<211> 73

<212> DNA

<213> Artificial sequence

<220><223> Forward Oligonucleotide of Figure 13-B

<400> 162

catgcgtaaa ctgactgatg agtccgtgag gacgaaacga gtaagctcgt cagtttcagt 60

acggttctag ttt 73

<210> 163

<211> 75

<212> DNA

<213> Artificial sequence

<220>

<223> Reverse Oligonucleotide of Figure 13-B

<400> 163

tctaaaacta gaaccgtact gaaactgaga cgagcttact cgtttcgtcc tcacggactc 60

atcagtcagt ttacg 75

<210> 164

<211> 211

<212> RNA

<213> Artificial Sequence

<220><223> pre-sgRNA URA3.1 (RGR-URA3.1)

<400> 164

gaacagcuga ugaguccgug aggacgaaac gaguaagcuc guccuguuca gagacaguuu 60

ccuguuuuag agcuagaaau agcaaguuaa aauaaggcua guccguuauc aacuugaaaa 120

aguggcaccg agucggugcu uuuggccggc auggucccag ccuccucgcu ggcgccggcu 180

gggcaacaug cuucggcaug gcgaauggga c 211

<210> 165

<211> 20

<212> DNA

<213> Yarrowia lipolytica

<400> 165

ctgttcagag acagtttcct 20

<210> 166

<211> 211

<212> RNA

<213> Artificial Sequence

<220><223> pre-sgRNA URA3.2 (RGR-URA3.2

<400> 166

auguuacuga ugaguccgug aggacgaaac gaguaagcuc gucuaacauc cagagaagca 60

cacguuuuag agcuagaaau agcaaguuaa aauaaggcua guccguuauc aacuugaaaa 120

aguggcaccg agucggugcu uuuggccggc augguccccag ccuccucgcu ggcgccggcu 180

gggcaacaug cuucggcaug gcgaauggga c 211

<210> 167

<211> 20

<212> DNA

<213> Yarrowia lipolytica

<400> 167

taacatccag agaagcacac 20

<210> 168

<211> 838

<212> DNA

<213> Artificial Sequence

<220><223> FBA1L promoter

<400> 168

gtttaaacag tgtacgcagt actatagagg aacaattgcc ccggagaaga cggccaggcc 60

gcctagatga caaattcaac aactcacagc tgactttctg ccattgccac tagggggggg 120

ccttttttata tggccaagcc aagctctcca cgtcggttgg gctgcaccca acaataaatg 180

ggtagggttg caccaacaaa gggatgggat ggggggtaga agatacgagg ataacggggc 240

tcaatggcac aaataagaac gaatactgcc attaagactc gtgatccagc gactgacacc 300

attgcatcat ctaagggcct caaaactacc tcggaactgc tgcgctgatc tggacaccac 360

agaggttccg agcactttag gttgcaccaa atgtcccacc aggtgcaggc agaaaacgct 420

ggaacagcgt gtacagtttg tcttaacaaa aagtgagggc gctgaggtcg agcagggtgg 480

tgtgacttgt tatagccttt agagctgcga aagcgcgtat ggatttggct catcaggcca 540

gattgagggt ctgtggacac atgtcatgtt agtgtacttc aatcgccccc tggatatagc 600

cccgacaata ggccgtggcc tcatttttttt gccttccgca catttccatt gctcggtacc 660

cacaccttgc ttctcctgca cttgccaacc ttaatactgg tttacattga ccaacatctt 720

acaagcgggg ggcttgtcta gggtatatat aaacagtggc tctcccaatc ggttgccagt 780

ctctttttttc ctttctttcc ccacagattc gaaatctaaa ctacacatca caccatgg 838

<210> 169

<211> 1839

<212> DNA

<213> Artificial Sequence

<220><223> acetohydroxyacid synthase gene

<400> 169

atggacgagt ccttcatcgg aatgtctgga ggagagatct tccacgagat gatgctgcga 60

cacaacgtcg acactgtctt cggttacccc ggtggagcca ttctccccgt ctttgacgcc 120

attcacaact ctgaatactt caactttgtg ctccctcgac acgagcaggg tgccggccac 180

atggccgagg gctacgctcg agcctctggt aagcccggtg tcgttctcgt cacctctggc 240

cccggtgcca ccaacgtcat caccccccatg caggacgctc tttccgatgg tacccctatg 300

gttgtcttca ccggtcaggt cctgacctcc gttatcggca ctgacgcctt ccaggaggcc 360

gatgttgtcg gcatctcccg atcttgcacc aagtggaacg tcatggtcaa gaacgttgct 420

gagctccccc gacgaatcaa cgaggccttt gagattgcta cttccggccg acccggtccc 480

gttctcgtcg atctgcccaa ggatgttact gctgccatcc tgcgagagcc catccccacc 540

aagtccacca ttccctcgca ttctctgacc aacctcacct ctgccgccgc caccgagttc 600

cagaagcagg ctatccagcg agccgccaac ctcatcaacc agtccaagaa gcccgtcctt 660

tacgtcggac agggtatcct tggctccgag gagggtccta agctgcttaa ggagctggct 720

gagaaggccg agattcccgt caccactact ctgcagggtc ttggtgcctt tgacgagcga 780

gaccccaagt ctctgcacat gctcggtatg cacggttccg gctacgccaa catggccatg 840

cagaacgctg actgtatcat tgctctcggc gcccgatttg atgaccgagt taccggctcc 900

atccccaagt ttgcccccga ggctcgagcc gctgcccttg agggtcgagg tggtattgtt 960

cactttgaga tccaggccaa gaacatcaac aaggttgttc aggccaccga agccgttgag 1020

ggagacgtta ccgagtctgt ccgacagctc atccccctca tcaacaaggt ctctgccgct 1080

gagcgagctc cctggactga gactatccag tcctggaagc agcagttccc cttcctcttc 1140

gaggctgaag gtgaggatgg tgttatcaag ccccagtccg tcattgctct gctctctgac 1200

ctgacagaga acaacaagga caagaccatc atcaccaccg gtgttggtca gcatcagatg 1260

tggactgccc agcatttccg atggcgacac cctcgaacca tgatcacttc tggtggtctt 1320

ggaactatgg gttacggcct gcccgccgct atcggcgcca aggttgcccg acctgactgc 1380

gacgtcattg acattgatgg tgacgcttct ttcaacatga ctctgaccga gctgtccacc 1440

gccgttcagt tcaacattgg cgtcaaggct attgtcctca acaacgagga acagggtatg 1500

gtcacccagc tgcagtctct cttctacgag aaccgatact gccacactca tcagaagaac 1560

cccgacttca tgaagctggc cgagtcgatg ggcatgaagg gtatccgaat cactcacatt 1620

gaccagctgg aggccggtct caaggagatg ctcgcataca agggccctgt gctcgttgag 1680

gttgttgtcg acaagaagat ccccgttctt cctatggttc ccgctggtaa ggctttgcat 1740

gagttccttg tctacgacgc tgacgccgag gctgcttctc gacccgatcg actgaagaat 1800

gcccccgccc ctcacgtcca ccagaccacc tttgagaac 1839

<210> 170

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> RHO705

<400> 170

gtacagacct cggccgacaa ttatgatatc 30

<210> 171

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> RHO719

<400> 171

gttccgaggt agttttgagg cccttagatg 30

<210> 172

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> RHO733

<400> 172

gttgggagag gggactagga actccttgta 30

<210> 173

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> RHO734

<400> 173

ctacgaagct cgagctaacg tccacaagtc 30

<210> 174

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> RHO707

<400> 174

gtcggtggct acaatacgtg agtcagaagg 30

<210> 175

<211> 53

<212> DNA

<213> Yarrowia lipolytica

<400> 175

cgtcctcctt cttctgttca gagacagttt cctcggcacc agctcgcagg cca 53

<210> 176

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> fragment of Plasmid URA3 from colony 1

<400> 176

cgtcctcctt cttctgttca gagacagttc ctcggcacca gctcgcaggc ca 52

<210> 177

<211> 51

<212> DNA

<213> Artificial Sequence

<220><223> fragment of Plasmid URA3 from colony 2

<400> 177

cgtcctcctt cttctgttca gagacagtcc tcggcaccag ctcgcaggcc a 51

<210

> 178

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> fragment of Plasmid URA3 from colony 3

<400> 178

cgtcctcctt cttctgttca gagacagttc ctcggcacca gctcgcaggc ca 52

<210> 179

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> fragment of Plasmid URA3 from colony 5

<400> 179

cgtcctcctt cttctgttca gagacagttc ctcggcacca gctcgcaggc ca 52

<210> 180

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> fragment of Plasmid URA3 from colony 6

<400> 180

cgtcctcctt cttctgttca gagacagttc ctcggcacca gctcgcaggc ca 52

<210> 181

<211> 39

<212> DNA

<213> Artificial Sequence

<220><223> fragment of Genomic URA3 from colony 1

<220><221> misc_feature

<222> (29)..(29)

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

<220><221> misc_feature

<222> (32)..(32)

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

<400> 181

cgtcctcctt cttctgttca gagacagtnc cncggcccc 39

<210> 182

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> fragment of Genomic URA3 from colony 2

<400> 182

cgtcctcctt cttctgttca gagacagttt acctcggcac cagctcgcag gcca 54

<210> 183

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> fragment of Genomic URA3 from colony 3

<400> 183

cgtcctcctt cttctgttca gagacagttt cagctcgcag gcca 44

<210> 184

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> fragment of Genomic URA3 from colony 5

<220><221> misc_feature

<222> (23)..(23)

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

<220><221> misc_feature

<222> (38)..(38)

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

<400> 184

cgtcctcctt cttctgttca ganacagttt ggcaccanct cgcaggcca 49

<210> 185

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> fragment of Genomic URA3 from colony 6

<400> 185

cgtcctcctt cttctgttca gagacagttt ccctcggcac cagctcgcag gcca 54

<210> 186

<211> 1029

<212> DNA

<213> Artificial Sequence

<220><223> hygromycin antibiotic resistant selection marker

<400> 186

atggccaaaa agcctgaact caccgcgacg tctgtcgaga agtttctgat cgaaaagttc 60

gacagcgtct ccgacctgat gcagctctcg gagggcgaag aatctcgtgc tttcagcttc 120

gatgtaggag ggcgtggata tgtcctgcgg gtaaatagct gcgccgatgg tttctacaaa 180

gatcgttatg tttatcggca ctttgcatcg gccgcgctcc cgattccgga agtgcttgac 240

attggggagt tcagcgagag cctgacctat tgcatctccc gccgtgcaca gggtgtcacg 300

ttgcaagacc tgcctgaaac cgaactgccc gctgttctgc agccggtcgc ggaggctatg 360

gatgcgatcg ctgcggccga tcttagccag acgagcgggt tcggcccatt cggaccgcaa 420

ggaatcggtc aatacactac atggcgtgat ttcatatgcg cgattgctga tccccatgtg 480

tatcactggc aaactgtgat ggacgacacc gtcagtgcgt ccgtcgcgca ggctctcgat 540

gagctgatgc tttgggccga ggactgcccc gaagtccggc acctcgtgca cgcggatttc 600

ggctccaaca atgtcctgac ggacaatggc cgcataacag cggtcattga ctggagcgag 660

gcgatgttcg gggattccca atacgaggtc gccaacatct tcttctggag gccgtggttg 720

gcttgtatgg agcagcagac gcgctacttc gagcggaggc atccggagct tgcaggatcg 780

ccgcggctcc gggcgtatat gctccgcatt ggtcttgacc aactctatca gagcttggtt 840

gacggcaatt tcgatgatgc agcttgggcg cagggtcgat gcgacgcaat cgtccgatcc 900

ggagccggga ctgtcgggcg tacacaaatc gcccgcagaa gcgcggccgt ctggaccgat 960

ggctgtgtag aagtactcgc cgatagtgga aaccgacgcc ccagcactcg tccgagggca 1020

aaggaatag 1029

<210> 187

<211> 565

<212> DNA

<213> Artificial Sequence

<220><223> TDH1 or GPD promoter

<400> 187

ggttgcggga tagacgccga cggagggcaa tggcgctatg gaaccttgcg gatatccata 60

cgccgcggcg gactgcgtcc gaaccagctc cagcagcgtt ttttccgggc cattgagccg 120

actgcgaccc cgccaacgtg tcttggccca cgcactcatg tcatgttggt gttgggaggc 180

cacttttaa gtagcacaag gcacctagct cgcagcaagg tgtccgaacc aaagaagcgg 240

ctgcagtggt gcaaacgggg cggaaacggc gggaaaaagc cacgggggca cgaattgagg 300

cacgccctcg aatttgagac gagtcacggc cccattcgcc cgcgcaatgg ctcgccaacg 360

cccggtcttt tgcaccacat caggttaccc caagccaaac ctttgtgtta aaaagcttaa 420

catattatac cgaacgtagg tttgggcggg cttgctccgt ctgtccaagg caacatttat 480

ataagggtct gcatcgccgg ctcaattgaa tctttttttct tcttctcttc tctatattca 540

ttcttgaatt aaacacacat caacc 565

<210> 188

<211> 43

<212> DNA

<213> Artificial Sequence

<220><223> RHO804

<400> 188

gatcaacgta cgagtgtacg cagtactata gaggaacaat tgc 43

<210> 189

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> RHO805

<400> 189

gatcaacgta cgccccaagc ttgtcccatt cgccatgccg aagc 44

<210> 190

<211> 746

<212> DNA

<213> Artificial Sequence

<220><223> TDH1 promoter-RGR-URA3.3 fusion

<400> 190

gatcaacgta cgatacgccg cggcggactg cgtccgaacc agctccagca gcgtttttttc 60

cgggccattg agccgactgc gaccccgcca acgtgtcttg gcccacgcac tcatgtcatg 120

ttggtgttgg gaggccactt tttaagtagc acaaggcacc tagctcgcag caaggtgtcc 180

gaaccaaaga agcggctgca gtggtgcaaa cggggcggaa acggcgggaa aaagccacgg 240

gggcacgaat tgaggcacgc cctcgaattt gagacgagtc acggccccat tcgcccgcgc 300

aatggctcgc caacgcccgg tcttttgcac cacatcaggt taccccaagc caaacctttg 360

tgttaaaaag cttaacatat tataccgaac gtaggtttgg gcgggcttgc tccgtctgtc 420

caaggcaaca tttatataag ggtctgcatc gccggctcaa ttgaatcttt tttcttcttc 480

tcttctctat attcattctt gaattaaaca cacatcaaca atgacagttc tgatgagtcc 540

gtgaggacga aacgagtaag ctcgtcaact gttagaggtt agactagttt tagagctaga 600

aatagcaagt taaaataagg ctagtccgtt atcaacttga aaaagtggca ccgagtcggt 660

gcttttggcc ggcatggtcc cagcctcctc gctggcgccg gctgggcaac atgcttcggc 720

atggcgaatg ggaccgtacg agtcag 746

<210> 191

<211> 211

<212> RNA

<213> Artificial Sequence

<220><223> pre-sgRNA URA3.3 (RGR-URA3.3)

<400> 191

acaguucuga ugaguccgug aggacgaaac gaguaagcuc gucaacuguu agagguuaga 60

cuaguuuuag agcuagaaau agcaaguuaa aauaaggcua guccguuauc aacuugaaaa 120

aguggcaccg agucggugcu uuuggccggc augguccccag ccuccucgcu ggcgccggcu 180

gggcaacaug cuucggcaug gcgaauggga c 211

<210> 192

<211> 43

<212> DNA

<213> Artificial Sequence

<220><223> RHO610

<400> 192

ggcggcttaa ttaagttgcg acacatgtct tgatagtatc ttg 43

<210> 193

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> RHO611

<400> 193

ggcggcttaa ttaacgagta tctgtctgac tcgtcattgc cgcc 44

<210> 194

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> RHO704

<400> 194

catatacttc actgccccag ataaggttcc 30

<210> 195

<211> 39

<212> DNA

<213> Yarrowia lipolytica

<400> 195

aaaccaacct gtgtgcttct ctggatgtta ccaccacca 39

<210> 196

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Fragment of URA3 sequence from colony 3

<220><221> misc_feature

<222> (17)..(17)

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

<220><221> misc_feature

<222> (19)..(23)

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

<400> 196

aaaccaacct gtgttgntnn nnng 24

<210> 197

<211> 37

<212> DNA

<213> Artificial Sequence

<220><223> Fragment of URA3 sequence from colony 4

<220><221> misc_feature

<222

> (21)..(21)

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

<220><221> misc_feature

<222> (24)..(25)

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

<220><221> misc_feature

<222> (27)..(29)

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

<400> 197

aaaccaacct gtgtgtttct nggnntnnnc ccccccc 37

<210> 198

<211> 39

<212> DNA

<213> Artificial Sequence

<220><223> Fragment of URA3 sequence from colony 5

<220><221> misc_feature

<222> (38)..(39)

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

<400> 198

aaaccaacct gtgtgtgtct ctggatgtta ccaccacnn 39

<210> 199

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Fragment of URA3 sequence from colony 6

<220><221> misc_feature

<222> (21)..(22)

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

<220><221> misc_feature

<222> (25)..(27)

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

<400> 199

aaaccaacct gtgtgttttc nnggnnnt 28

<210> 200

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Fragment of URA3 sequence from colony 9

<400> 200

aaaccaacct gtgttgcttc tctggatgt 29

<210> 201

<211> 39

<212> DNA

<213> Artificial Sequence

<220><223> Fragment of URA3 sequence from colony 10

<220><221> misc_feature

<222> (14)..(15)

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

<220><221> misc_feature

<222> (19)..(22)

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

<220><221> misc_feature

<222> (24)..(27)

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

<220><221> misc_feature

<222> (29)..(29)

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

<220><221> misc_feature

<222> (32)..(32)

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

<220><221> misc_feature

<222> (38)..(38)

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

<400> 201

aaaccaacct gtgnnttcnn nngnnnntnc cnccccna 39

<210> 202

<211> 351

<212> DNA

<213> Yarrowia lipolytica

<400> 202

aaaccaacct gtgtgcttct ctggatgtta ccaccaccaa ggagctcatt gagcttgccg 60

ataaggtcgg accttatgtg tgcatgatca agacccatat cgacatcatt gacgacttca 120

cctacgccgg cactgtgctc cccctcaagg agcttgctct taagcacggt ttcttcctgt 180

tcgaggacag aaagttcgca gatattggca acactgtcaa gcaccagtac aagaacggtg 240

tctaccgaat cgccgagtgg tccgatatca ccaacgccca cggtgtaccc ggagccggaa 300

tcattgctgg cctgcgagct ggtgccgagg aaactgtctc tgaacagaag a 351

<210> 203

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Fragment of URA3 sequence from colony 23

<400> 203

aaaccaacct gtgaaactgt ctctgaacag aaga 34

<210> 204

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> Fragment of URA3 sequence from colony 24

<400> 204

aaaccaacct gtgaaactgt ctctgaacag aaga 34

<210> 205

<211> 803

<212> DNA

<213> Yarrowia lypolitica

<400> 205

aaaccaacct gtgtgcttct ctggatgtta ccaccaccaa ggagctcatt gagcttgccg 60

ataaggtcgg accttatgtg tgcatgatca agacccatat cgacatcatt gacgacttca 120

cctacgccgg cactgtgctc cccctcaagg agcttgctct taagcacggt ttcttcctgt 180

tcgaggacag aaagttcgca gatattggca acactgtcaa gcaccagtac aagaacggtg 240

tctaccgaat cgccgagtgg tccgatatca ccaacgccca cggtgtaccc ggagccggaa 300

tcattgctgg cctgcgagct ggtgccgagg aaactgtctc tgaacagaag aaggaggacg 360

tctctgacta cgagaactcc cagtacaagg agttcctggt cccctctccc aacgagaagc 420

tggccagagg tctgctcatg ctggccgagc tgtcttgcaa gggctctctg gccactggcg 480

agtactccaa gcagaccatt gagcttgccc gatccgaccc cgagtttgtg gttggcttca 540

ttgcccagaa ccgacctaag ggcgactctg aggactggct tattctgacc cccggggtgg 600

gtcttgacga caagggagac gctctcggac agcagtaccg aactgttgag gatgtcatgt 660

ctaccggaac ggatatcata attgtcggcc gaggtctgta cggccagaac cgagatccta 720

ttgaggaggc caagcgatac cagaaggctg gctgggaggc ttaccagaag attaactgtt 780

agaggttaga ctatggatat gta 803

<210> 206

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Fragment of URA3 sequence from colony 27

<400> 206

aaaccaacct gtgtctatgg atatgta 27

<210> 207

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Fragment of URA3 sequence from colony 36

<400> 207

aaaccaacct gtggatatgt a 21

<210> 208

<211> 1346

<212> DNA

<213> Artificial Sequence

<220><223> ARS18 sequence

<400> 208

aattcatgtc acacaaaccg atcttcgcct caaggaaacc taattctaca tccgagagac 60

tgccgagatc cagtctacac tgattaattt tcgggccaat aatttaaaaa aatcgtgtta 120

tataatatta tatgtattat atatatacat catgatgata ctgacagtca tgtcccattg 180

ctaaatagac agactccatc tgccgcctcc aactgatgtt ctcaatattt aaggggtcat 240

ctcgcattgt ttaataataa acagactcca tctaccgcct ccaaatgatg ttctcaaaat 300

atattgtatg aacttatttt tattacttag tattattaga caacttactt gctttatgaa 360

aaacacttcc tatttaggaa acaatttata atggcagttc gttcatttaa caatttatgt 420

agaataaatg ttataaatgc gtatgggaaa tcttaaatat ggatagcata aatgatatct 480

gcattgccta attcgaaatc aacagcaacg aaaaaaatcc cttgtacaac ataaatagtc 540

atcgagaaat atcaactatc aaagaacagc tattcacacg ttactattga gattattatt 600

ggacgagaat cacacactca actgtctttc tctcttctag aaatacaggt acaagtatgt 660

actattctca ttgttcatac ttctagtcat ttcatcccac atattccttg gatttctctc 720

caatgaatga cattctatct tgcaaattca acaattataa taagatatac caaagtagcg 780

gtatagtggc aatcaaaaag cttctctggt gtgcttctcg tatttatttt tattctaatg 840

atccattaaa ggtatatatt tatttcttgt tatataatcc ttttgtttat tacatgggct 900

ggatacataa aggtattttg atttaatttt ttgcttaaat tcaatccccc ctcgttcagt 960

gtcaactgta atggtaggaa attaccatac ttttgaagaa gcaaaaaaaa tgaaagaaaa 1020

aaaaaatcgt atttccaggt tagacgttcc gcagaatcta gaatgcggta tgcggtacat 1080

tgttcttcga acgtaaaagt tgcgctccct gagatattgt acatttttgc ttttacaagt 1140

acaagtacat cgtacaacta tgtactactg ttgatgcatc cacaacagtt tgttttgttt 1200

ttttttgttt ttttttttte taatgattca ttaccgctat gtatacctac ttgtacttgt 1260

agtaagccgg gttattggcg ttcaattaat catagactta tgaatctgca cggtgtgcgc 1320

tgcgagttac ttttagctta tgcatg 1346

<210> 209

<211> 564

<212> DNA

<213> Artificial Sequence

<220><223> Yarrowia codon optimized P. aeruginosa Csy4

<400> 209

atggaccact acctggatat cagactccga cccgacccag agttccctcc tgcccagctc 60

atgtccgtct tgtttggcaa gctgcaccaa gctctcgtgg cccagggtgg agaccgaatt 120

ggcgtgtcgt tccccgattt ggacgagtcc cgttctcgac ttggagaaag actccgtatt 180

catgcttctg cagacgatct cagagctctg cttgcccgac cctggctgga gggtctccga 240

gatcatctgc agttcggcga gcctgccgtg gttccccatc ctaccccata ccgacaggtg 300

tctcgggttc aggccaaaag caaccccgag cgactcagac ggcgtcttat gcgaagacac 360

gacctgtccg aggaggaagc ccgaaagcgg atccccgaca ccgttgctcg agcgttggac 420

cttcctttcg tcacactgcg atctcaatcg actggtcagc actttcgact gttcatcaga 480

cacggacccc tgcaggtcac cgcagaggaa ggcggtttta cttgctatgg actgtccaag 540

ggtggctttg tcccctggtt ctaa 564

<210> 210

<211> 530

<212> DNA

<213> Artificial Sequence

<220><223> Yarrowia FBA1 promoter

<400> 210

catcatctaa gggcctcaaa actacctcgg aactgctgcg ctgatctgga caccacagag 60

gttccgagca ctttaggttg caccaaatgt cccaccaggt gcaggcagaa aacgctggaa 120

cagcgtgtac agtttgtctt aacaaaaagt gagggcgctg aggtcgagca gggtggtgtg 180

acttgttata gcctttagag ctgcgaaagc gcgtatggat ttggctcatc aggccagatt 240

gagggtctgt ggacacatgt catgttagtg tacttcaatc gccccctgga tatagccccg 300

acaataggcc gtggcctcat ttttttgcct tccgcacatt tccattgctc ggtacccaca 360

ccttgcttct cctgcacttg ccaaccttaa tactggttta cattgaccaa catcttacaa 420

gcgggggggct tgtctagggt atatataaac agtggctctc ccaatcggtt gccagtctct 480

tttttccttt ctttccccac agattcgaaa tctaaactac acatcacacc 530

<210> 211

<211> 659

<212> DNA

<213> Artificial Sequence

<220><223> TDH1:28bp-gCAN1-28bp

<400> 211

cggcggactg cgtccgaacc agctccagca gcgtttttttc cgggccattg agccgactgc 60

gaccccgcca acgtgtcttg gcccacgcac tcatgtcatg ttggtgttgg gaggccactt 120

tttaagtagc acaaggcacc tagctcgcag caaggtgtcc gaaccaaaga agcggctgca 180

gtggtgcaaa cggggcggaa acggcgggaa aaagccacgg gggcacgaat tgaggcacgc 240

cctcgaattt gagacgagtc acggccccat tcgcccgcgc aatggctcgc caacgcccgg 300

tcttttgcac cacatcaggt taccccaagc caaacctttg tgttaaaaag cttaacatat 360

tataccgaac gtaggtttgg gcgggcttgc tccgtctgtc caaggcaaca tttatataag 420

ggtctgcatc gccggctcaa ttgaatcttt tttcttcttc tcttctctat attcattctt 480

gaattaaaca cacatcaaca atggttcact gccgtatagg cagctaagaa atcaaacgat 540

tacccaccct cgttttagag ctagaaatag caagttaaaa taaggctagt ccgttatcaa 600

cttgaaaaag tggcaccgag tcggtgcttt tgttcactgc cgtataggca gctaagaaa 659

<210> 212

<211> 28

<212> RNA

<213> P. aeruginosa

<400> 212

guucacugcc guauaggcag cuaagaaa 28

<210> 213

<211> 156

<212> RNA

<213> Artificial Sequence

<220><223> Csy4 recognition sequence flanked sgRNA

<400> 213

guucacugcc guauaggcag cuaagaaauc aaacgauuac ccacccucgu uuuagagcua 60

gaaauagcaa guuaaaauaa ggcuaguccg uuaucaacuu gaaaaagugg caccgagucg 120

gugcuuuugu ucacugccgu auaggcagcu aagaaa 156

<210> 214

<211> 23

<212> DNA

<213> Yarrowia lipolytica

<400> 214

tcaaacgatt acccaccctc cgg 23

<210> 215

<211> 45

<212> DNA

<213> Yarrowia lipolytica

<400> 215

gaaaagacat tttcaaacga ttacccaccc tccgggactg aggcc 45

<210> 216

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> fragment of CAN1 from colony 14

<400> 216

gaaaagacat tttcaaacga ttacccacct ccgggactga ggcc 44

<210> 217

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> fragment of CAN1 from colony 16

<400> 217

gaaaagacat tttcaaacga ttacccacct ccgggactga ggcc 44

<210> 218

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> fragment of CAN1 from colony 18

<400> 218

gaaaagacat tttcaaacga ttacccacct ccgggactga ggcc 44

<210> 219

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> fragment of CAN1 from colony 19

<400> 219

gaaaagacat tttcaaacga ttacccacct ccgggactga ggcc 44

<210> 220

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> fragment of CAN1 from colony 24

<400> 220

gaaaagacat tttcaaacga ttacccacct ccgggactga ggcc 44

<210> 221

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> fragment of CAN1 from colony 25

<400> 221
gaaaagacat tttcaaacga ttacccacct ccgggactga ggcc 44

<210> 222
<211> 128
<212> RNA
<213> Artificial Sequence
<220><223> gRNA processed by Csy4
<400> 222
cuaagaaauc aaacgauuac ccacccucgu uuuagagcua gaaauagcaa guuaaaauaa 60
ggcuaguccg uuaucaacuu gaaaaagugg caccgagucg gugcuuuugu ucacugccgu 120
auaggcag 128

<210> 223
<211> 8
<212> RNA
<213> Artificial Sequence
<220><223
> 5'-flanking sequence after Csy4 cleavage
<400> 223
cuaagaaa 8

<210> 224
<211> 20
<212> RNA
<213> Artificial Sequence
<220><223> 3'-flanking sequence after Csy4 cleavage
<400> 224
guucacugcc guauaggcag 20