



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2021년04월07일

(11) 등록번호 10-2237151

(24) 등록일자 2021년04월01일

(51) 국제특허분류(Int. Cl.)  
C12N 15/66 (2006.01) C12N 15/10 (2017.01)  
C12N 15/64 (2006.01)

(52) CPC특허분류  
C12N 15/66 (2013.01)  
C12N 15/1031 (2013.01)

(21) 출원번호 10-2018-7002009(분할)

(22) 출원일자(국제) 2015년06월23일

심사청구일자 2020년06월17일

(85) 번역문제출일자 2018년01월22일

(65) 공개번호 10-2018-0011858

(43) 공개일자 2018년02월02일

(62) 원출원 특허 10-2017-7001792

원출원일자(국제) 2015년06월23일

심사청구일자 2017년01월20일

(86) 국제출원번호 PCT/US2015/037199

(87) 국제공개번호 WO 2015/200334

국제공개일자 2015년12월30일

(30) 우선권주장

62/015,809 2014년06월23일 미국(US)

(뒷면에 계속)

(56) 선행기술조사문헌

US20100035768 A1

JP2011512140 A

(73) 특허권자

리제너론 파마슈티칼스 인코포레이티드

미합중국 뉴욕주 10591 타리타운 올드 소우 밀 리버 로드 777

(72) 발명자

원해르 크리스

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드 777 리제너론 파마슈티칼스 인코포레이티드 내

맥휘터 존

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드 777 리제너론 파마슈티칼스 인코포레이티드 내

(뒷면에 계속)

(74) 대리인

특허법인와이에스장

전체 청구항 수 : 총 63 항

심사관 : 최준호

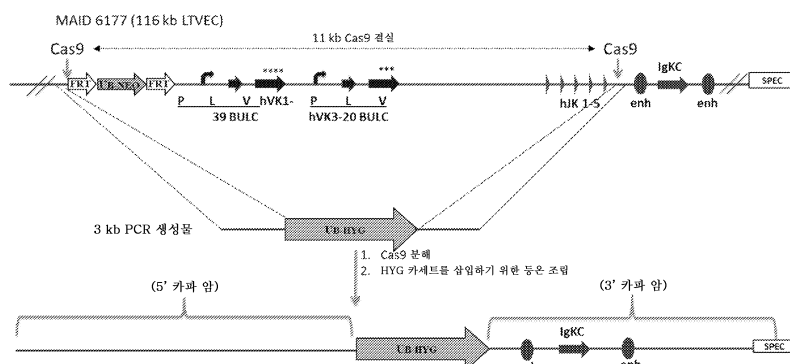
(54) 발명의 명칭 뉴클레아제-매개 DNA 조립

## (57) 요약

서열 특이적 뉴클레아제 제제 (예를 들어, gRNA-Cas 복합체)를 사용하여 상보성을 갖는 말단 서열을 생성한 다음, 중첩하는 상보적 서열을 조립하여, 2개 이상의 핵산을 조립하는 방법이 본 명세서에서 제공된다. 뉴클레아제 제제 (예를 들어, gRNA-Cas 복합체)는 중첩 말단 서열을 생성하기 위해 dsDNA에 이중 가닥 절단을 생성하게

(뒷면에 계속)

대표도 - 도1



나, 각각의 가닥 상에 틸을 생성하여 상보적 돌출 말단 서열을 생성할 수 있다. 본 명세서에 기재된 방법을 사용하는 조립은 중첩 서열을 갖는 임의의 핵산을 조립하거나, 조이너 올리고를 사용하여, 상보적 말단없이 서열을 조립할 수 있다.

(52) CPC특허분류

**C12N 15/64** (2013.01)

(72) 발명자

(30) 우선권주장

62/016,400 2014년06월24일 미국(US)

62/036,983 2014년08월13일 미국(US)

**모몬트 코레이**

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드  
777 리제너론 파마슈티칼스 인코포레이티드 내

**맥도날드 린**

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드  
777 리제너론 파마슈티칼스 인코포레이티드 내

**머피 앤드류 제이.**

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드  
777 리제너론 파마슈티칼스 인코포레이티드 내

**워쇼 그렉 에스.**

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드  
777 리제너론 파마슈티칼스 인코포레이티드 내

**로하스 호세 에프.**

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드  
777 리제너론 파마슈티칼스 인코포레이티드 내

**라이 카-만 비너스**

미국 뉴욕 10591 테리타운 웨스트 메인 스트리트  
127 유닛 201

**발렌주엘라 데이비드 엠.**

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드  
777 리제너론 파마슈티칼스 인코포레이티드 내

**몬타냐 케이틀린**

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드  
777 리제너론 파마슈티칼스 인코포레이티드 내

## 명세서

### 청구범위

#### 청구항 1

시험관 내에서 2개 이상의 이중 가닥 핵산을 끊김 없이 조립하기 위한 방법으로서, 상기 방법은

(a) 제1 핵산을 제1 뉴클레아제 제제와 접촉시키는 단계로서, 제1 뉴클레아제 제제는 제1 핵산을 제1 표적 부위에서 절단하고, 상기 절단에 의해 이중 가닥 단편이 제1 핵산의 말단으로부터 제거되어, 제1 분해된 핵산을 생성하는, 단계;

(b) 제1 분해된 핵산을 제2 핵산, 제1 조이너 올리고, 및 엑소뉴클레아제와 접촉시키는 단계로서, 제1 조이너 올리고는 50 bp 내지 400 bp인 선형 이중 가닥 DNA이고

(i) 제1 분해된 핵산에 상보적인 제1 상보적 서열;

(ii) 스페이서; 및

(iii) 제2 핵산에 상보적인 제2 상보적 서열

을 포함하며,

상기 스페이서는 제1 핵산의 말단으로부터 제거된 이중 가닥 단편과 동일한 서열을 포함하고,

제1 상보적 서열 및 제1 핵산의 말단으로부터 제거된 이중 가닥 단편과 동일한 서열 사이에 핵산 염기가 존재하지 않고, 제2 상보적 서열 및 제1 핵산의 말단으로부터 제거된 이중 가닥 단편과 동일한 서열 사이에 핵산 염기가 존재하지 않고,

상기 엑소뉴클레아제가 제1 및 제2 상보적 서열을 노출시키는,

단계; 및

(c) 제1 조이너 올리고를 제1 분해된 핵산 및 제2 핵산과 조립하는 단계로서, 조립은 제1 핵산의 말단으로부터 제거된 이중 가닥 단편을 재구성하는, 단계

를 포함하는, 방법.

#### 청구항 2

제1 항에 있어서, 단계 (c)의 조립 단계는

(i) 제1 조이너 올리고의 제1 상보적 서열을 제1 분해된 핵산에 어닐링하고 제1 조이너 올리고의 제2 상보적 서열을 제2 핵산에 어닐링하는 단계; 및

(ii) 제1 조이너 올리고를 제1 분해된 핵산 및 제2 핵산에 라이게이션하는 단계

를 포함하는 것을 특징으로 하는 방법.

#### 청구항 3

제2 항에 있어서, 단계 (c)(i)은 제1 분해된 핵산의 3' 말단을 연장하는 단계, 제2 핵산의 3' 말단을 연장하는 단계, 또는 제1 분해된 핵산 및 제2 핵산의 3' 말단을 연장하는 단계를 더 포함하는 것을 특징으로 하는 방법.

#### 청구항 4

제1 항 내지 제3 항 중 어느 한 항에 있어서, 단계 (a)는

(i) 제2 핵산을 제2 뉴클레아제 제제와 접촉시키는 단계로서, 제2 뉴클레아제 제제는 제2 표적 부위에서 제2 핵산을 절단하여 제1 조이너 올리고의 제2 상보적 서열에 상보적인 뉴클레오티드 서열을 포함하는 제2 분해된 핵산을 생산하고, 제1 분해된 핵산이 제2 분해된 핵산에 조립되는, 단계; 또는

(ii) 제2 핵산을 제한 효소와 접촉시키는 단계로서, 상기 제한 효소는 제2 핵산을 절단하여 제1 조이너 올리고의 제2 상보적 서열에 상보적인 뉴클레오티드 서열을 포함하는 제2 분해된 핵산을 생산하고, 제1 분해된 핵산이 제2 분해된 핵산에 조립되는, 단계

를 더 포함하는 것을 특징으로 하는 방법.

#### 청구항 5

제1 항 내지 제3 항 중 어느 한 항에 있어서,

(I) 제1 조이너 올리고의 제1 상보적 서열은 15 내지 120개의 상보적 염기이고, 제1 조이너 올리고의 제2 상보적 서열은 15 내지 120개의 상보적 염기인 것;

(II) 제1 핵산의 말단으로부터 제거된 이중 가닥 단편은 적어도 20 bp인 것;

(III) 제1 조이너 올리고의 스페이서는 20 bp 내지 120 bp인 것; 또는

(IV) 제1 조이너 올리고는 100 bp 내지 300 bp인 것

을 특징으로 하는 방법.

#### 청구항 6

제5 항에 있어서, 제1 조이너 올리고의 제1 상보적 서열은 20 내지 80개의 상보적인 염기이고, 제1 조이너 올리고의 제2 상보적 서열은 20 내지 80개의 상보적인 염기인 것을 특징으로 하는 방법.

#### 청구항 7

제1 항 내지 제3 항 중 어느 한 항에 있어서, 제1 조이너 올리고는 동시에 제1 분해된 핵산 및 제2 핵산에 조립되는 것을 특징으로 하는 방법.

#### 청구항 8

제1 항 내지 제3 항 중 어느 한 항에 있어서, 적어도 3개의 핵산이 조립되는 것을 특징으로 하는 방법.

#### 청구항 9

시험관 내에서 2개 이상의 핵산을 조립하기 위한 방법으로서,

(a) 제1 핵산을 제1 뉴클레아제 제제 및 제2 뉴클레아제 제제와 접촉시키는 단계로서, 제1 뉴클레아제 제제는 제1 표적 부위에서 제1 핵산을 절단하고 제2 뉴클레아제 제제는 제2 표적 부위에서 제1 핵산을 절단하여 제1 분해된 핵산을 생성하는, 단계;

(b) 제1 분해된 핵산을 제1 조이너 올리고, 제2 핵산, 제2 조이너 올리고, 및 엑소뉴클레아제와 접촉시키는 단계로서,

제1 조이너 올리고는 50 bp 내지 400 bp인 선형 이중 가닥 DNA이거나, 제2 조이너 올리고는 50 bp 내지 400 bp인 선형 이중 가닥 DNA이거나, 또는 제1 조이너 올리고와 제2 조이너 올리고 모두 50 bp 내지 400 bp인 선형 이중 가닥 DNA이고,

제1 조이너 올리고는

(i) 제1 분해된 핵산에 상보적인 제1 상보적 서열; 및

(ii) 제2 핵산에 상보적인 제2 상보적 서열

을 포함하고;

제2 조이너 올리고는

(i) 제2 핵산에 상보적인 제1 상보적 서열; 및

(ii) 제1 분해된 핵산에 상보적인 제2 상보적 서열

을 포함하고;

상기 엑소뉴클레아제가 제1 조이너 올리고, 제2 조이너 올리고, 제1 분해된 핵산, 및 제2 핵산의 상보적 서열을 노출시키는,

단계; 및

(c) 제1 분해된 핵산, 제1 조이너 올리고, 제2 핵산, 및 제2 조이너 올리고를 조립하는 단계를 포함하는, 방법.

#### 청구항 10

제9 항에 있어서, 단계 (c)의 조립 단계는

(i) 제1 조이너 올리고의 제1 상보적 서열을 제1 분해된 핵산에 어닐링하고, 제1 조이너 올리고의 제2 상보적 서열을 제2 핵산에 어닐링하고, 제2 조이너 올리고의 제1 상보적 서열을 제2 핵산에 어닐링하고, 제2 조이너 올리고의 제2 상보적 서열을 제1 분해된 핵산에 어닐링하는 단계; 및

(ii) 제1 분해된 핵산을 제1 조이너 올리고에 라이게이션하고, 제1 조이너 올리고를 제2 핵산에 라이게이션하고, 제2 핵산을 제2 조이너 올리고에 라이게이션하고, 제2 조이너 올리고를 제1 분해된 핵산에 라이게이션하는 단계

를 포함하는 것을 특징으로 하는 방법.

#### 청구항 11

제10 항에 있어서, 단계 (c)(i)은 어닐링된 상보적 서열의 3' 말단을 연장하는 단계를 더 포함하는 것을 특징으로 하는 방법.

#### 청구항 12

제9 항 내지 제11 항 중 어느 한 항에 있어서, 상기 2개 이상의 핵산은 이중 가닥 핵산이고, 제1 뉴클레아제 제제는 제1 표적 부위에서 제1 핵산을 절단하여 제1 이중 가닥 절단을 생성하고, 제2 뉴클레아제 제제는 제2 표적 부위에서 제1 핵산을 절단하여 제2 이중 가닥 절단을 생성하는 것을 특징으로 하는 방법.

#### 청구항 13

제9 항 내지 제11 항 중 어느 한 항에 있어서,

(I) 제1 조이너 올리고는 선형 이중 가닥 DNA이고, 제2 조이너 올리고는 선형 이중 가닥 DNA인 것;

(II) (i) 제1 조이너 올리고의 제1 상보적 서열 및 제2 상보적 서열은 각각 15 내지 120개의 상보적 염기이거나, (ii) 제2 조이너 올리고의 제1 상보적 서열 및 제2 상보적 서열은 각각 15 내지 120개의 상보적 염기이거나, 또는 (iii) 제1 조이너 올리고의 제1 상보적 서열 및 제2 상보적 서열은 각각 15 내지 120개의 상보적 염기이고, 제2 조이너 올리고의 제1 상보적 서열 및 제2 상보적 서열은 각각 15 내지 120개의 상보적 염기인 것; 또는

(III) 제1 조이너 올리고는 선형 이중 가닥 DNA이고, 제2 조이너 올리고는 선형 이중 가닥 DNA이고; 및 (i) 제1 조이너 올리고의 제1 상보적 서열 및 제2 상보적 서열은 각각 15 내지 120개의 상보적 염기이거나, (ii) 제2 조이너 올리고의 제1 상보적 서열 및 제2 상보적 서열은 각각 15 내지 120개의 상보적 염기이거나, 또는 (iii) 제1 조이너 올리고의 제1 상보적 서열 및 제2 상보적 서열은 각각 15 내지 120개의 상보적 염기이고, 제2 조이너 올리고의 제1 상보적 서열 및 제2 상보적 서열은 각각 15 내지 120개의 상보적 염기인 것

을 특징으로 하는 방법.

#### 청구항 14

제13 항에 있어서, (i) 제1 조이너 올리고의 제1 상보적 서열 및 제2 상보적 서열은 각각 20 내지 80개의 상보적 염기인 것, (ii) 제2 조이너 올리고의 제1 상보적 서열 및 제2 상보적 서열은 각각 20 내지 80개의 상보적 염기인 것, 또는 (iii) 제1 조이너 올리고의 제1 상보적 서열 및 제2 상보적 서열은 각각 20 내지 80개의 상보적 염기이고, 제2 조이너 올리고의 제1 상보적 서열 및 제2 상보적 서열은 각각 20 내지 80개의 상보적 염기인 것을 특징으로 하는 방법.

#### 청구항 15

제9 항 내지 제11 항 중 어느 한 항에 있어서, 제1 조이너 올리고는 50 bp 내지 400 bp의 선형 이중 가닥 DNA이고, 제2 조이너 올리고는 50 bp 내지 400 bp의 선형 이중 가닥 DNA인 것을 특징으로 하는 방법.

#### 청구항 16

제15 항에 있어서, 제1 조이너 올리고는 100 bp 내지 300 bp 이고, 제2 조이너 올리고는 100 bp 내지 300 bp 인 것을 특징으로 하는 방법.

#### 청구항 17

제9 항 내지 제11 항 중 어느 한 항에 있어서, 제1 조이너 올리고는 제1 상보적 서열과 제2 상보적 서열 사이에 스페이서를 더 포함하거나; 제2 조이너 올리고는 제1 상보적 서열과 제2 상보적 서열 사이에 스페이서를 더 포함하거나; 또는 제1 조이너 올리고는 제1 상보적 서열과 제2 상보적 서열 사이에 스페이서를 더 포함하고 제2 조이너 올리고는 제1 상보적 서열과 제2 상보적 서열 사이에 스페이서를 더 포함하는 것을 특징으로 하는 방법.

#### 청구항 18

제17 항에 있어서,

(I) 제1 조이너 올리고의 스페이서, 제2 조이너 올리고의 스페이서, 또는 둘 다는 성공적인 조립을 확인하기 위해 약물 저항성 유전자, 리포터 유전자, 또는 1개 이상의 제한 효소 부위를 포함하는 것; 또는

(II) 제1 조이너 올리고의 스페이서는 20 bp 내지 120 bp 이거나; 제2 조이너 올리고의 스페이서는 20 bp 내지 120 bp 이거나; 또는 제1 조이너 올리고의 스페이서는 20 bp 내지 120 bp 이고, 제2 조이너 올리고의 스페이서는 20 bp 내지 120 bp 인 것

을 특징으로 하는 방법.

#### 청구항 19

제9 항 내지 제11 항 중 어느 한 항에 있어서, 동시에 제1 분해된 핵산은 제1 조이너 올리고에 조립되고, 제1 조이너 올리고는 제2 핵산에 조립되고, 제2 핵산은 제2 조이너 올리고에 조립되고, 제2 조이너 올리고는 제1 분해된 핵산에 조립되는 것을 특징으로 하는 방법.

#### 청구항 20

제9 항 내지 제11 항 중 어느 한 항에 있어서, 제1 분해된 핵산은 제2 핵산에 끊김 없이 조립되는 것을 특징으로 하는 방법.

#### 청구항 21

제20 항에 있어서, 제1 뉴클레아제 제제 또는 제2 뉴클레아제 제제에 의한 절단에 의해 이중 가닥 단편이 제1 핵산의 말단으로부터 제거되며,

제1 조이너 올리고는 제1 상보적 서열과 제2 상보적 서열 사이에 스페이서를 더 포함하거나, 제2 조이너 올리고는 제1 상보적 서열과 제2 상보적 서열 사이에 스페이서를 더 포함하고,

상기 스페이서는 제1 핵산의 말단으로부터 제거된 이중 가닥 단편과 동일한 서열을 포함하며, 제1 상보적 서열 및 제1 핵산의 말단으로부터 제거된 이중 가닥 단편과 동일한 서열 사이에 핵산 염기가 존재하지 않고, 제2 상보적 서열 및 제1 핵산의 말단으로부터 제거된 이중 가닥 단편과 동일한 서열 사이에 핵산 염기가 존재하지 않는 것을 특징으로 하는 방법.

#### 청구항 22

제9 항 내지 제11 항 중 어느 한 항에 있어서, 단계 (a)는 제2 핵산을 제3 뉴클레아제 제제와 접촉시키는 단계를 더 포함하며, 제3 뉴클레아제 제제는 제3 표적 부위에서 제2 핵산을 절단하여 제2 분해된 핵산을 생성하는 것을 특징으로 하는 방법.

#### 청구항 23

제9 항 내지 제11 항 중 어느 한 항에 있어서, 단계 (a)는 제2 핵산을 제3 뉴클레아제 제제 및 제4 뉴클레아제 제제와 접촉시키는 단계를 더 포함하며, 제3 뉴클레아제 제제는 제3 표적 부위에서 제2 핵산을 절단하고 제4 뉴클레아제 제제는 제4 표적 부위에서 제2 핵산을 절단하여 제2 분해된 핵산을 생성하는 것을 특징으로 하는 방법.

#### 청구항 24

제9 항 내지 제11 항 중 어느 한 항에 있어서, 제1 조이너 올리고 및 제2 조이너 올리고 중 적어도 하나는 단일 가닥 DNA인 것을 특징으로 하는 방법.

#### 청구항 25

제1 항 내지 제3 항 중 어느 한 항에 있어서, 제1 뉴클레아제 제제는 Cas 단백질 및 가이드 RNA (gRNA) (gRNA-Cas 복합체), 징크 핑거 뉴클레아제, 또는 전사 활성인자-유사 이펙터 뉴클레아제 (TALEN)를 포함하는 것을 특징으로 하는 방법.

#### 청구항 26

제25 항에 있어서, 제1 뉴클레아제 제제는 Cas 단백질 및 gRNA를 포함하고, Cas 단백질은 Cas9 단백질이고, gRNA는 규칙적으로 사이 간격을 두고 분포하는 짧은 회문구조 반복 서열 (CRISPR) RNA (crRNA) 및 트랜스-활성화 CRISPR RNA (tracrRNA)를 암호화하는 핵산 서열을 포함하고, 제1 뉴클레아제 제제는 프로토스페이서 인접 모티프 (PAM) 서열에 바로 플랭킹된 표적 부위를 표적화하는 것을 특징으로 하는 방법.

#### 청구항 27

제1 항 내지 제3 항 중 어느 한 항에 있어서,

(I) 제1 핵산, 제2 핵산, 또는 이들 핵산 모두는 길이가 20 kb 내지 400 kb인 벡터인 것; 또는

(II) 조립된 핵산은 길이가 30 kb 내지 1 Mb인 것

을 특징으로 하는 방법.

#### 청구항 28

제1 항 내지 제3 항 중 어느 한 항에 있어서, 제1 핵산, 제2 핵산, 또는 이들 핵산 모두는 적어도 10 kb인 것을 특징으로 하는 방법.

#### 청구항 29

제1 항 내지 제3 항 중 어느 한 항에 있어서,

(I) 제1 핵산, 제2 핵산, 또는 이들 핵산 모두는 박테리아 인공 염색체를 포함하는 것;

(II) 제1 핵산, 제2 핵산, 또는 이들 핵산 모두는 인간 DNA, 설치류 DNA, 합성 DNA, 또는 이들의 조합을 포함하는 것; 또는

(III) 제1 핵산은 제1 박테리아 인공 염색체를 포함하고, 제2 핵산은 제2 박테리아 인공 염색체를 포함하고, 관심 유전자는 제1 및 제2 박테리아 인공 염색체로부터 유래되고, 상기 조립하는 단계에 의해 관심 유전자의 서열이 형성되는 것

을 특징으로 하는 방법.

#### 청구항 30

제1 항 내지 제3 항 중 어느 한 항에 있어서,

(I) 제1 핵산은 원형 핵산 또는 선형 핵산인 것; 또는

(II) 조립된 핵산은 원형 핵산 또는 선형 핵산인 것

을 특징으로 하는 방법.

### 청구항 31

제1 항 내지 제3 항 중 어느 한 항에 있어서, 방법은 제1 분해된 핵산, 제2 핵산 및 제1 조이너 올리고를 리가아제, 엑소뉴클레아제, DNA 폴리머라아제, 및 뉴클레오티드와 조합하는 단계, 및 일정한 온도에서 인큐베이션하는 단계를 포함하는 것을 특징으로 하는 방법.

### 청구항 32

제31 항에 있어서, 조립은 1 단계 등은 반응에서 발생하는 것을 특징으로 하는 방법.

### 청구항 33

시험관 내에서 3개 이상의 핵산을 조립하기 위한 방법으로서,

(a) 제1 핵산을 제1 뉴클레아제 제제 및 제2 뉴클레아제 제제와 접촉시키는 단계로서, 제1 뉴클레아제 제제는 제1 표적 부위에서 제1 핵산을 절단하고 제2 뉴클레아제 제제는 제2 표적 부위에서 제1 핵산을 절단하여 제1 분해된 핵산을 생성하는, 단계;

(b) 제1 분해된 핵산을 제1 조이너 올리고, 제2 핵산, 제2 조이너 올리고, 제3 핵산, 제3 조이너 올리고, 및 엑소뉴클레아제와 접촉시키는 단계로서,

제1 조이너 올리고는

(i) 제1 분해된 핵산에 상보적인 제1 상보적 서열; 및

(ii) 제2 핵산에 상보적인 제2 상보적 서열

을 포함하고;

제2 조이너 올리고는

(i) 제2 핵산에 상보적인 제1 상보적 서열; 및

(ii) 제3 핵산에 상보적인 제2 상보적 서열

을 포함하고;

제3 조이너 올리고는

(i) 제3 핵산에 상보적인 제1 상보적 서열; 및

(ii) 제1 분해된 핵산에 상보적인 제2 상보적 서열

을 포함하며,

상기 엑소뉴클레아제가 제1 조이너 올리고, 제2 조이너 올리고, 제3 조이너 올리고, 제1 분해된 핵산, 제2 핵산, 및 제3 핵산의 상보적 서열을 노출시키는,

단계를; 및

(c) 제1 분해된 핵산, 제1 조이너 올리고, 제2 핵산, 제2 조이너 올리고, 제3 핵산, 및 제3 조이너 올리고를 조립하는 단계

를 포함하는, 방법.

### 청구항 34

시험관 내에서 4개 이상의 핵산을 조립하기 위한 방법으로서,

(a) 제1 핵산을 제1 뉴클레아제 제제 및 제2 뉴클레아제 제제와 접촉시키는 단계로서, 제1 뉴클레아제 제제는 제1 표적 부위에서 제1 핵산을 절단하고 제2 뉴클레아제 제제는 제2 표적 부위에서 제1 핵산을 절단하여 제1 분해된 핵산을 생성하는, 단계;

(b) 제1 분해된 핵산을 제1 조이너 올리고, 제2 핵산, 제2 조이너 올리고, 제3 핵산, 제3 조이너 올리고, 제4



핵산, 제4 조이너 올리고, 및 엑소뉴클레아제와 접촉시키는 단계로서,

제1 조이너 올리고는

(i) 제1 분해된 핵산에 상보적인 제1 상보적 서열; 및

(ii) 제2 핵산에 상보적인 제2 상보적 서열

을 포함하고;

제2 조이너 올리고는

(i) 제2 핵산에 상보적인 제1 상보적 서열; 및

(ii) 제3 핵산에 상보적인 제2 상보적 서열

을 포함하고;

제3 조이너 올리고는

(i) 제3 핵산에 상보적인 제1 상보적 서열; 및

(ii) 제4 핵산에 상보적인 제2 상보적 서열

을 포함하고;

제4 조이너 올리고는

(i) 제4 핵산에 상보적인 제1 상보적 서열; 및

(ii) 제1 분해된 핵산에 상보적인 제2 상보적 서열

을 포함하고;

상기 엑소뉴클레아제가 제1 조이너 올리고, 제2 조이너 올리고, 제3 조이너 올리고, 제4 조이너 올리고, 제1 분해된 핵산, 제2 핵산, 제3 핵산, 및 제4 핵산의 상보적 서열을 노출시키는,

단계; 및

(c) 제1 분해된 핵산, 제1 조이너 올리고, 제2 핵산, 제2 조이너 올리고, 제3 핵산, 제3 조이너 올리고, 제4 핵산, 및 제4 조이너 올리고를 조립하는 단계

를 포함하는, 방법.

### 청구항 35

제33 항에 있어서, 단계 (c)의 조립 단계는

(i) 제1 분해된 핵산, 제2 핵산, 제3 핵산, 제1 조이너 올리고, 제2 조이너 올리고 및 제3 조이너 올리고를 어닐링하는 단계; 및

(ii) 제1 분해된 핵산, 제2 핵산, 제3 핵산, 제1 조이너 올리고, 제2 조이너 올리고 및 제3 조이너 올리고를 라이게이션하는 단계

를 포함하는 것을 특징으로 하는 방법.

### 청구항 36

제35 항에 있어서, 단계 (c)(i)은 어닐링된 서열의 3' 말단을 연장하는 단계를 더 포함하는 것을 특징으로 하는 방법.

### 청구항 37

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 제1 핵산, 제2 핵산 및 제3 핵산은 이중 가닥 핵산이고, 제1 뉴클레아제 제제는 제1 표적 부위에서 제1 핵산을 절단하여 제1 이중 가닥 절단을 생성하는 것을 특징으로 하는 방법.

#### 청구항 38

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 제1 뉴클레아제 제제는 Cas 단백질 및 가이드 RNA (gRNA) (gRNA-Cas 복합체), 징크 핑거 뉴클레아제, 또는 전사 활성인자-유사 이펙터 뉴클레아제 (TALEN)를 포함하는 것을 특징으로 하는 방법.

#### 청구항 39

제38 항에 있어서, 제1 뉴클레아제 제제는 Cas 단백질 및 gRNA를 포함하고, Cas 단백질은 Cas9 단백질이고, gRNA는 규칙적으로 사이 간격을 두고 분포하는 짧은 회문구조 반복 서열 (CRISPR) RNA (crRNA) 및 트랜스-활성화 CRISPR RNA (tracrRNA)를 암호화하는 핵산 서열을 포함하고, 제1 뉴클레아제 제제는 프로토스페이서 인접 모티프 (PAM) 서열에 의해 바로 플랭킹된 표적 부위를 표적화하는 것을 특징으로 하는 방법.

#### 청구항 40

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 방법은 제1 핵산, 제2 핵산, 제3 핵산, 제1 조이너 올리고, 제2 조이너 올리고 및 제3 조이너 올리고를 리가아제, 엑소뉴클레아제, DNA 폴리머라아제, 및 뉴클레오티드와 조합하는 단계 및 일정한 온도에서 인큐베이션하는 단계를 포함하는 것을 특징으로 하는 방법.

#### 청구항 41

제40 항에 있어서, 조립은 1 단계 등은 반응에서 발생하는 것을 특징으로 하는 방법.

#### 청구항 42

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 조이너 올리고 중 하나 이상 또는 전부는 단일 가닥 DNA인 것을 특징으로 하는 방법.

#### 청구항 43

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 조이너 올리고 중 하나 이상 또는 전부는 선형 이중 가닥 DNA인 것을 특징으로 하는 방법.

#### 청구항 44

제43 항에 있어서, 조이너 올리고 중 하나 이상 또는 전부는 50 bp 내지 400 bp인 것을 특징으로 하는 방법.

#### 청구항 45

제44 항에 있어서, 조이너 올리고 중 하나 이상 또는 전부는 100 bp 내지 300 bp인 것을 특징으로 하는 방법.

#### 청구항 46

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 조이너 올리고 중 하나 이상 또는 전부의 제1 상보적 서열 및 제2 상보적 서열은 각각 15 내지 120개의 상보적 염기인 것을 특징으로 하는 방법.

#### 청구항 47

제46 항에 있어서, 조이너 올리고 중 하나 이상 또는 전부의 제1 상보적 서열 및 제2 상보적 서열은 각각 20 내지 80개의 상보적 염기인 것을 특징으로 하는 방법.

#### 청구항 48

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 조이너 올리고 중 하나 이상 또는 전부는 제1 상보적 서열 및 제2 상보적 서열 사이에 스페이서를 더 포함하는 것을 특징으로 하는 방법.

#### 청구항 49

제48 항에 있어서, 스페이서는 성공적인 조립을 확인하기 위해 약물 저항성 유전자, 리포터 유전자, 또는 하나 이상의 제한 효소 부위를 포함하는 것을 특징으로 하는 방법.

#### 청구항 50

제48 항에 있어서, 스페이서는 20 bp 내지 120 bp인 것을 특징으로 하는 방법.

#### 청구항 51

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 하나 이상의 또는 모든 핵산은 적어도 10 kb인 것을 특징으로 하는 방법.

#### 청구항 52

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 하나 이상의 또는 모든 핵산은 길이가 20 kb 내지 400 kb의 벡터인 것을 특징으로 하는 방법.

#### 청구항 53

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 조립된 핵산은 길이가 30 kb 내지 1 Mb인 것을 특징으로 하는 방법.

#### 청구항 54

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 동시에 제1 분해된 핵산은 제1 조이너 올리고에 조립되고, 제1 조이너 올리고는 제2 핵산에 조립되고, 제2 핵산은 제2 조이너 올리고에 조립되고, 제2 조이너 올리고는 제3 핵산에 조립되고, 제3 핵산은 제3 조이너 올리고에 조립되는 것을 특징으로 하는 방법.

#### 청구항 55

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 하나 이상의 또는 모든 핵산은 박테리아 인공 염색체를 포함하는 것을 특징으로 하는 방법.

#### 청구항 56

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 하나 이상의 또는 모든 핵산은 인간 DNA, 설치류 DNA, 합성 DNA, 또는 이들의 조합을 포함하는 것을 특징으로 하는 방법.

#### 청구항 57

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 제1 분해된 핵산은 제2 핵산에 끊김 없이 조립되는 것을 특징으로 하는 방법.

#### 청구항 58

제57 항에 있어서, 제1 뉴클레아제 제제에 의한 절단에 의해 이중 가닥 단편이 제1 핵산의 말단으로부터 제거되고,

제1 조이너 올리고는 제1 상보적 서열 및 제2 상보적 서열 사이에 스페이서를 더 포함하고,

상기 스페이서는 제1 핵산의 말단으로부터 제거된 이중 가닥 단편과 동일한 서열을 포함하고, 제1 상보적 서열 및 제1 핵산의 말단으로부터 제거된 이중 가닥 단편과 동일한 서열 사이에 핵산 염기가 존재하지 않고, 제2 상보적 서열 및 제1 핵산의 말단으로부터 제거된 이중 가닥 단편과 동일한 서열 사이에 핵산 염기가 존재하지 않는 것을 특징으로 하는 방법.

#### 청구항 59

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 각각의 핵산은 박테리아 인공 염색체를 포함하고, 관심 유전자는 상기 박테리아 인공 염색체로부터 유래되고, 상기 조립하는 단계에 의해 관심 유전자의 서열이 형성되는 것을 특징으로 하는 방법.

#### 청구항 60

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 제1 핵산은 원형 핵산인 것을 특징으로 하는 방법.

#### 청구항 61

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 제1 핵산은 선형 핵산인 것을 특징으로 하는 방법.

#### 청구항 62

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 조립된 핵산은 원형 핵산인 것을 특징으로 하는 방법.

#### 청구항 63

제33 항, 제35 항 및 제36 항 중 어느 한 항에 있어서, 조립된 핵산은 선형 핵산인 것을 특징으로 하는 방법.

#### 청구항 64

삭제

#### 청구항 65

삭제

#### 청구항 66

삭제

### 발명의 설명

#### 기술 분야

[0001] 관련 출원과의 상호 참조

[0002] 본 출원은 2014년 6월 23일에 출원된 미국 가출원 제62/015,809호, 2014년 6월 24일에 출원된 미국 가출원 제62/016,400호 및 2014년 8월 13일에 출원된 미국 가출원 제62/036,983호의 이익을 주장하며, 이들 각각은 전체적으로 본 명세서에 참고로 포함된다.

[0003] EFS 웹을 통한 텍스트 파일로 제출된 서열 목록의 참조

[0004] 본 서열 목록의 공식 사본은 2015년 6월 23일자로 작성되고 크기가 66KB이며 본 명세서와 함께 제출된, 파일명 461002SEQLIST.TXT의 ASCII 포맷된 서열 목록으로서 EFS-웹을 통해 전자적으로 제출되었다. 상기 ASCII 포맷된 서류에 포함된 서열 목록은 본 명세서의 일부이고 전체적으로 본 명세서에 참조로 포함된다.

#### 배경 기술

[0005] 역사적으로, 중첩 연장은 중첩 합성 올리고뉴클레오타이드로부터 더 큰 이중 가닥 DNA 분자, 특히 유전자의 합성 수단으로 사용될 수 있다. 그러나, 이러한 방법은 큰 DNA 분자를 신속한 방식으로 효과적으로 조합할 수 없다. 또한, 중첩 서열을 사용하는 큰 핵산의 부위 특이적 조합은 종종 조합될 핵산 내의 원하는 위치에서의 중첩 서열의 이용 가능성에 의해 제한된다. 특정 DNA 서열을 표적화하도록 설계된 유전자 조작된(engineered) 뉴클레아제 효소는 표적 유전자의 결실, 대체 및 수복뿐만 아니라 외인성 서열을 삽입할 수 있게 하는 유전자 조작을 위한 강력한 도구로서 주목을 받고 있다. 그러나, 기존 기술은 예측할 수 없는 오프-타겟 효과(off-target effect)와 시간 소모적인 다단계 반응으로 이어질 수 있는, 제한된 정밀도(limited precision)의 문제를 겪는다.

#### 발명의 내용

[0006] 중첩 서열을 갖는 핵산을 조립하는 방법이 본 명세서에서 제공된다. 이러한 방법은 2개 이상의 핵산을 조립하는 방법을 포함하며, 이는 (a) 제1 핵산을 제1 뉴클레아제 체제와 접촉시키는 단계 (여기서, 제1 뉴클레아제 체제는 제1 핵산을 제1 표적 부위에서 절단하여, 제1 분해된 핵산과 제2 핵산 사이의 중첩 말단 서열을 포함하는 제1 분해된 핵산을 생성한다); (b) 제1 분해된 핵산과 제2 핵산을 엑소뉴클레아제와 접촉시켜, 제1 분해된 핵산과 제2 핵산 사이의 상보적 서열을 노출시키는 단계; 및 (c) 단계 (b)로부터 생성된 2개의 핵산 단편을 조립하는 단계를 포함한다. 이러한 일부 방법에서, 단계 (c)는 (i) 노출된 상보적 서열을 어닐링(annealing)하는 단계; (ii) 어닐링된 상보적 서열의 3' 말단을 연장하는 단계; 및 (iii) 제1 핵산과 제2 핵산을 라이게이션

(ligating)하는 단계를 추가로 포함한다.

- [0007] 일부 방법에서, 단계 (a)는 제2 핵산을 제2 뉴클레아제 제제와 접촉시키는 단계를 추가로 포함하며, 여기서, 제2 핵산은 중첩 말단 서열을 포함하지 않고, 제2 뉴클레아제 제제는 제2 핵산을 제2 표적 부위에서 절단하여, 제1 분해된 핵산과 제2 분해된 핵산 사이의 중첩 말단 서열을 포함하는 제2 분해된 핵산을 생성하며, 단계 (b)의 제2 핵산은 제2 분해된 핵산이다. 일부 방법에서, 중첩 말단 서열은 20 bp 내지 200 bp 범위의 길이이다.
- [0008] 일부 방법에서, 제1 뉴클레아제 제제 또는 제2 뉴클레아제 제제 중 적어도 하나는 제1 표적 부위 또는 제2 표적 부위를 표적화하는 Cas 단백질 및 가이드 RNA (gRNA) (gRNA-Cas 복합체)를 포함한다. 예를 들어, Cas 단백질은 Cas9 단백질일 수 있다. Cas9 단백질은 RuvC 도메인과 HNH 도메인을 포함할 수 있고, 이들 중 적어도 하나는 엔도뉴클레아제 활성이 결여되어 있다. 일부 실시 형태에서, gRNA는 규칙적으로 사이 간격을 두고 분포하는 짧은 회문구조 반복 서열 (Clustered Regularly Interspaced Short Palindromic Repeats; CRISPR) RNA (crRNA) 및 트랜스-활성화 CRISPR RNA (tracrRNA)를 암호화하는 핵산 서열을 포함한다. 제1 표적 부위 및/또는 제2 표적 부위는 프로토스페이스 인접 모티프 (Protospacer Adjacent Motif; PAM) 서열에 의해 플랭킹(flanked)될 수 있다. 일부 방법에서, 뉴클레아제 제제는 징크 핑거 뉴클레아제 또는 전사 활성화인자-유사 이펙터 뉴클레아제 (Transcription Activator-Like Effector Nuclease; 탈렌(TALEN))를 포함한다.
- [0009] 일부 방법에서, 제1 핵산, 제2 핵산 또는 이들 모두는 박테리아 인공 염색체로부터 유래된다. 박테리아 인공 염색체는 인간 DNA, 설치류 DNA, 합성 DNA 또는 이들의 조합을 포함할 수 있다. 박테리아 인공 염색체는 인간 서열을 포함할 수 있다.
- [0010] 본 명세서에 개시된 방법은 2개 이상의 핵산을 조립하는 방법을 포함하며, 이는 (a) 제1 핵산을 제1 뉴클레아제 제제 및 제2 뉴클레아제 제제와 접촉시켜, 제1 분해된 핵산을 생성하는 단계 (여기서, 제1 뉴클레아제 제제는 제1 핵산의 제1 가닥 상의 제1 표적 부위에서 틈(nick)을 생성하고, 제2 뉴클레아제 제제는 제1 핵산의 제2 가닥 상의 제2 표적 부위에서 틈을 생성하여, 이의 말단 중 하나에서 5' 또는 3' 돌출 서열(overhanging sequence)을 포함하는 제1 분해된 핵산을 생성한다); (b) 5' 또는 3' 돌출 서열에 상보적 서열을 포함하는 제1 분해된 핵산과 제2 핵산을 어닐링하는 단계; 및 (c) 제1 분해된 핵산과 제2 핵산을 라이게이션하는 단계를 포함한다. 일부 방법에서, 단계 (b)는 제2 가닥을 주형으로 사용하여 제1 가닥의 3' 말단을 연장하는 단계 및 제1 가닥을 주형으로 사용하여 제2 가닥의 3' 말단을 연장하는 단계를 추가로 포함한다. 일부 방법에서, 제1 표적 부위는 제2 표적 부위로부터 적어도 4 bp만큼 떨어져 있다.
- [0011] 일부 방법에서, 제1 뉴클레아제 제제 또는 제2 뉴클레아제 제제 중 적어도 하나는 제1 표적 부위 또는 제2 표적 부위를 표적화하는 Cas9 단백질 및 가이드 RNA (gRNA) (gRNA-Cas 복합체)를 포함한다. gRNA는 규칙적으로 사이 간격을 두고 분포하는 짧은 회문구조 반복 서열 (CRISPR) RNA (crRNA) 및 트랜스-활성화 CRISPR RNA (tracrRNA)를 암호화하는 핵산 서열을 포함할 수 있다. 일부 방법에서, 제1 표적 부위 및 제2 표적 부위 중 적어도 하나는 프로토스페이스 인접 모티프 (PAM) 서열에 의해 플랭킹된다. Cas9 단백질은 RuvC 도메인과 HNH 도메인을 포함할 수 있고, 이들 중 적어도 하나는 엔도뉴클레아제 활성이 결여되어 있다.
- [0012] 일부 방법에서, 제2 핵산은 제1 분해된 핵산의 5' 또는 3' 돌출 서열에 대한 상보적 서열을 포함하지 않고, 단계 (a)는 제1 분해된 핵산 및 제2 분해된 핵산을 조이너 올리고(joiner oligo)와 접촉시키는 단계를 추가로 포함하며, 여기서 조이너 올리고는 (i) 제1 분해된 핵산의 5' 또는 3' 돌출 서열에 대한 제1 상보적 서열; 및 (ii) 제2 분해된 핵산의 5' 또는 3' 돌출 서열에 대한 제2 상보적 서열을 포함한다. 일부 방법에서, 제1 핵산, 제2 핵산 또는 이들 모두는 박테리아 인공 염색체로부터 유래된다. 박테리아 인공 염색체는 인간 DNA, 설치류 DNA, 합성 DNA 또는 이들의 조합을 포함할 수 있다. 박테리아 인공 염색체는 인간 폴리뉴클레오타이드 서열을 포함할 수 있다. 일부 방법에서, 제2 핵산은 박테리아 인공 염색체를 포함한다.
- [0013] 제공된 방법은 또한 둘 이상의 핵산 단편을 조립하는 방법을 포함하며, 이는 (a) 제1 핵산을 적어도 하나의 뉴클레아제 제제와 접촉시켜, 제1 분해된 핵산을 생성하는 단계; (b) 제1 분해된 핵산을 제2 핵산, 조이너 올리고 및 엑소뉴클레아제와 접촉시키는 단계 (여기서, 조이너 올리고는 (i) 제1 분해된 핵산에 상보적인 제1 상보적 서열; (ii) 스페이서; 및 (iii) 제2 핵산에 상보적인 제2 상보적 서열을 포함하고; 엑소뉴클레아제는 제1 상보적 서열 및 제2 상보적 서열을 노출시킨다); 및 (c) 조이너 올리고를 제1 분해된 핵산 및 제2 핵산과 조립하는 단계를 포함한다. 이러한 일부 방법에서, 단계 (c)의 조립하는 단계는 (i) 조이너 올리고의 제1 상보적 서열을 제1 분해된 핵산에, 조이너 올리고의 제2 상보적 서열을 제2 핵산에 어닐링하는 단계; 및 (ii) 조이너 올리고를 제1 분해된 핵산과 제2 핵산에 라이게이션하는 단계를 포함한다.

- [0014] 일부 방법에서, 조이너 올리고의 제1 상보적 서열 및 제2 상보적 서열은 15 내지 120개의 상보적 염기를 포함한다. 일부 방법에서, 조이너 올리고의 스페이서는 비상보적 핵산을 포함한다. 일부 실시 형태에서, 제1 분해된 핵산은 제2 핵산에 끊김 없이(seamlessly) 조립된다.
- [0015] 일부 방법에서, 뉴클레아제 제제는 끊김 없는 조립이 발생할 제1 핵산의 말단으로부터 적어도 20 bp 단편을 절단하도록 설계되며, 여기서 조이너 올리고의 스페이서는 상기 적어도 20 bp 단편과 동일한 서열을 포함하고, 제1 상보적 서열과 상기 적어도 20 bp 단편 사이에 핵산 염기가 부재하며, 제2 상보적 서열과 상기 적어도 20 bp 단편 사이에 핵산 염기가 부재하여, 상기 제1 핵산과 상기 조이너 올리고 및 상기 제2 핵산의 조립은 적어도 20 bp 단편을 재구성하고, 제1 핵산과 제2 핵산을 끊김 없이 조립한다. 일부 방법에서, 동일한 방법이 스페이서 서열로서 제2 핵산으로부터 적어도 20 bp 단편을 사용하여 수행된다. 일부 방법에서, 스페이서는 약 20 bp 내지 약 120 bp를 포함한다. 일부 방법에서, 제2 핵산을 제2 뉴클레아제 제제 및 엑소뉴클레아제와 접촉시키며, 여기서 제2 뉴클레아제 제제는 제2 핵산을 절단하여, 조이너 올리고의 제2 상보적 서열과 상보적인 뉴클레오티드 서열을 포함하는 제2 분해된 핵산을 생성하고, 제1 분해된 핵산은 제2 분해된 핵산에 조립된다. 일부 방법에서, 제2 핵산을 제한 효소 또는 메가뉴클레아제(meganuclease) 및 엑소뉴클레아제와 접촉시키며, 여기서, 제한 효소 또는 메가뉴클레아제는 제2 핵산을 절단하여 올리고의 제2 상보적 서열과 상보적인 뉴클레오티드 서열을 포함하는 제2 분해된 핵산을 생성하고, 제1 분해된 핵산은 제2 분해된 핵산에 조립된다. 일부 방법에서, 제1 분해된 핵산 및/또는 제2 분해된 핵산의 3' 말단은 단계 (b)에서 연장된다. 조이너 올리고는 상기 제1 핵산 및 상기 제2 핵산에 동일한 반응에서 또는 순차적으로 조립될 수 있다. 일부 방법에서, 제1 핵산, 제2 핵산 또는 이들 모두는 박테리아 인공 염색체로부터 유래되고/되거나 적어도 10 kb이고/이거나, 인간 DNA, 설치류 DNA, 합성 DNA 또는 이들의 조합을 포함한다.
- [0016] 일부 방법에서, 적어도 하나의 뉴클레아제 제제 또는 제2 뉴클레아제 제제는 제1 표적 부위 또는 제2 표적 부위를 표적화하는 Cas 단백질 및 가이드 RNA (gRNA) (gRNA-Cas 복합체)를 포함한다. 예를 들어, Cas 단백질은 Cas9 단백질일 수 있다. Cas9 단백질은 RuvC 도메인과 HNH 도메인을 포함할 수 있고, 이들 중 적어도 하나에는 엔도뉴클레아제 활성이 결여되어 있다. 일부 실시 형태에서, gRNA는 규칙적으로 사이 간격을 두고 분포하는 짧은 회문구조 반복 서열 (CRISPR) RNA (crRNA) 및 트랜스-활성화 CRISPR RNA (tracrRNA)를 암호화하는 핵산 서열을 포함한다. 제1 표적 부위 및/또는 제2 표적 부위는 프로토스페이서 인접 모티프 (PAM) 서열에 의해 플랭킹될 수 있다. 일부 방법에서, 적어도 하나의 뉴클레아제 제제 및/또는 제2 뉴클레아제 제제는 징크 핑거 뉴클레아제 또는 전사 활성인자-유사 이펙터 뉴클레아제 (탈렌)를 포함한다.
- [0017] 일부 실시 형태에서, 조이너 올리고는 g블록(gBlock)을 포함한다. 일부 이러한 방법에서, g블록은 선택 카세트(selection cassette)를 포함하지 않는다.
- [0018] 둘 이상의 핵산을 조립하는 방법이 또한 제공되며, 이는 (a) 제1 핵산을 적어도 하나의 뉴클레아제 제제와 접촉시켜, 제1 분해된 핵산을 생성하는 단계; (b) 제2 핵산을 제2 뉴클레아제 제제와 접촉시켜, 제2 분해된 핵산을 생성하는 단계; (c) 제1 분해된 핵산 및 제2 분해된 핵산을 조이너 올리고 및 엑소뉴클레아제와 접촉시키는 단계 (여기서, 조이너 올리고는 (i) 제1 분해된 핵산에 상보적인 제1 상보적 서열; (ii) 스페이서; 및 (iii) 제2 분해된 핵산에 상보적인 제2 상보적 서열을 포함하고; 엑소뉴클레아제는 제1 상보적 서열 및 제2 상보적 서열을 노출시킨다); 및 (d) 조이너 올리고를 제1 분해된 핵산 및 제2 핵산과 조립하는 단계를 포함한다.
- [0019] 중첩 서열을 갖는 핵산을 조립하는 방법이 본 명세서에서 제공된다. 이러한 방법은 2개 이상의 핵산 단편을 조립하는 방법을 포함하며, 이는 (a) 중첩 서열을 포함하는 제1 핵산과 제2 핵산을 적어도 하나의 gRNA-Cas 복합체 및 엑소뉴클레아제와 접촉시킴으로써, 이들의 말단 중 하나에 상보적 서열을 포함하는 2개의 분해된 핵산 단편을 생성하는 단계; (b) 단계 (a)로부터 생성된 2개의 핵산 단편을 조립하는 단계를 포함한다. 일부 방법에서, 적어도 하나의 gRNA-Cas 복합체는 제1 핵산을 제1 표적 부위에서 절단하여, 제1 분해된 핵산과 제2 핵산 사이에 상보적 말단 서열을 포함하는 제1 분해된 핵산을 생성한다. 소정 방법에서, 단계 (b)는 (i) 노출된 상보적 서열을 어닐링하는 단계; (ii) 어닐링된 상보적 서열의 3' 말단을 연장하는 단계; 및 (iii) 제1 핵산과 제2 핵산을 라이게이션하는 단계를 추가로 포함한다. 일부 방법에서, 단계 (a)는 제2 핵산을 제2 gRNA-Cas 복합체와 접촉시키는 단계를 추가로 포함하며, 여기서, 제2 핵산은 중첩 말단 서열을 포함하지 않고, 제2 gRNA-Cas 복합체는 제2 핵산을 절단하여 제1 분해된 핵산과 제2 분해된 핵산 사이에 중첩 말단 서열을 포함하는 제2 분해된 핵산을 생성한다. 예를 들어, gRNA-Cas 복합체는 Cas9 단백질을 포함한다. Cas9 단백질은 RuvC 도메인과 HNH 도메인을 포함할 수 있고, 이들 중 적어도 하나에는 엔도뉴클레아제 활성이 결여되어 있다. 일부 방법에서, 중첩 서열은 20 bp 내지 200 bp 범위의 길이이다. 청구항 1항 내지 제7항 중 어느 하나의 방법에서, 제1 핵산, 제2 핵산 또는 이들 모두는 박테리아 인공 염색체로부터 유래된다. 일부 방법에서, 박테리아 인공 염색



체는 인간 DNA, 설치류 DNA, 합성 DNA 또는 이들의 조합을 포함한다. 박테리아 인공 염색체는 인간 서열을 포함할 수 있다.

[0020] 제공된 방법은 또한 둘 이상의 핵산 단편을 조립하는 방법을 포함하며, 이는 (a) 제1 핵산과 제2 핵산을 적어도 하나의 gRNA-Cas 복합체에 노출시켜, 이들의 말단 중 하나에 5' 또는 3' 돌출 서열을 포함하는 제1 분해된 핵산과 제2 분해된 핵산을 생성하는 단계; (b) 단계 (a)로부터 생성된 2개의 핵산 단편을 조립하는 단계를 포함한다. 일부 방법에서, 조립하는 단계 (b)는 (i) 5' 및 3' 돌출 서열을 어닐링하는 단계; 및 (ii) 제1 분해된 핵산과 제2 분해된 핵산을 라이게이션하는 단계를 포함한다. 일부 방법에서, 5' 및/또는 3' 돌출 서열은 적어도 4개의 상보적 염기를 포함한다. 일부 방법에서, 단계 (b)는 제1 분해된 핵산과 제2 분해된 핵산의 3' 말단을 연장하는 단계를 추가로 포함한다. 일부 방법에서, 제2 핵산은 제1 분해된 핵산의 5' 또는 3' 돌출 서열에 대한 상보적 서열을 포함하지 않고, 단계 (a)는 제1 분해된 핵산 및 제2 분해된 핵산을 조이너 올리고와 접촉시키는 단계를 추가로 포함하며, 여기서 조이너 올리고는 (i) 제1 분해된 핵산의 5' 또는 3' 돌출 서열에 대한 제1 상보적 서열; 및 (ii) 제2 분해된 핵산의 5' 또는 3' 돌출 서열에 대한 제2 상보적 서열을 포함한다. 일부 방법에서, gRNA-Cas 단백질 복합체는 RuvC 도메인과 HNH 도메인 (이들 중 하나는 엔도뉴클레아제 활성이 결여되어 있다)을 포함하는 Cas9 단백질을 포함한다. 일부 방법에서, gRNA-Cas 복합체는 crRNA, tracrRNA 및 Cas 단백질로서 별도로 제공된다. 일부 방법에서, 제1 핵산 및 제2 핵산은 프로토스페이서 인접 모티프 (PAM) 서열을 포함한다. 일부 방법에서, 제1 핵산, 제2 핵산 또는 이들 모두는 박테리아 인공 염색체로부터 유래된다. 일부 방법에서, 박테리아 인공 염색체는 인간 DNA, 설치류 DNA, 합성 DNA 또는 이들의 조합을 포함한다. 예를 들어, 박테리아 인공 염색체는 인간 폴리뉴클레오티드 서열을 포함할 수 있다.

[0021] 둘 이상의 핵산을 조립하는 방법이 또한 제공되며, 이는 (a) 제1 핵산을 적어도 하나의 gRNA-Cas 복합체와 접촉시켜, 제1 분해된 핵산을 생성하는 단계; (b) 제1 분해된 핵산을 제2 핵산, 조이너 올리고 및 엑소뉴클레아제와 접촉시키는 단계 (여기서, 조이너 올리고는 (i) 제1 분해된 핵산에 상보적인 제1 상보적 서열; (ii) 스페이서; 및 (iii) 제2 핵산에 상보적인 제2 상보적 서열을 포함하고; 엑소뉴클레아제는 제1 상보적 서열 및 제2 상보적 서열을 노출시킨다); 및 (c) 조이너 올리고를 제1 분해된 핵산 및 제2 핵산과 조립하는 단계를 포함한다. 일부 방법에서, 조립하는 단계 (c)는 (i) 조이너 올리고의 제1 상보적 서열을 제1 분해된 핵산에, 조이너 올리고의 제2 상보적 서열을 제2 핵산에 어닐링하는 단계; 및 (ii) 조이너 올리고를 제1 분해된 핵산과 제2 핵산에 라이게이션하는 단계를 포함한다. 일부 방법에서, 조이너 올리고의 제1 상보적 서열 및 제2 상보적 서열은 15 내지 120개의 상보적 염기를 포함한다. 일부 방법에서, 조이너 올리고의 스페이서는 비상보적 핵산을 포함한다.

[0022] 조이너 올리고를 사용하여, 제1 분해된 핵산은 제2 핵산에 끊김 없이 조립될 수 있다. 일부 방법에서, gRNA-Cas 복합체는 끊김 없는 조립이 발생할 제1 핵산의 말단으로부터 적어도 20 bp 단편을 절단하도록 설계되며, 여기서 조이너 올리고의 스페이서는 상기 적어도 20 bp 단편과 동일한 서열을 포함하고, 제1 상보적 서열과 상기 적어도 20 bp 단편 사이에 핵산 염기가 부재하며, 제2 상보적 서열과 상기 적어도 20 bp 단편 사이에 핵산 염기가 부재하여, 상기 제1 핵산과 상기 조이너 올리고 및 상기 제2 핵산의 조립은 적어도 20 bp 단편을 재구성하고, 제1 핵산과 제2 핵산을 끊김 없이 조립한다. 일부 방법에서, 동일한 방법이 스페이서 서열로서 제2 핵산으로부터 적어도 20 bp 단편을 사용하여 수행된다. 일부 방법에서, 스페이서는 약 20 bp 내지 약 120 bp를 포함한다. 일부 방법에서, 제2 핵산을 제2 gRNA-Cas 복합체 및 엑소뉴클레아제와 접촉시키며, 여기서 제2 gRNA-Cas 복합체는 제2 핵산을 절단하여, 조이너 올리고의 제2 상보적 서열과 상보적인 뉴클레오티드 서열을 포함하는 제2 분해된 핵산을 생성하고, 제1 분해된 핵산은 제2 분해된 핵산에 조립된다. 일부 방법에서, 제2 핵산을 제한 효소 또는 메가뉴클레아제 및 엑소뉴클레아제와 접촉시키며, 여기서, 제한 효소 또는 메가뉴클레아제는 제2 핵산을 절단하여 조이너 올리고의 제2 상보적 서열과 상보적인 뉴클레오티드 서열을 포함하는 제2 분해된 핵산을 생성하고, 제1 분해된 핵산은 제2 분해된 핵산에 조립된다. 일부 방법에서, 제1 분해된 핵산 및/또는 제2 분해된 핵산의 3' 말단은 단계 (b)에서 연장된다. 조이너 올리고는 상기 제1 핵산 및 상기 제2 핵산에 동일한 반응에서 또는 순차적으로 조립될 수 있다. 일부 방법에서, gRNA-Cas 복합체는 Cas9 단백질을 포함한다. 일부 방법에서, 제1 핵산, 제2 핵산 또는 이들 모두는 박테리아 인공 염색체로부터 유래되고/되거나 적어도 10 kb이고/이거나, 인간 DNA, 설치류 DNA, 합성 DNA 또는 이들의 조합을 포함한다.

### 도면의 간단한 설명

[0023] 도 1은 BAC에 특이적이게 설계된 중첩을 갖는 PCR 생성물로의 BAC의 조립을 도시한다. PCR에 의해 50 bp 중첩을 HYG 카세트에 첨가하였다.

도 2는 각각의 BAC 상의 2개의 Cas9 표적 부위를 사용하여 중첩 서열을 갖는 2개의 BAC의 조립을 도시한다. 조

립 과정은 본 명세서에 개시된 방법을 사용하여 2일이 걸렸다.

도 3은 종래의 방법을 사용하여 중첩 서열을 갖는 2개의 BAC의 조립을 도시한다. 조립 과정은 종래의 방법을 사용하여 4주 걸렸다.

도 4는 Cas9/등온 조립 방법의 클로닝 효율 및 BAC 클로닝 단계에 필요한 시간을 나타낸다.

도 5는 CRISPR/Cas9 시스템 및 등온 조립을 사용하여 큰 표적화 벡터 (LTVEC)의 구축을 도시한다. CRISPR/Cas9로 절단된 DNA 단편은 하나 이상의 조이너 올리고 및 등온 조립을 사용하여 끊임 없이 조립된다.

도 6은 Cas9 절단 후 핵산을 끊임 없이 조립하기 위한 링커(linker) (조이너 올리고)를 사용하기 위한 전략을 나타낸다. gRNA/Cas9 복합체는 관심 영역의 5' 상류에 위치한 표적 부위 (화살표)를 절단하여, 제1 Cas9-분해된 DNA 단편 (5' DNA)을 생성하도록 설계된다. 이어서, 5' DNA의 결실 부분 (슬래시된 박스)은 조이너 올리고의 5' 와 3' 중첩 서열 사이의 스페이서로 사용된다. 3개의 성분은 등온 조립 반응에서 조립된다: (a) 제1 Cas9-분해된 DNA 단편 (5' DNA); (b) 조이너 올리고; 및 (c) 제2 DNA 단편 (3' DNA). 조이너 올리고는 5' 에서 3' 까지 (1) 5' DNA와의 중첩 서열, (2) 제1 분해된 단편의 결실 부분을 함유하는 스페이서 및 (3) 3' DNA와의 중첩 서열을 포함한다. 5' DNA의 결실 부분은 조립 단계 동안 재구성된다.

도 7은 CRISPR/Cas9 시스템 및 등온 조립을 사용하는 DNA 벡터의 구축을 나타낸다.

도 8은 CRISPR/Cas9 시스템 및 등온 조립을 사용하는 큰 표적화 벡터의 구축을 나타낸다.

도 9는 등온 조립 및 2개의 링커 (조이너 올리고)를 사용하여 BAC 벡터의 일부를 카세트로 대체하기 위한 표적화 벡터의 구축을 나타낸다. 단편 또는 링커에 대한 mBAC의 다양한 비의 결과가 패널 #1, #2, #3 및 #4에 제시된다.

도 10은 2개의 링커를 사용하는 mBAC (BAC ID: RP23-399M19)와 카세트 사이의 조립 반응의 양 접합부에 걸친 끊임 없는 조립의 서열 확인을 나타낸다.

도 11은 Cas9 및 등온 조립을 사용하는 2개의 mBAC의 조립을 도시한다. bMQ50f19 벡터와 히그로마이신 저항성 유전자 유비퀴틴 프로모터를 포함하는 카세트 사이의 조립은 끊임 없이 없었다.

도 12는 링커 1에서의 끊임 없는 조립의 서열 확인 및 링커 2과 링커 3에서 의도적으로 끊기는 조립의 서열 확인을 나타낸다.

도 13은 4개의 링커 및 등온 조립을 사용하는 큰 인간 유전자 단편의 mBAC로의 삽입을 도시한다. Cas9는 hBAC1로부터의 hGene 단편 A, hBAC2로부터의 hGene 단편 B 및 mBAC를 절단하여, mGene 단편을 제거하였다.

도 14는 Cas9 및 등온 조립을 사용하는 인간 서열의 BAC 벡터로의 삽입을 도시한다.

도 15는 Cas9 및 등온 조립을 사용하는 메가뉴클레아제 부위를 포함하는 g블록의 삽입을 도시한다. 도 15a는 PI-SceI 부위를 포함하는 g블록의 삽입을 도시하고; 도 15b는 MauBI 부위를 포함하는 g블록의 삽입을 도시한다.

도 16은 3개의 조이너 올리고, Cas9 및 등온 조립을 사용하는 표적화 벡터의 직접 인간화(direct humanization)의 예를 도시한다.

도 17은 업 앤드 다운(up and down) 조이너 올리고를 갖는 공여자(donor), Cas9 및 등온 조립을 사용하는 표적화 벡터의 간접 인간화의 예를 도시한다.

도 18은 Cas9 및 등온 조립을 사용하여 점 돌연변이를 도입한 예를 도시한다.

도 19는 Cas9 및 등온 조립에 의한 BAC 트리밍의 예를 도시한다. 이러한 예에서, 트리밍은 Ori 서열을 제거한다. Ori 서열은 2개의 조이너 올리고 및 등온 조립을 사용하여 벡터에 재삽입된다.

## 발명을 실시하기 위한 구체적인 내용

### I. 정의

본 명세서에서 상호 교환적으로 사용되는 용어 "단백질", "폴리펩티드", 및 "펩티드"는 암호화 및 비암호화된 아미노산, 및 화학적으로 또는 생화학적으로 변형되거나 유도체화된 아미노산을 비롯하여, 임의의 길이의 아미노산의 폴리머 형태를 포함한다. 상기 용어는 또한 변형된 폴리머, 예컨대 변형된 펩티드 골격을 갖는 폴리펩티드를 포함한다.



- [0026] 본 명세서에서 상호 교환적으로 사용되는 용어 "핵산" 및 "폴리뉴클레오티드"는 리보뉴클레오티드, 데옥시리보뉴클레오티드, 또는 이들의 유사체 또는 변형된 형태를 비롯한 임의의 길이의 뉴클레오티드의 폴리머 형태를 포함한다. 이들은 단일 가닥, 이중 가닥, 및 다중 가닥 DNA 또는 RNA, 게놈 DNA, cDNA, DNA-RNA 하이브리드, 및 퓨린 염기, 피리미딘 염기, 또는 다른 천연, 화학적으로 변형된, 생화학적으로 변형된, 비천연 또는 유도체화된 뉴클레오티드 염기를 포함하는 폴리머를 포함한다.
- [0027] "코돈 최적화"는 일반적으로 천연 아미노산 서열을 유지하면서 천연 서열의 적어도 하나의 코돈을 숙주 세포의 유전자에서 더욱 빈번하게 또는 가장 빈번하게 사용되는 코돈으로 치환함으로써 특정 숙주 세포에서의 발현 향상을 위해 핵산 서열을 변형시키는 방법을 포함한다. 예를 들어, Cas 단백질을 암호화하는 핵산은 자연 발생 핵산 서열과 비교하여, 박테리아 세포, 효모 세포, 인간 세포, 비인간 세포, 포유류 세포, 설치류 세포, 마우스 세포, 래트 세포, 햄스터 세포, 또는 임의의 다른 숙주 세포를 비롯한 주어진 원핵 또는 진핵 세포에서의 사용 빈도가 높은 코돈으로 치환되도록 변형될 수 있다. 코돈 사용 빈도 표는 예를 들어, "코돈 사용 빈도 데이터베이스(Codon Usage Database)"에서 용이하게 이용할 수 있다. 이러한 표는 다양한 방법으로 조정될 수 있다. 문헌[Nakamura *et al.* (2000) *Nucleic Acids Research* 28:292]을 참조한다. 특정 숙주 세포에서의 발현을 위해 특정 서열의 코돈 최적화를 위한 컴퓨터 알고리즘도 이용 가능하다 (예를 들어, 문헌[Gene Forge]을 참조한다).
- [0028] "작동 가능한 연결" 또는 "작동 가능하게 연결된"은 2개 이상의 성분 (예를 들어, 프로모터 및 다른 서열 요소)의 병렬 배치(juxtaposition)를 포함하는데, 두 성분이 정상적으로 기능하고, 이들 성분 중에서 적어도 하나가 적어도 하나의 기타 성분에 미치는 기능을 매개할 수 있는 가능성을 허용하도록 한다. 예를 들어, 프로모터는 프로모터가 하나 이상의 전사 조절 인자의 유무에 대응하여 암호화 서열의 전사 레벨을 제어하는 경우에, 암호화 서열에 작동 가능하게 연결될 수 있다.
- [0029] 핵산의 "상보성"은 하나의 핵산 가닥의 뉴클레오티드 서열이 이의 핵산 염기 그룹의 배향으로 인해, 대향하는 핵산 가닥 상의 다른 서열과 수소 결합을 형성한다는 것을 의미한다. DNA의 상보적 염기는 전형적으로 A와 T, C와 G이다. RNA에서, 이들은 전형적으로 C와 G, U와 A이다. 상보성은 완전하거나 상당하거나/충분할 수 있다. 2개의 핵산 사이의 완전한 상보성은 2개의 핵산이, 이중나선 구조의 모든 염기가 왓슨-크릭 염기 결합(Watson-Crick pairing)에 의해 상보적 염기에 결합되는 이중 나선 구조를 형성할 수 있음을 의미한다. "상당한" 또는 "충분한" 상보성은 한 가닥의 서열이 대향하는 가닥의 서열에 전적으로 및/또는 완전히 상보적이지 않지만, 일련의 혼성화 조건 (예를 들어, 염 농도 및 온도)에서 안정한 하이브리드 복합체를 형성하도록 두 가닥 상의 염기 사이에 충분한 결합이 일어남을 의미한다. 이러한 조건은 혼성화된 가닥의 T<sub>m</sub>을 예측하는 서열 및 표준 수학적 계산을 이용하여 예측될 수 있거나, 일상적인(routine) 방법을 이용하여 T<sub>m</sub>의 경험적 결정에 의해 예측될 수 있다. T<sub>m</sub>은 2개의 핵산 가닥 사이에 형성된 혼성화 복합체 집단이 50% 변성된 온도를 포함한다. T<sub>m</sub>보다 낮은 온도에서는 혼성화 복합체의 형성이 촉진되는 반면에, T<sub>m</sub>보다 높은 온도에서는 혼성화 복합체의 가닥들의 용융 또는 분리가 촉진된다. 기타 공지된 T<sub>m</sub> 계산은 핵산 구조적 특성을 고려하지만, T<sub>m</sub>은 예를 들어, T<sub>m</sub>=81.5+0.41(% G+C)을 사용하여 1 M NaCl 수용액 중의 기지의 G+C 함량을 갖는 핵산에 대하여 추정될 수 있다.
- [0030] "혼성화 조건"은 하나의 핵산 가닥이 상보적 가닥 상호작용 및 수소 결합에 의해 제2 핵산 가닥에 결합하여 혼성화 복합체를 생산하는 누적 환경을 포함한다. 이러한 조건은 핵산을 함유하는 수용액 또는 유기 용액의 화학 성분 및 이의 농도 (예를 들어, 염, 킬레이트제, 포름아미드), 및 혼합물의 온도를 포함한다. 다른 인자, 예컨대 배양 기간 또는 반응실 치수가 환경에 기여할 수 있다. 예를 들어, 문헌[Sambrook *et al.*, Molecular Cloning, A Laboratory Manual, 2.sup.nd ed., pp. 1.90-1.91, 9.47-9.51, 11.47-11.57 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989)]을 참조한다.
- [0031] 혼성화는 염기 간의 미스매치(mismatch)가 가능하더라도, 2개의 핵산이 상보성 서열을 포함하는 것을 필요로 한다. 2개의 핵산 간의 혼성화에 적합한 조건은 본 기술 분야에 잘 알려진 변수인 핵산 길이 및 상보성 정도에 따라 다르다. 2개의 뉴클레오티드 서열 간의 상보성 정도가 클수록, 이들 서열을 갖는 핵산의 하이브리드에 대한 용융 온도(T<sub>m</sub>)의 값이 커진다. 짧은 스트레치(stretch)의 상보성 (예를 들어, 35개 이하, 30개 이하, 25개 이하, 22개 이하, 20개 이하 또는 18개 이하의 뉴클레오티드에 대한 상보성)을 갖는 핵산 간의 혼성화의 경우, 미스매치 위치가 중요해진다 (문헌[Sambrook *et al.*, supra, 11.7-11.8]을 참조한다). 전형적으로, 혼성화 가능한 핵산 길이는 적어도 약 10개의 뉴클레오티드로 되어 있다. 혼성화 가능한 핵산의 최소 길이의 예로는 적어도 약 15개의 뉴클레오티드, 적어도 약 20개의 뉴클레오티드, 적어도 약 22개의 뉴클레오티드, 적어도 약 25개의 뉴클레오티드 및 적어도 약 30개의 뉴클레오티드를 포함한다. 게다가, 온도 및 세정액 염 농도는 필요에

따라, 인자, 예컨대 상보성 영역의 길이 및 상보성 정도에 따라 조절될 수 있다.

[0032] 폴리뉴클레오티드의 서열은 특이적으로 혼성화 가능하게 되는 이의 표적 핵산의 서열에 대하여 100% 상보성을 나타낼 필요는 없다. 게다가, 폴리뉴클레오티드는 개재 또는 인접 세그먼트가 혼성화 이벤트 (예를 들어, 루프 구조 또는 헤어핀 구조)에 관여하지 않도록 하나 이상의 세그먼트에 대하여 혼성화할 수 있다. 폴리뉴클레오티드 (예를 들어, gRNA)는 이들이 표적화하는 표적 핵산 서열 내의 표적 부위에 대하여, 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 99% 또는 100% 서열 상보성을 포함할 수 있다. 예를 들어, 20개 중 18개의 뉴클레오티드가 표적 부위에 상보적이므로, 특이적으로 혼성화되는 gRNA는 90% 상보성을 나타낼 것이다. 이러한 예에서, 나머지 비상보적 뉴클레오티드는 상보적 뉴클레오티드와 클러스터를 이루거나 그 사이에 상보적 뉴클레오티드가 배치될 수 있으며, 서로 인접하거나 상보적 뉴클레오티드에 인접할 필요는 없다.

[0033] 핵산 내의 핵산 서열의 특정 스트레치 간의 상보성 비율(percent complementarity)은 통상, 본 기술 분야에 알려진 BLAST 프로그램 (기본 국소 정렬 검색 도구(basic local alignment search tools)) 및 PowerBLAST 프로그램 (문헌[Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410]; 문헌[Zhang and Madden (1997) *Genome Res.* 7:649-656])을 사용하거나, 스미스(Smith) 및 와트만(Waterman)의 알고리즘 (문헌[Adv. Appl. Math., 1981, 2, 482-489])을 사용하는 디폴트 설정을 이용한 Gap 프로그램 (위스콘신 서열 분석 패키지, 유닉스 용 버전 8, 제네틱스 컴퓨터 그룹(Genetics Computer Group), 유니버시티 리서치 파크(University Research Park), 위스콘신 주 메디슨 소재)을 사용하여 측정될 수 있다.

[0034] 본 명세서에 제공된 방법 및 조성물은 다양한 상이한 성분을 사용한다. 발명의 상세한 설명을 통해, 일부 성분이 활성 변이체 및 단편을 가질 수 있는 것으로 인식된다. 이러한 성분은 예를 들어, Cas 단백질, CRISPR RNA, tracrRNA, 및 가이드 RNA를 포함한다. 각각의 이들 성분에 대한 생물학적 활성은 본 명세서의 다른 곳에 기재되어 있다.

[0035] 2개의 폴리뉴클레오티드 또는 폴리펩티드 서열과 관련하여, "서열 동일성" 또는 "동일성"은 특정 비교창에 대하여, 최대 일치도로 정렬되는 경우에 동일한 2개의 서열의 잔기를 참조한다. 서열 동일성 비율이 단백질에 관하여 사용되는 경우에, 동일하지 않은 잔기 위치가 종종 보존 아미노산 치환에 따라 다르며, 여기서 아미노산 잔기가 유사한 화학적 특성 (예를 들어, 전하 또는 소수성)을 갖는 다른 아미노산 잔기 대신에 사용되므로, 분자의 기능 특성을 변경시키지 않음이 인지된다. 서열이 보존적 치환에 있어서 상이한 경우에는, 서열 동일성 비율은 치환의 보존적 성질을 보정하기 위해 상향 조정될 수 있다. 이러한 보존적 치환에 따라 다른 서열은 "서열 유사성" 또는 "유사성"을 갖는다고 한다. 이러한 조정을 행하는 수단은 당업자에게 잘 알려져 있다. 전형적으로, 이는 보존적 치환을 완전 미스매치보다는 부분 미스매치로서 스코어를 주어(scoring), 서열 동일성 비율을 증가시키는 것을 포함한다. 따라서, 예를 들어, 동일한 아미노산은 스코어 1로 주어지고, 비보존적 치환은 스코어 0으로 주어지며, 보존적 치환은 스코어 0 내지 1로 주어진다. 보존적 치환의 스코어는 예를 들어, 프로그램 PC/GENE(캘리포니아주 마운틴 뷰 소재의 인텔리제네틱스(Intelligenetics))으로 실행되는 바와 같이 계산된다.

[0036] "서열 동일성 비율"은 비교창에 대하여 최적으로 정렬된 2개의 서열을 비교함으로써 결정된 값을 포함하며, 여기서 비교창 내의 폴리뉴클레오티드 서열의 부분은 2개의 서열의 최적 정렬을 위한 참조 서열 (부가 또는 결실을 포함하지 않음)과 비교하여, 부가 또는 결실 (즉, 간극)을 포함할 수 있다. 동일한 핵산 염기 또는 아미노산 잔기가 두 서열에 나타나는 위치의 개수를 결정하여 매칭된 위치의 개수를 산출하고, 매칭된 위치의 개수를 비교창 내의 위치의 총 개수로 나누고, 그 결과에 100을 곱하여 서열 동일성 비율을 산출함으로써 그 비율을 계산한다.

[0037] 달리 명시되지 않는 한, 서열 동일성/유사성 값은 하기 파라미터를 사용하여 GAP 버전 10을 이용하여 구한 값을 포함한다: GAP 중량(Weight) 50 및 길이 중량(Length Weight) 3, 및 nwsgapdna.cmp 스코어링 매트릭스(scoring matrix)를 이용한 뉴클레오티드 서열의 % 동일성 및 % 유사성; GAP 중량 8 및 길이 중량 2, 및 BLOSUM62 스코어링 매트릭스를 이용한 아미노산 서열의 % 동일성 및 % 유사성; 또는 이의 임의의 등가 프로그램. "등가 프로그램"은 임의의 서열 비교 프로그램을 포함하는데, 당해 임의의 2개의 서열의 경우, GAP 버전 10에 의해 얻은 대응 정렬과 비교하여, 동일한 뉴클레오티드 또는 아미노산 잔기 매치 및 동일한 % 서열 동일성을 갖는 정렬을 얻는다.

[0038] 하나 이상의 열거된 요소를 "포함하는(comprising 또는 including)" 조성물 또는 방법은 구체적으로 열거되지 않은 다른 요소를 포함할 수 있다. 예를 들어, 단백질을 "포함하는" 조성물은 단백질을 단독으로 포함하거나, 다른 성분과 조합하여 포함할 수 있다.

- [0039] 다양한 값의 지정은 해당 범위 내 또는 그 범위를 한정하는 모든 정수, 및 그 범위 내의 정수로 한정되는 모든 부분적인 범위를 포함한다.
- [0040] 문맥상 달리 명백하지 않는 한, 용어 "약"은 표시된 값의 측정 표준 오차 (예를 들어, SEM) 내의 값을 포함한다.
- [0041] 문맥상 달리 명확히 지시하지 않는 한, 관사의 단수형 ("a", "an," 및 "the")은 복수형 지시 대상(plural reference)을 포함한다. 예를 들어, 용어 "Cas 단백질" 또는 "적어도 하나의 Cas 단백질"은 이들의 혼합물을 비롯하여, 복수의 Cas 단백질을 포함할 수 있다.
- [0042] II. 일반 사항
- [0043] 핵산을 조립하는 종래의 방법은 제한 효소를 사용하는 통상적인 효소적 분해, 핵산의 클로닝 및 핵산을 함께 라이게이션하는 단계의 시간 소모적인 단계를 사용한다 (종래의 방법 및 타임라인의 예시에 대한 도 3 및 도 4를 참조한다). 이러한 방법은 큰 단편 또는 벡터를 함께 조립하는 경우 더욱 어려워진다. 본 명세서에 제공된 방법은 뉴클레아제의 가단성 표적 특이성 (예를 들어, 가이드 RNA 및 Cas9 뉴클레아제)을 이용하여, 핵산을 신속한 조립 반응에 사용하기 적합한 형태로 전환한다.
- [0044] 예컨대 가이드 RNA (gRNA)에 의해 특정 표적 부위로 지정되는 뉴클레아제 제제 (예를 들어, 가이드 RNA (gRNA)에 의해 특정 표적 부위로 지정되는 Cas 단백질을) 사용하여 2개 이상의 핵산을 조립하는 방법이 본 명세서에 제공된다. 부위-지정(site-directed) 뉴클레아제 제제, 예를 들어, 가이드 RNA-지정 Cas 단백질은 그들의 엔도 뉴클레아제 활성에 의해 생성된 말단 서열을 선택하고 조합함으로써 핵산의 신속하고 효율적인 조합이 가능하게 한다. 본 명세서에 제공된 방법은 제1 폴리뉴클레오티드를 원하는 표적 부위에 특이적인 뉴클레아제 제제 (예를 들어, gRNA-Cas 복합체) 및 엑소뉴클레아제와 조합한다. 표적 부위는 뉴클레아제가 핵산을 절단할 때, 절단에 의해 생성된 결과로서 얻어진 말단이 제2 핵산의 말단에 상보적인 영역 (예를 들어, 중첩하는 말단)을 갖도록 선택될 수 있다. 이어서, 이러한 상보적 말단은 조립되어, 단일의 조립된 핵산을 생성한다. 뉴클레아제 제제 (예를 들어, gRNA-Cas 복합체)는 각각의 표적 부위에 특이적이므로, 본 발명의 방법은 정확한 부위-지정 방식으로 핵산을 변형할 수 있게 한다. 본 발명의 방법은 또한 뉴클레아제 절단에 의해 생성된 중첩하는 핵산 말단을 조립하기 위해 특수하게 설계되거나, 조립 반응을 위해 설계되고 합성된, 신속하고 효율적인 조립 방법을 이용함으로써 뉴클레아제 제제, 예를 들어, gRNA-Cas 복합체 특이성을 이용한다. 예를 들어, 절단 시에, 제2 핵산의 말단 서열에 상보적인 말단 서열이 생성되도록 표적 부위 특이적인 뉴클레아제 제제 (예를 들어, gRNA-Cas 복합체)를 선택함으로써, 등은 조립을 사용하여 결과로서 얻어진 분해된 핵산을 조립할 수 있다. 따라서, 중첩 말단 서열을 만드는 핵산 및 뉴클레아제 제제 (예를 들어, gRNA-Cas 복합체)를 선택함으로써, 핵산은 신속하고 조합적인 방법으로 조합되어, 최종 조립된 핵산을 신속하고 효율적인 방식으로 생성할 수 있다. 대안적으로, 상보적 말단을 포함하지 않는 핵산은 각각의 핵산에 상보적 말단을 가지도록 설계된 조이너 올리고와 조립될 수 있다. 조이너 올리고를 사용함으로써, 둘 이상의 핵산은 끊임 없이 조립됨으로써, 결과로서 얻어진 조립된 핵산에서 불필요한 서열을 감소시킬 수 있다.
- [0045] III. 뉴클레아제 제제
- [0046] 본 방법은 폴리뉴클레오티드의 부위-지정 절단을 위한 뉴클레아제 제제를 사용한다. 구체적으로, 확인된 표적 부위에서의 폴리뉴클레오티드의 엔도뉴클레아제 절단은, 이어서 제2 폴리뉴클레오티드와 결합하여, 둘 이상의 폴리뉴클레오티드를 부위 특이적인 방식으로 조립할 수 있는 말단을 포함하는 분해된 폴리뉴클레오티드를 생성한다.
- [0047] "뉴클레아제 제제"는 DNA 절단을 위한 활성을 갖는 분자를 포함한다. 본 명세서에 개시된 방법에 사용하기 위한 뉴클레아제 제제의 특정 예는 RNA-유도 CRISPR-Cas9 시스템, 징크 핑거 단백질, 메가뉴클레아제, TAL 도메인, 탈렌, 효모 조립체, 재조합효소, 류신 지퍼(leucine zipper), CRISPR/Cas, 엔도뉴클레아제 및 본 기술 분야의 숙련인에게 알려진 다른 뉴클레아제 제제를 포함한다. 뉴클레아제 제제는 주어진 표적 부위에서 절단에 있어서의 특이성을 위해 선택되거나 설계될 수 있다. 예를 들어, 뉴클레아제 제제는 절단된 폴리뉴클레오티드와 상이한 폴리뉴클레오티드 사이에 중첩하는 말단을 생성하는 표적 부위에서의 절단을 위해 선택될 수 있다. CRISPR-Cas9에서와 같이 단백질 및 RNA 요소를 모두 갖는 뉴클레아제 제제는 뉴클레아제 제제로써 이미 복합체를 형성한 제제를 공급받을 수 있거나, 별개의 단백질 및 RNA 요소를 공급받을 수 있으며, 이러한 경우에 이들은 본 명세서에 기재된 반응 혼합물에서 복합체를 형성하여 뉴클레아제 제제를 형성한다.
- [0048] 용어 "뉴클레아제 제제에 대한 인식 부위"는 틸 또는 이중 가닥 절단이 뉴클레아제 제제에 의해 유도되는 DNA

서열을 포함한다. 뉴클레아제 제제에 대한 인식 부위는 세포에 내인성 (또는 고유한 것)일 수 있거나, 인식 부위는 세포에 외인성일 수 있다. 구체적인 실시 형태에서, 인식 부위는 세포에 외인성이므로, 세포의 계놈에서 자연적으로 발생하지 않는다. 또 다른 추가의 실시 형태에서, 인식 부위는 세포 및 관심 대상의 폴리뉴클레오티드에 외인성이며, 표적 유전자좌에 위치되고자 한다. 추가의 실시 형태에서, 외인성 또는 내인성 인식 부위는 숙주 세포의 계놈에 단 한 번만 존재한다. 구체적인 실시 형태에서, 계놈 내에 단 한 번만 발생하는 내인성 또는 고유 부위가 동정된다. 이어서, 이러한 부위는 내인성 인식 부위에서 틸 또는 이중 가닥 절단을 일으킬 뉴클레아제 제제를 설계하는 데 사용될 수 있다.

[0049] 인식 부위의 길이는 달라질 수 있으며, 예를 들어, 징크 핑거 뉴클레아제 (ZFN) 쌍에 대하여 약 30 내지 36 bp (즉, 각 ZFN에 대하여 약 15 내지 18 bp), 전사 활성화 인자 유사 이펙터 뉴클레아제(탈렌)에 대하여 약 36 bp, 또는 CRISPR/Cas9 가이드 RNA에 대하여 약 20 bp인 인식 부위를 포함한다.

[0050] 예시된 인식 부위의 활성 변이체 및 단편도 제공된다. 이러한 활성 변이체는 주어진 인식 부위에 대하여, 적어도 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상의 서열 동일성을 포함할 수 있으며, 여기서 활성 변이체는 생물학적 활성을 보유하므로, 서열 특이적 방법으로 뉴클레아제 제제에 의해 인식되고 절단될 수 있다. 뉴클레아제 제제에 의한 인식 부위의 이중 가닥 절단을 측정하기 위한 분석은 당업계에 공지되어 있다 (예를 들어, 이의 전체 내용이 본 명세서에 참조로 포함되는 문헌[TaqMan® qPCR assay, Frenthewey D. *et al.*, Methods in Enzymology, 2010, 476:295-307]).

[0051] 구체적인 실시 형태에서, 인식 부위는 선택 마커를 암호화하는 폴리뉴클레오티드 내에 위치된다. 이러한 위치는 선택 마커의 암호화 영역 내에 또는 선택 마커의 발현에 영향을 주는 조절 영역 내에 위치할 수 있다. 따라서, 뉴클레아제 제제의 인식 부위는 선택 마커의 인트론, 프로모터, 인핸서, 조절 영역, 또는 선택 마커를 암호화하는 폴리뉴클레오티드의 임의의 비-단백질 암호화 영역에 위치할 수 있다. 구체적인 실시 형태에서, 인식 부위에서의 틸 또는 이중 가닥 절단은 선택 마커의 활성을 파괴한다. 기능적 선택 마커의 존재 또는 부재의 분석 방법은 알려져 있다.

[0052] 틸 또는 이중 가닥 절단을 원하는 인식 부위로 유도하는 임의의 뉴클레아제 제제가 본 명세서에 개시된 방법 및 조성물에 사용될 수 있다. 자연 발생 또는 천연 뉴클레아제 제제가 원하는 인식 부위에서 틸 또는 이중 가닥 절단을 유도하는 한, 상기 뉴클레아제 제제가 사용될 수 있다. 대안적으로, 변형되거나 유전자 조작된 뉴클레아제 제제가 사용될 수 있다. "유전자 조작된 뉴클레아제 제제"는 원하는 인식 부위에서 틸 또는 이중 가닥 절단을 특이적으로 인식하고 유도하도록 이의 천연 형태로부터 유전자 조작된 (변형되거나 유래된) 뉴클레아제를 포함한다. 따라서, 유전자 조작된 뉴클레아제 제제는 천연의 자연 발생 뉴클레아제 제제로부터 유래할 수 있거나, 인공적으로 생성되거나 합성될 수 있다. 뉴클레아제 제제의 변형은 단지 단백질 절단제의 하나의 아미노산 또는 핵산 절단제의 하나의 뉴클레오티드에서 일어날 수 있다. 일부 실시 형태에서, 유전자 조작된 뉴클레아제는 인식 부위에서 틸 또는 이중 가닥 절단을 유도하며, 여기서 인식 부위는 천연 (유전자 조작되지 않거나 변형되지 않은) 뉴클레아제 제제에 의해 인식되는 서열이 아니었다. 인식 부위 또는 다른 DNA에서 틸 또는 이중 가닥 절단을 형성하는 것은 본 명세서에서 인식 부위 또는 다른 DNA의 "컷팅" 또는 "절단"으로 명명될 수 있다.

[0053] 이어서, 이러한 절단은 비-상동 말단 결합 및 상동성-지정 수복 (상동 재조합) 중 하나의 방법으로 세포에 의해 수복될 수 있다. 비-상동 말단 결합 (NHEJ)에서, 이중 가닥 절단은 절단 말단을 서로 직접 라이게이션하여 수복된다. 이로 인해, 새로운 핵산 물질이 부위로 삽입되지 않지만, 일부 핵산 물질은 손실되어, 결실로 이어질 수 있다. 상동성-지정 수복에서, 절단된 표적 DNA 서열과 상동성인 공여자 폴리뉴클레오티드는 절단된 표적 DNA 서열의 수복을 위한 주형으로서 사용될 수 있는데, 이는 공여자 폴리뉴클레오티드로부터 표적 DNA로 유전 정보를 이동시킨다. 따라서, 새로운 핵산 물질이 부위로 삽입/복제될 수 있다. NHEJ 및/또는 상동성-지정 수복으로 인한 표적 DNA의 변형은 유전자 보정(gene correction), 유전자 치환, 유전자 표지, 외래 도입 유전자(transgene) 삽입, 뉴클레오티드 결실, 유전자 파괴, 유전자 돌연변이 등에 사용될 수 있다.

[0054] 일 실시 형태에서, 뉴클레아제 제제는 전사 활성화 인자 유사 이펙터 뉴클레아제 (탈렌)이다. TAL 이펙터 뉴클레아제는 원핵생물 또는 진핵생물의 계놈의 특정 표적 서열에서 이중 가닥 절단을 형성하는 데 사용될 수 있는 서열 특이적 뉴클레아제의 부류이다. TAL 이펙터 뉴클레아제는 천연 또는 유전자 조작된 전사 활성화 인자 유사(TAL) 이펙터 또는 이의 기능적 부분을 엔도뉴클레아제, 예를 들어 *FokI*의 촉매 도메인에 융합시킴으로써 생성된다. 독특한 모듈러 TAL 이펙터 DNA 결합 도메인은 잠재적으로 주어진 DNA 인식 특이성을 가진 단백질을 설계할 수 있다. 따라서, TAL 이펙터 뉴클레아제의 DNA 결합 도메인은 특정 DNA 표적 부위를 인식하도록 유전자 조작될 수 있으므로, 원하는 표적 서열에서 이중 가닥 절단을 형성하는 데 사용될 수 있다. 모두 그 전체 내용



이 본 명세서에 참조로 포함되는 WO 2010/079430; 문헌[Morbitzer *et al.* (2010) *PNAS* 10.1073/pnas.1013133107]; 문헌[Scholze & Boch (2010) *Virulence* 1:428-432]; 문헌[Christian *et al. Genetics* (2010) 186:757-761]; 문헌[Li *et al.* (2010) *Nuc. Acids Res.* (2010) doi:10.1093/nar/gkq704]; 및 문헌[Miller *et al.* (2011) *Nature Biotechnology* 29:143-148]을 참조한다.

[0055] 적절한 TAL 뉴클레아제의 예 및 적절한 TAL 뉴클레아제의 제조 방법이며, 예를 들어, 미국 특허 출원 제 2011/0239315 A1호, 제2011/0269234 A1호, 제2011/0145940 A1호, 제2003/0232410 A1호, 제2005/0208489 A1호, 제2005/0026157 A1호, 제2005/0064474 A1호, 제2006/0188987 A1호 및 제2006/0063231 A1호 (각각, 본 명세서에 참조로 포함됨)에 개시되어 있다. 다양한 실시 형태에서, TAL 이펙터 뉴클레아제는, 예를 들어, 관심 계층 유전자좌의 표적 핵산 서열이나 그 주변에서 절단되는 것으로 유전자 조작되며, 여기서 표적 핵산 서열은 표적화 벡터에 의해 변형되는 서열 또는 그 주변에 존재한다. 본 명세서에 제공된 다양한 방법 및 조성물에서의 사용에 적합한 TAL 뉴클레아제에는 본 명세서에 기재된 표적화 벡터에 의해 변형되는 표적 핵산 서열에 또는 그 주변에 결합하도록 특별히 설계된 것들이 포함된다.

[0056] 일 실시 형태에서, 탈렌의 각 단량체는 2개의 초가변 잔기(hypervariable residue)를 통해 단일 염기쌍을 인식하는 33 내지 35개의 TAL 반복체를 포함한다. 일 실시 형태에서, 뉴클레아제 제제는 독립(independent) 뉴클레아제에 작동 가능하게 연결된 TAL 반복체 기반 DNA 결합 도메인을 포함하는 키메라 단백질이다. 일 실시 형태에서, 독립 뉴클레아제는 FokI 엔도뉴클레아제이다. 일 실시 형태에서, 뉴클레아제 제제는 제1 TAL 반복체 기반 DNA 결합 도메인 및 제2 TAL 반복체 기반 DNA 결합 도메인을 포함하며, 여기서 제1 및 제2 TAL 반복체 기반 DNA 결합 도메인은 각각, FokI 뉴클레아제 서브유닛에 작동 가능하게 연결되고, 제1 및 제2 TAL 반복체 기반 DNA 결합 도메인은 가변 길이 (12 내지 20 bp)의 스페이서 서열에 의해 분리된 표적 DNA 서열의 각 가닥에서 2개의 인접한 표적 DNA 서열을 인식하고, FokI 뉴클레아제 서브유닛은 표적 서열에서 이중 가닥 절단을 형성하는 활성 뉴클레아제를 생성하도록 이량체화한다.

[0057] 본 명세서에 개시된 다양한 방법 및 조성물에 사용되는 뉴클레아제 제제는 징크 핑거 뉴클레아제 (ZFN)를 추가로 포함할 수 있다. 일 실시 형태에서, ZFN의 각 단량체는 3개 이상의 징크 핑거 기반 DNA 결합 도메인을 포함하며, 여기서 각 징크 핑거 기반 DNA 결합 도메인은 3 bp 서브사이트(subsite)에 결합한다. 다른 실시 형태에서, ZFN은 독립 뉴클레아제에 작동 가능하게 연결된 징크 핑거 기반 DNA 결합 도메인을 포함하는 키메라 단백질이다. 일 실시 형태에서, 독립 엔도뉴클레아제는 FokI 엔도뉴클레아제이다. 일 실시 형태에서, 뉴클레아제 제제는 제1 ZFN 및 제2 ZFN을 포함하며, 여기서 제1 ZFN 및 제2 ZFN은 각각, FokI 뉴클레아제 서브유닛에 작동 가능하게 연결되고, 제1 및 제2 ZFN은 약 5 내지 7 bp 스페이서에 의해 분리된 표적 DNA 서열의 각 가닥에서 2개의 인접한 표적 DNA 서열을 인식하고, FokI 뉴클레아제 서브유닛은 이중 가닥 절단을 형성하는 활성 뉴클레아제를 생성하도록 이량체화한다. 예를 들어, 각각, 본 명세서에 참조로 포함되는 US20060246567; US20080182332; US20020081614; US20030021776; WO/2002/057308A2; US20130123484; US20100291048; WO/2011/017293A2; 및 문헌[Gaj *et al.* (2013) *Trends in Biotechnology*, 31(7):397-405]을 참조한다.

[0058] 본 명세서에 제공된 방법의 일 실시 형태에서, 뉴클레아제 제제는 (a) FokI 엔도뉴클레아제에 융합된 징크 핑거 기반 DNA 결합 도메인을 포함하는 키메라 단백질; 또는 (b) FokI 엔도뉴클레아제에 융합된 전사 활성인자-유사 이펙터 뉴클레아제 (탈렌)를 포함하는 키메라 단백질을 포함한다.

[0059] 또 다른 실시 형태에서, 뉴클레아제 제제는 메가뉴클레아제이다. 메가뉴클레아제는 보존 서열 모티프에 기초한 4개의 패밀리로 분류되고, 패밀리는 LAGLIDADG (서열 번호: 16), GIY-YIG, H-N-H 및 His-Cys 박스 패밀리이다. 이러한 모티프는 금속 이온의 배위 및 인산다이에스테르 결합의 가수분해에 관여한다. HEase는 이들의 긴 인식 부위와, 이들의 DNA 기질에서의 약간의 서열 다형성에 대한 내성에 주목할 만하다. 메가뉴클레아제 도메인, 구조 및 기능은 공지되어 있으며, 예를 들어, 문헌[Guhan and Muniyappa (2003) *Crit Rev Biochem Mol Biol* 38:199-248]; 문헌[Lucas *et al.*, (2001) *Nucleic Acids Res* 29:960-9]; 문헌[Jurica and Stoddard, (1999) *Cell Mol Life Sci* 55:1304-26]; 문헌[Stoddard, (2006) *Q Rev Biophys* 38:49-95]; 및 문헌[Moure *et al.*, (2002) *Nat Struct Biol* 9:764]을 참조한다. 일부 예에서, 자연 발생 변이체 및/또는 유전자 조작된 유도체 메가뉴클레아제가 사용된다. 동력학, 보조 인자 상호 작용, 발현, 최적 조건 및/또는 인식부위 특이성을 변형시키고 활성을 스크리닝하는 방법이 공지되어 있으며, 예를 들어, 문헌[Epinat *et al.*, (2003) *Nucleic Acids Res* 31:2952-62]; 문헌[Chevalier *et al.*, (2002) *Mol Cell* 10:895-905]; 문헌[Gimble *et al.*, (2003) *Mol Biol* 334:993-1008]; 문헌[Seligman *et al.*, (2002) *Nucleic Acids Res* 30:3870-9]; 문헌[Sussman *et al.*, (2004) *J Mol Biol* 342:31-41]; 문헌[Rosen *et al.*, (2006) *Nucleic Acids Res* 34:4791-800]; 문헌[Chames *et al.*, (2005) *Nucleic Acids Res* 33:e178]; 문헌[Smith *et al.*, (2006) *Nucleic Acids Res* 34:e149]; 문헌

[Gruen *et al.*, (2002) *Nucleic Acids Res* 30:e29]; 문헌[Chen and Zhao, (2005) *Nucleic Acids Res* 33:e154]; WO2005105989; WO2003078619; WO2006097854; WO2006097853; WO2006097784; 및 WO2004031346을 참조한다.

[0060] I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAIIP, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-TliI, I-PpoI, PI-PspI, F-SceI, F-SceII, F-SuvI, F-TevI, F-TevII, I-AmaI, I-AniI, I-ChuI, I-CmoeI, I-CpaI, I-CpaII, I-CsmI, I-CvuI, I-CvuAIP, I-DdiI, I-DdiII, I-DirI, I-DmoI, I-HmuI, I-HmuII, I-HsNIP, I-LlaI, I-MsoI, I-NaaI, I-NanI, I-NcIIP, I-NgrIP, I-NitI, I-NjaI, I-Nsp236IP, I-PakI, I-PboIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PobIP, I-PorI, I-PorIIP, I-PbpIP, I-SpBetaIP, I-ScaI, I-SexIP, I-SneIP, I-SpomI, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdeIP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MtuI, PI-MtuHIP, PI-MtuHIIP, PI-PfuI, PI-PfuII, PI-PkoI, PI-PkoII, PI-Rma43812IP, PI-SpBetaIP, PI-SceI, PI-TfuI, PI-TfuII, PI-ThyI, PI-TliI, PI-TliII, 또는 이들의 임의의 활성 변이체 또는 단편을 포함하나, 이에 한정되지 않는 임의의 메가뉴클레아제가 본 명세서에 사용될 수 있다.

[0061] 일 실시 형태에서, 메가뉴클레아제는 12 내지 40개의 염기쌍의 이중 가닥 DNA 서열을 인식한다. 일 실시 형태에서, 메가뉴클레아제는 게놈의 하나의 완전히 매칭된 표적 서열을 인식한다. 일 실시 형태에서, 메가뉴클레아제는 호밍(homing) 뉴클레아제이다. 일 실시 형태에서, 호밍 뉴클레아제는 LAGLIDAG (서열 번호: 16) 패밀리 of 호밍 뉴클레아제이다. 일 실시 형태에서, LAGLIDAG (서열 번호: 16) 패밀리의 호밍 뉴클레아제는 I-SceI, I-CreI 및 I-DmoI로부터 선택된다.

[0062] 뉴클레아제 제제는 타입 I, 타입 II, 타입 III 및 타입 IV 엔도뉴클레아제를 포함하는 제한 엔도뉴클레아제 (제한 효소)를 추가로 포함할 수 있다. 타입 I 및 타입 III 제한 엔도뉴클레아제는 특이적 인식 부위를 인식하지만, 전형적으로 뉴클레아제 결합 부위로부터 가변 위치에서 절단하며, 이는 절단 부위 (인식 부위)로부터 수백 염기쌍 떨어질 수 있다. 타입 II 시스템에서, 제한 활성은 임의의 메틸라아제 활성과 무관하며, 절단은 전형적으로 결합부위 내에 또는 그 부근의 특정 부위에서 일어난다. 대부분의 타입 II 효소는 회문 서열을 절단하지만, 타입 IIa 효소는 비회문 인식 부위를 인식하고, 인식 부위의 외측을 절단하며, 타입 IIb 효소는 인식 부위 외측의 양측 부위에서 서열을 2회 절단하고, 타입 IIc 효소는 비대칭 인식 부위를 인식하며, 인식 부위로부터 약 1 내지 20개의 뉴클레오타이드의 한정된 거리에서 한쪽에서 절단한다. 타입 IV 제한효소는 메틸화 DNA를 표적으로 한다. 제한효소는 추가로, 예를 들어, 리베이스(REBASE) 데이터베이스에 기재되어 분류되어 있다 (문헌 [webpage at rebase.neb.com]; 문헌[Roberts *et al.*, (2003) *Nucleic Acids Res* 31:418-20], 문헌[Roberts *et al.*, (2003) *Nucleic Acids Res* 31:1805-12] 및 문헌[Belfort *et al.*, (2002) in *Mobile DNA II*, pp. 761-783, Eds. Craigie *et al.*, (ASM Press, Washington, DC)]. 구체적인 실시 형태에서, 적어도 2개의 엔도뉴클레아제 효소가 뉴클레아제 제제로써 선택될 수 있고, 여기서 효소는 상용성이거나 상보적인, 점착 말단을 생성한다.

[0063] 다양한 방법 및 조성물에 사용되는 뉴클레아제 제제는 또한 CRISPR/Cas 시스템을 포함할 수 있다. 이러한 시스템은 Cas9 뉴클레아제를 사용할 수 있는데, 경우에 따라서는 그것이 발현되는 원하는 세포 유형에 대하여 코돈 최적화된다. 상기 시스템은 또한, 코돈 최적화된 Cas9와 함께 기능하는 융합된 crRNA-tracrRNA 구축물을 사용한다. 이러한 단일 RNA는 종종 가이드 RNA 또는 gRNA로 명명된다. gRNA 내에서, crRNA 부분은 주어진 인식 부위에 대한 '표적 서열'로 특정되며, tracrRNA는 종종 '스캐폴드(scaffold)'로 명명된다. 이러한 시스템은 다양한 진핵 세포 및 원핵 세포에서 기능하는 것으로 나타났다. 간략하게, 표적 서열을 포함하는 짧은 DNA 단편은 가이드 RNA 발현 플라스미드에 삽입된다. gRNA 발현 플라스미드는 표적 서열 (일부 실시 형태에서, 약 20개의 뉴클레오타이드), tracrRNA 서열 (스캐폴드)의 형태뿐만 아니라, 세포에서 활성인 적절한 프로모터 및 진핵 세포에서의 적절한 프로세싱에 필요한 요소를 포함한다. 많은 시스템은 어닐링되어 이중 가닥 DNA를 형성한 다음에, gRNA 발현 플라스미드에 클로닝된 맞춤형(custom) 상보적 올리고에 의존한다. 이어서, gRNA 발현 카세트 및 Cas9 발현 카세트가 세포로 도입된다. 예를 들어, 각각 본 명세서에 참조로 포함되는, 문헌[Mali P *et al.* (2013) *Science* 2013 Feb 15; 339 (6121):823-6]; 문헌[Jinek M *et al.* *Science* 2012 Aug 17;337(6096):816-21]; 문헌[Hwang WY *et al.* *Nat Biotechnol* 2013 Mar;31(3):227-9]; 문헌[Jiang W *et al.* *Nat Biotechnol* 2013 Mar;31(3):233-9]; 및 문헌[Cong L *et al.* *Science* 2013 Feb 15;339(6121):819-23]을 참조한다.

[0064] 본 명세서에 개시된 방법 및 조성물은 세포 내의 게놈을 변형시키기 위해 CRISPR/CRISPR-관련 (Cas) 시스템 또는 이러한 시스템의 성분을 이용할 수 있다. CRISPR/Cas 시스템은 Cas 유전자의 발현 또는 활성의 유도에 관여

하는 전사물 및 다른 요소를 포함한다. CRISPR/Cas 시스템은 타입 I, 타입 II 또는 타입 III 시스템일 수 있다. 본 명세서에 개시된 방법 및 조성물은 핵산의 부위-지정 절단을 위해 CRISPR 복합체 (Cas 단백질과 복합체를 형성한 가이드 RNA(gRNA)를 포함함)를 이용함으로써 CRISPR/Cas 시스템을 사용한다.

[0065] 본 명세서에 개시된 방법에 사용되는 일부 CRISPR/Cas 시스템은 자연 발생적이지 않다. "자연적으로 발생하지 않는"시스템은 인간의 손의 관여를 나타내는 모든 것, 예를 들어, 이들의 자연 발생 상태에서부터 변경되거나 돌연변이되거나, 이들이 실제로는 자연적으로 결합하여 있는 적어도 하나의 다른 성분을 적어도 실질적으로 함유하지 않거나, 이들이 자연적으로 결합하여 있지 않은 적어도 하나의 다른 구성 요소와 결합된 시스템의 하나 이상의 성분을 포함한다. 예를 들어, 일부 CRISPR/Cas 시스템은 동시에 자연적으로 발생하지 않는 gRNA 및 Cas 단백질을 포함하는 자연적으로 발생하지 않는 CRISPR 복합체를 사용한다.

[0066] 뉴클레아제 제제의 활성 변이체 및 단편 (즉, 유전자 조작된 뉴클레아제 제제)이 또한 제공된다. 이러한 활성 변이체는 천연 뉴클레아제 제제에 대하여 적어도 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상의 서열 동일성을 포함할 수 있으며, 여기서 활성 변이체는 원하는 인식 부위에서 절단하는 능력을 보유하므로, 틸 또는 이중 가닥 절단 유도 활성을 보유한다. 예를 들어, 본 명세서에 기재된 임의의 뉴클레아제 제제는 천연 엔도뉴클레아제 서열로부터 변형되고, 천연 뉴클레아제 제제에 의해 인식되지 않는 인식 부위에서 틸 또는 이중 가닥 절단을 인식하고 유도하도록 설계될 수 있다. 따라서, 일부 실시 형태에서, 유전자 조작된 뉴클레아제는 상응하는 천연 뉴클레아제 제제 인식 부위와 상이한 인식 부위에서 틸 또는 이중 가닥 절단을 유도하는 특이성을 갖는다. 틸 또는 이중 가닥 절단 유도 활성에 대한 분석이 알려져 있고, 일반적으로 인식 부위를 함유하는 DNA 기질 상의 엔도뉴클레아제의 전반적인 활성 및 특이성을 측정한다.

[0067] IV. CRISPR/Cas 시스템 (gRNA-Cas 복합체)

[0068] 본 발명의 방법은 핵산의 부위-지정 절단을 위한 CRISPR/Cas 시스템 (예를 들어, gRNA-Cas 복합체)을 사용할 수 있다. 구체적으로, gRNA에 의해 확인된 표적 부위로 유도된 핵산의 Cas 절단은, 이어서 제2 핵산과 결합하여, 둘 이상의 핵산을 부위 특이적 방식으로 조립할 수 있는 말단을 포함하는 분해된 핵산을 생성한다.

[0069] "gRNA-Cas 복합체"는 Cas 단백질과 gRNA의 복합체를 포함한다. gRNA는 절단된 핵산과 상이한 핵산 사이에 중첩하는 말단을 생성하는 표적 부위로 Cas 절단을 유도하도록 설계되거나 선택될 수 있다. gRNA-Cas 복합체는 이미 복합체를 형성한 제제를 공급받을 수 있거나, 별개의 단백질 및 RNA 요소를 공급받을 수 있으며, 이러한 경우에 이들은 본 명세서에 기재된 방법 및 반응 혼합물에서 복합체를 형성하여 gRNA-Cas 복합체를 형성한다.

[0070] A. Cas RNA-유도 엔도뉴클레아제

[0071] Cas 단백질은 일반적으로 적어도 하나의 RNA 인식 또는 결합 도메인을 포함한다. 이러한 도메인은 가이드 RNA (gRNA, 이하에 상세히 기술됨)와 상호작용할 수 있다. Cas 단백질은 또한 뉴클레아제 도메인 (예를 들어, DNase 또는 RNase 도메인), DNA 결합 도메인, 헬리카제 도메인, 단백질-단백질 상호작용 도메인, 이량체화 도메인 및 기타 도메인을 포함할 수 있다. 뉴클레아제 도메인은 핵산 절단을 위한 촉매 활성을 갖는다. 절단은 핵산 분자의 공유 결합의 파괴를 포함한다. 절단은 평할 말단 또는 스테거된 말단을 생성할 수 있으며, 단일 가닥 또는 이중 가닥으로될 수 있다.

[0072] Cas 단백질의 예에는 Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5e (CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9 (Csn1 또는 Csx12), Cas10, Cas10d, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1 (CasA), Cse2 (CasB), Cse3 (CasE), Cse4 (CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 및 Cu1966, 및 이들의 상동체 또는 변형된 버전(modified version)이 포함된다.

[0073] 틸 또는 이중 가닥 절단을 원하는 인식 부위로 유도하는 임의의 Cas 단백질이 본 명세서에 개시된 방법 및 조성물에 사용될 수 있다. 자연 발생 또는 천연 Cas 단백질이 원하는 인식 부위에서 이중 가닥 절단을 유도하는 한, 상기 Cas 단백질이 사용될 수 있다. 대안적으로, 변형되거나 유전자 조작된 Cas 단백질이 사용될 수 있다. "유전자 조작된 Cas 단백질"은 원하는 인식 부위에서 틸 또는 이중 가닥 절단을 특이적으로 인식하고 유도하도록 이의 천연 형태로부터 유전자 조작된 (변형되거나 유래된) Cas 단백질을 포함한다. 따라서, 유전자 조작된 Cas 단백질은 천연의 자연 발생 Cas 단백질로부터 유래할 수 있거나, 인공적으로 생성되거나 합성될 수 있다.

[0074] 특정 실시 형태에서, Cas 단백질은 Cas9이다. Cas9 단백질은 전형적으로 보존 구조를 갖는 4개의 주요 모티프를 공유한다. 모티프 1, 2 및 4는 RuvC 유사 모티프이고, 모티프 3은 HNH 모티프이다. Cas9의 뉴클레아제 활

성은 표적 DNA를 절단하여, 이중 가닥 절단을 생성한다. 이어서, 이러한 절단은 비-상동 말단 결합 및 상동성-지정 수복 (상동 재조합) 중 하나의 방법으로 세포에 의해 수복될 수 있다. 비-상동 말단 결합 (NHEJ)에서, 이중 가닥 절단은 절단 말단을 서로 직접 라이게이션하여 수복된다. 이로 인해, 새로운 핵산 물질이 부위로 삽입되지 않지만, 일부 핵산 물질은 손실되어, 결실로 이어질 수 있다. 상동성-지정 수복에서, 절단된 표적 DNA 서열과 상동성인 공여자 폴리뉴클레오티드는 절단된 표적 DNA 서열의 수복을 위한 주형으로서 사용될 수 있는데, 이는 공여자 폴리뉴클레오티드로부터 표적 DNA로 유전 정보를 이동시킨다. 따라서, 새로운 핵산 물질이 부위로 삽입/복제될 수 있다. NHEJ 및/또는 상동성-지정 수복으로 인한 표적 DNA의 변형은 유전자 보정, 유전자 치환, 유전자 표지, 외래 도입 유전자 삽입, 뉴클레오티드 결실, 유전자 파괴, 유전자 돌연변이 등에 사용될 수 있다.

[0075] Cas 단백질은 타입 II CRISPR/Cas 시스템으로부터 유래될 수 있다. 예를 들어, Cas 단백질은 Cas9 단백질이거나, Cas9 단백질로부터 유래될 수 있다. Cas9 단백질은 전형적으로 보존 구조를 갖는 4개의 주요 모티프를 공유한다. 모티프 1, 2 및 4는 RuvC 유사 모티프이고, 모티프 3은 HNH 모티프이다. Cas9 단백질은 예를 들어, 화농연쇄구균(*Streptococcus pyogenes*), 스트렙토코커스 써모필러스(*Streptococcus thermophilus*), 스트렙토코커스 속(*Streptococcus sp.*), 황색포도상구균(*Staphylococcus aureus*), 노카르디옵시스 다손빌레이(*Nocardiopsis dassonvillei*), 스트렙토마이세스 프리스티네스피랄리스(*Streptomyces pristinaespiralis*), 스트렙토마이세스 비리도크로모게네스(*Streptomyces viridochromogenes*), 스트렙토마이세스 비리도크로모게네스(*Streptomyces viridochromogenes*), 스트렙토스포랑기움 로세움(*Streptosporangium roseum*), 스트렙토스포랑기움 로세움(*Streptosporangium roseum*), 알리사이클로바클루스 아시도칼다리우스(*Alicyclobacillus acidocaldarius*), 바실러스 슈도마이코이데스(*Bacillus pseudomycoides*), 바실러스 셀레니티레두센스(*Bacillus selenitireducens*), 엑시구오박테리움 시비리쿰(*Exiguobacterium sibiricum*), 락토바실러스 델브루에키(*Lactobacillus delbrueckii*), 락토바실러스 살리바리우스(*Lactobacillus salivarius*), 마이크로실라 마리나(*Microscilla marina*), 부르크홀데리아레스 박테리움(*Burkholderiales bacterium*), 폴라로모나스 나프탈레니보란스(*Polaromonas naphthalenivorans*), 폴라로모나스 속(*Polaromonas sp.*), 크로코스파에라 왓슨이(*Crocospaera watsonii*), 시아노테세 속(*Cyanothece sp.*), 마이크로시스티스 아에루기노사(*Microcystis aeruginosa*), 시네코코커스 속(*Synechococcus sp.*), 아세토할로비움 아라바티쿰(*Acetohalobium arabaticum*), 암모니팩스 데젠시(*Ammonifex degensii*), 칼디셀룰로시립토 베시(*Caldicelulosiruptor beccsii*), 칸디다투스 데술포루디스(*Candidatus Desulforudis*), 클로스트리듐 보툴리눔(*Clostridium botulinum*), 클로스트리듐 디피실레(*Clostridium difficile*), 피네골디아 마그나(*Finegoldia magna*), 나트라나에로비우스 써모필러스(*Natranerobius thermophilus*), 펠로토마쿨럼 써모프로피오니쿰(*Pelotomaculum thermopropionicum*), 아시디티오바실러스 칼두스(*Acidithiobacillus caldus*), 아시디티오바실러스 페로옥시단스(*Acidithiobacillus ferrooxidans*), 알로크로마티움 비노숨(*Allochromatium vinosum*), 마리노박터 속(*Marinobacter sp.*), 니트로소코커스 할로필러스(*Nitrosococcus halophilus*), 니트로소코커스 왓슨이(*Nitrosococcus watsoni*), 슈도알테로모나스 할로플란크티스(*Pseudoalteromonas haloplanktis*), 크테도노박테르 라세미페르(*Ktedonobacter racemifer*), 메타노할로비움 에베스티가툼(*Methanohalobium evestigatum*), 아나베나 바리아빌리스(*Anabaena variabilis*), 노둘라리아 스푸미게나(*Nodularia spumigena*), 노스톡 속(*Nostoc sp.*), 아르트로스피라 맥시마(*Arthrospira maxima*), 아르트로스피라 플라텐시스(*Arthrospira platensis*), 아르트로스피라 속(*Arthrospira sp.*), 링비아속(*Lyngbya sp.*), 마이크로콜레우스 크토노플라스테스(*Microcoleus chthonoplastes*), 오실라토리아 속(*Oscillatoria sp.*), 페트로토가 모빌리스(*Petrotoga mobilis*), 써모시포 아프리카누스(*Thermosiphon africanus*) 또는 아카리오크로리스 마리나(*Acaryochloris marina*)로부터 유래될 수 있다. Cas9 패밀리 멤버의 추가의 예는 그 전체 내용이 본원에 참조로 포함되는 WO 2014/131833에 기재되어 있다. 화농연쇄구균 유래의 Cas9 단백질 또는 이의 유도체는 바람직한 효소이다. 화농연쇄구균 유래의 Cas9 단백질에는 SwissProt 수탁 번호 Q99ZW2가 할당되어 있다.

[0076] Cas 단백질은 야생형 단백질(즉, 천연에서 발생하는 것), 변형된 Cas 단백질(즉, Cas 단백질 변이체) 또는 야생형 또는 변형된 Cas 단백질의 단편일 수 있다. Cas 단백질은 또한 야생형 또는 변형된 Cas 단백질의 활성 변이체 또는 단편일 수 있다. 활성 변이체 또는 단편은 야생형 또는 변형된 Cas 단백질 또는 이의 부분에 대하여, 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상의 서열 동일성을 포함할 수 있으며, 여기서 활성 변이체는 원하는 절단 부위에서 절단하는 능력을 보유하므로, 틸 유도 또는 이중 가닥 절단 유도 활성을 보유한다. 틸 유도 또는 이중 가닥 절단 유도 활성에 대한 분석이 알려져 있고, 일반적으로 절단 부위를 함유하는 DNA 기질에 대한 Cas 단백질의 전반적인 활성 및 특이성을 측정한다.

[0077] Cas 단백질은 핵산 결합 친화성, 핵산 결합 특이성 및 / 또는 효소 활성을 증가시키거나 감소 시키도록 변형될 수 있다. Cas 단백질은 또한 단백질의 임의의 다른 활성 또는 성질, 예컨대 안정성을 변화시키기 위해 변형될



수 있다. 예를 들어, Cas 단백질의 하나 이상의 뉴클레아제 도메인은 변형되거나, 결실되거나, 비활성화될 수 있거나, 또는 Cas 단백질은 단백질의 기능에 본질적이지 않은 도메인을 제거하거나 Cas 단백질의 활성을 최적화(예를 들어, 향상 또는 감소)시키기 위해 절단될 수 있다.

[0078] 일부 Cas 단백질은 적어도 2개의 뉴클레아제 도메인, 예컨대 DNase 도메인을 포함한다. 예를 들어, Cas9 단백질은 RuvC-유사 뉴클레아제 도메인 및 HNH-유사 뉴클레아제 도메인을 포함할 수 있다. RuvC 및 HNH 도메인 각각은 이중-가닥 DNA의 상이한 가닥을 잘라서 DNA에 이중-가닥 절단을 만들 수 있다. 예를 들어, 그 전체 내용이 본 명세서에 참조로 포함되는 문헌[Jinek *et al.* (2012) *Science* 337:816-821]을 참조한다.

[0079] 뉴클레아제 도메인들 중 하나 또는 둘 다는 더 이상 기능적이지 않거나 뉴클레아제 활성이 감소되도록 결실되거나 돌연변이될 수 있다. 한쪽의 뉴클레아제 도메인이 결실되거나 돌연변이되는 경우, 생성된 Cas 단백질(예를 들어, Cas9)은 니카아제(nickase)로 지칭될 수 있으며, 이중 가닥 DNA 내의 CRISPR RNA 인식 서열에서 단일 가닥 절단을 생성할 수 있지만, 이중 가닥 절단을 생성할 수 없다(즉, 상보적 가닥 또는 비상보적 가닥 중 어느 한쪽만 절단할 수 있음). 뉴클레아제 도메인 둘 다가 결실되거나 돌연변이된다면, 생성되는 Cas 단백질(예를 들어, Cas9)의 이중-가닥 DNA의 가닥 둘 다를 절단하는 능력은 감소될 것이다. Cas9를 니카아제로 변환시키는 돌연변이의 예는 화농연쇄구균 유래의 Cas9의 RuvC 도메인에서의 D10A(Cas9의 위치 10에서 아스파르트산을 알라닌으로) 돌연변이이다. 마찬가지로, 화농연쇄구균 유래의 Cas9의 HNH 도메인에서의 H939A(아미노산 위치 839에서 히스티딘을 알라닌으로) 또는 H840A(아미노산 위치 840에서 히스티딘을 알라닌으로)가 Cas9를 니카아제로 변환시킬 수 있다. Cas9를 니카제로 변환시키는 돌연변이의 다른 예는 S. 써모필러스 유래의 Cas9에 상응하는 돌연변이를 포함한다. 예를 들어, 이들 각각이 전체적으로 본 명세서에 참고로 포함되는 문헌[Sapranaukas *et al.* (2011) *Nucleic Acids Research* 39:9275-9282] 및 WO 2013/141680호를 참조한다. 이러한 돌연변이는 부위-지정 돌연변이생성, PCR-매개 돌연변이생성 또는 전체 유전자 합성과 같은 방법을 사용하여 생성될 수 있다. 니카아제를 생성하는 다른 돌연변이의 예는, 예를 들어, 각각 본 명세서에 참조로 포함되는, WO/2013/176772A1 및 WO/2013/142578A1에서 찾을 수 있다.

[0080] Cas 단백질은 또한 융합 단백질일 수 있다. 예를 들어, Cas 단백질은 절단 도메인, 후성적 변형 도메인, 전사 활성화 도메인 또는 전사 억제인자 도메인에 융합될 수 있다. 그 전체 내용이 본 명세서에 참조로 포함되는 WO 2014/089290을 참조한다. Cas 단백질은 또한 안정성 향상 또는 저하를 제공하는 이중 폴리펩티드에 융합될 수 있다. 융합된 도메인 또는 이중 폴리 펩티드는 N 말단, C 말단 또는 Cas 단백질 내부에 위치할 수 있다.

[0081] Cas 단백질은 세포 내 국재화(subcellular localization)를 제공하는 이중 폴리펩티드에 융합될 수 있다. 이러한 이중 펩티드는, 예를 들어, 핵을 표적화하기 위한 SV40 NLS와 같은 핵 국재화 신호(NLS), 미토콘드리아를 표적화하기 위한 미토콘드리아 국재화 신호, ER 보유 신호(retention signal) 등을 포함한다. 예를 들어, 문헌[Lange *et al.* (2007) *J. Biol. Chem.* 282:5101-5105]을 참조한다. 이러한 세포 내 국재화 신호는 N 말단, C 말단 또는 Cas 단백질 내의 어디에도 위치할 수 있다. NLS는 염기성 아미노산의 스트레치를 포함할 수 있으며, 모노파르타이트(monopartite) 서열 또는 바이파르타이트(bipartite) 서열일 수 있다.

[0082] Cas 단백질은 또한 세포 투과성 도메인에 연결될 수 있다. 예를 들어, 세포 투과성 도메인은 HIV-1 TAT 단백질, 인간 B형 간염 바이러스 유래의 TLM 세포 투과성 모티프, MPG, Pep-1, VP22, 단순 헤르페스 바이러스 유래의 세포 투과성 펩티드, 또는 폴리아르기닌 펩티드 서열로부터 유래될 수 있다. 예를 들어, 그 전체 내용이 본 명세서에 참조로 포함되는 WO 2014/089290을 참조한다. 세포 투과성 도메인은 N 말단, C 말단 또는 Cas 단백질 내의 어디에도 위치할 수 있다.

[0083] Cas 단백질은 또한 추적 또는 정제를 용이하게 하기 위한 이중 폴리펩티드, 예를 들어 형광 단백질, 정제 태그 또는 에피토프 태그를 포함할 수 있다. 형광 단백질의 예로는 녹색 형광 단백질(예를 들어, GFP, GFP-2, tagGFP, turboGFP, eGFP, 에메랄드(Emerald), 아자미 그린(Azami Green), 모노머릭 아자미 그린(Monomeric Azami Green), CopGFP, AceGFP, ZsGreen1), 황색 형광 단백질(예를 들어, YFP, eYFP, 시트린(Citrine), 비너스, YPet, PhiYFP, ZsYellow1), 청색 형광 단백질(예를 들어, eBFP, eBFP2, 아주라이트(Azurite), mKalamal, GFPuv, 사파이어(Sapphire), T-사파이어), 시안 형광 단백질(예를 들어, eCFP, 세룰리언(Cerulean), CyPet, AmCyan1, 미도리이시-시안(Midoriishi-Cyan)), 적색 형광 단백질(mKate, mKate2, mPlum, DsRed 모노머, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-모노머, HcRed-탠덤(HcRed-Tandem), HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred), 오렌지색 형광 단백질(mOrange, mKO, 쿠사비라-오렌지(Kusabira-Orange), 모노머릭 쿠사비라-오렌지, mTangerine, tdTomato), 및 임의의 다른 적절한 형광 단백질을 들 수 있다. 태그의 예로는 글루타티온-S-트랜스페라아제(GST), 키틴 결합 단백질(CBP), 말토스 결합 단백질, 티오

레독신(TRX), 폴리(NANP), 탠덤 친화성 정제(tandem affinity purification; TAP) tag, myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, 적혈구응집소(HA), nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, 히스티딘(His), 바이오틴 카르복실 운반 단백질(BCCP), 및 칼모듈린을 들 수 있다.

[0084] 일부 실시 형태에서, Cas 단백질은 결과로서 얻어진 뉴클레아제 활성이 변경되도록 변형될 수 있다. Cas 내에 특정 돌연변이는 표적 DNA의 상보적 가닥과 비-상보적 가닥 모두를 절단하는 뉴클레아제의 능력을 감소시킬 수 있다. 예를 들어, Cas 단백질은 알려진 위치에서 돌연변이화되어, 뉴클레아제 활성을 상보적 가닥 또는 비-상보적 가닥의 절단으로 제한할 수 있다. 구체적으로, D10A를 갖는 Cas9 (Cas9의 아미노산 위치 10에서 아스파르트레이트가 알라닌으로) 돌연변이는 표적 DNA의 상보적 가닥을 절단할 수 있지만, 표적 DNA의 비-상보적 가닥을 절단하는 능력은 감소된다. 일부 실시 형태에서, H840A를 갖는 Cas9 (아미노산 위치 840에서 히스티딘이 알라닌으로) 돌연변이는 표적 DNA의 비-상보적 가닥을 절단할 수 있지만, 표적 DNA의 상보적 가닥을 절단하는 능력은 감소된다. D10A 또는 H840A 돌연변이를 갖는 Cas9의 뉴클레아제 활성은 DSB 대신에 단일 가닥 절단 (SSB)을 생성할 것이다. 다른 잔기를 돌연변이화하여 동일한 효과를 얻을 수 있다 (즉, 어느 하나의 뉴클레아제 부분을 불활성화시킨다). 비제한적인 실시예로서, 잔기 D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986 및/또는 A987이 있다 (즉, 치환됨). 또한, 알라닌 이외의 아미노산을 치환하는 적합할 수 있다. 일부 실시 형태에서, 뉴클레아제 활성이 감소된 경우 (예를 들어, Cas9 단백질이 D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986 및/또는 A987 돌연변이, 예컨대 D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A 및/또는 D986A를 갖는 경우), 뉴클레아제는 gRNA와 상호작용하는 능력을 보유하는 한 gRNA에 의해 여전히 표적 DNA 서열로 유도되므로, 여전히 표적 DNA에 부위 특이적 방식으로 결합할 수 있다.

[0085] 일부 실시 형태에서, Cas는 뉴클레아제가 표적 DNA의 상보적 또는 비-상보적 가닥을 절단하지 않도록 변경된다. 예를 들어, D10A 및 H840A 돌연변이를 갖는 Cas9의 표적 DNA의 상보적 및 비-상보적 가닥을 절단하는 능력은 감소된다. 다른 잔기는 돌연변이화되어 동일한 효과를 얻을 수 있다 (즉, 어느 하나의 뉴클레아제 부분을 불활성화시킨다). 비제한적인 실시예로서, 잔기 D10, G12, G17, E762, H840, N854, N863, H982, H983, A984 및/또는 D986이 실질적으로 뉴클레아제 활성을 제거하도록 치환될 수 있다. 추가로, 알라닌 치환 이외의 돌연변이가 적합할 수 있다.

[0086] 용어 "표적 부위" 또는 "표적 서열"은 상호 교환적으로 사용될 수 있으며, 결합을위한 충분한 조건이 존재한다면, gRNA의 DNA-표적화 세그먼트가 결합할 표적 DNA에 존재하는 핵산 서열을 포함한다. 예를 들어, 표적 DNA 내의 표적 부위 (또는 표적 서열)는 Cas 단백질 또는 gRNA에 의해 표적화된다 (또는 이에 의해 결합되거나 이와 혼성화되거나 이에 상보적이다). 적합한 DNA/RNA 결합 조건은 보통 세포 내에 존재하는 생리학적 조건을 포함한다. 다른 적합한 DNA/RNA 결합 조건 (예를 들어, 무세포계에서의 조건)이 본 기술 분야에 알려져 있다 (예를 들어, 문헌[Molecular Cloning: A Laboratory Manual, 3rd Ed. (Sambrook *et al.*, Harbor Laboratory Press 2001)]을 참조한다). Cas 단백질 또는 gRNA와 상보적이고 혼성화되는 표적 DNA의 가닥은 "상보적 가닥"으로 명명되고, "상보적 가닥"과 상보적인 표적 DNA의 가닥 (따라서 Cas 단백질 또는 gRNA에 상보적인 것은 아니다)은 "비상보적 가닥" 또는 "주형 가닥"으로 명명된다.

[0087] Cas 단백질은 표적 서열 내의 또는 표적 서열 외부의 부위에서 핵산을 절단할 수 있다. "절단 부위"는 Cas 단백질이 단일-가닥 절단 또는 이중 가닥 절단을 생성하는 핵산의 위치를 포함한다. Cas 단백질이 이중 가닥 절단을 생성하는 경우, 절단 부위는 핵산의 두 가닥 상의 동일한 위치에 있을 수 있거나 (평활 말단을 생성함), 각각의 가닥 상의 상이한 부위에 있을 수 있다 (점착 또는 부착(cohesive) 말단을 생성함). 점착 말단은 또한 각각의 가닥 상의 절단 부위에서 단일-가닥 절단을 생성하는 2개의 Cas 단백질을 사용함으로써 생성될 수 있다. Cas9에 의한 표적 DNA의 부위 특이적 절단은 (i) 가이드 RNA와 표적 DNA 사이의 염기쌍 상보성; 및 (ii) 표적 DNA 내의 프로토스페이서 인접 모티프 (PAM)로 불리는 짧은 모티프에 의해 결정된 위치에서 발생할 수 있다. 예를 들어, Cas9의 절단 부위는 PAM 서열의 상류에서 약 1 내지 약 10개 또는 약 2 내지 약 5개의 염기쌍 (예를 들어, 3 염기쌍)일 수 있다. 일부 실시 형태에서 (예를 들어, 화농연쇄구균 유래의 Cas9 또는 밀접하게 관련된 Cas9가 사용된 경우), 비상보적 가닥의 PAM 서열은 5' -XGG-3' 일 수 있으며, 여기서 X는 임의의 DNA 뉴클레오타이드이고, X는 표적 DNA의 비상보적 가닥의 표적 서열의 3' 부근(immediately 3')이다. 이와 같이, 상보적 가닥의 PAM 서열은 5' -CCY-3' 일 것이며, 여기서 Y는 임의의 DNA 뉴클레오타이드이고 Y는 표적 DNA의 상보적 가닥의 표적 서열의 5' 부근이다. 이러한 일부 실시 형태에서, X와 Y는 상보적일 수 있고, X-Y 염기쌍은 임의의 염기쌍일 수 있다 (예를 들어, X=C 및 Y=G; X=G 및 Y=C; X=A 및 Y=T, X=T 및 Y=A).

[0088] Cas 단백질은 임의의 형태로 제공될 수 있다. 예를 들어, Cas 단백질은 단백질의 형태, 예컨대 gRNA와 복합체를 형성한 Cas 단백질의 형태로 제공될 수 있다. 대안적으로, Cas 단백질은 RNA(예를 들어, 메신저 RNA

(mRNA)) 또는 DNA와 같은, Cas 단백질을 암호화하는 핵산의 형태로 제공될 수 있다. 임의로, Cas 단백질을 암호화하는 핵산은 특정 세포 또는 유기체에서 단백질로의 효율적인 번역을 위해 코돈 최적화될 수 있다. 예를 들어, Cas 단백질을 암호화하는 핵산은 자연 발생 폴리뉴클레오타이드 서열과 비교하여, 박테리아 세포, 효모 세포, 인간 세포, 비인간 세포, 포유류 세포, 설치류 세포, 마우스 세포, 래트 세포 또는 관심 대상의 임의의 다른 숙주 세포에서의 사용 빈도가 높은 코돈으로 치환되도록 변형될 수 있다. Cas 단백질을 암호화하는 핵산이 세포로 도입되는 경우, Cas 단백질은 세포 내에서 일시적으로, 조건부로 또는 구성적으로 발현될 수 있다.

[0089] Cas 단백질을 암호화하는 핵산은 세포의 게놈에 안정하게 통합될 수 있으며, 세포에서 활성인 프로모터에 작동 가능하게 연결될 수 있다. 대안적으로, Cas 단백질을 암호화하는 핵산은 발현 구축물의 프로모터에 작동가능하게 연결될 수 있다. 발현 구축물은 대상으로 하는 유전자 또는 다른 핵산 서열 (예를 들어, Cas 유전자)의 발현을 지시할 수 있는 임의의 핵산 구축물을 포함하며, 대상으로 하는 그러한 핵산 서열을 표적 세포로 전달할 수 있다. 예를 들어, Cas 단백질을 암호화하는 핵산은 핵산 삽입물을 포함하는 표적화 벡터 및/또는 gRNA를 암호화하는 DNA를 포함하는 벡터 내에 있을 수 있거나, 또는 핵산 삽입물을 포함하는 표적화 벡터와는 별개이고/이거나 gRNA를 암호화하는 DNA를 포함하는 벡터와 별개인 벡터 또는 플라스미드에 존재할 수 있다. 발현 구축물에서 사용될 수 있는 프로모터는 예를 들어, 만능성 래트, 진핵생물, 포유동물, 비인간 포유류, 인간, 설치류, 마우스, 또는 햄스터 세포에서 활성인 프로모터를 포함한다. 이러한 프로모터는, 예를 들어, 조건적 프로모터, 유도성 프로모터, 구성적 프로모터 또는 조직-특이적 프로모터일 수 있다. 다른 프로모터의 예는 본 명세서의 다른 부분에 기재되어 있다.

[0090] B. 가이드 RNA (gRNA)

[0091] "가이드 RNA" 또는 "gRNA"는, Cas 단백질에 결합하고 표적 DNA 내의 특이적 위치로 Cas 단백질을 표적화하는 RNA 분자를 포함한다. 가이드 RNA (gRNA)는 2개의 세그먼트를 포함할 수 있다: "DNA-표적화 세그먼트" 및 "단백질-결합 세그먼트". "세그먼트"는 RNA의 뉴클레오타이드의 연속 스트레치와 같은, 분자의 세그먼트, 섹션 또는 영역을 포함한다. 일부 gRNA는 2개의 분리된 RNA 분자, 즉, "액티베이터(activator)-RNA"와 "타게터(targeter)-RNA"를 포함한다. 다른 gRNA는 "단일 분자 gRNA", "단일 가이드 RNA" 또는 "sgRNA"로도 지칭될 수 있는 단일 RNA 분자 (단일 RNA 폴리뉴클레오타이드)이다. 예를 들어, 각각 본 명세서에 참고로 인용되는, WO/2013/176772A1, WO/2014/065596A1, WO/2014/089290A1, WO/2014/093622A2, WO/2014/099750A2, WO/2013142578A1 및 WO 2014/131833A1을 참조한다. 용어 "가이드 RNA" 및 "gRNA"는 이중 분자 gRNA와 단일 분자 gRNA를 포함한다.

[0092] 예시적인 2분자 gRNA는 crRNA 유사 ("CRISPR RNA" 또는 "타게터-RNA" 또는 "crRNA" 또는 "crRNA 반복체") 분자 및 대응하는 tracrRNA 유사 ("트랜스 작용 CRISPR RNA" 또는 "액티베이터-RNA" 또는 "tracrRNA" 또는 "스캐폴드") 분자를 포함한다. crRNA는 gRNA의 DNA 표적화 세그먼트 (단일 가닥), 및 gRNA의 단백질 결합 세그먼트의 dsRNA 이중 가닥의 절반을 형성하는 뉴클레오타이드의 스트레치를 포함한다. 대응하는 tracrRNA (액티베이터-RNA)는 gRNA의 단백질 결합 세그먼트의 dsRNA 이중 가닥의 다른 절반을 형성하는 뉴클레오타이드의 스트레치를 포함한다. crRNA의 뉴클레오타이드의 스트레치는 tracrRNA의 뉴클레오타이드의 스트레치에 상보적이며, 이것과 혼성화하여, gRNA의 단백질 결합 도메인의 dsRNA 이중 가닥을 형성한다. 이와 같이, 각 crRNA는 대응하는 tracrRNA를 갖는다고 할 수 있다. crRNA는 추가로 단일 가닥 DNA-표적화 세그먼트를 제공한다. 따라서, gRNA는 표적 서열로 혼성화되는 서열 및 tracrRNA를 포함한다.

[0093] crRNA와 상응하는 tracrRNA (상응하는 쌍으로서)는 혼성화되어 gRNA를 형성한다. 게다가, crRNA는 CRISPR RNA 인식 서열에 혼성화하는 단일 가닥 DNA 표적화 세그먼트를 제공한다. 세포 내에서의 변형에 사용된다면, 주어진 crRNA 또는 tracrRNA 분자의 완전 서열은 RNA 분자가 사용될 종류에 특이적하도록 설계될 수 있다. 예를 들어, 각각 본 명세서에 참조로 포함되는, 문헌[Mali P *et al.* (2013) *Science* 2013 Feb 15;339(6121):823-6]; 문헌[Jinek M *et al.* *Science* 2012 Aug 17;337(6096):816-21]; 문헌[Hwang WY *et al.* *Nat Biotechnol* 2013 Mar;31(3):227-9]; 문헌[Jiang W *et al.* *Nat Biotechnol* 2013 Mar;31(3):233-9]; 및 문헌[Cong L *et al.* *Science* 2013 Feb 15;339(6121):819-23]을 참조한다.

[0094] 주어진 gRNA의 DNA 표적화 세그먼트 (crRNA)는 표적 DNA의 서열과 상보적인 뉴클레오타이드 서열을 포함한다. gRNA의 DNA 표적화 세그먼트는 혼성화를 통해 서열 특이적 방식으로 표적 DNA와 상호 작용한다 (즉, 염기쌍 형성). 이와 같이, DNA 표적화 세그먼트의 뉴클레오타이드 서열은 달라질 수 있으며, gRNA 및 표적 DNA가 상호 작용할 표적 DNA 내의 위치를 결정한다. 대상 gRNA의 DNA 표적화 세그먼트는 표적 DNA 내의 임의의 원하는 서열에 혼성화되도록 변형될 수 있다. 자연 발생적인 crRNA는 Cas9 시스템 및 유기체에 따라 다르지만, 종종 21 내

지 46개의 뉴클레오타이드 길이의 2개의 직접 반복체(DR)가 플랭킹된 21 내지 72개의 뉴클레오타이드 길이의 표적화 세그먼트를 포함한다 (예를 들어, WO2014/131833 참조함). 화농연쇄구균의 경우, DR은 36개의 뉴클레오타이드 길이이고, 표적화 세그먼트는 30개의 뉴클레오타이드 길이이다. 3' 위치의 DR은 대응하는 tracrRNA에 상보적이며, 이것과 혼성화하여, 결국 Cas9 단백질에 결합한다.

[0095] DNA 표적화 세그먼트는 약 12개의 뉴클레오타이드 내지 약 100개의 뉴클레오타이드 길이를 가질 수 있다. 예를 들어, DNA 표적화 세그먼트는 약 12개의 뉴클레오타이드(nt) 내지 약 80 nt, 약 12 nt 내지 약 50 nt, 약 12 nt 내지 약 40 nt, 약 12 nt 내지 약 30 nt, 약 12 nt 내지 약 25 nt, 약 12 nt 내지 약 20 nt, 또는 약 12 nt 내지 약 19 nt의 길이를 가질 수 있다. 대안적으로, DNA 표적화 세그먼트는 약 19 nt 내지 약 20 nt, 약 19 nt 내지 약 25 nt, 약 19 nt 내지 약 30 nt, 약 19 nt 내지 약 35 nt, 약 19 nt 내지 약 40 nt, 약 19 nt 내지 약 45 nt, 약 19 nt 내지 약 50 nt, 약 19 nt 내지 약 60 nt, 약 19 nt 내지 약 70 nt, 약 19 nt 내지 약 80 nt, 약 19 nt 내지 약 90 nt, 약 19 nt 내지 약 100 nt, 약 20 nt 내지 약 25 nt, 약 20 nt 내지 약 30 nt, 약 20 nt 내지 약 35 nt, 약 20 nt 내지 약 40 nt, 약 20 nt 내지 약 45 nt, 약 20 nt 내지 약 50 nt, 약 20 nt 내지 약 60 nt, 약 20 nt 내지 약 70 nt, 약 20 nt 내지 약 80 nt, 약 20 nt 내지 약 90 nt, 또는 약 20 nt 내지 약 100 nt의 길이를 가질 수 있다.

[0096] 표적 DNA의 뉴클레오타이드 서열 (CRISPR RNA 인식 서열)에 상보적인 DNA 표적화 세그먼트의 뉴클레오타이드 서열은 적어도 약 12 nt의 길이를 가질 수 있다. 예를 들어, DNA-표적화 서열 (예를 들어, 표적 DNA 내의 CRISPR RNA 인식 서열에 상보적인, DNA-표적화 세그먼트 내의 서열)의 길이는 약 12 nt 이상, 약 15 nt 이상, 약 18 nt 이상, 약 19 nt 이상, 약 20 nt 이상, 약 25 nt 이상, 약 30 nt 이상, 약 35 nt 이상 또는 약 40 nt 이상일 수 있다. 대안적으로, 표적 DNA의 표적 서열에 상보적인 DNA-표적화 세그먼트의 DNA-표적화 서열은 약 12 뉴클레오타이드 (nt) 내지 약 80 nt, 약 12 nt 내지 약 50 nt, 약 12 nt 내지 약 45 nt, 약 12 nt 내지 약 40 nt, 약 12 nt 내지 약 35 nt, 약 12 nt 내지 약 30 nt, 약 12 nt 내지 약 25 nt, 약 12 nt 내지 약 20 nt, 약 12 nt 내지 약 19 nt, 약 19 nt 내지 약 20 nt, 약 19 nt 내지 약 25 nt, 약 19 nt 내지 약 30 nt, 약 19 nt 내지 약 35 nt, 약 19 nt 내지 약 40 nt, 약 19 nt 내지 약 45 nt, 약 19 nt 내지 약 50 nt, 약 19 nt 내지 약 60 nt, 약 20 nt 내지 약 25 nt, 약 20 nt 내지 약 30 nt, 약 20 nt 내지 약 35 nt, 약 20 nt 내지 약 40 nt, 약 20 nt 내지 약 45 nt, 약 20 nt 내지 약 50 nt, 또는 약 20 nt 내지 약 60 nt의 길이를 가질 수 있다. 표적 DNA의 뉴클레오타이드 서열 (표적 서열)에 상보적인 DNA 표적화 세그먼트의 뉴클레오타이드 서열 (DNA 표적화 서열)은 적어도 약 12 nt의 길이를 가질 수 있다. 일부 경우에, DNA-표적화 서열은 적어도 약 20 nt의 길이를 가질 수 있다.

[0097] tracrRNA는 임의의 형태 (예를 들어, 전장 tracrRNA 또는 활성 부분 tracrRNA) 및 다양한 길이일 수 있다. 이는 일차 전사물 또는 가공된 형태를 포함할 수 있다. 예를 들어, tracrRNA (단일 가이드 RNA의 일부로서 또는 2 분자 gRNA의 일부로서 분리된 분자로서)는 야생형 tracrRNA 서열의 전부 또는 일부 (예를 들어, 야생형 tracrRNA 서열의 약 20, 26, 32, 45, 48, 54, 63, 67, 85개 또는 그 이상의 뉴클레오타이드)를 포함하거나 이로 구성될 수 있다. 화농연쇄구균 유래의 야생형 tracrRNA 서열의 예는 171-뉴클레오타이드, 89-뉴클레오타이드, 75-뉴클레오타이드 및 65-뉴클레오타이드 버전을 포함한다. 예를 들어, 각각 그 전체 내용이 본원에 참조로 포함되는, 문헌[Deltcheva *et al.* (2011) *Nature* 471:602-607]; WO 2014/093661을 참조한다. 단일 가이드 RNA(sgRNA) 내의 tracrRNA의 예로는 sgRNA의 +48, +54, +67 및 +85 버전 내에 발견되는 tracrRNA 세그먼트를 들 수 있으며, 여기서 "+n"은 야생형 tracrRNA의 + n 이하의 뉴클레오타이드가 sgRNA에 포함된다는 것을 나타낸다. 그 전체 내용이 본 명세서에 참조로 포함되는 US 8,697,359를 참조한다.

[0098] 표적 DNA 내의 DNA 표적화 서열과 CRISPR RNA 인식 서열 사이의 상보성 비율은 적어도 60% (예를 들어, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100%)일 수 있다. 표적 DNA 내의 DNA-표적화 서열과 CRISPR RNA 인식 서열 사이의 상보성 비율은 표적 DNA의 상보적 가닥의 표적 서열의 7개의 인접 5' -최말단 뉴클레오타이드(contiguous 5' -most nucleotide)에 대해 100%이다. 소정 실시 형태에서, 표적 DNA 내의 DNA-표적화 서열과 CRISPR RNA 인식 서열 사이의 상보성 비율은 약 20개의 인접 뉴클레오타이드에 대해 적어도 60%일 수 있다. 예로서, 표적 DNA 내의 DNA-표적화 서열과 CRISPR RNA 인식 서열 사이의 상보성 비율은 표적 DNA의 상보적 가닥 내의 CRISPR RNA 인식 서열의 5' -최말단에서 14개의 인접 뉴클레오타이드에 대해 100%이고 나머지에 대해 0%만큼 낮다. 이러한 경우에, DNA 표적화 서열은 길이가 14개의 뉴클레오타이드인 것으로 간주될 수 있다. 다른 예로서, 표적 DNA 내의 DNA-표적화 서열과 CRISPR RNA 인식 서열 사이의 상보성 비율은 표적 DNA의 상보적 가닥 내의 CRISPR RNA 인식 서열의 5' -최말단에서 7개의 인접 뉴클레오타이드에 대해 100%이고 나머지에 대해 0%만큼 낮다. 이러한 경우



에, DNA 표적화 서열은 길이가 7개의 뉴클레오티드인 것으로 간주될 수 있다.

- [0099] 핵산의 상보성은 하나의 핵산 가닥의 뉴클레오티드 서열이 이의 핵산 염기 그룹의 배향으로 인해, 대향하는 핵산 가닥 상의 다른 서열과 수소 결합한다는 것을 의미한다. 상보적 염기는 전형적으로 DNA에서 A와 T, C와 G이고, RNA에서, C와 G 및 U와 A이다. 상보성은 완전하거나 상당하거나/충분할 수 있다. 2개의 핵산 사이의 완전한 상보성은 2개의 핵산이, 이중나선 구조의 모든 염기가 왓슨-크릭 염기 결합에 의해 상보적 염기에 결합되는 이중 나선 구조를 형성할 수 있음을 의미한다. "상당한" 또는 "충분한" 상보성은 한 가닥의 서열이 대향하는 가닥의 서열에 전적으로 맞/또는 완전히 상보적이지 않지만, 일련의 혼성화 조건 (예를 들어, 염 농도 및 온도)에서 안정한 하이브리드 복합체를 형성하도록 두 가닥 상의 염기 사이에 충분한 결합이 일어남을 의미한다. 이러한 조건은 혼성화된 가닥의  $T_m$ 을 예측하는 서열 및 표준 수학적 계산을 이용하여 예측될 수 있거나, 일상적인 방법을 이용하여  $T_m$ 의 경험적 결정에 의해 예측될 수 있다.  $T_m$ 은 2개의 핵산 가닥 사이에 형성된 혼성화 복합체 집단이 50% 변성된 온도를 말한다.  $T_m$ 보다 낮은 온도에서는 혼성화 복합체의 형성이 촉진되는 반면에,  $T_m$ 보다 높은 온도에서는 혼성화 복합체의 가닥들의 용융 또는 분리가 촉진된다. 기타 공지된  $T_m$  계산은 핵산 구조적 특성을 고려하지만,  $T_m$ 은 예를 들어,  $T_m=81.5+0.41(\% G+C)$ 을 사용하여 1 M NaCl 수용액 중의 기지의 G+C 함량을 갖는 핵산에 대하여 추정될 수 있다.
- [0100] "혼성화 조건"은 하나의 핵산 가닥이 상보적 가닥 상호작용 및 수소 결합으로 제2 핵산 가닥에 결합하여 혼성화 복합체를 생산하는 누적 환경을 말한다. 이러한 조건은 핵산을 함유하는 수용액 또는 유기 용액의 화학 성분 및 이의 농도 (예를 들어, 염, 킬레이트제, 포름아미드), 및 혼합물의 온도를 포함한다. 다른 인자, 예컨대 배양 기간 또는 반응실 치수가 환경에 기여할 수 있다 (예를 들어, 문헌[Sambrook et al., Molecular Cloning, A Laboratory Manual, 2.sup.nd ed., pp. 1.90-1.91, 9.47-9.51, 1 1.47-11.57 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989)]).
- [0101] 혼성화는 염기 간의 미스매치가 가능하더라도, 2개의 핵산이 상보성 서열을 포함하는 것을 필요로 한다. 2개의 핵산 간의 혼성화에 적합한 조건은 본 기술 분야에 잘 알려진 변수인 핵산 길이 및 상보성 정도에 따라 다르다. 2개의 뉴클레오티드 서열 간의 상보성 정도가 클수록, 이들 서열을 갖는 핵산의 하이브리드에 대한 용융 온도 ( $T_m$ )의 값이 커진다. 짧은 스트레치의 상보성 (예를 들어, 35개 이하, 30개 이하, 25개 이하, 22개 이하, 20개 이하 또는 18개 이하의 뉴클레오티드에 대한 상보성)을 갖는 핵산 간의 혼성화의 경우, 미스매치 위치가 중요해진다 (문헌[Sambrook et al., supra, 11.7-11.8]을 참조한다). 전형적으로, 혼성화 가능한 핵산 길이는 적어도 약 10개의 뉴클레오티드로 되어 있다. 혼성화 가능한 핵산의 최소 길이의 예로는 적어도 약 15개의 뉴클레오티드, 적어도 약 20개의 뉴클레오티드, 적어도 약 22개의 뉴클레오티드, 적어도 약 25개의 뉴클레오티드 및 적어도 약 30개의 뉴클레오티드이다. 게다가, 온도 및 세정액 염 농도는 필요에 따라, 인자, 예컨대 상보성 영역의 길이 및 상보성 정도에 따라 조절될 수 있다.
- [0102] 폴리뉴클레오티드의 서열은 특이적으로 혼성화 가능하게 되는 이의 표적 핵산의 서열에 대하여 100% 상보성을 나타낼 필요는 없다. 게다가, 폴리뉴클레오티드는 개재 또는 인접 세그먼트가 혼성화 이벤트 (예를 들어, 루프 구조 또는 헤어핀 구조)에 관여하지 않도록 하나 이상의 세그먼트에 대하여 혼성화할 수 있다. 폴리뉴클레오티드 (예를 들어, gRNA)는 이들이 표적화하는 표적 핵산 서열 내의 표적 부위에 대하여, 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 99% 또는 100% 서열 상보성을 포함할 수 있다. 예를 들어, 20개 중 18개의 gRNA의 뉴클레오티드가 표적 부위에 상보적이므로, 특이적으로 혼성화되는 gRNA는 90% 상보성을 나타낼 것이다. 이러한 예에서, 나머지 비상보적 뉴클레오티드는 상보적 뉴클레오티드와 클러스터를 이루거나 그 사이에 상보적 뉴클레오티드가 배치될 수 있으며, 서로 인접하거나 상보적 뉴클레오티드에 인접할 필요는 없다. 핵산 내의 핵산 서열의 특정 스트레치 간의 상보성 비율은 통상, 본 기술 분야에 알려진 BLAST 프로그램 (기본 국소 정렬 검색 도구) 및 PowerBLAST 프로그램 (문헌[Altschul et al., J. Mol. Biol., 1990, 215, 403-410]; 문헌[Zhang and Madden, Genome Res., 1997, 7, 649-656])을 사용하거나, 스미스 및 와트만의 알고리즘 (문헌[Adv. Appl. Math., 1981, 2, 482-489])을 사용하는 디폴트 설정을 이용한 Gap 프로그램 (위스콘신 서열 분석 패키지, 유닉스 용 버전 8, 제네텍스 컴퓨터 그룹, 유니버시티 리서치 파크, 위스콘신주 메디슨 소재)을 사용하여 측정될 수 있다.
- [0103] 대상 gRNA의 단백질-결합 세그먼트는 Cas 단백질과 상호작용한다. 대상 gRNA는 결합된 폴리펩티드를 DNA-표적화 세그먼트를 통해 표적 DNA 내의 특정 뉴클레오티드 서열로 유도한다. 대상 gRNA의 단백질-결합 세그먼트는 서로 상보적인 2개의 뉴클레오티드의 스트레치를 포함할 수 있다. 단백질 결합 세그먼트의 상보적 뉴클레오티드는 혼성화되어, 이중 가닥 RNA 이중 가닥(dsRNA)를 형성한다. 대상 gRNA의 단백질-결합 세그먼트는 Cas 단백질과 상호작용하고, gRNA는 결합된 Cas 단백질을 DNA-표적화 세그먼트를 통해 표적 DNA 내의 특정 뉴클레오티드

서열로 유도한다.

- [0104] 소정 실시 형태에서, 본 명세서에 기재된 바와 같은 gRNA는 2개의 별개의 RNA 분자를 포함한다. 대상 gRNA의 2개의 RNA 분자 각각은 2개의 RNA 분자의 상보적 뉴클레오티드가 혼성화하여 단백질-결합 세그먼트의 이중 가닥 RNA 이중 가닥 (예를 들어, 헤어핀)을 형성하도록 서로 상보적인 뉴클레오티드의 스트레치를 포함한다. 대상 gRNA는 임의의 상응하는 crRNA와 tracrRNA 쌍을 포함할 수 있다. 본 명세서에 기재된 방법에서, gRNA는 crRNA와 tracrRNA의 복합체 (예를 들어 gRNA-Cas 복합체)로서 사용될 수 있거나, crRNA 및 상응하는 tracrRNA는 별도로 전달될 수 있다. 예를 들어, 다수의 gRNA가 절단 반응에 사용되는 경우, 각각의 표적 부위에 특이적인 각각의 crRNA는 각각의 crRNA와 복합체를 형성할 수 있는 표준 tracrRNA와는 별도로 전달될 수 있다. 이러한 방법에서, crRNA는 표준 tracrRNA와 복합체를 형성하여 Cas 단백질을 표적 부위로 유도할 수 있다.
- [0105] 가이드 RNA는 추가의 바람직한 특징(예를 들어, 변형되거나 조절된 안정성; 세포내 표적화; 형광 표지를 이용한 추적; 단백질 또는 단백질 복합체에 대한 결합 부위 등)을 제공하는 변형 또는 서열을 포함할 수 있다. 이러한 변형의 비제한적인 예는, 예를 들어, 5' 캡 (예를 들어, 7-메틸구아닐레이트 캡 (m7G)); 3' 폴리아데닐화된 꼬리 (즉, 3' 폴리(A) 꼬리); (예를 들어, 단백질 및/또는 단백질 복합체에 의한 조절된 안정성 및/또는 조절된 접근성을 허용하기 위한) 리보스위치(riboswitch) 서열; 안정성 제어 서열; dsRNA 이중 가닥 (즉, 헤어핀)을 형성하는 서열; RNA를 세포내 부위 (예를 들어, 핵, 미토콘드리아, 엽록체 등)로 표적화시키는 변형 또는 서열; 추적 (예를 들어, 형광 분자에 대한 직접 접합(conjugation), 형광 검출을 촉진하는 모이어티에 대한 접합, 형광 검출을 허용하는 서열 등)을 제공하는 변형 또는 서열; 단백질 (예를 들어, 전사 활성화인자, 전사 억제인자, DNA 메틸트랜스퍼라제, DNA 데메틸라제, 히스톤 아세틸트랜스퍼라제, 히스톤 데아세틸라제 등을 포함하는, DNA 상에서 작용하는 단백질)에 대한 결합 부위를 제공하는 변형 또는 서열; 및 이들의 조합을 포함한다.
- [0106] 가이드 RNA는 임의의 형태로 제공될 수 있다. 예를 들어, gRNA는 2개의 분자 (별개의 crRNA 및 tracrRNA) 또는 1개의 분자 (sgRNA)로서의 RNA 형태 및 임의로 Cas 단백질과의 복합체 형태로 제공될 수 있다. gRNA는 또한 RNA를 암호화하는 DNA의 형태로 제공될 수 있다. gRNA를 암호화하는 DNA는 단일 RNA 분자(sgRNA) 또는 분리된 RNA 분자(예를 들어, 분리된 crRNA 및 tracrRNA)를 암호화할 수 있다. 후자의 경우, gRNA를 암호화하는 DNA는 각각, crRNA 및 tracrRNA를 암호화하는 분리된 DNA 분자로서 제공될 수 있다.
- [0107] gRNA를 암호화하는 DNA는 세포의 게놈에 안정하게 통합될 수 있으며, 세포에서 활성인 프로모터에 작동가능하게 연결될 수 있다. 대안적으로, gRNA를 암호화하는 DNA는 발현 구축물의 프로모터에 작동가능하게 연결될 수 있다. 예를 들어, gRNA를 암호화하는 DNA는 핵산 삽입물을 포함하는 표적화 벡터 및/또는 Cas 단백질을 암호화하는 핵산을 포함하는 벡터 내에 있을 수 있거나, 또는 핵산 삽입물을 포함하는 표적화 벡터와는 별개이고/이거나 Cas 단백질을 암호화하는 핵산을 포함하는 벡터와 별개인 벡터 또는 플라스미드에 존재할 수 있다. 이러한 프로모터는 예를 들어, 만능성 래트, 진핵생물, 포유동물, 비인간 포유류, 인간, 설치류, 마우스, 또는 햄스터 세포에서 활성을 나타낼 수 있다. 이러한 프로모터는, 예를 들어, 조건적 프로모터, 유도성 프로모터, 구성적 프로모터 또는 조직-특이적 프로모터일 수 있다. 경우에 따라서는, 프로모터는 RNA 폴리머라아제 III 프로모터, 예를 들어 인간 U6 프로모터, 래트 U6 폴리머라아제 III 프로모터 또는 마우스 U6 폴리머라아제 III 프로모터이다. 다른 프로모터의 예는 본 명세서의 다른 부분에 기재되어 있다. gRNA를 암호화하는 DNA가 세포로 도입되는 경우, gRNA는 세포 내에서 일시적으로, 조건부로 또는 구성적으로 발현될 수 있다.
- [0108] 대안적으로, gRNA는 다양한 다른 방법으로 제조될 수 있다. 예를 들어, gRNA는 예를 들어, T7 RNA 폴리머라아제를 사용하여, 시험관내 전사에 의해 제조될 수 있다 (예를 들어, WO 2014/089290 및 WO 2014/065596를 참조함). 가이드 RNA는 화학 합성에 의해 제조되는 합성적으로 생산된 분자일 수도 있다.
- [0109] C. CRISPR RNA 인식 서열
- [0110] 용어 "CRISPR RNA 인식 서열"은 결합을 위한 충분한 조건이 존재한다면, gRNA의 DNA 표적화 세그먼트가 결합할 표적 DNA에 존재하는 핵산 서열을 포함한다. 예를 들어, CRISPR RNA 인식 서열은 가이드 RNA가 상보성을 갖도록 설계된 서열을 포함하며, 여기서 CRISPR RNA 인식 서열과 DNA 표적 서열 사이의 혼성화는 CRISPR 복합체의 형성을 촉진시킨다. 혼성화를 일으켜 CRISPR 복합체의 형성을 촉진시키기에 충분한 상보성이 있다면, 완전 상보성이 반드시 필요한 것은 아니다. CRISPR RNA 인식 서열은 또한 하기에서 보다 상세히 설명되는, Cas 단백질에 대한 절단 부위를 포함한다. CRISPR RNA 인식 서열은, 예를 들어, 세포의 세포핵 또는 세포질에, 또는 미토콘드리아 또는 엽록체와 같은 세포 소기관 내에 위치할 수 있는 임의의 폴리뉴클레오티드를 포함할 수 있다.
- [0111] 표적 DNA 내의 CRISPR RNA 인식 서열은 Cas 단백질 또는 gRNA에 의해 표적화될 수 있다(즉, 결합되거나 혼성화

되거나 상보적일 수 있다). 적합한 DNA/RNA 결합 조건은 보통 세포 내에 존재하는 생리학적 조건을 포함한다. 다른 적합한 DNA/RNA 결합 조건 (예를 들어, 무세포계에서의 조건)이 본 기술 분야에 알려져 있다 (예를 들어, 문헌[Molecular Cloning: A Laboratory Manual, 3rd Ed. (Sambrook *et al.*, Harbor Laboratory Press 2001)]을 참조한다). Cas 단백질 또는 gRNA와 상보적이고 혼성화되는 표적 DNA의 가닥은 "상보적 가닥"으로 명명될 수 있고, "상보적 가닥"과 상보적인 표적 DNA의 가닥 (따라서 Cas 단백질 또는 gRNA에 상보적인 것은 아니다)은 "비상보적 가닥" 또는 "주형 가닥"으로 명명될 수 있다.

[0112] Cas 단백질은 gRNA의 DNA 표적화 세그먼트가 결합할 표적 DNA에 존재하는 핵산 서열의 내부 또는 외부의 위치에서 핵산을 절단할 수 있다. "절단 부위"는 Cas 단백질이 단일 가닥 절단 또는 이중 가닥 절단을 일으키는 핵산의 위치를 포함한다. 예를 들어, CRISPR 복합체 (CRISPR RNA 인식 서열에 혼성화되어, Cas 단백질과 복합체를 형성한 gRNA를 포함함)의 형성은 gRNA의 DNA 표적화 세그먼트가 결합될 표적 DNA 내에 존재하는 핵산 서열 또는 그 부근에 (예를 들어, 상기 핵산 서열로부터 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50개 또는 그 이상의 염기쌍 내에) 한 가닥 또는 두 가닥의 절단을 일으킬 수 있다. 절단 부위가 gRNA의 DNA 표적화 세그먼트가 결합할 핵산 서열 외부에 있다면, 절단 부위는 여전히 "CRISPR RNA 인식 서열" 내에 있는 것으로 간주된다. 절단 부위는 핵산의 한 가닥에만 또는 두 가닥에 있을 수 있다. 절단 부위는 핵산의 두 가닥에서 동일한 위치에 있을 수 있거나(평행 말단을 생성함), 각 가닥에 상이한 부위에 있을 수 있다(스태거된(staggered) 말단을 생성함). 스태거된 말단은 예를 들어, 2개의 Cas 단백질을 사용하여 생성될 수 있으며, 각각은 각 가닥의 상이한 절단 부위에서 단일 가닥 절단을 일으켜 이중 가닥 절단을 일으킨다. 예를 들어, 제1 니카아제는 이중 가닥 DNA (dsDNA)의 제1 가닥에 단일 가닥 절단을 일으킬 수 있고, 제2 니카아제는 돌출 서열이 생성되도록 dsDNA의 제2 가닥에 단일 가닥 절단을 일으킬 수 있다. 경우에 따라서는, 제1 가닥의 니카아제의 CRISPR RNA 인식 서열은 제2 가닥의 니카아제의 CRISPR RNA 인식 서열로부터 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 75, 100, 250, 500 또는 1,000개의 염기쌍으로 분리되어 있다.

[0113] Cas9에 의한 표적 DNA의 부위 특이적 절단은 표적 DNA에서, (i) gRNA와 표적 DNA 사이의 염기쌍 형성 상보성과 (ii) 프로토스페이서 인접 모티프 (PAM)로 지칭되는 짧은 모티프에 의해 결정되는 위치에서 발생할 수 있다. PAM은 CRISPR RNA 인식 서열을 플랭킹할 수 있다. 임의로, CRISPR RNA 인식 서열은 PAM이 플랭킹될 수 있다. 예를 들어, Cas9의 절단 부위는 PAM 서열의 상류 또는 하류에 약 1 내지 약 10개, 또는 약 2 내지 약 5개의 염기쌍(예를 들어, 3개의 염기쌍)으로 될 수 있다. 경우에 따라서는(예를 들어, 화농연쇄구균 유래의 Cas9 또는 근연종의 Cas9가 사용되는 경우), 비상보적 가닥의 PAM 서열은 5' -N<sub>1</sub>GG-3' 수 있으며, 여기서 N<sub>1</sub>은 임의의 DNA 뉴클레오티드이고, 표적 DNA의 비상보적 가닥의 CRISPR RNA 인식 서열의 3' 바로 옆에 위치한다. 이와 같이, 상보적 가닥의 PAM 서열은 5' -CC N<sub>2</sub>-3' 일 것이며, 여기서 N<sub>2</sub>는 임의의 DNA 뉴클레오티드이고, 표적 DNA의 상보적 가닥의 CRISPR RNA 인식 서열의 5' 바로 옆에 위치한다. 일부 이러한 경우에는, N<sub>1</sub> 및 N<sub>2</sub>는 상보적일 수 있고, N<sub>1</sub>-N<sub>2</sub> 염기쌍은 임의의 염기쌍일 수 있다 (예를 들어, N<sub>1</sub>=C 및 N<sub>2</sub>=G; N<sub>1</sub>=G 및 N<sub>2</sub>=C; N<sub>1</sub>=A 및 N<sub>2</sub>=T, N<sub>1</sub>=T 및 N<sub>2</sub>=A).

[0114] CRISPR RNA 인식 서열의 예로는 gRNA의 DNA 표적화 세그먼트에 상보적인 DNA 서열, 또는 PAM 서열 이외의 이러한 DNA 서열을 포함한다. 예를 들어, 표적 모티프는 Cas 단백질에 의해 인식되는 NGG 모티프 직전의 20-뉴클레오티드 DNA 서열, 예컨대 GN<sub>19</sub>NGG (서열 번호: 8) 또는 N<sub>20</sub>NGG (서열 번호: 24)일 수 있다 (예를 들어, WO 2014/165825를 참조한다). 5' 말단의 구아닌은 세포에서의 RNA 폴리머라아제에 의한 전사를 촉진시킬 수 있다. CRISPR RNA 인식 서열의 다른 예는 시험관내에서의 T7 폴리머라아제에 의한 효율적인 전사를 용이하게 하기 위해 5' 말단에 2개의 구아닌 뉴클레오티드(예를 들어, GGN<sub>20</sub>NGG; 서열 번호 25)를 포함할 수 있다. 예를 들어, WO 2014/065596을 참조한다. 다른 CRISPR RNA 인식 서열은 5' G 또는 GG 및 3' GG 또는 NGG를 포함하는 서열 번호 8, 24 및 25의 길이가 4 내지 22개인 뉴클레오티드를 가질 수 있다. 또 다른 CRISPR RNA 인식 서열은 서열 번호 8, 24 및 25의 길이가 14 내지 22개인 뉴클레오티드를 가질 수 있다.

[0115] CRISPR RNA 인식 서열은 세포에 대하여 내인성이거나 외인성인 임의의 핵산 서열일 수 있다. CRISPR RNA 인식 서열은 유전자 산물(예를 들어, 단백질)을 암호화하는 서열 또는 비암호화 서열(예를 들어, 조절 서열)일 수 있거나, 두 가지를 포함할 수 있다.

[0116] 일 실시 형태에서, Cas 단백질은 타입 I Cas 단백질이다. 일 실시 형태에서, Cas 단백질은 타입 II Cas 단백질이다. 일 실시 형태에서, 타입 II Cas 단백질은 Cas9이다. 일 실시 형태에서, 제1 핵산 서열은 인간 코돈 최적화된 Cas 단백질을 암호화한다.



- [0117] 일 실시 형태에서, gRNA는 crRNA와 a tracrRNA를 암호화하는 핵산 서열을 포함한다. 구체적인 실시 형태에서, Cas 단백질은 Cas9이다. 일부 실시 형태에서, gRNA는 (a) 핵산 서열 5' -GUUUUAGAGCUAGAAAUAGCAAGUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3' (서열 번호: 1)의 키메라 RNA; 또는 (b) 핵산 서열 5' -GUUUUAGAGCUAGAAAUAGCAAGUAAAAU AAGGCUAGUCCG-3' (서열 번호: 2)의 키메라 RNA를 포함한다. 다른 실시 형태에서, crRNA는 5' -GUUUUAGAGCUAGAAAUAGCAAGUAAAAU-3' (서열 번호: 3); 5' -GUUUUAGAGCUAGAAAUAGCAAGUAAAAU AAG (서열 번호: 4); 또는 5' -GAGUCCGAGCAGAAGAAGAAGUUUA-3' (서열 번호: 5)을 포함한다. 또 다른 실시 형태에서, tracrRNA는 5' -AAGGCUAGUCCG-3' (서열 번호: 6) 또는 5' -AAGGCUAGUCCGU UAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3' (서열 번호: 7)을 포함한다.
- [0118] V. 폴리뉴클레오티드의 조립
- [0119] 본 명세서에 개시된 방법은 실질적으로 온전하거나 끊김이 없는 이중 가닥 DNA 분자를 형성하도록 DNA 분자를 결합시키기 효과적인 조건 하에서 2개 이상의 핵산을 조립할 수 있다. 중첩 서열을 갖는 관심 대상의 임의의 핵산은 본 명세서에 개시된 방법에 따라 조립될 수 있다. 예를 들어, 자연 발생된 DNA, 클로닝된 DNA 분자, 합성으로 생성된 DNA 등을 포함하는, 중첩 서열을 갖는 관심 대상의 임의의 DNA 분자가 조립될 수 있다. 결합된 DNA 분자는, 원한다면, 본 발명의 방법을 사용하여 벡터로 클로닝될 수 있다 (예를 들어, 삽입). 2개의 핵산을 조립하는 것은 2개의 핵산의 가닥을 결합시키는 임의의 방법을 포함한다. 예를 들어, 조립은 각각의 핵산의 가닥이 다른 핵산과 어닐링하고 연장(여기서, 각각의 가닥은 다른 핵산의 연장을 위한 주형으로서 역할한다)하도록 분해된 핵산을 결합시키는 것을 포함한다.
- [0120] 일부 실시 형태에서, 각각의 핵산이 함께 직접적으로 조립되는 대신에 조이너 올리고에 조립되도록 핵산은 조이너 올리고와 조립된다. 조이너 올리고를 사용한 조립은 조립되는 핵산의 일부가 아니지만 조이너 올리고의 일부인 조립되는 핵산 사이에 핵산 염기를 배치할 수 있다. 따라서, 핵산 사이에 여분의 염기가 남아 있는 경우라도 핵산은 성공적으로 조립될 수 있다. 대안적으로, 조이너 올리고는 끊김 없는 조립을 위해 사용될 수 있으며, 여기서 조립되는 핵산 사이에 여분의 염기가 남아 있지 않다.
- [0121] 일부 실시 형태에서, 핵산은 Cas 단백질, 제한 효소 (제한 엔도뉴클레아제) (예를 들어, 본 명세서의 다른 곳에 기재된 임의의 다양한 제한 엔도뉴클레아제), 메가뉴클레아제 (예를 들어, 본 명세서의 다른 곳에 기재된 임의의 다양한 메가뉴클레아제) 또는 이들의 임의의 조합을 사용한 절단으로 조립을 위해 제조될 수 있다. 예를 들어, 조립되는 핵산 중 하나는 Cas 단백질로 절단될 수 있고, 조립되는 다른 핵산은 Cas 단백질, 제한 효소, 메가뉴클레아제 또는 이들의 임의의 조합으로 절단될 수 있다. 뉴클레아제로 절단 한 후, 분해된 핵산은 중첩 말단 서열을 갖는 다른 분해된 핵산에 직접 조립되거나, 분해되지 않았지만 중첩 말단 서열을 갖는 핵산에 조립될 수 있다. 분해된 핵산은 또한 조이너 올리고를 사용하여 다른 핵산에 조립될 수 있다.
- [0122] 뉴클레아제 제제 (예를 들어, Cas 단백질)를 사용하여 2개의 핵산 분자 사이에 중첩 말단 서열을 생성하는 실시 형태에서, 신속하고 조합적인 방법을 사용하여, 분해된 핵산을 조립할 수 있다. 예를 들어, 중첩하는 말단을 갖는 제1 핵산과 제2 핵산을 리가아제, 엑소뉴클레아제, DNA 폴리머라아제 및 뉴클레오티드와 조합하고 50℃에서와 같은 일정한 온도에서 인큐베이션할 수 있다. 구체적으로, T5 엑소뉴클레아제를 사용하여 상보적 돌출부를 생성하는 dsDNA의 5' 말단으로부터 뉴클레오티드를 제거할 수 있다. 이어서, 50℃에서 상보적 단일 가닥 DNA 돌출부를 어닐링하고, DNA 폴리머라아제를 사용하여 간극을 채우고 Taq DNA 리가아제를 사용하여 결과로서 얻어진 틈을 밀봉할 수 있다. 따라서, 중첩 말단 서열을 공유하는 2개의 핵산은 1 단계 등은 반응에서 공유 결합된 분자로 결합될 수 있다. 예를 들어, 전체적으로 본 명세서에 참고로 포함되는 문헌[Gibson, *et al.* (2009) *Nature Methods* 6(5): 343-345]을 참조한다. 일부 실시 형태에서, 프로테이나아제 K 또는 페놀/클로로포름/아이소아밀알코올 (PCI) 정제를 사용하여 반응 혼합물로부터 뉴클레아제 제제 (예를 들어, Cas 단백질)를 제거한다. 일부 실시 형태에서, 뉴클레아제 제제 (예를 들어, Cas 단백질)는 실리카 겔-기반 컬럼 정제로 반응 혼합물로부터 제거될 수 있다.
- [0123] 소정 실시 형태에서, 본 명세서에 개시된 방법은 선형 폴리뉴클레오티드를 사용하여 벡터를 조립한다. 다른 실시 형태에서, 본 명세서에 개시된 방법은 2개 이상의 벡터, 예컨대 2개의 BAC 벡터를 조립한다. 용어 "BAC 벡터"는 임의의 박테리아 인공 염색체를 포함한다. 구체적인 실시 형태에서, BAC는 선형 핵산 또는 다른 벡터, 예를 들어, 또 다른 BAC의 뉴클레오티드 서열 영역과 중첩되는 뉴클레오티드 서열을 갖는 영역을 함유하도록 변형된다.
- [0124] 제1 및 제2 단일 가닥 핵산은 각각의 말단이 서로 상보적인 경우 중첩하는 말단을 갖는다. 제1 및 제2 이중 가닥 핵산은 제1 핵산 가닥의 5' 말단이 제2 핵산 가닥의 3' 말단에 상보적인 경우 중첩하는 말단을 가지며, 그



반대도 마찬가지이다. 예를 들어, 이중 가닥 중첩 말단 서열에 대해, 하나의 핵산 가닥은 상응하는 다른 핵산 가닥과 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성을 가질 수 있다. 본 명세서에 개시된 방법에서, 조립되는 dsDNA 분자의 가닥의 5' 말단은 다른 dsDNA 분자의 가닥의 3' 말단과 중첩 말단 서열을 공유한다. 용어 "중첩 말단 서열"은 dsDNA 분자의 두 가닥을 모두 포함한다. 따라서, 중첩 서열의 상보적 영역이 조립되는 두 폴리뉴클레오티드의 5' 및 3' 말단으로부터 단일 가닥 돌출부로 제공되는 경우, 중첩 영역의 하나의 가닥은 이의 상보적 가닥에 특이적으로 혼성화될 수 있다. 일부 실시 형태에서, 엑소뉴클레아제를 사용하여, 5' 또는 3' 말단으로부터 뉴클레오티드를 제거하여 돌출 말단 서열을 생성한다. 일부 실시 형태에서, 제1 및/또는 제2 핵산의 중첩 영역은 Cas 단백질로 분해되기 전까지는 5' 또는 3' 말단 상에 존재하지 않는다. 즉, 중첩 영역은 내부 영역일 수 있으며, 이는 이어서 Cas 단백질로 내부 중첩 영역을 함유하는 핵산(들)을 분해한 후 중첩 말단 서열로 전환된다. Cas 단백질은 중첩 영역 내의 또는 중첩 영역 외부의 표적 부위 (예를 들어, 절단 부위)에서 절단할 수 있다.

[0125] 중첩 영역의 길이는 바람직하게는 상기 영역이 조립되는 임의의 핵산 내에서 한 번만 발생하도록 충분한 길이를 갖는다. 이러한 방식으로, 다른 폴리뉴클레오티드는 말단 서열과 어닐링되는 것이 방지되고, 조립은 표적 핵산에 특이적일 수 있다. 중첩 영역의 길이는 최소 약 10 염기쌍 (bp) 내지 약 300 bp 이상으로 달라질 수 있다. 일반적으로, 중첩의 길이는 조립되는 폴리뉴클레오티드의 대략의 크기보다 작거나 같지만, 약 10bp 이상 1000bp 이하인 것이 바람직하다. 2 또는 3개의 폴리뉴클레오티드의 결합을 위해서는 약 20 내지 30 bp의 중첩이 충분할 수 있다. 10 초과의 단편을 위해서는 약 80 bp 내지 약 300 bp의 중첩이 바람직하다. 일 실시 형태에서, 중첩 영역은 합성 방법, 예를 들어, 약 40 bp에 의해 용이하게 생성될 수 있는 길이를 갖는다. 구체적인 실시 형태에서, 중첩 영역의 길이는 약 20 내지 200 bp일 수 있다. 중첩의 길이는 약 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950 또는 1,000 bp일 수 있다. 일부 실시 형태에서, 중첩 영역의 길이 20 내지 200 bp이다. 본 명세서에 개시된 방법의 구체적인 실시 형태에서, 2개 이상의 폴리뉴클레오티드가 조립될 수 있으며, 여기서 적어도 하나의 폴리뉴클레오티드의 중첩 영역은 뉴클레아제 제제 (예를 들어, gRNA-Cas 복합체)와의 접촉에 의해 생성된다. 예를 들어, 제1 폴리뉴클레오티드의 엔도뉴클레아제 분해는 제2 폴리뉴클레오티드의 말단 서열과 중첩하는 서열을 생성할 수 있으며, 이어서 중첩 말단 서열이 조립된다.

[0126] 본 명세서에 개시된 방법에서, 중첩 서열을 엑소뉴클레아제와 접촉시켜, 중첩 서열 사이의 상보적 서열 (예를 들어, 상보적 단일 가닥 서열)을 노출시킬 수 있다. 엑소뉴클레아제 분해는 상보성의 노출된 단일 가닥 영역의 특이적인 어닐링을 허용하기에 충분한 수의 뉴클레오티드를 제거 ("쉼백(chew back)")하기에 효과적인 조건 하에서 수행된다. 일반적으로, 중첩 영역의 일부 또는 중첩 영역의 전체가 쉼백되어, 중첩 영역의 일부 또는 중첩 영역의 전체를 포함하는 돌출부를 남긴다. 일부 방법에서, 엑소뉴클레아제 분해는 dNTP (예를 들어, T5 DNA 폴리머라아제)의 부재 하에 폴리머라아제에 의해 수행될 수 있는 반면, 다른 방법에서, 엑소뉴클레아제 분해는 폴리머라아제 (예를 들어, 엑소뉴클레아제 III) 활성이 결핍된 dNTP의 존재 하에 엑소뉴클레아제에 의해 수행될 수 있다.

[0127] 임의의 다양한 5' 에서 3' 의 이중 가닥 특이적 엑소데옥시리보뉴클레아제가 본 명세서에 개시된 방법에서 핵산의 말단을 쉼백하는 데 사용될 수 있다. 용어 "5' 엑소뉴클레아제"는 때때로 5' 에서 3' 엑소데옥시리보뉴클레아제를 지칭하기 위해 본 명세서에서 사용된다. 본 명세서에서 사용되는 바와 같은 "비-진행적" 엑소뉴클레아제는 각각의 DNA 결합 이벤트 동안 제한된 수의 (예를 들어, 불과 몇 안되는) 뉴클레오티드를 분해하는 엑소뉴클레아제이다. 5' 엑소뉴클레아제를 사용한 분해는 DNA 분자에 3' 단일 가닥 돌출부를 생성한다. 무엇보다도 5' 엑소뉴클레아제에 바람직한 특성은, 이것이 3' 엑소뉴클레아제 활성이 결핍되고, 5' 포스페이트 말단을 생성하며, 5' -인산화 말단 및 비인산화 말단 모두로부터 분해를 개시한다는 것이다. 이것이 평활 말단이든지 적은 5' 또는 3' 오목 말단(recessed end)을 가지든지 간에, 효소가 분자의 5' 말단으로부터 분해를 개시할 수 있다는 것이 또한 바람직하다. 적합한 엑소뉴클레아제가 숙련된 작업자에게 명백할 것이다. 이는, 예를 들어, 파지 T5 엑소뉴클레아제 (파지 T5 유전자 D15 생성물), 파지 람다 엑소뉴클레아제, Rac 프로파지의 RecE, 대장균 유래의 엑소뉴클레아제 VIII, 파지 T7 엑소뉴클레아제 (파지 T7 유전자 6 생성물) 또는 상동 재조합 반응에 관여하는 임의의 다양한 5' 엑소뉴클레아제를 포함한다. 본 발명의 일 실시 형태에서, 엑소뉴클레아제는 T5 엑소뉴클레아제 또는 람다 엑소뉴클레아제이다. 다른 실시 형태에서, 엑소뉴클레아제는 T5 엑소뉴클레아제이다. 다른 실시 형태에서, 엑소뉴클레아제는 파지 T7 엑소뉴클레아제가 아니다. 발명의 방법에 사용되는 엑소뉴클레아제 및 다른 효소를 제조하고 사용하는 방법은 통상적이며; 많은 것들이 USB 코포레이션 (26111 Miles Road, Cleveland, Ohio 44128) 또는 뉴 잉글랜드 바이오랩스, 인코포레이티드(New England Biolabs, Inc. (NEB); 240

County Road, Ipswich, Mass. 01938-2723)와 같은 상업적 공급처로부터 입수가 가능하다.

- [0128] 특히, 중첩 영역이 매우 긴 실시 형태에서, 이와 같이 생성된 단일 가닥 돌출부가 반응 조건 하에서 특이적으로 어닐링하기 충분한 길이 및 염기 함량인 한, (예를 들어, 중첩 영역의 절반보다 많은) 영역의 일부를 추백하는 것만이 필요할 수 있다. 용어 "특이적인 어닐링"은 특정 한쌍의 단일 가닥 돌출부가 반응 혼합물 중에 존재하는 다른 단일 가닥 돌출부 (예를 들어, 비상보적 돌출부)보다는 서로 우선적으로 (또는 독점적으로) 어닐링하는 상황을 포함한다. "우선적으로"는 적어도 약 95%의 돌출부가 상보적 돌출부에 어닐링한다는 것을 의미한다. 숙련된 작업자는 주어진 반응 조건 세트 하에서 관심 서열의 특이적인 어닐링을 달성하기 위한 최적 길이를 용이하게 결정할 수 있다. 일반적으로, 중첩의 상동성 영역 (단일 가닥 돌출부 또는 그들의 상보 가닥)은 동일한 서열을 함유한다. 그러나, 단일 가닥 돌출부가 반응 조건 하에서 특이적으로 어닐링될 수 있는 한, 부분적으로 동일한 서열이 사용될 수 있다.
- [0129] 소정 실시 형태에서, 뉴클레아제 제제 (예를 들어, Cas 단백질)는 dsDNA의 두 가닥을 절단하지 않고 표적 부위에서 단일 가닥 절단 (즉, "틈")을 생성할 수 있다. "닉카아제(nickase)"는 dsDNA에 틈을 생성하는 뉴클레아제 제제 (예를 들어, Cas 단백질)를 포함한다. 이러한 방식으로, dsDNA의 각각의 가닥 상의 표적 부위에 특이적인 2개의 별개의 뉴클레아제 제제 (예를 들어, Cas 단백질) (예를 들어, 닉카아제)는 다른 핵산 상의 돌출 서열 또는 동일한 핵산 상의 별개의 영역에 상보적인 돌출 서열을 생성할 수 있다. dsDNA의 두 가닥 상의 표적 부위에 특이적인 2개의 닉카아제와 핵산을 접촉시켜 생성된 돌출 말단은 5' 또는 3' 돌출 말단일 수 있다. 예를 들어, 제1 닉카아제는 dsDNA의 제1 가닥 상에 단일 가닥 절단을 생성할 수 있는 반면, 제2 닉카아제는 dsDNA의 제2 가닥 상에 단일 가닥 절단을 생성할 수 있어, 돌출 서열이 생성된다. 단일 가닥 절단을 생성하는 각각의 닉카아제의 표적 부위는 생성된 돌출 말단 서열이 제2 핵산 상의 돌출 말단 서열에 상보적이며 선택될 수 있다. 따라서, 제1 핵산과 제2 핵산의 상보적 돌출 말단은 본 명세서에 개시된 방법으로 어닐링될 수 있다. 일부 실시 형태에서, 제1 가닥 상의 닉카아제의 표적 부위는 제2 가닥 상의 닉카아제의 표적 부위와 상이하다. dsDNA의 별개의 가닥 상의 상이한 표적 부위는 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 75, 100, 250, 500 또는 1,000 염기쌍에 의해 분리된 단일 가닥 절단을 초래한다.
- [0130] 소정 실시 형태에서, 제2 핵산은 또한 제2 핵산 상의 제1 표적 부위에서 틈을 생성하는 제1 닉카아제 및 제2 핵산 분자 상의 제2 표적 부위에서 틈을 생성하는 닉카아제와 접촉한다. 제2 핵산 상의 2개의 상이한 부위에서의 틈에 의해 생성된 돌출 말단 서열은 제1 핵산 상의 2개의 상이한 부위에서의 틈에 의해 생성된 돌출 말단 서열에 상보적이어서, 상보적 돌출 말단 서열이 어닐링될 수 있다.
- [0131] 일부 실시 형태에서, 관심 유전자의 핵산 서열은 둘 이상의 BAC에 걸쳐있다. 이러한 경우에, 본 명세서에 제공된 방법을 사용하여, 특이적으로 설계된 뉴클레아제 제제는 둘 이상의 BAC를 원하는 위치에서 절단할 수 있어, 결과로 얻어진 핵산 단편은 함께 결합하여 관심 유전자의 서열을 형성한다.
- [0132] 일부 실시 형태에서, 제1 핵산의 두 가닥 상의 상이한 표적 부위에서 틈에 의해 생성된 돌출 말단은 제2 핵산의 두 가닥 상의 상이한 표적 부위에서 틈에 의해 생성된 돌출 말단에 상보적이지 않다. 다른 실시 형태에서, 조립되는 핵산은 상보적 말단을 가지지 않아, 비상보적 말단을 조립하는데 별도의 핵산이 필요하다. 조이너 올리고를 사용하여 2개의 핵산의 비상보적 말단을 결합시킬 수 있다. "조이너 올리고"는 상이한 폴리뉴클레오티드 또는 핵산의 말단에 상보적 서열을 갖는 폴리뉴클레오티드 또는 핵산을 포함하는 상보적 암(arm)을 포함한다. 일부 실시 형태에서, 조이너 올리고는 5' 말단 중앙 부분 (스페이서)에서 제1 핵산에 상보적인 암 및 3' 말단에서 제2 핵산에 상보적인 암을 갖는다. 따라서, 서로 비상보적인 말단 서열을 갖는 핵산은 각각의 핵산을 엑소뉴클레아제 처리한 후에 동일한 조이너 올리고에 어닐링하여 조립될 수 있다. 구체적인 실시 형태에서, 조이너 올리고는 제1 분해된 핵산의 5' 또는 3' 말단 서열에 상보적인 제1 암 및 제2 분해된 핵산의 5' 또는 3' 서열에 상보적인 제2 암을 갖는다. 조이너 올리고는 평활이거나 5' 또는 3' 돌출 서열을 갖는 비상보적 말단 서열을 결합할 수 있다.
- [0133] 조이너 올리고의 상보적 암 서열의 길이는 엑소뉴클레아제 처리 후에 조립되는 핵산에 어닐링하기 충분해야 한다. 예를 들어, 조이너 올리고의 상보적 암 서열의 길이는 적어도 약 10, 20, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, 110, 120, 130, 140, 150 bp 또는 그 이상일 수 있다. 구체적인 실시 형태에서, 상보적 암은 15 내지 120 bp, 20 내지 100 bp, 30 내지 90 bp, 30 내지 60 bp 또는 20 내지 80 bp이다. 하나의 특정 실시 형태에서, 조이너 올리고의 상보적 암 서열의 길이는 40 bp이다. 조이너 올리고의 상보적 암 각각은 상이한 길이를 가질 수 있다. 조립되는 핵산에 상보적인 말단 서열 사이의 조이너 올리고의 스페이서는 적어도 약 20 bp, 30 bp, 35 bp, 40 bp, 45 bp, 50 bp, 55 bp, 60 bp, 65 bp, 70 bp, 75 bp, 80 bp, 90 bp, 100 bp,

250 bp, 500 bp, 750 bp, 1000 bp, 2000 bp, 3000 bp, 4000 bp, 5000 bp, 8000 bp, 10 kb, 15kb, 20 kb 또는 그 이상일 수 있다. 예를 들어, 조이너 올리고의 스페이서는 BAC 벡터 또는 LTVEC를 포함할 수 있다. 일부 실시 형태에서, 조이너 올리고의 스페이서는 성공적인 조립을 확인하기 위해 검출에 특이적인 서열 또는 PCR에 적합한 서열을 가지도록 설계될 수 있다. 일부 실시 형태에서, 조이너 올리고의 스페이서는 하나 이상의 제한 효소 부위를 도입하도록 설계될 수 있다. 일부 실시 형태에서, 조이너 올리고의 스페이서는 약물 저항성 유전자 또는 리포터 유전자를 도입하도록 설계될 수 있다. 다른 실시 형태에서, 스페이서는 핵산을 끊김 없이 조립하기 위해 조립되는 핵산의 말단 부분으로부터 적어도 20 bp를 함유할 수 있다. 예를 들어, 끊김 없는 조립을 위해 스페이서는 약 45 bp일 수 있다.

[0134] 일부 실시 형태에서, 핵산 대 조이너 올리고(들)의 몰비는 약 1:1 내지 약 1:200일 수 있다. 일부 실시 형태에서, 핵산 대 조이너 올리고(들)의 몰비는 약 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:11, 1:12, 1:13, 1:14, 1:15, 1:16, 1:17, 1:18, 1:19, 1:20, 1:30, 1:40, 1:50, 1:60, 1:70, 1:80, 1:90, 1:100, 1:120, 1:140, 1:160, 1:180 또는 1:200이다. 구체적인 실시 형태에서, 핵산 대 조이너 올리고(들)의 몰비는 약 1:6 내지 약 1:20일 수 있다. 일 실시 형태에서, 몰비는 약 1:6이다. 다른 실시 형태에서, 몰비는 약 1:20이다.

[0135] 구체적인 실시 형태에서, 조이너 올리고를 사용하여, 적어도 2개의 핵산을 끊김 없이 조립한다. "끊김 없는" 조립은 조립되는 핵산의 인접한 말단 사이에 개재 핵산 염기가 존재하지 않는 2개의 핵산의 조립을 말한다. 예를 들어, 끊김 없이 조립된 핵산에는 조립되는 핵산의 일부가 아닌 핵산 염기가 존재하지 않는다. 2개의 핵산을 끊김 없이 조립하기 위해, 조이너 올리고의 스페이서는 조립되는 제1 또는 제2 핵산의 말단 부분과 동일한 핵산 서열을 포함해야 한다. 이러한 말단 부분은 조이너 올리고로 조립되기 전에 핵산으로부터 제거되어야 한다. 예를 들어, 말단 부분은 핵산의 말단으로부터 적어도 20bp, 예컨대 핵산의 말단으로부터 적어도 40bp 또는 적어도 45bp의 뉴클레아제 제제 (예를 들어, gRNA-Cas 복합체)로 절단될 수 있다. 대안적으로, 말단 부분은 조립되는 핵산의 말단으로부터 적어도 2, 적어도 4, 적어도 6, 적어도 8, 적어도 10, 적어도 12, 적어도 15, 적어도 20, 적어도 25, 적어도 30, 적어도 35, 적어도 37, 적어도 40, 적어도 42, 적어도 45, 적어도 48, 적어도 50, 적어도 55, 적어도 60, 적어도 65, 적어도 70, 적어도 80, 적어도 100, 적어도 110, 적어도 120, 적어도 130, 적어도 140, 적어도 150 bp의 뉴클레아제 제제 (예를 들어, gRNA-Cas 복합체)로 절단될 수 있다.

[0136] 일 실시 형태에서, 조이너 올리고는 5' 말단부터 3' 말단까지 하기를 포함할 수 있다: 5' 핵산에 중첩하는 약 15 내지 120 bp, 5' 핵산의 3' 말단 영역의 약 20 내지 50 bp 및 3' 핵산에 중첩하는 약 15 내지 120bp. 일 실시 형태에서, 조이너 올리고는 5' 말단부터 3' 말단까지 하기를 포함할 수 있다: 5' 핵산에 중첩하는 약 15 내지 120 bp, 3' 핵산의 5' 말단 영역의 약 20 내지 50 bp 및 3' 핵산에 중첩하는 약 15 내지 120bp. 따라서, 조이너 올리고가 제1 핵산 및 제2 핵산과 조립되는 경우, 조이너 올리고로부터의 스페이서는 조립 전에 핵산으로부터 제거된 부분을 재구성한다. 도 5 및 도 6을 참조한다. 용어 "재구성"은 조이너 올리고에 조립된 경우 완전하게 조립된 핵산을 제공하기 위해 절단된 핵산의 말단 부분의 대체를 포함한다. 예를 들어, 절단된 핵산을 재구성하면 핵산의 절단된 부분은 절단된 부분의 서열과 동일한 서열을 갖는 조이너 올리고의 스페이서에 포함된 핵산으로 대체된다.

[0137] 조이너 올리고는 제1 핵산과 제2 핵산 분자에 동시에 또는 순차적으로 조립될 수 있다. 동시에 조립하는 경우, 조이너 올리고는 제1 핵산 및 제2 핵산과 동일한 반응 혼합물에서 접촉되어, 결과로서 얻어진 조립된 핵산은 제1 핵산, 조이너 올리고 및 제2 핵산을 포함할 수 있다. 순차적으로 조립하는 경우, 조이너 올리고는 제2 핵산이 아닌 조이너 올리고에 조립된 제1 핵산 조립을 포함하는 조립된 핵산을 생성하는 조립 반응에서 제1 핵산과 접촉된다. 이어서, 이러한 조립된 핵산은 제1 핵산, 조이너 올리고 및 제2 핵산을 포함하는 조립된 핵산을 생성하는 별개의 조립 반응에서 제2 핵산과 접촉될 수 있다. 다른 실시 형태에서, 조이너 올리고는 제1 핵산이 아닌 조이너 올리고에 조립된 제2 핵산 조립을 포함하는 조립된 핵산을 생성하는 조립 반응에서 제2 핵산과 접촉된다. 이어서, 이러한 조립된 핵산은 제1 핵산, 조이너 올리고 및 제2 핵산을 포함하는 조립된 핵산을 생성하는 별개의 조립 반응에서 제1 핵산과 접촉될 수 있다.

[0138] 임의의 수의 조이너 올리고는 본 명세서의 방법에서 핵산 분자를 조립하기 위해 사용될 수 있다. 예를 들어, 1개의 조이너 올리고를 사용하여 2개의 핵산 분자를 조립할 수 있거나, 2개의 조이너 올리고를 사용하여 3개의 핵산 분자를 조립할 수 있거나, 3개의 조이너 올리고를 사용하여 4개의 핵산 분자를 조립할 수 있거나, 4개의 조이너 올리고를 사용하여 5개의 핵산 분자를 조립할 수 있거나, 5개의 조이너 올리고를 사용하여 6개의 핵산 분자를 조립할 수 있다. 조이너 올리고의 수는 조립되는 핵산 분자의 수에 따라 1, 2, 3, 4, 5, 6, 7, 8, 9,

10 또는 그 이상일 수 있다.

- [0139] 일부 실시 형태에서, 조이너 올리고는 g블록 DNA를 포함한다. "g블록"은 선형 이중 가닥 DNA 단편이다. g블록은 약 50 bp 내지 약 2000 bp일 수 있다. g블록은 약 50 bp 내지 약 100 bp, 약 100 bp 내지 약 200 bp, 약 200 bp 내지 약 300 bp, 약 300 bp 내지 약 400 bp, 약 400 bp 내지 약 500 bp, 약 500 bp 내지 약 600 bp, 약 600 bp 내지 약 800 bp, 약 800 bp 내지 약 1000 bp, 약 1000 bp 내지 약 1250 bp, 약 1250 bp 내지 약 1500 bp, 약 1500 bp 내지 약 1750 bp 또는 약 1750 bp 내지 약 2000 bp일 수 있다.
- [0140] 둘 이상의 핵산과 g블록의 조립은, 예를 들어, 본 명세서의 다른 부분에 기재된 (예를 들어, 실시예 10) PCR 분석에 의해 스크리닝될 수 있다. 일부 경우에, g블록은 선택 카세트를 포함하지 않는다. 이러한 방법은 간단한 PCR 분석에 의해 스크리닝될 수 있는 둘 이상의 핵산 분자의 신속한 결합을 가능하게 한다. g블록은 임의의 관심 핵산 서열을 포함할 수 있다. 일부 경우에, g블록은 뉴클레아제 제제에 대한 표적 부위 또는 본 명세서에 제공된 임의의 다양한 메가뉴클레아제 또는 제한 효소에 대한 표적 부위를 포함할 수 있다. 다른 실시 형태에서, g블록은 선택 카세트를 포함할 수 있다. 일부 실시 형태에서, g블록은 관심 DNA 서열을 포함한다. 일 실시 형태에서, g블록은 인간 DNA 서열을 포함한다.
- [0141] 조립되는 핵산 또는 임의의 다양한 조이너 올리고는 또한 선택 카세트 또는 리포터 유전자를 포함할 수 있다. 선택 카세트는 선택 마커를 암호화하는 핵산 서열을 포함할 수 있으며, 여기서 핵산 서열은 프로모터에 작동가능하게 연결된다. 프로모터는 관심 원핵 세포에서 활성일 수 있고/있거나 관심 진핵 세포에서 활성일 수 있다. 이러한 프로모터는 유도성 프로모터, 리포터 유전자 또는 세포에 내인성인 프로모터, 리포터 유전자 또는 세포에 이중성인 프로모터, 세포 특이적 프로모터, 조직 특이적 프로모터 또는 발달 단계 특이적 프로모터일 수 있다. 일 실시 형태에서, 선택 마커는 네오마이신 포스포트랜스페라아제(neo<sup>r</sup>), 하이그로마이신 B 포스포트랜스페라아제(hyg<sup>r</sup>), 푸로마이신-N-아세틸트랜스페라아제(puro<sup>r</sup>), 블라스티사이드인 S 테아미나아제(bsr<sup>r</sup>), 잔틴/구아닌 포스포리보실 트랜스페라아제(gpt) 및 단순 헤르페스 바이러스 티미딘 키나아제(HSV-k) 및 이들의 조합으로부터 선택된다. 표적화 벡터의 선택 마커는 상류 및 하류 상동성 암이 플랭킹 될 수 있거나, 상동성 암에 대하여 5' 또는 3' 인 것으로 발견될 수 있다.
- [0142] 일 실시 형태에서, 조립되는 핵산 또는 임의의 다양한 조이너 올리고는 프로모터에 작동가능하게 연결된 리포터 유전자를 포함하며, 여기서 리포터 유전자는 LacZ, mPlum, mCherry, tdTomato, mStrawberry, J-Red, DsRed, mOrange, mKO, mCitrine, 비너스, YPet, 고감도 황색 형광 단백질(EYFP), 에메랄드, 고감도 녹색 형광 단백질(EGFP), CyPet, 시안 형광 단백질(CFP), 세룰리언, T-사파이어, 루시페라아제, 알칼리 포스파타아제 및 이들의 조합으로 이루어진 군으로부터 선택되는 리포터 단백질을 암호화한다. 이러한 리포터 유전자는 세포에서 활성인 프로모터에 작동가능하게 연결될 수 있다. 이러한 프로모터는 유도성 프로모터, 리포터 유전자 또는 세포에 내인성인 프로모터, 리포터 유전자 또는 세포에 이중성인 프로모터, 세포 특이적 프로모터, 조직 특이적 프로모터 방식 또는 발달 단계 특이적 프로모터일 수 있다.
- [0143] 단일 가닥 DNA (예를 들어, 결합되는 DNA 분자가 dsDNA인 경우 엑소뉴클레아제의 작용으로 생성된 돌출부 또는 각각의 가닥 상의 상이한 표적 부위에 틈을 만들어 생성된 돌출부)의 어닐링 후에, 엑소뉴클레아제에 의해 남겨진 단일 가닥 간극은 적합한, 비가닥-변위(non-strand-displacing) DNA 폴리머라아제를 사용하여 채워지고, 이로 인해 형성된 틈은 리가아제를 사용하여 밀봉된다. 본 명세서에서 사용되는 바와 같은 "비가닥-변위 DNA 폴리머라아제"는 dsDNA 분자의 복사를 진행하면서 이의 경로에 놓인 DNA 가닥을 직면하는 경우 DNA의 합성을 종결하거나, 진행하면서 직면한 DNA 가닥을 분해하면서, 동시에 이로 인해 생성된 간극을 채움으로써 "움직이는 틈" (틈 번역)을 생성하는 DNA 폴리머라아제이다.
- [0144] 일부 실시 형태에서, 중첩 말단 서열은 각각의 폴리뉴클레오티드의 단일 가닥 상보적 말단을 어닐링시키기 위해 중첩 영역 사이에 충분한 상보성을 갖는다. 제1 폴리뉴클레오티드의 단일 가닥을 제2 폴리뉴클레오티드의 상보적 가닥에 어닐링한 후에, 제1 폴리뉴클레오티드의 3' 말단은 제2 폴리뉴클레오티드 가닥의 주형에 기초하여 연장될 수 있고, 제2 폴리뉴클레오티드 가닥의 3' 말단은 제1 폴리뉴클레오티드 가닥의 주형에 기초하여 연장될 수 있다. 각각의 폴리뉴클레오티드의 상보적 3' 말단을 연장함으로써, 폴리뉴클레오티드는 조립될 수 있다. 조립 후, 하나의 단편으로부터의 가닥의 연장된 3' 말단과 다른 단편으로부터의 가닥의 인접 5' 말단 사이의 틈은 라이게이션에 의해 밀봉될 수 있다. 더욱 구체적으로, 제1 폴리뉴클레오티드의 연장된 3' 말단의 하이드록실기를 제2 폴리뉴클레오티드의 5' 말단의 포스페이트기로 그리고 제2 폴리뉴클레오티드의 연장된 3' 말단의 하이드록실기를 제1 폴리뉴클레오티드의 5' 말단의 포스페이트기로 라이게이션한다.



- [0145] 라이게이션 반응은 임의의 다양한 적합한 열안정성 DNA 리가아제에 의해 수행될 수 있다. 적합한 리가아제 중에는, 예를 들어, Taq 리가아제, Amp 리가아제 열안정성 DNA 리가아제 (에피센트리 바이오테크놀로지스 (Epicentre Biotechnologies)), 미국 특허 제6,576,453호에 개시된 열안정성 리가아제, 바이오니아, 인코포레이티드(Bioneer, Inc.)로부터의 열안정성 Tfi DNA 리가아제가 있다.
- [0146] 반응 혼합물 중의 적합한 양의 군집제(crowding agent), 예컨대 PEG는 분자 군집을 허용하거나, 향상시키거나, 촉진시킨다. 특정 메커니즘에 구애되기를 바라지는 않지만, 분자 군집을 허용하고, 용액 중의 물을 결합하여 묶어, 용액의 성분이 서로 더 가깝게 접촉할 수 있게 하는 군집제가 제안된다. 예를 들어, 재조합되는 DNA 분자는 더 근접하게 될 수 있어, 단일 가닥 돌출부의 어닐링을 촉진시킨다. 또한, 물 분자의 제거로 효소가 그들의 DNA 기질과 더 가깝게 접촉할 수 있고 안정화될 수 있다는 것을 시사한다. 다양한 적합한 군집제가 숙련된 작업자에게 명백할 것이다. 이들은, 다양한 잘 알려진 거대분자, 예를 들어, 폴리머, 예컨대 폴리에틸렌 글리콜 (PEG); 피콜, 예컨대 피콜 70; 텍스트란, 예컨대 텍스트란 70; 등을 포함한다. 본 출원의 논의의 대부분은 PEG에 관한 것이다. 그러나, 논의 내용은 다른 적합한 군집제에도 적용되는 것으로 의도된다. 숙련된 작업자는 다른 군집제의 사용을 수용하기 위해 방법에서 통상적인 변경을 구현하는 방법을 인지할 것이다.
- [0147] 반응 혼합물 중의 적합한 양의 군집제, 예컨대 PEG는 분자 군집을 허용하거나, 향상시키거나, 촉진시킨다. 예를 들어, 군집제는 재조합되는 DNA 분자가 더 근접하게 되고, 이로 인해 단일 가닥 돌출부의 어닐링이 촉진되는 것을 도울 수 있다. 또한, 물 분자의 제거로 효소가 그들의 DNA 기질과 더 가깝게 접촉할 수 있고 안정화될 수 있다는 것을 시사한다. 다양한 적합한 군집제가 숙련된 작업자에게 명백할 것이다. 이들은, 다양한 잘 알려진 거대분자, 예를 들어, 폴리머, 예컨대 폴리에틸렌 글리콜 (PEG); 피콜, 예컨대 피콜 70; 텍스트란, 예컨대 텍스트란 70; 등을 포함한다. 일반적으로, PEG가 사용되는 경우, 약 5%(중량/부피)의 농도가 최적이다. 그러나, PEG의 양은, 예를 들어, 약 3 내지 약 7%의 범위일 수 있다. 예를 들어, 약 PEG-200 (예를 들어, PEG-4000, PEG-6000 또는 PEG-8000) 내지 약 PEG-20,000 또는 그보다 더 큰 범위의 임의의 적합한 크기의 PEG가 사용될 수 있다. 본 명세서의 실시예에서는 PEG-8000이 사용되었다. 군집제는 어닐링 반응을 향상시키는 것 외에도 라이게이션을 향상시킬 수 있다.
- [0148] 조립 반응 혼합물에 존재하는 반응 성분 (예를 들어, 염, 완충액, 적합한 에너지 공급원 (예를 들어, ATP 또는 NAD), 반응 혼합물의 pH 등)은 개별 효소 (엑소뉴클레아제, 폴리머라아제 및 리가아제)에 대해 최적이지 아닐 수 있으며; 오히려, 전체 반응의 세트에 효과적인 절충안으로서 역할을 한다. 예를 들어, 발명자에 의해 확인된 하나의 적합한 완충 시스템(때로는 본 명세서에서 ISO (등온) 완충액으로 지칭됨)은 전형적으로 0.1 M의 Tris-HCl pH 7.5; 10 mM의 MgCl<sub>2</sub>, 0.2 mM의 각각의 dGTP, dATP, dTTP 및 dCTP, 10 mM의 DTT, 5% PEG-8000 및 1 mM의 NAD를 포함한다.
- [0149] 본 명세서에 개시된 방법에서, 2개 이상의 핵산은 핵산을 조립하기 효과적인 조건 하에서 Cas 단백질 및 다른 효소와 접촉하여, 중첩 영역의 단일 카피를 보유하는 조립된 이중 가닥 DNA 분자를 형성한다. 기재된 방법을 사용하여, 자연 발생 DNA, 클로닝된 DNA 분자, 합성으로 생성된 DNA, 등을 포함하는 임의의 관심 DNA 분자를 결합할 수 있다. 결합된 DNA 분자는, 원한다면, (예를 들어, 본 발명의 방법을 사용하여) 벡터로 클로닝될 수 있다. 일부 실시 형태에서, 조립되는 핵산은 관심 세포 (예를 들어, 설치류 세포, 마우스 세포, 래트 세포, 인간 세포, 포유류 세포, 미생물 세포, 효모 세포, 등)에서의 도입 및 발현에 최적화된 코돈이다.
- [0150] 임의의 길이의 DNA 분자는 본 명세서에 개시된 방법으로 결합될 수 있다. 예를 들어, 약 100 bp 내지 약 750 또는 1,000 또는 그 이상을 갖는 핵산이 결합될 수 있다. 본 명세서에 기재된 방법에 따른 하나 또는 여러 조립 단계에서 조립될 수 있는 핵산의 수는 예를 들어, 약 2 내지 약 30개의 핵산의 범위의 적어도 약 2, 3, 4, 6, 8, 10, 15, 20, 25, 50, 100, 200, 500, 1,000, 5,000 또는 10,000개의 DNA 분자일 수 있다. 조립 단계의 수는 약 2, 4, 6, 8, 10 또는 그 이상일 수 있다. 단일 단계에서 조립된 분자 조립의 수는 약 2 내지 약 10개의 분자의 범위일 수 있다. 본 발명의 방법을 사용하여 DNA 분자 또는 카세트들 함께 결합할 수 있으며, 각각은 최소한 또는 최대한 약 40 bp, 60 bp, 80 bp, 100 bp, 500 bp, 1 kb, 3 kb, 5 kb, 6 kb, 10 kb, 18 kb, 20 kb, 25 kb, 32 kb, 50 kb, 65 kb, 75 kb, 150 kb, 300 kb, 500 kb, 600 kb, 1 Mb 또는 그 이상의 시작 크기를 갖는다. 조립된 최종 생성물은 적어도 약 500 bp, 1 kb, 3 kb, 5 kb, 6 kb, 10 kb, 18 kb, 20 kb, 25 kb, 32 kb, 50 kb, 65 kb, 75 kb, 150 kb, 300 kb, 500 kb, 600 kb, 1Mb 또는 그 이상일 수 있으며, 예를 들어 30 kb 내지 1 Mb의 범위일 수 있다.
- [0151] 일부 실시 형태에서, 조립된 핵산은 원을 형성하고/하거나 벡터로 라이게이션되어 원을 형성한다. 원형을 만들기 위한 dsDNA의 크기의 하한은 약 200개의 염기쌍이다. 따라서, 결합된 단편의 총 길이 (일부 경우에, 벡터의

길이를 포함함)는 적어도 약 200 bp 길이이다. 크기의 실질적인 상한은 없고, 수백 킬로베이스 (kb) 쌍 또는 그 이상의 결합된 DNA가 본 명세서에 개시된 방법으로 생성될 수 있다. 결합된 핵산은 원형 또는 선형 분자 형태를 취할 수 있다.

[0152] 본 명세서에 기재된 방법을 사용하여 선형 단편과 다른 선형 단편, 선형 단편과 원형 핵산 분자, 원형 핵산 분자와 다른 원형 핵산 분자, 또는 선형과 원형 핵산의 임의의 조합을 조립할 수 있다. "벡터"는 임의의 원형 핵산 분자를 포함한다. 소정 실시 형태에서, 본 명세서에 개시된 방법으로 조립된 벡터는 박테리아 인공 염색체 (BAC)이다. 벡터 (예를 들어, BAC)는 인간 DNA, 설치류 DNA, 합성 DNA 또는 이들의 임의의 조합을 포함할 수 있다. 예를 들어, BAC는 인간 폴리뉴클레오타이드 서열을 포함할 수 있다. DNA 분자의 혼합물을 연결하는 경우, DNA가 대략 등물량으로 존재하는 것이 바람직하다.

[0153] 본 명세서에 개시된 방법에 의한 조립에서 사용되는 핵산은 큰 표적화 벡터일 수 있다. 용어 "큰 표적화 벡터" 또는 "LTVEC"는 세포에서 상동성 표적화를 위해 사용된 핵산 서열에 상응하고 이로부터 유래되는 상동성 암을 포함하고/하거나 세포에서 상동 재조합 표적화를 수행하도록 의도된 핵산 서열을 포함하는 삽입 핵산을 포함하는 벡터를 포함한다. 예를 들어, LTVEC는 이들의 크기 제한 때문에 전통적인 플라스미드-기반 표적화 벡터로 수용될 수 없는 큰 유전자좌의 변형을 가능하게 한다. 구체적인 실시 형태에서, LTVEC의 상동성 암 및/또는 삽입 핵산은 진핵 세포의 게놈 서열을 포함한다. LTVEC의 크기는 너무 커서 통상적인 분석, 예를 들어, 서던 블로팅 및 장거리 (예를 들어, 1kb 내지 5kb) PCR에 의한 표적화 이벤트의 스크리닝을 가능하게 할 수 없다. LTVEC의 예로는 박테리아 인공 염색체 (BAC), 인간 인공 염색체 또는 효모 인공 염색체(YAC)로부터 유래된 벡터를 들 수 있으나, 이에 한정되는 것은 아니다. LTVEC의 비제한적인 예 및 이들을 제조하는 방법은 예를 들어, 각각 본 명세서에 참조로 포함되는, 미국 특허 제6,586,251호, 제6,596,541호, 제7,105,348호, WO 2002/036789(PCT/US01/45375) 및 US 2013/0137101에 기재되어 있다.

[0154] 일부 실시 형태에서, 카세트는 나중에 제거될 수 있는 벡터로 삽입될 수 있다. 다양한 형태의 카세트는 특정 세포 또는 조직 유형에서, 특정 발달 단계에서 또는 유도 시에 결실을 허용하도록 구축될 수 있다. 이러한 카세트는 카세트가 재조합효소 인식 부위에 의해 양측에 플랭킹되는 재조합효소 시스템을 사용할 수 있고, 원하는 세포 유형에서 발현되거나, 원하는 발달 단계에서 발현되거나, 또는 유도 시에 발현되거나 활성화되는 재조합효소를 사용하여 제거될 수 있다. 이러한 카세트는 또한 그 전체 내용이 참조로 포함되는 US 2011/0104799에 기재된 바와 같이, 비대립유전자, 조건부 대립유전자 또는 조합 조건부/비대립유전자가 생성될 수 있도록 배치된 상이한 재조합효소 인식 부위 쌍의 어레이를 포함하도록 구축될 수 있다. 재조합효소 유전자의 조절은 다양한 방식으로, 예컨대 재조합효소 유전자를 세포-특이적, 조직-특이적 또는 발생학적으로 조절된 프로모터 (또는 다른 조절 요소)에 작동가능하게 연결시킴으로써 또는 재조합효소 유전자를 특정 세포 유형, 조직 유형 또는 발달 단계에서만 전사되는 miRNA를 위한 인식 부위를 포함하는 3' -UTR에 작동가능하게 연결시킴으로써 제어될 수 있다. 재조합효소는 또한, 예를 들어, 재조합효소를 이펙터 또는 대사 산물 (예를 들어, 활성이 타목시펜으로 양으로 조절되는(positively controlled) CreER<sup>T2</sup>)의 제어 하에 두는 융합 단백질을 사용함으로써, 또는 재조합효소 유전자를 유도성 프로모터 (예를 들어, 활성이 독시사이클린 및 TetR 또는 TetR 변이체로 조절되는 것)의 제어 하에 둠으로써 조절될 수 있다. 다양한 형태의 카세트의 예 및 재조합효소 유전자의 조절 방법, 예를 들어, 각각이 전체적으로 참조로 포함되는 US 8,518,392; US 8,354,389; 및 US 8,697,851에서 제공된다.

[0155] 본 명세서에 개시된 바와 같이 조립하는 데 사용되는 벡터 (예를 들어, LTVEC)는 약 20kb 내지 약 400kb, 약 20kb 내지 약 30kb, 약 30kb to 40kb, 약 40kb 내지 약 50kb, 약 50kb 내지 약 75kb, 약 75kb 내지 약 100kb, 약 100kb to 125kb, 약 125kb 내지 약 150kb, 약 150kb 내지 약 175kb, 약 175kb 내지 약 200kb, 약 200kb 내지 약 225kb, 약 225kb 내지 약 250kb, 약 250kb 내지 약 275kb 또는 약 275kb 내지 약 300kb, 약 300kb 내지 약 300kb, 약 300kb 내지 약 350kb, 약 350kb 내지 약 400kb, 약 350kb 내지 약 550kb를 포함하나, 이에 한정되지 않는 임의의 길이의 것일 수 있다. 일 실시 형태에서, LTVEC는 약 100kb이다.

[0156] 본 명세서에 제공된 핵산을 조립하는 방법은 약 5kb 내지 약 10kb, 약 10kb 내지 약 20kb, 약 20kb 내지 약 40kb, 약 40kb 내지 약 60kb, 약 60kb 내지 약 80kb, 약 80kb 내지 약 100kb, 약 100kb 내지 약 150kb 또는 약 150kb 내지 약 200kb, 약 200kb 내지 약 300kb, 약 300kb 내지 약 400kb, 약 400kb 내지 약 500kb, 약 500kb 내지 약 1Mb, 약 1Mb 내지 약 1.5Mb, 약 1.5Mb 내지 약 2Mb, 약 2Mb 내지 약 2.5Mb 또는 약 2.5Mb 내지 약 3Mb의 결실을 허용하도록 설계될 수 있다.

[0157] 다른 경우에, 본 명세서에 제공된 방법은 약 5kb 내지 약 10kb, 약 10kb 내지 약 20kb, 약 20kb 내지 약 40kb, 약 40kb 내지 약 60kb, 약 60kb 내지 약 80kb, 약 80kb 내지 약 100kb, 약 100kb 내지 약 150kb, 약 150kb 내

지 약 200kb, 약 200kb 내지 약 250kb, 약 250kb 내지 약 300kb, 약 300kb 내지 약 350kb 또는 약 350kb 내지 약 400kb 범위의 외인성 핵산 서열의 삽입을 허용하도록 설계된다. 일 실시 형태에서, 삽입 폴리뉴클레오티드는 약 130 kb 또는 약 155kb이다.

[0158] 선형 핵산은 본 명세서에 개시된 방법으로 서로 또는 벡터에 조립될 수 있다. 선형 분자는 엔도뉴클레아제 (예를 들어, Cas 단백질) 또는 임의의 합성, 인공 또는 자연 발생 선형 핵산에 의해 분해된 벡터일 수 있다. 소정 실시 형태에서, 선형 핵산은 말단 서열이 다른 핵산의 영역과 중첩되도록 생성된다. 선형 핵산의 중첩 말단 서열은 맞춤 핵산 서열을 생성하기 위해 본 기술 분야에 알려진 임의의 방법으로 도입될 수 있다. 예를 들어, 말단 서열은 합성에 의해 제조된 분자의 부분일 수 있고, PCR에 의해 도입될 수 있거나, 전통적인 클로닝 기술에 의해 도입될 수 있다.

[0159] 실시예

[0160] 하기 실시예는 당업자에게 본 발명을 제조하고 사용하는 방법의 완전한 개시 및 설명을 제공하기 위해 제시되며, 발명자가 그들의 발명으로 간주하는 것의 범주를 제한하고자 하는 것이 아니고, 하기 실험이 수행되는 모든 또는 유일한 실험인 것을 나타내는 것으로 의도되지 않는다. 사용된 숫자 (예를 들어, 양, 온도 등)와 관련하여 정확성을 확보하기 위한 노력이 이루어졌으나, 일부 실험 오차 및 편차가 고려되어야 한다. 다른 언급이 없다면, 부는 중량부이고, 분자량은 중량 평균 분자량이며, 온도는 섭씨 단위이고, 압력은 대기압 또는 대기압 근처이다.

[0161] 실시예 1: CAS9로 BAC를 분해한 후에 선택 카세트와의 조립

[0162] 인공 crRNA와 인공 tracrRNA는 3 kb PCR 생성물 (UB-HYG)과의 조립을 위해 MAID 6177 (116 kb LTVEC)의 특정 서열을 표적화하도록 설계되었다. PCR 생성물은 벡터와 중첩되는 50 bp를 함유하였다. 먼저 crRNA와 tracrRNA를 듀플렉스 완충액(Duplex Buffer) (30 mM HEPES, pH 7.5, 100 mM 아세트산칼륨)에서 100  $\mu$ M로 용해시킨다. RNA를 어닐링하기 위해, 10  $\mu$ l의 100  $\mu$ M crRNA와 10  $\mu$ l의 100  $\mu$ M tracrRNA를 80  $\mu$ l의 어닐링 완충액에 첨가한다. RNA를 90°C 온도 블록에서 가열한 다음 블록을 히터로부터 제거하여 벤치 상에서 냉각시킨다. RNA의 최종 농도는 약 10  $\mu$ M이다.

[0163] BAC를 분해하기 위해, 깨끗한 맥시프랩(maxiprep) BAC DNA가 사용되고, BAC를 하기 혼합물에 의해 분해한다.

[0164] 1X

[0165]	BAC DNA (500ng)	X $\mu$ l
[0166]	BSA (100x)	0.5 $\mu$ l
[0167]	RNA	2 $\mu$ l (각 1 $\mu$ l의 tracr:crRNA 하이브리드)
[0168]	Cas9 (4.5mg/ml)	1 $\mu$ l
[0169]	10x 완충액	1.5 $\mu$ l
[0170]	H <sub>2</sub> O	15 $\mu$ l가 되도록 첨가함

[0171] 37°C에서 1시간 동안 분해한 후 30분 동안 탈염한다. 최종 반응 완충액은 하기를 함유한다: 20 mM Tris 7.5; 100-150 mM NaCl; 10 mM MgCl<sub>2</sub>; 1 mM DTT; 0.1 mM EDTA; 100  $\mu$ g/ml BSA; 최종 부피는 15  $\mu$ l가 되도록 함.

[0172] BAC를 조립하고 삽입하기 위해, 플라스미드를 분해하거나 PCR을 수행하여 삽입물을 생성한다. PCR 반응의 경우, 적은 분취액을 겔 상에 실행하고 단일 생성물을 찾는다. 생성물이 단일 밴드이면 겔 추출 대신에 PCR 클리닝을 수행한다. BAC:삽입물의 1:1 내지 1:6의 물비가 요구된다. 보통, 50 ng의 정제된 삽입물이 효과적인 것이다. 하기 반응 혼합물이 사용될 수 있다:

[0173]	BAC 분해물	4 $\mu$ l
[0174]	삽입물	1 $\mu$ l
[0175]	조립 혼합물	15 $\mu$ l

[0176] DNA와 혼합물을 50°C에서 얼음 상에 또는 직접 PCR 기계로 첨가한다. 50°C에서 1시간 동안 인큐베이션한다. 0.5  $\mu$ l의 프로테이나아제 K (20mg/ml)를 첨가하고 50°C에서 1시간 동안 인큐베이션한다. 30분 동안 탈염하고

DH10B 세포에 8  $\mu$ l 의 반응물을 전기천공한다. 10  $\mu$ l의 BAC 분해물을 펄스장 겔(pulse-field gel) 상에서 실행하여 분해 효율을 확인할 수 있다. RNase가 없는 물과 완충액을 사용한다.

[0177] 조립 반응을 다음과 같이 수행한다: 등은 완충액: 3 mL 1M Tris-HCL (pH 7.5); 150  $\mu$ l 2M  $MgCl_2$ ; 60  $\mu$ l 100 mM 각각: dGTP, dATP, dTTP, dCTP; 300  $\mu$ l 1M DTT; 1.5 g PEG 8000; 300  $\mu$ l 100 mM NAD. 등은 완충액은 320  $\mu$ l 분취액으로 -20℃에서 보관된다. 마스터 믹스(Master Mix)를 다음과 같이 제조한다: 320  $\mu$ l 등은 완충액; 0.64  $\mu$ l T5 엑소뉴클레아제 (스톡 농도=10 U/ $\mu$ l); 20  $\mu$ l 퓨전(Phusion) DNA 폴리머라아제 (스톡 농도=2 U/ $\mu$ l); 160  $\mu$ l Taq DNA 리가아제 (스톡 농도=40 U/ $\mu$ l); 699.36  $\mu$ l  $H_2O$ ;를 함께 혼합하고, 15  $\mu$ l 또는 30  $\mu$ l에서 분액하여 -20℃에서 저장한다. 총 부피 20  $\mu$ l의 반응물에 15  $\mu$ l 마스터 믹스 (MM)를 사용한다.

[0178] 실시예에서 사용된 tracr RNA 서열은 다음과 같다:

[0179] CAAAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUC (서열 번호: 9). 이러한 CRISPR RNA (crRNA)는 다음을 함유한다: (1) 표적 서열에 상보적인 약 20 뉴클레오타이드의 RNA 및 (2) tracrRNA에 어닐링되는 꼬리 서열 (GUUUUAGAGCUAUGCUGUUUUG (서열 번호: 10)).

[0180] 이러한 단계가 도 1에 요약되어 있다.

[0181] 실시예 2: 2개의 중첩하는 BAC를 함께 재봉(sewing): 마우스 MHC II 유전자좌에서 인간화 HLA-DQ + 인간화 HLA-DR (H2-A/H2-E)

[0182] 인공 crRNA와 인공 tracrRNA는 인간화 HLA-DR BAC와의 조립을 위해 인간화 HLA-DQ BAC의 특정 서열을 표적화하도록 설계되었다. 벡터는 각각의 벡터 상의 2개의 부위에서 Cas9 절단에 의해 생성된 서로 중첩하는 ~70bp를 함유하였다 (도 2를 참조한다). crRNA와 tracrRNA를 하이브 완충액(Hybe Buffer)에서 100  $\mu$ M로 용해시킨다. RNA를 어닐링하기 위해, 10  $\mu$ l의 100  $\mu$ M crRNA와 10  $\mu$ l의 100  $\mu$ M tracrRNA를 80  $\mu$ l의 어닐링 완충액에 첨가한다. RNA를 90℃ 히트 블록(heat block)에 두고, 이어서 히트로부터 블록을 제거하여 벤치 상에서 냉각시킨다. RNA의 최종 농도는 약 10  $\mu$ M이다.

[0183] BAC를 분해하기 위해, 깨끗한 맥시프랩 BAC DNA가 사용될 수 있다. 각각의 BAC는 하기 혼합물에 의해 개별적으로 분해될 수 있다:

[0184]	BAC DNA 2.5 $\mu$ g	X $\mu$ l
[0185]	BSA (100x)	0.5 $\mu$ l
[0186]	RNA	4 $\mu$ l (각 2 $\mu$ l의 tracr:crRNA 하이브리드)
[0187]	Cas9 (4.5mg/ml)	1 $\mu$ l
[0188]	10x 완충액	5 $\mu$ l
[0189]	$H_2O$	50 $\mu$ l가 되도록 첨가함

[0190] BAC 벡터를 37℃에서 C 1시간 동안 분해한 다음, 20분 동안 65℃에서 열 불활성화시킨다(heat inactivated). 30분 동안 탈염한다. 분해된 DNA를 페놀/클로로포름/아이소아밀알코올 (PCI) 추출을 통해 정제한 다음, 35  $\mu$ l TE 완충액에 재현탁시켰다.

[0191] 벡터를 조립하기 위해, 다음과 같은 조립 반응에 2.5  $\mu$ l의 BAC를 사용한다:

[0192]	분해된 BAC	5 $\mu$ l (총)
[0193]	조립 혼합물	15 $\mu$ l

[0194] DNA와 혼합물을 50℃에서 얼음 상에 또는 직접 PCR 기계로 첨가한다. 50℃에서 1시간 동안 인큐베이션한다. 30분 동안 탈염하고 DH10B 세포에 8  $\mu$ l 의 조립된 DNA를 전기천공한다. RNase가 없는 물과 완충액을 사용한다.

[0195] 실시예에서 사용된 tracr RNA 서열은 다음과 같다: CAAAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUC (서열 번호: 9). 이러한 CRISPR RNA (crRNA)는 다음을 함유한다: (1) 표적 서열에 상보적인 약 20 뉴클레오타이드의 RNA 및 (2) tracrRNA에 어닐링되는 꼬리 서열 (GUUUUAGAGCUAUGCUGUUUUG (서열 번호: 10)).



- [0196] 이러한 단계가 도 2에 요약되어 있다.
- [0197] 실시예 3: 링커를 사용한 2개의 상이한 플라스미드로부터의 2개의 Cas9-절단된 단편의 조립
- [0198] 표적화 벡터를 구축하기 위해, pMJ8502x를 2개의 동일한 crRNA로 절단하여 400 bp 단편과 2283 bp Amp 골격을 빼냈다. (도 7). 퀴아젠(Qiagen) 컬럼을 사용하여 전체 반응을 정제하였다. 이어서, R6KZenUbiNeo를 2개의 상이한 crRNA로 절단하여 Neo 저항성 (1086 bp)과 골격 (5390 bp)으로 분리하였다. 퀴아젠 컬럼을 사용하여 전체 반응을 정제하였다. (도 7). 절단 반응: 1170 ng DNA, 30  $\mu$ l 완충액, 4  $\mu$ l 어닐링된 RNA (@100  $\mu$ M), 1.7  $\mu$ l Cas9 (@0.89 ng/ $\mu$ l), 60  $\mu$ l가 되도록 H<sub>2</sub>O 첨가. 혼합물을 37°C에서 1시간 동안 인큐베이션하고 퀴아젠 컬럼 상에서 정제한 후, 30  $\mu$ l 용리 완충액에서 용리시켰다.
- [0199] 이어서, 절단된 단편을 2개의 링커로 조립하여, 하기 반응 혼합물에 의해 꿰김 없이 조립하였다: 0.5  $\mu$ l 링커1 (5 ng), 0.5  $\mu$ l 링커2 (5 ng), 2  $\mu$ l Neo 절단 (~60ng), 2  $\mu$ l Amp 절단 (~60ng), 15  $\mu$ l 조립 마스터 믹스. 혼합물을 50°C에서 1시간 동안 인큐베이션하고, 반응물을 H<sub>2</sub>O에 대해 투석시켰다. 10  $\mu$ l의 반응물을 전기적격성(electrocompetent) Pir 세포에 전기천공한 다음, Carb/Kan 플레이트 상에 플레이트하였다. 접합부에 걸친 PCR은 6/8 선택된 콜로니가 정확한 것으로 나타났고, 시퀀싱에 의해 확인되었다.
- [0200] 실시예 4: 링커를 사용한 BAC의 부분의 카세트로의 교체
- [0201] 녹아웃 마우스 표적화 벡터를 구축하기 위해, 40 kb의 BAC 표적화 벡터를 재조합 인식 부위에 의해 플랭킹된 선택 카세트에 교체하였다. (도 8) 2개의 링커를 mBAC로부터 관심 영역을 삭제하고, 선택 카세트를 삽입하도록, 하나는 5' 하나는 3' 을 위해 설계하였다. 링커는 mBAC와 중첩하는 40 bp 및 선택 카세트와 중첩하는 40 bp를 가졌다. 먼저, 206 kb 표적화 벡터 (mBAC)의 39.5 kb를 하기 반응에 의해 절단하였다: 500  $\mu$ l 반응물 (H<sub>2</sub>O를 주입함): 1  $\mu$ l Cas9 (@0.89  $\mu$ g/ $\mu$ l), 각 2  $\mu$ l RNA 듀플렉스 (@50  $\mu$ M), 250  $\mu$ l 완충액, 220  $\mu$ l (12.5 ng) BAC 맥시프랩을 첨가하고, 37°C에서 1시간 동안 인큐베이션하였다. 분해된 DNA를 페놀/클로로포름/아이소아밀알코올 (PCI) 추출을 통해 정제한 다음, 55  $\mu$ l TE 완충액에 재현탁시켰다. mBAC 절단의 PCI 클린업 후, 조립을 50°C에서 1시간 동안 수행하고, 10  $\mu$ l의 반응물을 DH10B 세포로 전기천공하였다. (도 9). 접합부에 걸친 시퀀싱으로 정확한 조립을 확인하였다. (도 10). 링커 1 (조이너 올리고 1)은 mBAC 서열에서 카세트 서열까지 꿰김이 없다 (서열 번호: 12). 링커 2 (조이너 올리고 2)는 카세트 서열에서 mBAC 서열까지 꿰김이 없다 (서열 번호: 13).
- [0202] 실시예 5: 링커 (조이너 올리고)를 사용한 2개의 BAC 벡터의 조립
- [0203] Cas9/등온 조립에 의한 2개의 mBAC의 봉합(stitching)을 사용하여, 마우스 게놈 영역에 대한 상동성 압 및 BAC 라이게이션으로 인간 유전자를 삽입하기 위한 제한 부위를 함유하는 표적화 벡터를 제조하였다. 이러한 표적화 벡터를 BAC 라이게이션에서 사용하여, 인간화 표적화 벡터를 제조하였다. mBAC를 하기 반응물에 의해 절단하였다: 12.5  $\mu$ g DNA, 2  $\mu$ l 각각의 어닐링된 RNA (@50  $\mu$ M), 10  $\mu$ l Cas9 (@0.89  $\mu$ g/ $\mu$ l), 250  $\mu$ l 완충액, 500  $\mu$ l가 되도록 H<sub>2</sub>O 첨가. 혼합물을 37°C에서 1시간 동안 인큐베이션하고; 페놀/클로로포름/아이소아밀알코올 (PCI) 추출로 클린업하여; 20  $\mu$ l TE에 재현탁시켰다. 이어서, 2개의 마우스 BAC를 하기 반응물에 의해 링커 (도 11)와 함께 조립하였다: 6  $\mu$ l (2  $\mu$ g) bMQ-208A16 절단, 5.6  $\mu$ l (2  $\mu$ g) bMQ-50F19 절단, 0.25  $\mu$ l 각각의 링커 (@50  $\mu$ M), 4.3  $\mu$ l (100 ng) 선택 카세트 (Ubi-Hyg) 카세트, 12  $\mu$ l 고농도 조립 마스터 믹스, 11.35  $\mu$ l H<sub>2</sub>O. 반응 혼합물을 50°C에서 1시간 동안 인큐베이션하고, 30°C에서 H<sub>2</sub>O에 대해 투석하였다. 10  $\mu$ l 또는 30  $\mu$ l의 투석된 반응물을 사용하여 DH10B 세포를 형질전환하였다. 생거(Sanger) 시퀀싱으로 모든 접합부를 확인하였다. 일루미나(Illumina) 시퀀싱으로 모든 접합부를 재확인하였다 (도 12 및 서열 번호: 17). 링커 1은 mBAC에서 카세트까지 꿰김이 없다 (서열 번호: 14). 링커 2는 카세트에서 mBAC까지 꿰김이 있었다. 이것은 프로젝트 설계에 따라 인간 스페이서 서열을 포함한다. 링커 3은 mB2에서 mB3까지 꿰김이 있었다. 이것은 PCR 검증에 사용되는 독특한 서열을 포함한다. 이러한 영역을 ES 전기천공을 위해 선형화될 때 제거하였다 (서열 번호: 15).
- [0204] 도 13은 4개의 링커 및 등온 조립을 사용하여 큰 인간 유전자 단편을 mBAC로 삽입하는 4개의 조이너 올리고 (링커)를 사용하는 예를 도시한다.
- [0205] 실시예 6: 절단 및 조립을 위한 시약 및 반응 혼합물
- [0206] Crispr RNA (crRNA) (ssRNA로 주문됨)는 다음을 함유한다: (1) 절단하기 위한 표적 영역에 상보적인 RNA의 20

개의 뉴클레오티드; (2) 및 tracr RNA에 어닐링하는 꼬리: <20nt crisprRNA>GUUUUAGAGCUAUGCUGUUUUG (서열 번호: 10).

[0207] Tracr RNA (ssRNA로 주문됨): GUUGGAACCAUUCAAAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU (서열 번호: 11).

[0208] 모든 RNA를 H<sub>2</sub>O에서 100 μM로 재현탁시켰다. 2.5 μl의 각각의 crRNA 및 tracrRNA를 5 μl의 어닐링 완충액 (최종 농도: 10 mM Tris pH 7.5 내지 8.0, 50 mM NaCl, 1 mM EDTA)과 혼합한다. 이어서, 혼합물을 95°C에서 5 분 동안 인큐베이션하고, 1시간에 걸쳐 실온으로 서서히 냉각시킨다. Cas9 2X 절단 완충액은 40 mM HEPES pH7.5 (최종= 20 mM); 300 mM KCl (최종= 150 mM); 1mM DTT (최종= 0.5 mM); 0.2mM EDTA (최종= 0.1 mM); 20 mM MgCl<sub>2</sub> (최종= 10 mM)를 함유한다.

[0209] 대규모 Cas9 절단 반응: 실온에서 순서대로: 500 μl가 되도록 H<sub>2</sub>O 첨가, 250 μl 2x 절단 완충액, 12.5 μg DNA, 2 μl의 각각의 RNA (50 μM 농도), 10 μl Cas9 (0.89 mg/ml 농도)를 첨가하고, 37°C에서 1시간 동안 인큐베이션한다.

[0210] 이러한 반응은 필요에 따라 크기 조정될 수 있다. 예를 들어, 50 μl가 되도록 H<sub>2</sub>O 첨가, 25 μl 완충액, 125 ng DNA, 2 μl 각각의 RNA (5 μM 농도), 1 μl Cas9 (0.89mg/ml 농도)를 첨가하고, 37°C에서 1시간 동안 인큐베이션한다.

[0211] 조립 반응을 다음과 같이 수행한다: 등은 완충액: 3 mL 1M Tris-HCL (pH 7.5); 150 μl 2M MgCl<sub>2</sub>; 60 μl 100 mM 각각: dGTP, dATP, dTTP, dCTP; 300 μl 1M DTT; 1.5 g PEG 8000; 300 μl 100 mM NAD. 등은 완충액은 320 μl 분취액으로 -20°C에서 보관된다. 마스터 믹스를 다음과 같이 제조한다: 320 μl 등은 완충액; 0.64 μl T5 엑소뉴클레아제 (스톡 농도=10 U/μl); 20 μl 퓨전 DNA 폴리머라아제 (스톡 농도=2 U/μl); 160 μl Taq DNA 리가아제 (스톡 농도=40 U/μl); 699.36 μl H<sub>2</sub>O를 함께 혼합하고, 15 μl 또는 30 μl에서 분액하여 -20°C에서 저장한다. 총 부피 20 μl의 반응물에 15 μl 마스터 믹스 (MM)를 사용한다.

[0212] 대안적으로, 고농도 마스터 믹스 (GA MM HC)를 다음과 같이 제조할 수 있다: 320 μl 등은 완충액; 0.64 μl T5 엑소뉴클레아제 (스톡 농도=10 U/μl); 20 μl 퓨전 DNA 폴리머라아제 (스톡 농도=2 U/μl); 160 μl Taq DNA 리가아제 (스톡 농도=40 U/μl)를 함께 혼합하고, 6 μl 또는 12 μl에서 분액하여 -20°C에서 저장한다. 총 부피 20 μl의 반응물에 6 μl 마스터 믹스 (MM)를 사용한다.

[0213] 모든 조립 반응에 대해, DNA의 농도는 측정되어야 하고 (예를 들어, 나노 드롭(Nano Drop)에 의해) 1:6 몰비 (벡터 대 삽입(들))가 사용된다. 표준 농도의 경우, 15 μl의 조립 마스터 믹스가 사용된다. 200 μl PCR 튜브에서, 최종 부피 20 μl로 DNA와 물을 첨가한다. 반응을 유전자 증폭기(thermocycler)에서 50°C에서 1시간 동안 수행한다. 이어서, 반응물을 -20°C에서 저장할 수 있다. 고농도의 경우, 6 μl의 고농도 조립 마스터 믹스가 사용된다. 200 μl PCR 튜브에서, 최종 부피 20 μl로 DNA와 물을 첨가한다. 반응을 유전자 증폭기(thermocycler)에서 50°C에서 1시간 동안 수행한다. 이어서, 반응물을 -20°C에서 저장할 수 있다. 반응 완료 시, 10 μl를 물에 대해 30분 동안 투석하고, 적절한 전기적격성 세포 (예를 들어, DH10B 또는 Pir+ 세포)에 전기천공한다.

[0214] Cas9/등은 조립 반응: Cas9 분해를 위해, 2.5 μg의 각각의 DNA (예를 들어, BAC DNA), 4 μl의 10 μM 가이드/tracr RNA 각각 및 5 μl의 Cas9 단백질 (0.89 mg/ml)을 37°C에서 2시간 동안 분해한다. 반응을 65°C에서 20 분 동안 열 불활성화시키고, 페놀 클로로포름 추출하여 (예를 들어, Cas9 단백질을 제거하기 위해), 70% 에탄올로 1회 세정하여, DNA를 35 μl 물에서 재현탁시킨다. 등은성 조립은 본 명세서의 다른 부분에 기재된 바와 같이, 15 μl의 마스터 믹스 (MM)와 함께 혼합된 5 μl의 DNA를 사용하여 수행되고, 50°C에서 1시간 동안 인큐베이션된다. 반응을 30분 동안 탈염하고 8 μl의 반응물을 세포로 전기천공할 수 있다.

[0215] 실시예 7: 인간 서열을 BAC 벡터로 삽입하기 위한 Cas9/등은 조립

[0216] 인간화 표적화 벡터를 구축하기 위해, MAID 6236을 gRNA-Cas 복합체로 절단하여, 중첩 서열을 갖는 절단된 단편을 생성하였다. VI568을 또한 gRNA-Cas 복합체로 절단하여, MAID6236의 단편과 중첩하는 서열을 생성하였다. Cas9/ 등은 조립을 상기한 바와 같이 수행하여, 인간화 유전자좌를 벡터 (VI599)로 삽입하였다. 이러한 과정이 도 14에 요약되어 있다.

[0217] 실시예 8: 선택 없이 g블록을 사용한 Cas9/등은 조립

[0218] Cas9 분해 및 조립은 또한 선택없이, 예를 들어, g블록 DNA 단편을 사용함으로써 수행될 수 있다. 선택 카세트 없이 이중 가닥 DNA를 유전자좌로 첨가할 가능성을 시험하기 위해, g블록 DNA 단편을 합성하여 구축물로 삽입하였다. 도 15a 및 도 15b에 도시된 바와 같이, 4.4 kb 단편을 제거하기 위해 TCR 베타 유전자좌 내의 2개의 부위를 표적화하도록 Cas9/gRNA를 설계하였다. g블록을 메가뉴클레아제 인식 부위를 구축물로 도입하도록 설계하였다. 선택 마커를 사용하지 않고 g블록을 구축물로 삽입할 수 있었다. 도 15a는 PISceI g블록의 삽입을 도시하고, 도 15b는 MauBI g블록의 삽입을 도시한다.

[0219] 표 1에 나타난 프라이머를 사용하여 PCR 접합 스크린(PCR junction screen)에 의한 각각의 g블록의 성공적인 삽입에 대해 최종 구축물을 확인하였다. 접합 스크린에 대한 프로토콜은 다음과 같다: PCR 반응은 하기를 함유한다: 1μL DNA, 0.5μL 프라이머 1, 0.5μL 프라이머 2, 1μL DMSO, 4μL dNTPs, 2.5μL 10x 완충액, 0.5μL Ex-Taq 및 15μL 물. 반응을 유전자 증폭기에서 95℃에서 3분 동안, 95℃에서 30초 동안, 55℃에서 30초 동안 25 사이클, 이어서 72℃에서 30초 동안 그리고 72℃에서 5분 동안 수행하였다. 접합 서열을 시퀀싱에 의해 확인하였다.

[0220] [표 1]

PI-SceI g 블록 또는 MauBI g 블록을 이용한 MAID1715의 접합 스크리닝을 위한 프라이머

MAID1715+PISceI G블록		
프라이머명	서열	접합 크기
(m380)5' 302p18 검출	GGAAAGCCACCCTGTATGCT (서열 번호: 18)	796 bp
3'다운 검출 302p18(m41)	CTTGGCCAACAGTGGATGG (서열 번호: 19)	
Cas9 프라이머명	서열	DNA 표적 서열
1715 표적-5'	CUAAAAUGAUUCUCAUCUGC GUUUUAGAGCUAUGCUGUUUUG (서열 번호: 20)	CTAAATGATTCTCATCTGC(AGG) (서열 번호: 22)
1715 표적-3'	GCUCUCAACUUCACCCUUUC GUUUUAGAGCUAUGCUGUUUUG (서열 번호: 21)	GCTCTCAACTTCACCCTTTC(TGG) (서열 번호: 23)
MAID1715+MauBI G블록		
프라이머명	서열	접합 크기
(m380)5' 302p18 검출	GGAAAGCCACCCTGTATGCT (서열 번호: 18)	759bp
3'다운 검출 302p18(m41)	CTTGGCCAACAGTGGATGG (서열 번호: 19)	
Cas9 프라이머명	서열	DNA 표적 서열
1715 표적-5'	CUAAAAUGAUUCUCAUCUGC GUUUUAGAGCUAUGCUGUUUUG (서열 번호: 20)	CTAAATGATTCTCATCTGC(AGG) (서열 번호: 22)
1715 표적-3'	GCUCUCAACUUCACCCUUUC GUUUUAGAGCUAUGCUGUUUUG (서열 번호: 21)	GCTCTCAACTTCACCCTTTC(TGG) (서열 번호: 23)

[0221]

[0222] 실시예 9: 조이너 올리고를 사용하여 인간 서열을 BAC 벡터로 삽입하기 위한 Cas9/등은 조립

[0223] 도 16은 Cas9/등은 조립 및 조이너 올리고를 사용한 직접 인간화의 예를 제공한다. 인간 단편과 마우스 결실은 Cas9에 의해 빠진다 (각각의 BAC는 2개의 crispr RNA를 사용한다). 인간 단편과 마우스 골격이 3개의 링커 (조이너 올리고)와 선택 카세트를 사용하는 김슨(Gibson) 조립 반응으로 함께 연결된다.

[0224] 도 17은 큰 표적화 벡터 (LTVEC)로의 조립을 위한 Cas9/등은 조립 및 조이너 올리고를 사용하는 간접 인간화의 예를 제공한다. hBAC 상의 인간 단편은 2개의 crispr RNA를 사용하여 Cas9에 의해 절단된다. 공여자는 업 앤드 다운 조이너 올리고 및 선택 카세트를 포함한다. Cas9에 의한 hBAC 절단 후, 단편은 혼입된 상보성 돌출부를 갖는 합성 공여자를 사용하여 김슨 조립에 의해 "포획"된다. 표적화 벡터 구축이 김슨 조립 또는 BHR에 의해 완료된다.

[0225] 실시예 10: Cas9/등은 조립에 의한 점 돌연변이 도입

[0226] 도 18은 Cas9/등은 조립을 이용한 점 돌연변이의 도입의 예를 제공한다. 공여자는 전통적인 클로닝으로 제조된

다. 선택 카세트가 링커 중첩 및 점 돌연변이를 함유하는 합성 DNA 단편으로 삽입된다. mBAC를 Cas9로 절단하고, 서열을 mBAC로부터 제거하여, mBAC를 공여자에 집는 조립하여, 점 돌연변이와 선택 카세트를 포함하는 구축물 (LTVEC)을 제조한다.

[0227] 실시예 11: Cas9/등온 조립에 의한 BAC 트리밍

[0228] 도 19는 Cas9/등온 조립 방법을 사용하는 BAC 트리밍의 예를 제공한다. LTVEC로부터 제거해야 할 영역을 Cas9를 사용하여 트리밍한다. 이러한 실시예에서, BAC 트리밍은 Ori 서열을 제거한다. Ori는 2개의 링커 (조이너 올리고)를 사용하여 집는 조립 반응에서 대체된다.

[0229] 실시예 12: CAS9를 사용하여 BAC를 분해한 후 조립하기 위한 다른 방법

[0230] 다른 방법이 하기를 포함하는 본 명세서에 제공된 방법에 사용될 수 있다: 합성 또는 시험관 내-전사된 tracrRNA와 crRNA를 95°C로 가열하고 실온으로 서서히 냉각시킴으로써 반응 전에 미리 어닐링하였다. 고유 또는 선형화된 플라스미드 DNA (300 ng (약 8 nM))를, 10 mM MgCl<sub>2</sub>의 유무에 관계 없이, Cas9 플라스미드 절단 완충액 (20 mM HEPES pH 7.5, 150 mM KCl, 0.5 mM DTT, 0.1 mM EDTA) 중의 정제된 Cas9 단백질 (50-500 nM) 및 tracrRNA:crRNA 듀플렉스 (50-500 nM, 1:1)와 함께 37°C에서 60분 동안 인큐베이션하였다. 반응을 250 mM EDTA를 함유하는 5X DNA 로딩 완충액으로 중지시키고, 0.8 또는 1% 아가로스 겔 전기영동으로 분해하여, 브롬화 에티듐 염색으로 가시화하였다. Cas9 돌연변이 절단 분석을 위해, 반응을 아가로스 겔에 로딩하기 전에 5X SDS 로딩 완충액 (30% 글리세롤, 1.2% SDS, 250 mM EDTA)으로 중단시켰다.

[0231] 인공 crRNA와 인공 tracrRNA는 3 kb PCR 생성물 (UB-HYG)과의 조립을 위해 MAID 6177 (116 kb LTVEC)의 특정 서열을 표적화하도록 설계되었다. PCR 생성물은 벡터와 중첩되는 50 bp를 함유하였다. 다음과 같이 3' 엑소뉴클레아제 활성이 결여된, 단리된 열안정성이 아닌 5'에서 3' 엑소뉴클레아제의 사용에 기초하여 등온 1 단계 조립을 사용하였다. 반응이 다음을 함유하도록 설정하였다: 100 fmol 각각의 dsDNA 기질, 16 µl 5X ISO 완충액, 16 µl T5 엑소뉴클레아제 (0.2 U/µl, 에피센트리), 8.0 µl Taq DNA 리가아제 (40 U/µl, NEB), 1.0 µl 퓨전™ DNA 폴리머라아제 (2 U/µl, NEB) 및 80 µl가 되도록 물 첨가. 5× ISO (등온) 완충액은 25% PEG-8000, 500 mM Tris-Cl, 50 mM MgCl<sub>2</sub>, 50 mM DTT, 5 mM NAD 및 1000 µM 각각의 dNTP (pH 7.5)이었다.

[0232] 이는 조립될 1.25 fmol/µl 각각의 dsDNA (또는 45 fmol/µl 각각의 ssDNA), 5% PEG-8000, 100 mM Tris-Cl pH 7.5, 10 mM MgCl<sub>2</sub>, 10 mM DTT, 200 mM 각각의 dNTP, 1 mM NAD, 0.02 U/µl T5 엑소뉴클레아제, 4 U/µl Taq DNA 리가아제 및 0.03 U/µl 퓨전 DNA 폴리머라아제의 최종 농도를 제공하였다.

[0233] 방법은 20-80 bp로 중첩하는 기질에 대해 1.64 µl (0.2 U/µl)의 T5 엑소뉴클레아제를 사용하고, 더 많이 중첩하는 (예를 들어, 200 bp) 기질에 대해 1.6 µl (1 U/µl)의 T5 엑소뉴클레아제를 사용하였다. T5 엑소뉴클레아제를 10 U/µl T5 엑소뉴클레아제 (에피센트리) 농축 효소 스톱으로부터 (T5 엑소뉴클레아제 저장 완충액 중의) 1:50 희석하여 사용하였다. 이어서, 반응을 50°C에서 15분 동안 인큐베이션하였다.

[0234] 실시예 13: 2개의 중첩하는 BAC를 함께 재봉하는 다른 방법

[0235] 다른 방법이 하기를 포함하는 본 명세서에 제공된 방법에 사용될 수 있다: 합성 또는 시험관 내-전사된 tracrRNA와 crRNA를 95°C로 가열하고 실온으로 서서히 냉각시킴으로써 반응 전에 미리 어닐링하였다. 고유 또는 선형화된 플라스미드 DNA (300 ng (약 8 nM))를, 10 mM MgCl<sub>2</sub>의 유무에 관계 없이, Cas9 플라스미드 절단 완충액 (20 mM HEPES pH 7.5, 150 mM KCl, 0.5 mM DTT, 0.1 mM EDTA) 중의 정제된 Cas9 단백질 (50-500 nM) 및 tracrRNA:crRNA 듀플렉스 (50-500 nM, 1:1)와 함께 37°C에서 60분 동안 인큐베이션하였다. 반응을 250 mM EDTA를 함유하는 5X DNA 로딩 완충액으로 중지시키고, 0.8 또는 1% 아가로스 겔 전기영동으로 분해하여, 브롬화 에티듐 염색으로 가시화하였다. Cas9 돌연변이 절단 분석을 위해, 반응을 아가로스 겔에 로딩하기 전에 5X SDS 로딩 완충액 (30% 글리세롤, 1.2% SDS, 250 mM EDTA)으로 중단시켰다.

[0236] 인공 crRNA와 인공 tracrRNA는 인간화 HLA-DR BAC와의 조립을 위해 인간화 HLA-DQ BAC의 특정 서열을 표적화하도록 설계되었다. 벡터는 각각의 벡터 상의 2개의 부위에서 Cas9 절단에 의해 생성된 서로 중첩하는 ~70bp를 함유하였다 (도 2를 참조한다). 다음과 같이 3' 엑소뉴클레아제 활성이 결여된, 단리된 열안정성이 아닌 5'에서 3' 엑소뉴클레아제의 사용에 기초하여 등온 1 단계 조립을 사용하였다. 반응이 대략 다음을 함유하도록 설정하였다: 100 fmol 각각의 dsDNA 기질, 16 µl 5X ISO 완충액, 16 µl T5 엑소뉴클레아제 (0.2 U/µl, 에피센트리), 8.0 µl Taq DNA 리가아제 (40 U/µl, NEB), 1.0 µl 퓨전™ DNA 폴리머라아제 (2 U/µl, NEB) 및 80



$\mu\text{l}$ 가 되도록 물 첨가.  $5\times$  ISO (등온) 완충액은 25% PEG-8000, 500 mM Tris-Cl, 50 mM  $\text{MgCl}_2$ , 50 mM DTT, 5 mM NAD 및 1000  $\mu\text{M}$  각각의 dNTP (pH 7.5)이었다.

[0237] 이는 조립될 약 1.25 fmol/ $\mu\text{l}$  각각의 dsDNA (또는 45 fmol/ $\mu\text{l}$  각각의 ssDNA), 5% PEG-8000, 100 mM Tris-Cl pH 7.5, 10 mM  $\text{MgCl}_2$ , 10 mM DTT, 200 mM 각각의 dNTP, 1 mM NAD, 0.02 U/ $\mu\text{l}$  T5 엑소뉴클레아제, 4 U/ $\mu\text{l}$  Taq DNA 리가아제 및 0.03 U/ $\mu\text{l}$  퓨전 DNA 폴리머라아제의 최종 농도를 제공하였다.

[0238] 방법은 20-80 bp로 증첩하는 기질에 대해 1.64  $\mu\text{l}$  (0.2 U/ $\mu\text{l}$ )의 T5 엑소뉴클레아제를 사용하고, 더 많이 증첩하는 (예를 들어, 200 bp) 기질에 대해 1.6  $\mu\text{l}$  (1 U/ $\mu\text{l}$ )의 T5 엑소뉴클레아제를 사용하였다. T5 엑소뉴클레아제를 10 U/ $\mu\text{l}$  T5 엑소뉴클레아제 (에피센트리) 농축 효소 스톡으로부터 (T5 엑소뉴클레아제 저장 완충액 중의) 1:50 희석하여 사용하였다. 이어서, 반응을 50°C에서 15분 동안 인큐베이션하였다.

[0239] 실시예 14: BAC 벡터로 삽입물을 조립하는 다른 방법

[0240] 다른 방법이 하기를 포함하는 본 명세서에 제공된 방법에 사용될 수 있다: crRNA와 tracrRNA를 하이브 완충액 (10X 완충액: 20 mM Tris 7.5, 100-150 mM NaCl, 10 mM  $\text{MgCl}_2$ , 1 mM DTT, 0.1 mM EDTA, 100  $\mu\text{g/ml}$  BSA)에서 100  $\mu\text{M}$ 로 용해시킨다. RNA를 어닐링하기 위해, 10  $\mu\text{l}$ 의 100  $\mu\text{M}$  crRNA와 10  $\mu\text{l}$ 의 100  $\mu\text{M}$  tracrRNA를 80  $\mu\text{l}$ 의 어닐링 완충액에 첨가한다. RNA를 90°C 온도 블록에서 가열한 다음 블록을 히터로부터 제거하여 벤치 상에서 냉각시킨다. RNA의 최종 농도는 약 10  $\mu\text{M}$ 이다.

[0241] BAC를 분해하기 위해, 깨끗한 맥스프랩 BAC DNA이 사용되고, BAC를 하기 혼합물에 의해 분해한다.

[0242]		1X
[0243]	BAC DNA 500ng	X $\mu\text{l}$
[0244]	BSA	0.5 $\mu\text{l}$
[0245]	RNA	2 $\mu\text{l}$ (각 1 $\mu\text{l}$ 의 tracr:crRNA 하이브리드)
[0246]	Cas9 (4.5mg/ml)	1 $\mu\text{l}$
[0247]	10x 완충액	1.5 $\mu\text{l}$
[0248]	H <sub>2</sub> O	15 $\mu\text{l}$ 가 되도록 첨가함

[0249] 37°C에서 1시간 동안 분해한 후 30분 동안 탈염한다.

[0250] BAC를 조립하고 삽입하기 위해, 플라스미드를 분해하거나 PCR을 수행하여 삽입물을 생성한다. PCR 반응의 경우, 적은 분취액을 겔 상에 실행하고 깨끗한 생성물을 찾는다. 생성물이 깨끗하지 않다면, 겔 추출 대신에 PCR 클리핑을 수행한다. BAC:삽입물의 1:1 내지 1:6의 물비가 요구된다. 보통, 50 ng의 정제된 삽입물이 효과적일 것이다. 하기 반응 혼합물이 사용될 수 있다:

[0251]	BAC 분해물	4 $\mu\text{l}$
[0252]	삽입물	1 $\mu\text{l}$
[0253]	조립 혼합물	15 $\mu\text{l}$

[0254] DNA와 혼합물을 50°C에서 직접 PCR 기계로 또는 얼음 상에 첨가한다. 50°C에서 1시간 동안 인큐베이션한다. 0.5  $\mu\text{l}$ 의 프로테이나아제 K (20mg/ml)를 첨가하고 50°C에서 1시간 동안 인큐베이션한다. 30분 동안 탈염하고 DH10B 세포에 8  $\mu\text{l}$ 의 반응물을 전기천공한다. 10  $\mu\text{l}$ 의 BAC 분해물을 펄스장 겔 상에서 실행하여 분해 효율을 확인할 수 있다. RNase가 없는 물과 완충액을 사용한다. 최종 반응 완충액은 하기를 함유한다: 20 mM Tris 7.5; 100-150 mM NaCl; 10 mM  $\text{MgCl}_2$ ; 1 mM DTT; 0.1 mM EDTA; 100  $\mu\text{g/ml}$  BSA; 최종 부피는 15  $\mu\text{l}$ 가 되도록 함.

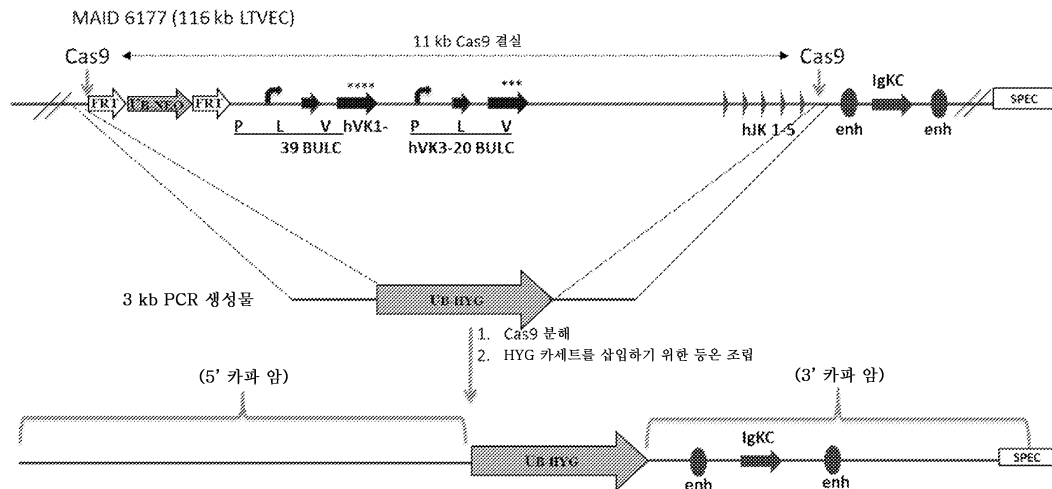
[0255] 실시예에서 사용된 tracr RNA 서열은 다음과 같다:

[0256] CAAAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUC (서열 번호: 9). 이러한 CRISPR RNA (crRNA)는 다음을 함유한다: (1) 표적 서열에 상보적인 약 20 뉴클레오타이드의 RNA 및 (2) tracrRNA에 어닐링되는 꼬리 서열 (GUUUUAGAGCUAUGCUGUUUUG (서열 번호: 10)).

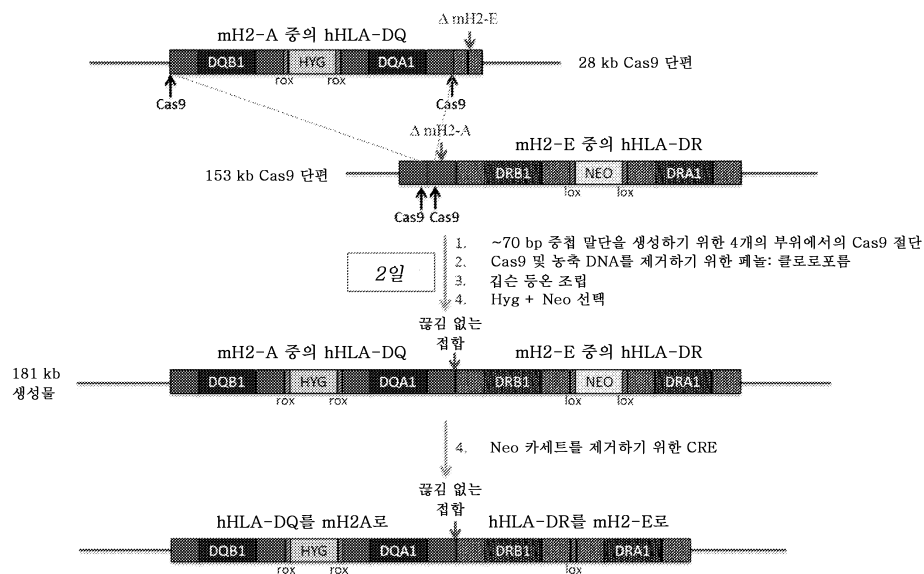


## 도면

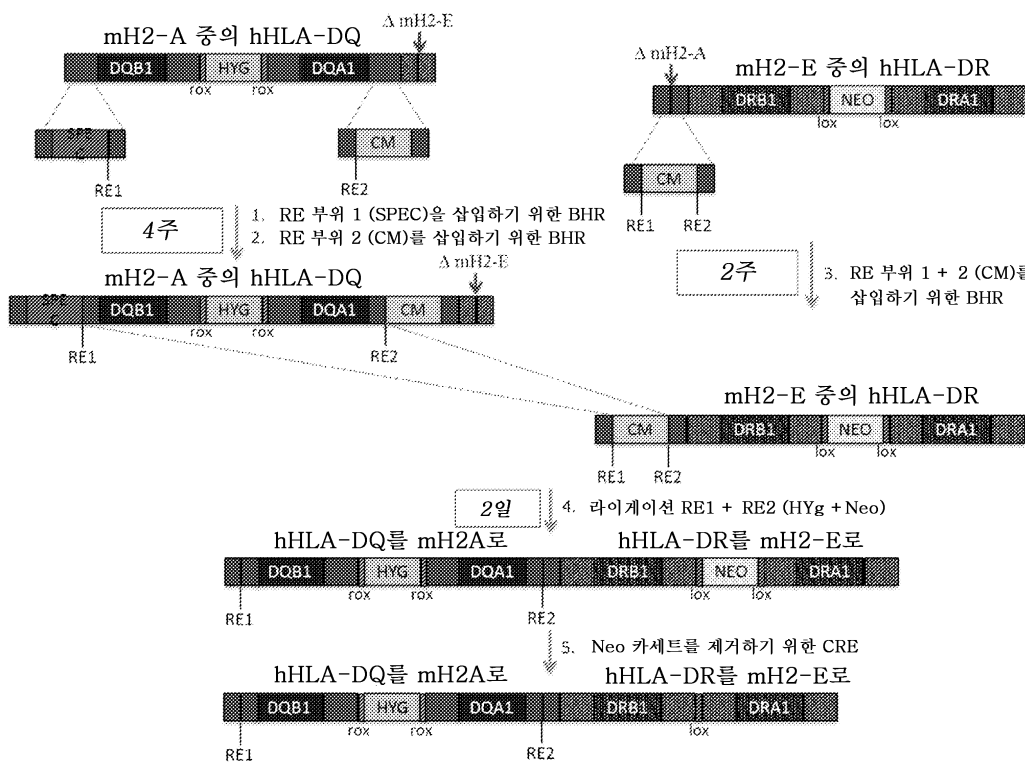
### 도면1



### 도면2



도면3



도면4

Cas9 / 집슨 조립 방법에 대한 클로닝 효율				
구축물	사용된 Cas9	Cas9를 제거하는 방법	콜로니의 총수	정확한 클론 수(%)
6177(BDLC) + HYG	Cas9-6xHis	프로테이나아제 K	3	3 {100}
HLA-DQ + HLA-DR	6xHis-MBP-Cas9	프로테이나아제 K, 페놀: 클로로포름	1	1 {100}
		SDS, 열 불활성화(heat inact), 페놀:클로로포름	1	1 {100}
		열 불활성화, 페놀:클로로포름	16	16 {100}

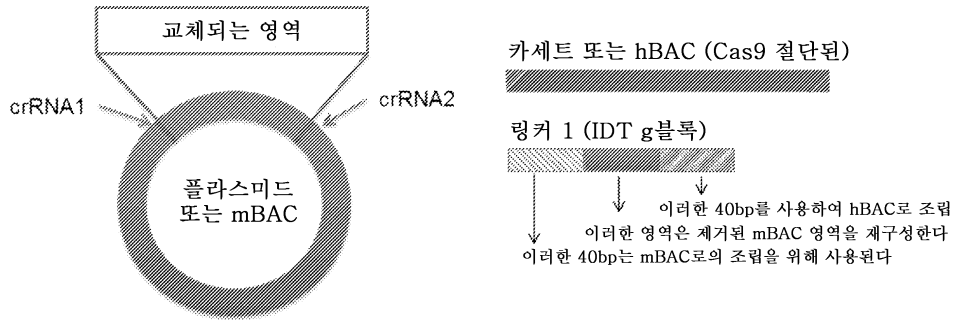
BAC 클로닝 단계에 필요한 시간 §

방법	시간
BHR	~ 1 주
BHR + BAC 라이제이션	> 2 주
2 BHRs + BAC 라이제이션	~ 5 주
Cas9/집슨 조립	2일

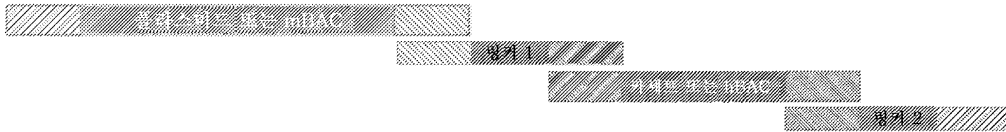
§ BAC 맵시프랩 DNA로부터 시작하여  
대장균 콜로니를 사용하여 종료  
BHR: 박테리아 상동성 제조함

## 도면5

Cas9-절단된 DNA 단편을 생성하도록 crRNA를 설계

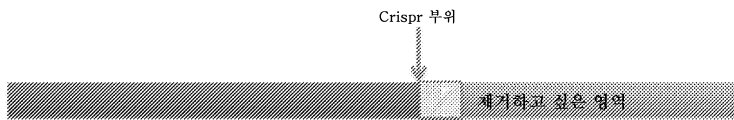


Cas9-절단된 DNA 단편을 DNA 링커로 등은 조립을 사용하여 재구성한다

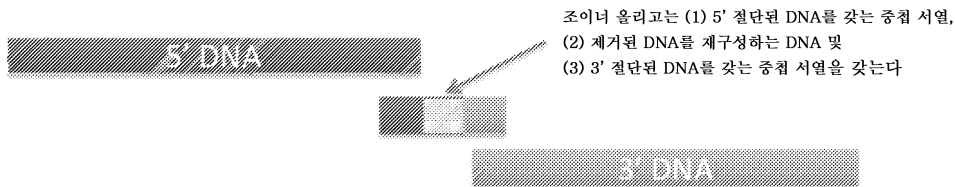


## 도면6

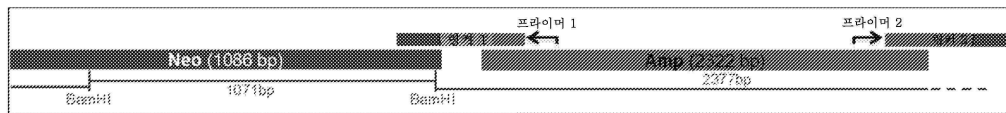
관심 영역으로부터 5' 상류를 표적화하도록 crRNA를 설계:



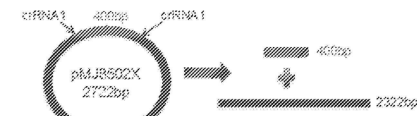
제거된 부분은 조립 동안 조이너 올리고에 의해 재구성될 것이다



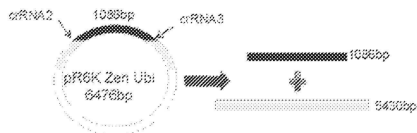
도면7



- 2개의 동일한 crRNA로 pMJ8502X를 절단:

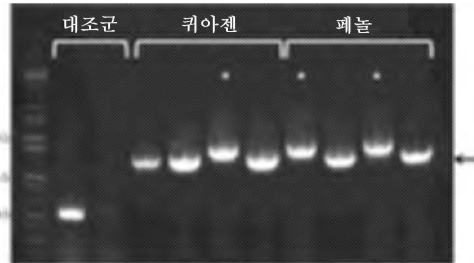


### 2. 2개의 상이한 crRNA로 R6K Zen Ubi를 절단



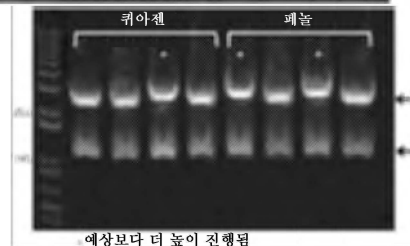
PCR 분석:

정확한 삽입  
=> 1327bp  
삽입없음  
=> 562bp

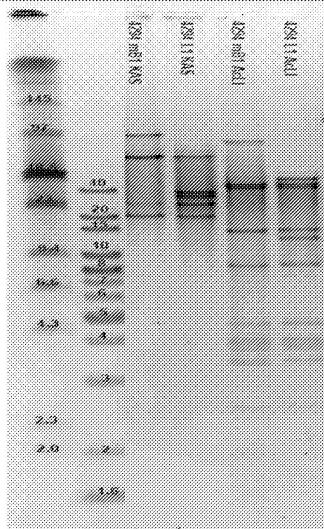
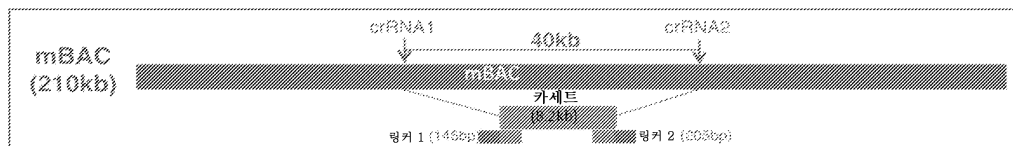


BamHI 분해물:

정확한 삽입  
 $\Rightarrow 1071+2377$  bp  
 pR6K Zen Ubi 단독  
 $\Rightarrow 36+2174+4268$  bp  
 pMJ8502X 단독  
 $\Rightarrow 2683$  bp



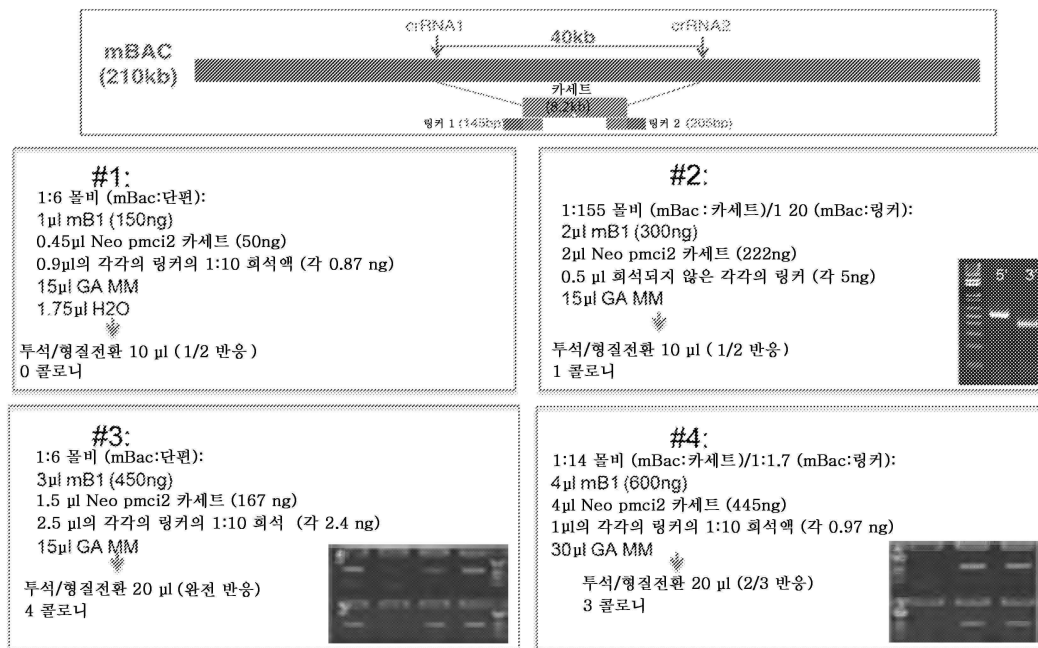
도면8



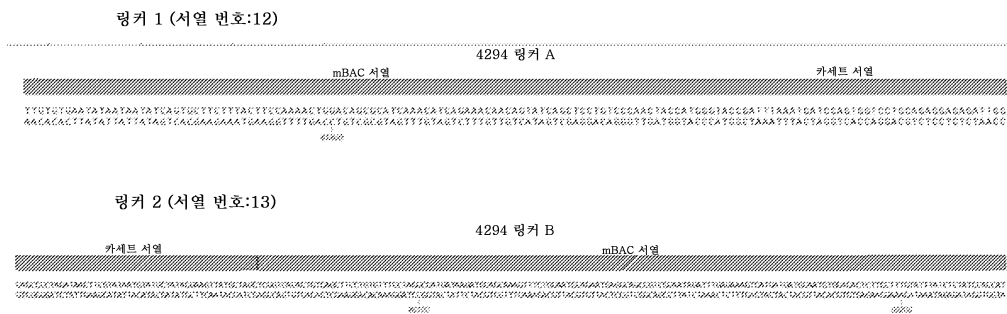
## RE 분해/PFG 점정 (염기의 예상 크기)

	mBAC:	표적 배타:
Kasi:	96399	69632
	66654	33900
	23800	29927
	16848	23800
		16836
		933
		451
		13
Acil:	88026	45466
	40507	40507
	38186	38186
	13075	13075
	7974	11373
	4416	7974
	3845	4394
	3687	3845
	3342	3687
	2477	3342
	1166	2477
		1166

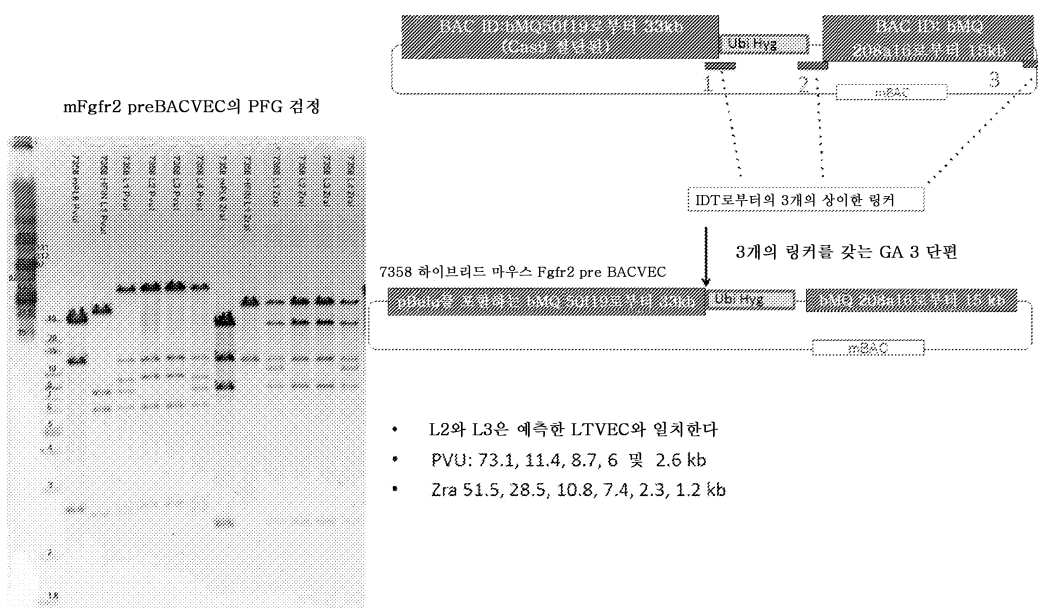
도면9



도면10



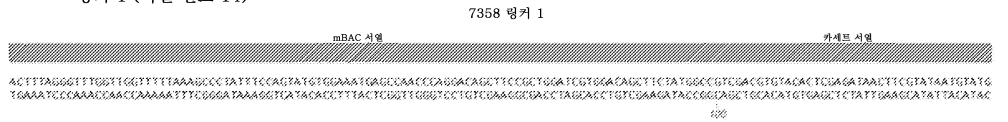
도면11





도면12

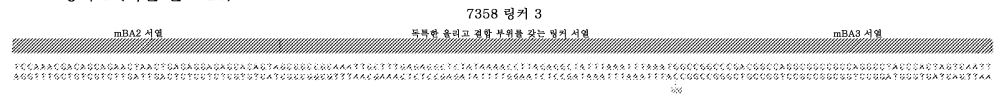
링커 1 (서열 번호:14)



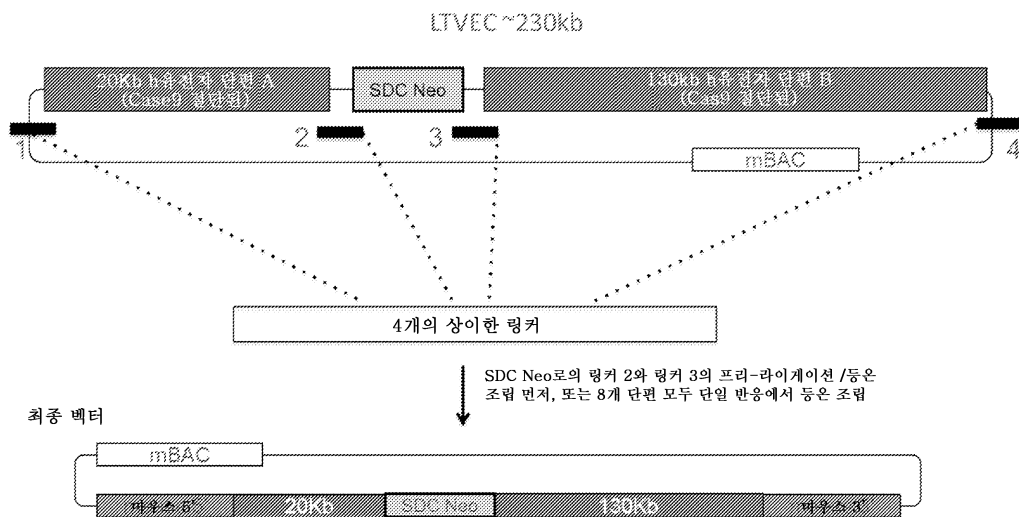
## 링커 2



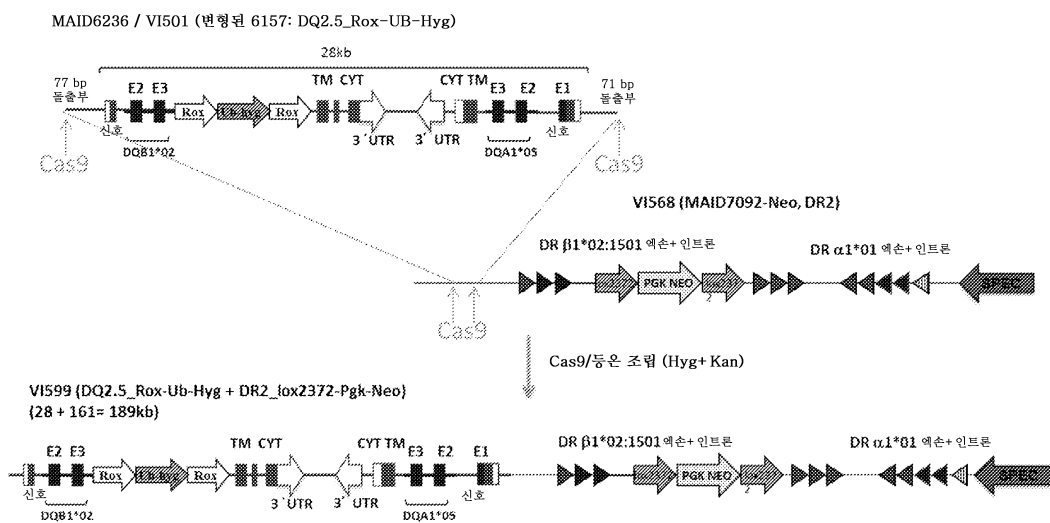
링커 3 (서열 번호:15)



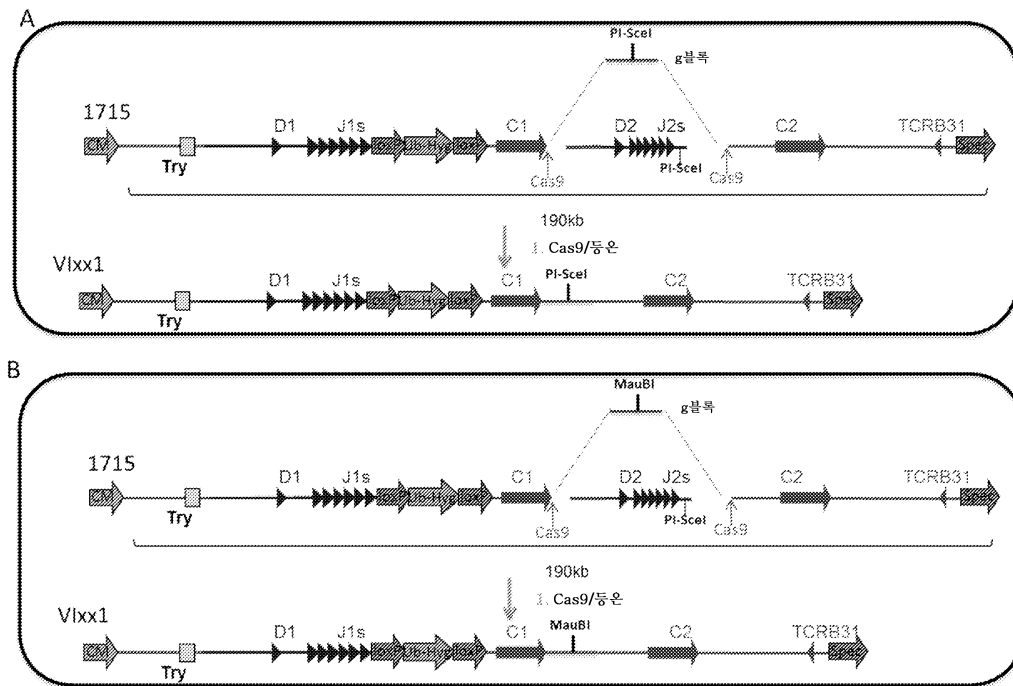
도면 13



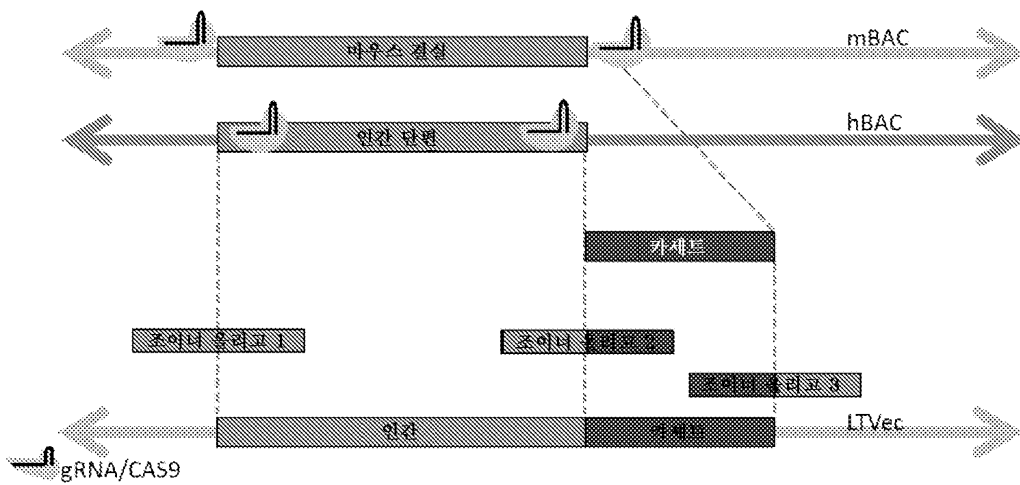
도면14



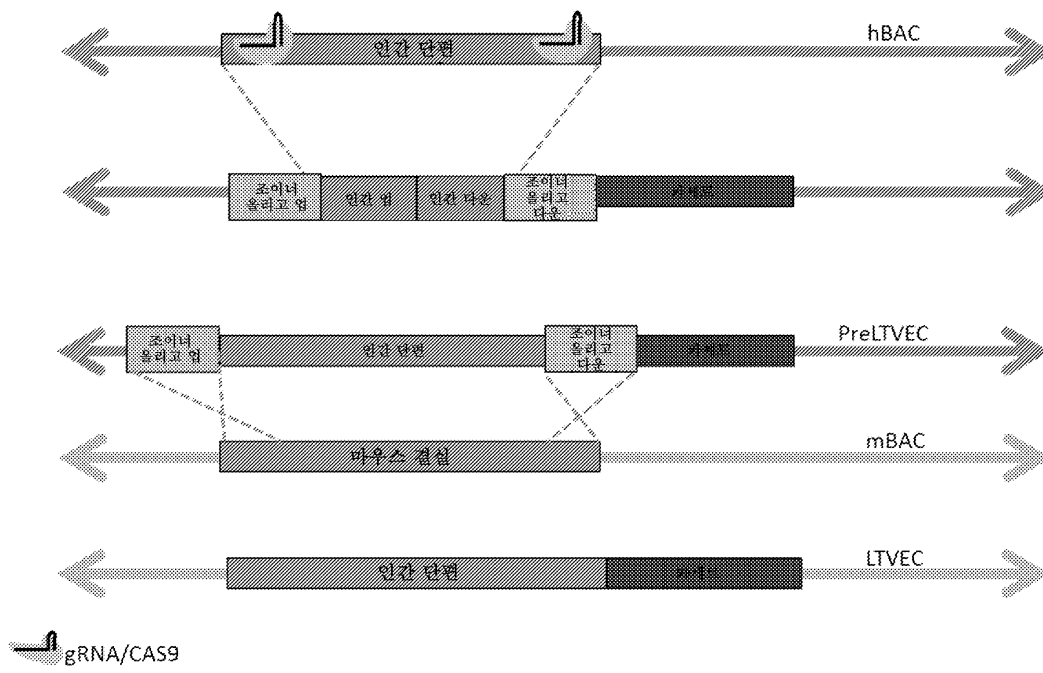
도면15



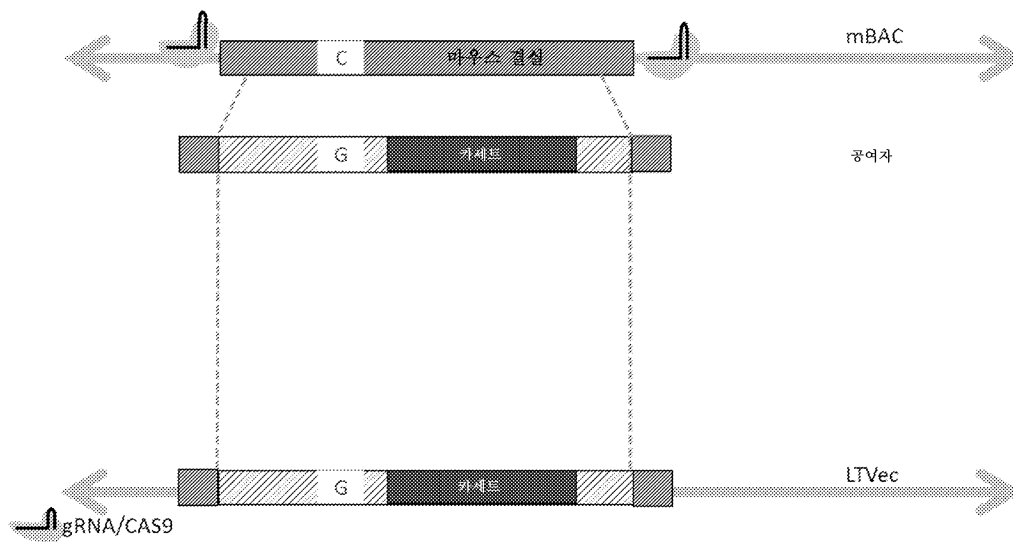
도면16



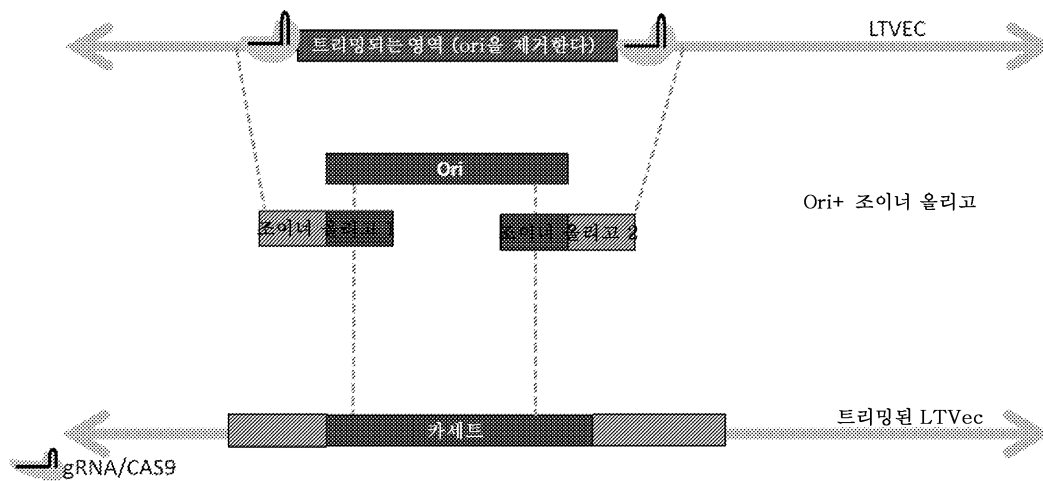
도면17



도면18



도면19



## 서열 목록

### SEQUENCE LISTING

<110> Schoenherr, Chris

McWhirter, John

Momont, Corey

Macdonald, Lynn

Murphy, Andrew

Warshaw, Gregg S.

Rojas, Jose F.

Lai, Ka-Man Venus

Valenzuela, David M.

Montagna, Caitlin

<120> Nuclease-Mediated DNA Assembly

<130> 057766-461002

<140> PCT/US2015/037199

<141> 2015-06-23

<150> US 62/036,983

<151> 2014-08-13

<150> US 62/016,400

<151> 2014-06-24

<150> US 62/015,809

<151>

2014-06-23

<160> 25

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 80

<212> RNA

<213> Artificial Sequence

<220>

<223> Synthetic chimeric gRNA

<400> 1

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60

ggcaccgagu cggugcuuuu 80

<210> 2

<211> 42

<212> RNA

<213> Artificial Sequence

<220>

<223> Synthetic chimeric gRNA

<400> 2

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cg 42

<210> 3

<211> 30

<212> RNA

<213> Artificial Sequence

<220>

<223> Synthetic crRNA

<400> 3

guuuuagagc uagaaauagc aaguuaaaau 30

<210> 4

<211> 33

<212> RNA

<213> Artificial Sequence

<220>

<223> Synthetic crRNA

<400> 4



guuuuagagc uagaaauagc aaguuaaaau aag	33
<210> 5	
<211> 26	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Synthetic crRNA	
<400> 5	
gaguccgagc agaagaagaa guuuua	26
<210> 6	
<211> 12	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Synthetic tracrRNA	
<400> 6	
aaggcuaguc cg	12
<210> 7	
<211> 50	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> Synthetic tracrRNA	
<400> 7	
aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu	50
<210> 8	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> a target locus that is linked to a guide RNA	
(gRNA)	
<220>	

<221> misc\_feature

<222> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,  
19, 20, 21

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

<400> 8

gnnnnnnnnnn nnnnnnnnnn ngg 23

<210> 9

<211> 41

<212> RNA

<213> Artificial Sequence

<220>

<223> Synthetic tracrRNA

<400> 9

caaaacagca uagcaaguua aaauaaggcu aguccguuau c 41

<210> 10

<211> 22

<212> RNA

<213> Artificial Sequence

<220>

<223> Synthetic crRNA region complementary to tracrRNA

<400> 10

guuuuagagc uaugcuguuu ug 22

<210> 11

<211> 89

<212> RNA

<213> Artificial Sequence

<220>

<223> Synthetic tracrRNA

<400> 11

guuggaacca uucaaaacag cauagcaagu uaaaauaagg cuaguccguu aucaacuuga 60

aaaaguggca ccgagucggu gcuuuuuuu 89

<210> 12

<211> 145

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic -Confirmation of seamless assembly from  
mBAC to cassette

<400> 12

ttgtgtgaat ataataatat cagtgttctt ttacttccaa aactggacag cgcatacaac 60  
atcagaaaca acagtatcag ctctgtccc aactaccatg ggtaccgatt taaatgatcc 120

agtggtcctg cagaggagag attgg 145

<210> 13

<211> 205

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic -confirmation of seamless assembly from  
cassette to mBAC

<400> 13

cagcccctag ataacttctg ataatgtatg ctatacgaag ttatgctagc tcggtcacac 60  
tgtcagcttc ctgtgtttcc taggcatga taagatgcag caaagtttct gcaatgcaca 120  
atgaggcagc cgtcgggaata gatttgagaa agtcatgatg atgcaatgtg cacactcttc 180  
ctttgtatct atctctatcc accat 205

<210> 14

<211> 138

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic -confirmation of seamless assembly from  
mBAC to cassette

<400> 14

acttttaggt ttggttggtt tttaaagccc tatttccagt atgtggaaat gagccaaccc 60  
aggacagctt ccgttgatc gtggacagct tctatggccg tcgacgtgta cactcgagat 120  
aacttcgtat aatgtatg 138

<210> 15

<211> 147

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 15

tccaaacgac agcagaacta actgagagga gagcacagta gcggccgcaa attgctttga 60

gaggctctat aaaaccttag aggctattta aatttaaag gccggcccga cggccaggcg 120

gccgccaggc ctaccacta gtcaatt 147

<210> 16

<211> 9

<212> PRT

<213> Unknown

<220>

<223> Synthetic

<400> 16

Leu Ala Gly Leu Ile Asp Ala Asp Gly

1 5

<210> 17

<211> 49631

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<220>

<221> misc\_feature

<222> (22396)...(22533)

<223> Linker 1

<220>

<221> misc\_feature

<222>

> (22494)...(25426)

<223> Cassette Sequence

<220>

```

<221> misc_feature
<222> (25427)...(25595)
<223> Human Spacer Sequence
<220>
<221> misc_feature
<222> (25596)...(40791)
<223> BMQ-208A16 sequence

<220>
<221> misc_feature
<222> (25387)...(25672)
<223> Linker 2
<220>
<221> misc_feature
<222> (40792)...(40858)
<223> Unique additional sequence of linker 3

<220>
<221> misc_feature
<222> (40752)...(40898)
<223> Linker 3
<220>
<221>
> misc_feature
<222> (1)...(22395)
<223> bmq-50F19
<220>
<221> misc_feature
<222> (40899)...(49631)
<223> bmq-50F19

<400> 17
gctggagtgt ggtcaggcaa catcccaaa gggatggaga tgccgggacg acacctttag 60
ggaggcagtg gctctggtcc gggattccgg tgctggccat ccctcaccag ccacagcgtt 120
tggcgcagga gggatcgccg cgcgctggg gctagggggc gaactggacc gacttttctt 180

```



agttcgcta gctgctccga ccgctgccgc gccgagatgt tgaaagcaca ggcgagttct 240  
 aacttgcgcg ctattcttt cagcgcgggg gaatcggtcg agggccctgc gtggcgctgg 300  
  
 ctccaccct cgcggccagg gggcaggcgc gggaggccgg cttcggtcc gtgcccctgc 360  
 aaacttccca agaccttct tctcccccc acctcacccc ccagttcaat aaaatctacc 420  
 cttaaaggca gacttgcttt caaatccacg gcacccatta tgtgtttggt gtgaaacgct 480  
 atcaacattt aaaactctat tgteccaagc gtcccaaate cctgtaaatc ttccaccagc 540  
 ctggactcat ttcatctga aaagcctgtt tagtttgaat agaaaagcaa tcaggcgccc 600  
 ctctcgctct cgttggaaatg tcaattaaaa tgcagatttc tcagagctct ttagcgcccc 660  
 aagaagtggg acaaacagg atatttcagg ctgacaaatg aaagaaatgc tacaatgaag 720  
 tgggtggcg atgtgcacc caaactgctt ggagtacca ctgaaagagt aggtcaggga 780  
  
 ttatggtctt acttacgaca gcttatattt ttgggtttc gttgtgttta gggccccccc 840  
 ttggtgtccc cccccccat gagcccatga cagctccctt ccctattcag ccccgaggag 900  
 aagtaaggga gccttgaacc agggtagaga ggctacattt agtattaacc tgggagtgtt 960  
 gacttctccc aggagtaate cacttgagaa caaatgccca attgctctgc ccgctgaggt 1020  
 atcctggaac taccctttaa ggtagcagta ccgctgcac cgccccctc cccaagggc 1080  
 ttgcctaaa ttaacctgcc ttcttgagg acaggggaga gtgtgtaaac gtgtataaca 1140  
 ctgcgaagc tcaccagccg ggccctttcg gccgggtccc ttgcctgtc ttggaggca 1200  
 gacttgtgtg gagatgacce caaggggcgg gtggccgtga agagccatcc gtcagagtga 1260  
  
 gggtagggac tctccctcg taggtgaga agagagtatc ctttcagggg gaaaaataaa 1320  
 cacgtgggg ctttctctgg ggttcagcct ccaggaagga ttatggtatt gaaggcagga 1380  
 agctgggatt gtggcccca gcagcatgtt ggcctgtgt tccaacacg gaggccttggg 1440  
 acctaattat cctgcctagg aggtcgctca gcacttttgt ccaactcggg gaggagctgt 1500  
 gcagacctgc tgcgctact tctgcctta cagaggtttg aggagggggc tctgtgggg 1560  
 gctgggactt cgaagaacga acgttcaagt tgagtcagcc tggggcactg gccatcttc 1620  
 tcattcagct ggagctgagg tactcctggg tagtggttag tagagacagt gggcccagca 1680  
 ctctgttca agacctactg ggacctgaga ttgcaaagt gctggagagg ggagtttacc 1740  
  
 tgcattctga aagtcttag gaaatcaacg agaatgtttg tgcactttcc ttgacttgtt 1800  
 atgtagaaat agacaaggaa ttatcttttg tgactcttgg cttaagaag aaagaagact 1860  
 tgggggaaca aaaatcctc cagccaacta aaaacactgg gcacctaact gctcatatac 1920  
 ccctggcttt tttgtttage tataaccattc tacctgtgct taaaaaaca accaaacagc 1980  
 agcttctat tcccctcttg gagatggtac gtctctctg ccttagtctc agtgaaggct 2040

gaaaggaaca gatttttagga cggagggttct ggagtggtcg aaatcctgtg tcataattga 2100  
aagcatcaaa agcgacacggg attagaattc tttttctctt tctctctttt tcattaaaac 2160  
gtcaccccat cccagctctc ataaaaatggg catcccagca tccaaagccc atggttttgt 2220

gcgatccttt ccigccattg gtttcagcag attctctaaa gctcgtgcat tctgactcaa 2280  
agattagtca ctgaagacac tgaacaaaca taaagttatt tgtactgtgg taagcttttt 2340  
tttttgggaa attctctgct ttggatctag taaattgagt gcccccttgt aactgatact 2400  
tgggagggtt agccaatagg ttagcgtatt gaaagttccc aggccaatca cataccaggg 2460  
cagcttgtag gtatcatcac cattactaat aaaatcttga attattcattc aagggttgta 2520  
tctttacccc ttgacgtcg gttgcagata tttagttagt atgcctgtac actgccttgt 2580  
agtcagtggg agggaaattca ggctttgaat cccccggttg gattaaactc actctttgta 2640  
agtggctgct tggcggaaga ttgaaataca cgctgcatt cgaaaatgaa ttctgacaag 2700

tgtaactgg tgggaatgtt tttgaagcct tcctgagatt ctttgattct gttggtctcc 2760  
tttctttctg agaaccgttc tgaagcgagg acgtgccgt cagctcagct gaaatgcggt 2820  
tctcagagca gacccttctt ccagtcagcg tcttaaaggc cagctggaat aagagacgtt 2880  
aatgaggctg gccatgcaa gccagcgtt ttaaactcag gtttttctgc agttgccctt 2940  
gaaaggaatg aaggtcaagt tgcttcagca accttcagc tttgatagt gacggaaggg 3000  
cacgtcgcag agctgggttg ctgggtccta cagtgatgtt ttatcttgcg tctcttaaaa 3060  
gtaagcttaa aaaaaaaaaa attagcctac tgcagcttgt ggactagcct ggaaacacct 3120  
gggacgctga ggtgaggatg gaagctttt ccgataatga gaaagaatgt gtttgcaat 3180

gtattgagag gctgagaaat ggttttatcc catctgggtt taagcaagtt ggcactgggg 3240  
gaaaaaactg aatctggctg aatctctctc ttctagtggc agccacagca gcagcagcag 3300  
cagcagtggg agcaggaaca gcagtaacaa cagcaacagc agcacagcgg cctcagagct 3360  
ttggctcctg agccccctgt gggctgaagg cattgcaggt agcccatggt ctgagaagaa 3420  
gtgtgcagat gggattaccg tccacgtgga gatatggaag aggaccaggg attggcactg 3480  
tgaccatggt cagctggggg cgcttcattc gcctggcttt ggtcaccatg gcaaccttgt 3540  
ccctggcccc gccctccttc agtttagttg aggataccac tttagaacca gaaggtaagt 3600  
tcatgcgtgc cattttaagg gtaccaagtc gttttgggga tgtgtctggg ggaagtggc 3660

tttaagtggg aggcctgttt cagccggctg ccatatgagt agtctctctc cgcatacat 3720  
cggagcttag aaggagggtt cttgtctccc aggcattgagt ggagtgggtt ggtttgctct 3780  
gttctttgtg cttgggtgag ggaagcagtg gcagttcttg tttagccagt gccttacagc 3840  
actctggagg ggacgtacct tggcagggtg actgtgcctt tctgcagttg ctctctagat 3900

tgagggaata gccttgaatc acactatctt ttggctaaag gaaataggca gcctctgaaa 3960  
 gctgactttt tttttctttt tccgcattgt ttaagagaaa agaaggttct gaagttgagc 4020  
 atggagagcc gtgcatgct ggatcggttt ttaagctggt gtaagctctt tgtgctttca 4080  
 cccggcatca cagagtgggc aggtttcatg ttgggaagat tggaaagtga atttgccaag 4140  
  
 agtcttcccc catctgggga aaagccagat ttcactagtg tgtttggctt tgcacacttg 4200  
 gtgcaaatg tgagaagcta gttgtgagga ggacgtggct gaaatccgga gctgggcaaa 4260  
 gcgctggctc ttctcccagg tcttcagag acgtggctctg tggccaagcc tctctccttg 4320  
 gtgccgcacg ggaatctgtc atcaggaggg aatattggtg ggcgagttat tttttgagt 4380  
 gtaatccgag cgtgacactg cagatcgag cactcatcgc cacttaatga acgtgtttgc 4440  
 tgagggccca cctgggtccg gctggctttg gactccgtca cggctctgag tgctggcagg 4500  
 tcagctgagt tgctgtggct atgcacactg aatcagggtc ctgattcatc cagatcatcc 4560  
 agagggggat tgiaggaggg acaggacccc tcccccaagg gtgacctcaa ggagggtat 4620  
  
 glacccatct gagaggaggg cttgagaaat gggtccccag taagatccac ccagacagac 4680  
 actctccctg gctttgtgtg tatgtcgggc cacacagatg cctggaaatg ttataaatta 4740  
 ccaggtatct ttggaaagga aatgaggtag gacttttgtg catgaggtgt gttcaacata 4800  
 cagctcagc tcttttccg gaaccacctc tctgtgactt atcctgtgac gtcagggaga 4860  
 gtgtaatctg caacagtac atgttttcaa agggcttaat gtgaggggga aaggattggg 4920  
 tttctgaaag tctggctcgc acttctttaa tttgttaat aattaaaatg gatgtcccc 4980  
 taattgccg ttgtccctgg agtgtgtggc tcagactaa ctaagggaagc tgagctagga 5040  
 tttctacag cgtgggcttc agaaacagcc ccggttagga aagaattgtc atttttcatt 5100  
  
 tggactctcg gggcagtgtt gctgtgagtt gatttcagtt gcagagtata aaatggtcct 5160  
 ggagggtttc ctggactgca tctaattacc tcagaaaggt tacaagatgt ttgtactcgc 5220  
 aaggaggagg caggtggggg agaggaagga cagtgggctg gactcccca aatggctctt 5280  
 tgtgtaagaa ccgatatcca acaatgctca cattgttgaa agcagatccc accacctggg 5340  
 gacctgtagg tacatgtaag gttaggagg gaggtgaga agtctccgaa gttgtaggtc 5400  
 acactttgcc aatgcccctg ggtacacttt gctaggctca gagtttgcag gaggttcgaa 5460  
 tcacatatag agttgggaga cgctaagaaa aagaaaagaa aaagaaaaaa ggaaaaaagg 5520  
 aaaatgtctc aagggtgga gtttcaccag agcaagcttg ggaaatgcag agaaacccca 5580  
  
 gagccttgat tgggtgggac tctttatcaa tagtactga acagtagtac catccccaga 5640  
 tgctttctga ggaccagctc aagagattta gttttacca gtgacctgga cagaaagcag 5700  
 aaagcacagc tctggcatt gatggtggcc ttggccatcc ccatccccag caagctgggg 5760

acaagggggt gcacagttct cagtcgagca aacacggtac cctgagatga atgttgcttt 5820  
 tggatggagg aggtggatgat gctggatttc ggcagggtct gtgctcactc tccttgctctg 5880  
 ttagaccaac attgccactg acatccaggc catcaagcta gaggctaggc tccatgctag 5940  
 gctctgggtg ctaatgtgtg cataitgtga tctctccagc cgccatattt gatgcagcca 6000  
 ggacttcagc taacactgag ttcagcttct gtctcctgaa gctttacat ggaaggcatc 6060

cgtttgctaa tttagaagct cagtttagat aatgtctatt gggccggaca aatatgtaat 6120  
 caggaagttc ctgaaagag cctgtgcctc actactaagg agcccttttg accctctagg 6180  
 gagatgttat gttcagtcac gtagttctgt gcagtgtatg tagccatgca atgtatgtcc 6240  
 tcaccccgaa tctatcctg tccgtgtgtc tctggacact ttctcaagtg gcagcagcag 6300  
 gattgggtca agtcagttga cctaagaagg cagtcactc tgtaagattt tcctcggat 6360  
 ttcagaatag aaatgattgt atccagctgg tcatccctgt gacaaaggac aacagtatca 6420  
 acagttgggg acttcggagg ggtgtgtccg attctaagta ctgttctgtt gattcaaadc 6480  
 ctgaatgttc ccagtgtagt caagcttgat tactgccagt ctcggctctt actttcagat 6540

tccccctgac gttgtcacct gctctgggtta attaagtcac tgttgacatt aagggaatct 6600  
 gtttacccca gccagtagg agctaaaata aaagggtttt cccaaacca aacccttaat 6660  
 tactttccca cctctgcta agtgcaaagg gacggcctgg ggggtgggtg ggggtgggag 6720  
 tgaggagta attacatgc cttaaaaaac acccaccatt tcttgggcag tcttctgggt 6780  
 tgatgtgtgt ccgattgaa gtgagccagc gaaaacctcg tgagtgtgag gtctactgtg 6840  
 agacctgtg aagggttccc cccccctcac caccaccacc acagggtagt tcaagtcctt 6900  
 tgtcagagag tatcctacat gctgtggtct cagccccca catttaattg ccagttagaa 6960  
 gaagagaaaa gaaatatctg cgtgtgtgcaa gtggatgatt taacaggagt cttgtgtttc 7020

ctattatctg cattttttgt tcctcagtgat gactgtgaat atttaagagt tgactgtaac 7080  
 ttgatagttt cgctgaggaa caagggttta ttcttggtca attaaaccaa atgcaggcgc 7140  
 atgtgtttaa acacacaatg aagtacacat tctttattag aatatagtgt atttcacaat 7200  
 tcatgggcaa ggaactgtgt ttaatattac ttctagagca aaaatctggc cagcccagaa 7260  
 aattggcatt tatataactc ttcttctgtg gcttcactg atctgaatag agcaagtttg 7320  
 tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtacc tgtctgtcta tgttgccgtg 7380  
 ttatttcaat cagttaaaag actaaataac ttacttaaaa aaaaatagcc accccttctc 7440  
 agatagcctc taaagacttt atgtgttttt taagccttat ttttaatta ttttaattg 7500

gggtctgtgt aggcctgtgc acatgagtgc aggtccccct ggaggccaga ggtttttgat 7560  
 gctagcccgg gatttggttc tggaagtga tgtggatcct ctgcaagaac aatatagact 7620

tacaaaaact gagctatctc ttctgctcca atacagtcca acattttttc ttttttcttt 7680  
 ttcttttttt ctgttgaggg tgagaattca aacaaccaag cattccaata tgagatgttt 7740  
 ccaatatctg atttaaatgaa taatcacatg gttatgaaat aactggggta gtgttttaac 7800  
 aggaagggga tgittaatgt tcacattctc tgtggagtgc gtgcagtagc cccgtgcctg 7860  
 cagtagccct gtgccacact tatagacagt ttggctactt acatagttag ggtggtcagt 7920  
 aaaagacaac taagtcctt tcatacagct cgtcttaac tttccattt ctgattgaat 7980

ccctgggatc gatccagcag ggtgtcttgi cttggtcagc agctaggagt tattttgggg 8040  
 gaagggatgc tgcaggctat ttacagata attatgggtt tctgtgcag aactgtccct 8100  
 gtaggggctg gagcaagtga tgattctgtg attaagagca cttctcctt ttgcagagga 8160  
 ccagtgttgg gtccccaaca cccacaggga agtactcctg attgtcttta tctccagggtg 8220  
 ctggggggcca gcgcctctga cctactgttg cctgcagggtg ccttctcat gtgggctgcc 8280  
 ctgcccctta ctgtctttgt ggcttcgtag agaattgatgg gaaaaaattc caagatggta 8340  
 gtcccactgg tgactaaagg tgttttagtag tagctttttc ttaaataca gttggtagc 8400  
 aagatagtgg tgcacacctt taatccaac actaggagg cagaagcggg tggttctttg 8460

agattgaggc tagcctggtc tacagagtga gtgccaggac tacacacaca cacacacga 8520  
 cacacacacc ccagaaagct tgaagttgta gttttacgaa agtgtattta accgtcagga 8580  
 ctaactatga tctttctttt gggctggtag ctgatggttt ggtttttttt ttttttagatg 8640  
 ggcatctccc acagcctggc ttgggatttg cctttagct caggtcggtc tagaactttc 8700  
 aatcccccta cctcaacttc cactcctaatt tgcggcat ccttgaagag catgtgtctt 8760  
 gattttctgt aattttgaaa aacttggcct cggattttat ggcttactta tctttatgtg 8820  
 tatctttatg tgttttgctt gcatgtatgt atgtgtacca tgtatgttt tagtgcttgc 8880  
 agagaccaga agaggacttg cggctccctg gaaatagagt tatctacgtg gttttgagct 8940

agcaccaggt ggtcttctact tccccgtgg cctccagcc cttggattaa atgtggaatg 9000  
 tgctgtttgc ttgcttgaag ccacatagg cagtgcagg ctttgggac tttctacact 9060  
 ctgagaataa atgaaagtc acttgcttgc tcttggttgg gtcaagtcag ggagctaaac 9120  
 tatcacatac ctctcttta acttcttgc caactaaaga atcatgaatc ccaagccgtt 9180  
 tctggacaga gagaattcca ggttatggtg accatgtttt atgaggatgt taaaaatagc 9240  
 tcctaaggag gatgctgaca gattcaggaa ggagaacccg gcctcatgtt tatttgggtg 9300  
 ttattttatg agcatgttcc tgagacatct caatcctgag cactaaggaa gtcaacacat 9360  
 tgtttcctaa ccttgaact tgtttttcac ttcttatacc tgacagtta caaatactgg 9420

ttcccccccc ccccatgtg tggccaagt ttttaaaggt atctaacacc gaaaatggcc 9480

aatttgggtg gctgttatag atcaaaagga gatctttgag actagagatc tctgtcaagt 9540  
ttattctctt tggaaccct tcaagttcac attgagagct gacagttggc tagccctaga 9600  
gtcatgtggc ttgcttcaag cgcctctcc cccattccac ctcaaccct tggactgcca 9660  
ctaagactgt tgcttagctg attgtagcag gtaccttgct gaatgtgtaa cctgtataga 9720  
ttatgtgttt cagatttaaa accactcagg tctttaaga ctaaggatc tgatccgaca 9780  
tttgacttaa aattttaagt agaaactaag taaagttgtt ttgaatagta tgtgttgtgt 9840  
tttctggagg tacagtctca taggaaatcg cccttgggtg ctgagtttga atgtgcctac 9900  
  
tatctacttg accttagtca agtgagataa cctggttgaa attccaagat aatatctgtc 9960  
taaattgcac agattgaata cacactggac tgtagttcct ggcccagtg aggcgcaggg 10020  
taagtgtga tttctccca cccacacct ttgtcaaact aaataaaacc cacatctcaa 10080  
agacctaata tgatgcttc cttgtaatct ataagtata atgtcagatt ttcagacctt 10140  
aggccttctt ttatccaact cttttttgc cctcgggtt ttgcaagccc cctggtgttt 10200  
agacatgtga ccctttatct gcttacagtc taggtgttca aggttgactt tttttttt 10260  
tcttctgtta aggaagtcaa ccgtagccac ccagcacata gtgagaatat gtcattgtca 10320  
  
tgggtatatt ttggcaggag agtcctctgt ttgaggtttt caaataatcg atgtaggcca 10380  
gtgaagggtg gtagagaggt tgggtgtgagc ttggttgggt gtgttgggtg tgagcttggt 10440  
tgggtgtgtt ggatgtgtta tagtgtctg tgcctgtcc aagccagtga agaaccatc 10500  
caccattgca ggtgttgctt gtcttttgc atcttctct gtaatgccac catccatttg 10560  
cctgccaggg gagctagggt ctgggcttcc ggtgggctgt atggcaggga gttcacagaa 10620  
ctgtgtctgg gtccagacta gtggaagagc tggacattca tgtgcatggt tcctctaaga 10680  
ggggcttgtg atggcagagg ctcagggtga gatcgtgtcc ttcaactcag tccttgggt 10740  
  
aatgatggtt tccatgaaga cacttagct cctgctcttt gctccgtgcc ttgtgataag 10800  
atgtgaagg tgcagatgct gagagcgcca ggcctttatt aagtcctgt aagcggtca 10860  
catgtgctag ggatgttgac aaattgcccc ttccaacaa acaggcagat cccaggatcc 10920  
catttcatga ataaaatttt tgcaattctt agagatgctg tggtttccgg acaccttcac 10980  
agtgaccaca caccaccct ttaggtgaac taattggtgg aagatggatt tcacagctca 11040  
ttctctctc ctacgaaga ggatagatat ttgatggagt gttaggcacc cctctgttt 11100  
ttttttttt ttaccacct actttggact ttaaaactca gaggacaggc tggttggttc 11160  
  
tgtttctct tcacctccc acaccactt ccttaagtcc tttgaagaag agtttcaggc 11220  
aataaaaatt ttctagcact tatattctgt agttctggtg cgatgtaggg agttggtcca 11280  
taccctctgc taccgtgggg accagtggga cacagcacag agtcctagac gctgacttat 11340



gctgagtcac tggagaaaag ctccagaaca gaagggccac cttgctcctg cctgactgtt 11400  
cctcatcggt aggtcttctt ttctcggat cctccagacc ttagcttcat tgagttgctg 11460  
ttttgatagc atttcaagct tctcttttca gcattttctt ctttttgcaa caaggtggga 11520  
aatcaaaggc cacctggact ggactacctg gaactcctt aggctgtggt catgaaaagg 11580

acagggtgtg aggcctttcc gggaactttt ttctccagag attagggact cacctatctt 11640  
ctctccatct ctatctctc ccctctcccc caggaggaaa aagaaaagaa aaaaattcca 11700  
agaacagaaa gtgtggccct aggggcaaaa gaagccagga aatgaagccg ttttaaaagc 11760  
cagcaaaagct cactttgtg actttaaaaa aaaaaaagt gacctctggt cgcagctggg 11820  
tatggagggt acagtactg actaggatac tgatcttgt agaggtcatt tgtgaaatgg 11880  
gtggggatgc tcagagacag caggtatgaa gtaaggcaag gtcactgctg aagggaaga 11940  
ccccccaca tcagcttccc tcagagctgt acagcctttg catataacga ccacttcca 12000

ggctggtaga gagaagatgg catctctaga tgtcttttc tagtctcagg gtaattagtt 12060  
ccctttgagt cagtttccca acttattggc tgattggtt acctagagtc tcatgtagcc 12120  
cttgaacgct caattcttct gtctctacct cctgagtgt ggaattacag gcaggcacca 12180  
ctttagccag ttcccatctc atctttgttg tagaaaagtg ttacccttg aagggtggc 12240  
cagtctgagg aagctgcacc gcgctgtagc ttcccttga cgtctcttc ctggcacttc 12300  
actctgatgg ctttcttgc tagccatcat ggaggcaagg aaatggccag ggctgagagg 12360  
ccagaaaacc ctgcttctc ttgggcagag taatgatgac ttccctgcct ggcacagtga 12420

cacacctgt cctcgggaag ccacaatgt tgggcacctc gcctggatct tccagactc 12480  
agtggctgag tcigaaggga gccactttc agatttgctt gctttctgaa agccttcct 12540  
ccaggcaaag ctgaggtctg tgggcagga gaggaaggtt aacctggtg ctgccatctt 12600  
aatttgaac ttccaagcag atgtggcttt cagtcctcct ggatggcatc cccaggcaga 12660  
ggcagagagt cctgtgtcca tccgtccgtc cgtcccccc aacgcaaac actacagaaa 12720  
agtatcctt ctggctctgg cctaccttc taggtcctgt ggtgttaacc agctgggatg 12780  
gtgtggcccc ccgtccaga acgatccctg cctctcctg aaagcagctt tctgtgaggt 12840

cattgtgtc cagcaactg cggaccattc ctccagcaga gattccctc cagcttccat 12900  
gcaggcctga gctaactgag cccagcaac aggatcaaac ccattccaag aggaaggcca 12960  
tctgttctc agcctccagc tgctggccct tcatttgcaa ttggctggga agctttggag 13020  
gggtcaggtc ctggggacac atctgcagtc tctgaatggt gttataactg gggtcctgct 13080  
gagcagagaa agccaagcc ctttaataa acttgctgaa caatacccc ccaaagggtg 13140  
agagtcagag aagcaggagg cagctttgcc ctttagctaa ctctaacct tggttttgta 13200

gccagggcac ttgaaaacta tttctttatt cagaaagtac ttaccaagcg gagaaggag 13260

gggctgctct gaacaaggaa aatgtcatat aggatttggg catcgatctg ccccttaagg 13320

gaattagagg gcaataatct ccacttgagt gtatgccatt tattgaatat ttacctcagt 13380

gtcaaaggagg tgagcttggt ccagatgcag ctgttaaaga gccacaggca gcatgaagtc 13440

ctctcgaact tgcctctgga atgcagttca gccttgggaa caccagccaa tctccctagt 13500

tcattgcaag caggteccca agctgtagct gctttaggtc ctgtggttct tgggcctgtc 13560

tgtagtttgg tgttagggcc cttatttctt gcatccgggg gcctatgata acttagccta 13620

atgctctagg gactttctat agggaccagg ctgtaatcgg gcgtgtgact tcattgagtg 13680

gatgaatggc agtatgcag gtgtgttcca ccttggtttt attgacaggg tctctccctg 13740

actgggaact taccacatag gttaagctgg ctggtgagca ggctccctgg agatgtctct 13800

gtctcattca ctacatctag gtttttttgt ttgtttgttt ttgttttttt ggtttttttt 13860

gtttttgttt ttttgttttt tttttttttt ggtgggttct gaggaattaa agtgatgctt 13920

gcaaggcaag aactttatgg actgagctat ctgtcagcta tctgtcagcc cagccccag 13980

aggtacaata acctgtgggc cctttggctc actggtttct tgaagcaggt attagccctg 14040

gtctgtatga gacggagcct tcaggacctg cagatgttta gttccacttg agaactttgc 14100

aggaatcctc gctcagggaa ggctgtgata taagatgtga cagattttatt cacttgaaag 14160

aaagccctgg ttiggagtca gaggcatgca agtggatatt ctcatggggc catcttaacc 14220

ctctgtctgac tcatctactg acctgttaga atcaggctgt gaccataaa accaagcccc 14280

aggtggctcc cggttgggtg aatatgtctg cagagcttca ggtagagcat ttgcctact 14340

gtgcacagag tgtttctct cagtgtgtc ctcacatcag ggtcagtgag ggacttaaca 14400

gaaagccttg ggttccctct ttgtgccacc gtttgccagt agctggcctt tctggtgtct 14460

caggacaga gggggccgtt cagtacgacc acgttcattt tggacagcag caagccttaa 14520

gcttttggct ttggacaaag ggtttctgag ctggcgggtgc catcctcagc tgggagccca 14580

ggagcaccca gccagagcac tcaggccatt caggaggctg accttgggtg gaggtcctta 14640

tgcacgataa acctcggta ttgcgttcat tttcttctt cccaccttct cagaatgtct 14700

ccacagaca gttgggtgag aatgaatatg tctgcgtgtt ctacgtggat aaaacatagg 14760

ctgtgacatc atggggatgg ggtgacggca tgtgtcataa tgggaaactg gaaatcttat 14820

agaagagaca ttiaggtttt gaaaactgca caggagcctc tcaggtagag aaacagttaa 14880

ggtacaggga acaggacag gggacagagg acagacatac cgtctggcta ggcaagccac 14940

catgtgaatg aacgggggga agaggggaaa ctgggggaat gtggtactcg gtaatgatgt 15000  
 aaagatttcc tagagagaca ctcatatag gttgggtaca ttccattcag gcctttgcct 15060  
 ctttaggagc ccctatagca ttctttagtg ttgtagctac gaggagcagc aacctggccc 15120  
 caaaagagat tcaacagact ttcccagtgg cttttgtctg cctgtggatc cagccctaga 15180  
 tggcaaggtt tgggactagt gtgtcctaag gagtcctgca gaccttgggg agcctgtgct 15240  
 ttctcttgca agtgcgcctt caggacgcag gaggcctggg cctggctggc cagacctcgg 15300  
 atacagacgc ctctttgtgc ctctgagcca cgagtgtctgg gtactttgac ataacttgta 15360

atgccagttt ctacttctcg ggtgctatgg aatctaattg ctgagttctc tgggacatgc 15420  
 tctctcagaa caaaaggttc cattttccag ttcttgctca agcaaagcat caacagctag 15480  
 gggatttgtg tagctgcgca gatttgatct ctctctcgt cttgggtggc cagtgggaat 15540  
 ttcagtcttg ggagtgtatg aattgagtgc gtatgttgtg accaggcgcc tctgtcattt 15600  
 ggacactatc gtcgcatgac aggattgggg gggagagagg tgcgggtggg taaggagcta 15660  
 agctgccgcc gctttgagtc taggtaccgg gtgacacaat gattcttagg cccttttgcc 15720  
 tttctgcat ttttatttct tcttgggctc aggcataatt tgtttcaaac tggagggctg 15780

tccacctgt ttctcaaagc caaacctaaa ttacgagggg tgtgcctaaa tatgaaatat 15840  
 gtaatggttc ccatattgaa acatttgcta ctttctagtc ctctccgatg ggcggcttga 15900  
 gccagccag agtttctggg gctgtccgac tactgcagct gaggtagcta ttgggtgggg 15960  
 tgatgctaac aggaacgtgt ctgaagagat gctccagcta ttggttgtaa acaaagagcc 16020  
 tgggcagcct gctcacctct ctctctccc tagcctcacc atctgcctt ccccccccc 16080  
 ctttttttat gcagccgat ttcttgaggt tgaaaacttc catctttgtc ctgtatgggt 16140  
 gttggcccc tctctcttt caggatgagt tgtacagagg ccttataagg atgctatcag 16200

gatgtgcaag ttggcacact ggtaaagggg aaactttgaa agagtaggag ctgcagcagc 16260  
 cagctctggg atgtcgtctt tgtgtctggg gacaaggcta gctaggccgc tcttcttct 16320  
 gactccacca aaggaccca ttgtcttaa tatcttttat actgaactct ggtgccagct 16380  
 ccatgctgac agtgccatgc aaaaatatgt acaggagagg ctcttccaag gtcccagctt 16440  
 tgccaggtgt caccggtttc taaaagccta ggtggacatt ccagtlacat gtgccctgca 16500  
 ttctgggtgt ccttgatttg aagttacaaa gaacctttca agttctgtac cctgttctat 16560  
 ggccagtgc cacagctcac caggcccatg gattggcagg gcatctttat ggctcggagg 16620

gcagagtggg tcaacccttt gccactcacc tgttatgaac ccagtgctct gtgactttgc 16680  
 agtgacattt ggcagctcga tccccattct ccgtcaagac ttttggcagt cctgtggctt 16740  
 tgctgtttat ttgtcttgta ttagatggca ctgtctggga gaacgccggg ccatggtatt 16800

gtcctcgcc cagggttcct gtgcagtcct actgggctag aggagtgcctg ggaggtgggg 16860  
 acagcttagc tgggcagccc cgtcccttga caggacatgc ctgctgaagc tgtgccttct 16920  
 cctccacct cctcctcccc ttccctcctt gcctcctctc tctcttctc tctcctcgt 16980  
 cccctttctt ccttgtacgt ccctcttctg ggtgaatcta ctctgattct gctttgtcct 17040  
  
 ttccagaaga atgtgttttg ggatctgatt gtgccctgtg gggagcccc ctaagtgggg 17100  
 ctgtttgagg taccctactg tatctttaac tcagatcctt tagacgtga ctaaagaagt 17160  
 cattctgggg acacctaga agtggcttgg tgtgggtcga ggtgatttgt tggccagag 17220  
 gtggttgga gaagtggctc ctctccctg cgatgggtggg aagctgccat gtgatctgtg 17280  
 ggagacgatt ggccaggga ggacttggac gcccatctgt tctctgtttg cagttggcg 17340  
 ccatttcaga aaccacaggg gaaaagtta taggcaaaca tgataaaaag tgacagtctg 17400  
 aagtgtgct atcgtggct tggcaactta aagcattacc tgaagcagct tctaacttc 17460  
  
 agacgtcta gctgaacgg gaacccaag atggccatcc tgtgggcgt ggggaagatt 17520  
 tcgtttgtgc gcagtgggt gtcttagtct cgccccatc tacttctga aggtccctt 17580  
 tctagggtga ctacacgaat agcaagggtgt cataccctc cccctagct tacaggaagg 17640  
 taaatacaag ctgtcactag tgacatcagg tgagggtcca ccagagggt gtgacctact 17700  
 tggatctgta gaaggacttg gagaagggtc aggaagattc tgcctcagtt tccctttcgc 17760  
 ctgggtctga agccctctc atttctaat cctattacc tcccaggga tagtggcttg 17820  
 aggaatcttt gggaagaaag agggctcatg gcagggtaac agtcagccac gtgtgcggaa 17880  
  
 ttttaagac agaattcac tacatagccc aggttggtt tgactgcct cactcagtag 17940  
 cttagtaggc ggtaactct gaagccgatg caggcttga acttatgac ttcctaacc 18000  
 accatgtgcc accatacccc accactgttg atgtttcat tattggattt gatgctgtga 18060  
 aggaacccct ttatcttttg gtttgtttgt ttcttgagta tcagagtagt cagctcactg 18120  
 aaaatatgac cagtatatg gaaactgctg gcatgtctca agggtttgta acctgtgggt 18180  
 agaaacacag ctaagcctc acacaggaga gcctctggcc actgttgtgt ttgtcgcagg 18240  
 tagaaacagc tgagcagagc ctccagaa agtaaacatg tcgccttgtt tgttcagaga 18300  
  
 gtttaggtaa caatgacagt gtatggcca gctccatgc atcttccaa gtttccattt 18360  
 aattatgaaa aatgtatgag aacagacttt ctgtctgcgg aaacctga aagagcattt 18420  
 ggtgcctctg ctctagctt ctggaacttt ctccccactg tgctgtgcag agtcagagg 18480  
 gtggaacttg gaagcgtgtg ctccggtgag ccacggcatc agaaatgta aatccaggaa 18540  
 atgttgatat tgcataaaa gagactgttt ggatttccca gggagtccct tgtcctgtgt 18600  
 caattgtcac gtgttacaca gagcagcttg gcagagtcgg gcaaggagtg gcctgtgtgg 18660

agaggccatc ttagtgggag agacaggtgg ggtgtggcga gcacagtcct tggtagccttg 18720

gccccttata ggacactatg aggtgggttac aatatggagt tgtaacacca caggactctt 18780

aagagcaggc agtgattggg aggagccagt cccgaagcct ggtgaaggat ttaggcacag 18840

aagagaagcc tttagctctc aagtctccag ggctaggcgg gagcaggatg gcatcttttc 18900

agcatgccac ttgggttcca tgttcttagt gccctgggcc gtgatgtatc tcatgtgtga 18960

tccatttgca gggagctacc aactgcatct gtgtcctggg atgctgttgg gttggctttt 19020

tcttctcacc cccttattat aatcctgctc tctcctgttt cttccccctc tacggtattt 19080

gacctctcc tttctttctg ccttttcttt tctgtattc acccaatctc cctactccct 19140

aggatcacca aggaggaggt aacattgcct tctgctgacg ctgctgacct ctaagtgggg 19200

cctcttgaga gaaggtcact agggagtgtg gcattctgcc tatccaagc agataccttg 19260

gaggaggcct tggcgttagg atggcttgat ttcatagata cttatctttc tgacgtgctt 19320

gcagatgata ctctatactg tccccaaagc cagtcgtctt cctgggaaac tagagagttt 19380

cccattttgc ccatgccaac ctggcctcac cattgactga gtgagatggg agcccatcag 19440

tgaagtctt gagattaaaa atccagtgtt tctgaagac agtggagcac cacagtata 19500

gcttgagaac aacggcggat gactgacatt ggttgtggct ggaagatcaa gtatacagc 19560

ggtggctccc aggcacctcc cgtataatgc cttcttgtat gttggtggtt ggggatcttg 19620

tggctgagag gctatgcagg gcagagagga aatgagccca gtgtccctgt acccagggca 19680

gtgtcccttt accaaacatc cagtgtcctg tctacctga gaccctctt cttctgtgtt 19740

cctcacgca tggatgata gtatggtaga attggtccag catggtccag tagtgacgt 19800

aaatttcaat gagtcttggc ctttgttga tgttgggtgg aggaagggtt tctccgtgga 19860

tgggtgagac ttaaggctc catcattctt aacattgtac gaatcttgg tttaaagatg 19920

ttaagaccag actggcagat ggtatgagac ttaggttcaa atggaacccc ctttccct 19980

ccttatttct cttcctcatc cttaaaaata tgaacccttt gtttacttg ttgttgctgt 20040

tgttcattat tctcagtgtt agtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgatgt 20100

gtgagtgcac gtagacaaat cagccatgct gtgtaaaggc cagagtgcag tctggtagag 20160

tcaattcctt cttcctctc taccttttta agggtttctg ggaaccaaac ttggctttag 20220

gcaagcatgc cttcctctc ggagccatca ttccagaact gcctctccct cattactta 20280

gccactcagg tcagctgcct cttggtttaa atgggcaggg aaaggcctga gctgagacct 20340

ttccaactga attctcaatg tctttcaaac ttggttctgt gtagtgccac aggtgtctg 20400

ctacttcttg gaggagactc ctatccctc ctgcaacaga agctgaaaca ccttcggtga 20460  
 ggggccacgc tatcagtgtt tggggcttgt agaccatgag attttttttt ttttttcaat 20520  
 gactctggtc tgcccgata acacaaaagc agccctagac aatacatacc caagtatgta 20580  
 ttgagtatgg cactgctcca agaagtcctt gtttacaaaa gcaagtggct gacttgtccc 20640  
 tcaggccatg ctttctggc tcttctgcc cacggggcct tcgccaccg tttccacatg 20700  
 aacggctacc tacctgcctc acccttaaga ctcccttaca cacttctat tttctctgag 20760  
 gttttcttc actttcattt gcccactgc aatggagggt ccaccagggc agggatatgg 20820

ctaccctcct gttgcttctt gagtgtacag aacaaagctt ggctgtggt aggtatgcaa 20880  
 taaacagagg gcacatgaga taaacaagcc cttgaaacct tacctggctg tcagtgggt 20940  
 ttgctttctg cccctgctt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt 21000  
 gtgtggtgtt ggtggggttt gtgttccatc aacttctgtt ttcttcccta tgtgggtttt 21060  
 acttttgtgt tctgtactg ttaacatctg tgccctctt ggctgtgtgc atttgaagt 21120  
 ggggtccct gtgagaagcc tcaggccctt tgtgttggt gctgtgcgc ttcttgacc 21180  
 agatgtttat taaatagcag gactgaaaca tgaattgact gtattctagt cgtgagagaa 21240

tttgttctt ggagtgggt ctgggcagaa taatcgctt gtgatgtgc tgccagatc 21300  
 tggaacctgc ccagtgtgg gaaggaagca ttgtgtttt caggcttggg actctgggt 21360  
 ccattcacag ctctactgt gggatgaaag cttatttcat gagccctct ggccacctc 21420  
 ggccctgagc aaggtcagga gcttcttcc tctcacttt tttgggagaa gctgggaggt 21480  
 tggatcatag ttggtttcat tctgccctgt ctttagagga aggcaatgtc tgccttctc 21540  
 gtgtacagca aagatatcca gtgtagggt gggcgtgggc acaatgacct atcagaactg 21600  
 agctttctga tgtgaaggtt tctcttgaa gtcaggacac ccataggcaa tgtgtctatt 21660

tcagtgttt gaggtatagg gtaggcagat ggactttaga gtgggagaga cccctttagt 21720  
 ttccagccag gtgactgatg cagagtgatg gatcatggag ggccatggtt gacctgggca 21780  
 tcagaggagg aactgggcta aacgggagt agagggagga ccttgtgtt ataaagaaga 21840  
 gcaggatgct tgacggagat cagggactct ggggtagtgg tgggttggtg ggcaggatgg 21900  
 atctggctcc accagtggaa tgctgggtag tagtcatgc tacttatcca gtacatgtag 21960  
 tctatgtgta tacatggctg gtttatggta tagggccatt aagtccagt aattccttac 22020  
 ttttcttct ttggacgtta aaggacccc agcatctgtc attttgagga agatggaatg 22080

tcccagctcg ccagaacag atctagctca gtctgatcg ggcccaaga gcacataaaa 22140  
 acaatcaagc caatagctgc ctcttccaa gtggtgaaga gtaattttgt agatgggtct 22200  
 gtttgcctc tgaatttgag acattttatt tatattgaaa agcttggttc tgtgagaaca 22260



ggcaaagtga aatatgaata agtagctaag tcagtgtgag aacgtgtatg tacgtgtgca 22320  
tgtatcacat atacagtcac gctggatggc tagcttggaa atcaacttta cagttttctt 22380  
gtggatTTTT cttccacttt agggtttggg ttgttttttaa agccctatTT ccagtatgtg 22440  
gaaatgagcc aaccaggac agcttccgct ggatcgtgga cagcttctat ggccgtcgac 22500

gtgtacactc gagataactt cgtataatgt atgctatacg aagttatatg catggcctcc 22560  
gcgccgggtt ttggcgctc ccgcgggcgc cccctcctc acggcgagcg ctgccacgtc 22620  
agacgaaggg cgacgcgagc gtcctgatcc ttccgcccgg acgctcagga cagcgccccg 22680  
ctgctcataa gactcggcct tagaacccca gtatcagcag aaggacatTT taggacggga 22740  
cttgggtgac tctagggcac tggttttctt tccagagagc ggaacaggcg aggaaaagta 22800  
gtcccttctc ggcgattctg cggagggatc tccgtggggc ggtgaacgcc gatgattata 22860  
taaggacgcg ccgggtgtgg cacagctagt tccgtcgag ccgggatttg ggtcgcggtt 22920

cttgtttgtg gatcgctgtg atcgtcactt ggtgagtagc gggctgctgg gctggccggg 22980  
gctttcgtgg ccgccgggcc gctcgggtggg acggaagcgt gtggagagac cgccaagggc 23040  
tgtagtctgg gtccgcgagc aaggttgccc tgaactgggg gttgggggga gcgcagcaaa 23100  
atggcggctg ttcccgagtc ttgaatggaa gacgcttgtg aggcgggctg tgaggctcgtt 23160  
gaaacaaggt ggggggcatg gtggcgggca agaaccaag gtcttgagc cttcgctaatt 23220  
gcgggaaagc tcttattcgg gtgagatggg ctggggcacc atctggggac cctgacgtga 23280  
agtttgcac tgactggaga actcggtttg tcgtctgttg cggggcgcg agttatggcg 23340

gtgccgttgg gcagtgcacc cgtacctttg ggagcgcgcg ccctcgctgt gtcgtgacgt 23400  
caccgttct gttagcttat aatgcagggt gggggcacct gccggtaggt gtgcgtagg 23460  
cttttctcc tcgcaggacg cagggttcgg gcctagggtg ggctctctg aatcgacagg 23520  
cgccggacct ctggtgaggg gagggataag tgaggcgtca gtttcttttg tcggttttat 23580  
gtacctatct tcttaagtag ctgaagctcc ggttttgaac tatgcgctcg gggttggcga 23640  
gtgtgttttg tgaagttttt taggcacctt ttgaaatgta atcatttggg tcaatatgta 23700  
attttcagtg ttagactagt aaattgtccg ctaaattctg gccgtttttg gcttttttgt 23760

tagacgtgtt gacaattaat catcggcata gtatatcggc atagtataat acgacaaggt 23820  
gaggaactaa accatgaaaa agcctgaact caccgcgacg tctgtcgaga agtttctgat 23880  
cgaaaagttc gacagcgtgt ccgacctgat gcagctctcg gagggcgaag aatctcgtgc 23940  
tttcagcttc gatgtaggag ggcgtggata tgcctgcgg gtaaatagct gcgccgatgg 24000  
tttctacaaa gatcgttatg ttatcgga ctttgcatcg gccgcgtcc cgattccgga 24060  
agtgcctgac attgggggaat tcagcgagag cctgacctat tgcactctcc gccgtgcaca 24120

gggtgtcacg ttgcaagacc tgcctgaaac cgaactgccc gctgttctgc agccggtcgc 24180

ggaggccatg gatgcgattg ctgcggccga tcttagccag acgagcgggt tcggccatt 24240

cggaccgcaa ggaatcggtc aatacactac atggcgtgat ttcatatgcg cgattgctga 24300

tccccatgtg tatcactggc aaactgtgat ggacgacacc gtcagtgcgt ccgtcgcga 24360

ggctctcgat gagctgatgc tttgggccga ggactgcccc gaagtccggc acctcgtga 24420

cgcggatttc ggctccaaca atgtcctgac ggacaatggc cgcataacag cggtcattga 24480

ctggagcgag gcgatgttcg gggattccca atacgaggtc gccaacatct tcttctggag 24540

gccgtggttg gcttgtatgg agcagcagac gcgctacttc gagcggaggc atccggagct 24600

tgcaggatcg ccgcggctcc gggcgtatat gctccgcatt ggtcttgacc aactctatca 24660

gagcttggtt gacggcaatt tcgatgatgc agcttgggcg cagggtcgat gcgacgcaat 24720

cgccgatcc ggagccggga ctgtcgggcg tacacaaatc gcccgagaa gcgcggccgt 24780

ctggaccgat ggctgtgtag aagtactcgc cgatagtga aaccgacgcc ccagcactcg 24840

tccgagggca aaggaatagg gggatccgt gtaagtctgc agaaattgat gatctattaa 24900

acaataaaga tgtccactaa aatggaagtt tttcctgtca tactttgtta agaagggtga 24960

gaacagagta cctacatctt gaatggaagg attggagcta cgggggtggg ggtgggtgg 25020

gattagataa atgctgtctc ttactgaag gctctttact attgctttat gataatgttt 25080

catagttgga tatcataatt taaacaagca aaaccaaatt aagggccagc tcattcctcc 25140

cactcatgat ctatagatct atagatctct cgtgggatca ttgttttct cttgattccc 25200

actttgtgtt tctaagtact gtggtttcca aatgtgtcag tttcatagcc tgaagaacga 25260

gatcagcagc ctctgttcca catacattc attctcagta ttgttttgcc aagtttctaat 25320

tccatcagac ctgcacctgc agcccctagc ccgggataac ttcgtataat gtatgctata 25380

cgaagtatg ctagtaacta taacggctct aaggtagcga gctagccac cttgccttga 25440

gaatggtcgt cgccttttgg ttcctttggt tgtgctatga tgcgtcagtc tgggtgtgcta 25500

actctatggc ctgcttatct gtctctctc ctgtgatctg caatctagcg cctggaagag 25560

aaaaggagat tacagcttcc ccagactacc tggagatagc tatttactgc ataggggtct 25620

tcttaatcgc ctgcatggtg gtgacagtca tcttttgccg aatgaagacc acgaccaaga 25680

agccagactt cagcagccag ccagctgtgc acaagctgac caagcgcac cccctgcgga 25740

gacaggtaac agaaagtaga taaagagttt gaagaaattt actccctcc cccaccagc 25800

cagctcttgg atcttcttcc ctctgatttt cccctaact tctgtgagct ccagaactgc 25860

aggcaattct aatctgccac tgtgtggagg ttcagtcagc ggttgggact aaagagcatt 25920  
aagtcacaat gctgctctgg gcttggtagg ctggctctgg ttttaaagga caagagtgtg 25980  
aagactggag ctgcccagtg ggatgggcag aggaggccat gccctctgcg cccctcaagc 26040  
tcacggctcc ttgggagaa caagcatttg gtcctgctcc attgcttctg tatgaggcca 26100  
gatgttcggg ttcaagtttt acccttcata ggaaagagag tttaatcttc ttgtatttac 26160  
tattttaagt agagatcaga aacagaggat ggaggtatac ctgaactaat gcttgcataa 26220  
aagtggctctg tgatgtcttc taaactgggt ttggctgat ttgtctggt ttttaaacg 26280

ctgtatcgct atagtttatt gttacaggtt tggctaggga ttcagtata ggatgattgt 26340  
gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt 26400  
atttaggtta taagtacatg tgtgcagggt cctgtggaga ccaaagaga gtgtcagggtc 26460  
cactagtct agagtttatt agcatagggt gtaggaattg aactctggtc ttctagaagt 26520  
gcagcaagca ctctttaacc actgagccag ttctccagcc cccagatacg atgattgcta 26580  
tgtagaacag ggagaaaatt acccttaacc ttgagcttga tctttgatgg ctggctttgg 26640  
gggaggttaag gcaatagaac ctccctctg ccataaaaca aagcccttca aaggtggata 26700

aggaaaaat gcttgacttc tgtacttgct cctggattcc aagagccagg catgtgtggg 26760  
tgtaaatctt tatgataaga ttccgaactt gattctgata agattgtcac tatttttttt 26820  
aaattagcaa tggaaatgaa caacctggcc tgtgctatgg ggaggtgcat cttagtgttt 26880  
gttaaaactg catattcatt agtttcaacc ctgaaattc ttatttagta cttcttgaat 26940  
ggatctgtaa gagtctgcat tttaaacact ttctcgggtg atactgtgta ataccttaag 27000  
aatctctggg ttcaacccaa ccctgccttt cctgggccct ttctgtggac aaggtgggaa 27060  
ctagcagggtc agtagtggtt tggacacagg gccttggtct ttctcaacct agcttcacac 27120

tacaggctga gcaggtcctt gtcaacgtcc ttgaagcctc gtttccaaca ggtgtttctt 27180  
gcagagggtt aagataaat ttggagcata tgcagatgc agcctttggc cagctgttga 27240  
atgtggagtc aaaaaggctc agttgggttc cttttaatcc tgagaatgct gtgctacttt 27300  
gagtgcacc actgtcattt gtgggacat agaagctaga tgggtgctgaa gttaaagtgt 27360  
gttgctgaa tgagtgtctg gaagagtcct taataaaact cttacctggc tagatagtgt 27420  
taaggcttca ggctgaatgg ccactccttt ggccactcct ttggctactc cttcacagcc 27480  
tctctgac tcttagccct gggccattct taacatctgg actctggtct agggagttaa 27540

agtaaaggga gcaatgtcct gtctcatttg tttttataat agagaaaaga agtaaaatcc 27600  
ataagtttag gtagataggc cattgacct taattatttc atcatttaaa aaactgatgt 27660  
gtgtgtgtgt gtgtttatc atgtgtgctg gagctcttgg aaagctggaa gagagcactg 27720

gatccctcag agctggaggg ccggtagtgtg tgagctctct gatgtggtgc tgggactaga 27780  
actcgggtcc tctgcaagag cagcaagcgc tcctaatac tgagccatct atctctccag 27840  
cctgcatcac ttttaaagaa aactctttct atttctccat tttccatttc catttccatc 27900  
tttttacatt tatatattac atttatatat ttatgtagct tgggcacgtg tgcgtttgtg 27960

tgggggcatg cacatggcat agcaagggtt gtgaagggtc ttttgaatg gtggccaggg 28020  
gacaacttac atgggagcca gcacctttct ctaccacgtg attcccagga atcaaacagg 28080  
tcagggtcag tcagggtcac gtatctgatg accgattttg ttgactccat cgctttttaa 28140  
gaaaaaaaga attaacacct attacagcgc tcttctttt gcttcattgt aaaaagacag 28200  
aggccctgga gtccccaggc acatggattc agcatgtctt ctttctgtt tgtccaactg 28260  
agtttcttca ttttctgtcc acctaagctg tccattttgt ttgtttttat attccctgtg 28320  
tgaccggagg gaaaagtgtt tttttttttt tttcatttac ctccctttct tcttgatttc 28380

attgttattt actgagtcca cagtttcttt tagtgcatgg gcctaaatca ggactcttgg 28440  
gctgggagtg tggtcagtg ctggcatgct cgcctagctt gtccaagcct cagtatcacc 28500  
aaagaaaata attaacccag tttctgttca gagaaagcct gccattttgc cactggctgc 28560  
aaggttagt aaggctcttt tgaaatgttt ctcatgctg acgctggata acaaatgtgt 28620  
gaggcccagg ctctctgcat gaggaagcct ctgggagata aatgggttga aaaggtactg 28680  
ataatacccc agcatttctt agaagtcag gggaagtatg gtactaactg cctcttccca 28740  
aaggatttcc caaagcttag gccactggga ggaggaggag gaggaaaagg aggagaagga 28800

ggaggggaga tgcttatcat gagtctggat aaagagggtt ttgggctgta gctggaggcc 28860  
tgcagatagg ttaatgacag agtgaattcc tcagggatgc caagcatgcc ttacctggcg 28920  
acagatgagc ctgtaatcag atgtctggag gacgggtgct cccaggcact aagaggctag 28980  
gctttatttt gtgtaggccc aagcttctat atgatgcagc atccatgccc tggcccttgc 29040  
ccaggacggc gaggaggcgc atagcctctc tccattcact ccatctttgt ctttgtttta 29100  
gaacgagaaa agttggtttg tttattcatg ctgttttttt ccatgtgcac aagcgctg 29160  
tcggaaagtg tgtaggtgtg tacagaaagt gtgctgagge caaatgataa ccttgggtgt 29220

cattcctcag gtgccgtcca ctcttactc tttgtgtgtg tatgtgtgtg tgtgtgtgtg 29280  
tgtgtgtgtg ctaagtttct cacctgcttg cctgaactag ccaagtaage taggttatct 29340  
ggctctgtgag cccagggtc ccaattgtc ctcttctcc tctctctgt taggattcca 29400  
agtgtcggct cccaagcctg actcttcttt tttttctga gacagggttc ctctgtgtag 29460  
ccctggctgt cctggaactc actctgtaga ccaggctggc cttgaactca gaaatccgcc 29520  
tgcctctgcc tccaagtgtg gggattaaag gcatggacca ccaccgccg gctttttttt 29580

tttttttttt tttttttttt taatatagct cctagggatt gctctcaggt caaggaaggc 29640  
  
 aggcgttttg atccttcttt ctctctgagg ttgcttccc tgccctgaac ttgtttaaaa 29700  
 caggcatttc actttaaaaa ggtagggctct ttttttttgt gggtaggagg ggtggggggt 29760  
 gttttctgta tcaaataaat tctttatagt ctttctagta aacattaatt ttgggagaca 29820  
 ttgtgcttgg agtaagatac gcaacttttt ggtgggacag cctggtaggt agcctgtggg 29880  
 atctctaagg aggagtcac tctctcacc aaggctagga ctgggcactt tgtaagcgt 29940  
 tgcgcacttg cctctacttc ttggtacctt gtgttaaatg gcaatagtca gtctagagaa 30000  
 gggcaccttg tgaccaact ggaccatcag tggctactgg gcagtggctt ttgtgtactt 30060  
  
 ctgagtccaa gtggaaagat ttgccttctg tgatttccac aagtccttg ttggggaggt 30120  
 gggccgtatg tgagtgcaga gcgggggtga ggaagccttg ttctgtggag tgcttgtttg 30180  
 tggaggagct ttctgggta ggttcagctt ctttctggag ccagaagttt gcttagggca 30240  
 agatggagat ccatctgtct gtgtccagat gagtgcatac cctaccgat ccccgagctt 30300  
 cacacaggac ttagtgagtg ttgttcccag cctcagccat tgacatgggt agctgagaaa 30360  
 accagagagc aatttcaata tgcgtttgag acctatggtt attcagggtg ggctgggggg 30420  
 aacacttaat tccagagctg ttctcagggc aatgtattcg tggctctaga gtatatgaaa 30480  
  
 ctgagtgaat gtgagtgtg actgcttagc atcccagcac cgtgacctgg aatctccatc 30540  
 gtacgaggtg tagtcgatcc agagttgcag tglaccggtt ctgggaaaca ttggggcagc 30600  
 tggatagttg tggatgacct gagggtgagc ttgcttttcc tagggatcga tcgctttcct 30660  
 tgtccccagt ttggcctgtc ttttctctca gccccgaaa gacatgctgc cttggctgag 30720  
 atccacccta gacttttct gatgagctat aagtaggttc agaacacctg agtcaggtag 30780  
 ttttactgtt gtgacagggc attcaagagt ccagaggagg tagaagctgt ctaaggggca 30840  
 gtgtgagcaa ttacctagat ttgttatgg aaggaaaaac aaaacaaaac aaaacaaaac 30900  
  
 caaaaaactc cactcccaga aactctctga agcttggtgt ggtgcaggtt tttctgttgt 30960  
 ccatagaggt gtgtggggct agacttaaga tagaacacac tggccctctg ttctgatgtg 31020  
 gaaggctcca tctgctgctt gggagtcgga ggtgtgtctc agtctgctgt agtccaaggg 31080  
 catgtgtcaa ttctcaggaa taaagacaaa cttgactcac cttccccgt actgtctttg 31140  
 cttccgctg cgtgtgtgtc tgtgaggtcc cctctgaatg ttcagcttca tccagcataa 31200  
 agggagacgg ctatgacttg gtggctcttt aaaaagaaaa ggggagaaaa cccacttctt 31260  
 ccgttaactt cccatatgta ccgtggaaat atatgaaaag cacatttagt taaaagcttg 31320

atttatggca cgttgtaaag agatcccgcc atgtaaggct gccgaattgg agactgtgaa 31380  
gagtgtgcgg ctttctaaaa accgcctgcc aagatttggg gtggggaatg ggggtggggc 31440  
ggagcaacag ttactacag tgttagcgtt tatgtttat aagtgaactt ctaacagtgg 31500  
gatgttttta agtgcgttga aagggaact ccaaaatgga agtttctaga ttaaggattg 31560  
agaactatct gaggaggga gttataagta caagagaaag agaagaaagg aagtctgtaa 31620  
tacagtgggtg ttaggaacct tccaagggtg gcggtggggc cacaattcag aggaaggag 31680  
cccctgaaaa gccaggctcc tccagggacc cctgctgggg attttgccaa gccctccaga 31740

caggttgccct tctgaggag aggcgagtga agagaaagcc agtcatgctt tatagcccca 31800  
gagaggattt taaaagtata gtaaacgca tggaggtaga attaagatgg acctctgtag 31860  
acaggagag cagagtgtat gctcagagac ttgctgtatt tctttaccct tccccactct 31920  
gggtgttttt tacaaggta ttttcaggc ttgtacattg aacctgaatc tgcactgtgt 31980  
attgaacaaa attccacac atgaaggcag ttttacattt tgataccaat gtgcagcaac 32040  
gactgccaag gtgttttttt ttttcttcc gtattagttt agtttttttt ttttttttt 32100  
tctccccgtt ttccatttg aaaatgttgc ccttaaaacc ttgtggaggt gctctgttgt 32160

ggggtgggta tgcgtatgg aaacttgac cccaggcctg tgctgtgcat tctgtttggg 32220  
tcaaaggtec tccacagagt agttgatgtc agactggatg gtaaatctct ctgttttgag 32280  
gtaacccta agtcatggtc accagcggga cttgctgct ctatggtttt cttcttctcc 32340  
tctaattcct acattaaaa tataatatgt cttgcttact ggaactccag gctatcctgg 32400  
ctggcagttt agggteccat ttgttaaact agactcgcaa ttcagggtga tgccatctaa 32460  
aatcagaaca aactcacctt gtagagcaga ctggtagct atggctgtcc cagctcagca 32520  
ataagcactt gatgctgtct tcatctgtc ctgctaactc tgagaccacc tgagactcac 32580

atagaccccc ggaatctgac cttgacttca cggtaaccatt gaccaggatg tagcctgcca 32640  
gggcatcttg gccctgggtg atcaccaggi cacacattga aggatgcgga aacatcaca 32700  
aacagcctgg ggtggggggg acaaaaaaga agtgccatcg ggcgtcttgc tagtttctaa 32760  
actgaagtct gcataattca accctgtgcc ttttttctc gctgttcata tttattttat 32820  
tccaatgct attttgcta aagaagaat gtctactaaa acacaaagga aacacaagac 32880  
cagggttaata aaatctatat gatgtagaaa gttctagaat aagacctgtt tcctaccttg 32940  
ctccctatc ttgatctctc actctctctc cgaagggtgac cactgctaaa tccttagata 33000

tctttccaga aaacatttcc tgctttgctt cccaagtctt gatctctctc cccaaagggg 33060  
accatggcta aacccttaga tatcttcca gaaaatgcct gtggtcacaa cccatcctgt 33120  
aagcctctat gtgctgagta ctgactccca aggacaggcc acagaagctg cgatgtgcca 33180



ctagcctctg gccattacca tcattcagaa ctgtggctct ctgagatttc tcagcatccc 33240  
ctcctcactg gtcttagcac acagtgggtc ctaacaacta agctaggaac tttagggctc 33300  
agtgatgcag aggcaagctg atgatggccc tataaagagt atcctggcta cacacagtct 33360  
ctgttggctc ttigtccctt ggggtctgtg ttgtctcatt actgggcaga cttttacttg 33420

tttggctgta gcttcttgcc tctgattatc tgggtggaat ttttactata tttctactgg 33480  
gagatgattt ttgcctatit gtgtggaaag actgccagaa agatcttaaa aattaaaaaa 33540  
aattacatgc cttttgcaag cataacttgi gacccctgatt cagaatgagt cagggtgggtg 33600  
gttccacaga agcactatgg accagctcca ttccagaatc ttctgagtcc ctgtctgta 33660  
gatggagctc acgatgtttt tgtggccagt ggaaaatgga catcttgatg ttgtcaggaa 33720  
acttctgggt tctgatgcag cctgtcacc acagttaggc tggacaccat gcggacagtg 33780  
gaaggggctt gggagttatc tttgtctctg ctgggatgga atgcctatc tggacaagg 33840

caagtgggtt ctagaggcac tcgctgttc cctgtcacc tttccctgct tgcttctgctg 33900  
tttgccttag agattgggat ccttgaatgt atggctctct attacagaat taccaggttc 33960  
cttctcttct tttctctttt tttttaatta aaaaaaagc atcaattttt gttgtggcac 34020  
aaggagtaaa tgcctgtct gcatagtata atgtatatac agcttcttct tgggtacggg 34080  
tgagatggct caatggacaa aggcacttac gctgatgacc aagcctgacg ctctgtagtc 34140  
aatcttcaga gctatgtgg taggaaaaga gaactgacce tcagaagttg tcctctgacc 34200  
tccacactga tatgcacaca aacacatgca cacagatata ttttttttca tttaaaaaga 34260

aaatcacctt tctccttccc aaaagatact tagaaggttc agaaaagtcc ttatgtgtat 34320  
tttaataat aagatttcat atcaaaattt gcttactgat tttacattt ctttgtgggg 34380  
tttttttct tttgaggggg ggagatagg gtctctggga ttgagctcag tgggtagcca 34440  
aggataacca tcaactgact taatactgca aacacttttc ttcaattcta ttaagggtag 34500  
ttgggtttcc aaagagcaga agggcttgcc aatgggacag tcagtcctgg gaacaacata 34560  
ggaccttggg ttctctgat gagagtctag gatcccatg ggagagttcc tttggcttta 34620  
tctttgccag ctggattgag gagtgtgtat actcagcagg ggattgtcac ccatgtggga 34680

gctggaagcc tgggtgtgct gctgagtggc tctgtctaa cctcacacc atgtctccgg 34740  
gacaaaagcc tccgttggg tctgagttga aagcagtac cagcagccca ccatcacacc 34800  
aagatttgt agtcatacc aggcacaggc tttgtgtggg ctctgggtat attttcttc 34860  
gcagaaatca gccaaaggaga gacggtgtgt ttccagagata gacactgggt ctgacacagt 34920  
ctgtatata tcaaggcaaa cttggtgaag cctgtgtgc tgcgtgggtga gagaggacc 34980  
ttcccgtgtg gctctgagt aaagtatctt ttccctaacc cttggtctcc tgtattcact 35040

gctctgcttt ctgaagctaa agtgacaaga gtcagcccat tttcactata tggctctgggc 35100

atcatcaagt ttcagaagga ctggggagag atggagaata gcctccccgt gcctggaact 35160

ctggatttct tgaataaaaag acctttgagt taccagaatg ccttttcct gtgtcttagt 35220

taggatttta ttgctgcaaa gagacaacac aatgtaactt aaaaaatta tttatttgtt 35280

ttatgtatat gagtgcacca tgactctctt cagagacact agaagagggc atcagatccc 35340

attacagatg gttgtgagcc accatgtggt tgcctgggaat tgaactgagg acttctcaaa 35400

gagcagttgg tgctcttaac tactgagtca tctctccagc cccagtgca actcttataa 35460

agaaacacac ttaattgggg cttgcttaca gtttcagagg tttagttcat tattgtcatg 35520

gtggaaagca tggcagcttc ctggcagaca cagtgcctgga gagagaagaa gctgagagtt 35580

ctacatcttg atccacaggc agcagaagg gattgtgtgc catactcttt gaggtttgag 35640

caaaggaaac ctcaaagccc gccccacag tgagaaactc cctccaacaa ggccacatgt 35700

tctctagcaa ggccacacct cctaatacg cctatgggcc aggtattcaa accaccacac 35760

catacatatc ttacagctct ttccttgaga tctttcttta tactttggag gcaatggcag 35820

cacggatgac ctcaattgtt agatgtttgt gaatccctcc ctgctgactt gatatttgat 35880

gtgtttttat tttatgggtc tggacattgt acatgagaca agcatcctgt aattgagccc 35940

agcctttgag ttagtgatct atagctgag caaaaaacta taatgaagtc agtagagtct 36000

gtctgcacat tcttaagtgg ctgtcttaaa acaattaagg taaggggctg gagagatgct 36060

tcctcggtca agagcactgg ctgctcttcc agaggacctg ggttcagttc ccagcaccca 36120

tatggcagct cacaactgtc tatacctcca gttccagtct gacatcctca catagacata 36180

catgcaggca aaacaccaat gtacattaaa aaaaacacct aatttttaaa aagttcagat 36240

gaaaagaaga aatactatga ttaaacttct agaaacattt ctatttgtaa acttgacctc 36300

ccaaggtcaa ggatcctgtg acttctcatt tttgccctg tattttgttg ttgtttgtgt 36360

ttttgtttgt ttgtttgtg ttttgtttgt tgtttagttt agtttctcgt tgtttgtttt 36420

gtcctttcct ggttctctcc cctttctttg taagcactcc tgctctggct gggteccagc 36480

tcacttcag cctcctctga tggagccagc attacatctg ctgttttgca tttgtatac 36540

aggtttcggc cgagtcagc tcctccatga actccaacac cccgtggtg aggataacaa 36600

cgcgtctgtc ctcaacagcg gacacccga tgctagcagg ggtctccgag tatgagtgc 36660

cagaggatcc aaagtgggaa ttccccagag ataagtaagt actctcctc tgggagggtc 36720

gttgtctgca cctcctggga ctgagcgag gtcttggtt tgggagtctc cacctgtgtc 36780

ttgtaatca gggacctgtg tcttggtaat caggacatt cgaactgtaa actgtaaact 36840  
 glaaactgca gcaagatggt gcaattaaca gagctgctgg tgcacagggt aggctaccag 36900  
 cctgtgccct tgaggtggaa gaccaacctt agctctggga agtgaggatc ctggaaggct 36960  
 ggcagcttcc ttctttagg attagcgtct aaacagcttg agagtaacag aagggtgaaa 37020  
 aatgggctct ttctgcatca aagacacagg aatagctcc cagcttgctt gaagacaact 37080  
 cgtctgccta tcttgacatt ttttcagtgt ctctctaaga ttgttagtga tatgtttaac 37140  
 acacacagcg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tatacagaga 37200

ttcagaaaca caaagacaag ctttccaag tctgggtatg gccagctcta gttgaaagt 37260  
 gaaggaagct taggccccct tggagcttcg tgcacacta ttccaggtat ttgggtccc 37320  
 aacttagaga tctttacata tctcctaact caggcccgag gaactgtctg ttagtcatc 37380  
 ttatgtcagt ggaagaatt tccagtttct ttatatactg cagtgaagag agcattcatt 37440  
 cattccttca ttcatcccg gaatgttttag tggacatttt tgggtgagg tatgaaagaa 37500  
 aacacaaaga atggcttctt tacctataga attttagaa aggaaggtat gtttctcttg 37560  
 acccttgcca gcctctagc tgggtgtgat tatagaaata gctgggttcg tgtgcacatt 37620

cctatgtcta caaggctgga ctggggagt gtgctcatat gctaaaaact tgcagctttc 37680  
 gggttgagcc tgtgctttgg gccacctgtg acaccacagt caacagtgtg agtctgtgtt 37740  
 gccagcagc tccaccggg gggtagccaa gtgtggaaa tctgagcgt atgcatttcc 37800  
 aagcagtgtt tagcacaat gtaggtggag cacattcca atgaatgaag gctatttaga 37860  
 aatggtttat taggttgag gcaggagtta cacaagaga ggtttgttat ggtttagta 37920  
 acaggggcaa taggacataa attagccatt ttccaacga aatgtttatt ttctgccaag 37980  
 atgttcaatt taattttagt tcttgactgg aaagggggcg ccttcagcag agcaagtgt 38040

ggatccatt ttcttttct tttttgagct tttaaaagcg ttgcgtgctt actgcaaatg 38100  
 ctgtttactg aggacagctc aatacttgc tgtactgggc ttatggtttt ttggttttg 38160  
 ttttttttaa aacaaaacaa aacaaaacaa aacaaaacac cccgaaaagc atactcaggc 38220  
 tggagaggta tctctctggt ctaaagcaga cactgccctg taaaggacca gagttcggtt 38280  
 cccagcacc atgatgtgt gctcacaacc tctgtaacta cagatctggg gaatctgtct 38340  
 gggctcctta ggtgctgtg ctcacatgcc catacctct cccagaaac acacatacac 38400  
 ataataaga ataatgtga aaaaaataaa aataaaaaac acactcctaa gtattaattc 38460

caaagacttc cctgttctt ttggttcttg aacatctaaa ataatgtcag gtcatttgtc 38520  
 tgttgtgtat aaaacttaca tgcttagaaa tgtaacttgt gctgttttct atttttttt 38580  
 ttcttgttt actttgggta gtgataagga atcctaaact tatgtcaaaa aggtatcgtg 38640

cctgatttct agaagttttt cttaatgaga cagcataaat tatttgaaac gtgctgaaga 38700  
 ttcctaccct gcaactgggc aatcgatgta accataaaat ctaccggtat tgaataatag 38760  
 tgatttgaga gttgccactt tacagggaca gaaaataaga acagactttc actttttttt 38820  
 tcacctctgc gacattttta attataaata tttaatggc tcaagaccaa aagctcccta 38880

tggtctggcat gcaggagacc tgaccaccgc gctagcaagg acaccttcca taaagaaaaa 38940  
 gaaaataaat cgagaggaca aatgtgaaat ttaatagtcc ctccaacagt aattgacgtt 39000  
 ctggaaaaac atcactaaga aaatagcctg cgtgtgtatc ggaggctcat tggttccata 39060  
 tgcattgcctc tggaagattt ttatatattag ttctggaatt tccctccctg tgccctcgg 39120  
 ccagactcgc ggtgtgctaa tcccgatttt acacatttag gctgacgctg ggcaaacccc 39180  
 tgggggaagg ttgcttcggg caagtgtca tggtgaagc agtgggaatc gataaagaca 39240  
 aaccaagga ggcggtcacc gtggcagtga agatgttgaa aggtgagtgg gcggatgggc 39300

ggtcggggag gagagggtct tatcaggagc gagcgttctt tttgtgacat gtgaactctg 39360  
 cagggacgtg gggtcagaga gcacatactt gacctggcgg ttgagggggt tttcaggata 39420  
 aatgagcaaa tgagatggag gatttacctt gagctgtgtg tacttaaaaa gaaaagccag 39480  
 ttttagcaga agttgtagct tgctgggctg aaccggcttc taactcctta gaaaagggtc 39540  
 ccgattctct tcttttctgt gtgttcattg gtttagaaag tttagggggt ttatttagct 39600  
 ggttaaatth tggaccaga cttttaacat acaataaagg agaggtaggt gttggagtgg 39660  
 caactggaga cagaatgtca aatgtggat tcaaagagtc gcttagaagc caaaaaggag 39720

caaacaattg gaactgatgc agaatcccag ggacatgtaa acaataatgc cacgtataa 39780  
 atgcccgctt tgttcttttc ttttcttttc ttttcttttc tttttttttt ttttagggga 39840  
 ggggagggga gggggtctgg gaatttatcc acaaccttcc taacacagct tgatgatgac 39900  
 gcccaaggag cttaaatgac tttcaactat taacttatcc ttgcatgggt attcttttat 39960  
 cgaagagata aagggaagg tcacattata aatcctgttg ttggggaatc tcagaaagga 40020  
 gaaaggagcc atgttcaatg tttccctggc ttgtgggcag agaagtctgt cccgggcctg 40080  
 tgggatgtgg catgtttca ggagtcgac cttttctctc tttgatagga cacttaccac 40140

atccctccct gatgcagaca acaaagggcc aggacatggt tcattttgtc agttttagtt 40200  
 attgacctga gactccaggt gaaatctggg atgttctttt ctttgagac tgataccagg 40260  
 aaggagatag caagtatcgg ggcaccaggg cagaggcagc ccttggtacc tactggaagc 40320  
 tgtgggttgg gaaggatcag gcatacact gctttccaca gaacctctgg ttttagatc 40380  
 cctggagcta gtgcaaaagg gaggtttagg ggttggccct tccctttaag caagatcacc 40440  
 caccatcctt ttcatctgg tcagaggaca tgccttttca acattctttg tgacagccag 40500

aggatggctg aggtgtaagg aagacaagtg tactgagcca tgtgtctgtc catagtcctc 40560

tcttccctct tctctgtatt ggtcaggata gatTTTTTga tacctgtgcc tctatttcat 40620

ttttaaccct ttgtcttttc ttttagctca gatTTTTtct ttctaagtat ttctgtattg 40680

aattagctta gtgacagaac acttgctggg tgtgcacatg gtactgggtt tgcacccatg 40740

cattacaaga atccaaacga cagcagaact aactgagagg agagcacagt agcggccgca 40800

aattgctttg agaggtctta taaaacctta gaggtatatt aaatttaaat ggccggcccg 40860

acggccaggc ggccggcagg cctaccact agtcaattcg ggaggatcga aacggcagat 40920

cgcaaaaaac agtacatata gaaggagaca tgaacatgaa catcaaaaaa attgtaaaac 40980

aagccacagt tctgactttt acgactgcac ttctggcagg aggagcgact caagccttcg 41040

cgaaagaaaa taacaaaaa gcatacaaag aaacgtacgg cgtctctcat attacacgcc 41100

atgatatgct gcagatccct aaacagcagc aaaacgaaaa ataccaagtg cctcaattcg 41160

atcaatcaac gattaaaaat attgagtcctg caaaaggact tgatgtgtgg gacagctggc 41220

cgctgcaaaa cgctgacgga acagtagctg aatacaacgg ctatcacgtt gtgtttgtc 41280

ttgcgggaag cccgaaagac gctgatgaca catcaatcta catgttttat caaaaggctc 41340

gcgacaactc aatcgacagc tggaaaaacg cgggcccgtgt ctttaaagac agcgataagt 41400

tcgacgcaa cgatccgac ctgaaagatc agacgcaaga atggtccggt tctgcaacct 41460

ttacatctga cggaaaaatc cgtttattct aactgacta ttccggtaaa cattacggca 41520

aacaaagcct gacaacagcg caggtaaatg tgtcaaaatc tgatgacaca ctcaaaatca 41580

acggagtggg agatcacaaa acgatttttg acggagacgg aaaaacatat cagaacgttc 41640

agcagtttat cgatgaaggc aattatacat ccggcgacaa ccatacgctg agagaccctc 41700

actacgttga agacaaaggc cataaatacc ttgtattcga agccaacacg ggaacagaaa 41760

acggatacca agcgaagaa tctttattta acaaagcgta ctacggcggc ggcacgaact 41820

tcttccgtaa agaaagccag aagcttcagc agagcgctaa aaaacgcgat gctgagttag 41880

cgaacggcgc cctcggtatc atagagttaa ataagtatta cacattgaaa aaagtaatga 41940

agccgctgat cacttcaaac acggttaactg atgaaatcga gcgcgcgaat gttttcaaaa 42000

tgaacggcaa atgggtactg ttactgatt cacgcggttc aaaaatgacg atcgatggta 42060

ttaactcaaa cgatatttac atgcttggtt atgtatcaaa ctctttaacc ggcccttaca 42120

agccgctgaa caaacaggc cttgtgctgc aaatgggtct tgatccaaac gatgtgacat 42180

tcacttactc tcacttcgca gtgccgcaag ccaaaggcaa caatgtggtt atcacaagct 42240



acatgacaaa cagaggcttc ttcgaggata aaaaggcaac atttgcgcca agctttcttaa 42300  
 tgaacatcaa aggcaataaa acatccgttg tcaaaaacag catcctggag caaggacagc 42360  
 tgacagtcaa ctaataacag caaaaagaaa atgccgatac ttcattggca tttttcttta 42420  
 ttttctcaaca agatggtgaa ttgactagt ggtagatcca caggacgggt gtggtcgcca 42480  
 tgatcgcgta gtcgatatg gctccaagta gcgaagcgag caggactggg cggcggccaa 42540  
 agcggtcgga cagtgtccg agaacgggtg cgcatagaaa ttgcatcaac gcatatagcg 42600  
 ctagcagcac gccatagtga ctggcgatgc tgtcggaatg gacgatatcc cgcaagaggc 42660  
  
 ccggcagtac cggcataacc aagcctatgc ctacagcatc cagggtgacg gtgccgagga 42720  
 tgacgatgag cgcatgttta gatttcatac acgggtgcctg actgcgttag caatttaact 42780  
 gtgataaact accgcattaa agcttatcga tgataagctg tcaaacatga gaattgatcc 42840  
 ggaaccctta atataacttc gtataatgta tgctatacga agttattagg tccctcgact 42900  
 atagggtcac cgtcgacagc gacacacttg catcgatgc agcccgttta acgtgccggc 42960  
 acggcctggg taaccaggta tttgtccac ataaccgtgc gcaaatgtt gtggataagc 43020  
 aggacacagc agcaatccac agcaggcata caaccgcaca ccgaggttac tccgttctac 43080  
  
 aggttacgac gacatgtcaa tacttgcct tgacaggcat tgatggaatc gtagtctcac 43140  
 gctgatagtc tgatgacaa tacaagtggg accgtgttcc cagaccgata atcagaccga 43200  
 caacacgagt gggatcgtgg tcccagacta ataatcagac cgacgatac agtgggaccg 43260  
 tgggtccaga ctaataatca gaccgacgat acgagtggga ccgtggttcc agactaataa 43320  
 tcagaccgac gatacagtg ggaccgtggt cccagactaa taatcagacc gacgatacga 43380  
 gtgggacat ggtcccagac taataatcag accgacgata cgagtgggac cgtggtccca 43440  
 gtctgattat cagaccgacg atacgagtgg gaccgtggtc ccagactaat aatcagaccg 43500  
  
 acgatacgag tgggaccgtg gtcccagact aataatcaga ccgacgatac gagtgggacc 43560  
 gtgggtccag tctgattatc agaccgacga tacaagtgga acagtgggcc cagagagaat 43620  
 attcaggcca gttatgcttt ctggcctgta acaaaggaca ttaagtaaag acagataaac 43680  
 gtagactaaa acgtggtcgc atcagggtgc tggcttttca agttccttaa gaatggcctc 43740  
 aattttctct atacactcag ttggaacacg agacctgtcc aggttaagca ccattttatc 43800  
 gcccttatac aatactgtcg ctccaggagc aaactgatgt cgtgagctta aactagtctc 43860  
 tgatgcagat gacgttttaa gcacagaagt taaaagagtg ataacttctt cagcttcaaa 43920  
  
 tatcacccca gcttttttct gctcatgaag gttagatgcc tgctgcttaa gtaattcctc 43980  
 tttatctgta aaggcttttt gaagtgcac acctgaccgg gcagatagtt caccgggggtg 44040  
 agaaaaaaga gcaacaactg atttaggcaa ttggcgggtg ttgatacagc gggtataaat 44100

cttactgtgaa atatittccg catcagccag cgcagaaata tttccagcaa attcattctg 44160  
 caatcggtt gcataacgt gaccacgttc ataagcactt gttgggcgat aatcgttacc 44220  
 caatctggat aatgcagcca tctgtctatc atccagctcg ccaaccagaa cacgataatc 44280  
 actttcggta agtgcagcag ctttacgacg gcgactccca tcggcaattt ctatgacacc 44340

agatactctt cgaccgaacg cgggtgtctg ttgaccagtc agtagaaaag aagggatgag 44400  
 atcatccagt ggcctctcag taagcagctc ctggtcacgt tcattacctg accatacccg 44460  
 agaggcttct tcaacactat caccctggag cacttcaaga gtaaacttca catcccgacc 44520  
 acatacagc aaagtaatgg cattaccgag agccattact cctacgcgcg caattaacga 44580  
 atccaccatc ggggcagctg gtgtcgataa cgaagtatct tcaaccggtt gagtattgag 44640  
 cgtatgtttt ggaataacag gcgcacgctt cattatctaa tctcccagcg tggtttaatc 44700  
 agacgatcga aaatttcatt gcagacaggt tcccaaatag aaagagcatt tctccaggca 44760

ccagttgaag agcgttgatc aatggcctgt tcaaaaacag ttctcatccg gatctgacct 44820  
 ttaccaactt catccgttcc acgtacaaca ttttttagaa ccatgcttcc ccaggcatcc 44880  
 cgaatttgc tctccatcca cggggactga gagccattac tattgctgta ttggtaagc 44940  
 aaaatacgt catcaggctc gaaccttta agatcaacgt tcttgagcag atcacgaagc 45000  
 atatcgaata actgcagtgc ggaggtgtag tcaacaact cagcaggcgt gggaacaatc 45060  
 agcacatcag cagcacatc gacattaatc gtgccgatac ccaggttagg cgcgctgtca 45120  
 ataactatga catcatagtc atgagcaaca gtttcaatgg ccagtcggag catcaggtgt 45180

ggatcgggtg gcagtttacc ttcatcaaat ttgccattt actcagtttc aatacgggtc 45240  
 agagccagac aggaaggaat aatgtcaagc cccggccagc aagtgggctt tattgcataa 45300  
 gtgacatcgt ctttttccc aagatagaaa ggcaggagag tgtcttctgc atgaatatga 45360  
 agatctggta cccatccgtg atacattgag gctgttcctt gggggtcgtt accttccacg 45420  
 agcaaaacac gtagccctt cagagccaga tctgagcaa gatgaacaga aactgaggtt 45480  
 ttgtaaacgc cacctttatg ggcagcaacc ccgatcaccg gtggaatac gtcttcagca 45540  
 cgtcgcaatc gcgtaccaa ccatcacgc atatgattaa tttgttcaat tgtataacca 45600

acacgttgct caaccgttc tcgaatttcc atatccgggt gcggtagtcg cctgtcttc 45660  
 tcggcatctc tgatagcctg agaagaaacc ccaactaaat ccgtgcttc acctattctc 45720  
 cagcgccggg ttattttcct cgcttccggg ctgtcatcat taaactgtgc aatggcgata 45780  
 gccttcgtca tttcatgacc agcgtttatg cactggttaa gtgtttccat gagtttcatt 45840  
 ctgaacatcc tttaatcatt gctttgcgtt tttttattaa atcttgcaat ttactgcaa 45900  
 gcaacaacaa aatcgcaaag tcatcaaaaa accgcaaagt tgtttaaaat aagagcaaca 45960

ctacaaaagg agataagaag agcacatacc tcagtcactt attatcacta gcgctcgccg 46020

cagccgtgta accgagcata gcgagcgaac tggcgaggaa gcaaagaaga actgttctgt 46080

cagatagctc ttacgctcag cgcaagaaga aatatccacc gtgggaaaaa ctccaggtag 46140

aggtagcacac gcggatagcc aattcagagt aataaactgt gataatcaac cctcatcaat 46200

gatgacgaac taacccccga tatcagggtca catgacgaag ggaaagagaa ggaaatcaac 46260

tgtgacaaac tgcctcaaa tttggcttcc ttaaaaatta cagttcaaaa agtatgagaa 46320

aatccatgca ggctgaagga aacagcaaaa ctgtgacaaa ttaccctcag taggtcagaa 46380

caaatgtgac gaaccaccct caaatctgtg acagataacc ctcagactat cctgtcgtca 46440

tggaagtgat atcgcggaag gaaaatacga tatgagtcgt ctggcggcct tcttttttct 46500

caatgtatga gaggcgcatt ggagttctgc tgttgatctc attaacacag acctgcagga 46560

agcggcggcg gaagtcaggc atacgttgtt aactttgagg cagctggtaa cgctctatga 46620

tccagtcgat tttcagagag acgatgcctg agccatccgg cttacgatac tgacacaggg 46680

attcgtataa acgcatggca tacggattgg tgattttctt tgtttcacta agccgaaact 46740

gcgtaaacgg gttctgtaac ccgataaaga agggaaatgag atatgggttg atatgtacac 46800

tgtaaagccc tcgggatgga ctgtgcgcac gtttgataaa ccaaggaaaa gattcatagc 46860

ctttttcatt gccggcatcc tcttcagggc gataaaaaac cacttcttc cccgcgaaac 46920

tcttcaatgc ctgccgtata tcttiactgg ctccgcaga ggtcaatccg aatatttcag 46980

catatttagc aacatggatc tcgagatac cgtcatgttc ctgtagggtg ccatcagatt 47040

ttctgatctg gtcaacgaac agatacagca tacgtttttg atcccgggag agactatatg 47100

ccgcctcagt gaggtcgttt gactggacga ttccgagggt atttttacgt tctttgtgat 47160

tgataaccgc tgtttccgc atgacagatc catgtgaagt gtgacaagtt tttagattgt 47220

cacactaaat aaaaaagagt caataagcag ggataacttt gtgaaaaaac agcttcttct 47280

gagggaatt tgtcacaggg ttaaggga tttgtcacag acaggactgt catttgaggg 47340

tgatttgta cactgaaagg gcaatttgc acaacacctt ctctagaacc agcatggata 47400

aaggcctaca aggcgctcta aaaaagaaga tctaaaaact ataaaaaaaa taattataaa 47460

aatatccccg tgataagtg gataaccca agggaagttt tttcaggcat cgtgtgtaag 47520

cagaatatat aagtgtgtt cctcgtgtc tctcgtcga ctcagggtgt tcgacctgtc 47580

gtcgaactgc ggcgagcact actggctgta aaaggacaga ccacatcatg gttctgtgtt 47640

cattaggttg tttgtccat tgctgacata atccgtcca cttcaacgta acaccgcacg 47700

aagatttcta ttgttctga aggcataattc aaatcgtttt cgttaccgct tgcaggcatc 47760  
atgacagaac actatttctt ataaacgcta cacaggctcc tgagattaat aatgcggatc 47820  
tctacgataa tgggagattt tcccactgtt ttcgttcgct tctcagtgga taacagccag 47880  
cttctctgtt taacagacaa aaacagcata tccactcagt tccacatttc catataaagg 47940  
ccaaggcatt tattctcagg ataattgttt cagcatcgca accgcatcag actccggcat 48000  
cgcaaatgc acccgggtcc gggcagccac atccagcgca aaaaccttcg tgtagacttc 48060  
cgttgaactg atggacttat gtcccatcag gctttgcaga actttcagcg gtataccggc 48120

atacagcatg tgcacgcat aggaatggcg gaacgtatgt ggtgtgaccg gaacagagaa 48180  
cgtcacaccg tcagcagcag cggcggcaac cgcctcccca atccaggctc tgaccgttct 48240  
gtccgtcact tcccagatcc gcgctttctc tgccttctt gtgcgacggt tacgccgctc 48300  
catgagctta tcggaataa atacctgtga cggaagatca cttcgagaa taaataaatc 48360  
ctggtgtccc tgttgatacc gggaagccct gggccaactt ttggcgaaaa tgagacgttg 48420  
atcggcacgt aagaggctcc aactttcacc ataataaat aagatcacta ccgggcgtat 48480  
tttttgagtt atcgagattt tcaggagcta aggaagctaa aatggagaaa aaaatcactg 48540

gatataccac cgttgatata tccaatggc atcgtaaaga acattttgag gcatttcagt 48600  
cagttgtcga atgtacctat aaccagaccg ttcagctgga tattacggcc tttttaaaga 48660  
ccgtaaaaga aaataagcac aagttttatc cggcctttat tcacattctt gcccgcctga 48720  
tgaatgctca tccggagttc cgtatggcaa tgaagacgg tgagctggtg atatgggata 48780  
gtgttcacc ttgttacacc gttttccatg agcaactga aacgttttca tcgctctgga 48840  
gtgaatacca cgacgatttc cggcagtttc tacacatata ttcgcaagat gtggcgtgtt 48900  
acggtgaaaa cctggcctat ttcctaaag ggtttattga gaatatgttt ttcgtctcag 48960

ccaatccctg ggtgagtttc accagttttg atttaaact ggccaatatg gacaacttct 49020  
tcgccccctg tttcacatg ggcaaatatt atacgcaagg cgacaagggt ctgatgccgc 49080  
tggcgattca ggttcacat gccgtttgtg atggcttcca tgcggcaga atgcttaatg 49140  
aattacaaca gtactgcgat gagtggcagg gcggggcgta atttttttaa ggcagttatt 49200  
ggtgccctta aacgcctggt tgctacgcct gaataagtga taataagcgg atgaatggca 49260  
gaaattcgat gataagctgt caaacatgag aattggctga cggcgcgcca aagcttgcat 49320  
gcctgcagcc gcgtaacctg gcaaaatcgg ttacggttga gtaataaat gatgcctgc 49380

gtaagcgggg cacatttcat tactctttc tccgcaccg acatagataa taacttcgta 49440  
tagtatacat tatacgaagt tatctagtag acttaattaa ggatcgatcc ggcgcgccaa 49500  
tagtcatgcc ccgcgccac cggaaggagc tgactgggtt gaaggctctc aagggcacgc 49560

gtcagagcttg acattgtagg actatattgc tctaataaat ttgcggccgc taatacgact 49620  
cactataggg a 49631  
<210> 18  
<211> 20  
<212> DNA  
<213> Artificial Sequence  
<220>  
<223> Synthetic  
<400> 18  
ggaaagccac cctgtatgct 20  
<210> 19  
<211> 19  
<212> DNA  
<213> Artificial Sequence  
<220>  
<223> Synthetic  
<400> 19  
cttgccaac agtggatgg 19  
<210> 20  
<211> 42  
<212> RNA  
<213> Artificial Sequence  
<220>  
<223> Synthetic  
<400> 20  
cuaaaaugau ucucaucugc guuuuagagc uaugcuguuu ug 42  
<210> 21  
<211> 42  
<212> RNA  
<213> Artificial Sequence  
<220>  
<223> Synthetic  
<400> 21

gcucucaacu ucacccuuuc guuuuagagc uaugcuguuu ug 42

<210> 22

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 22

ctaaaatgat tctcatctgc agg 23

<210> 23

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic

<400> 23

gtcttcaact tcacccttgc tgg 23

<210> 24

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic- a target locus that is linked to a  
guide RNA (gRNA)

<220>

<221> misc\_feature

<222> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,  
18, 19, 20, 21

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

<400> 24

nnnnnnnnnn nnnnnnnnnn ngg 23

<210> 25

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic- a target locus that is linked to a  
guide RNA (gRNA)

<220>

<221> misc\_feature

<222> 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,  
19, 20, 21, 22, 23

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

<400> 25

ggnnnnnnnnn nnnnnnnnnn nnngg

25