



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2021년07월08일

(11) 등록번호 10-2274445

(24) 등록일자 2021년07월01일

(51) 국제특허분류(Int. Cl.)
C12Q 1/68 (2018.01) **C12N 15/11** (2006.01)
C12N 15/90 (2006.01) **C12N 9/22** (2006.01)
(52) CPC특허분류
C12Q 1/6897 (2018.05)
C12N 15/111 (2013.01)
(21) 출원번호 10-2016-7019329
(22) 출원일자(국제) 2014년12월19일
심사청구일자 2019년12월18일
(85) 번역문제출일자 2016년07월15일
(65) 공개번호 10-2016-0098481
(43) 공개일자 2016년08월18일
(86) 국제출원번호 PCT/US2014/071693
(87) 국제공개번호 WO 2015/095804
국제공개일자 2015년06월25일
(30) 우선권주장
61/918,625 2013년12월19일 미국(US)
61/937,444 2014년02월07일 미국(US)
(56) 선행기술조사문헌
W02013141680 A1
W02012149470 A1
JP2012531909 A
Gene, 58, p.201-216, 1987.

(73) 특허권자
아미리스 인코퍼레이티드
미국, 캘리포니아주 94608, 에머리빌, 스위트
100, 5885 홀리스 스트리트
(72) 발명자
호르위츠, 앤드류
미국, 캘리포니아 94608, 에머리빌, 스위트 100,
5885 홀리스 스트리트
호킨스, 크리스티 미셸
미국, 캘리포니아 94608, 에머리빌, 스위트 100,
5885 홀리스 스트리트
(뒷면에 계속)
(74) 대리인
특허법인한얼

전체 청구항 수 : 총 30 항

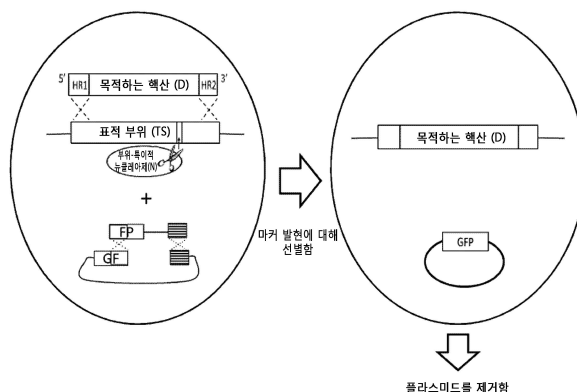
심사관 : 이재영

(54) 발명의 명칭 **게놈 삽입을 위한 방법**

(57) 요약

본원에는 하나 이상의 외인성 핵산을 숙주 세포 게놈의 하나 이상의 선택된 표적 부위 내로 삽입시키는 방법이 제공된다. 특정 실시형태에서, 상기 방법은 숙주 세포 게놈을, 게놈 표적 부위 내로 삽입되는 외인성 핵산, 게놈 표적 부위에서 절단을 유발할 수 있는 뉴클레아제, 및 그 자체로, 또는 세포 집단과 접촉된 하나 이상의 추가 (뒷면에 계속)

대표도



선형 핵산과 상동성 재조합할 수 있는 선형 핵산을 포함하는 하나 이상의 삼입 폴리뉴클레오타이드와 접촉시키는 것을 포함하고, 상기 상동성 재조합은 선별가능한 마커를 위한 코딩 서열을 포함하는 환상 염색체외 핵산의 형성을 제공한다. 일부 실시형태에서, 상기 방법은 선별가능한 마커를 발현하는 숙주 세포를 선별하는 것을 추가로 포함한다.

(52) CPC특허분류

C12N 15/902 (2013.01)

C12N 15/907 (2013.01)

C12N 9/22 (2013.01)

C12N 2310/10 (2013.01)

(72) 발명자

슈베르트, 맥스

미국, 캘리포니아 94608, 에머리빌, 스위트 100,
5885 홀리스 스트리트

스제토, 웨인

미국, 캘리포니아 94608, 에머리빌, 스위트 100,
5885 홀리스 스트리트

명세서

청구범위

청구항 1

외인성 핵산을 숙주 세포 계놈의 표적 부위 내로 삽입하는 시험관내 방법으로서,

(a) 하나 이상의 숙주 세포를 하기와 접촉시키는 단계:

(i) 숙주 세포 계놈의 표적 부위(TS)에서 상동성 재조합에 의해 재조합할 수 있는 외인성 핵산(ES)으로서,

ES가 제1 상동성 영역(HR1) 및 제2 상동성 영역(HR2)를 포함하고, HR1 및 HR2가 제3 상동성 영역(HR3) 및 제4 상동성 영역(HR4) 각각과 상동성 재조합에 의해 재조합할 수 있으며, HR3 및 HR4가 각각 TS에 존재하는, ES;

(ii) TS에서 절단을 생성할 수 있는 뉴클레아제(N); 및

(iii) 숙주 세포와 접촉된 하나 이상의 추가의 선형 핵산의 상동성 영역과 상동성 재조합할 수 있는 상동성 영역을 포함하는 선형 핵산으로서, 선형 핵산의 상동성 재조합에 의해 선별가능한 마커에 대한 코딩 서열을 포함하는 환상 염색체의 핵산을 형성하는, 선형 핵산; 및

(b) 선별가능한 마커를 발현하는 숙주 세포를 선별하는 단계를 포함하는, 방법.

청구항 2

제1항에 있어서, 선형 핵산이 선형 핵산으로부터 발현될 수 없는 선별가능한 마커에 대한 부분적, 단속적 및/또는 비-인접적 코딩 서열을 포함함으로써 환상 염색체의 핵산의 형성에 의해 환상 염색체의 핵산으로부터 발현될 수 있는 선별가능한 마커의 완전한 코딩 서열을 형성하는, 방법.

청구항 3

제1항에 있어서, (b) 단계의 선별하는 단계가 시각, 비색(colorimetric) 또는 형광 검출 방법을 통해 선별가능한 마커의 발현을 검출하는 단계를 포함하는, 방법.

청구항 4

제1항에 있어서, 숙주 세포를 회수하는 단계를 추가로 포함하고, ES는 TS에서 상동성 재조합되는, 방법.

청구항 5

제4항에 있어서,

(a) 상기 회수가 숙주 세포 계놈 내로 선별가능한 마커의 삽입을 필요로 하지 않거나; 또는

(b) 상기 회수가 스크리닝된 10, 9, 8, 7, 6, 5, 4, 3 또는 2개의 접촉된 숙주 세포 또는 이의 클론 집단당 적어도 약 1개의 빈도로 발생하는, 방법.

청구항 6

제1항에 있어서, 복수(n)의 외인성 핵산이 숙주 세포 계놈의 복수(n)의 표적 부위 내로 삽입되고, n이 적어도 2이며, (a) 단계가 숙주 세포를 하기와 접촉시키는 단계를 포함하는 방법:

(i) 상기 복수의 외인성 핵산으로서, x은 1 내지 n의 정수이고, 각각의 정수 x에 대해, 각각의 외인성 핵산(ES)_x은 상기 숙주 세포 계놈의 상기 복수(n)의 표적 부위로부터 선택된 표적 부위(TS)_x에서 상동성 재조합에 의해 재조합할 수 있는, 외인성 핵산;

(ii) 각각의 상기 표적 부위(TS)_x에 대해, (TS)_x에서 절단을 생성할 수 있는 뉴클레아제(N)_x.

청구항 7

제6항에 있어서,

(a) 단일 뉴클레아제가 (TS)_x 각각을 절단시킬 수 있거나; 또는

(b) n이 3, 4, 5, 6, 7, 8, 9 또는 10인, 방법.

청구항 8

제1항에 있어서,

(a) N이 TS에서 단일-가닥 절단 또는 이중-가닥 절단을 생성할 수 있거나; 또는

(b) ES가 선형인, 방법.

청구항 9

제1항에 있어서, ES가 목적하는 D의 핵산을 추가로 포함하고, D가 선별가능한 마커, 프로모터, 에피토프 태그를 코딩하는 핵산 서열, 목적하는 유전자, 리포터 유전자, 및 종결 코돈을 코딩하는 핵산 서열로 이루어진 그룹으로부터 선택되는, 방법.

청구항 10

제1항에 있어서, 환상 염색체의 핵산이 뉴클레아제에 대한 코딩 서열을 추가로 포함하는, 방법

청구항 11

제10항에 있어서, 뉴클레아제가 RNA-유도 DNA 엔도뉴클레아제 또는 Cas9 엔도뉴클레아제인, 방법.

청구항 12

제10항에 있어서, 환상 염색체의 핵산이 RNA-유도 DNA 엔도뉴클레아제에 의한 TS의 부위-특이적 인식 및 절단을 가능하게 하는 crRNA 활성 및 tracrRNA 활성을 코딩하는 서열을 추가로 포함하는, 방법.

청구항 13

제12항에 있어서, crRNA 활성 및 tracrRNA 활성이 단일 인접 RNA 분자로서 발현되는, 방법.

청구항 14

제1항에 있어서, 뉴클레아제가 엔도뉴클레아제, 징크 핑거 뉴클레아제, TAL-작동자 DNA 결합 도메인-뉴클레아제 융합 단백질(TALEN), 트랜스포사제 및 부위-특이적 재조합효소로 이루어진 그룹으로부터 선택되는, 방법.

청구항 15

제14항에 있어서,

(a) 징크 핑거 뉴클레아제가 조작된(engineered) 징크 핑거 결합 도메인에 융합된 TypeIIS 제한 엔도뉴클레아제의 절단 도메인을 포함하는 융합 단백질로서, TypeIIS 제한 엔도뉴클레아제가 HO 엔도뉴클레아제 및 Fok I 엔도뉴클레아제로 이루어진 그룹으로부터 선택되며, 징크 핑거 결합 도메인이 3, 5 또는 6개의 징크 핑거를 포함하거나; 또는

(b) 엔도뉴클레아제가 하기인, 방법:

(i) LAGLIDADG 호밍 (homing) 엔도뉴클레아제, HNH 호밍 엔도뉴클레아제, His-Cys box 호밍 엔도뉴클레아제, GIY-YIG 호밍 엔도뉴클레아제 및 시아노박테리아 호밍 엔도뉴클레아제로 이루어진 그룹으로부터 선택되는 호밍 엔도뉴클레아제;

(ii) 하기로 이루어진 그룹으로부터 선택되는 엔도뉴클레아제:

H-DreI, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAIIIP, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-TliI, I-PpoI, Pi-PspI, F-SceI, F-SceII, F-SuvI, F-CphI, 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-NclIIP, 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-Ssp68031, 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-MgaI, 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 및 PI-TliII; 또는

(iii) 내인성 게놈 서열에 특이적으로 결합하도록 변형된 엔도뉴클레아제로서, 변형된 엔도뉴클레아제가 이의 야생형 엔도뉴클레아제 인식 서열에 더 이상 결합하지 않고, (i) 또는 (ii)의 야생형 엔도뉴클레아제에서 유래하는, 엔도뉴클레아제.

청구항 16

제1항 내지 제15항 중 어느 한 항에 있어서, 숙주 세포가 하기인, 방법:

- (a) 원핵생물 세포;
- (b) 진핵생물 세포;
- (c) 진균 세포, 박테리아 세포, 식물 세포, 곤충 세포, 조류 세포, 어류 세포 및 포유동물 세포로 이루어진 그룹으로부터 선택되는 세포;
- (d) 설치류 세포, 영장류 세포 및 인간 세포로 이루어진 그룹으로부터 선택되는 포유동물 세포; 또는
- (e) 효모 세포 또는 사카로마이세스 세레비지애(*Saccharomyces cerevisiae*) 세포.

청구항 17

하기를 포함하는 단리된 숙주 세포:

- (i) 숙주 세포 게놈의 표적 부위(TS)에서 상동성 재조합에 의해 재조합할 수 있는 상동성 영역을 포함하는 외인성 핵산(ES)으로서, ES가 제1 상동성 영역(HR1) 및 제2 상동성 영역(HR2)을 포함하고, HR1 및 HR2가 제3 상동성 영역(HR3) 및 제4 상동성 영역(HR4) 각각과 상동성 재조합에 의해 재조합할 수 있으며, HR3 및 HR4가 각각 TS에 존재하는, ES;
- (ii) TS에서 절단을 생성할 수 있는 뉴클레아제(N); 및
- (iii) 서로 상동성 재조합할 수 있는 상동성 영역을 포함하는 적어도 두 개의 선형 핵산으로서, 상기 상동성 재조합에 의해 선별가능한 마커에 대한 코딩 서열을 포함하는 환상 염색체의 핵산을 형성하는, 적어도 두 개의 선형 핵산.

청구항 18

제17항에 있어서, 선형 핵산이, 선형 핵산으로부터 발현될 수 없는 선별가능한 마커에 대한 부분적, 단속적 및/또는 비-인접적 코딩 서열을 포함함으로써 환상 염색체의 핵산의 형성에 의해 환상 염색체의 핵산으로부터 발현될 수 있는 선별가능한 마커의 완전한 코딩 서열을 형성하는, 숙주 세포.

청구항 19

제17항에 있어서, 하기를 포함하는 숙주 세포:

- (i) 복수의 외인성 핵산으로서, x 는 1 내지 n 의 정수이고, 각각의 정수 x 에 대해, 각각의 외인성 핵산(ES) $_x$ 은 상기 숙주 세포 게놈의 복수(n)의 표적 부위로부터 선택된 표적 부위(TS) $_x$ 에서 상동성 재조합에 의해 재조합할 수 있는, 외인성 핵산;
- (ii) 각각의 상기 표적 부위(TS) $_x$ 에 대해, (TS) $_x$ 에서 절단을 생성할 수 있는 뉴클레아제(N) $_x$.

청구항 20

제17항에 있어서,

- (a) N이 TS에서 단일-가닥 절단 또는 이중-가닥 절단을 생성할 수 있거나; 또는
- (b) ES가 선형인, 숙주 세포.

청구항 21

제17항에 있어서, ES가 목적하는 D의 핵산을 추가로 포함하고, D가 선별가능한 마커, 프로모터, 에피토프 태그를 코딩하는 핵산 서열, 목적하는 유전자, 리포터 유전자, 및 종결 코돈을 코딩하는 핵산 서열로 이루어진 그룹으로부터 선택되는, 숙주 세포.

청구항 22

제17항에 있어서, 환상 염색체의 핵산이 뉴클레아제에 대한 코딩 서열을 추가로 포함하는, 숙주 세포.

청구항 23

제22항에 있어서, 뉴클레아제가 RNA-유도 DNA 엔도뉴클레아제 또는 Cas9 엔도뉴클레아제인, 숙주 세포.

청구항 24

제22항에 있어서, 환상 염색체의 핵산이 RNA-유도 DNA 엔도뉴클레아제에 의한 TS의 부위-특이적 인식 및 절단을 가능하게 하는 crRNA 활성 및 tracrRNA 활성을 코딩하는 서열을 추가로 포함하는, 숙주 세포.

청구항 25

제24항에 있어서, crRNA 활성 및 tracrRNA 활성이 단일 인접 RNA 분자로서 발현되는, 숙주 세포.

청구항 26

제17항에 있어서, 뉴클레아제가 엔도뉴클레아제, 징크 핑거 뉴클레아제, TAL-작동자 DNA 결합 도메인-뉴클레아제 융합 단백질(TALEN), 트랜스포사제 및 부위-특이적 재조합효소로 이루어진 그룹으로부터 선택되는, 숙주 세포.

청구항 27

제26항에 있어서,

(a) 징크 핑거 뉴클레아제가 조작된 징크 핑거 결합 도메인에 융합된 TypeIIS 제한 엔도뉴클레아제의 절단 도메인을 포함하는 융합 단백질로서, TypeIIS 제한 엔도뉴클레아제가 HO 엔도뉴클레아제 및 Fok I 엔도뉴클레아제로 이루어진 그룹으로부터 선택되며, 징크 핑거 결합 도메인이 3, 5 또는 6개의 징크 핑거를 포함하거나; 또는

(b) 엔도뉴클레아제가 하기인, 숙주세포:

(i) LAGLIDADG 호밍 (homing) 엔도뉴클레아제, HNH 호밍 엔도뉴클레아제, His-Cys box 호밍 엔도뉴클레아제, GIY-YIG 호밍 엔도뉴클레아제 및 시아노박테리아 호밍 엔도뉴클레아제로 이루어진 그룹으로부터 선택되는 호밍 엔도뉴클레아제;

(ii) 하기로 이루어진 그룹으로부터 선택되는 엔도뉴클레아제:

H-DreI, 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-CrepsbIIVP, I-TliI, I-PpoI, Pi-PspI, F-SceI, F-SceII, F-SuvI, F-CphI, F-TevI, F-TevII, I-AmaI, I-AniI, I-ChuI, I-CmoI, 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-NclIP, 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-Ssp68031, 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-MgaI, 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 및 PI-TliII; 또는

(iii) 내인성 게놈 서열에 특이적으로 결합하도록 변형된 엔도뉴클레아제로서, 변형된 엔도뉴클레아제가 이의

야생형 엔도뉴클레아제 인식 서열에 더 이상 결합하지 않고, (i) 또는 (ii)의 야생형 엔도뉴클레아제에서 유래하는, 엔도뉴클레아제.

청구항 28

- (a) 부위-특이적 뉴클레아제, 또는 부위-특이적 뉴클레아제에 대한 코딩 서열을 포함하는 핵산;
 - (b) 제1 선형 핵산 및 하나 이상의 추가의 선형 핵산으로서, 제1 및 제2 선형 핵산은 숙주 세포 내로 도입된 이후에 서로 상동성 재조합할 수 있음으로써 상기 상동성 재조합에 의해 선별가능한 마커에 대한 코딩 서열을 포함하는 환상 핵산을 형성하는, 선형 핵산; 및
 - (c) 숙주 세포 계놈의 표적 부위에서 상동성 재조합에 의해 재조합할 수 있는 상동성 영역을 포함하는 외인성 핵산
- 을 포함하는, 조성물.

청구항 29

세포 배양 배지, 및 제17항 내지 제27항 중의 어느 한 항의 숙주 세포를 포함하는, 세포 배양 조성물.

청구항 30

외인성 핵산을 숙주 세포 계놈의 표적 부위 내로 삽입시키는 시험관내 방법으로서,

- (a) 하나 이상의 숙주 세포를 하기와 접촉시키는 단계:
 - (i) 숙주 세포 계놈의 표적 부위(TS)에서 상동성 재조합에 의해 재조합할 수 있는 외인성 핵산(ES); 및
 - (ii) 숙주 세포와 접촉된 하나 이상의 추가의 선형 핵산과 상동성 재조합할 수 있는 선형 핵산으로서, 상기 상동성 재조합에 의해 선별가능한 마커에 대한 코딩 서열, 및 TS에서 절단을 생성할 수 있는 뉴클레아제(N)에 대한 코딩 서열을 포함하는 환상 염색체의 핵산을 형성하는, 선형 핵산; 및
- (b) 선별가능한 마커를 발현하는 숙주 세포를 선별하는 단계를 포함하는, 방법.

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

청구항 36

삭제

청구항 37

삭제

청구항 38

삭제

청구항 39

삭제

청구항 40

삭제

청구항 41

삭제

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

청구항 60

삭제

청구항 61

삭제

청구항 62

삭제

청구항 63

삭제

청구항 64

삭제

청구항 65

삭제

청구항 66

삭제

청구항 67

삭제

청구항 68

삭제

청구항 69

삭제

청구항 70

삭제

청구항 71

삭제

청구항 72

삭제

청구항 73

삭제

청구항 74

삭제

청구항 75

삭제

청구항 76

삭제

청구항 77

삭제

청구항 78

삭제

청구항 79

삭제

청구항 80

삭제

청구항 81

삭제

청구항 82

삭제

청구항 83

삭제

청구항 84

삭제

청구항 85

삭제

청구항 86

삭제

청구항 87

삭제

청구항 88

삭제

청구항 89

삭제

청구항 90

삭제

청구항 91

삭제

청구항 92

삭제

청구항 93

삭제

청구항 94

삭제

청구항 95

삭제

청구항 96

삭제

청구항 97

삭제

청구항 98

삭제

청구항 99

삭제

청구항 100

삭제

청구항 101

삭제

청구항 102

삭제

청구항 103

삭제

청구항 104

삭제

청구항 105

삭제

청구항 106

삭제

청구항 107

삭제

청구항 108

삭제

청구항 109

삭제

청구항 110

삭제

청구항 111

삭제

청구항 112

삭제

청구항 113

삭제

청구항 114

삭제

청구항 115

삭제

청구항 116

삭제

청구항 117

삭제

청구항 118

삭제

청구항 119

삭제

청구항 120

삭제

청구항 121

삭제

청구항 122

삭제

청구항 123

삭제

청구항 124

삭제

청구항 125

삭제

청구항 126

삭제

청구항 127

삭제

청구항 128

삭제

청구항 129

삭제

청구항 130

삭제

청구항 131

삭제

청구항 132

삭제

청구항 133

삭제

청구항 134

삭제

청구항 135

삭제

청구항 136

삭제

청구항 137

삭제

청구항 138

삭제

청구항 139

삭제

청구항 140

삭제

청구항 141

삭제

청구항 142

삭제

청구항 143

삭제

청구항 144

삭제

청구항 145

삭제

청구항 146

삭제

발명의 설명

기술 분야

[0001] 관련 출원의 상호 참조

[0002] 본 출원은 2013년 12월 19일자로 출원된 미국 가출원 제61/918,625호 및 2014년 2월 7일자로 출원된 미국 가출원 제61/937,444호의 이익을 주장하며, 이들 각각은 이의 전체가 참조로서 원용된다.

[0003] 1. 발명의 분야

[0004] 본원에 제공된 방법 및 조성물은 일반적으로 분자 생물학 및 유전자 공학의 분야에 관한 것이다.

배경 기술

[0005] 2. 배경

[0006] 표적화된 변형체를 숙주 세포 계놈 내로 도입하는 유전자 조작 기술은 다양한 분야에서의 용도를 발견한다. 기본적으로, 유전자형이 표현형에 어떻게 영향을 미치는지에 대한 결정은 천연 유전자 기능을 손상시키거나 폐지하기 위해 표적화된 삽입체 또는 결실체를 도입하는 능력에 의존한다. 합성 생물학 분야에서, 목적하는 화합물을 생산할 수 있는 유전자 유전자 변형된 미생물의 제조는 맞춤형 DNA 서열을 숙주 세포의 염색체 내로 삽입(integration)하는 것을 필요로 하고, 산업적 규모의 생산은 일반적으로 수십개의 유전자를 단일 숙주 계놈 내로 도입하는 것, 예를 들면, 전체 생합성 경로를 필요로 한다. 치료 문맥에서, 정확한 계놈 변형체를 도입하는 능력은 단일-유전자 결함에 기인하는 질환, 예를 들면, X-연결된 중증 조합 면역 결핍증(SCID), 혈우병 B, 베타-탈라세미아, 낭포성 섬유증, 근 이영양증 및 겸상 적혈구병에 대처하기 위한 거대한 가능성을 갖는다.

[0007] 계놈 조작에서의 최근의 진보는 다양한

[0008] 세포형 및 생물체에 걸쳐 실질적으로 임의의 유전자의 조작 및/또는 도입을 가능하게 했다. 특히, 부위 특이적 디자인어 뉴클레아제의 출현은 표적화된 절단을 숙주 세포 계놈 내로 도입함으로써, 즉 계놈 편집에 의해 부위-특이적 유전자 변형체를 가능하게 했다. 이들 뉴클레아제는 징크 핑거 뉴클레아제(ZFN), 전사 활성화인자-유사 작동자(effector) 뉴클레아제(TALEN) 및 클러스터된 조절 이격된 짧은 회귀성 반복체 CRISPR/Cas(CRISPR-연관된)-기반 RNA-유도 엔도뉴클레아제를 포함한다. ZFN은 특히 작물에서 표적 유전자좌를 변형시키기 위해[참조: Wright *et al.*, *Plant J* 44:693-705 (2005)], 치료학적 항체를 발현하는 포유동물 세포 배양계를 개선시키기 위해[참조: Malphettes *et al.*, *Biotechnol Bioeng* 106(5): 774-783 (2010)] 및 인간 계놈을 편집하여 HIV에 대한 내성을 유발하기 위해[참조: Urnov *et al.*, *Nat Rev Genet* 11(9):636-646 (2010)] 사용되어 왔다. 유사하게는, TALEN은 작물 식물[참조: Li, *et al.*, *Nat. Biotechnol.* 30: 390-392 (2012)], 인간, 소 및 마우스[참조: Xu *et al.*, *Molecular Therapy-Nucleic Acids* 2, e112 (2013)]의 계놈을 포함하여 다양한 계놈을 변형시키기 위해 사용되어 왔다. 보다 최근에, CRISPR은 박테리아[참조: Jiang *et al.*, *Nature Biotechnology* 31(3):233-239 (2013); Qi *et al.*, *Cell*, 5, 1173-1183 (2013)]; 효모[참조: DiCarlo *et al.*, *Nucleic Acids Res.*, 7, 4336-4343 (2013)]; 제브라피쉬[참조: Hwang *et al.*, *Nat. Biotechnol.*, 3, 227-229(2013)]; 과실 파리[참조: Gratz *et al.*, *Genetics*, 194, 1029-1035 (2013)]; 인간 세포[참조: Cong *et al.*, *Science* 6121, 819-823, (2013); Mali *et al.*, *Science*, 6121, 823-826 (2013); Cho *et al.*, *Nat. Biotechnol.*, 3, 230-232 (2013)]; 및 식물[참조: Jiang *et al.*, *Nucleic acids Research* 41(20):e188 (2013)]; Belhaj *et al.*, *Plant Methods* 9(39) (2013)]의 계놈을 편집하기 위해 성공적으로 사용되어 왔다.

[0009] 부위-특이적 뉴클레아제는 비-상동성 말단 결합(NHEJ), 단일-가닥 어닐링(SSA) 및 상동성-지향 수복(HDR)을 포함하여 숙주 세포의 세포 DNA 수복(repair) 메커니즘을 자극하는 염색체 DNA에서 절단을 유도한다. 뉴클레아제-유도된 이중-가닥 절단(DSB)의 NHEJ-매개된 수복은 표적화된 부위에서 작은 결실 또는 삽입의 도입을 유도하여, 예를 들면, 프레임쉬프트 돌연변이를 통해 유전자 기능의 손상 또는 폐지를 유도한다. 동일한 분자의 절단된 말단은 또 다른 DNA 분자를 수반하지 않는 다단계 효소적 프로세스에 의해 재결합된다. NHEJ는 에러가 발생하기 쉽거나 부정확하여, 수복 동안 절단-부위에서 가변 크기의 상이한 및 예상치 못한 삽입 및 결실을 갖는 돌연변이 대립인자를 생성한다. 유사하게는, SSA는 DSB에 인접하는 서열 반복체로부터의 상보성 쇄가 서로 어닐링되는 경우에 발생하여, DSB의 수복, 그러나 개재 서열의 결실을 제공한다. 대조적으로, HDR은, 이것이 절단의 수복을 돕기 위해 상동성 서열을 갖는 별개의 무손상 분자에 의존하기 때문에, 통상적으로 정확하게 복원된 분자를 유도한다. 세포에 고유한 상동성 공여체 서열의 2개 주요 공급원이 있다: 세포 사이클 전체에 걸쳐 이용가능한 상동성 염색체, 및 절단된 분자의 시스터 크로마티드(이는 DNA가 복제된 후에만 이용가능하다). 그러나, 계놈 조작 기술은 통상적으로 DSB의 표적 부위와 상동성인 영역을 포함하고 표적 부위와 재조합할 수 있는 외인성 공여체 DNA를 도입한다. 표적 서열에 대한 목적하는 변형을 외인성 공여체에 포함시킴으로써, 이들 변형은 HDR을 통해 원래 표적 서열 내로 삽입되어 대체될 수 있다.

[0010] DNA의 뉴클레아제-유도된 파손시, 수복 경로의 숙주 세포 선택은 다수의 요인에 의존하고, 결과는 목적하는 계놈 변형의 정확도를 결정할 수 있다. 이러한 요인은 숙주 세포의 DNA 손상 신호전달 경로, 절단의 성질, 크로마틴 리모델링, 특정 수복 단백질의 전사, 및 세포 사이클의 후기 단계에 존재하는 사이클린-의존성 키나제 활성을 포함한다[참조: Beucher *et al.*, *EMBO J* 28:3413-27(2009); Soensen *et al.*, *Nat Cell Biol* 7:195-201 (2005); Jazayeri *et al.*, *Nat Cell Biol* 8:37-45 (2006); Huertas *et al.*, *Nature* 455:689-92 (2008); Moyal *et al.*, *Mol Cell* 41:529-42 (2011); and Chernikova *et al.*, *Radiat Res* 174:558-65 (2010)]. 절단된 DNA에 대해 강력한 상동성을 갖는 공여체 DNA가 존재하는 경우, 상동성 재조합에 의한 공여체의 삽입 기회는 현저히 증가한다[참조: Moehle *et al.*, *Proc. Natl Acad. Sci. USA*, 9:3055-3060 (2007); Chen *et al.*, *Nat.*

Methods, 9, 753-755 (2011)]. 그러나, 상동성 공여체 DNA가 절단된 표적 부위 내로 HDR에 의해 삽입되는 전체 빈도는, NHEJ에 의해 표적 부위의 비-삽입 수복과 대조적으로, 여전히 매우 낮을 수 있다. 최근의 연구는 HDR-매개된 편집이 일반적으로 저효율 사상이고, 보다 덜 정확한 NHEJ는 DSB에 대한 수복 메카니즘으로서 우수할 수 있음을 시사한다.

[0011] 예를 들면, 말리 등[참조: Mali *et al.*, *Science* 339:823-826 (2013)]은 CRISPR(유도 RNA 및 Cas9 엔도뉴클레아제) 및 현재 공급된 단일-가닥 공여체 DNA를 사용하여 인간 K562 세포에서 유전자 변형을 시도했고, 2.0%의 빈도로 AAVS1 유전자좌에서 HDR-매개된 유전자 변형을 관찰한 반면, 동일한 유전자좌에서 NHEJ-매개된 표적화된 돌연변이유발은 38%의 빈도로 관찰되었다. 리 등[참조: Li *et al.*, *Nat Biotechnol.* (8):688-91 (2013)]은 CRISPR(유도 RNA 및 Cas9 엔도뉴클레아제) 및 현재 공급된 이중-가닥 공여체 DNA를 사용하여 식물 니코티아나 벤타미아나(*Nicotiana benthamiana*)에서 유전자 대체를 시도했고, 9.0%의 빈도로 HDR-매개된 유전자 교환을 관찰한 반면, NHEJ-매개된 표적화된 돌연변이유발은 14.2%의 빈도로 관찰되었다. 카스 등[참조: Kass *et al.*, *Proc Natl Acad Sci U S A.* 110(14): 5564-5569 (2013)]은 다양한 계통으로부터 유래된 초대(primary) 정상 체세포 유형에서 HDR을 연구했고, 포유동물 상피, 난소 및 신생아 뇌로부터 유래된 세포 뿐만 아니라 마우스 배 및 성인 섬유아세포가 대략 1%(0.65-1.7%)의 빈도로 I-SceI 엔도뉴클레아제-유도된 DSB에서 HDR을 겪었음을 관찰했다. 카스 등은 세포가 세포 사이클의 S 및 G2 단계에 존재하는 경우에 보다 높은 HDR 활성을 보고했다. 리 등[참조: Li *et al.*, *Nat Biotechnol.* (8):688-91 (2013)] 아라비도시스(*Arabidopsis*) CYCD3(사이클린 D-형 3), 세포 사이클의 마스터 활성화인자의 동시-발현에 의해 이소성 세포 분열을 유발함으로써 니코티아나 벤타미아나(*Nicotiana benthamiana*)에서 HDR을 증강시키는 가능성을 시험했지만, 이는 HDR의 속도를 거의 촉진시키지 않았다(9% - CYCD3으로부터 최대 11.1%). HDR 속도를 개선시키기 위한 전략은 또한 길항 NHEJ 수복 메카니즘을 해체하는 것(knocking out)을 포함했다. 예를 들면, 퀴 등[참조: Qi *et al.*, *Genome Res* 23:547-554 (2013)]은 *ku70* 돌연변이체의 경우에 아라비도시스(*Arabidopsis*)의 HDR-매개된 유전자 표적화에서 5 내지 6배의 증가 및 *lig4* 돌연변이의 경우에 3 내지 4배의 증가를 보고했다. 그러나, 전체 속도는 약 5% 이하, 대부분 1% 미만인 것으로 관찰되었다. 추가로, 목적하는 유전자-표적화 사상이 생성되는 경우, *ku70* 또는 *lig4* 돌연변이는 돌연변이체 식물 외부로 교차되어야 했다.

[0012] 대부분의 세포 유형에서 HDR-매개된 삽입의 비교적 낮은 속도를 고려하면, 염색체 내로 외인성 DNA의 삽입은 통상 선별가능한 마커의 동시 삽입을 필요로 하고, 이는 목적하는 삽입 사상을 겪는 형질전환된 세포를 위한 농후화를 필요로 한다. 그러나, 이는 하류 적용과 상용성이 아닐 수 있는 계놈 내로 외인성 서열을 도입하고, 마커의 연장된 발현은 또한 유해한 효과를 가질 수도 있다. 예를 들면, 인간 세포 계놈 내로 네오마이신 내성 유전자의 도입, 이어서 G418에서 연장된 시간의 배양은 세포의 특성에 대한 변화를 유발하는 것으로 보고되었고, 증강된 녹색 형광 단백질(EGFP) 및 기타 형광 단백질의 발현은 면역성 및 독성을 유발하는 것으로 보고되었다 [참조: Barese *et al.*, *Human Gene Therapy* 22:659-668 (2011); Morris *et al.*, *Blood* 103:492-499 (2004); and Hanazono *et al.*, *Human Gene Therapy* 8:1313-1319 (1997)]. 추가로, 유전자 변형된(GM) 식물에서 선별가능한-마커 유전자의 삽입은 다른 생물체로 수평 전달의 관심을 제기했고, 항생물질 내성 마커의 경우에, 이들 마커는 항생물질 내성 박테리아 균주에서 증가를 유도할 수 있음이 특히 중요하다. 유사한 관심은 제조제-내성 마커의 삽입 및 신규 공격적 잡초의 가능한 생성에 관한 것이다. 최저에서, 후기 단계에서 삽입된 마커 서열의 제거는 시간 및 노동 집약적이다. 이는 선별가능한 마커의 제한된 캐쉬만이 소정 숙주에서 이용가능하고 마커가 추가 조작 단계를 가능하게 하기 위해 재순환되어야 하는 경우에 특히 문제가 된다. 따라서, 특정한 적용은, 예를 들면, 안전성 및/또는 조절 준수를 위해 목적하는 표현형을 수행하는데 요구되는 최소 외인성 서열만을 도입하는 것이 타당하고, 궁극적으로 완전히 마커 삽입의 회피를 필요로 할 수 있다.

[0013] 따라서, 숙주 세포 계놈 내로 하나 이상의 외인성 핵산의 HDR-매개된 삽입의 효율 및/또는 선별을 개선시키는 방법 및 조성물에 대한 필요성이 존재한다. 더욱이, 선별가능한 마커에 대한 코딩 서열의 동시-삽입을 필요로 하지 않는 계놈 조작 전략에 대한 필요성이 존재한다. 이들 및 기타 필요성은 본원에 제공된 조성물 및 방법에 의해 충족된다.

발명의 내용

[0014] 3. 요약

[0015] 본원에 제공된 방법 및 조성물은 상동성 재조합(HR)-적격 숙주 세포를 선별하는 방법에 관한 것이다. 조작 이론에 구속되지 않지만, 세포 집단 중에서 HR-적격은 숙주 세포 내로 도입된 하나 이상의 선행 단편(들)을 상동성 재조합하여, 선별가능한 마커를 발현하는 환상 벡터를 형성할 수 있는 숙주 세포를 선별함으로써 선별될 수

있는 것으로 생각된다. 여기서, 이러한 특징은 숙주 세포 계놈 내로 HR에 의해 하나 이상의 외인성 핵산이 부위-특이적으로 삽입된 숙주 세포의 동정을 증강시키기 위해 개발되었다. 일부 실시형태에서, 부위-특이적 삽입은 숙주 세포 계놈을, 의도된 삽입 부위에서 절단을 생성할 수 있는 부위-특이적 뉴클레아제와 접촉시킴으로써 증강된다. 따라서, 숙주 세포에

- [0016] (i) 숙주 세포 계놈의 하나 이상의 표적 부위에 대한 상동성 영역을 갖는 하나 이상의 외인성 핵산;
- [0017] (ii) 의도된 표적 부위(들)에서 절단을 선택적으로 생성할 수 있는 하나 이상의 뉴클레아제; 및
- [0018] (iii) 그 자체, 또는 숙주 세포에 도입된 하나 이상의 추가 선형 DNA 단편과 상동성 재조합하여 선별가능한 마커가 발현되는 환상의 작용성 발현 벡터를 형성할 수 있는 선형 핵산을 도입하고,
- [0019] 선별가능한 마커의 발현을 위해 선별함으로써, 이들의 각각의 표적 부위(들) 내로 하나 이상의 외인성 핵산을 삽입시킨 세포의 동시-선택이 또한 달성된다. 본원에 제공된 방법 및 조성물에 의해 제공된 목적하는 삽입을 실시한 숙주 세포를 회수하는 증가된 빈도는 조작이 곤란하거나 난치성 숙주 세포의 유전자 조작을 가능하게 하고, 다중 삽입 등과 같이 보다 고차의 조작 디자인의 효율을 개선시킨다.
- [0020] 따라서, 한 가지 측면에서, 본원에는 숙주 세포 계놈의 하나 이상의 표적 부위 내로 하나 이상의 외인성 핵산을 삽입하는 방법이 제공되고, 상기 방법은 하나 이상의 숙주 세포를 숙주 세포 계놈의 하나 이상의 표적 부위(TS)에서 상동성 재조합에 의해 재조합할 수 있는 하나 이상의 외인성 핵산(ES), 및 각각의 TS에서 절단을 생성할 수 있는 하나 이상의 뉴클레아제(N)과 접촉시키고; 상동성 재조합을 위한 적격 숙주 세포를 선별하는 것을 포함한다. 일부 실시형태에서, 상기 선별은 외인성 핵산이 상동성 재조합되는 숙주 세포를 선별하는 것을 포함한다. 일부 실시형태에서, 상동성 재조합되는 핵산은 그 자체 또는 하나 이상의 추가 선형 핵산과 상동성 재조합할 수 있는 선형 핵산이다. 일부 실시형태에서, 선형 핵산(들)은 상동성 재조합시 환상 핵산을 형성한다. 일부 실시형태에서, 상동성 재조합되는 핵산은 선별가능한 마커를 코딩한다. 일부 실시형태에서, 환상 핵산을 형성하기 위한 선형 핵산의 상동성 재조합은 선별가능한 마커에 대한 코딩 서열을 형성한다.
- [0021] 또 다른 측면에서, 본원에는 상동성 재조합에 적격한 숙주 세포를 선별하는 방법이 제공되고, 상기 방법은
- [0022] (a) 하나 이상의 숙주 세포를, 그 자체, 또는 세포 집단과 접촉된 하나 이상의 추가의 선형 핵산과 함께, 상동성 재조합할 수 있는 선형 핵산과 접촉시킴으로써 상기 상동성 재조합에 의해 선별가능한 마커에 대한 코딩 서열을 포함하는 환상 염색체의 핵산을 형성하는 단계; 및
- [0023] (b) 선별가능한 마커를 발현시키는 숙주 세포를 선별하는 단계를 포함한다.
- [0024] 또 다른 측면에서, 본원에는 외인성 핵산을 숙주 세포 계놈의 표적 부위 내로 삽입하는 방법이 제공되고, 상기 방법은
- [0025] (a) 하나 이상의 숙주 세포를
- [0026] (i) 숙주 세포 계놈의 표적 부위(TS)에서 상동성 재조합에 의해 재조합할 수 있는 외인성 핵산(ES);
- [0027] (ii) TS에서 절단을 생성할 수 있는 뉴클레아제(N); 및
- [0028] (iii) 그 자체, 또는 숙주 세포와 접촉된 하나 이상의 추가의 선형 핵산과 함께, 상동성 재조합할 수 있는 선형 핵산과 접촉시, 상기 상동성 재조합에 의해 선별가능한 마커에 대한 코딩 서열을 포함하는 환상 염색체의 핵산을 형성하는 단계; 및
- [0029] (b) 선별가능한 마커를 발현하는 숙주 세포를 선별하는 단계를 포함한다.
- [0030] 일부 실시형태에서, 선형 핵산은 서로 상동성 재조합할 수 있는 2개의 내부 상동성 영역을 포함함으로써 상기 내부 상동성 영역의 상동성 재조합에 의해 선별가능한 마커를 발현하는 환상 염색체의 핵산을 형성한다. 일부 실시형태에서, 선형 핵산이, 숙주 세포와 접촉된 추가의 선형 핵산의 상동성 영역과 재조합할 수 있는 상동성 영역을 포함함으로써 2개의 선형 핵산의 상동성 재조합에 의해 선별가능한 마커를 발현하는 환상 염색체의 핵산의 형성을 제공한다. 일부 실시형태에서, 선형 핵산은, 상기 선형 핵산으로부터 발현될 수 없는 선별가능한 마커에 대한 부분적, 단속적(interrupted) 및/또는 비-인접적(non-contiguous) 코딩 서열을 포함함으로써 상기 환상 염색체의 핵산의 형성에 의해 환상 염색체의 핵산으로부터 발현될 수 있는 선별가능한 마커의 완전한 코딩 서열의 형성을 제공한다.
- [0031] 일부 실시형태에서, 접촉된 숙주 세포(들)은 상기 선별 전에 적어도 약 12, 24, 36, 48, 72 또는 72시간 초과

기간 동안 배양된다. 일부 실시형태에서, 상기 접촉된 세포는, 선별가능한 마커를 발현하지 않는 세포의 생존에 대해 선별하는 배양 조건하에 배양된다. 일부 실시형태에서, 단계(b)의 상기 선별은 시각, 비색 또는 형광 검출 방법을 통해 선별가능한 마커의 발현을 검출하는 것을 포함한다. 일부 실시형태에서, 상기 방법은 숙주 세포를 회수하는 단계를 추가로 포함하고, 여기서 ES는 TS에서 상동성 재조합된다. 일부 실시형태에서, 상기 회수는 숙주 세포 내로 선별가능한 마커의 삽입을 필요로 하지 않는다. 일부 실시형태에서, 상기 회수는 스크리닝된 10, 9, 8, 7, 6, 5, 4, 3 또는 2개의 접촉된 숙주 세포 또는 이의 클론 집단당 적어도 약 1개의 빈도로 발생한다. 일부 실시형태에서, 상기 방법은 환상 염색체의 핵산을 선별된 숙주 세포로부터 제거하는 단계를 추가로 포함한다.

[0032] 일부 실시형태에서, 상기 방법은, 복수(n)의 외인성 핵산을 숙주 세포 게놈의 복수(n)의 표적 부위 내로 삽입하는 것을 포함하고, 여기서 n이 적어도 2이며, 단계(a)는 숙주 세포를, 상기 복수의 외인성 핵산(여기서, x는 1 내지 n의 정수이고, 각각의 정수 x에 대해, 각각의 외인성 핵산(ES)_x은 상기 숙주 세포 게놈의 상기 복수(n)의 표적 부위로부터 선택된 표적 부위(TS)_x에서 상동성 재조합에 의해 재조합할 수 있다); 및 각각의 상기 표적 부위(TS)_x에 대해, (TS)_x에서 절단을 생성할 수 있는 뉴클레아제(N)_x와 접촉시키는 것을 포함한다. 일부 실시형태에서, 단일 뉴클레아제는 각각 (TS)_x를 절단시킬 수 있다. 일부 실시형태에서, n은 3, 4, 5, 6, 7, 8, 9 또는 10이다. 일부 실시형태에서, (ES)_x는 제1 상동성 영역(HR1)_x 및 제2 상동성 영역(HR2)_x를 포함하고, (HR1)_x 및 (HR2)_x는 각각 제3 상동성 영역(HR3)_x 및 제4 상동성 영역(HR4)_x와 상동성 재조합에 의해 재조합할 수 있으며, (HR3)_x 및 (HR4)_x는 각각 (TS)_x에 존재한다. 일부 실시형태에서, (N)_x는 (TS)_x에서 단일-가닥 절단 또는 이중-가닥 절단을 생성할 수 있다. 일부 실시형태에서, (ES)_x는 목적하는 핵산(D)_x을 추가로 포함한다. 일부 실시형태에서, (D)_x는 선별가능한 마커, 프로모터, 에피토프 태그를 코딩하는 핵산 서열, 목적하는 유전자, 리포터 유전자, 및 종결 코돈을 코딩하는 핵산 서열로 이루어진 그룹으로부터 선택된다. 일부 실시형태에서, (ES)_x는 선형이다.

[0033] 일부 실시형태에서, 환상 염색체의 핵산은 뉴클레아제에 대한 코딩 서열을 추가로 포함한다. 일부 실시형태에서, 상기 뉴클레아제는 RNA-유도 DNA 엔도뉴클레아제이다. 일부 실시형태에서, 상기 RNA-유도 DNA 엔도뉴클레아제는 Cas9 엔도뉴클레아제이다. 일부 실시형태에서, 환상 염색체의 핵산은, RNA-유도 DNA 엔도뉴클레아제에 의한 TS의 부위-특이적 인식 및 절단을 가능하게 하는 crRNA 활성화 및 tracrRNA 활성을 코딩하는 서열을 추가로 포함한다. 일부 실시형태에서, crRNA 활성화 및 tracrRNA 활성화는 단일 인접(contiguous) RNA 분자로서 발현된다.

[0034] 일부 실시형태에서, 상기 뉴클레아제는 엔도뉴클레아제, 징크 핑거 뉴클레아제, TAL-작동자 DNA 결합 도메인-뉴클레아제 융합 단백질(TALEN), 트랜스포사제 및 부위-특이적 재조합효소로 이루어진 그룹으로부터 선택된다. 일부 실시형태에서, 상기 징크 핑거 뉴클레아제는, 조작된 징크 핑거 결합 도메인에 융합된 TypeIIS 제한 엔도뉴클레아제의 절단 도메인을 포함하는 융합 단백질이다. 일부 실시형태에서, TypeIIS 제한 엔도뉴클레아제는 HO 엔도뉴클레아제 및 Fok I 엔도뉴클레아제로 이루어진 그룹으로부터 선택된다. 일부 실시형태에서, 상기 징크 핑거 결합 도메인은 3, 5 또는 6개의 징크 핑거를 포함한다.

[0035] 일부 실시형태에서, 상기 엔도뉴클레아제는, LAGLIDADG(서열번호 1) 호밍 엔도뉴클레아제, HNH 호밍 엔도뉴클레아제, His-Cys box 호밍(homing) 엔도뉴클레아제, GIY-YIG(서열번호 2) 호밍 엔도뉴클레아제 및 시아노박테리아 호밍 엔도뉴클레아제로 이루어진 그룹으로부터 선택된 호밍 엔도뉴클레아제이다. 일부 실시형태에서, 상기 엔도뉴클레아제는 H-DreI, 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-CphI, F-TevI, F-TevII, I-AmaI, I-AniI, I-ChuI, I-CmoI, 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-NclIP, 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-Ssp68031, 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-MgaI, PI-MtuI, PI-MtuHIIP, PI-MtuHIIP, PI-PfuI, PI-PfuII, PI-PkoI, PI-PkoII, PI-Rma43812IP, PI-SpBetaIP, PI-SceI, PI-TfuI, PI-TfuII, PI-ThyI, PI-TliI 또는 PI-TliII로 이루어진 그룹으로부터 선택된다. 일부 실시형태에서, 상기 엔도뉴클레아제는 변형되어 내인성 게놈 서열에 특이적으로 결합하고, 상기 변형된 엔도뉴클레아제는 이의 야생형 엔도뉴클레아제

아제 인식 서열에 더 이상 결합하지 않는다. 일부 실시형태에서, 상기 변형된 엔도뉴클레아제는, LAGLIDADG(서열번호 1) 호밍 엔도뉴클레아제, HNH 호밍 엔도뉴클레아제, His-Cys box 호밍 엔도뉴클레아제, GIY-YIG(서열번호 2) 호밍 엔도뉴클레아제 및 시아노박테리아 호밍 엔도뉴클레아제로 이루어진 그룹으로부터 선택된 호밍 엔도뉴클레아제로부터 유래한다. 일부 실시형태에서, 상기 변형된 엔도뉴클레아제는 H-DreI, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAIIIP, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-TliI, I-PpoI, Pi-PspI, F-SceI, F-SceII, F-SuvI, F-CphI, F-TevI, F-TevII, I-AmaI, I-AniI, I-ChuI, I-CmoI, 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-NclIP, 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-Ssp68031, I-SthPhiIP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdeIP, I-TevI, I-TevII, I-TevIII, i-UarAP, i-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MgaI, 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 또는 PI-TliII로 이루어진 그룹으로부터 선택된 엔도뉴클레아제로부터 유래한다.

[0036] 일부 실시형태에서, 상기 숙주 세포는 원핵생물 세포이다. 일부 실시형태에서, 상기 숙주 세포는 진핵생물 세포이다. 일부 실시형태에서, 상기 숙주 세포는 균류 세포, 박테리아 세포, 식물 세포, 곤충 세포, 조류 세포, 어류 세포 및 포유동물 세포로 이루어진 그룹으로부터 선택된다. 일부 실시형태에서, 상기 숙주 세포는 설치류 세포, 영장류 세포 및 인간 세포로 이루어진 그룹으로부터 선택된 포유동물 세포이다. 일부 실시형태에서, 상기 숙주 세포는 효모 세포이다. 일부 실시형태에서, 상기 효모는 사카로마이세스 세레비시애(*Saccharomyces cerevisiae*)이다.

[0037] 또 다른 측면에서, 본원에는 숙주 세포가 제공되고, 상기 숙주 세포는, 숙주 세포 게놈의 표적 부위(TS)에서 상동성 재조합에 의해 재조합할 수 있는 외인성 핵산(ES); TS에서 절단을 생성할 수 있는 뉴클레아제(N); 및 그 자체 또는 숙주 세포 내에서 하나 이상의 추가의 선형 핵산과 함께, 상동성 재조합할 수 있는 선형 핵산(이로써, 상기 상동성 재조합에 의해 선별가능한 마커에 대한 코딩 서열을 포함하는 환상 염색체의 핵산을 형성한다)을 포함한다. 일부 실시형태에서, 상기 선형 핵산은, 서로 상동성 재조합할 수 있는 2개의 내부 상동성 영역을 포함함으로써 상기 내부 상동성 영역의 상동성 재조합에 의해 선별가능한 마커를 발현하는 환상 염색체의 핵산을 형성한다. 일부 실시형태에서, 상기 선형 핵산은, 숙주 세포 내에서 추가의 선형 핵산의 상동성 영역과 재조합할 수 있는 상동성 영역을 포함함으로써 상기 2개의 선형 핵산의 상동성 재조합에 의해 선별가능한 마커를 발현하는 환상 염색체의 핵산의 형성을 제공한다. 일부 실시형태에서, 상기 선형 핵산은, 상기 선형 핵산으로부터 발현될 수 없는 선별가능한 마커에 대한 부분적, 단속적 및/또는 비-인접적 코딩 서열을 포함함으로써 상기 환상 염색체의 핵산의 상기 형성에 의해 환상 염색체의 핵산으로부터 발현될 수 있는 선별가능한 마커의 완전한 코딩 서열의 형성을 제공한다.

[0038] 또 다른 측면에서, 본원에는, 부위-특이적 뉴클레아제, 또는 부위-특이적 뉴클레아제에 대한 코딩 서열을 포함하는 핵산; 및 숙주 세포에서 서로 상동성 재조합할 수 있는 2개의 내부 상동성 영역을 포함하는 선형 핵산(여기서, 상기 내부 상동성 영역의 상동성 재조합은 선별가능한 마커에 대한 코딩 서열을 포함하는 환상 핵산의 형성을 제공한다)을 포함하는, 조성물이 제공된다. 일부 실시형태에서, 상기 선형 핵산은, 숙주 세포에서 상기 선형 핵산으로부터 발현될 수 없는 선별가능한 마커에 대한 부분적, 단속적 및/또는 비-인접적 코딩 서열을 포함하고, 상기 환상 핵산의 상기 형성은, 숙주 세포에서 환상 핵산으로부터 발현될 수 있는 선별가능한 마커의 완전한 코딩 서열의 형성을 제공한다.

[0039] 또 다른 측면에서, 본원에는, 부위-특이적 뉴클레아제, 또는 부위-특이적 뉴클레아제에 대한 코딩 서열을 포함하는 핵산; 및 제1 선형 핵산 및 하나 이상의 추가 선형 핵산(여기서, 상기 제1 및 제2 선형 핵산은 숙주 세포에서 서로 상동성 재조합할 수 있고, 상기 상동성 재조합은 선별가능한 마커에 대한 코딩 서열을 포함하는 환상 핵산의 형성을 제공한다)을 포함하는, 조성물이 제공된다. 일부 실시형태에서, 각각의 선형 핵산은, 숙주 세포에서 각각의 선형 핵산으로부터 발현될 수 없는 선별가능한 마커에 대한 부분적, 단속적 및/또는 비-인접적 코딩 서열을 포함함으로써 상기 환상 핵산의 형성에 의해 숙주 세포에서 환상 핵산으로부터 발현될 수 있는 선별가능한 마커의 완전한 코딩 서열의 형성을 제공한다. 일부 실시형태에서, 상기 환상 핵산은 부위-특이적 뉴클레아제에 대한 코딩 서열을 추가로 포함한다. 일부 실시형태에서, 상기 부위-특이적 뉴클레아제는 RNA-유도 DNA 엔도뉴클레아제이다. 일부 실시형태에서, 상기 RNA-유도 DNA 엔도뉴클레아제는 Cas9 엔도뉴클레아제이다. 일부 실시형태에서, 상기 조성물은, crRNA 활성을 포함하는 리보핵산 및 tracrRNA 활성을 포함하는 리보핵산;

또는 crRNA 활성을 포함하는 리보핵산을 코딩하는 데옥시리보핵산 및 tracrRNA 활성을 포함하는 리보핵산을 코딩하는 데옥시리보핵산을 추가로 포함한다. 일부 실시형태에서, 상기 환상 핵산은, crRNA 활성을 포함하는 리보핵산을 코딩하는 데옥시리보핵산 및 tracrRNA 활성을 포함하는 리보핵산을 코딩하는 데옥시리보핵산을 추가로 포함한다. 일부 실시형태에서, 상기 crRNA 활성 및 tracrRNA 활성을 코딩하는 상기 데옥시리보핵산은 단일 인접 RNA 분자 상에서 상기 활성을 코딩한다. 일부 실시형태에서, 상기 부위-특이적 뉴클레아제는 엔도뉴클레아제, 징크 핑거 뉴클레아제, TAL-작동자 DNA 결합 도메인-뉴클레아제 융합 단백질(TALEN), 트랜스포사제 및 부위-특이적 재조합효소로 이루어진 그룹으로부터 선택된다. 또한, 본원에는 상술된 임의의 조성물을 포함하는 숙주 세포가 제공된다. 또한, 본원에는, 세포 배양 배지, 및 본원에 기재된 임의의 숙주 세포를 포함하는, 세포 배양 조성물이 제공된다. 일부 실시형태에서, 상기 세포 배양 조성물은 선별가능한 마커의 발현을 위해 선별하는 화합물을 추가로 포함한다.

[0040] 또 다른 측면에서, 본원에는, 제1 상동성 영역(HR1) 및 제2 상동성 영역(HR2)을 포함하는 선형 핵산(여기서, HR1 및 HR2는 상동성 재조합에 의해 서로 재조합할 수 있고, HR1과 HR2의 상동성 재조합은 선별가능한 마커에 대한 코딩 서열을 포함하는 환상 핵산의 형성을 제공한다)이 제공된다. 일부 실시형태에서, HR1은 선별가능한 마커의 제1 불완전 코딩 서열을 포함하고, HR2가 선별가능한 마커의 제2 불완전 코딩 서열을 포함하고, HR1과 HR2의 상동성 재조합이 선별가능한 마커에 대한 완전 코딩 서열의 재구성을 제공한다. 일부 실시형태에서, 상기 선형 핵산은 본원에 기재된 부위-특이적 뉴클레아제에 대한 코딩 서열을 추가로 포함한다.

[0041] 또한, 본원에는 부위-특이적 RNA 유도된 엔도뉴클레아제(RGEN), 예를 들면, CRISPR/Cas9에 의해 매개된 숙주 세포 게놈 내로 하나 이상의 공여체 DNA의 게놈 삽입을 위한 방법 및 조성물이 제공된다. 한 가지 측면에서, 본원에는 하나 이상의 외인성 핵산을 숙주 세포 게놈의 하나 이상의 표적 부위에 삽입하는 방법이 제공되고, 상기 방법은

[0042] (a) 하나 이상의 숙주 세포를,

[0043] (i) 숙주 세포 게놈의 하나 이상의 표적 부위(TS)에서 상동성 재조합에 의해 재조합할 수 있는 하나 이상의 외인성 공여체 핵산(ES);

[0044] (ii) RNA-유도 엔도뉴클레아제(RGEN); (iii) RGEN에 의한 하나 이상의 TS의 부위-특이적 인식 및 절단을 가능하게 하는 하나 이상의 리보핵산; 및

[0045] (iv) 그 자체, 또는 숙주 세포와 접촉된 하나 이상의 추가의 선형 예비-재조합 핵산과 함께, 상동성 재조합함으로써 상기 상동성 재조합에 의해 선별가능한 마커에 대한 코딩 서열을 포함하는 환상 염색체의 핵산을 형성하는, 선형 예비-재조합 핵산

[0046] 과 접촉시키는 단계; 및

[0047] (b) 선별가능한 마커를 발현시키는 숙주 세포를 선별하고, 이에 의해 하나 이상의 외인성 핵산이 숙주 세포 게놈의 하나 이상의 표적 부위에 삽입된 세포를 선별하는 단계를 포함한다.

[0048] 일부 실시형태에서, 상기 상동성 재조합에 의해 상기 환상 염색체의 핵산 내에서 선별가능한 마커의 완전한 코딩 서열의 형성을 제공한다. 일부 실시형태에서, 적어도 하나의 선형 예비-재조합 핵산은, RNA-유도 DNA 엔도뉴클레아제에 의한 TS의 부위-특이적 인식 및 절단을 가능하게 하는 하나 이상의 리보핵산을 코딩하는 서열을 포함한다. 일부 실시형태에서, 상기 하나 이상의 리보핵산은 단일 인접 유도 RNA(gRNA) 분자 상에 crRNA 활성 및 tracrRNA 활성을 포함한다. 일부 실시형태에서, 적어도 하나의 선형 예비-재조합 핵산은, RNA-유도 DNA 엔도뉴클레아제를 코딩하는 서열을 포함한다. 일부 실시형태에서, 상기 RNA-유도 DNA 엔도뉴클레아제는 Cas9이다.

[0049] 일부 실시형태에서, 상기 환상 염색체의 핵산의 형성은 2 또는 3개의 선형 예비-재조합 핵산의 상동성 재조합으로부터 발생한다. 일부 실시형태에서, 하나 이상의 선형 예비-재조합 핵산은, 하나 이상의 예비-재조합 핵산을 포함하는 하나 이상의 환상 핵산의 RGEN 절단에 의해 생체내에서 생성된다. 일부 실시형태에서, 복수(n)의 외인성 핵산은 숙주 세포 게놈의 복수(n)의 표적 부위 내로 삽입되고, 여기서 n은 적어도 2이고, 단계(a)는 숙주 세포를

[0050] (i) 상기 복수의 외인성 핵산(여기서, x는 1 내지 n의 정수이고, 각각의 정수 x에 대해, 각각의 외인성 핵산(ES)_x은 상기 숙주 세포 게놈의 상기 복수(n)의 표적 부위로부터 선택된 표적 부위(TS)_x에서 상동성 재조합에 의해 재조합할 수 있다);

- [0051] (ii) 각각의 상기 표적 부위(TS)_x에 대해, RGEN에 의한 (TS)_x의 부위-특이적 인식 및 절단을 가능하게 하는 유도 RNA(gRNA)_x
- [0052] 와 접촉시키는 것을 포함한다.
- [0053] 일부 실시형태에서, 상기 선별가능한 마커는 약물 내성 마커, 형광 단백질 또는 비색 또는 형광 검출 방법에 의해 검출가능한 단백질이다. 일부 실시형태에서, ES는 목적하는 핵산 D를 추가로 포함한다. 일부 실시형태에서, D는 선별가능한 마커, 프로모터, 에피토프 태그를 코딩하는 핵산 서열, 목적하는 유전자, 리포터 유전자, 및 종결 코돈을 코딩하는 핵산 서열로 이루어진 그룹으로부터 선택된다. 일부 실시형태에서, 상기 숙주 세포는 진균 세포, 박테리아 세포, 식물 세포, 곤충 세포, 조류 세포, 어류 세포 및 포유동물 세포로 이루어진 그룹으로부터 선택된다. 일부 실시형태에서, 접촉된 숙주 세포(들)은 상기 선별 전에 적어도 약 12, 24, 36, 48, 72시간 또는 72시간 초과 기간 동안 배양된다. 일부 실시형태에서, 상기 접촉된 세포는, 선별가능한 마커를 발현하지 않는 세포의 생존에 대해 선별하는 배양 조건하에 배양된다. 일부 실시형태에서, 단계(b)의 상기 선별은 시각, 비색 또는 형광 검출 방법을 통해 선택가능한 마커의 발현을 검출하는 것을 포함한다.
- [0054] 또 다른 측면에서, 본원에는 하나 이상의 외인성 핵산을 숙주 세포 계놈의 하나 이상의 표적 부위 내로 삽입시키기 위한 조성물이 제공되고, 상기 조성물은
- [0055] (a) 숙주 세포 계놈의 하나 이상의 표적 부위(TS)에서 상동성 재조합에 의해 재조합할 수 있는 하나 이상의 외인성 공여체 핵산(ES);
- [0056] (b) RNA-유도 엔도뉴클레아제(RGEN), 또는 상기 RGEN을 코딩하는 핵산;
- [0057] (c) RGEN에 의한 하나 이상의 TS의 부위-특이적 인식 및 절단을 가능하게 하는 하나 이상의 리보핵산, 또는 상기 하나 이상의 리보핵산을 코딩하는 하나 이상의 핵산; 및
- [0058] (d) 그 자체, 또는 조성물 중의 하나 이상의 추가의 선형 예비-재조합 핵산과 함께, 생체내에서 상동성 재조합할 수 있음으로써 상기 생체내 상동성 재조합에 의해 선별가능한 마커에 대한 코딩 서열을 포함하는 환상 염색체의 핵산을 형성하는, 선형 예비-재조합 핵산을 포함한다.
- [0059] 일부 실시형태에서, 상기 상동성 재조합에 의해 환상 염색체의 핵산 내에서 선별가능한 마커의 완전한 코딩 서열의 형성을 제공한다. 일부 실시형태에서, 적어도 하나의 선형 예비-재조합 핵산은, RNA-유도 DNA 엔도뉴클레아제에 의한 TS의 부위-특이적 인식 및 절단을 가능하게 하는 하나 이상의 리보핵산을 코딩하는 서열을 포함한다. 일부 실시형태에서, 하나 이상의 리보핵산 분자는 단일 인접 유도 RNA(gRNA) 분자 상에 crRNA 활성화 및 tracrRNA 활성을 포함한다. 일부 실시형태에서, 적어도 하나의 선형 예비-재조합 핵산은, RNA-유도 DNA 엔도뉴클레아제를 코딩하는 서열을 포함한다. 일부 실시형태에서, 상기 RNA-유도 DNA 엔도뉴클레아제는 Cas9이다. 일부 실시형태에서, 상기 조성물은, 상동성 재조합하여 환상 염색체의 핵산을 형성할 수 있는 2 또는 3개의 선형 예비-재조합 핵산을 포함한다. 일부 실시형태에서, 상기 하나 이상의 선형 예비-재조합 핵산은, 하나 이상의 예비-재조합 핵산을 포함하는 하나 이상의 환상 핵산의 RGEN 절단에 의해 생체내에서 생성된다.
- [0060] 일부 실시형태에서, 상기 조성물은
- [0061] (a) 숙주 세포 계놈의 복수(n)의 표적 부위 내로 삽입될 수 있는 복수(n)의 외인성 핵산(여기서, n은 적어도 2이고, x는 1 내지 n의 정수이고, 각각의 정수 x에 대해, 각각의 외인성 핵산(ES)_x은 상기 숙주 세포 계놈의 상기 복수(n)의 표적 부위로부터 선택된 표적 부위(TS)_x에서 상동성 재조합에 의해 재조합할 수 있다); 및
- [0062] (b) 각각의 상기 표적 부위(TS)_x에 대해, RGEN에 의한 (TS)_x의 부위-특이적 인식 및 절단을 가능하게 하는 유도 RNA(gRNA)_x
- [0063] 를 포함한다.
- [0064] 일부 실시형태에서, 상기 선별가능한 마커는 약물 내성 마커, 형광 단백질, 또는 비색 또는 형광 검출 방법에 의해 검출할 수 있는 단백질이다. 일부 실시형태에서, ES는 목적하는 핵산 D를 추가로 포함한다. 일부 실시형태에서, D는 선별가능한 마커, 프로모터, 에피토프 태그를 코딩하는 핵산 서열, 목적하는 유전자, 리포터 유전자, 및 종결 코돈을 코딩하는 핵산 서열로 이루어진 그룹으로부터 선택된다.
- [0065] 또 다른 측면에서, 본원에는 본원에 기재된 숙주 세포 계놈의 하나 이상의 표적 부위 내로 하나 이상의 외인성

핵산의 RGEN-매개된 삽입을 위한 임의의 조성물을 포함하는 숙주 세포가 제공된다. 일부 실시형태에서, 상기 숙주 세포는 진균 세포, 박테리아 세포, 식물 세포, 곤충 세포, 조류 세포, 어류 세포 및 포유동물 세포로 이루어진 그룹으로부터 선택된다.

도면의 간단한 설명

4. 도면의 간단한 설명

도 1은 선별가능한 마커(GFP)에 대한 코딩 서열을 포함하는 환상 플라스미드를 형성하기 위해 부위-특이적 뉴클레아제(N) 및 서로 상동성 재조합(HR)할 수 있는 2개의 예비-재조합 분자를 사용하여 외인성 핵산(D)의 숙주 세포 중의 게놈 삽입에 대한 예시적 양태를 제공한다. 이 예에서, 녹색 형광 단백질(GFP)을 위한 코딩 서열은 2개의 예비-재조합 분자 사이에서 분할되고, 코딩 서열은 2개 분자 사이의 중첩 상동성 영역의 HR 사에서 생체내 재구성된다. 선별가능한 마커의 발현에 대한 선별은 또한 이의 표적 부위 내로 외인성 핵산을 삽입시키 세포에 대해 선별하고, 선별 후, 선별가능한 마커를 포함하는 플라스미드를 제거할 수 있다. HR1 - 상류 상동성 영역; HR2 - 하류 상동성 영역; TS - 표적 부위; N - 부위-특이적 뉴클레아제; D - 목적하는 핵산.

도 2는 선별가능한 마커(GFP)에 대한 코딩 서열을 포함하는 환상 플라스미드를 형성하기 위해 부위-특이적 뉴클레아제 및 서로 상동성 재조합할 수 있는 2개의 예비-재조합 분자를 사용하여 복수의 외인성 핵산의 숙주 세포 중의 동시 게놈 삽입에 대한 예시적 양태를 제공한다. 이 예에서, 2개의 예비-재조합 분자는 3개의 외인성 공여체 DNA와 함께 도입되고, 각각은 숙주 세포 게놈에서 고유 표적 부위에 특이적인 상동성 영역, 및 3개의 표적 부위에서 절단할 수 있는 하나 이상의 뉴클레아제를 갖는다. 선별가능한 마커의 발현에 대한 선별은 또한 이의 각각의 표적 부위 내로 각 외인성 핵산을 삽입시킨 세포에 대해 선별한다. HR1 - 상류 상동성 영역; HR2 - 하류 상동성 영역; TS - 표적 부위; N -부위 특이적 뉴클레아제; D - 목적하는 핵산.

도 3은 예비-재조합 분자의 HR-매개된 조립 및 외인성 공여체 DNA의 표적화된 게놈 삽입을 가능하게 하는 세포를 선별하는 2개의 예시적 실시형태를 제공한다. **도 3A**는 형광-기반 선별가능한 마커를 포함하는 플라스미드의 HR-매개된 형성에 기반하는 선별 전략을 도시한다. 숙주 세포는 하나 이상의 외인성 공여체 DNA로 형질전환되고, 이는, 생체내 HR-매개된 조립시, 형광-기반 선별가능한 마커(예: GFP), 및 임의로 숙주 세포 게놈의 하나 이상의 표적 부위를 절단할 수 있는 하나 이상의 부위-특이적 뉴클레아제를 포함하는 환상 플라스미드를 형성한다. HR 적격 세포는 형광에 의해 표지되고, 유동 세포분석 등의 표준 기술을 사용하여 숙주 세포 집단으로부터 단리될 수 있다. **도 3B**는 약물-기반 선별가능한 마커의 HR-매개된 형성을 기반으로 하는 선별 전략을 도시한다. 이 실시형태에서, HR-적격 세포는, 적절한 선별제를 함유하는 배지에서 배양하는 경우에 약물 내성 및 생존물에 의해 표지되고, 비-HR 적격 약물-감수성 세포는 제거된다. 어느 하나의 선별 방식하에 단리된 세포 또는 클론 세포 집단을 확장시킬 수 있고, 예를 들면, 게놈 표적 영역의 PCR 및/또는 서열분석에 의해 하나 이상의 외인성 공여체 DNA의 표적화된 삽입을 함유하는 것을 확인할 수 있다.

도 4는 본원에 제공된 게놈 삽입 방법에 유용한 예시적 예비-재조합 조성물을 제공한다. 본원에 제공된 임의의 예비-재조합 조성물에 있어서, 상기 조성물은 선형 핵산 분자로서 숙주 세포 내로 직접 형질전환될 수 있거나, 달리, 예비-재조합 분자를 포함하는 모 (parent) 환상 분자는 숙주 세포 내로 도입되고 하나 이상의 뉴클레아제에 의해 생체내에서 절단되어 예비-재조합 분자를 방출할 수 있다. HR-적격 숙주 세포에서, 선형 예비-재조합 분자(들)은 상동성 재조합하여, 선별가능한 마커를 포함하는 환상 벡터를 형성한다. **도 4A**는 마커 플라스미드의 1-피스 생체내 조립을 위한 예비-재조합 분자에 대한 2개의 예시적 실시형태를 도시한다. (L) 단일 선형 예비-재조합 분자는, 예를 들면, 선별가능한 마커의 무손상 코딩 서열(GFP에 의해 제시됨)의 외측, 즉 이를 포함하지 않는 2개의 중첩 상동성 영역(수직 스트립 박스에 의해 제시됨)을 포함할 수 있다. (C) 또는, 단일 선형 예비-재조합 분자는 선별가능한 마커의 부분 코딩 서열을 각각 포함하는 2개의 중첩 상동성 영역(각각 GF 및 FP; 중첩은 회색으로 음영)을 포함할 수 있다. (R) 두 실시형태에 있어서, 그 자체와 단일 선형 예비-재조합 분자의 생체내 상동성 재조합은 선별가능한 마커의 완전 코딩 서열을 포함하는 환상 플라스미드의 형성을 제공한다. 일부 실시형태에서, 단일 선형 예비-재조합 분자는 부위-특이적 뉴클레아제에 대한 코딩 서열을 추가로 포함할 수 있다(도시되지 않음).

도 4B는 마커 플라스미드의 2-피스 생체내 조립을 위한 예비-재조합 조성물에 대한 2개의 예시적 실시형태를 도시한다. 2개의 선형 예비-재조합 분자는 각각 2개의 비-중첩 상동성 영역(각각 수직 및 수평 스트립 박스로 제시됨)을 포함할 수 있고, 각각의 상동성 영역은 다른 예비-재조합 분자의 상동성 영역과 상동성이다. (L) 2개의 선형 예비-재조합 분자 중의 하나는 2개의 비-중첩 상동성 영역으로부터 분리된 선별가능한 마커의 무손상 코딩 서열(GFP에 의해 제시됨)을 포함할 수 있다. (C) 또는, 2개의 예비-재조합 분자 중의 하나는 다른 예비-

재조합 분자 상의 부분 마커 코딩 서열과 상동성을 갖는 선별가능한 마커의 부분 코딩 서열을 포함할 수 있다 (각각 GF 및 FP, 중첩은 회색으로 음영). (R) 두 실시형태에 있어서, 서로 2개의 선형 예비-재조합 분자의 생체 내 상동성 재조합은 선별가능한 마커의 완전 코딩 서열을 포함하는 환상 플라스미드의 형성을 제공한다. 일부 실시형태에서, 선형 예비-재조합 분자 중의 하나는 부위-특이적 뉴클레아제에 대한 완전 코딩 서열을 추가로 포함할 수 있거나, 달리, 2개 예비-재조합 분자 각각은 다른 예비-재조합 분자 상의 부분 뉴클레아제 코딩 서열과 상동성을 갖는 부분 뉴클레아제 코딩 서열을 포함할 수 있다(제시되지 않음). 이러한 실시형태는 뉴클레아제 발현이, 예를 들면, 비-HR-적격 세포에서 뉴클레아제-매개된 NHEJ의 발생을 감소시키기 위해 HR-적격 세포에서만 요구되는 경우에 유용할 수 있다.

도 4C는 마커 플라스미드의 3-피스 생체내 조립을 위한 예비-재조합 부자 조성물에 대한 2개의 예시적 실시형태를 도시한다. 3개의 선형 예비-재조합 분자는 각각 2개의 비-중첩 상동성 영역(각각 수직 및 수평 스트립 박스; 수직 스트립 박스 및 다이아몬드-충전 박스; 및 수평 스트립 박스 및 다이아몬드-충전 박스로 제시됨)을 포함할 수 있고, 각각의 상동성 영역은 다른 예비-재조합 분자 중의 하나 상의 상동성 영역과 상동성이다. (L) 3개의 선형 예비-재조합 분자 중의 하나는 2개의 비-중첩 상동성 영역으로부터 분리된 선별가능한 마커의 무손상 코딩 서열(GFP로 제시됨)을 포함할 수 있다. (C) 또는, 적어도 2개의 예비-재조합 분자 각각은 하나의 다른 예비-재조합 분자 상의 부분 마커 코딩 서열과 상동성을 갖는 선별가능한 마커의 부분 코딩 서열을 포함할 수 있다(각각 GF 및 FP). (R) 두 실시형태에 있어서, 서로 3개 선형 예비-재조합 분자의 상동성 재조합은 선별가능한 마커의 완전 코딩 서열을 포함하는 환상 플라스미드의 형성을 제공한다. 일부 실시형태에서, 2-피스 조립에서와 같이, 선형 예비-재조합 분자 중의 하나는 부위-특이적 뉴클레아제에 대한 완전 코딩 서열을 추가로 포함할 수 있거나, 달리, 적어도 2개 예비-재조합 분자 각각은 하나의 다른 예비-재조합 분자 상의 부분 뉴클레아제 코딩 서열과 상동성을 갖는 부분 뉴클레아제 코딩 서열을 포함할 수 있다(제시되지 않음).

도 5는 본원에 제공된 게놈 삽입 방법에 대한 RNA-유도 DNA 엔도뉴클레아제(RGEN) 특이적 실시형태에 유용한 예시적 예비-재조합 조성물을 제공한다. **도 5A: 도 4A**에 제시된 1-피스 생체내 조립의 일부 실시형태에서, 단일 예비-재조합 q분자는 RGEN(예: CRISPR/Cas9)에 의해 게놈 표적 부위의 부위-특이적 인식 및 절단을 가능하게 하는 crRNA 활성 및 tracrRNA 활성(예: 유도 RNA(gRNA) 서열)을 코딩하는 하나 이상의 서열을 추가로 포함할 수 있다. 일부 실시형태에서, 예비-재조합 분자는 RGEN(예: Cas9; 제시되지 않음)을 위한 코딩 서열을 추가로 포함할 수 있다. **도 5B: 도 4B**에 도시된 2-피스 생체내 조립의 일부 실시형태에서, 2개 예비-재조합 분자 중의 하나는 gRNA 서열을 추가로 포함할 수 있다. 다른 실시형태에서, 2개 예비-재조합 분자 중의 하나는 RGEN의 완전 코딩 서열을 추가로 포함할 수 있거나, 달리, 2개 예비-재조합 분자 중의 하나는 다른 예비-재조합 분자 상의 부분 뉴클레아제 코딩 서열과 상동성을 갖는 부분 뉴클레아제 코딩 서열을 포함할 수 있다(제시하지 않음). **도 5C:** 3-피스 생체내 조립의 일부 실시형태에서, 3개 예비-재조합 분자 중의 하나는 gRNA 서열을 추가로 포함할 수 있다. 다른 실시형태에서, 3개 예비-재조합 분자 중의 하나는 RGEN의 완전 코딩 서열을 추가로 포함할 수 있거나, 달리, 적어도 2개 예비-재조합 분자 각각은 하나의 다른 예비-재조합 분자 상의 부분 뉴클레아제 코딩 서열과 상동성을 갖는 부분 뉴클레아제 코딩 서열을 포함할 수 있다(제시하지 않음).

도 6은 RGEN-매개된 다중 게놈 삽입에 유용한 예시적 선형 예비-재조합 조성물을 제공한다. **도 6A:** 조립에 관여하는 2개 예비-재조합 분자 중의 하나가 gRNA 서열을 포함하는 2-피스 생체내 조립의 일부 실시형태에서, 이들 분자 중의 몇몇은 한번에 제공될 수 있고(예를 들면, 3: gRNA-1, gRNA-2, gRNA-3), 각각은 상이한 게놈 표적 부위를 표적화하는 고유 gRNA 서열을 포함한다. 이러한 실시형태에서, 다른 예비-재조합 분자는 선별가능한 마커(L)에 대한 완전 코딩 서열, 또는 각각의 gRNA 함유 단편(C)에 통상적인 부분 코딩 서열과 상동성인 부분 마커 코딩 서열을 나타낸다. (R) HR-적격 세포는 고유 gRNA 서열을 각각 포함하는 3개의 상이한 마커 플라스미드를 재구성하기 위해 각각의 고유 gRNA 함유 단편을 통상의 벡터 골격과 재조합시킬 수 있다. **도 6B:** 다른 실시형태에서, 2개 선형 예비-재조합 분자 중의 하나는 RGEN(예: Cas9)의 완전 코딩 서열을 추가로 포함할 수 있다. 달리, 2개 예비-재조합 분자 각각은 다른 예비-재조합 분자 상의 부분 뉴클레아제 코딩 서열과 상동성을 갖는 부분 뉴클레아제 코딩 서열을 포함할 수 있다(제시하지 않음). 다중 게놈 삽입은 조성물의 하나 이상의 예비-재조합 분자 내에 위치된 복수의 고유 gRNA 카세트와 조합하여 2-피스, 3-피스 또는 보다 고차의 예비-재조합 조성물로 수행할 수 있다.

도 7은 실시예 1에 기재된 바와 같은 CRISPR/Cas-9 매개된 다중 공여체 DNA 삽입을 위한 gRNA 전달의 최적 모드를 결정하는데 사용된 조성물을 도시한다. (L) 고유 gRNA 카세트는 달모양(crescents)으로 도시되고, 고유 약물 선별가능한 마커는 직사각형으로 도시된다. (R) gRNA 카세트를 다음으로서 숙주 세포에 도입했다: (1) 3개 고유 gRNA 카세트 각각이 고유 선별가능한 마커를 포함하는 플라스미드 내로 클로닝된 환상 벡터; (2) 3개 고유

gRNA 카세트 각각이 동일한 선별가능한 마커를 포함하는 플라스미드 내로 클로닝된 환상 벡터; (3) 3개 선형 gRNA 카세트가 선별가능한 마커를 포함하는 환상 플라스미드로 동시-형질전환된 선형 발현 카세트; 및 (4) 선별가능한 마커를 포함하는 동시-형질전환된 선형 플라스미드의 말단과 상동성이고 따라서 각각의 gRNA 및 통상의 선별가능한 마커를 포함하는 환상 플라스미드의 HR-매개된 생체내 조립을 가능하게 하는 말단을 각각 갖는 선형 발현 카세트.

도 8은 실시예 1에 기재된 gRNA 전달의 최적 모드를 결정하기 위한 실험 결과를 제공한다. Cas9-발현 숙주 세포(*S. cerevisiae*)는 RHR2, HO 및 ADH5 개방 판독 프레임의 동시 마커-레스 삽입/결실을 위한 공여체 DNA, 및 각각의 유전자좌를 표적화하는 gRNA 작제물로 형질전환시켰다. gRNA 전달 모드는 1) 3개의 상이한 선별가능한 마커를 갖는 3개 플라스미드, 2) 동일한 마커를 갖는 3개 플라스미드, 3) 3개 선형 gRNA 카세트를 갖는 단일 마커 플라스미드, 및 4) 3개 선형 gRNA 카세트의 겹 수복을 위한 인접 서열을 갖는 단일 선형화 마커 플라스미드였다. 콜로니는 결실 작제물 외부의 상류 전방 프라이머, 및 각 개방 판독 프레임 위치에 삽입된 짧은 링커 서열에 결합하는 역방향 프라이머를 사용하여 cPCR에 의해 검정했다. 11개 콜로니는 네가티브 대조군("N")으로 사용되는 모 콜로니 뿐만 아니라 각 전달 모드에 대해 검정했다.

도 9는, 각각 RHR2, HO 및 ADH5 유전자좌 내로 공여체 DNA의 CRISPR/Cas-9 매개된 단일 삽입에 대하여, gRNA 발현에 대해 선별하는 잇점과 분리된, 마커 벡터의 겹 수복의 잇점을 결정하기 위한 실험 결과(실시예 2에 기재됨)를 제공한다. Cas9-발현 숙주 세포(*S. cerevisiae*)는 RHR2, HO 또는 ADH5 개방 판독 프레임의 마커-레스 결실을 위한 공여체 DNA 및 각 유전자좌를 표적화하는 gRNA 작제물로 형질전환시켰다. 적절한 공여체 DNA에 추가하여, 선형 gRNA 카세트는 1) 밀폐된 (closed) 마커 벡터(A) 또는 2) 추가 공급된 단편의 겹 수복이 벡터를 밀폐하고 마커 카세트를 재구성하는데 요구되도록 선형화 및 절단된 동일한 벡터(B)로 동시-형질전환시켰다. 콜로니는 결실 작제물 외부의 상류 전방 프라이머, 및 각 개방 판독 프레임의 위치에 삽입된 짧은 링커 서열에 결합하는 역방향 프라이머를 사용하여 PCR에 의해 검정했다. 23개 콜로니는, 네가티브 대조군("N")으로 사용되는 모 콜로니 뿐만 아니라 각 전달 모드에 대해 검정했다. 실험은 3회 반복했고; 단일 실험으로부터의 결과가 제시되어 있다.

도 10은 **도 9**에서 입증된 3개 겹 수복 실험 결과의 요약을 제공한다. 3개 실험 각각에 있어서, 23개 콜로니를 검정했다.

도 11은 **도 9**에서 입증된 3개 겹 수복 실험의 박스 플롯 요약을 제공한다.

도 12는 마커 플라스미드의 겹-수복 조립을 통해 배수 증가; **도 9**에서 입증된 3개 겹 수복 실험의 요약을 제공한다.

도 13은, 각각 Gal80, HO 및 ADH5 유전자좌 내로 3개 공여체 DNA의 CRISPR/Cas-9 매개된 동시 삽입에 대해 마커/gRNA 벡터의 2-피스 생체내 조립 대 3-피스 생체내 조립의 잇점을 결정하기 위한 실험 결과를 제공한다. Cas9-발현 숙주 세포(반수체 *S. cerevisiae*)는, Gal80, HO 및 ADH5 개방 판독 프레임의 동시 마커-레스 결실을 위한 공여체 DNA, 각 유전자좌를 표적화하는 gRNA 작제물, 및 2 또는 3 피스 마커/gRNA 벡터 조립을 위한 예비-재조합 분자로 형질전환시켰다. 콜로니는 결실 작제물 외부의 상류 전방 프라이머, 및 각 개방 판독 프레임의 위치에 삽입된 짧은 링커 서열에 결합하는 역방향 프라이머를 사용하여 cPCR에 의해 검정했다. 11개 콜로니는 네가티브 대조군("N")으로 사용되는 모 콜로니 뿐만 아니라 각 전달 모드에 대해 검정했다.

도 14는, 각각 Gal80, HO 및 ADH5 유전자좌 내로 3개 공여체 DNA의 CRISPR/Cas-9 매개된 동시 삽입에 대해 마커/gRNA 벡터의 2-피스 생체내 조립 대 3-피스 생체내 조립의 잇점을 결정하기 위한 실험 결과를 제공한다. 이 배체 효모 균주 CAT-1(*S. cerevisiae*)의 Cas9-발현 세포는, Gal80, HO 및 ADH5 개방 판독 프레임의 동시 범-대립인자 (pan-allelic) 마커-레스 결실을 위한 공여체 DNA, 각 유전자좌를 표적화하는 gRNA 작제물, 및 2 또는 3 피스 마커/gRNA 벡터 조립을 위한 예비-재조합 분자로 형질전환시켰다. 콜로니는 결실 작제물 외부의 상류 전방 프라이머, 및 각 개방 판독 프레임의 위치에 삽입된 짧은 링커 서열에 결합하는 역방향 프라이머를 사용하여 PCR에 의해 검정했다. 실험은 세포가 선별 플라스미드를 생성하기 위해 2 또는 3 상동성 재조합 사상을 사용하여 형질전환된 DNA 시약을 처리해야 하는 선별 방식을 사용했다. 동시 범-대립인자 삼중 삽입의 비율은 3개 사상이 요구되는 경우에 거의 10배 더 높았다. 실험으로부터 회수된 콜로니의 수는 또한 3개 사상이 요구되는 경우에 거의 10배 더 낮았고(도시되지 않음), 이는 선별 방식이 삼중 삽입의 증가율에 관여하였음을 나타낸다.

도 15는 "이종성 블록"의 문맥에서 점 돌연변이의 도입을 위한 개략을 제공한다. 표적화된 아미노산을 박스로 되어 있고, 인접한 절단 부위는 절단 부위 및 PAM 서열(상부 패널)로 주석되어 있다. 침묵 코돈 변화의 이종성

블록 및 인접한 상동성의 문맥에서 목적하는 점 돌연변이를 함유하는 공여체 DNA는 60-머 올리고(중간 패널)를 어닐링 및 확장시키거나 보다 큰 클론 작제물로 합성에 의해 생성할 수 있다. 공여체 DNA의 삽입은 목적하는 점 돌연변이(하부 패널)를 제공한다.

도 16은 마커/gRNA 벡터의 2-피스 생체내 조립과 병용하여 CRISPR을 사용하여 공여체 DNA에 코딩된 단일 점 돌연변이를 도입하기 위한 실험 결과를 제공한다. 후보 콜로니(1-11) 및 모 네가티브 대조군(c)은 이중성 블록 및 인접 서열(좌측 패널, 및 표)에 대해 콜로니 PCR에 의해 검정했다. 선택된 포지티브 콜로니는 삽입 유전자 좌에 걸쳐 보다 큰 PCR 생성물을 서열분석함으로써 확인했다.

도 17은 마커/gRNA 벡터의 2-피스 생체내 조립과 병용하여 CRISPR을 사용하여 공여체 DNA에서 코딩된 점 돌연변이를 다중 방식으로 도입하기 위한 실험 결과를 제공한다. ECM38, PGD1 및 ADH2 유전자좌는 3개 점 돌연변이의 동시 도입을 위해 표적화했다. 공여체 DNA는, 각 표적 부위에 인접하는 500bp의 상류 및 하류 상동성을 사용하여 클로닝했다. 후보 콜로니는 이중성 블록 및 인접 서열(좌측 패널 및 표)에 대해 콜로니 PCR에 의해 동정했다. 10/11 콜로니(90.9%)는 모든 3개 이중성 블록의 삽입에 대해 포지티브였다.

도 18은 마커/gRNA 벡터의 2-피스 생체내 조립과 병용하여 CRISPR을 사용하여 공여체 DNA에서 코딩된 점 돌연변이를 다중 방식으로 도입하기 위한 실험 결과를 제공한다. ADH2, PGD1, ECM38, SIN4 및 CYS4 유전자좌는 5개 점 돌연변이의 동시 도입을 위해 표적화시켰다. 공여체 DNA는, 각 표적 부위에 인접하는 상류 및 하류 상동성의 500bp를 사용하여 이 경우에 클로닝시켰다. 후보 콜로니는 이중성 블록 및 인접 서열(좌측 패널 및 표)에 대해 콜로니 PCR에 의해 동정했다. 2/11 콜로니(18.2%)는 모든 5개 이중성 블록의 삽입에 대해 포지티브였다 (클론 # 4 및 9).

도 19는 반수체 CENPK2(A) 중의 GAL80 유전자좌에서 짧은 링커 서열의 삽입, 및 이배체 산업용 균주 CAT-1(B) 및 이배체 산업용 균주 PE-2(C) 중의 GAL80 유전자좌의 동일한 작제물의 범-대립인자 삽입을 입증하는 실험 결과를 제공한다. 각각의 콜로니는 야생형 대립인자(작수 레인)의 존재 뿐만 아니라 짧은 링커 서열(홀수 레인)의 삽입에 대해 검정했다. 각 겔 상의 최종 2개 레인은 모(네가티브) 대조군이다.

도 20은 마커/gRNA 벡터의 2-피스 생체내 조립과 병용하여 CRISPR을 사용하여 이소프레노이드 파르네센의 생성을 위한 12-유전자 생합성 경로(초 약 30kb)를 다중 방식으로 도입하기 위한 실험 결과를 제공한다. Gal80, HO 및 BUD9 유전자좌는 파르네센 경로 성분을 위한 코딩 서열을 포함하는 3개 공여체 DNA의 동시 도입을 위해 표적화했다(공여체 1: 전사 조절인자 GAL4; 아르테미시아 안누아(*Artemisia annua*)로부터의 파르네센 신타제(2 카피); 아세틸-CoA 티올라제를 코딩하는 ERG10; 및 HMG-CoA 신타제를 코딩하는 ERG13; 공여체 2: HMG-CoA 리덕타제를 코딩하는 tHMG1(2 카피); 및 공여체 3: 메발로네이트 키나제를 코딩하는 ERG12; 포스포메발로네이트 키나제를 코딩하는 ERG8; 메발로네이트 피로포스페이트 데카복실라제를 코딩하는 ERG19; 이소펜테닐 피로포스페이트 이소머라제를 코딩하는 IDI1; 및 파르네실 피로포스페이트 신타제를 코딩하는 ERG20). 공여체 DNA는 각 표적 부위에 인접한 상류 및 하류 상동성의 500bp로 클로닝시켰다. 후보 콜로니는 내부 링커 서열 및 삽입 표적 부위에 인접하는 서열에 대해 콜로니 PCR에 의해 동정했다. 11/47 콜로니(23.4%)는 전체 경로의 삽입에 대해 포지티브였다.

도 21은 파르네센 경로를 완전히 삽입한 것으로 cPCR에 의해 동정된 11개 클론에 대해 배치 슈크로즈 플레이트 모델 검정으로 파르네센 생성을 확인하기 위한 실험 결과를 제공한다. 각각의 cPCR 포지티브 클론은 약 0.1 내지 1.5g/파르네센 L 범위의 양으로 파르네센을 생성했다.

도 22는 단일 및 다중 대립인자 교환(swap)의 높은 비율을 입증하는 대립인자 교환 cPCR의 결과를 제공한다.

도 23(A) 내지 (F)는 CRISPR 표시 상승작용 표현형을 사용하여 생성한 다중 대립인자 교환을 입증하는 실험 결과를 제공한다. **(A)** ACE2의 절단은 불완전 세포 분열 및 응집을 제공한다. **(B)** 분비 및 세포 사이클 돌연변이는 37°C에서 성장하지 않는다. **(C)** 세포 사이클 돌연변이는 비-허용 온도에서 G1에서 정지한다. **(D)** SEC3-GFP는 허용 온도(23°C)에서 배아에 정확하게 국지화되지만, 분비 돌연변이에서 승온에서 잘못 국지화된다. **(E)** 2개 대립인자는 개별적으로 내열성을 증가시키고, 함께 보다 높은 내열성 균주를 생성한다. **(F)** 몇몇 돌연변이는 에탄올 내성을 부여하지만, 모든 대립인자는 함께 추가로 증가된 에탄올 내성을 위해 상승작용한다. 모든 5개 변화는 CRISPR을 사용하여 동시에 이루어졌다.

도 24는 단일 형질전환으로 천연 효모 균주 내로 전체 류코산 생합성 경로의 삽입을 입증하는 결과를 제공한다. **(A)** 류코산 생합성 경로의 도식. **(B)** 류코산 경로는 총 28kb의 공여체 DNA의 6개 피스에 의해 3개 분리 유전자 좌 내로 도입되었다. 각 피스는 중첩 상동성(중양)을 갖는 공여체 DNA의 또 다른 피스 뿐만 아니라 표적화된

유전자좌(말단)의 상동성 상류(US) 또는 하류(DS)의 영역을 통해 게놈 내로 재조합되었다. (C) 상기 경로의 1 단계 삽입은 경로 병목의 신속한 진단을 가능하게 했다: AroY. 삽입된 경로를 사용한 균주는 약 3g/L PCA를 생성한다(하부로부터 2번째 라인). 카테콜을 공급하는 경우(하부로부터 1번째 라인), 이들 균주는 모든 이용가능한 카테콜을 뮌콘산으로 완전히 전환시킨다(하부로부터 3, 4 및 5번째 라인). (D) 뮌콘산 경로는 또한 단일 단계로 K. 락티스 중의 3개 분리 유전자좌 내로 도입했다(10kb). (E) 삽입된 경로를 갖는 K. 락티스 균주는 약 1g/L PCA를 생성하고(하부로부터 1번째 라인), 이는 에스. 세레비시아(*S. cerevisiae*)와 동일한 경로 병목을 나타낸다. 카테콜을 공급하는 경우, 이들 균주는 또한 모든 이용가능한 카테콜을 뮌콘산으로 완전히 전환시킨다(하부로부터 2번째 라인).

도 25는 CRISPR 시약 및 공여체 DNA를 사용한 형질감염 후에 293T 세포 중의 표적화된 게놈 유전자좌의 앰플리콘에 대한 RFLP 검정의 결과를 제공한다. 세포는 다음과 같이 형질감염시켰다: (2) 밀폐된 "gRNA 비함유" 플라스미드 + 선형 공여체; (3) 개방 "gRNA 비함유" 플라스미드; (4) 개방 "gRNA 비함유" 플라스미드 + CD4 캡 단편; (6) 밀폐된 gRNA 플라스미드 + 선형 공여체; (8) 개방 gRNA 플라스미드 + 전체 캡 + 선형 공여체.

발명을 실시하기 위한 구체적인 내용

[0067]

5. 실시형태의 상세한 기재

[0068]

5.1 정의

[0069]

본원에서 사용된 바와 같이, 뉴클레아제, 예를 들면, 호밍 엔도뉴클레아제, 징크-핑거 뉴클레아제 또는 TAL-작동자 뉴클레아제 또는 RNA-유도 DNA 엔도뉴클레아제(예: CRISPR/Cas9)와 관련하여 용어 "절단하다 (cleave)", "절단" 및/또는 "절단하는"은 특정 핵산에서 절단을 생성하는 작용을 지칭한다. 절단 (break)은 당해 기술분야의 숙련가에 의해 이해되는 바와 같이 블런트 말단 또는 스틱키 말단(즉, 5' 또는 3' 오버행)을 남길 수 있다. 상기 용어는 또한 단일 가닥 DNA 절단 [니 ("nick")] 및 이중 가닥 DNA 절단을 포함한다.

[0070]

본원에 사용된 바와 같이, 용어 "조작된 숙주 세포"는 유전자 조작 기술(즉, 재조합 기술)을 사용하여 모 세포를 유전적으로 변형함으로써 생성되는 숙주 세포를 지칭한다. 조작된 숙주 세포는 모 세포의 게놈에 대한 뉴클레오티드 서열의 부가, 결실 및/또는 변형을 포함할 수 있다.

[0071]

본원에서, 용어 "이종성"은 천연에서 보통 발견되지 않는 것들을 지칭한다. 용어 "이종성 뉴클레오티드 서열"은 천연에서 소정 세포에서 보통 발견되지 않는 뉴클레오티드 서열을 지칭한다. 이와 같이, 이종성 뉴클레오티드 서열은, (a) 이의 숙주 세포에 대해 외래성(즉, 세포에 대해 "외인성")이거나; (b) 숙주 세포에서 천연으로 발견되지만(즉, "내인성"), 상기 세포에서 비천연적인 양으로 존재하거나(즉, 숙주 세포에서 천연적으로 발견되는 양보다 더 많거나 더 적음); 또는 (c) 숙주 세포에서 천연적으로 발견되지만 이의 천연적인 유전자좌의 외부에 위치할 수 있다.

[0072]

본원에서 사용된 바와 같이, 용어 "상동성"은 2개 이상의 핵산 서열, 또는 2개 이상의 아미노산 서열 사이의 동일성을 지칭한다. 서열 동일성은 동일성(또는 유사성 또는 상동성) %로 측정할 수 있으며; %가 높을수록, 상기 서열은 서로 동일성이 더 높다. 핵산 또는 아미노산 서열의 호모로그 또는 오르소로그(ortholog)는, 표준 방법을 이용해 정렬하는 경우, 서열 동일성이 상대적으로 높다. 비교를 위해 서열을 정렬하는 방법은 당해 기술분야에 잘 알려져 있다. 여러 가지 프로그램 및 정렬 알고리즘은 문헌[참조: Smith & Waterman, *Adv. Appl. Math.* 2:482, 1981; Needleman & Wunsch, *J. Mol. Biol.* 48:443, 1970; Pearson & Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444, 1988; Higgins & Sharp, *Gene*, 73:237-44, 1988; Higgins & Sharp, *CABIOS* 5:151-3, 1989; Corpet *et al.*, *Nuc. Acids Res.* 16:10881-90, 1988; Huang *et al.* *Computer Appls. Biosc.* 8, 155-65, 1992; 및 Pearson *et al.*, *Meth. Mol. Bio.* 24:307-31, 1994. Altschul *et al.*, *J. Mol. Biol.* 215:403-10, 1990]에 기술되어 있으며, 서열 정렬 방법 및 상동성 계산법에 관한 상세 사항이 기술되어 있다. 문헌[참조: NCBI Basic Local Alignment Search Tool (BLAST), Altschul *et al.*, *J. Mol. Biol.* 215:403-10, 1990]은 서열 분석 프로그램인 blastp, blastn, blastx, tblastn 및 tblastx와 함께 사용하기 위해 미국 생물학 정보 센터(National Center for Biological Information) (NCBI, National Library of Medicine, Building 38A, Room 8N805, Bethesda, Md. 20894) 및 인터넷을 포함하는 다양한 공급원에서 이용가능하다. 부가적인 정보는 NCBI 웹 사이트에서 찾을 수 있다.

[0073]

본원에서 사용된 바와 같이, 용어 "마커레스(markerless)"는, 선별가능한 마커의 삽입 없이, 숙주 세포 게놈내 표적 부위로 공여체 DNA를 삽입하는 것을 지칭한다. 일부 실시형태에서, 상기 용어는 또한, 선별가능한 마커를 숙주 세포 게놈에 삽입하는 것에 좌우되는 선별 방식을 이용하지 않고서 이러한 숙주 세포를 회수하는 것을 지

칭한다. 예를 들면, 특정 실시형태에서, 에피솜(episomal) 또는 염색체의 선별 마커를 이용하여, 게놈 표적 부위를 절단할 수 있는 뉴클레아제를 코딩하는 플라스미드를 포함하는 세포를 선별할 수 있다. 이러한 용도는, 선별가능한 마커가 숙주 세포 게놈에 삽입되지 않는 한, "마커레스"로 간주될 것이다.

[0074] 본원에 사용된 바와 같이, 용어 "작동적으로 연결된"은 서열이 목적하는 기능을 코딩하도록 핵산 서열 사이의 기능적 연결을 지칭한다. 예를 들면, 목적하는 유전자, 예를 들면, 선별가능한 마커에 대한 코딩 서열은, 연결된 프로모터 및/또는 조절 영역이 코딩 서열의 발현을 기능적으로 조절하는 경우, 이의 프로모터 및/또는 조절 서열과 작동적 연결 상태에 있다. 이는 또한 이들이 동일한 연결 프로모터 및/또는 조절 영역에 의해 조절될 수 있도록 코딩 서열 사이의 연결을 지칭하고; 코딩 서열 사이의 이러한 연결은 또한 프레임 또는 동일한 코딩 프레임에서 연결되는 것으로 지칭될 수 있다. "작동적으로 연결된"은 또한 기능적, 그러나 비-코딩 서열, 예를 들면, 자가 증식 서열 또는 복제 기원의 연결을 지칭한다. 이러한 서열은, 이들이 이들의 정상 기능을 수행할 수 있는 경우, 예를 들면, 숙주 세포에서 상기 서열을 포함하는 벡터의 복제, 증식 및/또는 분리를 가능하게 하는 경우, 작동적 연결 상태에 있다.

[0075] 본원에서 사용된 바와 같이, 용어 "선별가능한 마커를 발현하는 숙주 세포를 선별하는"은 또한 형질전환된 세포의 집단으로부터 선별가능한 마커를 발현하는 숙주 세포를 농후화하는 것을 포괄한다.

[0076] 본원에서 사용된 바와 같이, 용어 "선별가능한 마커"는 본원에 기재된 바와 같은 마커 벡터를 포함하는 숙주 세포를 선별하는 지침으로서 기능하는 유전자를 지칭한다. 선별가능한 마커는, 이로써 제한되지 않지만, 형광 마커, 발광 마커 및 약물 선별가능한 마커를 포함할 수 있다. 형광 마커는, 이로써 제한되지 않지만, 형광 단백질, 예를 들면, 녹색 형광 단백질(GFP), 시안 형광 단백질(CFP), 황색 형광 단백질(YFP), 적색 형광 단백질(dsRFP) 등을 코딩하는 유전자를 포함할 수 있다. 발광 마커는, 이로써 제한되지 않지만, 루시페라제 등의 발광 단백질을 코딩하는 유전자를 포함할 수 있다. 본원에 제공된 방법 및 조성물과 사용하기에 적합한 약물 선별가능한 마커는, 이로써 제한되지 않지만, 항생물질에 대한 내성 유전자, 예를 들면, 암피실린, 스트렙토마이신, 젠타마이신, 카나마이신, 하이드로마이신, 테트라사이클린, 클로람페니콜 및 네오마이신을 포함한다. 일부 실시형태에서, 선별은 포지티브 선별일 수 있고; 즉, 마커를 발현하는 세포는 집단으로부터 단리되어, 예를 들면, 선별가능한 마커를 포함하는 농후화된 세포 집단을 생성한다. 다른 예에서, 선별은 네가티브 선별일 수 있고; 즉, 집단은 세포로부터 단리되어, 예를 들면, 선별가능한 마커를 포함하지 않는 농후화된 세포 집단을 생성한다. 분리는 사용된 선별가능한 마커에 적절한 임의의 편리한 분리 기술에 의해 이루어질 수 있다. 예를 들면, 형광 마커가 사용되는 경우, 세포는 형광 활성화된 세포 분류(sorting)에 의해 분리될 수 있는 반면, 세포 표면 마커가 삽입되는 경우, 세포는 친화성 분리 기술, 예를 들면, 자기 분리, 친화성 크로마토그래피, 고체 매트릭스에 부착된 친화성 시약을 사용한 "패닝" 또는 기타 편리한 기술에 의해 이중성 집단으로부터 분리할 수 있다.

[0077] 본원에 사용된 바와 같이, 다중 삽입과 관련하여, 용어 "동시"는, 숙주 세포가 뉴클레아제, 예를 들면, 뉴클레아제를 코딩하는 플라스미드, 및 숙주 세포 게놈에 삽입될 하나 이상의 공여체 DNA로 동시-형질전환되는 시점에서 개시하여, 각각의 표적 유전자좌에 상기 공여체 DNA가 성공적으로 삽입되었는지에 대해 상기 형질전환된 숙주 세포, 또는 이의 클론 집단을 스크리닝하는 시점으로 종결하는 기간을 포함한다. 일부 실시형태에서, "동시"에 포함되는 기간은 적어도, 표적 서열 뉴클레아제가 숙주 세포의 염색체(들) 내의 이의 표적 서열에 결합하여 절단하는 데 필요한 시간이다. 일부 실시형태에서, "동시"에 포함되는 기간은, 숙주 세포가 뉴클레아제, 예를 들면, 뉴클레아제를 코딩하는 플라스미드 및 하나 이상의 공여체 DNA를 동시-형질전환하는 시점에서 개시하여, 적어도 6, 12, 24, 36, 48, 60, 72, 96시간 또는 96시간 초과이다.

[0078] 5.2 외인성 핵산을 삽입하는 방법

[0079] 본원에는 숙주 세포 게놈의 하나 이상의 선별된 표적 부위로 하나 이상의 외인성 핵산을 삽입하는 방법이 제공된다. 특정 실시형태에서, 상기 방법은 숙주 세포를, 게놈 표적 부위 내로 삽입될 외인성 핵산을 포함하는 하나 이상의 삽입 폴리뉴클레오티드, 즉 공여체 DNA; 게놈 표적 부위 근처 또는 내부에 이중-가닥 절단을 야기할 수 있는 하나 이상의 뉴클레아제; 및 그 자체 또는 숙주 세포와 접촉된 하나 이상의 추가 선형 핵산과 상동성 재조합할 수 있는 선형 핵산과 접촉시키는 것을 포함하고, 숙주 세포에서 선형 핵산의 상기 상동성 재조합은 선별가능한 마커에 대한 코딩 서열을 포함하는 환상 염색체의 핵산의 형성을 제공한다. 특정 실시형태에서, 접촉된 숙주 세포는 이어서 선별 조건하에 성장시킨다. 조작 이론에 구속되지 않지만, 본원에 기재된 방법에 따라 선별되도록, 숙주 세포가 HR에 의해 발현 벡터를 환화(circularize)시키는 것은 선별된 세포가 외인성 DNA의 하나 이상의 의도된 HR-매개된 게놈 삽입을 성공적으로 수행할 가능성을 증가시키는 것으로 생각된다.

- [0080] 특정 측면에서, 본원에는 숙주 세포 계놈의 표적 부위 내로 외인성 핵산의 마커레스 삽입을 위한 방법이 제공되고, 상기 방법은
- [0081] (a) 하나 이상의 숙주 세포를,
- [0082] (i) 제1 상동성 영역(HR1) 및 제2 상동성 영역(HR2)를 포함하는 외인성 핵산(ES)(여기서, 상기 (HR1) 및 (HR2)는 상기 숙주 세포 숙주 계놈의 (TS)에서 숙주 세포 매개된 상동성 재조합을 개시할 수 있다);
- [0083] (ii) 각각의 상기 표적(TS)에 대해, (TS)에서 절단할 수 있는 뉴클레아제(N)(여기서, 상기 절단은 (TS)에서 (ES)의 상동성 재조합을 제공한다);
- [0084] 및 (iii) 그 자체, 또는 숙주 세포와 접촉된 하나 이상의 추가 선형 핵산과 상동성 재조합할 수 있는 선형 핵산(여기서, 상기 상동성 재조합은 선별가능한 마커에 대한 코딩 서열을 포함하는 환상 염색체의 핵산의 형성을 제공한다)과 접촉시키는 단계; 및
- [0085] (b) 선별가능한 마커를 발현하는 숙주 세포를 선별하는 단계를 포함한다.
- [0086] 일부 실시형태에서, 상기 방법은 (TS)에서 삽입된 (ES)를 갖는 숙주 세포를 회수하는 것을 포함하고, 상기 회수는 선별가능한 마커의 삽입을 필요로 하지 않는다.
- [0087] 도 1은, 환상 마커 발현 벡터를 형성하기 위해, 숙주 세포 매개된 HR에 의해 생체내에서 조립할 수 있는 부위 특이적 뉴클레아제 및 예비-재조합 조성물을 사용하여 외인성 핵산의 계놈 삽입에 대한 예시적 실시형태를 제공한다. 공여체 폴리뉴클레오티드는 숙주 세포에 도입되고, 상기 폴리뉴클레오티드는 제1 상동성 영역(HR1) 및 제2 상동성 영역(HR2)의 측면에 목적하는 핵산(D)을 포함한다. HR1 및 HR2는 계놈 표적 부위(TS)의 각각 5' 및 3' 영역과 상동성을 공유한다. 부위-특이적 뉴클레아제(N)은 또한 숙주 세포에 도입되고, 상기 뉴클레아제는 표적 부위 내의 고유 서열을 인식 및 절단할 수 있다. 또한, 세포에 예비-재조합 조성물이 도입되고, 이는 이 예에서 서로 상동성 재조합할 수 있는 2개의 상동성 영역을 각각 포함하는 2개의 선형 예비-재조합 분자를 포함한다. 이러한 예에서, 상동성 영역은 각 예비-재조합 분자의 5' 및 3' 말단에 위치된다. 각 예비-재조합 분자의 하나의 상동성 영역은 선별가능한 마커(각각 GF 및 FP)에 대한 부분 코딩 서열을 포함하고, 따라서, 2개 상동성 영역 사이의 HR시에, 선별가능한 마커(GFP)의 완전한 및 작동성 코딩 서열은 환상 마커 발현 벡터 상에 재구성된다. 일반적으로, 이러한 환상화는, 예를 들면, 배양 배지를, 선별가능한 마커가 약물 내성 마커인 선별제(예: 항생물질)와 함께 공급하거나 비색 또는 형광 검출 방법에 의해 검출가능한 마커를 발현하는 세포를 분류함으로써 선별가능한 마커의 발현을 선별하는 조건하에 세포를 배양함으로써 선별된다. 동시에, HR에 적극적인 세포에서, 부위-특이적 뉴클레아제에 의한 표적 부위 내에 이중-가닥 절단의 유도는 절단된 표적 부위에서 목적하는 공여체 핵산의 HR-매개된 삽입을 촉진시킨다. 숙주 세포가 HR에 의해 발현 벡터를 환상화하여 선별되도록 하는 요건을 제공함으로써, 또한 외인성 공여체 DNA의 HR-매개된 삽입을 또한 수행한 세포의 회수가 또한 증가된다. 이러한 증가된 회수 빈도는 재조합 사상을 겪는 형질전환체를 선별하기 위해 선별가능한 마커를 동시-삽입할 필요성을 제거한다. 예를 들면, 조작된 미생물의 작제 동안, 선별가능한 마커에 대한 필요성을 제거함으로써, 숙주 세포 계놈을 조작하는데 필요한 시간이 크게 감소된다. 또한, 조작 전략은 소정 숙주 생물체에 이용가능한 마커의 제한된 캐시(cache)로 인해 선별가능한 마커를 재순환시킬 필요성에 의해 더 이상 제한되지 않는다.
- [0088] 일부 실시형태에서, 성공적으로 삽입된 외인성 핵산을 포함하는 형질전환된 세포의 마커레스 회수는, 스크리닝된, 1000, 900, 800, 700, 600, 500, 400, 300, 200 또는 100개 접촉된 숙주 세포 또는 이의 클론 집단당 약 1개의 빈도로 발생한다. 특정 실시형태에서, 성공적으로 삽입된 외인성 핵산을 포함하는 형질전환된 세포의 마커레스 회수는, 스크리닝된, 90, 80, 70, 60, 50, 40, 30, 20 또는 10개 접촉된 숙주 세포 또는 이의 클론 집단당 약 1개의 빈도로 발생한다. 보다 특정 실시형태에서, 성공적으로 삽입된 외인성 핵산을 포함하는 형질전환된 세포의 마커레스 회수는, 스크리닝된, 9, 8, 7, 6, 5, 4, 3 또는 2개 접촉된 숙주 세포 또는 이의 클론 집단당 약 1개의 빈도로 발생한다.
- [0089] 다양한 방법은 선별가능한 마커의 사용 없이 표적 부위에서 또는 부근에서 변형된 계놈을 갖는 이들 세포를 동정하기 위해 이용가능하다. 일부 실시형태에서, 이러한 방법은 표적 부위에서 임의의 변화를 검출하고자 하고, 이로써 제한되지 않지만, PCR 방법, 서열분석 방법, 뉴클레아제 소화, 예를 들면, 제한 맵핑, 서던 블롯, 및 이의 임의의 조합을 포함한다. 표현형 판독치, 예를 들며, 기능의 예측된 획득 또는 상실은 의도된 계놈 변형(들)을 실시하기 위한 프록시로서 또한 사용될 수 있다.

- [0090] 또 다른 측면에서, 본원에는 숙주 세포 계놈 내로 복수의 외인성 핵산을 삽입하는 방법이 제공되고, 상기 방법은
- [0091] (a) 숙주 세포를
- [0092] (i) 복수의 외인성 핵산(여기서, 각각의 외인성 핵산(ES_x)는 제1 상동성 영역($HR1_x$) 및 제2 상동성 영역($HR2_x$)를 포함하고, 상기 ($HR1_x$) 및 ($HR2_x$)는 상기 숙주 세포 계놈의 표적 부위(TS_x)에서 (ES_x)의 숙주 세포 매개된 상동성 재조합을 개시할 수 있다);
- [0093] (ii) 각각의 상기 표적 부위(TS_x)에 대해, (TS_x)에서 절단할 수 있는 뉴클레아제(N_x)(여기서, 상기 절단은 (TS_x)에서 (ES_x)의 상동성 재조합을 제공한다); 및
- [0094] (iii) 그 자체 또는 숙주 세포와 접촉된 하나 이상의 추가 선형 핵산과 상동성 재조합할 수 있는 선형 핵산(여기서, 상기 상동성 재조합은 선별가능한 마커에 대한 코딩 서열을 포함하는 환상 염색체의 핵산의 형성을 제공한다)과 접촉시키는 단계, 및
- [0095] (b) 선별가능한 마커를 발현하는 숙주 세포를 선별하는 단계를 포함한다.
- [0096] 일부 실시형태에서, 상기 방법은 숙주 세포를 회수하는 것을 추가로 포함하고, 각각 선별된 외인성 핵산(ES_x)는 각각의 선별된 표적 서열(TS_x)에 삽입되고,
- [0097] 여기서, x 는 1 내지 n 의 정수이고, n 은 적어도 2이다.
- [0098] 도 2는 복수의 부위-특이적 뉴클레아제를 사용하여 복수의 외인성 핵산의 동시 계놈 삽입에 대한 예시적 실시형태를 제공한다. 이 예에서, 3개의 상이한 공여체 폴리뉴클레오티드가 숙주 세포에 도입되고, 각각의 폴리뉴클레오티드는 목적하는 핵산(D_x)을 포함하는 외인성 핵산(ES_x)(여기서, x 는 1, 2 또는 3이다)를 포함한다. 각각의 (D_x)는 제1 상동성 영역($HR1_x$) 및 제2 상동성 영역($HR2_x$)의 측면에 존재한다. ($HR1_x$) 및 ($HR2_x$)는 상기 계놈의 전체 3개 고유 표적 부위중 선별된 표적 부위(TS_x)의 각각의 5' 및 3' 영역과 상동성을 공유한다. 하나 이상의 부위-특이적 뉴클레아제(N_x)(예를 들면, 고유 인식 부위를 갖는 하나 이상(예: "x" 수)의 엔도뉴클레아제; 또는 하나 이상(예: "x" 수)의 유도 RNA와 함께 RNA-유도된 엔도뉴클레아제)가 또한 숙주 세포에 도입되고, 여기서 각각의 뉴클레아제(N_x)는 이의 상응하는 표적 부위(TS_x) 내의 고유 서열을 인식 및 절단할 수 있다. 또한, 세포에 예비-재조합 조성물이 도입되고, 이는 이러한 예에서 서로 상동성 재조합할 수 있는 2개의 상동성 영역을 각각 포함하는 2개의 선형 예비-재조합 분자를 포함한다. 이러한 예에서, 상동성 영역은 각 예비-재조합 분자의 5' 및 3' 말단에서 위치된다. 각 예비-재조합 분자 중의 하나의 상동성 영역은 선별가능한 마커(각각 GF 및 FP)에 대한 부분 코딩 서열을 포함하고, 이에 의해, 2개 상동성 영역 사이의 HR시, 선별가능한 마커(GFP)의 완전한 및 작동성 코딩 서열은 환상 마커 발현 벡터 상에서 재구성된다. 이러한 환상화는 선별가능한 마커의 발현을 선별하는 조건하에 세포를 배양함으로써 선별된다. 동시에, HR에 적절한 세포에서, 이의 상응하는 부위-특이적 뉴클레아제(N_x)에 의한 표적 부위(TS_x)의 절단은 숙주 세포의 내인성 상동성 재조합 기작에 의해 (TS_x)에서 상응하는 목적하는 핵산(D_x)의 삽입을 촉진한다. 숙주 세포가 HR에 의해 발현 벡터를 환상화하여 선별되도록 하는 요건을 제공함으로써, 또한 외인성 공여체 DNA의 HR-매개된 삽입을 또한 수행한 세포의 회수가 또한 증가된다.
- [0099] 특정 실시형태에서, 목적하는 핵산(D_x)을 임의로 포함하는 각각의 외인성 핵산(ES_x)는 이의 각각의 계놈 표적 부위(TS_x)에 동시에 삽입되는데, 즉 복수의 폴리뉴클레오티드 및 복수의 뉴클레아제를 사용한 숙주 세포의 단일 형질전환으로 삽입된다. 일부 실시형태에서, 상기 방법은, 상술한 방법에 언급된 변형에 따라, 복수의 외인성 핵산(ES_x)(즉, x 는 1 내지 n 의 임의의 정수이고, n 은 적어도 2이다)을 동시에 삽입하는데 유용하다. 일부 실시형태에서, 본원에 제공된 동시 삽입 방법은 최대 10개 외인성 핵산(ES_x)을 10개 선택된 표적 부위(TS_x)(즉, x 는 1 내지 n 의 임의의 정수이고, n 은 2, 3, 4, 5, 6, 7, 8, 9 또는 10이다)에 동시 삽입하는데 유용하다. 일부 실시형태에서, 본원에 제공된 동시 삽입 방법은 최대 20개 외인성 핵산(ES_x)을 20개 선택된 표적 부위(TS_x)(즉, x 는 1 내지 n 의 임의의 정수이고, n 은 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20이다)에 동시 삽입하는데 유용하다. 일부 실시형태에서, n 은 2이다. 일부 실시형태에서, n 은 3이다. 일부 실

시형태에서, n은 4이다. 일부 실시형태에서, n은 5이다. 일부 실시형태에서, n은 6이다. 일부 실시형태에서, n은 7이다. 일부 실시형태에서, n은 8이다. 일부 실시형태에서, n은 9이다. 일부 실시형태에서, n은 10이다. 일부 실시형태에서, n은 11이다. 일부 실시형태에서, n은 12이다. 일부 실시형태에서, n은 13이다. 일부 실시형태에서, n은 14이다. 일부 실시형태에서, n은 15이다. 일부 실시형태에서, n은 16이다. 일부 실시형태에서, n은 17이다. 일부 실시형태에서, n은 18이다. 일부 실시형태에서, n은 19이다. 일부 실시형태에서, n은 20이다. 일부 실시형태에서, 본원에 제공된 동시 삽입 방법은 20개 이상의 외인성 핵산을 동시 삽입하는데 유용하다.

[0100] 단일 표적 부위에서 단일 외인성 핵산의 삽입과 같이, 이의 각각의 표적 부위에서 성공적으로 삽입된 각각의 외인성 핵산을 갖는 숙주 세포의 회수는, 숙주 세포를 본원에 기재된 하나 이상의 선형 예비-재조합 분자와 접촉하지 않고 선별가능한 마커의 발현을 선별하는 것과 비교하여, 실질적으로 보다 높은 빈도로 발생한다. 일부 실시형태에서, 이러한 증가된 삽입 빈도는 복수의 재조합 사상의 동정을 위한 하나 이상의 선별가능한 마커의 동시-삽입에 대한 요건을 제거한다. 일부 실시형태에서, 복수의 성공적으로 삽입된 외인성 핵산을 포함하는 형질전환된 세포의 마커레스 회수는, 스크리닝된, 1000, 900, 800, 700, 600, 500, 400, 300, 200 또는 100개 접촉된 숙주 세포 또는 이의 클론 집단당 약 1개의 빈도로 발생한다. 특정 실시형태에서, 마커레스 회수는, 스크리닝된, 90, 80, 70, 60, 50, 40, 30, 20 또는 10개 접촉된 숙주 세포 또는 이의 클론 집단당 약 1개의 빈도로 발생한다. 보다 특정 실시형태에서, 마커레스 회수는, 스크리닝된, 9, 8, 7, 6, 5, 4, 3 또는 2개 접촉된 숙주 세포당 약 1개의 빈도로 발생한다.

[0101] 5.2.1 환상 마커 발현 벡터의 HR-매개된 생체내 조립

[0102] 본원에 제공된 방법은 겹 수복에 의한 환상 발현 벡터의 숙주 세포 매개된 조립을 포함한다. 겹 수복은 재조합 DNA 분자를 생체내에서 조립하는 신속하고 효율적인 방법이다. 상기 기술은, 박테리아[*Escherichia coli*; 참조: Datta *et al.*, *Gene* 379:109-115 (2006)], 효모[*Saccharomyces cerevisiae*; 참조: Bessa *et al.*, *Yeast* 29:419-423 (2012)], 곤충[*Drosophila melanogaster*; 참조: Carreira-Rosario *et al.*, *J Vis Exp* 77:e50346 (2013)] 및 포유동물 세포[human cells; 참조: Adar *et al.*, *Nucleic Acids Research* 37(17):5737-5748 (2009)]를 포함하여 다수의 생물체에서 플라스미드를 조립 및/또는 수복하는데 효과적인 것으로 기재되어 있다. 겹 수복은 단일 선형 DNA의 2개 상동성 영역 사이 또는 2개 이상의 분리된 선형 DNA 단편 사이의 상동성 재조합에 의해 환상 DNA 분자를 생성할 수 있다. 전형적으로, 조립된 환상화 DNA는 복제 서열 및 선별 마커를 포함하는 벡터로서 작용한다[참조: Orr-Weaver *et al.*, *Methods Enzymol* 101:228-245 (1983)]. 도 4에 개략된 기술은 통상 선형 "겹화" 벡터 및 선형 DNA 단편(삽입체)의 동시-형질전환으로 개시한다[참조: Orr-Weaver *et al.*, 1983]. 상동성 재조합을 위한 적격 세포에서, 재조합은 벡터와 삽입체 사이의 상동성 서열의 인접 스트레치의 2개 쌍 사이에서 발생하고, 겹이 수복된 보다 큰 환상 벡터를 제공한다. 삽입체의 인접 상동성을 제공하는 단순한 방식은 테일 프라이머가 상동성 영역을 제공하는 폴리머라제 연쇄 반응(PCR)에 의한 것이다.

[0103] 본원에 제공된 방법 및 조성물의 한 가지 측면에서, 숙주 세포는 예비-재조합 벡터 중간체로서 작용하는 단일 인접 선형(겹) 핵산과 접촉된다. 본원에 사용된 바와 같이, 문구 "단일 핵산"은 동일한 핵산 분자의 복수 카피의 실시형태를 포함한다. 일부 실시형태에서, 예비-재조합 벡터는 자가-환상화하고, 서로 상동성 재조합할 수 있는 2개의 서열-특이적 재조합 영역을 포함하고, 따라서 재조합-적격 숙주 세포 내로의 도입은 환상 발현 벡터의 형성을 제공한다. 일부 실시형태에서, 재조합 영역은, 선형 예비-재조합 벡터의 말단에 또는 부근(예를 들면, 하나의 재조합 영역은 선형 벡터의 각각의 말단에 위치되고, 추가 서열은 상기 2개 영역을 개재한다), 상기 말단의 내부(예를 들면, 각각의 재조합 영역은 추가 서열에 의해 양 말단 상에 인접하여 있다) 또는 이의 조합(예를 들면, 하나의 재조합은 선형 벡터의 하나의 말단에 존재하고, 다른 재조합은 이의 내부에 존재하고 추가 서열에 의해 양 측면에 인접하여 있다)에 위치되어 있다. 일부 실시형태에서, 제1 및 제2 재조합 영역은 충분한 길의 임의의 뉴클레오타이드 서열을 포함하고, 서로 상동성 재조합을 가능하게 하는 임의의 서열 동일성을 공유할 수 있다. 일부 실시형태에서, "충분한 서열 동일성"은, 예를 들면, 적어도 15개 염기쌍, 적어도 20개 염기쌍, 적어도 50개 염기쌍, 적어도 100개 염기쌍, 적어도 250개 염기쌍, 적어도 500개 염기쌍 또는 500개 초과 염기쌍의 길이에 걸쳐, 재조합 영역 사이의 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 99% 또는 100% 동일성을 갖는 서열을 지칭한다. 서열 동일성의 정도는, 본원에 기재된 것들, 예를 들면, 디펄트 파라미터를 사용하여, BLAST 2.2.2 또는 FASTA 버전 3.0t78을 포함하는 임의의 컴퓨터 프로그램 및 관련 파라미터를 사용하여 결정할 수 있다. 재조합 영역 사이의 상동성의 유효 길이의 논의는 문헌[참조: Hasty *et al.*, *Mol Cell Biol* 11:5586-91 (1991)]을 참조한다.

[0104] 일부 실시형태에서, 제1 및 제2 재조합 영역은 적어도 25% 뉴클레오타이드 서열 동일성을 공유한다. 일부 실시형

태에서, 제1 및 제2 재조합 영역은 적어도 30% 뉴클레오타이드 서열 동일성을 공유한다. 일부 실시형태에서, 제1 및 제2 재조합 영역은 적어도 35% 뉴클레오타이드 서열 동일성을 공유한다. 일부 실시형태에서, 제1 및 제2 재조합 영역은 적어도 40% 뉴클레오타이드 서열 동일성을 공유한다. 일부 실시형태에서, 제1 및 제2 재조합 영역은 적어도 45% 뉴클레오타이드 서열 동일성을 공유한다. 일부 실시형태에서, 제1 및 제2 재조합 영역은 적어도 50% 뉴클레오타이드 서열 동일성을 공유한다. 일부 실시형태에서, 제1 및 제2 재조합 영역은 적어도 60% 뉴클레오타이드 서열 동일성을 공유한다. 일부 실시형태에서, 제1 및 제2 재조합 영역은 적어도 65% 뉴클레오타이드 서열 동일성을 공유한다. 일부 실시형태에서, 제1 및 제2 재조합 영역은 적어도 70% 뉴클레오타이드 서열 동일성을 공유한다. 일부 실시형태에서, 제1 및 제2 재조합 영역은 적어도 75% 뉴클레오타이드 서열 동일성을 공유한다. 일부 실시형태에서, 제1 및 제2 재조합 영역은 적어도 80% 뉴클레오타이드 서열 동일성을 공유한다. 일부 실시형태에서, 제1 및 제2 재조합 영역은 적어도 85% 뉴클레오타이드 서열 동일성을 공유한다. 일부 실시형태에서, 제1 및 제2 재조합 영역은 적어도 90% 뉴클레오타이드 서열 동일성을 공유한다. 일부 실시형태에서, 제1 및 제2 재조합 영역은 적어도 95% 뉴클레오타이드 서열 동일성을 공유한다. 일부 실시형태에서, 제1 및 제2 재조합 영역은 적어도 99% 뉴클레오타이드 서열 동일성을 공유한다. 일부 실시형태에서, 제1 및 제2 재조합 영역은 100% 뉴클레오타이드 서열 동일성을 공유한다.

[0105] 특정 실시형태에서, 각각의 제1 및 제2 재조합 영역은 약 50 내지 5,000개 뉴클레오타이드로 이루어진다. 특정 실시형태에서, 각각의 제1 및 제2 재조합 영역은 약 50 내지 5,000개 뉴클레오타이드를 포함한다. 특정 실시형태에서, 각각의 제1 및 제2 재조합 영역은 약 100 내지 2,500개 뉴클레오타이드로 이루어진다. 특정 실시형태에서, 각각의 제1 및 제2 재조합 영역은 약 100 내지 1,000개 뉴클레오타이드로 이루어진다. 특정 실시형태에서, 각각의 제1 및 제2 재조합 영역은 약 250 내지 750개 뉴클레오타이드로 이루어진다. 특정 실시형태에서, 각각의 제1 및 제2 재조합 영역은 약 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800, 3900, 4000, 4100, 4200, 4300, 4400, 4500, 4600, 4700, 4800, 4900 또는 5,000개 뉴클레오타이드로 이루어진다. 일부 실시형태에서, 각각의 제1 및 제2 재조합 영역은 약 500개 뉴클레오타이드로 이루어진다. 일부 실시형태에서, 각각의 제1 및 제2 재조합 영역은 적어도 18개 뉴클레오타이드 염기쌍을 포함한다. 일부 실시형태에서, 각각의 제1 및 제2 재조합 영역은 15 내지 500개 뉴클레오타이드 염기쌍으로 이루어진다. 일부 실시형태에서, 각각의 제1 및 제2 재조합 영역은 15 내지 500, 15 내지 495, 15 내지 490, 15 내지 485, 15 내지 480, 15 내지 475, 15 내지 470, 15 내지 465, 15 내지 460, 15 내지 455, 15 내지 450, 15 내지 445, 15 내지 440, 15 내지 435, 15 내지 430, 15 내지 425, 15 내지 420, 15 내지 415, 15 내지 410, 15 내지 405, 15 내지 400, 15 내지 395, 15 내지 390, 15 내지 385, 15 내지 380, 15 내지 375, 15 내지 370, 15 내지 365, 15 내지 360, 15 내지 355, 15 내지 350, 15 내지 345, 15 내지 340, 15 내지 335, 15 내지 330, 15 내지 325, 15 내지 320, 15 내지 315, 15 내지 310, 15 내지 305, 15 내지 300, 15 내지 295, 15 내지 290, 15 내지 285, 15 내지 280, 15 내지 275, 15 내지 270, 15 내지 265, 15 내지 260, 15 내지 255, 15 내지 250, 15 내지 245, 15 내지 240, 15 내지 235, 15 내지 230, 15 내지 225, 15 내지 220, 15 내지 215, 15 내지 210, 15 내지 205, 15 내지 200, 15 내지 195, 15 내지 190, 15 내지 185, 15 내지 180, 15 내지 175, 15 내지 170, 15 내지 165, 15 내지 160, 15 내지 155, 15 내지 150, 15 내지 145, 15 내지 140, 15 내지 135, 15 내지 130, 15 내지 125, 15 내지 120, 15 내지 115, 15 내지 110, 15 내지 105, 15 내지 100, 15 내지 95, 15 내지 90, 15 내지 85, 15 내지 80, 15 내지 75, 15 내지 70, 15 내지 65, 15 내지 60, 15 내지 55, 15 내지 50, 15 내지 45, 15 내지 40, 15 내지 35, 15 내지 30, 15 내지 25, 또는 15 내지 20개 뉴클레오타이드 염기쌍으로 이루어진다. 일부 실시형태에서, 각각의 제1 및 제2 재조합 영역은 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199 또는 200개 뉴클레오타이드 염기쌍으로 이루어진다.

[0106] 본원에 제공된 방법 및 조성물의 바람직한 실시형태에서, 상동성 영역 쌍의 각각의 상동성 영역은, 서로 상동성 재조합을 가능하게 하지만, 조립에 관여하는 예비-재조합 분자(들)의 다른 영역 또는 숙주 세포의 임의의 게놈

영역과 상동성 재조합을 가능하게 하지 않는 충분한 길이 및 서열 동일성의 뉴클레오티드 서열을 포함한다.

[0107] 일부 실시형태에서, 환상 발현 벡터는 자가-환상화되는 단일 예비-재조합 분자로부터 형성되지만, 다른 실시형태에서, 환상 발현 벡터는 2개 이상의 선형 예비-재조합 분자의 HR-매개된 조립에 의해 형성된다. 예를 들면, 환상 벡터는 2개의 선형 예비-재조합 분자로부터 조립될 수 있고, 여기서 제1 분자는 겹 벡터이고, 제2 분자는 겹 벡터 상의 2개 상동성 영역과 재조합할 수 있는 2개의 분리된 상동성 영역을 포함하는 삽입체이다. 각각의 겹 선형 벡터 및 선형 삽입체에 있어서, 재조합 영역은, 선형 예비-재조합 벡터의 말단에 또는 부근(예를 들면, 하나의 재조합 영역은 각각의 말단에 위치되고, 추가 서열은 상기 2개 영역을 개재한다), 상기 말단의 내부(예를 들면, 각각의 재조합 영역은 추가 서열에 의해 양 말단 상에 인접하여 있다) 또는 이의 조합(예를 들면, 하나의 재조합은 하나의 말단에 존재하고, 다른 재조합은 이의 내부에 존재하고 추가 서열에 의해 양 측면에 인접하여 있다)에 위치되어 있다. 여전히 다른 실시형태에서, 겹 벡터를 수복하는 삽입체는 서로에 대해 상동성 영역을 포함하는 적어도 2개 선형 핵산으로부터 자체로 조립될 수 있다. 예를 들면, 환상 벡터는 3개의 상이한 선형 예비-재조합 단편으로부터 형성될 수 있고, 여기서 제1 선형 분자는 상동성 영역 A_1 및 A_2 를 포함하고, 제2 선형 분자는 B_2 및 C_2 를 포함하고, 제3 선형 분자는 C_3 및 A_3 을 포함하고, 따라서 HR-적격 숙주 세포에서 각 단편의 상동성 영역 사이의 재조합(즉, A_1 과 A_3 , B_1 과 B_2 , 및 C_2 와 C_3)은 영역 $A \rightarrow B \rightarrow C$ 를 포함하는 환상 발현 벡터의 형성을 제공한다.

[0108] 여전히 다른 실시형태에서, 환상 벡터는 유사한 방식으로 적어도 4, 5, 6, 7, 8, 9 또는 10개의 상이한 선형 예비-재조합 단편으로부터 HR-적격 숙주 세포에서 조립된다. 조작 이론에 구속되지 않고서, 2개 이상의 선형 예비-재조합 분자로부터 조립되는 환상 발현 벡터를 필요로 하는 것은 상동성 재조합에 특히 능숙한 숙주 세포를 선별하는 것으로 생각된다. 따라서, 다중 예비-재조합 분자로부터 환상 발현 벡터의 조립은, 보다 고차의 삽입 사상이 요구되는 경우, 예를 들면, 다중 게놈 삽입(예를 들면, 2개 이상의 공여체 외인성 DNA), 또는 매우 낮은 비율의 HR을 갖는 것으로 공지되거나 의심되는 세포형 내로 게놈 삽입을 수행하는 경우에 바람직할 수 있다. 한 가지 예에서, 숙주 세포의 3개 각각의 게놈 표적 부위 내로 3개 외인성 공여체 핵산의 다중(즉, 동시) 삽입을 위해, 숙주 세포는 적어도 3개 선형 예비-재조합 단편을 조립하여 환상 발현 벡터를 형성하도록 "강제"된다. 선별가능한 마커를 발현하는 환상 벡터를 형성하기 위해 3개 단편을 성공적으로 재조합할 수 있는 유일한 세포는 선별을 생존, 즉 선별될 수 있고, 이들 세포는 이들의 각각의 게놈 표적 부위 내로 각각의 3개 외인성 공여체 핵산을 성공적으로 삽입할 가능성이 보다 높은 것이다. 일부 실시형태에서, 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 10개 초과 각각의 게놈 표적 부위 내로 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 10개 초과의 외인성 공여체 핵산의 다중 삽입이 요구되는 경우, 숙주 세포는 환상 발현 벡터를 형성하기 위해 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 10개 이상의 예비-재조합 단편을 조립, 즉 재조합하도록 강제되고, 이로부터의 선별가능한 마커는 숙주 세포에서 발현된다. 일부 실시형태에서, 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 10개 초과의 예비-재조합 단편의 생체내 조립은 하나 이상의 선별가능한 마커(예: 2, 3개, 또는 3개 이상의 상이한 선별가능한 마커)에 대한 코딩 서열을 포함하는 환상 발현 벡터를 형성한다.

[0109] 바람직한 실시형태에서, 환상 발현 벡터는, 조립되면, 선별가능한 마커에 대한 코딩 서열, 및 적합한 조절 서열, 예를 들면, 숙주 세포에서 마커의 발현을 가능하게 하는 프로모터 및/또는 터미네이터를 포함한다. 선별가능한 마커는 둘 다 선별가능한 마커를 함유하는 숙주 세포의 선별을 가능하게 하는 것들을 포함하고, 다른 실시형태는 선별가능한 마커를 함유하지 않는 세포의 선별을 가능하게 하며, 이는 선별 후에 환상 발현 벡터의 제거가 요구되는 양태에서의 용도를 발견한다(하기 언급된 바와 같음). 일부 실시형태에서, 환상 발현 벡터의 조립 전에, 선형 예비-재조합 분자, 또는 다중-단편 조립의 예비-재조합 단편 중의 적어도 하나는, 이의 조절 서열에 대한 작동적 연결로, HR-매개된 조립에 수반된 상동성 영역으로부터 분리된 및 떨어진, 선별가능한 마커에 대한 무손상 코딩 서열을 포함한다. 따라서, 환상 발현 벡터의 조립은 선별가능한 마커의 코딩 서열 또는 발현에 필요한 임의의 이의 조절 서열을 변경하지 않는다. 일부 이러한 실시형태에서, 환상화 사상은 마커의 코딩 서열의 증식 및 이로부터의 지속된 발현을 단지 가능하게 하는 반면, 비-환상 선형 벡터는 숙주 세포에서 증식 및/또는 유지될 수 없다. 바람직한 실시형태에서, 환상 발현 벡터를 포함하지 않는 숙주 세포는 선별 단계를 생존하지 않고/않거나 본원에 기재된 방법에서 선별되지 않는다. 조작 이론에 구속되지 않고서, 본원에 기재된 방법에 따라 선별되도록, 숙주 세포가 HR에 의해 발현 벡터를 환상화하도록 함으로써, 선별된 세포가 외인성 DNA의 하나 이상의 의도된 HR-매개된 게놈 삽입을 성공적으로 수행할 가능성을 증가시키는 것으로 생각된다.

[0110] 다른 실시형태에서, 선별가능한 마커를 코딩하는 서열 및/또는 발현에 필요한 이의 조절 서열은 무손상이 아니

고, 즉 임의의 단일 예비-재조합 분자에 작동적 연결로 아니다. 일부 이러한 실시형태에서, 발현 벡터가 환상화되는 경우에만 작동적으로 연결된 이의 조절 요소와 함께 마커 코딩 서열이다. 따라서, 마커를 코딩하는 서열 및/또는 이의 필요한 조절 서열은, 성분 예비-재조합 단편 사이의 HR이 이의 조절 서열과의 작동적 연결로 선별가능한 마커의 코딩 서열의 재구성을 제공하는 한, 조립에 관여하는 임의의 수의 단편 중에 분포된 임의의 수의 중첩 상동성 서열로 분할될 수 있다. 이들 실시형태는 숙주 세포의 선별을 회피하는데 특히 유용하고, 여기서 환상 발현 벡터의 형성은 비-HR 메카니즘, 예를 들면, 비-상동성 말단-결합 또는 단일-가닥 어닐링에 의해 예비-재조합 단편의 결합에 기인하고; 선별을 생존하는 이러한 세포는 펄스 포지티브를 나타낼 것이다. 원하지 않는 사상의 빈도는 포스포타제를 사용하여 예비-재조합 단편(들) 상의 5'-포스페이트 그룹을 제거함으로써 저하될 수 있고, 이는 시험관내 연결에 사용된 표준 방법이다. 벡터 재연결은 또한 Taq DNA 폴리머라제 및 dATP를 사용한 예비-재조합 단편(들)의 처리에 의해 회피할 수 있고, 이는 생체내에서 벡터 재-환상화를 방지하여 진정한 재조합체 클론의 스크리닝을 촉진하는데 특히 효과적인 것으로 보고되었다[참조: Bessa *et al.*, *Yeast* 29:419-423 (2012)]. 또한, 예비-환상 DNA의 잘못된 도입에 의해 유발된 펄스 포지티브는 엔도뉴클레아제 소화에 의해 환상 벡터를 선형화시키는 대신에 PCR에 의해 예비-재조합 단편(들)을 준비하고, 이어서 비-소화된 환상 주형에 걸쳐 운반할 수 있는 단편을 분리함으로써 회피할 수 있다. 그럼에도 불구하고, 조작 이론에 구속되지 않고서, 예를 들면, 적어도 2개의 부분 서열로부터, 숙주 세포가 HR에 의해 마커 코딩 서열을 재구성하도록 하는 것은, 선별 프로세스를 생존하기 위해, 본원에 기재된 방법에 따라 선별된 세포가 또한 외인성 DNA의 하나 이상의 의도된 HR-매개된 게놈 삽입을 성공적으로 수행할 가능성을 증가시키는 것으로 생각된다.

[0111] 본원에 제공된 방법의 일부 실시형태에서, 환상 발현 벡터는, 조립되면, 본원에 기재된 부위-특이적 뉴클레아제에 대한 코딩 서열, 및 적합한 조절 서열, 예를 들면, 숙주 세포에서 뉴클레아제의 발현을 가능하게 하는 프로모터 및/또는 터미네이터를 추가로 포함한다. 일부 실시형태에서, 뉴클레아제는 CRISPR/Cas-연관된 RNA-유도 엔도뉴클레아제, 메가뉴클레아제, 징크 핑거 뉴클레아제, TAL-작동자 DNA 결합 도메인-뉴클레아제 융합 단백질(TALEN), 트랜스포사제 및 부위-특이적 재조합효소로 이루어진 그룹으로부터 선택된다. 일부 실시형태에서, 뉴클레아제는 CRISPR-연관된 RNA-유도 엔도뉴클레아제이다. 일부 이러한 실시형태에서, 환상 발현 벡터는 RNA-유도 엔도뉴클레아제, 예를 들면, RNA-유도 DNA 엔도뉴클레아제에 의해 게놈 표적 DNA의 부위-특이적 인식 및 절단을 가능하게 하는 crRNA 활성 및 tracrRNA 활성을 위한 절대 (obligate) 유도 서열을 가능하게 하는 서열 또는 서열들을 추가로 포함한다. 일부 실시형태에서, crRNA 활성 및 tracrRNA 활성은 단일 인접 RNA 분자, 즉 키메라 유도 RNA(gRNA) 분자로서 환상 발현 벡터로부터 발현된다. 일부 실시형태에서, 환상 발현 벡터는, 또한 뉴클레아제에 대한 코딩 서열을 포함하지 않고서, RNA-유도 엔도뉴클레아제에 대한 유도 서열(들)(예: gRNA)을 코딩하는 하나 이상의 서열을 포함한다. 일부 이러한 실시형태에서, RNA-유도 뉴클레아제를 코딩하는 하나 이상의 서열은 게놈 내로 삽입된 분리 벡터 상에 공급될 수 있거나, 뉴클레아제는 단백질로서 세포에 도입될 수 있고, 예를 들면, 시험관내에서 발현 및 정제될 수 있다.

[0112] 세포에서 뉴클레아제의 발현 및 작동능에 필요한 뉴클레아제 코딩 서열 및/또는 추가 서열이 환상 발현 벡터에 포함되는 상술한 임의의 실시형태에서, 이들 서열은 조립 반응에 관여하는 임의의 하나 이상의 예비-재조합 선형 단편(들)에서 무손상(즉, 작동적으로 연결)일 수 있거나, 달리, 성분 예비-재조합 단편 사이의 HR이 이의 조절 서열과의 작동적 연결로 뉴클레아제의 코딩 서열의 재구성을 제공하는 한, 조립에 관여하는 임의의 수의 단편 중에 분포된 임의의 수의 중첩 상동성 서열로 분할될 수 있다. 유리하게는, 뉴클레아제의 코딩 서열(및/또는 뉴클레아제가 RNA-유도 뉴클레아제인 유도 RNA 서열을 코딩하는 서열)을 환상 발현 벡터에 결합시키는 것은 이들 서열의 발현이 HR-매개된 삽입 사상을 보조하기에 충분한 수준 및 기간으로 유지되는 것을 보장한다. 본원에 기재된 방법에 따라, 유전자 표적화의 효율은 의도된 삽입 부위 부근에 도입된 표적화 게놈 이중-가닥 절단(DSB)과 조합될 때에 개선될 수 있다[참조: Jasin, M., *Trends Genet* 12(6):224-228 (1996); and Urnov *et al.*, *Nature* 435(7042):646-651 (2005)]. 더욱이, 뉴클레아제의 코딩 서열 및/또는 연관 유도 서열(들)을 환상 발현 벡터에 결합시키는 것은 이들 서열의 발현을 실시하기 위해 다중 벡터를 숙주 세포에 도입할 필요성을 제거한다. 추가로, 이러한 결합은 목적하는 삽입을 수행한 숙주 세포의 선별 후에 뉴클레아제 및 표지된 플라스미드의 동시 제거를 가능하게 한다. 따라서, 뉴클레아제의 불필요하게 연장된 발현이 회피되고, 결과적으로, 이와 연관된 임의의 독성이 회피된다[참조: Cho *et al.*, *Genome Res*, "Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases," (2013); Sander *et al.*, "in silico abstraction of zinc finger nuclease cleavage profiles reveals an expanded landscape of off-target sites," *Nucleic Acids Research* 41(19):e181 (2013)].

[0113] 본원에 제공된 방법의 일부 실시형태에서, 환상 발현 벡터는, 조립되면, 하기 섹션 5.2.2에 기재된 하나 이상의 외인성 공여체 핵산을 추가로 포함한다. 일부 이러한 실시형태에서, 외인성 공여체 핵산은 숙주 세포 내로 또

한 도입된 뉴클레아제, 예를 들면, 환상 발현 벡터에 의해 또한 코딩된 뉴클레아제에 대한 인식 서열을 갖는 외인성 공여체 핵산을 플랜킹함으로써 환상 발현 벡터로부터 방출될 수 있다.

[0114] 당업자에게 명백한 바와 같이, 환상 발현 벡터는 바람직하게는, 숙주 세포 분열의 복수 주기 동안 발현 벡터가 복제, 증식 및 분리되는 것을 가능하게 하는 자가 증식 서열을 함유할 것이다. 자가 증식 서열은 원핵생물 또는 진핵생물일 수 있고, 복제 기원을 포함한다. 복제 기원은, 다량체 기원-결합 단백질에 의해 인식되고 기원 부위에서 DNA 복제 효소를 조립하는데 중요한 역할을 하는 다중 짧은 반복 서열을 포함하는 고유 폴리뉴클레오티드이다. 본원에 제공된 도입 및 조립 벡터에 사용하기에 적합한 복제 기원은, 이로써 제한되지 않지만, 이. 콜라이 oriC, colE1 플라스미드 기원, 2 μ 및 ARS(둘 다 효모 시스템에 유용함), sf1, SV40 EBV oriP(포유동물 시스템에 유용함), 또는 pSC101에서 발견된 것들을 포함한다. 발현 벡터의 특정 실시형태는 원핵생물 및 진핵생물 자가 증식 서열 둘 다를 포함한다. 이러한 서열은 조립 반응에 관여하는 임의의 하나 이상의 예비-재조합 선형 단편(들)에서 무손상(즉, 작동적 연결)일 수 있거나, 달리는, 성분 예비-재조합 단편이 자가 증식 서열의 재구성을 제공하는 한, 조립에 관여하는 임의의 수의 단편 중에 분포된 임의의 수의 중첩 자가 서열로 분열될 수 있다. 특정 실시형태에서, 자가 증식 서열은 무손상이 아니고, 즉 임의의 단일 예비-재조합 분자에 작동적 연결로 존재하지 않는다. 일부 이러한 실시형태에서, 발현 벡터가 예비-재조합 분자(들)의 재조합에 의해 환상화되는 경우만이 작동적으로 연결되는 자가 증식 서열이다. 이들 실시형태는, 환상 발현 벡터의 형성이, 예를 들면, 비-상동성 말단-결합 또는 단일-가닥 어닐링에 의해 무손상 자가 증식 서열을 포함하는 예비-재조합 단편의 비-HR 결합에 기인하는 숙주 세포의 선별을 회피하는데 특히 유용하고; 선별 생존 세포는 펄스 포지티브를 나타낼 것이다.

[0115] 당해 기술분야의 숙련자에게 명백한 바와 같이, 숙주 세포에서 환상 발현 벡터의 형성을 제공하는, 상기 기재된 하나 이상의 예비-재조합 DNA 단편 사이의 생체내 재조합은 또한 적절한(예: 동일한) 상동성 영역을 개시점으로 포함하는 환상 예비-재조합 핵산으로 달성될 수 있다. 본원에 기재된 임의의 예비-재조합 조성물에 있어서, 상기 조성물은 선형 핵산 분자로서 숙주 세포에 직접 형질전환될 수 있거나, 달리는 예비-재조합 분자를 포함하는 모 환상 분자는 숙주 세포에 도입되고 하나 이상의 뉴클레아제에 의해 생체내에서 절단되어 예비-재조합 분자를 유리시킬 수 있다. 일부 실시형태에서, 마커 벡터 조립에 관여하는 하나 이상의 예비-재조합 분자는 절단을 위해 하나 이상의 게놈 표적 부위를 표적화하는 하나 이상의 뉴클레아제에 의해 숙주 세포에서 생체내 절단을 통해 모 환상 플라스미드로부터 유리될 (liberate) 수 있다.

[0116] 또 다른 측면에서, 본원에는 본원에 제공된 삽입 방법의 실시예에 유용한 예비-재조합 발현 벡터 중간체를 제조하는 방법이 제공된다. 일부 실시형태에서, 자가 증식 서열, 제1 프라이머 결합 서열 및 제2 프라이머 결합 서열을 포함하는 기본 벡터는 적어도 제1 프라이머 및 제2 프라이머를 사용하여 증폭된다. 제1 프라이머는 통상 제1 서열-특이적 재조합 서열을 갖는 5' 부분, 및 기본 벡터 (base vector)의 제1 프라이머 결합 서열에 대해 실질적으로 상보성인 (즉, 목적하는 핵산의 증폭을 가능하게 하지만 다른 목적하지 않는 분자의 증폭을 불가능하게 하기에 충분한 상보성을 가짐) 프라이밍 부분을 갖는 3' 부분을 포함한다. 유사하게는, 제2 프라이머는 제2 서열-특이적 재조합 서열을 갖는 5' 부분, 및 기본 벡터의 제2 프라이머 결합 서열에 대해 실질적으로 상보성인 프라이밍 부분을 갖는 3' 부분을 포함한다. 기본 벡터의 증폭(이는 증폭 프로세스의 개시 전에 선형 또는 환상일 수 있음)은, 제1 서열-특이적 재조합 영역을 포함하는 제1 말단 및 제2 서열-특이적 재조합 영역을 포함하는 제2 말단을 갖는 선형 발현 벡터 중간체의 생성을 제공한다. 특정 실시형태에서, 기본 벡터는 플라스미드, 특히 당해 기술분야에 공지되고 다양한 박테리아- 또는 효모-유래된 염색체의 요소에 기반하는 플라스미드이다. 특정한 다른 실시형태에서, 기본 벡터는 하나 이상의 선별가능한 마커, 전사 개시 서열 및/또는 전사 종결 서열을 추가로 포함한다. 당해 기술분야의 숙련가가 인식하는 바와 같이, 표적 핵산에 함유된 유전자의 발현을 조절하도록 의도된 요소는, 환상 발현 벡터가 생체내에서 조립되는 경우, 발현되는 유전자(들)와 기능적 또는 작동적으로 연관되도록 발현 벡터에 위치되어야 한다. 이러한 요소의 특정 위치는, 상기 기재된 바와 같이, 숙주 세포에서 목적하는 삽입 수를 포함하여, 사용된 요소들, 숙주 세포, 발현되는 유전자(들) 및 기타 요인에 의존한다. 결과적으로, 본 발명의 기술에 따라 제조된 특정 발현 벡터의 최종 디자인은 선택의 문제이고, 특정한 적용에 의존한다.

[0117] 다른 측면은 상기 방법에 따라 제조된 발현 벡터 중간체, 및 이를 함유하는 숙주 세포에 관한 것이다. 또 다른 측면은 본 발명의 삽입 방법의 실시예에 유용한 복수의 상이한 발현 벡터 중간체를 제조하는 방법에 관한 것이다. 이러한 방법에서, 기본 벡터는, 상이한 삽입 핵산과의 상동성 재조합을 가능하게 하는 고유 서열-특이적 재조합 영역을 각각 갖는 2개 이상의 발현 벡터 중간체를 생성하기 위해 증폭된다. 이러한 증폭 반응은 바람직하게는 별개의 발현 벡터 중간체를 생성하기 위해 분리된 반응 혼합물에서 수행된다. 이러한 높은 처리량 방법의 특히

바람직한 실시형태에서, 필수 조작은 하나 이상의 단계가 컴퓨터-제어 장치에 의해 수행되는 자동화 방식으로 수행된다.

[0118] 일부 실시형태에서, 임의의 벡터는 본원에 제공된 바와 같이 예비-재조합 분자를 작제하기 위해 사용될 수 있다. 특히, 당해 기술분야에 공지된 벡터 및 상업적으로 이용가능한 것들(및 이의 변이체 또는 유도체)는 상기 기재된 바와 같이 재조합 영역을 포함하도록 조작될 수 있다. 이러한 벡터는, 예를 들면, 벡터 라보라토리즈 인크(Vector Laboratories Inc.), 인비트로젠(InVitrogen), 프로메가(Promega), 노바젠(Novagen), NEB, 클론테크(Clontech), 베링거 만하임(Boehringer Mannheim), 파마시아(Pharmacia), 에피센터(EpiCenter), 오리진스 테크놀로지스 인크(Origenes Technologies Inc.), 스트라타젠(Stratagene), 퍼킨 엘머(Perkin Elmer), 파밍겐(Pharmingen), 라이프 테크놀로지, 인크(Life Technologies, Inc.), 및 리서치 제네틱스(Research Genetics)으로부터 수득할 수 있다. 특정 목적하는 벡터의 일반 분류는 원핵생물 및/또는 진핵생물 클로닝 벡터, 발현 벡터, 융합 벡터, 2-하이브리드 또는 역 2-하이브리드 벡터, 상이한 숙주에 사용하기 위한 서플 벡터, 돌연변이 유발 벡터, 전사 벡터, 거대 삽입체를 수용하기 위한 벡터 등을 포함한다. 목적하는 기타 벡터는 바이러스 기원 벡터(M13 벡터, 박테리아 파지 λ 벡터, 아데노바이러스 벡터, 아데노-연관 바이러스 벡터(AAV) 및 레트로바이러스 벡터), 높은, 낮은 및 조절가능한 카피수 벡터, 단일 숙주와 조합하여 사용하기 위한 혼화성 레플리콘을 갖는 벡터(PACYC184 및 pBR322) 및 진핵생물 에피솜 복제 벡터(pCDM8)를 포함한다. 다른 실시형태에서, 예비-재조합 분자는, 클로닝된 DNA(예를 들면, DNA "라이브러리")로부터 당해 기술분야에 공지된 표준 절차에 의해, 화학적 합성에 의해, cDNA 클로닝에 의해, 또는 목적하는 세포로부터 정제된 게놈 DNA 또는 이의 단편의 클로닝에 의해 또는 PCR 증폭 및 클로닝에 의해 수득할 수 있다[참조: Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual*, 3d. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (2001); Glover, D.M. (ed.), *DNA Cloning: A Practical Approach*, 2d. ed., MRL Press, Ltd., Oxford, U.K. (1995)].

[0119] 5.2.1.1 선별가능한 마커

[0120] 바람직한 실시형태에서, 환상 발현 벡터는, 조합되는 경우, 선별가능한 마커에 대한 코딩 서열, 및 적합한 조절 서열, 예를 들면, 숙주 세포에서 마커의 발현을 가능하게 하는 프로모터 및/또는 터미네이터를 포함한다. 유용한 선별가능한 마커는 포지티브 및 네가티브 선별 시스템 둘 다에서 기능하는 것들을 포함한다.

[0121] 일부 실시형태에서, 목적하는 세포의 선별은 선별가능한 마커에 의해 코딩된 약물 내성을 선별하는 것에 기반한다(도 3B). 포지티브 선별 시스템은 형질전환된 세포의 성장을 촉진시키는 것들이다. 이들은 조건-포지티브 또는 비-조건-포지티브 선별 시스템으로 분할될 수 있다. 조건-포지티브 선별 시스템은, 비형질전환된 세포에 독성인 특정 기질에 내성을 부여하거나, 형질전환된 세포의 성장 및/또는 분화를 조장하는 단백질, 일반적으로 효소를 코딩하는 유전자로 이루어진다. 조건-포지티브 선별 시스템에서, 기질은 수개 방식 중의 하나로 작용할 수 있다. 이는 항생물질, 제초제, 약물 또는 대사 유사체 또는 탄소 공급 전구체일 수 있다. 각각의 경우에, 상기 유전자는 형질전환된 세포의 선택적 성장 및 증식을 조장하도록 기질에 대해 특이성을 갖는 효소를 코딩한다. 기질은 비형질전환된 세포에 대해 독성 또는 무독성일 수 있다. 단백질 합성을 억제함으로써 카나마이신 내성을 부여하는 *nptII* 유전자는 비형질전환된 세포에 독성인 시스템의 전형적 예이다. 포스포만노즈 이소머라제를 코딩하는 *manA* 유전자는 선별 기질이 독성이 아닌 조건-포지티브 선별 시스템의 예이다. 이 시스템에서, 기질 만노즈는 비형질전환된 세포에 대한 탄소 공급원으로 작용할 수 없지만, 이는 *manA*로 형질전환된 세포의 성장을 촉진시킬 것이다. 비-조건-포지티브 선별 시스템은 외부 기질을 필요로 하지 않고, 형질전환된 세포의 선별 성장 및 분화를 여전히 촉진시킨다. 식물에서의 예는 식물 호르몬 수준을 내인적으로 변형시킴으로써 슈트(shoot) 발달을 증강시키는 *ipt* 유전자이다.

[0122] 네가티브 선별 시스템은 형질전환된 세포의 사멸을 제공한다. 이들은 조건적 및 비-조건적 선별 시스템으로 기재될 수 있는 우성 선별가능한 마커 시스템이다. 선별 시스템이 기질 의존적이지 않는 경우, 이는 비-조건-네가티브 선별 시스템이다. 예는 독성 단백질, 예를 들면, 특정한 세포형을 제거하기 위한 리보뉴클레아제의 발현이다. 독성 유전자의 작용이 독성을 발현하는 기질을 필요로 하는 경우, 상기 시스템은 조건 네가티브 선별 시스템이다. 이들은 박테리아 *codA* 유전자를 포함하고, 이는 시토신 데아미나제, 박테리아 사이토크롬 P450 모노-옥시게나제 유전자, 박테리아 할로알칸 데할로게나제 유전자, 또는 아라비도시스(*Arabidopsis*) 알콜 데하이드로게나제 유전자를 코딩한다. 이들 각각은 무독성 체제를 독성 체제로 전환시켜 형질전환된 세포의 사멸을 제공한다. *codA* 유전자는 또한 염록체 형질전환을 위한 효과적인 우성 네가티브 선별 마커인 것으로 밝혀졌다. 아그로박테리움(*Agrobacterium*) *aux2* 및 *tms2* 유전자는, 이들이 포지티브 선별 시스템에서 또한 사용될 수 있다는 점에서 흥미롭다. 포지티브-네가티브 선별 시스템의 조합은 본원에 제공된 삽입 방법에 특히 유용한데, 포지티브 선별은 환상 발현 벡터를 성공적으로 재조합한(및 추정적으로는, 하나 이상의 의도된 HR-매개된 게놈 삽

입을 수행한) 세포를 농후화하기 위해 사용될 수 있고, 네가티브 선별은, 목적하는 계능 삽입이 확인되면, 동일한 집단으로부터 발현 벡터를 제거("치유")하기 위해 사용될 수 있기 때문이다.

[0123] 광범위한 선별가능한 마커는 당해 기술분야에 공지되어 있다[참조: Kaufinan, *Meth. Enzymol.*, 185:487 (1990); Kaufman, *Meth. Enzymol.*, 185:537 (1990); Srivastava and Schlessinger, *Gene*, 103:53 (1991); Romanos *et al.*, in *DNA Cloning 2: Expression Systems*, 2nd Edition, pages 123-167 (IRL Press 1995); Markie, *Methods Mol. Biol.*, 54:359 (1996); Pfeifer *et al.*, *Gene*, 188:183 (1997); Tucker and Burke, *Gene*, 199:25 (1997); Hashida-Okado *et al.*, *FEBS Letters*, 425:117 (1998)]. 일부 실시형태에서, 선별가능한 마커는 약물 내성 마커이다. 약물 내성 마커는 달리는 세포를 사멸시킬 수 있는 외인성 약물을 세포가 해독할 수 있게 한다. 약물 내성 마커의 구체적 예는, 이로써 제한되지 않지만, 항생물질, 예를 들면, 암피실린, 테트라사이클린, 케나마이신, 블레오마이신, 스트렙토마이신, 하이그로마이신, 네오마이신, ZeocinTM 등에 내성을 부여하는 것들을 포함한다. 다른 선별가능한 마커는 블레오마이신-내성 유전자, 메탈로티오네인 유전자, 하이드로마이신 B-포스포트랜스퍼라제 유전자, AURI 유전자, 아데노신 데아미나제 유전자, 아미노글리코사이드 포스포트랜스퍼라제 유전자, 디하이드로폴레이트 리덕타제 유전자, 티미딘 키나제 유전자, 크산틴-구아닌 포스포리보실트랜스퍼라제 유전자 등을 포함한다.

[0124] pBR 및 pUC-유래된 플라스미드는 선별가능한 마커로서 박테리아 약물 내성 마커 *AMP^r* 또는 *BLA* 유전자를 포함한다[참조: Sutcliffe, J. G., *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 75:3737 (1978)]. *BLA* 유전자는 효소 Tem-1을 코딩하고, 이는 베타-락타마제로서 기능하고 베타-락탐 항생물질, 예를 들면, 좁은 범위의 세팔로스포린, 세파마이신 및 카바페넴(에르타페넴), 세파만돌 및 세포페라존, 및 테모실린을 제외한 모든 항-그램-네가티브-박테리아 페니실린에 대한 박테리아 내성에 기여한다.

[0125] 다른 유용한 선별가능한 마커는, 이로써 한정되지 않지만, *NAT1*, *PAT*, *AURI-C*, *PDR4*, *SMR1*, *CAT*, 마우스 *dhfr*, *HPH*, *DSDA*, *KAN^R* 및 *SH BLE* 유전자를 포함한다. 에스. 노우르세이(*S. noursei*)의 *NAT1* 유전자는 노우르세오트리신 N-아세틸트랜스퍼라제를 코딩하고, 노우르세오트리신에 내성을 부여한다. 에스. 비리도크로모게네스(*S. viridochromogenes*) Tu94로부터의 *PAT* 유전자는 포스포노트리신 N-아세틸트랜스퍼라제를 코딩하고 비알로포스에 내성을 부여한다. 에스. 세레비지애로부터의 *AURI-C* 유전자는 에로바시딘 A(AbA), 버딩 효모 에스. 세레비지애에 독성이 있는 아우에로바시디움 폴루란스(*Aerobasidium pullulans*)에 의해 생성된 항균 항생물질에 내성을 부여한다. *PDR4* 유전자는 세룰레닌에 내성을 부여한다. *SMR1* 유전자는 설포메투론 메틸에 내성을 부여한다. Tn9 트랜스포손으로부터의 *CAT* 코딩 서열은 클로람페니콜에 내성을 부여한다. 마우스 *dhfr* 유전자는 메토틱세이트에 내성을 부여한다. 클레브시엘라 뉴모니아(*Klebsiella pneumonia*)의 *HPH* 유전자는 하이드로마이신 B 포스포트랜스퍼라제를 코딩하고, 하이드로마이신 B에 내성을 부여한다. 이. 콜라이(*E. coli*)의 *DSDA* 유전자를 D-세린 데아미나제를 코딩하고, 단일 질소 공급원으로 D-세린을 사용하여 플레이트 상에서 효모가 성장하게 한다. Tn903 트랜스포손의 *KAN^R* 유전자는 아미노글리코사이드 포스포트랜스퍼라제를 코딩하고 G418에 내성을 부여한다. 스트렙토알로테이쿠스 힌두스타누스(*Streptoalloteichus hindustanus*)로부터의 *SH BLE* 유전자는 제오신 결합 단백질을 코딩하고, 제오신(블레오마이신)에 내성을 부여한다.

[0126] 다른 실시형태에서, 선별가능한 마커는 영양요구성 (auxotrophic) 마커이다. 영양요구성 마커는 세포가 필수 성분(통상 아미노산)을 합성하게 하고, 당해 필수 성분을 결여한 배지에서 성장하게 한다. 선별가능한 영양요구성 유전자 서열은, 예를 들면, 히스티딘의 존재하에 히스티딘 비함유 배지에서 성장하는 *hisD*이다. 일부 실시형태에서, 선별가능한 마커는 숙주 균주에서 영양 영양요구성을 구제한다. 이러한 실시형태에서, 숙주 균주는, 영양요구성 표현형을 유발하는 숙주의 아미노산 생합성 경로의 하나 이상의 유전자, 예를 들면, *HIS3*, *LEU2*, *LYS1*, *MET15* 및 *TRP1* 중의 기능 파괴, 또는 영양요구성 표현형을 유발하는 숙주의 뉴클레오티드 생합성 경로의 하나 이상의 유전자, 예를 들면, *ADE2* 및 *URA3* 중의 기능 파괴를 포함한다. 특정 실시형태에서, 숙주 세포는 *URA3* 유전자 중에 기능 파괴를 포함한다. 영양요구성 표현형을 유발하는 숙주 세포 중의 기능 파괴는 점 돌연변이, 부분 또는 완전 유전자 결실, 또는 뉴클레오티드의 부가 또는 치환일 수 있다. 아미노산 또는 뉴클레오티드 생합성 경로 내의 기능 파괴는 숙주 균주가 영양요구성 돌연변이로 되도록 하고, 이는, 원영양 (prototrophic) 야생형 세포와 대조적으로, 하나 이상의 영양소의 보충 없이 배지에서 최적 성장할 수 없다. 이어서, 숙주 균주에서 기능적 파괴된 생합성 유전자는, 예를 들면, 파괴된 생합성 유전자의 기능적 카피를 포함하는 하나 이상의 플라스미드를 도입하면, 이후에 구조될 수 있는 영양요구성 유전자 마커로서 작용할 수 있다.

[0127] 효모에서, 선별가능한 마커로서 *URA3*, *TRP1* 및 *LYS2* 유전자의 이용은, 포지티브 및 네가티브 선별 둘 다 가능하기 때문에, 현저한 잇점을 갖는다. 포지티브 선별은 *URA3*, *TRP1* 및 *LYS2* 돌연변이의 영양요구성 상보성에 의해 수행되는 반면, 네가티브 선별은 특정의 억제제 5-플루오로-오로트산(FOA), 5-플루오로안트라닐산 및 a-아미노아디프산(aAA)에 각각 기반하고, 이는 원영양 균주의 성장을 방지하지만 각각 *URA3*, *TRP1* 및 *LYS2* 돌연변이체의 성장을 가능하게 한다. *URA3* 유전자는 오로티딘-5' 포스페이트 데카복실라제, 우라실의 생합성에 요구되는 효소를 코딩한다. *ura3*-(또는 *ura5*-) 세포는, 모든 *URA3*+를 사멸시키지만 *ura3*- 세포를 사멸시키지 않는 FOA 함유 배지 상에서 선별될 수 있는데, 이는 FOA가 데카복실라제의 작용에 의해 독성 화합물 5-플루오로우라실로 전환되는 것으로 생각되기 때문이다. FOA 배지 상에서 네가티브 선별은 매우 식별성이며, 통상 10^{-2} 미만의 FOA-내성 콜로니가 Ura+이다. FOA 선별 절차는 돌연변이에 의해 반수체 균주에서 *ura3* 마커를 생성하기 위해 및 보다 중요하게는, *URA3*-함유 플라스미드를 갖지 않는 세포를 선별하기 위해 사용될 수 있다. *TRP1* 유전자는 트립토판 생합성에서 제3 단계를 촉매하는 포스포리보실안트라닐레이트 이소머라제를 코딩한다. 5-플루오로안트라닐산을 사용한 역선별은, 트립토판으로 안트라닐산의 전환에 요구되는 효소를 결여하고 따라서 5-플루오로안트라닐산에 내성이 있는 균주에 의한 대사길항을 수반한다. *LYS2* 유전자는 아미노아디페이트 리덕타제, 리신의 생합성에 요구되는 효소를 코딩한다. 정상 균주가 아닌, *Lys2*- 및 *lys5*- 돌연변이체는, 정상 질소 공급원을 결여하지만 리신 및 aAA를 함유하는 배지 상에서 성장한다. 명백하게는, *lys2* 및 *lys5* 돌연변이는, 고도의 aAA에 의해 형성되는 리신 생합성의 독성 중간체의 축적을 유발하지만, 이들 돌연변이는 여전히 질소 공급원으로 aAA를 사용할 수 있다. FOA 선별 공정과 유사하게, *LYS2*-함유 플라스미드는 편리하게는 *lys2* 숙주로부터 추방될 수 있다. 다른 실시형태에서, 선별가능한 마커는 영양요구성 돌연변이를 구조하는 것 이외의 마커이다.

[0128] 본원에 기재된 임의의 방법 및 조성물에 있어서, 리포터 유전자, 예를 들면, 형질전환된 콜로니, 또는 형광 단백질, 예를 들면, 녹색, 적색 및 황색 형광 단백질의 청색/백색 선별을 촉진하는 *lac Z* 리포터 유전자는, 하나 이상의 예비-재조합 단편으로부터 환상 발현 벡터를 성공적으로 조립할 수 있는 HR-적격 숙주 세포의 선별을 촉진하기 위해 선별가능한 마커 유전자로서 사용될 수 있다(도 3A 참조). 이들 실시형태에서, 선별 화합물, 예를 들면, 항생물질을 함유하는 배지에서 형질전환된 세포를 성장시키는 것이 아니라, 세포는 리포터의 발현을 가능하게 하기에 충분한 조건하에 성장시키고, 선별은 리포터의 시각, 비색 또는 형광 검출에 의해 수행할 수 있다. 숙주 세포의 약물-비함유 및 선별 압력-비함유 세포 유지는 다수의 잇점을 제공할 수 있다. 예를 들면, 선별 약물 및 기타 선별 압력 인자는 종종 돌연변이성이거나 달리는 세포의 생리기능을 간섭하여 세포-기반 검정에서 왜곡된 결과를 유도한다. 예를 들면, 선별 약물은 아포토시스에 대한 감수성을 감소시킬 수 있고[참조: Robinson et al., Biochemistry, 36(37):11169-11178 (1997)], DNA 수복 및 약물 대사를 증가시킬 수 있고[참조: Deffie et al., Cancer Res. 48(13):3595-3602 (1988)], 세포 pH를 증가시킬 수 있고[참조: Thiebaut et al., J Histochem Cytochem. 38(5):685-690 (1990); Roepe et al., Biochemistry. 32(41):11042-11056 (1993); Simon et al., Proc Natl Acad Sci USA. 91(3):1128-1132 (1994)], 리소좀 및 엔도솜 pH를 감소시킬 수 있고[참조: Schindler et al., Biochemistry. 35(9):2811-2817 (1996); Altan et al., J Exp Med. 187(10):1583-1598 (1998)], 원형질 막 잠재성을 감소시킬 수 있고[참조: Roepe et al., Biochemistry. 32(41):11042-11056 (1993)], 클로라이드[참조: Gill et al., Cell. 71(1):23-32 (1992)] 및 ATP[참조: Abraham et al., Proc Natl Acad Sci USA. 90(1):312-316 (1993)]에 대한 원형질 막 전도성을 증가시킬 수 있고, 소포 수송의 속도를 증가시킬 수 있다[참조: Altan et al., Proc Natl Acad Sci USA. 96(8):4432-4437 (1999)]. 따라서, 본원에 제공된 방법은 선별 압력에 의해 유발된 인공물을 함유하지 않는 스크리닝을 가능하게 하는 약물-비함유 선별을 사용하여 실시할 수 있다.

[0129] 유동 세포분석 세포 분류기는 형광단에 결합되고 마커 단백질에 친화성을 갖는 형광 마커 또는 단백질(예: 항체)의 발현에 포지티브인 세포를 단리하기 위해 사용될 수 있다. 일부 실시형태에서, 복수 주기의 분류가 수행될 수 있다. 한 가지 실시형태에서, 유동 세포분석 세포 분류기는 FACS 기계이다. 고-생산량 스크리닝과 상용성인 것들을 포함하는 기타 형광 플레이트 판독기가 또한 사용될 수 있다. MACS(자기 세포 분류)는, 예를 들면, 자기 비드에 결합되고 마커 단백질에 친화성을 갖는 단백질을 사용하여 숙주 세포를 선별하기 위해 사용될 수 있다. 이는 선별가능한 마커가, 예를 들면, 막 단백질, 막관통 단백질, 막 고정 단백질, 세포 표면 항원 또는 세포 표면 수용체(예: 사이토킨 수용체, 면역글로불린 수용체 계열 멤버, 리간드-게이트 이온 채널, 단백질 키나제 수용체, G-단백질 결합된 수용체(GPCR), 핵 호르몬 수용체 및 기타 수용체; CD14(단핵구), CD56(천연 킬러 세포), CD335(NKp46, 천연 킬러 세포), CD4(T 헬퍼 세포), CD8(세포독성 T 세포), CD1c(BDCA-1, 혈액 수치상 세포 서브세트), CD303(BDCA-2), CD304(BDCA-4, 혈액 수치상 세포 서브세트), NKp80(천연 킬러 세포, 감마/델타 T 세포, 작동자/메모리 T 세포), "B11"(Va24/Vb11; 비변형 천연 킬러 T 세포), CD137(활성화된 T 세포),

CD25(조절 T 세포) 또는 CD138(혈장 세포)에 대해 고갈된, CD4, CD8, CD19, CD25, CD45RA, CD45RO)를 코딩하는 경우에 특히 유용할 수 있다. 따라서, 일부 실시형태에서, 선별가능한 마커는 숙주 세포 표면 상에 표시된 단백질질을 포함하고, 이는, 예를 들면, 형광단 또는 비색 또는 기타 시각 관독물에 결합된 항체로 용이하게 검출할 수 있다.

[0130] 5.2.1.2 세포 배양

[0131] 본원에 기재된 방법의 일부 실시형태에서, 하나 이상의 예비-제조합 단편으로 형질전환된 숙주 세포는 환상 발현 벡터로부터 선별가능한 마커의 발현에 충분한 기간 동안 배양된다.

[0132] 선별가능한 마커가 약물 내성 마커인 일부 실시형태에서, 배양은, 선별가능한 배지에서 마커를 발현하는 세포의 생존을 지지할 수 있는 마커 단백질의 양을 생성하기에 충분한 기간 동안 수행된다. 바람직한 실시형태에서, 이들 조건은 또한 선별가능한 마커를 발현하지 않는 세포의 생존에 대해 선별한다. 선별 압력은 당해 기술분야의 숙련가에게 공지되어 있는 다양한 화합물 또는 처리를 사용하여 세포에 적용될 수 있다. 이론에 구속되지 않고서, 선별 압력은 숙주 세포를 성장, 세포 사이클의 진행 또는 생존성에 준최적화된 또는 유해한 조건에 노출시킴으로써 적용할 수 있고, 이에 의해 이들 조건에 인성 또는 내성이 있는 세포는 이들 조건에 인성 또는 내성이 없는 세포와 비교하기 위해 선별된다. 선별 압력을 가하거나 적용하기 위해 사용될 수 있는 조건은, 이로써 한정되지 않지만, 항생물질, 약물, 돌연변이원, 생물학적 구성 블록의 세포 성장 또는 합성을 지연 또는 중단시키는 화합물, RNA, DNA 또는 단백질 합성을 파괴하는 화합물, 세포 성장 또는 배양 배지로부터 세포 성장 및 생존에 필요한 영양소, 아미노산, 탄수화물 또는 화합물의 결핍 또는 제한, 세포 성장에 준최적화된 조건, 예를 들면, 준최적화 온도에서, 대기 조건(예를 들면, 이산화탄소 %, 산소 또는 질소 또는 습도) 또는 결핍된 배지 조건하에 세포의 성장 또는 유지 등의 처리를 포함한다. 사용되는 선별 압력의 수준은 당해 기술분야의 숙련가에 의해 결정될 수 있다. 이는, 예를 들면, 대조군 세포 및 내성 마커 또는 유전자를 포함하는 세포를 선별 압력의 수준, 용량, 농도 또는 처리, 및 목적하는 시간 범위(예: 1 내지 24시간, 1 내지 3일, 3 내지 5일, 4 내지 7일, 5 내지 14일, 1 내지 3주, 2 내지 6주)에 걸쳐 유일하게 또는 우선적으로 네가티브 세포를 선별하는 범위를 증가시켜 시험하는 경우, 사멸 곡선 실험을 수행함으로써 이루어질 수 있다. 사용될 수 있는 선별 압력의 정확한 수준, 농도, 용량 또는 처리는 사용되는 세포, 목적하는 특성 자체, 마커, 선별 압력에 내성 또는 인성을 부여하는 인자 또는 유전자 및 선별되는 세포에 요구되는 목적하는 특성의 수준에 의존하고, 당해 기술분야의 숙련가는 이들 고려사항에 기반하여 적절한 범위를 어떻게 결정하는지를 용이하게 인지할 것이다.

[0133] 배양은, 이로써 한정되지 않지만, 세포 배양 플레이트, 플라스크 또는 발효조를 포함하는 적합한 용기에서 적합 배양 배지로 수행할 수 있다. 일부 실시형태에서, 배양 배지는 동화가능한 탄소, 질소 및 인산 공급원을 포함하는 수성 배지이다. 이러한 배지는 또한 적절한 염, 미네랄, 금속 및 기타 영양소를 포함할 수 있다. 일부 실시형태에서, 선별제 이외에, 적합한 배지는 하나 이상의 추가 제제, 예를 들면, 유도물질(예: 유전자 생성물을 코딩하는 하나 이상의 뉴클레오티드 서열이 유도성 프로모터의 조절하에 있는 경우), 억제인자(예: 유전자 생성물을 코딩하는 하나 이상의 뉴클레오티드 서열이 억제성 프로모터의 조절하에 있는 경우)가 보충된다. 세포 배양의 유지 및 성장을 위한 재료 및 방법은 미생물 또는 발효 과학의 기술분야의 숙련가에게 공지되어 있다 [참조: Bailey *et al.*, Biochemical Engineering Fundamentals, second edition, McGraw Hill, New York, 1986]. 고려사항은, 숙주 세포, 발효 및 프로세스의 특정 요건에 따라, 적절한 배양 배지, pH, 온도, 호기성, 미호기성 또는 혐기성 조건에 대한 요건에 제공되어야 한다. 일부 실시형태에서, 배양은 목적하는 세포 밀도가 도달될 때까지 형질전환된 집단이 복수의 배가를 경험하기에 충분한 기간 동안 수행된다. 일부 실시형태에서, 배양은 배양이 수행되는 발효조 또는 용기에서 숙주 세포 집단이 0.01 내지 400의 세포 밀도(OD_{600})에 도달하기에 충분한 기간 동안 수행된다. 일부 실시형태에서, 배양은 적어도 0.01의 OD_{600} 이 도달될 때까지 수행된다. 일부 실시형태에서, 배양은 적어도 0.1의 OD_{600} 이 도달될 때까지 수행된다. 일부 실시형태에서, 배양은 적어도 1.0의 OD_{600} 이 도달될 때까지 수행된다. 일부 실시형태에서, 배양은 적어도 10의 OD_{600} 이 도달될 때까지 수행된다. 일부 실시형태에서, 배양은 적어도 100의 OD_{600} 이 도달될 때까지 수행된다. 일부 실시형태에서, 배양은 0.01 내지 100의 OD_{600} 이 도달될 때까지 수행된다. 일부 실시형태에서, 배양은 0.1 내지 10의 OD_{600} 이 도달될 때까지 수행된다. 일부 실시형태에서, 배양은 1 내지 100의 OD_{600} 이 도달될 때까지 수행된다. 다른 실시형태에서, 배양은 적어도 12, 24, 36, 48, 60, 72, 84, 96시간 또는 96시간 초과 동안 수행된다. 일부 실시형태에서, 배양은 3 내지 20일의 기간 동안 수행된다. 일부 실시형태에서, 배양은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20일 또는 20일 초과 동안 수행된다.

[0134] 본원에 기재된 방법의 일부 실시형태에서, 상기 방법은, 예를 들면, 선별된 숙주 세포가 목적하는 게놈 성분(들)을 포함하는 것으로 동정되는 경우, 숙주 세포로부터 환상 발현 벡터, 즉 플라스미드를 제거하는 단계를 추가로 포함한다. 플라스미드-기반 시스템은 일반적으로 세포에 외래 DNA를 유지하기 위해 플라스미드 상에 선별 압력을 필요로 한다. 예를 들면, 효모에서 대부분의 플라스미드는, 효모 세포가 통상 각 유사분열 후에 세포에 함유된 플라스미드의 10%를 상실하기 때문에, 비교적 불안정하다. 따라서, 일부 실시형태에서, 선별된 세포로부터 선별된 마커를 코딩하는 플라스미드의 제거는, 플라스미드가 집단으로부터 효과적으로 희석되도록 선별된 세포가 충분한 유사분열을 받게 함으로써 달성될 수 있다. 또는, 플라스미드-비함유 세포는 플라스미드의 부재를 선별함으로써, 예를 들면, 카운터-선별가능한 마커(예: URA3)를 선별하거나 선별 배지 및 비-선별 배지 둘 다 상에 동일한 콜로니를 플레이팅한 다음, 선별 배지 상에서 성장하지 않고 비-선별 배지 상에서 성장하는 콜로니를 선별함으로써 선별할 수 있다.

[0135] 5.2.2 외인성 공여체 핵산

[0136] 유리하게는, 삽입 폴리뉴클레오티드, 즉 공여체 DNA는 숙주 세포 게놈의 선택된 표적 부위 내로 하나 이상의 외인성 핵산 작제물의 삽입을 촉진시킨다. 바람직한 실시형태에서, 삽입 폴리뉴클레오티드는 제1 상동성 영역(HR1)_x 및 제2 상동성 영역(HR2)_x, 및 임의로 (HR1)_x 및 (HR2)_x 사이에 위치한 목적하는 핵산을 포함하는 외인성 핵산(ES)_x을 포함한다. 일부 실시형태에서, 삽입 폴리뉴클레오티드는 선형 DNA 분자이다. 다른 실시형태에서, 삽입 폴리뉴클레오티드는 환상 DNA 분자이다. 일부 실시형태에서, 삽입 폴리뉴클레오티드는 단일-가닥 DNA 분자, 즉 올리고뉴클레오티드이다. 다른 실시형태에서, 삽입 폴리뉴클레오티드는 이중-가닥 DNA 분자이다.

[0137] 삽입 폴리뉴클레오티드는 당해 기술분야의 숙련자에게 명백한 임의의 기술에 의해 생성될 수 있다. 특정 실시형태에서, 삽입 폴리뉴클레오티드는 당해 기술분야에 공지된 폴리머라제 연쇄 반응(PCR) 및 분자 클로닝 기술을 사용하여 생성된다[참조: *PCR Technology: Principles and Applications for DNA Amplification*, ed. HA Erlich, Stockton Press, New York, N.Y. (1989); Sambrook *et al.*, 2001, *Molecular Cloning - A Laboratory Manual*, 3rd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; *PCR Technology: Principles and Applications for DNA Amplification*, ed. HA Erlich, Stockton Press, New York, N.Y. (1989); U.S. Patent No. 8,110,360].

[0138] 5.2.2.1 게놈 삽입 서열

[0139] 바람직한 실시형태에서, 삽입 폴리뉴클레오티드는 제1 상동성 영역(HR1)_x 및 제2 상동성 영역(HR2)_x를 포함하는 외인성 핵산(ES)_x을 포함하고, 여기서 (HR1)_x 및 (HR2)_x는 숙주 세포 게놈 내의 선별된 표적 부위(TS)_x에서 숙주 세포 매개된 상동성 재조합을 개시할 수 있다. 상동성 재조합에 의해 게놈 내로 외인성 핵산을 삽입하기 위해, 삽입 폴리뉴클레오티드는 바람직하게는 하나의 말단에 (HR1)_x 및 다른 말단에 (HR2)_x를 포함한다. 일부 실시형태에서, (HR1)_x는 선별된 게놈 표적 부위(TS)_x의 5' 영역과 상동성이며, 상기 (HR2)_x는 선별된 표적 부위(TS)_x의 3' 영역과 상동성이다. 일부 실시형태에서, (HR1)_x는 선별된 게놈 표적 부위(TS)_x의 5' 영역에 약 70%, 75%, 80%, 85%, 90%, 95% 또는 100% 상동성이다. 일부 실시형태에서, (HR2)_x는 선별된 표적 부위(TS)_x의 3' 영역에 약 70%, 75%, 80%, 85%, 90%, 95% 또는 100% 상동성이다.

[0140] 특정 실시형태에서, (HR1)_x는 목적하는 핵산(D)_x에 대해 5'에 위치한다. 일부 실시형태에서, (HR1)_x는 (D)_x의 5' 말단에 바로 인접하여 위치한다. 일부 실시형태에서, (HR1)_x는 (D)_x의 5'의 상류에 위치한다. 특정 실시형태에서, (HR2)_x는 목적하는 핵산(D)_x에 대해 3'에 위치한다. 일부 실시형태에서, (HR2)_x는 (D)_x의 3' 말단에 바로 인접하여 위치한다. 일부 실시형태에서, (HR2)_x는 (D)_x의 3'의 하류에 위치한다.

[0141] 특정 게놈 유전자좌에서 삽입 폴리뉴클레오티드의 삽입에 영향을 줄 수 있는 특성으로는, 이로써 제한되지 않지만, 게놈 삽입 서열의 길이, 절제가능한 핵산 작제물의 전체 길이, 및 게놈 삽입 유전자좌의 뉴클레오티드 서열 또는 위치를 포함한다. 예를 들면, 숙주 세포 게놈에서 게놈 삽입 서열의 한 가닥과 특정 유전자좌의 한 가닥 사이의 효과적인 헤테로듀플렉스 형성은 상기 게놈 삽입 서열의 길이에 의존할 수 있다. 게놈 삽입 서열의 효과적인 길이 범위는 50 내지 5,000개 뉴클레오티드이다. 게놈 삽입 서열과 게놈 유전자좌 사이의 상동성의 효과적인 길이에 관해서는 문헌[참조: Hasty *et al.*, *Mol Cell Biol* 11:5586-91 (1991)]을 참조한다.

[0142] 일부 실시형태에서, $(HR1)_x$ 및 $(HR2)_x$ 는, 효모 게놈 유전자좌에서 외인성 핵산(ES)_x의 게놈 삽입을 가능하게 하는, 충분한 길이 및 서열 동일성의 뉴클레오티드 서열을 포함한다. 특정 실시형태에서, $(HR1)_x$ 및 $(HR2)_x$ 는 각각 독립적으로 약 50 내지 5,000개 뉴클레오티드로 이루어진다. 특정 실시형태에서, $(HR1)_x$ 및 $(HR2)_x$ 는 각각 독립적으로 약 100 내지 2,500개 뉴클레오티드로 이루어진다. 특정 실시형태에서, $(HR1)_x$ 및 $(HR2)_x$ 는 각각 독립적으로 약 100 내지 1,000개 뉴클레오티드로 이루어진다. 특정 실시형태에서, $(HR1)_x$ 및 $(HR2)_x$ 는 각각 독립적으로 약 250 내지 750개 뉴클레오티드로 이루어진다. 특정 실시형태에서, $(HR1)_x$ 및 $(HR2)_x$ 는 각각 독립적으로 약 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800, 3900, 4000, 4100, 4200, 4300, 4400, 4500, 4600, 4700, 4800, 4900 또는 5,000개 뉴클레오티드로 이루어진다. 일부 실시형태에서, $(HR1)_x$ 및 $(HR2)_x$ 는 각각 독립적으로 약 500개 뉴클레오티드로 이루어진다.

[0143] 5.2.2.2 목적하는 핵산

[0144] 일부 실시형태에서, 삽입 폴리뉴클레오티드는 추가로 목적하는 핵산(D)_x을 포함한다. 상기 목적하는 핵산은 당해 기술분야의 숙련가에 의해 유용하게 여겨지는 DNA 세그먼트일 수 있다. 예를 들면, 상기 DNA 세그먼트는 숙주 게놈으로 "녹킹 인(knocked in)"될 수 있는 목적하는 유전자를 포함할 수 있다. 다른 실시형태에서, DNA 세그먼트는 "녹아웃(knockout)" 작제물로서 기능하는데, 이는 숙주 세포 게놈의 표적 부위로 작제물이 삽입될 때 표적 유전자를 특이적으로 방해하여, 방해된 유전자가 비-기능성으로 되게 할 수 있다. 목적하는 핵산(D)_x의 유용한 예로는, 이로써 한정되지 않지만, 단백질-코딩 서열, 리포터 유전자, 형광 마커 코딩 서열, 프로모터, 인핸서, 터미네이터, 전사 활성화인자, 전사 억제인자, 전사 활성화인자 결합 부위, 인트론, 엑손, 폴리-A 테일, 다중 클로닝 부위, 핵 위치화 신호, mRNA 안정화 신호, 삽입 유전자좌, 에피토프 태그 코딩 서열, 분해 신호, 또는 다른 천연 또는 합성 DNA 분자를 포함한다. 일부 실시형태에서, (D)_x는 천연 기원일 수 있다. 달리는, (D)_x는 완전히 시험관내에서 생성된 합성 기원일 수 있다. 추가로, (D)_x는 단리된 천연 DNA 분자의 조합, 또는 단리된 천연 DNA 분자와 합성 DNA 분자의 조합을 포함할 수 있다. 예를 들면, (D)_x는 단백질 코딩 서열에 작동적으로 연결된 이종성 프로모터, 폴리-A 테일에 연결된 단백질 코딩 서열, 에피토프 태그 코딩 서열과 프레임 내에서 연결된 단백질 코딩 서열 등을 포함할 수 있다. 목적하는 핵산(D)_x는 목적하는 세포로부터 정제된 게놈 DNA 또는 이의 단편의 화학적 합성에 의해, cDNA 클로닝에 의해, 또는 클로닝에 의해, 또는 PCR 증폭 및 클로닝에 의해, 당해 기술분야에 공지된 표준 절차에 의해 클로닝된 DNA (예를 들면, DNA "라이브러리")로부터 수득될 수 있다[참조: Sambrook *et al.*, *Molecular Cloning, Laboratory Manual*, 3d. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (2001); Glover, D.M. (ed.), *DNA Cloning: A Practical Approach*, 2d. ed., MRL Press, Ltd., Oxford, U.K. (1995)]

[0145] 특정 실시형태에서, 목적하는 핵산(D)_x는 선별가능한 마커를 코딩하는 핵산을 포함하지 않는다. 이들 실시형태에서, 본원에서 기재된 방법에 의해 제공되는 삽입의 고효율은, 선별 배지 상에서 형질전환된 세포를 성장시킬 필요 없이, 삽입 사상의 스크리닝 및 동정을 가능하게 한다. 그러나, 선별 배지 상에서의 성장이 바람직한 다른 실시형태에서, 목적하는 핵산(D)_x는 외인성 핵산의 숙주 게놈으로의 삽입을 선별하는데 사용할 수 있는 선별 가능한 마커를 포함한다.

[0146] 목적하는 핵산(D)_x는 약 300개 뉴클레오티드 내지 최대 약 100백만개 뉴클레오티드 염기쌍 범위의 크기를 포함하는 임의의 크기일 수 있다. 목적하는 이러한 핵산은 하나 이상의 유전자 및/또는 이들의 연관 조절 영역을 포함할 수 있다. 목적하는 이들 핵산은 조직-특이적, 정규화 및 기질 cDNA 라이브러리를 포함하는 임의의 공급원, 예를 들면, 게놈 공급원 또는 cDNA 라이브러리로부터 유래할 수 있다. 게놈 공급원은 병원성 생물체, 예를 들면, 바이러스(예: HIV 및 간염 바이러스) 및 세포 병원체를 포함하는 다양한 생물체의 게놈(또는 이의 단편)을 포함한다. 더욱이, 목적하는 핵산은, 이들이 진핵생물 또는 원핵생물이든지, 임의의 식물 또는 임의의 동물을 포함하는 임의의 생물체로부터 수득할 수 있다. 특정 실시형태에서, 목적하는 핵산은 질환-연관 유전자인 유전자, 즉 질환과 상관되거나 질환을 유발하는 변경된 형태의 존재, 부재, 발현, 발현 결핍, 변형된 형태의 발현 또는 변경된 형태의 존재를 코딩한다.

[0147] 일부 실시형태에서, 목적하는 핵산은 숙주 세포의 표적화 대립인자의 점 돌연변이를 코딩하고, 이는 표적화된 대립인자에 미스센스 SNP(즉, "대립인자 교환")의 도입에 사용될 수 있다. 이러한 일부 실시형태에서, 대립인자 교환을 위한 뉴클레아제(예: CRISPR/Cas9 및 gRNA) 표적 부위의 선별은 OFR 내에서의 결실 또는 삽입보다 상당히 더 제한된다. 바람직한 실시형태에서, 뉴클레아제 절단 부위는 게놈에 고유해야 하고, 가능한 표적 뉴클레오티드에 근접해야 하며, 이에 의해 재조합은 공여체 DNA의 인접 서열이 아니라 돌연변이 서열을 도입할 것이다. 이는 절단 부위를 수복하는 재조합이 목적하는 SNP의 도입을 필요로 하지 않고, 이의 포함 가능성이 절단 부위로부터의 거리에 따라 감소될 것으로 예상되기 때문이다. 추가로, 최적 효율을 위해, 공여체 DNA는 이것이 뉴클레아제(예: CRISPR/Cas9 및 gRNA)에 대한 표적이 아니도록 디자인되어야 한다. 따라서, 공여체 DNA가 절단에 면역으로 되도록 하기 위해 및 동시에 재조합 사상이 목적하는 SNP를 포함할 기회를 향상시키기 위해, 이중성 블록 방법을 사용하여 침묵 돌연변이가 표적 부위와 점 돌연변이 사이의 코돈 사이에서 수행되어, 목적하는 SNP를 생략하는 재조합 사상의 가능성을 감소시킨다. 공여체 DNA는 중앙의 "이중성 블록"을 둘러싸는 인접 상동성을 갖도록 디자인될 수 있다. 이중성 블록은 침묵 돌연변이를 뉴클레아제 표적 부위를 둘러싸는 서열에 도입하고, 몇가지 목적으로 작용한다. 먼저, 이는, 공여체 DNA가 절단되지 않도록, 뉴클레아제(예: CRISPR-Cas9) 인식에 중요한 거점을 제거한다. 추가로, 이중성 블록의 삽입은 PCR에 의해 후보 클론을 동정하기 위한 신규 프라이머 결합 부위를 제공한다. 도 15는 "이중성 블록"의 문맥에서 점 돌연변이의 CRISPR/Cas9-매개된 도입을 위한 도식을 제공한다. 표적 아미노산을 박스로 되어 있고, 인접한 절단 부위는 절단 부위 및 PAM 서열(상부 패널)로 주석되어 있다. 침묵 코돈 변화 및 인접 상동성의 이중성 블록의 문맥에서 목적하는 점 돌연변이를 함유하는 공여체 DNA는 60-머 올리고(중간 패널)를 어닐링 및 확장시킴으로써 또는 보다 큰 클로닝된 작제물로 합성에 의해 생성할 수 있다. 공여체 DNA의 삽입은 목적하는 점 돌연변이(하부 패널)를 제공한다.

[0148] 5.2.3 뉴클레아제

[0149] 본원에서 제공되는 방법의 일부 실시형태에서, 숙주 세포 게놈을, 절단, 즉, 선별된 표적 부위 내의 지정 영역에서 절단을 유도할 수 있는, 하나 이상의 뉴클레아제와 접촉시킨다. 일부 실시형태에서, 절단은 단일-가닥 절단이고, 즉 표적 부위의 하나의 DNA 가닥(두 DNA 가닥이 아님)이 절단된다(즉, "닉"). 일부 실시형태에서, 절단은 이중-가닥 절단이다. 일부 실시형태에서, 절단 유도제는 특정 폴리뉴클레오티드 인식 서열을 인식 및/또는 결합하여, 상기 인식 서열에 또는 그 근처에 절단을 생성하는 임의의 체제이다. 절단 유도제의 예로는, 이로써 제한되지 않지만, 엔도뉴클레아제, 부위-특이적인 재조합효소, 트랜스포사제, 토포아이스오머라제, 및 징크 핑거 뉴클레아제를 포함하고, 이의 변형된 유도제, 변이체 및 단편을 포함한다.

[0150] 일부 실시형태에서, 하나 이상의 뉴클레아제의 각각은 선별된 표적 부위(TS)_x 내의 지정된 영역에서 절단을 유발할 수 있다. 일부 실시형태에서, 상기 뉴클레아제는, (HR1)_x 및 (HR2)_x가 각각 상동성을 공유하는 (TS)_x의 5' 영역과 3' 영역 사이에 위치한 영역에서 절단을 유도할 수 있다. 다른 실시형태에서, 상기 뉴클레아제는 (TS)_x의 5' 영역과 3' 영역의 상류 또는 하류에 위치한 영역에서 절단을 유발할 수 있다.

[0151] 인식 서열은 절단 유도제에 의해 특이적으로 인식 및/또는 결합되는 임의의 폴리뉴클레오티드 서열이다. 인식 부위 서열의 길이는 다양할 수 있으며, 예를 들면, 적어도 10, 12, 14, 16, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70개 이상의 뉴클레오티드 길이인 서열이다.

[0152] 일부 실시형태에서, 인식 서열은 회문식(palindromic)으로, 즉, 한쪽 가닥의 서열은 상보성 가닥의 반대 방향과 동일하게 판독된다. 일부 실시형태에서, 닉/절단 부위가 상기 인식 서열 내에 존재한다. 다른 실시형태에서, 상기 닉/절단 부위는 상기 인식 서열의 외부에 존재한다. 일부 실시형태에서, 절단은 블런트 말단을 생성한다. 다른 실시형태에서, 절단은 단일-가닥 오버행, 즉, "스티키성 말단"을 생성하고, 이는 5' 오버행 또는 3' 오버행일 수 있다.

[0153] 일부 실시형태에서, 선별된 표적 부위 내의 인식 서열은 상기 숙주 세포 게놈에 대해 내인성 또는 외인성일 수 있다. 인식 부위가 내인성 서열인 경우, 이는 천연, 또는 고유한 절단 유도제에 의해 인식되는 인식 서열일 수 있다. 달리, 내인성 인식 부위는, 내인성 인식 서열을 특이적으로 인식하도록 설계 또는 선별된 변형된 또는 조작된 절단 유도제에 의해 인식 및/또는 결합되어, 절단을 생성할 수 있다. 일부 실시형태에서, 변형된 절단 유도제는 고유의, 천연 절단 유도제로부터 유래된다. 다른 실시형태에서, 변형된 절단 유도제는 인공적으로 생성되거나 또는 합성된다. 이러한 변형된 또는 조작된 절단 유도제를 선별하는 방법은 당해 기술분야에 알려져

있다. 예를 들면, 단백질(들)의 아미노산 서열 변이체는 DNA에서 돌연변이에 의해 제조될 수 있다. 돌연변이 생성(mutagenesis) 및 뉴클레오티드 서열 변경 방법은, 예를 들면, 문헌[참조: Kunkel, (1985) *Proc Natl Acad Sci USA* 82:488-92; Kunkel, *et al.*, (1987) *Meth Enzymol* 154:367-82; 미국 특허 제4,873,192 호; Walker and Gaastra, eds. (1983) *Techniques in Molecular Biology* (MacMillan Publishing Company, New York)] 및 상기 문헌들에서 언급한 참조문헌들을 포함한다. 단백질의 생물학적 활성에 영향을 미치지 않을 아미노산 교환에 관한 지침은, 예를 들면, 문헌[참조: Dayhoff, *et al.*, (1978) *Atlas of Protein Sequences and Structure* (Natl Biomed Res Found, Washington, D.C.)]의 모델에 나타나 있다. 보존적 치환, 예를 들면, 하나의 아미노산을 유사한 특성을 갖는 또 다른 아미노산으로 바꾸는 것이 바람직할 수 있다. 보존적 결실, 삽입, 및 아미노산 치환은 단백질의 특성에 급진적 변화를 생성하지 않을 것으로 예상되며, 치환, 결실, 삽입, 또는 이들의 조합의 효과는 통상의 스크리닝 분석법으로 평가할 수 있다. 이중-가닥 절단 유도 활성에 대한 검정법이 알려져 있으며, 이는 일반적으로 인식 부위를 포함하는 DNA 기질에 대한 상기 제제의 전반적 활성 및 특이성을 측정한다.

[0154] 5.2.3.1 클러스터 조절 공간 짧은 회문 반복체(CRISPR)

[0155] 본원에 기재된 방법의 일부 실시형태에서, 뉴클레아제는 CRISPR/Cas-유래 RNA-유도 엔도뉴클레아제이다. CRISPR은 유형 II 원핵생물 CRISPR(clustered regularly interspersed short palindromic repeats) 적응 면역 시스템에 기반한 게놈 편집 도구이다. 진정세균 및 고세균에서 CRISPR 시스템은 침략 외래 DNA를 표적화 및 절단하기 위해 작은 RNA 및 CRISPR-연관(Cas) 엔도뉴클레아제를 사용한다[참조: Bhaya *et al.*, *Annu Rev Genet* 45:273-297 (2011); Terns *et al.*, *Curr Opin Microbiol* 14(3):321-327 (2011); and Wiedenheft *et al.*, *Nature* 482 (7385):331-338]. 박테리아에서, CRISPR 유전자좌는 스페이서로 불리우는 외인성 DNA(약 30bp 길이)의 세그먼트에 의해 분리된 일련의 반복체로 구성된다. 반복체-스페이스서 정렬은 긴 전구체로서 전사되고, CRISPR 뉴클레아제에 의해 절단된 표적 서열(또한 프로토스페이스서로서 공지됨)을 지정하는 작은 crRNA를 생성하기 위해 반복 서열 내에서 처리된다. 이어서, CRISPR 스페이서는 RNA 또는 DNA 수준에서 외인성 유전자 요소를 인식 및 침묵시키기 위해 사용된다. 절단을 위한 필수요소는 프로토스페이스서-인접 모티프(PAM)으로 공지된, 표적 영역의 3' 말단의 바로 하류의 서열 모티프이다. PAM은 표적 DNA에 존재하지만, 이를 표적화하는 crRNA에 존재하지 않는다.

[0156] 가장 단순한 CRISPR 시스템 중의 하나는 스트렙토코쿠스 피오그네스(*Streptococcus pyogenes*)로부터의 유형 II CRISPR 시스템이다. CRISPR-연관 Cas9 엔도뉴클레아제 및 2개의 작은 RNA, 표적-상보성 CRISPR RNA(crRNA); 및 거래 (transacting) crRNA(tracrRNA)는 외래 DNA의 RNA-유도 절단에 충분하다. Cas9 단백질, 즉 유형 II CRISPR-Cas 시스템의 홀마크 단백질은 RuvC 및 HNH 뉴클레아제에 상동성인 2개의 뉴클레아제 도메인을 함유하는 거대 단량체성 DNA 뉴클레아제이다. Cas9는 crRNA:tracrRNA 복합체에 의해 PAM(프로토스페이스서 인접 모티프) 서열 모티프에 인접한 DNA 표적 서열로 유도된다. tracrRNA에 대한 성숙 crRNA 염기쌍은 Cas9를 표적 DNA로 유도하는 2개-RNA 구조를 형성한다. crRNA-유도 서열에 상보성인 부위에서, Cas9 HNH 뉴클레아제 도메인은 상보성 가닥을 절단하고, Cas9 RuvC-유사 도메인은 비상보성 가닥을 절단하여, 표적 DNA에서 이중 가닥 절단을 제공한다[참조: Deltcheva *et al.*, *Nature* 47(7340):602-607 (2011)].

[0157] 최근의 연구는 crRNA:tracrRNA 복합체를 모방하는 단일 유도 RNA(gRNA) 키메라가 시험관내에서 부위 특이적 DNA 이중-가닥 절단을 도입하기 위해 Cas9를 유도하는 게놈 편집 도구로서 Cas9와 함께 사용될 수 있음을 나타낸다. 표적 게놈 내의 절단 특이성은 키메라 유도 RNA 분자(gRNA)의 스페이서-유사 잔기에 의해 결정되고, 이는 천연 crRNA:tracrRNA 복합체를 모방한다. 따라서, 기능적 CRISPR/Cas 시스템에서 성분의 최소 수는 2개: Cas9 및 sgRNA이다. 이의 5' 말단에 위치한 sgRNA 유도 서열은 DNA 표적 특이성을 부여한다. 따라서, 유도 서열을 변형시킴으로써 상이한 표적 특이성을 갖는 sgRNA를 생성할 수 있다. 유도 서열의 표준 길이는 20bp이다. 결과적으로, DNA 표적은 또한 20bp이고, 컨센서스 NGG를 수반하는 PAM 서열을 수반한다. 이러한 변형된 CRISPR 시스템의 사용은 시험관내[참조: Jinek *et al.*, *Science* 337(6096):816-821 (2012)], 포유동물 세포주[참조: Mali *et al.*, *Science* 339(6121):823-826 (2013), Jinek *et al.*, *Elife* 2:e00417 (2013); Cong *et al.*, *Science* 339(6121):819-823 (2013); and Cho *et al.*, *Nat Biotechnol* 31(3):230-232 (2013)], 박테리아[참조: Jiang *et al.*, *Nat Biotechnol* 31(3):233-239 (2013); and Gasiunas *et al.*, *Proc Natl Acad Sci USA* 109(39):E2579-E2586. (2012)], 효모[참조: DiCarlo *et al.*, *핵산 Res* 41(7):4336-4343 (2013)], 제브라피쉬[참조: Hwang *et al.*, *Nat Biotechnol* 31(3):227-229 (2013); and Chang *et al.*, *Cell Res* 23(4):465-472 (2013)], 마우스[참조: Wang *et al.*, *Cell* 153(4):910-918 (2013)] 및 식물[참조: Belhaj *et al.*, *Plant Methods* 9:39 (2013)]에서 입증되었다.

- [0158] Cas9 뉴클레아제는 (1) 이중성 숙주 내에서 증가된 발현을 위한 코돈 최적화; (2) 적절한 구획화를 위한 핵 국지화 신호(NLS)와의 융합; 및 (3) 뉴클레아제를 가닥-특이적 Nickase(nickage)로 전환시키기 위해 HNH 또는 RuvC 도메인의 부위 지향된 돌연변이유발에 의해 변형시킬 수 있다. RuvC- 또는 HNH-모티프에서 Cas9의 부위-지향된 돌연변이유발은 가닥 절단 특이성을 나타내고, 이에 의해 야생형 엔도뉴클레아제 이외에 DNA의 표적화 단일-가닥 절단을 가능하게 하는 2개의 가닥-특이적 Nickase를 제공한다[참조: Jinek et al., Science 337(6096):816-821 (2012), and Gasiunas et al., Proc Natl Acad Sci USA 109(39):E2579-E2586. (2012)] 징크 핑거 뉴클레아제 및 TALEN에 대해 보고된 바와 같이, 하나의 가닥만을 절단하는 Nickase로서 기능하도록 뉴클레아제를 변형시키는 것은 표적-외 절단으로부터 독성을 감소시키고, 또한 비-HR 메카니즘(예: NHEJ)을 통한 절단 수복의 속도를 저하시킬 수 있다[참조: Jinek et al., Science 337(6096):816-821 (2012)].
- [0159] 당해 기술분야에 공지된 CRISPR/Cas 시스템은 본원에 기재된 방법 및 조성물에서 뉴클레아제로서의 사용을 발견한다. 고도로 다양한 CRISPR-Cas 시스템은 3개의 주요 유형으로 분류되고, 이는 코어 요소 함량 및 서열에 기반하여 10개 아형으로 추가로 세분된다[참조: Makarova et al., Nat Rev Microbiol 9:467-77 (2011)]. 외래 핵산의 crRNA-매개된 침묵에 수반된 핵단백질 복합체의 구조적 조직 및 기능은 상이한 CRISPR/Cas 유형마다 상이하다[참조: Wiedenheft et al., Nature 482:331-338 (2012)]. 에스케리키아 콜라이(*Escherichia coli*)에 의해 예시된 유형 1-E 시스템에서, crRNA는 캐스케이드로 불리우는 다중아단위 작동자 복합체(항바이러스 방어를 위한 CRIPR-연관 복합체) 내로 도입되고[참조: Brouns et al., Science 321 : 960-4 (2008)], 이는 표적 DNA에 결합하고 서명 Cas3 단백질에 의한 분해를 유발한다[참조: Sinkunas et al., EMBO J 30:1335-2 (2011); Beloglazova et al., EMBO J 30:616-27 (2011)]. 설포로부스 설파타리쿠스(*Sulfolobus solfataricus*) 및 피로콕쿠스 푸리오수스(*Pyrococcus furiosus*)의 유형 III CRISPR/Cas 시스템에서, Cas RAMP 모듈(Cmr) 및 crRNA 복합체는 시험관내에서 합성 RNA를 인식하고 절단하며[참조: Hale et al., Mol Cell 45:292-302 (2012); Zhang et al., Mol Cell, 45:303-13 (2012)], 스태필로콕쿠스 에피데르미디스(*Staphylococcus epidermidis*)의 CRISPR/Cas 시스템은 생체내에서 DNA를 표적화한다[참조: Marraffini & Sontheimer, Science. 322:1843-5 (2008)]. 유형 II CRISPR/Cas 시스템, 보다 특히는 스트렙토콕쿠스 테르모필루스(*Streptococcus thermophilus*) DGCC7710의 CRISPR3/Cas 시스템[참조: Horvath & Barrangou, Science 327:167-70 (2010)]에 의한 DNA 침묵에 수반된 RNP 복합체는 4개 cas 유전자: *cas9*, *cas1*, *cas2* 및 *csn2*로 이루어지고, 이는 12개 반복-스페이서 단위의 상부에 위치되어 있다. Cas9(이전에 *cas5* 또는 *csn1*로 지칭됨)은 유형 II 시스템을 위한 서명 유전자이다[참조: Makarova et al., Nat Rev Microbiol 9:467-77 (2011)].
- [0160] 본원에 제공된 방법 및 조성물에서의 용도를 발견하는 CRISPR 시스템은 또한 국제 공개공보 제WO 2013/142578 A1호 및 제WO 2013/098244 A1호에 기재된 것들을 포함하고, 이의 내용은 이들의 전체가 본원에서 인용된다.
- [0161] **5.2.3.2 전사 활성화인자-유사 작동자 뉴클레아제(TALEN)**
- [0162] 본원에 제공된 방법의 일부 실시형태에서, 하나 이상의 뉴클레아제는 TAL-작동자 DNA 결합 도메인-뉴클레아제 융합 단백질(TALEN)이다. 잔토모나스(*Xanthomonas*) 속(genus)의 식물 병원성 박테리아의 TAL 작동자는 숙주 DNA에 결합하고 작동자-특이적 숙주 유전자를 활성화함으로써, 질환에서 중요한 역할을 하거나 또는 방어를 유발한다[참조: Gu et al. (2005) Nature 435:1122-5; Yang et al., (2006) Proc. Natl. Acad. Sci. USA 103:10503-8; Kay et al., (2007) Science 318:648-51; Sugio et al., (2007) Proc. Natl. Acad. Sci. USA 104:10720-5; Romer et al., (2007) Science 318:645-8; Boch et al., (2009) Science 326(5959):1509-12; and Moscou and Bogdanove, (2009) 326(5959):1501]. TAL 작동자는, 하나 이상의 탠덤 반복 도메인(tandem repeat domain)을 통해 서열-특이적 방식으로 DNA와 상호작용하는 DNA 결합 도메인을 포함한다. 상기 반복된 서열은 전형적으로 34개 아미노산을 포함하며, 상기 반복체는 전형적으로 서로 91 내지 100% 상동성이다. 상기 반복체의 다형성은 통상 12번 및 13번 위치에 위치하며, 12번 및 13번 반복 가변성-2잔기(diresidue)의 동일성과, TAL-작동자의 표적 서열 내의 인접 뉴클레오티드의 동일성 사이에 1:1 대응이 존재하는 것으로 보인다.
- [0163] TAL-작동자 DNA 결합 도메인은 목적하는 표적 서열에 결합하도록 변형되며, 예를 들면, II형 제한효소 엔도뉴클레아제의 뉴클레아제 도메인, 전형적으로 II형 제한효소 엔도뉴클레아제, 예를 들면, FokI의 비특정 절단 도메인에 융합될 수 있다[참조: Kim et al. (1996) Proc. Natl. Acad. Sci. USA 93:1156-1160 참조]. 다른 유용한 엔도뉴클레아제로는, 예를 들면, HhaI, HindIII, Nod, BbvCI, EcoRI, BglI, 및 A1wI을 포함할 수 있다. 따라서, 바람직한 실시형태에서, TALEN은, TALEN이 특정 뉴클레오티드 서열 내의 또는 이에 인접한 표적 DNA를 절단하도록, 복수의 TAL 작동자 반복 서열을 포함하는, 즉, 조합하여 표적 DNA 서열에서 특이적 뉴클레오티드 서열에 결합하는 TAL 작동자 도메인을 포함한다. 본원에서 제공하는 방법에 유용한 TALENs는 제WO10/079430호

및 미국 출원 공개 제2011/0145940호에 기재된 것들을 포함한다.

[0164] 일부 실시형태에서, 표적 DNA 내의 특이적 뉴클레오티드 서열에 결합하는 TAL 작동자 도메인은 10개 이상의 DNA 결합 반복체, 및 바람직하게는 15개 이상의 DNA 결합 반복체를 포함할 수 있다. 일부 실시형태에서, 각각의 DNA 결합 반복체는 반복 가변성-2잔기(RVD)를 포함하며, 이는 표적 DNA 서열의 염기쌍의 인식을 결정하며, 상기 각각의 DNA 결합 반복체는 표적 DNA 서열의 하나의 염기쌍을 인식하는데 관여하며, 상기 RVD는, C를 인식하는 HD; T를 인식하는 NG; A를 인식하는 NI; G 또는 A를 인식하는 NN; A 또는 C 또는 G 또는 T를 인식하는 NS; C 또는 T를 인식하는 N*(여기서, *는 RVD의 제2 위치에서의 갭을 나타냄); T를 인식하는 HG; T를 인식하는 H*(여기서, *는 RVD의 제2 위치에서의 갭을 나타냄); T를 인식하는 IG; G를 인식하는 NK; C를 인식하는 HA; C를 인식하는 ND; C를 인식하는 HI; G를 인식하는 HN; G를 인식하는 NA; G 또는 A를 인식하는 SN; 및 T를 인식하는 YG 중의 하나 이상을 포함한다.

[0165] 본원에서 제공하는 방법의 일부 실시형태에서, 하나 이상의 뉴클레아제는 부위-특이적 재조합효소이다. 재조합 효소로도 지칭되는 부위-특이적인 재조합효소는, 혼화성 재조합 부위 사이의 보존적 부위-특이적인 재조합을 촉매하며, 천연 폴리펩타이드 뿐만 아니라, 활성을 유지하는 이의 유도체, 변이체 및/또는 단편, 및 재조합 효소를 코딩하는 천연 폴리뉴클레오티드, 활성을 유지하는 유도체, 변이체, 및/또는 단편을 포함하는 폴리펩타이드이다. 부위-특이적인 재조합효소 및 이들의 인식 부위에 관해서는 문헌[참조: Sauer (1994) *Curr Op Biotechnol* 5:521-7; 및 Sadowski, (1993) *FASEB* 7:760-7]을 참조한다. 일부 실시형태에서, 재조합효소는 세린 재조합효소 또는 티로신 재조합효소이다. 일부 실시형태에서, 재조합효소는 인테그라제(Integrase) 또는 레솔라제(Resolvase) 계열 유래이다. 일부 실시형태에서, 재조합효소는 FLP, Cre, 람다 인테그라제, 및 R로 이루어진 그룹으로부터 선택되는 인테그라제이다. 인테그라제 계열의 다른 구성원들에 관해서는, 예를 들면, 문헌[참조: Esposito, *et al.*, (1997) *Nucleic Acids Res* 25:3605-14 및 Abremski, *et al.*, (1992) *Protein Eng* 5:87-91]를 참조한다. 키네틱스, 보조인자 상호작용 및 요건, 발현, 최적 상태, 및/또는 인지 부위 특이성을 변형시키고, 재조합효소 및 변이체의 활성을 스크리닝하는 방법이 공지되어 있다[참조: Miller, *et al.*, (1980) *Cell* 20:721-9; Lange-Gustafson and Nash, (1984) *J Biol Chem* 259:12724-32; Christ, *et al.*, (1998) *J Mol Biol* 288:825-36; Lorbach, *et al.*, (2000) *J Mol Biol* 296:1175-81; Vergunst, *et al.*, (2000) *Science* 290:979-82; Dorgai, *et al.*, (1995) *J Mol Biol* 252:178-88; Dorgai, *et al.*, (1998) *J Mol Biol* 277:1059-70; Yagu, *et al.*, (1995) *J Mol Biol* 252:163-7; Scimiente, *et al.*, (2001) *Nucleic Acids Res* 29:5044-51; Santoro and Schultze, (2002) *Proc Natl Acad Sci USA* 99:4185-90; Buchholz and Stewart, (2001) *Nat Biotechnol* 19:1047-52; Voziyanov, *et al.*, (2002) *Nucleic Acids Res* 30:1656-63; Voziyanov, *et al.*, (2003) *J Mol Biol* 326:65-76; Klippel, *et al.*, (1988) *EMBO J* 7:3983-9; Arnold, *et al.*, (1999) *EMBO J* 18:1407-14; WO03/08045; WO99/25840; 및 WO99/25841]. 인식 부위의 범위는 약 30개 뉴클레오티드의 최소 부위부터 수백개 뉴클레오티드에 이른다. 천연 발생 부위 및 변이체를 포함하여, 재조합효소에 대한 인식 부위를 사용할 수 있다. 변이체 인지 부위는 공지되어 있다[참조: Hoess, *et al.*, (1986) *Nucleic Acids Res* 14:2287-300; Albert, *et al.*, (1995) *Plant J* 7:649-59; Thomson, *et al.*, (2003) *Genesis* 36:162-7; Huang, *et al.*, (1991) *Nucleic Acids Res* 19:443-8; Siebler and Bode, (1997) *Biochemistry* 36:1740-7; Schlake and Bode, (1994) *Biochemistry* 33:12746-51; Thygarajan, *et al.*, (2001) *Mol Cell Biol* 21:3926-34; Umlauf and Cox, (1988) *EMBO J* 7:1845-52; Lee and Saito, (1998) *Gene* 216:55-65; WO01/23545; WO99/25821; WO99/25851; WO01/11058; WO01/07572 및 미국 특허 제5,888,732].

[0166] 본원에서 제공하는 방법의 일부 실시형태에서, 하나 이상의 뉴클레아제는 트랜스포사제이다. 트랜스포사제는, 게놈의 한 위치에서 또 다른 위치로 트랜스포존(transposon)이 전위되는 것을 매개하는 폴리펩타이드이다. 트랜스포사제는 전형적으로, 트랜스포존을 절제하기 위해 이중 가닥 절단을 유도하며, 서브말단 반복체를 인식하고, 절제된 트랜스포존의 말단들을 함께 연결하며, 일부 시스템에서는 전위 동안에 상기 말단들을 함께 연결하기 위해 다른 단백질들이 필요하다. 트랜스포존 및 트랜스포사제의 예는, 이로써 한정되지 않지만, 옥수수 유래의 Ac/Ds, Dt/rdt, Mu-M1/Mn, 및 Spm(En)/dSpm 요소, 금어초(snapdragon) 유래의 Tam 요소, 박테리오파지 유래의 Mu 트랜스포존, 박테리아 트랜스포존(Tn) 및 삽입 서열(IS), 효모의 Ty 요소(레트로트랜스포존), 아라비도시스(*Arabidopsis*) 유래의 Ta1 요소(레트로트랜스포존), 드로소필라(*Drosophila*) 유래의 P 요소 트랜스포존[참조: Gloor, *et al.*, (1991) *Science* 253:1110-1117], 드로소필라 유래의 Copia, Mariner 및 Minos 요소, 집파리 유래의 Hermes 요소, *트리크플루시아 니(Trichoplusia ni)* 유래의 PiggyBack 요소, *씨. 엘레강스(C. elegans)* 유래의 Tc1 요소, 및 마우스 유래의 IAP 요소(레트로트랜스포존)를 포함한다.

[0167] 5.2.3.3 징크 핑거 뉴클레아제(ZFN)

[0168] 본원에서 제공하는 방법의 일부 실시형태에서, 하나 이상의 뉴클레아제는 징크-핑거 뉴클레아제(ZFN)이다. ZFN은 징크 핑거 DNA 결합 도메인 및 절단 유도제 도메인으로 이루어진 조작된 절단 유도제이다. 조작된 ZFN은 2개의 징크 핑거 정렬(ZFA)로 이루어지며, 이들의 각각은 비-특정 엔도뉴클레아제의 단일 아단위, 예를 들면, FokI 효소 유래의 뉴클레아제 도메인에 융합되며, 이량체화시 활성화된다. 전형적으로, 단일 ZFA는 3 또는 4개의 징크 핑거 도메인으로 이루어지며, 이의 각각은 특정 뉴클레오티드 3중체(GGC, GAT 등)를 인식하도록 디자인된다. 따라서, "3-핑거" ZFA 2개로 이루어진 ZFN은 18개 염기쌍 표적 부위를 인식할 수 있으며; 18개 염기쌍 인식 서열은 일반적으로, 인간 및 식물과 같은 거대 게놈 내에서도 독특하다. 2개의 FokI 뉴클레아제 단량체의 공동-국지화 및 이량체화를 지향함으로써, ZFN은, 표적화된 유전자좌에서 DNA에 절단을 형성하는 기능성 부위-특이적인 엔도뉴클레아제를 생성한다.

[0169] 유용한 징크-핑거 뉴클레아제로는, 공지된 것들과, 본원에서 기재한 하나 이상의 표적 부위(TS)에 대한 특이성을 갖도록 조작된 것들을 포함한다. 징크 핑거 도메인은, 예를 들면, 숙주 세포 게놈의 표적 부위 내에서 선별된 폴리뉴클레오티드 인식 서열에 특이적으로 결합하는 폴리펩타이드를 디자인하는데 순응성이다. ZFN은, 비-특이적 엔도뉴클레아제 도메인, 예를 들면, HO 또는 FokI와 같은 II형 엔도뉴클레아제 유래의 뉴클레아제 도메인에 연결된 조작된 DNA-결합 징크 핑거 도메인으로 이루어진다. 또는, 조작된 징크 핑거 DNA 결합 도메인은 DNA 니킹(nicking)/절단 활성을 유지하는, 다른 절단 유도제 또는 이들의 유도체이다. 예를 들면, 이러한 유형의 융합을 사용하여, 상이한 표적 부위로 절단 유도제를 위치시키거나, Nick 또는 절단 부위의 위치를 변경하거나, 더 짧은 표적 부위로 유도제를 유도하거나, 또는 더 긴 표적 부위로 유도제를 유도할 수 있다. 일부 실시예에서, 징크 핑거 DNA 결합 도메인은 부위-특이적 재조합효소, 트랜스포사제, 또는 니킹/절단 활성을 보유하는 이들의 유도체에 융합된다. 추가적 관능성은 전사 활성화인자 도메인, 전사 억제인자 도메인, 및 메틸라제를 포함하는 징크-핑거 결합 도메인에 융합될 수 있다. 일부 실시형태에서, 뉴클레아제 도메인의 이량체화가 절단 활성화에 필요하다.

[0170] 각각의 징크 핑거는 표적 DNA 내에 3개의 연속 염기쌍을 인식한다. 예를 들면, 3개 핑거 도메인은 9개의 인접 뉴클레오티드의 서열을 인식하였으며, 뉴클레아제의 이량체화 요건과 함께, 징크 핑거 3중체의 2개 세트를 사용하여 18개의 뉴클레오티드 인식 서열에 결합시킨다. 유용한 디자이너 징크 핑거 모듈로는, 다양한 GNN 및 ANN 3중체를 인식하는 것들[참조: Dreier, *et al.*, (2001) *J Biol Chem* 276:29466-78; Dreier, *et al.*, (2000) *J Mol Biol* 303:489-502; Liu, *et al.*, (2002) *J Biol Chem* 277:3850-6], 뿐만 아니라 다양한 CNN 또는 TNN 3중체를 인식하는 것들[참조: Dreier, *et al.*, (2005) *J Biol Chem* 280:35588-97; Jamieson, *et al.*, (2003) *Nature Rev Drug Discov* 2:361-8]을 포함한다. 또한, 문헌[참조: Durai, *et al.*, (2005) *Nucleic Acids Res* 33:5978-90; Segal, (2002) *Methods* 26:76-83; Porteus and Carroll, (2005) *Nat Biotechnol* 23:967-73; Pabo, *et al.*, (2001) *Ann Rev Biochem* 70:313-40; Wolfe, *et al.*, (2000) *Ann Rev Biophys Biomol Struct* 29:183-212; Segal and Barbas, (2001) *Curr Opin Biotechnol* 12:632-7; Segal, *et al.*, (2003) *Biochemistry* 42:2137-48; Beerli and Barbas, (2002) *Nat Biotechnol* 20:135-41; Carroll, *et al.*, (2006) *Nature Protocols* 1:1329; Ordiz, *et al.*, (2002) *Proc Natl Acad Sci USA* 99:13290-5; Guan, *et al.*, (2002) *Proc Natl Acad Sci USA* 99:13296-301; WO2002099084; WO00/42219; WO02/42459; WO2003062455; US20030059767; 미국 특허 출원 공개 제2003/0108880호; 미국 특허 제6,140,466호, 제6,511,808호 및 제6,453,242호]를 참조한다. 유용한 징크-핑거 뉴클레아제는 또한 문헌[참조: WO03/080809; WO05/014791; WO05/084190; WO08/021207; WO09/042186; WO09/054985; 및 WO10/065123]에 기술된 것들을 포함한다.

[0171] 5.2.3.4 엔도뉴클레아제

[0172] 본원에서 제공하는 방법의 일부 실시형태에서, 하나 이상의 뉴클레아제는 엔도뉴클레아제이다. 엔도뉴클레아제는 폴리뉴클레오티드 사슬 내의 포스포다이에스테르 결합을 절단하는 효소로서, 염기에 손상을 주지 않으면서 특정 부위에서 DNA를 절단하는 제한효소 엔도뉴클레아제를 포함한다. 제한효소 엔도뉴클레아제는 I형, II형, III형, 및 IV형 엔도뉴클레아제를 포함하며, 서브타입을 추가로 포함한다. 제한효소 엔도뉴클레아제는, 예를 들면, REBASE 데이터베이스[참조: webpage at rebase.neb.com; Roberts, *et al.*, (2003) *Nucleic Acids Res* 31:418-20), Roberts, *et al.*, (2003) *Nucleic Acids Res* 31:1805-12, 및 Belfort, *et al.*, (2002) in *Mobile DNA II*, pp. 761-783, Eds. Craigie, *et al.*, ASM Press, Washington, D.C.]에서 추가로 기재 및 예시되어 있다.

[0173] 본원에서 사용된 바와 같이, 엔도뉴클레아제는 또한, 제한효소 엔도뉴클레아제와 같이 특정 인식 서열에 결합하여 절단하는 호밍 엔도뉴클레아제를 포함한다. 그러나, 호밍 엔도뉴클레아제에 대한 인식 부위는 전형적으로 더 길며, 예를 들면, 약 18bp 이상이다. 메가뉴클레아제로도 알려져 있는 호밍 엔도뉴클레아제는 보존된 서열

모티프를 토대로 하기 계열로 분류된다: LAGLIDAG(서열번호 1) 호밍 엔도뉴클레아제, HNH 호밍 엔도뉴클레아제, His-Cys box 호밍 엔도뉴클레아제, GIY-YIG(서열번호 2) 호밍 엔도뉴클레아제, 및 시아노박테리아 호밍 엔도뉴클레아제[참조: Stoddard, *Quarterly Review of Biophysics* 38(1): 49-95 (2006)]. 이들 계열은, 이들의 보존된 뉴클레아제 활성-부위 코어 모티프 및 촉매 메카니즘, 생물학적 및 계층 분포, 및 비-호밍 뉴클레아제 시스템과의 보다 광범위한 관계에서 크게 상이하다[참조: Guhan and Muniyappa (2003) *Crit Rev Biochem Mol Biol* 38:199-248; Lucas, *et al.*, (2001) *Nucleic Acids Res* 29:960-9; Jurica and Stoddard, (1999) *Cell Mol Life Sci* 55:1304-26; Stoddard, (2006) *Q Rev Biophys* 38:49-95; 및 Moure, *et al.*, (2002) *Nat Struct Biol* 9:764]. 이들 계열의 유용한 특정 호밍 엔도뉴클레아제의 예는, 이로써 한정되지 않지만, I-CreI[참조: Rochaix *et al.*, *Nucleic Acids Res.* 13: 975-984 (1985), I-MsoI (Lucas *et al.*, *Nucleic Acids Res.* 29: 960-969 (2001)], I-SceI[참조: Foury *et al.*, *FEBS Lett.* 440: 325-331 (1998)], I-SceIV[참조: Moran *et al.*, *Nucleic Acids Res.* 20: 4069-4076 (1992)], H-DreI[참조: Chevalier *et al.*, *Mol. Cell* 10: 895-905 (2002)], I-HmuI[참조: Goodrich-Blair *et al.*, *Cell* 63: 417-424 (1990); Goodrich-Blair *et al.*, *Cell* 84: 211-221 (1996)], I-PpoI[참조: Muscarella *et al.*, *Mol. Cell. Biol.* 10: 3386-3396 (1990)], I-DirI[참조: Johansen *et al.*, *Cell* 76: 725-734 (1994); Johansen, *Nucleic Acids Res.* 21: 4405 (1993)], I-NjaI[참조: Elde *et al.*, *Eur. J. Biochem.* 259: 281-288 (1999); De Jonckheere *et al.*, *J. Eukaryot. Microbiol.* 41: 457-463 (1994)], I-NanI[참조: Elde *et al.*, *S. Eur. J. Biochem.* 259: 281-288 (1999); De Jonckheere *et al.*, *J. Eukaryot. Microbiol.* 41: 457-463 (1994)], I-NitI[참조: De Jonckheere *et al.*, *J. Eukaryot. Microbiol.* 41: 457-463 (1994); Elde *et al.*, *Eur. J. Biochem.* 259: 281-288 (1999)], I-TevI[참조: Chu *et al.*, *Cell* 45: 157-166 (1986)], I-TevII[참조: Tomaschewski *et al.*, *Nucleic Acids Res.* 15: 3632-3633 (1987)], I-TevIII[참조: Eddy *et al.*, *Genes Dev.* 5: 1032-1041 (1991)], F-TevI[참조: Fujisawa *et al.*, *Nucleic Acids Res.* 13: 7473-7481 (1985)], F-TevII[참조: Kadyrov *et al.*, *Dokl. Biochem.* 339: 145-147 (1994); Kaliman, *Nucleic Acids Res.* 18: 4277 (1990)], F-CphI[참조: Zeng *et al.*, *Curr. Biol.* 19: 218-222 (2009)], PI-MgaI[참조: Saves *et al.*, *Nucleic Acids Res.* 29:4310-4318 (2001)], I-CsmI[참조: Colleaux *et al.*, *Mol. Gen. Genet.* 223:288-296 (1990)], I-CeuI[참조: Turmel *et al.*, *J. Mol. Biol.* 218: 293-311 (1991)], 및 PI-SceI[참조: Hirata *et al.*, *J. Biol. Chem.* 265: 6726-6733 (1990)]를 포함한다.

[0174] 본원에서 제공되는 방법의 일부 실시형태에서, 호밍 엔도뉴클레아제의 천연 변이체, 및/또는 조작된 유도체를 사용한다. 키네틱스, 보조인자 상호작용, 발현, 최적 조건 및/또는 인지 부위 특이성, 및 활성 스크리닝을 변형시키는 방법은 공지되어 있다[참조: Epinat, *et al.*, (2003) *Nucleic Acids Res* 31:2952-62; Chevalier, *et al.*, (2002) *Mol Cell* 10:895-905; Gimble, *et al.*, (2003) *Mol Biol* 334:993-1008; Seligman, *et al.*, (2002) *Nucleic Acids Res* 30:3870-9; Sussman, *et al.*, (2004) *J Mol Biol* 342:31-41; Rosen, *et al.*, (2006) *Nucleic Acids Res* 34:4791-800; Chames, *et al.*, (2005) *Nucleic Acids Res* 33:e178; Smith, *et al.*, (2006) *Nucleic Acids Res* 34:e149; Gruen, *et al.*, (2002) *Nucleic Acids Res* 30:e29; Chen and Zhao, (2005) *Nucleic Acids Res* 33:e154; WO2005105989; WO2003078619; WO2006097854; WO2006097853; WO2006097784; 및 WO2004031346]. 유용한 호밍 엔도뉴클레아제는 또한 문헌[참조: WO04/067736; WO04/067753; WO06/097784; WO06/097853; WO06/097854; WO07/034262; WO07/049095; WO07/049156; WO07/057781; WO07/060495; WO08/152524; WO09/001159; WO09/095742; WO09/095793; WO10/001189; WO10/015899; 및 WO10/046786]에 기술한 것들을 포함한다.

[0175] 임의의 호밍 엔도뉴클레아제는 이중-가닥 절단 유도체로서 사용될 수 있고, 이로써 한정되지 않지만, H-DreI, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAIP, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-TliI, I-PpoI, Pi-PspI, F-SceI, F-SceII, F-SuvI, F-CphI, F-TevI, F-TevII, I-Amal, 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-NclIP, 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-Ssp68031, 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-MgaI, 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 또는 PI-TliII, 또는 이의 임의의 변이체 또는 유도체를 포함한다.

[0176] 일부 실시형태에서, 엔도뉴클레아제는 고유한 또는 내인성 인식 서열에 결합한다. 다른 실시형태에서, 엔도뉴클레아제는, 비-고유한 또는 외인성 인식 서열에 결합하고 고유한 또는 내인성 인식 서열에는 결합하지 않는,

변형된 엔도뉴클레아제이다.

[0177] 5.2.3.5 게놈 표적 부위

[0178] 본원에서 제공하는 방법에서, 뉴클레아제는, 게놈 표적 부위 근처 또는 그 내에 이중-가닥 절단을 유발할 수 있는 숙주 세포로 도입되며, 상기 절단 부위에서 또는 그 근처에서 상동성 재조합의 빈도를 크게 증가시킨다. 바람직한 실시형태에서, 뉴클레아제에 대한 인식 서열은 숙주 세포 게놈 내의 표적 부위에만 존재하여, 상기 뉴클레아제에 의한 표적-외 게놈 결합 및 절단을 최소화한다.

[0179] 일부 실시형태에서, 게놈 표적 부위는 숙주 세포에 대해 내인성, 예를 들면, 고유한 유전자좌이다. 일부 실시형태에서, 고유한 게놈 표적 부위는 본원에서 제공하는 삽입 방법에 사용되는 뉴클레아제의 유형에 따라 선택된다.

[0180] 이용되는 뉴클레아제가 CRISPR/Cas-유래 RNA-유도된 엔도뉴클레아제인 경우, 최적 표적 부위는 사용되는 특정 CRISPR-Cas 엔도뉴클레아제의 표적 인식을 위한 요건에 따라 선별될 수 있다. 예를 들면, Cas9 표적 인식은 표적 DNA 중의 "프로토스페이서" 서열 및 crRNA 중의 잔류 스페이서 서열 사이에서 상보성의 검출시에 발생한다. Cas9는 정확한 프로토스페이서-인접 모티프(PAM)이 또한 3' 말단에 존재하는 경우에만 DNA를 절단한다. 상이한 유형 II 시스템은 상이한 PAM 요건을 갖는다. 에스. 피오게네스(*S. pyogenes*) 시스템은 NGG 서열을 필요로 하고, N은 임의의 뉴클레오티드일 수 있다. 에스. 테르모필루스(*S. thermophilus*) 유형 II 시스템은 각각 NGGNG 및 NNAGAAW를 필요로 하고, 상이한 에스. 무탄스(*S. mutans*) 시스템은 NGG 또는 NAAR에 내성이다. 생물정보 분석은, 새로운 PAM을 동정하고 CRISPR-표적화가능한 서열을 증식시키기 위해 사용할 수 있는 다양한 박테리아에서 CRISPR 유전자좌의 광범위한 데이터베이스를 생성했다[참조: Rho *et al.*, *PLoS Genet.* 8, e1002441 (2012); and D. T. Pride *et al.*, *Genome Res.* 21, 126 (2011)]. 에스. 테르모필루스(*S. thermophilus*)에서, Cas9는 프로토스페이서의 상류에 블런트-말단 이중-가닥 절단 3bp를 생성하고, 이 공정은 Cas9 단백질 중의 2개 촉매적 도메인: DNA의 상보성 가닥을 절단하는 HNH 도메인 및 비-상보성 가닥을 절단하는 RuvC-유사 도메인에 의해 매개된다.

[0181] 사용되는 뉴클레아제가 징크 핑거 뉴클레아제인 경우, 최적의 표적 부위는 공개적으로 이용가능한 온라인 리소스를 이용해 선택될 수 있다[참조: Reyon *et al.*, *BMC Genomes* 12:83 (2011), 이는 이의 전체가 참조로서 본원에 원용됨]. 예를 들면, 올리고머화된 풀 조작(Oligomerized Pool Engineering, OPEN)은 징크 핑거 정렬을 높은 특이성 및 생체내 기능성으로 조작하기 위한 매우 확실(robust)하고 공개적으로 이용가능한 프로토콜이고, 식물, 제브라피쉬, 및 인간의 체세포 및 만능 줄기 세포에서 효율적으로 기능하는 ZFN을 생성하는데 성공적으로 사용되어 왔다. OPEN은, 후보 ZFA의 예비-구축된 랜덤화된 풀을 스크리닝하여 목적하는 표적 서열에 대해 높은 친화성 및 특이성을 갖는 것들을 동정하는, 선별-기반 방법이다. ZFNGenome은 OPEN-생성된 ZFN에 대한 잠재적인 표적 부위를 동정 및 시각화하기 위한, GBrowse-기반 도구이다. ZFNGenome은 모델 생물의 서열화되고 주석된 게놈에서 잠재적 ZFN 표적 부위의 개요를 제공한다. ZFNGenome은 현재 총 1억 1600만개 초과인 잠재적 ZFN 표적 부위를 포함하며, 7가지의 모델 생물체인 에스. 세레비시아(*S. cerevisiae*), 씨. 레인하르트티(*C. reinhardtii*), 에이. 탈리아나(*A. thaliana*), 디. 멜라노가스터(*D. melanogaster*), 디. 레리오(*D. rerio*), 씨. 엘레강스(*C. elegans*), 및 에이치. 사피엔스(*H. sapiens*)의 완전히 서열분석된 게놈 내에서 지도화된다. 부가적인 모델 생물체로는 3가지 식물 종인 글리시네 막스(*Glycine max*)(대두), 오리자 사티바(*Oryza sativa*)(쌀), 제아 마이스(*Zea mays*)(옥수수)를 포함하며, 3가지 동물 종인 트리볼리움 카스타네움(*Tribolium castaneum*)(붉은밀가루갑충(red flour beetle)), 무스 무스쿨루스(*Mus musculus*)(마우스), 라투스 노르베기쿠스(*Rattus norvegicus*)(갈색 래트)가 장래에 추가될 것이다. ZFNGenome은 전사 개시 부위(들)에 관한 이의 염색체의 자리 및 위치를 포함하여 각각의 잠재적 ZFN 표적 부위에 대한 정보를 제공한다. 사용자들은 몇몇 서로 다른 범주를 사용하여 ZFNGenome에 조회할 수 있다(예를 들면, 유전자 ID, 전사체 ID, 표적 부위 서열).

[0182] 사용되는 뉴클레아제가 TAL-작동자 뉴클레아제인 경우, 일부 실시형태에서, 최적의 표적 부위는 문헌[참조: Sanjana *et al.*, *Nature Protocols*, 7:171-192 (2012), 원용에 의해 본원에 포함됨]에 기재된 방법에 따라 선별할 수 있다. 간략하게는, TALEN은 이량체로서 기능하며, 좌측 및 우측 TALEN이라고도 지칭하는 한 쌍의 TALEN은 DNA의 반대측 가닥들의 서열을 표적화한다. TALEN은 TALE DNA-결합 도메인 및 단량체성 FokI 촉매성 도메인의 융합체로서 조작된다. FokI 이량체화를 촉진하기 위해, 좌측 및 우측 TALEN 표적 부위는 대략 14-20 염기로 간격을 두어 선택된다. 따라서, 각각이 20-bp 서열을 표적화하는 한 쌍의 TALEN의 경우, 최적의 표적 부위는 5'-TN¹⁹N¹⁴⁻²⁰N¹⁹A-3'를 형성하며, 여기서, 좌측 TALEN은 5'-TN¹⁹-3'를 표적화하고, 우측 TALEN은 5'-N¹⁹A-3' (N = A, G, T 또는 C)의 안티센스 가닥을 표적화한다.

[0183] 본원에서 제공하는 방법의 다른 실시형태에서, 게놈 표적 부위는 숙주 세포에 대해 외인성이다. 예를 들면, 하나 이상의 게놈 표적 부위는, 본원에서 기재된 삽입 방법을 수행하기 전에, 전형적인 방법, 예를 들면, 유전자 표적화를 사용하여 숙주 세포 게놈으로 조작될 수 있다. 일부 실시형태에서, 동일 표적 서열의 다중 카피를 상이한 유전자좌에서 숙주 세포 게놈으로 조작하여, 표적 서열을 특이적으로 인식하는 단일 뉴클레아제만을 사용하는 동시 다중 삽입 사상을 촉진한다. 다른 실시형태에서, 복수의 상이한 표적 서열은 상이한 유전자좌에서 숙주 세포 게놈으로 조작된다. 일부 실시형태에서, 조작된 표적 부위는, 숙주 세포의 고유한 게놈에서는 나타나지 않는 표적 서열을 포함한다. 예를 들면, 호밍 엔도뉴클레아제는 통상 인트론 또는 인테인(intein)에 매립된 거대 인지 부위(12-40 bp)를 표적화하며, 이와 같이, 이들의 인식 부위는 극히 드물며, 이들 부위 중 어느 것도 또는 소수만이 포유동물-크기의 게놈에 존재한다. 따라서, 일부 실시형태에서, 외인성 게놈 표적 부위는 호밍 엔도뉴클레아제에 대한 인식 서열이다. 일부 실시형태에서, 호밍 뉴클레아제는, H-DreI, 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-CrepsbIIVP, I-TliI, I-PpoI, Pi-PspI, F-SceI, F-SceII, F-SuvI, F-CphI, F-TevI, F-TevII, I-AmaI, I-AniI, I-ChuI, I-CmoI, 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-NclIP, 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-Ssp68031, I-SthPhiIP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdeIP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MgaI, 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, 또는 PI-TliII, 또는 이들의 임의의 변이체 또는 유도체로 이루어진 그룹으로부터 선택된다. 특정 실시형태에서, 상기 외인성 게놈 표적 부위는 I-SceI, VDE(PI-SceI), F-CphI, PI-MgaI 또는 PI-MtuII에 대한 인식 서열이며, 이들은 각각 하기에 제공된다.

표 1

표 1: 호밍 엔도뉴클레아제(homing endonuclease) 선별을 위한 인지 및 절단 부위

뉴클레아제	인지 서열
I-SceI	TAGGGATAACAGGGTAAT (서열번호:52)
VDE (PI-SceI)	TATGTCGGGTGCGGAGAAAGAGGTAATGAAA (서열번호:53)
F-CphI	GATGCACGAGCGCAACGCTCACAA (서열번호:54)
PI-MgaI	GCGTAGCTGCCCAGTATGAGTCAG (서열번호:55)
PI-MtuII	ACGTGCACTACGTAGAGGGTCGCACCGCACCGATCTACAA (서열번호:56)

[0184]

[0185] 5.2.3.6 전달

[0186] 일부 실시형태에서, 본원에서 제공되는 방법에 유용한 하나 이상의 뉴클레아제는, 정제된 단백질로서 숙주 세포로 제공, 예를 들면, 전달될 수 있다. 다른 실시형태에서, 상기 하나 이상의 뉴클레아제는 뉴클레아제를 코딩하는 핵산을 포함하는 폴리뉴클레오티드(들)를 통해 제공된다. 다른 실시형태에서, 상기 하나 이상의 뉴클레아제는, 숙주 세포의 핵에서 직접 번역될 수 있는 정제된 RNA로서 숙주 세포로 도입된다.

[0187] 특정 실시형태에서, 삽입 폴리뉴클레오티드, 뉴클레아제를 코딩하는 폴리뉴클레오티드, 또는 상술한 바와 같은 정제된 뉴클레아제 단백질, 또는 이들의 조합은 통상의 기법을 사용하여 숙주 세포 내로 도입하여, 당해 기술분야에 공지된 세포 내로 외인성 단백질 및/또는 핵산을 도입할 수 있다. 이러한 방법은, 이로써 한정되지 않지만, 용액으로부터 세포에 의해 분자를 직접 취하는 것, 또는 예를 들면, 리포솜 또는 면역리포솜을 사용하는 리

포펙틴(lipofection); 입자-매개성 형질감염 등을 통한 촉진화된 취득 (uptake)을 포함한다[참조: 미국 특허 제 5,272,065호; Goeddel *et al.*, eds, 1990, *Methods in Enzymology*, vol. 185, Academic Press, Inc., CA; Krieger, 1990, *Gene Transfer and Expression -- A Laboratory Manual*, Stockton Press, NY; Sambrook *et al.*, 1989, *Molecular Cloning -- A Laboratory Manual*, Cold Spring Harbor Laboratory, NY; and Ausubel *et al.*, eds., Current Edition, *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley Interscience, NY]. 세포를 형질전환시키는 특정 방법들은 당해 기술분야에 공지되어 있다[참조: Hinnen *et al.*, *Proc. Natl. Acad. Sci. USA* 75:1292-3 (1978); Cregg *et al.*, *Mol. Cell. Biol.* 5:3376-3385 (1985)]. 예시적 기술로는, 이로써 한정되지 않지만, 스페로플라스팅(spheroplasting), 전기천공, PEG 1000 매개된 형질전환, 및 리튬 아세테이트 또는 리튬 클로라이드 매개된 형질전환을 포함한다.

[0188] 일부 실시형태에서, 바이오리스틱스(biolistics)를 사용하여, 삽입 폴리뉴클레오티드, 뉴클레아제를 코딩하는 폴리뉴클레오티드, 정제된 뉴클레아제 단백질, 또는 이들의 조합을 숙주 세포, 특히 식물과 같이 통상적인 기술을 사용하여 형질전환/형질감염되기 어려운 숙주 세포로 도입한다. 바이오리스틱스는, 형질전환 반응물을 현미경 금 입자에 결합시킨 다음, 상기 입자들을 표적 세포에서 압축 가스를 사용하여 추진함으로써, 작동한다.

[0189] 일부 실시형태에서, 뉴클레아제를 코딩하는 핵산을 포함하는 폴리뉴클레오티드는 숙주 세포에서 뉴클레아제가 발현될 수 있게 하는 발현 벡터이다. 적절한 발현 벡터는, 이로써 한정되지 않지만, 에스케리키아 콜라이(*Escherichia coli*), 효모, 또는 포유류 세포에서 유전자를 발현하는데 사용되는 것으로 공지된 것들을 포함한다. 에스케리키아 콜라이 발현 벡터의 예는, 이로써 한정되지 않지만, pSCM525, pDIC73, pSCM351, 및 pSCM353을 포함한다. 효모 발현 벡터의 예는, 이로써 한정되지 않지만, pPEX7 및 pPEX408을 포함한다. 적절한 발현 벡터의 다른 예는, CEN.ARS 서열 및 효모 선별가능한 마커를 포함하는 효모-에스케리키아 콜라이 pRS 시리즈의 서를 벡터; 및 2 μ 플라스미드를 포함한다. 일부 실시형태에서, 뉴클레아제를 코딩하는 폴리뉴클레오티드는, 천연 폴리뉴클레오티드 서열과 비교하여 숙주 세포에서의 사용 빈도가 더 높은 코돈을 치환하도록 변형될 수 있다. 예를 들면, 뉴클레아제를 코딩하는 폴리뉴클레오티드는, 천연 폴리뉴클레오티드 서열과 비교하여, 에스. 세레비지에에서 사용 빈도가 더 높은 코돈을 치환하도록 변형될 수 있다.

[0190] 뉴클레아제가 각각의 단량체의 개별 발현을 필요로 하는 헤테로이량체로서 기능하는 일부 실시형태에서, 징크 핑거 뉴클레아제 및 TAL-작동자 뉴클레아제의 경우에서와 같이, 상기 헤테로이량체의 각각의 단량체는 동일한 발현 플라스미드, 또는 상이한 플라스미드로부터 발현될 수 있다. 다중 뉴클레아제가 상이한 표적 부위에서 이중-가닥 절단을 할 수 있도록 세포에 도입되는 실시형태에서, 상기 뉴클레아제는 단일 플라스미드 또는 개별 플라스미드에서 코딩될 수 있다.

[0191] 특정 실시형태에서, 뉴클레아제 발현 벡터는, 상기 발현 벡터를 포함하는 숙주 세포를 선별할 수 있게 하는 선별가능한 마커를 추가로 포함한다. 이러한 선별은, 뉴클레아제가 충분한 양으로 발현하는데 필요한 기간 동안, 예를 들면, 12, 24, 36, 48, 60, 72, 84, 96시간, 또는 96시간 초과 동안, 숙주 세포에서 벡터가 유지되는 것을 도울 수 있으며, 이후 상기 숙주 세포는 발현 벡터가 더 이상 유지되지 않는 조건하에 배양될 수 있다. 특정 실시형태에서, 상기 선별가능한 마커는, URA3, 하이그로마이신 B 포스포트랜스퍼라제, 아미노글라이코시드 포스포트랜스퍼라제, 제오신 내성, 및 포스포노트리신 N-아세틸트랜스퍼라제로 이루어진 그룹으로부터 선택된다. 일부 실시형태에서, 뉴클레아제 발현 벡터는, 하나 이상의 공여체 핵산 분자의 삽입, 이어서 발현 벡터를 함유하지 않는 숙주 세포를 선별할 수 있게 하는 카운터-선별가능한 마커를 포함할 수 있다. 사용되는 뉴클레아제 발현 벡터는 또한, 선별 마커를 갖지 않는 일시적 벡터일 수 있으며, 또는 선별되지 않는 벡터이다. 특정 실시형태에서, 일시적 뉴클레아제 발현 벡터를 포함하는 숙주 세포의 자손은 시간 경과에 따라 상기 벡터를 상실한다.

[0192] 특정 실시형태에서, 발현 벡터는 전사 종결 서열, 및 뉴클레아제를 코딩하는 뉴클레오티드 서열에 작동적으로 연결된 프로모터를 추가로 포함한다. 일부 실시형태에서, 상기 프로모터는 구성적 프로모터이다. 일부 실시형태에서, 상기 프로모터는 유도성 프로모터이다. 효모 세포에서 사용하기에 적절한 프로모터의 예시적 예는, 이로써 한정되지 않지만, 케이. 락티스(*K. lactis*)의 TEF1 유전자의 프로모터, 사카로마이세스 세레비지에의 PGK1 유전자의 프로모터, 사카로마이세스 세레비지에의 TDH3 유전자의 프로모터, 억제성 프로모터, 예를 들면, 사카로마이세스 세레비지에의 CTR3 유전자의 프로모터, 및 유도성 프로모터, 예를 들면, 사카로마이세스 세레비지에의 갈락토스 유도성 프로모터(예를 들면, GAL1, GAL7, 및 GAL10 유전자의 프로모터)를 포함한다.

[0193] 일부 실시형태에서, 핵 국지화 서열(NLS)을 포함하는 추가의 뉴클레오티드 서열은 뉴클레아제를 코딩하는 뉴클레오티드 서열의 5'에 연결된다. NLS는 보다 큰 뉴클레아제(>25 kD)의 핵 국지화를 촉진할 수 있다. 일부 실

시형태에서, 핵 국지화 서열은 SV40 핵 국지화 서열이다. 일부 실시형태에서, 핵 국지화 서열은 효모 핵 국지화 서열이다.

[0194] 뉴클레아제 발현 벡터는 당해 기술분야의 숙련자에게 명백한 기술에 의해 제조할 수 있다. 특정 실시형태에서, 상기 벡터는 당해 기술분야에 공지된 폴리머라제 연쇄 반응(PCR) 및 분자 클로닝 기술을 사용하여 제조된다[참조: *PCR Technology: Principles and Applications for DNA Amplification*, ed. HA Erlich, Stockton Press, New York, N.Y. (1989); Sambrook *et al.*, 2001, *Molecular Cloning : A Laboratory Manual*, 3rd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY].

[0195] 5.3 숙주 세포

[0196] 다른 측면에서, 본원에는, 본원에 기재된 하나 이상의 외인성 핵산을 게놈에 삽입하는 임의의 방법에 의해 생성되는 변형된 숙주 세포가 제공된다. 적절한 숙주 세포는, 목적하는 핵산 또는 "공여체의 DNA"를 염색체 또는 에피솜 유전자좌로 삽입하는 것이 바람직한 임의의 세포를 포함한다. 일부 실시형태에서, 상기 세포는 상동성 재조합을 수행하는 능력을 가진 유기체의 세포이다. 예시적 실시형태의 몇몇이 효모(*예. 세레비지애*)에서 입증되어 있지만, 본원에서 제공된 게놈 변형 방법은, 재조합 시스템이 효모에서만 효율적이지 않은 경우라도, 기능성 재조합 시스템을 갖는 모든 생물학적 생물체에서 수행될 수 있다. 기능성 상동성 재조합 시스템을 갖는 다른 세포 또는 세포 유형은, 바실러스 서브틸리스(*Bacillus subtilis*) 및 이.콜라이(*E. coli*) 등의 박테리아 [참조: RecE RecT 재조합 효율성임; Muirers *et al.*, *EMBO rep.* 1: 239-243, 2000]; 원생동물[참조: 플라스모디움(*Plasmodium*), 톡소플라즈마(*Toxoplasma*)]; 기타 효모[참조: 스킴조사카로마이세스 폼베(*Schizosaccharomyces pombe*)]; 사상균[참조: 아시브야 고씨피이(*Ashbya gossypii*)]; 필로스코미트렐라 파텐스(*Physcomitrella patens*) 이끼 등의 식물[참조: Schaefer and Zryd, *Plant J.* 11: 1195-1206, 1997]; 및 포유류 세포 및 치킨 DT40 세포 등의 동물 세포[참조: Dieken *et al.*, *Nat. Genet.* 12:174-182, 1996]를 포함한다.

[0197] 일부 실시형태에서, 숙주 세포는 원핵생물 세포이다. 일부 실시형태에서, 숙주 세포는 진핵생물 세포이다. 일부 실시형태에서, 상기 세포는 진균 세포, 박테리아 세포, 식물 세포, 곤충 세포, 조류 세포, 어류 세포 및 포유동물 세포이다. 일부 실시형태에서, 포유동물 세포는 설치류 세포, 영장류 세포 및 인간 세포로 이루어진 그룹으로부터 선택된다. 일부 실시형태에서, 세포는 진균 세포(예: 효모 세포), 박테리아 세포, 식물 세포 또는 동물 세포(예: 치킨 세포)이다. 일부 실시형태에서, 숙주 세포는 차이니즈 햄스터 난소(CHO) 세포, COS-7 세포, 마우스 섬유아세포, 마우스 배아 암세포, 또는 마우스 배아 줄기세포이다. 일부 실시형태에서, 숙주 세포는 곤충 세포이다. 일부 실시형태에서, 숙주 세포는 S2 세포, Schneider 세포, S12 세포, 5B1-4 세포, Tn5 세포, 또는 Sf9 세포이다. 일부 실시형태에서, 숙주 세포는 단세포성 진핵 유기체 세포이다.

[0198] 특정 실시형태에서, 숙주 세포는 효모 세포이다. 유용한 효모 숙주 세포는 미생물 수탁기관(예를 들면, IFO, ATCC, *etc.*)에 기탁된 효모 세포를 포함하며, 특히 아키쿨로코니디움(*Aciculoconidium*), 암브로시오지마(*Ambrosiozyma*), 아르트로아스쿠스(*Arthroascus*), 아룩시오지마(*Arxiozyma*), 아시브야(*Ashbya*), 바베비아(*Babjevia*), 벤싱토니아(*Bensingtonia*), 보트리오아스쿠스(*Botryoascus*), 보트리오지마(*Botryozyma*), 브레타노마이세스(*Brettanomyces*), 불레라(*Bullera*), 불레라마이세스(*Bulleromyces*), 칸디다(*Candida*), 시테로마이세스(*Citeromyces*), 클라비스포라(*Clavispora*), 크립토코커스(*Cryptococcus*), 키스토필로바시디움(*Cystofilobasidium*), 데바리오마이세스(*Debaryomyces*), 데카라(*Dekkara*), 디포다스콕시스(*Dipodascopsis*), 디포다스쿠스(*Dipodascus*), 에에니엘라(*Eeniella*), 엔도마이콕셀라(*Endomycopsella*), 에레마스쿠스(*Eremascus*), 에레모테시움(*Eremothecium*), 에리트로바시디움(*Erythrobasidium*), 펠로마이세스(*Fellomyces*), 필로바시디움(*Filobasidium*), 갈락토마이세스(*Galactomyces*), 게오텐리쿰(*Geotrichum*), 구일리에르몽델라(*Guilliermondella*), 한세니아스포라(*Hanseniaspora*), 한세놀라(*Hansenula*), 하세가와에아(*Hasegawaea*), 홀테르만니아(*Holtermannia*), 호르모아스쿠스(*Hormoascus*), 하이포피키아(*Hyphopichia*), 이삐트첸키아(*Issatchenkia*), 클로에케라(*Kloeckera*), 클로에케라스포라(*Kloeckerasporea*), 클루이베로마이세스(*Kluyveromyces*), 콘도아(*Kondoa*), 쿠라이시아(*Kuraishia*), 쿠르즈마노마이세스(*Kurtzmanomyces*), 레우코스포르리디움(*Leucosporidium*), 리포마이세스(*Lipomyces*), 로데로마이세스(*Lodderomyces*), 말라세지아(*Malassezia*), 메트쉬니코비아(*Metschnikowia*), 프라키아(*Mrakia*), 마이속지아(*Myxozyma*), 나드소니아(*Nadsonia*), 나카자와에아(*Nakazawaea*), 네마토스포라(*Nematospora*), 오가타에아(*Ogataea*), 오오스포리디움(*Oosporidium*), 파키솔렌(*Pachysolen*), 파키티코스포라(*Phachytichospora*), 파피아(*Phaffia*), 피키아(*Pichia*), 로도스포리디움(*Rhodospiridium*), 로도토룰라(*Rhodotorula*), 사카로마이세스(*Saccharomyces*), 사카로마이코데스(*Saccharomycodes*), 사카로마이콕시스(*Saccharomycopsis*), 사이토엘라(*Saitoella*), 사카구키아(*Sakaguchia*),

사투르노스포라(*Saturnospora*), 스킴조블라스토스포르리온(*Schizoblastosporion*), 쉬조사카로마이세스(*Schizosaccharomyces*), 쉬반니오마이세스(*Schwanniomyces*), 스포리디오볼루스(*Sporidiobolus*), 스포로볼로마이세스(*Sporobolomyces*), 스포로파키테르미아(*Sporopachydermia*), 스테파노아스쿠스(*Stephanosascus*), 스테리그마토마이세스(*Sterigmatomyces*), 스테리그마토스포리디움(*Sterigmatosporidium*), 심보비오타프리카(*Symbiotaphrina*), 심포디오마이세스(*Sympodiomyces*), 심포디오마이콥시스(*Sympodiomycopsis*), 토룰라스포라(*Torulaspora*), 트리코스포리엘라(*Trichosporiella*), 트리코스포론(*Trichosporon*), 트리코넵시스(*Trigonopsis*), 츠키야에아(*Tsuchiyaea*), 우데니오마이세스(*Udeniomyces*), 왈토마이세스(*Waltomyces*), 위케르하미아(*Wickerhamia*), 위케르하미엘라(*Wickerhamiella*), 윌리옵시스(*Williopsis*), 야마다지마(*Yamadazyma*), 야로위아(*Yarrowia*), 자이고아스쿠스(*Zygoascus*), 자이고사카로마이세스(*Zygosaccharomyces*), 자이고윌리옵시스(*Zygowilliopsis*) 및 자이고자이마(*Zygozima*) 속에 속한다.

[0199] 일부 실시형태에서, 효모 숙주 세포는 사카로마이세스 세레비지에 세포, 피키아 파스토리스(*Pichia pastoris*) 세포, 스킴조사카로마이세스 폼베(*Schizosaccharomyces pombe*) 세포, 데케라 브룩셀렌시스(*Dekkera bruxellensis*) 세포, 클루이베로마이세스 락티스(*Kluyveromyces lactis*) 세포, 아룩술라 아테니노란스(*Arxula adeninivorans*) 세포, 또는 한세놀라 폴리모र्फ(*Hansenula polymorpha*) (현재 피키아 안구스타(*Pichia angusta*))로 알려져 있음) 세포이다. 특정 실시형태에서, 효모 숙주 세포는 사카로마이세스 세레비지에 세포이다. 일부 실시형태에서, 효모 숙주 세포는 사카로마이세스 프라길리스(*Saccharomyces fragilis*) 세포 또는 클루이베로마이세스 락티스(과거 사카로마이세스 락티스(*Saccharomyces lactis*))로 불렸음) 세포이다. 일부 실시형태에서, 효모 숙주 세포는 칸디다 속에 속하는 세포, 예를 들면, 칸디다 리폴리티카(*Candida lipolytica*), 칸디다 구일리에르몽디이(*Candida guilliermondii*), 칸디다 크루세이(*Candida krusei*), 칸디다 슈도트로피칼리스(*Candida pseudotropicalis*), 또는 칸디다 유틸리스(*Candida utilis*)이다. 다른 특정 실시형태에서, 효모 숙주 세포는 클루이베로마이세스 마룩시아누스(*Kluyveromyces marxianus*) 세포이다.

[0200] 특정 실시형태에서, 효모 숙주 세포는, 빵 효모 세포, CBS 7959 세포, CBS 7960 세포, CBS 7961 세포, CBS 7962 세포, CBS 7963 세포, CBS 7964 세포, IZ-1904 세포, TA 세포, BG-1 세포, CR-1 세포, SA-1 세포, M-26 세포, Y-904 세포, PE-2 세포, PE-5 세포, VR-1 세포, BR-1 세포, BR-2 세포, ME-2 세포, VR-2 세포, MA-3 세포, MA-4 세포, CAT-1 세포, CB-1 세포, NR-1 세포, BT-1 세포, 및 AL-1 세포로 이루어진 그룹으로부터 선택되는 사카로마이세스 세레비지에 세포이다. 일부 실시형태에서, 숙주 세포는 PE-2 세포, CAT-1 세포, VR-1 세포, BG-1 세포, CR-1 세포, 및 SA-1로 이루어진 그룹으로부터 선택되는 사카로마이세스 세레비지에 세포이다. 특정 실시형태에서, 상기 사카로마이세스 세레비지에 숙주 세포는 PE-2 세포이다. 다른 특정 실시형태에서, 상기 사카로마이세스 세레비지에 숙주 세포는 CAT-1 세포이다. 다른 특정 실시형태에서, 상기 사카로마이세스 세레비지에 숙주 세포는 BG-1 세포이다.

[0201] 일부 실시형태에서, 효모 숙주 세포는 바이오에탄올 발효 등의 산업적 발효에 적절한 세포이다. 특정 실시형태에서, 상기 세포는 높은 용매 농도, 고온, 확장된 기질 용도, 영양 제한, 삼투성 스트레스 요인, 산성, 설파이트 및 박테리아 오염, 또는 이들의 조합(이들은 산업적 발효 환경의 스트레스 조건인 것으로 인지됨)하에 생존하도록 조건화된다.

[0202] 5.4 적용

[0203] 5.4.1 유전자 및 유전자 치료

[0204] 본원에 기재된 방법 및 조성물은, 예를 들면, 대상체로부터 유래된 세포 집단에서 생체외에서 유전자 결함을 수정하고자 하는 치료학적 적용에서 잇점을 제공한다. 예를 들면, 슈반크 등[참조: Schwank *et al.*, Cell Stem Cell 13:653-658 (2013)]은 최근들어 인간 CF 환자의 배양된 장 줄기 세포에서 상동성 재조합에 의해 CFTR 유전자좌를 수정하기 위해 CRISPR/Cas9 게놈 편집 시스템의 이용을 보고했다. 수정된 대립인자는, 클론 확장된 오르가노이드에서 측정된 바와 같이 발현되고 충분히 기능적이며, 따라서 이러한 보고는 단일-유전자 유전적 결함을 갖는 환자로부터 유래된 초대 성인 줄기 세포에서 상동성 재조합에 의한 유전자 수정을 위한 개념 입증을 제공한다. 그러나, 배양된 줄기 세포에서 CFTR 유전자좌의 수정은 공여체 DNA와 함께 푸로마이신 내성 카세트의 게놈 삽입, 이어서 푸로마이신에서의 선별을 필요로 했다. 이는 인간 세포 게놈 내로 네오마이신 내성 유전자의 도입, 이어서 G418에서 연장된 배양 시간이 세포의 특성에 대한 변화를 유발하는 것으로 보고했고, 증강된 녹색 형광 단백질(EGFP) 및 기타 형광 단백질의 발현이 면역원성 및 독성을 유발하는 것으로 보고되었다[참조: Barese *et al.*, Human Gene Therapy 22:659-668 (2011); Morris *et al.*, Blood 103:492-499 (2004); and Hanazono *et al.*, Human Gene Therapy 8:1313-1319 (1997)]. 따라서, 본원에 제공된 방법 및 조성물은 선별가

능한 마커의 삽입 필요성 없이 초대 성인 줄기 세포에서 상동성 재조합에 의한 유전자 수정을 수행하는데 사용될 수 있다.

[0205] 유전자 질환의 HR-매개된 수정의 또 다른 보고에서, 우 등[참조: Wu *et al.*, *Cell Stem Cell* 13:659-662 (2013)]은 백내장을 유발하는 *Crygc* 유전자에서 우성 돌연변이를 갖는 마우스가 돌연변이 대립인자를 표적화하는 Cas9 mRNA 및 단일-유도 RNA(sgRNA)의 접합자 내로 동시주입에 의해 구제될 수 있음을 입증했다. 수정은 외인성 공급된 올리고뉴클레오티드 또는 내인성 ST 대립인자에 기반하여 상동성-지향된 수복(HDR)에 의해 발생했고, 수득되는 마우스는 가임성이고 이들의 자손에 수정된 대립인자를 전달할 수 있다. 그러나, HDR-매개된 수복의 비율은 NHEJ에 의한 수복 또는 비-수복의 발생율보다 훨씬 낮았다[참조: Wu *et al.*, 표 1]. 따라서, 본원에 제공된 방법 및 조성물은 HR-매개된 유전자 변형을 선별하는 유용한 선별 메커니즘을 제공함으로써 유전자 수정의 효율을 개선시키기 위해 사용될 수 있다.

[0206] 5.4.2 대사 경로 조작 방법

[0207] 본원에서 기재하는 방법 및 조성물은 최적화된 생합성 경로, 예를 들면, 바이오매스를 바이오연료, 약제 또는 생체적합물질로 전환하는 경로를 포함하는 재조합 생물체를 작제하는 특정 이점들을 제공한다. 기능성 비-천연 생물학적 경로는 항말라리아 약물 아르테미시닌(artemisinin)으로의 전구체[참조: Martin *et al.*, *Nat Biotechnol* 21:796-802 (2003)]; 지방산 유래의 연료 및 화합물질[예를 들면, 지방산 에스테르, 지방 알코올 및 왁스; 예를 들면, Steen *et al.*, *Nature* 463:559-562 (2010)]; 메틸 할라이드-유래의 연료 및 화합물질[참조: Bayer *et al.*, *J Am Chem Soc* 131:6508-6515 (2009)]; 콜레스테롤 저하 약물을 제조하는 폴리케타이드 신타제[참조: Ma *et al.*, *Science* 326:589-592 (2009)]; 및 폴리케타이드[참조: Kodumal, *Proc Natl Acad Sci USA* 101:15573-15578 (2004)]를 제조하기 위한 미생물 숙주에서 성공적으로 작제되었다.

[0208] 전형적으로, 대사 조작, 및 특히 생합성 경로의 작제는, 경로 구성분이 한번에 단일 유전자좌로 숙주 세포 계층에 삽입되는 한번에 한 단계식의 방법으로 진행된다. 본원에서 제공하는 삽입 방법은, 신규 대사 경로, 즉, 숙주 세포에 의해 내인성으로 생성되지 않는 대사물질을 생성하는 대사 경로의 효소를 코딩하는 하나 이상의 이중성 뉴클레오티드 서열을 포함하도록 숙주 세포, 예를 들면, 미생물 세포를 조작하는데 전형적으로 필요한 시간을 감소시키는데 사용할 수 있다. 다른 특정 실시형태에서, 본원에서 제공된 삽입 방법은 숙주 세포에 내인성인 대사 경로, 즉, 숙주 세포에 의해 내인성으로 생성되는 대사물질을 생성하는 대사 경로의 효소를 코딩하는 하나 이상의 이중성 뉴클레오티드 서열을 포함하도록 숙주 세포를 효율적으로 조작하는데 사용할 수 있다. 한 가지 예에서, 디자인 전략은 숙주 세포의 3가지 고유 유전자를 상보적 외인성 경로로 대체하고자 할 수 있다. 현재의 당해 기술분야를 사용하는 이들 3가지 내인성 유전자좌를 변형시키는 것은 3가지 분리된 형질전환을 필요로 한다. 이와는 달리, 본원에서 제공된 동시 다중 삽입 방법은 3가지 삽입 모두가 단일 형질전환으로 수행될 수 있도록 하여, 필요한 조작 주기를 1/3로 줄일 수 있다. 더욱이, 상기 방법은 숙주 세포 캐시의 다중 부위에 삽입된 최적화된 경로 성분을 포함하는 DNA 어셈블리를 제2 숙주 세포 캐시의 유사한 부위로 포팅(porting)될 수 있게 한다. 목적하는 유전자형을 조작하는데 필요한 주기의 수를 감소시킴으로써, 대사 경로의 작제의 속도를 상당히 증가시킨다.

[0209] 5.4.2.1 아이소프레노이드 경로 조작

[0210] 일부 실시형태에서, 본원에서 제공하는 방법은, 조작된 숙주 세포의 생성물 프로파일을 변형시키기 위해 생합성 경로 중 하나 이상의 성분을 동시에 도입 또는 대체하는데 사용할 수 있다. 일부 실시형태에서, 상기 생합성 경로는 아이소프레노이드 경로이다.

[0211] 테르펜은 많은 생물체에서 생성되는 탄화수소의 큰 분류이다. 테르펜이 (예를 들면, 산화 또는 탄소 골격의 재배열을 통해) 화학적으로 개질되는 경우, 생성되는 화합물은 일반적으로 테르페노이드로 지칭되며, 이는 아이소프레노이드로도 공지되어 있다. 아이소프레노이드는 많은 중요한 생물학적 역할을 하는데, 예를 들면, 전자 수송 연쇄반응에서 퀸온으로서, 막 성분으로서, 단백질 프레닐화를 통한 조절 및 아세포성 표적화에서, 카로테노이드, 클로로필을 포함하는 광합성 색소로서, 호르몬 및 보조인자로서, 및 모노테르펜, 세스퀴테르펜, 및 다이테르펜과 함께 식물 방어 화합물로서 작용한다. 이들은 항생제, 호르몬, 향암 약물, 살충제, 및 화합물질로서 산업적으로 유용하다.

[0212] 테르펜은 아이소프렌(C_5H_8)의 연결 단위에 의해 유래되며, 아이소프렌 단위 존재수에 의해 분류된다. 헤미테르펜은 단일 아이소프렌 단위로 이루어진다. 아이소프렌 자체는 유일한 헤미테르펜으로 간주된다. 모노테르펜은 2개의 아이소프렌 단위로 이루어지며, 분자식 $C_{10}H_{16}$ 을 가진다. 모노테르펜의 예는 게라니올(geraniol), 리모넨

(limonene), 및 테르피네올(terpineol)이다. 세스퀴테르펜은 3개의 아이소프렌 단위로 이루어지며, 분자식 $C_{15}H_{24}$ 을 가진다. 세스퀴테르펜의 예는 파르네센 및 파르네솔(farnesol)이다. 다이테르펜은 4개의 아이소프렌 단위로 이루어지며, 분자식 $C_{20}H_{32}$ 을 가진다. 다이테르펜의 예는 카페스톨(cafestol), 카웨올(kahweol), 켈브렌(cembrene) 및 탁사다이엔(taxadiene)이다. 세스터테르펜은 5개의 아이소프렌 단위로 이루어지며, 분자식 $C_{25}H_{40}$ 을 가진다. 세스터테르펜의 예는 게라닐파르네솔(geranyl farnesol)이다. 트리테르펜은 6개의 아이소프렌 단위로 이루어지며, 분자식 $C_{30}H_{48}$ 을 가진다. 테트라테르펜은 8개의 아이소프렌 단위로 이루어지며, 분자식 $C_{40}H_{64}$ 을 가진다. 생물학적으로 중요한 테트라테르펜은, 아크릴산 라이펜(acyclic lycopene), 모노사이클릭 감마-카로텐, 및 바이사이클릭 알파-카로텐 및 베타-카로텐을 포함한다. 폴리테르펜은 다수의 아이소프렌 단위의 장쇄로 이루어진다. 천연 고무는, 이중 결합이 cis형인 폴리아이소프렌으로 이루어진다.

[0213] 테르펜은 아이소펜테닐 파이로포스페이트(아이소펜테닐 다이포스페이트 또는 IPP) 및 이의 이성질체 다이메틸알릴 파이로포스페이트(다이메틸알릴 다이포스페이트 또는 DMAPP)의 축합을 통해 생합성된다. IPP 및 DMAPP를 생성하는 2가지 경로, 즉 진핵생물의 메발로네이트-의존성(MEV) 경로(도 3), 및 원핵생물의 메발로네이트-비의존성 또는 데옥시리보실물로스-5-포스페이트(DXP) 경로가 공지되어 있다. 식물은 MEV 경로와 DXP 경로 둘 다를 사용한다. 즉, IPP 및 DMAPP는 프레닐 디스포스페이트 신타제(예를 들면, 각각 GPP 신타제, FPP 신타제, 및 GGPP 신타제)의 작용을 통해 폴리프레닐 다이포스페이트(예를 들면, 게라닐 디스포스페이트 또는 GPP, 파네실 다이포스페이트 또는 FPP, 및 게라닐게라닐 다이포스페이트 또는 GGPP)로 축합된다. 폴리프레닐 다이포스페이트 중간체는 테르펜 신타제에 의해 보다 복잡한 아이소프레노이드 구조로 전환된다.

[0214] 테르펜 신타제는 다중 생성물을 형성하는 큰 유전자 계열로 구성된다. 테르펜 신타제의 예는, GPP를 모노테르펜으로 전환하는 모노테르펜 신타제; GGPP를 다이테르펜으로 전환하는 다이테르펜 신타제; 및 FPP를 세스퀴테르펜으로 전환하는 세스퀴테르펜 신타제를 포함한다. 세스퀴테르펜 신타제의 예는 파르네센 신타제로, FPP를 파르네센으로 전환한다. 테르펜 신타제는, 이들이 대사 분지점에서 작동하여 세포에 의해 생성되는 아이소프레노이드의 유형을 지시하기 때문에, 아이소프레노이드로의 경로 유동의 조절에 중요하다. 더욱이, 테르펜 신타제는 이러한 테르펜의 고수율 생성에 중요하다. 이와 같이, 이중성 아이소프레노이드 생성을 위해 조작된 숙주에서 경로 유동을 개선하기 위한 하나의 전략은 테르펜 신타제를 코딩하는 핵산의 다중 카피를 도입하는 것이다. 예를 들면, 파르네센 등의 세스퀴테르펜의 생성이 요구되는 MEV 경로를 포함하는 조작된 미생물에서, 세스퀴테르펜 신타제, 예를 들면, 파르네센 신타제를 상기 경로의 말단 효소로서 사용하고, 파르네센 신타제 유전자의 다중 카피를 파르네센 생성에 최적화된 균주의 생성을 위해 숙주 세포로 도입할 수 있다.

[0215] 아이소프레노이드의 생합성은 프레닐 디스포스페이트 신타제 및 테르펜 신타제의 상류 성분과 동일한 경로에 의존하기 때문에, 일단 숙주 "플랫폼(platform)" 균주로 조작된 이들 경로 성분을 세스퀴테르펜의 생성에 사용할 수 있으며, 세스퀴테르펜의 동일성은 숙주 세포로 도입되는 특정 세스퀴테르펜 신타제에 의해 지시될 수 있다. 더욱이, 상이한 아이소프렌 단위를 갖는 테르펜, 예를 들면, 세스퀴테르펜 대신에 모노테르펜의 생성이 바람직한 경우, 프레닐 다이포스페이트 신타제 및 테르펜 신타제 둘 다는 상기 경로의 상류 성분을 여전히 사용하면서 상이한 테르펜을 생성하도록 대체될 수 있다.

[0216] 따라서, 본원에서 제공되는 방법 및 조성물을 사용하여, 아이소프레노이드 생성 경로, 예를 들면, 목적하는 아이소프레노이드를 생성하는 MEV 경로를 포함하는 숙주 세포를 효율적으로 변형할 수 있다. 일부 실시형태에서, 숙주 세포는 MEV 경로를 포함하며, 본원에서 제공되는 동시 다중 삽입 방법을 사용하여, 프레닐 다이포스페이트 신타제 및/또는 테르펜 신타제의 다중 카피를 동시에 도입하여, 숙주 세포의 테르펜 생성물 프로파일을 규명할 수 있다. 일부 실시형태에서, 상기 프레닐 다이포스페이트 신타제는 GPP 신타제이며, 상기 테르펜 신타제는 모노테르펜 신타제이다. 일부 실시형태에서, 상기 프레닐 다이포스페이트 신타제는 FPP 신타제이며, 상기 테르펜 신타제는 세스퀴테르펜 신타제이다. 일부 실시형태에서, 상기 프레닐 다이포스페이트 신타제는 GGPP 신타제이며, 상기 테르펜 신타제는 다이테르펜 신타제이다. 다른 실시형태에서, 숙주 세포는 MEV 경로 및 프레닐 다이포스페이트 신타제 및/또는 제1 유형 테르펜, 예를 들면, 파르네센의 생성을 위한 테르펜 신타제를 포함하며, 본원에서 제공되는 동시 다중 삽입 방법은 프레닐 다이포스페이트 신타제 및/또는 제2 유형 테르펜, 예를 들면, 아모파다이엔을 생성하기 위한 테르펜 신타제의 하나 이상의 카피를 동시에 대체할 수 있다. 이들 실시형태는 하기 실시예 3 및 4에 예시되어 있다. 본원에서 제공하는 방법은 경로 성분의 다중 카피를 사용하는 생합성 경로의 작제 및/또는 변형을 위해 유사하게 사용할 수 있으며, 생성물 프로파일이 단일 경로 성분의 다중 카피의 부가 또는 교환으로 용이하게 변형될 수 있는 숙주 세포를 조작하는 데 특히 유용하다.

[0217]

6. 실시예

[0218]

6.1 실시예 1: 결실 작제물의 삽입에 의해 CRISPR를 사용하여 3개 유전자의 동시 결실에 대한 gRNA 전달의 복수의 모드의 비교

[0219]

본 실시예는 에스. 세레비지애(*S. cerevisiae*)에서 RHR2, HO 및 ADH5 ORF(짧은 링커 서열의 삽입 포함)의 동시 결실을 위해 CRISPR의 사용을 입증하는 결과를 제공한다. 요약하면, 키메라 gRNA는 RHR2, HO 및 ADH5의 ORF에 함유된 표적화 고유 서열을 생성하고, 스트렙토코쿠스 피오게네스의 II형 세균 CRISPR 시스템으로부터 Cas9 단백질을 발현시키는 숙주 세포에서 다양한 구성으로 형질전환시켰다. 형질전환된 콜로니는 짧은 링커 서열을 갖는 1, 2 또는 3개의 ORF의 교환을 위해 콜로니 PCR(cPCR)에 의해 스크리닝했다.

[0220]

6.1.1. Cas9p의 구성적 발현

[0221]

야생형 사카로마이세스 세레비지애 균주(CEN.PK2, Mat a, *ura3*-, TRP1+, *leu2*-, MAL2-8C, SUC2)를 스트렙토코쿠스 피오게네스로부터의 Cas9p의 구성적 발현을 위해 숙주로서 사용했다. 중간 강도 FBA1 프로모터(서열번호 3)의 조절하에 Cas9에 대한 코딩 서열을 최적화한 효모 코돈을 함유하는 작제물의 게놈 삽입은 GRE3 유전자좌로 표적화했다.

[0222]

6.1.2. RHR2, HO 및 ADH5 ORF에서 CRISPR 표적 부위의 선별

[0223]

표적화 ORF 내부의 후보 CRISPR 표적은 PAM 서열 N₍₁₉₎NGG의 존재에 기반하여 동정되었다. NGG 서열은 PAM 서열로서 명명되고, PAM 서열에 선행하는 DNA의 8개 염기쌍은 특히 특이성을 실시하는데 중요하다[참조: Fu et al., *Nat Biotechnol* 31(9):822-826 (2013); Ran et al., *Nat Protoc* 8(11):2281-2308 (2013)]. 이어서, 후보 부위는 게놈 중의 표적 서열의 독자성에 기초하여 랭크되고, 게놈 중의 다른 부위에 대해 최저 유사성을 갖는 부위는 표적-외 절단의 위험을 최소화하기 위해 선택되었다. 표적 부위는 표 2에 나타낸다.

표 2

CRISPR 표적 부위

ORF	표적 서열 (NGG 생략)
RHR2	ACCTCTGGTACCCGTGACA (서열번호 4)
HO	CCGGCTTGATCGACTCAGA (서열번호 5)
ADH5	GGGTCATTGGTATCGATGG (서열번호 6)

[0225]

6.1.3. gRNA 전달 모드

[0226]

Cas9p는 일반적 구조 RNA 및 특이적 표적화 RNA와 관련하여 부위를 절단하기 위해 표적화한다. 현재 표준 "키메라" 구성은 이 작업에 채택되었고, 여기서, 표적화 및 구조적 RNA는 단일 가이드 RNA, 또는 gRNA를 생성하기 위해 융합한다[참조: Ran, Hsu et al. 2013]. gRNA 작제물(들)의 발현은 *SUP4* 터미네이터와 함께 *SNR52* 폴리머라제 III 프로모터에 의해 구동시켰다[참조: DiCarlo, Norville et al. 2013]. 각각 RHR2, HO 및 ADH5 유전자좌를 표적화하는 gRNA 작제물의 서열은 본원에서 서열번호 7, 8 및 9로 제공된다. gRNA 전달의 몇몇 모드가 상기 섹션 5.2.1에 기재된 바와 같이 사용되었고, 도 7에 묘사되었다. pRS4XX-시리즈 2μ 벡터[참조: Sikorski and Hieter 1989]로부터 gRNA 카세트의 발현은 1) Cas9-발현 효모 균주로 형질전환 이전에 완성된 환상 플라스미드(도 7.1 및 7.2)를 생성하기 위한 표준 클로닝 방법에 의해 또는 2) gRNA 카세트를 생체내 캡 수복에 의해 환상 플라스미드로 조립하고, 카세트를 선형화 벡터 골격과 함께 Cas9-발현 효모 균주로 직접 형질전환시킴으로써[참조: Orr-Weaver, Szostak et al. 1983](도 7.4) 달성될 수 있다. gRNA 카세트와 선형 벡터 골격(서열번호 10) 사이의 상동성(약 500bp)의 영역은 생체내 환상 gRNA 플라스미드의 조립을 촉진시킨다. 또는, 3) gRNA는 형질전환된 세포를 선택하기 위해 NatA(스트렙토마이세스 노우르세이로로부터 노우르세오트리신 아세틸트랜스퍼라제) 선별가능한 마커(서열번호 11)를 포함하는 밀폐된 플라스미드로 동시-형질전환된 선형 카세트로부터 직접 발현시켰다(도 7.3). 최종적으로, 제3 방법의 변형으로, 4) 선형 카세트는 NatA 마커의 중간 코딩 서열이 분실되도록 NatA 마커(서열번호 12) 내부의 PCR에 의해 선형화된 벡터로 동시-형질전환시키고; 완전 NatA 발현

카세트를 포함하는 밀폐된 플라스미드를 재생성하기 위해 갭 수복을 통해 재결합할 수 있는 보완적 중복 NatA ORF 단편(서열번호 13)도 또한 동시-형질전환시켰다(도 9b; 이하 실시예 2에서 논의됨).

[0227] 환상 gRNA 플라스미드(전달 모드 1)를 생성하기 위해, CRISPR 씨드 서열 및 카세트에 대한 20bp의 상류/하류 상동성을 함유하는 어닐링된 올리고뉴클레오타이드는 이. 콜라이에서 선형화 골격으로 갭 수복시켰고[참조: Mandecki 1986], 정확한 클론을 동정하였고, 완성된 플라스미드를 숙주 균주로 형질전환시켰다. 선형화 벡터(전달 모드 2, 3 및 4)에 대해 약 500bp 인접 상동성을 갖는 전장 선형 gRNA 카세트를 제조하기 위해, PCR 조립 방법이 사용되었다. 일반적 gRNA 카세트를 주형으로 사용하여, 1/2 카세트는 고유한 CRISPR 씨드 서열을 포함하는 22개 염기쌍의 중앙 중복을 생성하기 위한 프라이머를 사용하여 증폭시켰다. 이어서, 2개의 1/2 카세트를 전장 gRNA 발현 카세트를 생성하기 위해 제2 PCR 반응으로 조립했다. 형질전환당 10 μ l의 미정제된 PCR 조립체(전형적으로, DNA 마커 래더와 비교로 측정된 20-60ng/ μ l의 농도) 및 150ng의 선형화 2 μ 벡터를 사용했다.

[0228] 6.1.4. 선형 공여체 DNA의 생성

[0229] 선형 공여체 DNA는 중앙 링커(CGCTCGTCCAACGCCGGCGGACCT)에 인접하는 각 ORF를 표적화하는 500bp 상류 및 하류 상동성 영역을 포함하고, 미국 특허 제8,221,982호에 기재된 폴리뉴클레오타이드 조립 방법에 의해 생성했다. 각각 RHR2, HO 및 ADH5 유전자좌에서의 삽입을 위한 공여체 DNA 서열은 본원에서 서열번호 14, 15 및 16으로 제공된다.

[0230] 6.1.5. CRISPR를 사용하여 ORF의 동시 결실 및 짧은 링커 서열의 삽입

[0231] 공여체 DNA(약 1mg) 및 적합한 gRNA 시약은 42°C에서 열 충격 이전에 30°C에서 세포의 30분 배양을 첨가하면서 최적화된 LiAc 방법[참조: Gietz and Woods 2002]을 사용하여 각각의 Cas9 발현 균주로 동시-형질전환시켰다. 세포는 gRNA 또는 마커 플라스미드를 유지시키기 위해 선택적 항생제 함유 배지(노우르세오트리신, 100 μ g/ml)에 플레이팅하기 전에 비선택적 YPD(효모 추출 펩톤 텍스트로즈) 배지에서 밤새 회수했다. 마커-레스 삽입은, 삽입 링커 서열(표 3)에 5' 삽입 측면의 상류를 결합하는 프라이머를 사용하는 콜로니 PCR(cPCR)이 표적 삽입 사실을 나타내는 결과인 정확한 앰플리콘을 생성하는 경우에 포지티브로서 스코어링했다.

표 3

RHR2, HO 및 ADH5 유전자좌에서 링커 삽입의 cPCR 변동을 위한 프라이머 서열

프라이머 유형	설명	서열	서열번호
RHR2-US-F	RHR2 유전자 좌 US FOR	gggtgcgaagtaccaccacgtttcttttcatctct	서열번호 17
HO-US-F	HO 유전자 좌 US FOR	acgtgtgtgtctcatggaaattgatgcagttgaagaca	서열번호 18
ADH5-US-F	ADH5 유전자 좌 US FOR	ggcgttatatccaaacatttcagacagaagatt	서열번호 19
R5	링커 REV	AGGTCCGCGCGGCGTTGGACGAGC G	서열번호 20

[0232]

[0233] 6.1.6. 결과 및 논의

[0234] 갭 수복 전달 모드: gRNA 전달의 4개 모드는 짧은 링커 서열의 삽입과 함께 RHR2, HO 및 ADH5 개방 판독 프레임의 동시 결실의 효율에 대해 평가했다. 삼중 삽입을 평가하기 위한 콜로니 PCR 결과는 도 8에 나타내고, 비율은 표 4에 요약한다.

표 4

gRNA 작제물의 가변적 전달에 의한 삼중 삽입 비율

설명	삼중 비율
삼중 선택, 플라스미드 gRNAs	0.91
단일 선택, 플라스미드 gRNAs	0
마커 플라스미드, 선형 gRNAs	0
갭 회복 선형 gRNAs	0.64

[0235]

[0236] gRNA 전달의 제1 모드에서, 3개의 gRNA(각각 RHR2, HO 및 ADH5를 표적화하는) 각각을 플라스미드 포함 고유 마커(NatA, URA3 카세트 및 HIS3 카세트; 참조: 도 7.1)에 공급하고, 세포를 모두 3개의 플라스미드로 형질전환시키고, 각 마커의 발현을 위해 삼중으로 선별했다. (삽입을 통해) 삼중 결실의 매우 높은 효율(91%)이 관찰되었다(도 8, 패널 1). 이러한 높은 빈도는 아마도 이들의 카세트를 포함하는 플라스미드에 대해 삼중 선별로 모두 3개의 gRNA의 지속적 발현으로부터 발생할 것이다. 대조적으로, gRNA가 동일한 마커(NatA; 도 7.2 참조)를 포함하는 3개의 플라스미드 상에 공급된 제2 모드는 어떤 삼중 결실도 생성하지 못했다(도 8, 패널 2). 대신에, 단일 결실이 지배적이었고, 이는 3개의 플라스미드 중 단지 1개만을 유지시키기 위한 선별 조건과 일치한다. 제3 모드에서, gRNA는 선형 카세트로서 공급되었고, NatA 마킹 플라스미드는 형질전환된 세포를 선별하기 위해 포함된다(참조: 도 7.3). 어떤 삼중 결실도 관찰되지 않았고, 임의의 결실 사상의 매우 낮은 비율이 관찰되었다(도 8, 패널 3). 이러한 전달 모드는 gRNA 작제물의 일시적 발현을 유도할 것으로 예상되고, 이는 지속 발현에 비해 열등한 것으로 보인다. 조사된 제4 전달 모드는 NatA 마커를 포함하는 선형화 벡터에서 3개의 gRNA 카세트의 겹 수복을 필요로 한다(참조: 도 7.4). 본 발명자들은 64%의 콜로니가 삼중으로 결실되었다는 것을 관찰했다(도 8, 패널 4). 이는 이러한 전달 모드가 제1 전달 모드의 경우와 같이 모두 3개의 gRNA의 지속적 발현을 실시하지 않았기 때문에 놀라운 결과이다. 사실, 제2 모드로부터의 결과는 명백하게 3개의 유사 표시된 gRNA 플라스미드의 동시-형질전환이 유효하지 않다는 것을 나타내고, 제3 모드로부터의 결과는 선형 카세트로부터의 gRNA의 일시적 발현이 또한 비기능적임을 나타낸다. 따라서, gRNA 카세트의 전달 모드로서 겹 수복과 관련된 CRISPR/Cas-9 매개된 게놈 삽입 사상에 대한 예상치 않은 유리한 이점이 존재한다.

[0237] 6.2 실시예 2: 겹 수복을 통한 HR 적격 세포의 선택

[0238] 이 실시예는 gRNA 발현을 위한 선택 이점과 독립적으로, 뉴클레아제 매개된 삽입 사상의 효율을 향상시키기 위한 겹 수복의 이점을 입증한다.

[0239] 겹 수복이 CRISPR(또는 임의의 부위 특이적 뉴클레아제)에 의해 유전자 조작된 클론의 회복을 향상시킬 가능성이 있는 하나의 메커니즘은 상동성 재조합 (HR)에 대해 능숙한 세포를 위한 추가 선별을 실시하는 것이다. HR 능력은 비동기 세포 집단에서 광범위하게 다를 수 있고[참조: Brnzei and Foiani, *Nat Rev Mol Cell Bio* 9(4):297-308 (2008)], 선별가능한 마커를 포함하는 플라스미드의 겹 수복을 달성할 수 있는 세포의 선별은 특별히 HR 능숙한 집단을 선택할 수 있다. 이는 실시예 1에서 논의된, gRNA 전달의 제4 모드(도 7.4 및 8.4)의 놀라운 성공을 설명할 수 있었다. 선별 메커니즘으로서 겹 수복의 효과를 gRNA 중의 적어도 하나의 지속 발현의 것로부터 분리하기 위해, 본 발명자들은 적합한 공여체 DNA 및 선형 gRNA 카세트(상기 실시예 1에서 기재됨) 및 두 마커 벡터 중의 하나의 동시-형질전환에 의해 RHR2, HO 및 ADH5 유전자좌의 단일 결실의 비율을 평가했다. 제1 마커 벡터는 밀폐되었고, 즉 겹 수복은 형질전환된 세포에서 NatA 마커의 발현에 필요하지 않다(도 9a). NatA 마커의 중앙 부분이 소실되도록 제2 마커가 선형되었지만, 완전 NatA 발현 카세트를 포함하는 밀폐된 기능적 마커 벡터를 생성하기 위해 중복 단편의 겹 수복에 의해 보완될 수 있었다(도 9b). 두 경우에, gRNA의 발현은 단지 일시적이다. 3개의 독립적 실험에 대해, 본 발명자들은 마커 플라스미드가 겹 수복을 필요로 할 때(도 9, 비율은 도 10 내지 12에 요약됨) 단일 유전자좌 결실(삽입) 비율의 8배 이하(평균 약 3 내지 5배 범위) 개선을 주시했다.

[0240] 이 결과는 겹 수복이 HR에 능숙한 세포를 위한 추가의 선별로서 기능할 수 있고, 따라서 대부분 성공적인 뉴클레아제 원래 유전자 조작 및 특히, 공여체 DNA의 표적화 삽입을 가능하게 할 수 있다는 가설을 지지한다. 본 발명자들은 에스. 세레비지에에는 특히 HR에, 그리고 NHEJ 및 다른 불완전한 수복 방법을 선호하는 세포(예: 포유동물 세포)에서 능숙하다는 것을 주목하고, 본 발명자들은 겹 수복이 하나 이상의 표적화 삽입을 포함하는 세포의 회복을 증가시키는데 특히 효과적일 수 있음을 제안한다.

[0241] 6.3 실시예 3: 복수 조각 겹 수복을 통해 HR-적격 세포의 증강된 선별

[0242] 본 실시예는 벡터를 추가로 단편화하여 마커 벡터의 겹 수복의 복잡성을 증가시킴으로써, 뉴클레아제 원래 유전자 조작의 효율성을 추가로 증가시킬 수 있다는 것을 입증한다.

[0243] Cas9-발현 반수체 효모 세포(에스. 세레비지에)는 Gal80, HO 및 ADH5 개방 판독 프레임, 각 유전자좌를 표적화하는 gRNA 작제물의 동시 마커 부재 결실을 위해 공여체 DNA로 형질전환시키고, 실시예 1에 기재된 바와 같이, 형질전환의 추가로 벡터 골격을 선형화하고, 이에 의해 벡터 골격은 각 조각이 서로(47 bp)에 대해 뿐만 아니라 gRNA 카세트(약 500bp)에 대해 중복 상동성 영역을 포함하는 2개의 조각으로 단편화했다. 이는 gRNA 카세트를

도입하는 환상 NatA 마커의 겹 수복을 통해 3-피스 생체내 조립을 가능하게 한다. NatA 마커 ORF는 두 골격 단편 중의 하나 위에 구성된 반면, NatA 발현을 구동하는 프로모터(*K. lactis* Tef1 프로모터)는 나머지 위에 구성되었고, 따라서 NatA 발현은 단편의 HR-매개된 조립체에서만 가능하다. 각각 Gal80, HO 및 ADH5 유전자좌의 각각을 표적화하는 공여체 DNA; 각 유전자좌를 표적화하는 gRNA 작제물; 및 2-피스 및 3-피스 생체내 조립체를 위한 마커 플라스미드 단편의 서열은 본원에서 서열번호 21 내지 29로서 제공된다.

[0244] Gal80, HO 및 ADH5 유전자좌의 표적 부위는 각각 표 5에 나타낸다.

표 5

CRISPR 표적 부위

ORF	표적 서열 (NGG 생략)	서열번호
Gal80	TAAGGCTGCTGCTGAACGT	서열번호 30
HO	CCGGCTTGATCGACTCAGA	서열번호 31
ADH5	GGGTCATTGGTATCGATGG	서열번호 32

[0245]

세포를 형질전환시키고, 실시예 1에 기재된 바와 같이 배양하고, 선별시 나타나는 콜로니는 결실 작제물의 외부에 결합하는 상류 전방 프라이머, 및 각각의 개방 판독 프레임 대신에 삽입된 짧은 링커 서열에 결합하는 역방 프라이머를 사용하여 cPCR에 의해 각 유전자좌에서 결실 작제물의 삽입에 대해 검정했다. 11개의 콜로니, 뿐만 아니라 네가티브 대조군("N")으로서 작용하는 모 콜로니 THAT를 각 전달 모드에 대해 검정했다.

[0246]

표 6

Gal80, HO 및 ADH5 유전자좌에서 링커 삽입의 cPCR 검증을 위한 프라이머 서열

프라이머 명칭	설명	서열	서열번호
Gal80-US-F	Gal80 유전자 좌 US FOR	CAAACGGCCGCTCTGCCATGGC AAAGAATGCTTTCCA	서열번호 33
HO-US-F	HO 유전자 좌 US FOR	ACGTGTGTGTCTCATGGAAATTG ATGCAGTTGAAGACA	서열번호 34
ADH5-US-F	ADH5 유전자 좌 US FOR	GGCGTTATATCCAAACATTTTCAG ACAGAAGATT	서열번호 35
R_	링커 REV	AGGTCGCGCCGGCGTTGGACGAGC G	서열번호 36

[0247]

도 13에 도시된 바와 같이, 3개의 유전자좌 모두에서 동시 삼중 삽입 비율은 단지 2개의 HR 사상이 필요할 때와 비교하여 마커 벡터를 조립하기 위해 3개의 HR 사상이 필요할 때 실질적으로 더 높았다. 특히, 마커 벡터의 2-피스 생체내 조립체의 경우, 6/11 콜로니는 Gal80 유전자좌에서 삽입을 가졌고, 10/11은 HO 유전자좌에서 삽입을 가졌고, 7/11 콜로니는 ADH5 유전자좌에서 삽입을 가졌고, 5/11 콜로니(45.4%)는 3개 유전자좌 모두에서 삽입을 가졌다. 이에 비해, 마커 벡터의 3-피스 생체내 조립체의 경우, 9/11 콜로니는 Gal80 유전자좌에서 삽입을 가졌고, 10/11은 HO 유전자좌에서 삽입을 가졌고, 10/11 콜로니는 ADH5 유전자좌에서 삽입을 가졌고, 9/11 콜로니(81.8%)는 3개 유전자좌 모두에서 삽입을 가졌다. 따라서, 2-피스 겹 수복 대신 마커 벡터의 3-피스 겹 수복을 필요로 하는 것은 삼중 삽입의 비율을 거의 2배로 개선시켰다.

[0248]

다중 삽입 비율에서 이러한 개선이 또한 예스. 세레비지아의 이배체 균주에서도 볼 수 있는지를 판단하기 위해, 이배체 효모 균주 CAT-1의 Cas9-발현 세포는 유사하게 Gal80, HO 및 ADH5 개방 판독 프레임, 각 유전자좌를 표적화하는 gRNA 작제물, 및 2 또는 3 중복 피스로 단편화된 선택적 DNA의 동시 범-대립유전자 마커 부재 결실을 위한 공여체 DNA로 형질전환시켰다. 콜로니를 결실 작제물이 외부 상류 전방 프라이머 및 각 개방 판독 프레임 대신 삽입된 짧은 링커 서열에 결합하는 역방 프라이머를 사용하여 cPCR에 의해 검정했다(표 6).

[0249]

- [0250] 도 14에 도시된 바와 같이, 이배체 균주의 3개 유전자좌 모두에서 동시 삼중 삽입 비율은 또한 마커 벡터를 조립하기 위해 3개의 HR 사상이 필요할 때 실질적으로 더 높았다. 특히, 마커 벡터의 2-피스 생체내 조립체의 경우, 3/24 콜로니는 Gal80 유전자좌에서 삽입을 가졌고, 7/24는 HO 유전자좌에서 삽입을 가졌고, 2/24 콜로니는 ADH5 유전자좌에서 삽입을 가졌고, 1/24 콜로니(4.2%)는 3개 유전자좌 모두에서 삽입을 가졌다. 이에 비해, 마커 벡터의 3-피스 생체내 조립체의 경우, 3/8 콜로니는 Gal80 유전자좌에서 삽입을 가졌고, 5/8은 HO 유전자좌에서 삽입을 가졌고, 3/8 콜로니는 ADH5 유전자좌에서 삽입을 가졌고, 3/8 콜로니(37.5%)는 3개 유전자좌 모두에서 삽입을 가졌다. 따라서, 2-피스 겹 수복 대신 마커 벡터의 3-조각 겹 수복을 필요로 하는 것은 이배체 균주에서 삼중 삽입의 비율을 거의 10배로 개선시켰다. 실험으로부터 회복된 콜로니의 수는 또한 3개 사상이 필요했을 때(데이터는 나타내지 않음) 대략 10배 적어, 마커 벡터의 보다 높은 차수의 조립체를 필요로 하는 것은 단지 집단의 대부분의 HR 적격 세포를 위해 선택함을 시사한다.
- [0251] **6.4 실시예 4: 겹 수복과 함께 CRISPR을 사용하여 단일 및 다중 점 돌연변이의 도입**
- [0252] 이 실시예는 정확한 점 돌연변이 또는 점 돌연변이에 대한 수정을 도입하기 위한 실시예 1에 기재된 바와 같이 최적화된 프로토콜의 적용(모드 4: gRNA 카세트(들)의 마커 벡터 골격으로의 생체내 HR-매개 도입)을 입증한다.
- [0253] 현재, 단일 유전자좌에 점 돌연변이의 도입은 지루한 공정이다. *Delitto Perfetto* 방법은 점 돌연변이의 마커 부재 도입을 가능하게 하지만, 표적 부위의 근방에 유도성 메가뉴클레아제를 함유하는 마킹된 카세트의 삽입을 필요로 한다[참조: Storici *et al.*, *Proc Natl Acad Sci USA* 100(25):14994-9 (2003)]. 또는, URA3 마커를 포함하는 복잡한 삽입 카세트가 설계되어 5-FOA 카운터 선택에 의한 URA3의 후속적 루프 아웃이 포함된 점 돌연변이를 갖는 유전자좌를 재구성하도록 표적 부위에 삽입될 수 있다. 이러한 방법 모두는 필수 유전자에 문제가 되고, 유전 공학의 적어도 두 라운드를 필요로 하고, 다중화에 적합하지 않다.
- [0254] CRISPR에 의한 점 돌연변이(즉, 미스센스 SNP)의 도입을 위해 몇몇 고려사항이 있다. 첫째, 게놈에서 고유한 것 이외에, 절단을 위해 표적화된 부위는 목적하는 SNP의 부위에 가능한 한 근접해야 한다(도 15). 이는 절단 부위를 수복하기 위한 재조합이 목적하는 SNP의 도입을 필요로 하지 않기 때문이고, 이의 포함 가능성은 절단 부위로부터 거리를 감소시킬 것으로 예상된다. 둘째로, 최적 효율을 위해, 공여체 DNA는 그것이 CRISPR을 위한 표적이 아니도록 설계되어야 한다. 사실, 이 문제는 이들 실험에서 관찰된 CRISPR에 의한 SNP 삽입의 낮은 비율을 설명하기 위해 디카를로 등(DiCarlo *et al.*)에 의해 인용되었다[참조: DiCarlo *et al.*, *Nucleic Acids Res.*, 7, 4336-4343 (2013)]. 절단을 피하기 위해, 목적하는 SNP는 대부분의 유전자좌에서 충족시키기에 불가능한 요건인 공여체 DNA 중의 CRISPR 표적 부위를 파괴할 필요가 있다. 절단에 대한 공여체 DNA 면역을 수행하고 동시에 재조합 사상이 목적하는 SNP를 포함하는 가능성을 향상시키기 위해, 이중성 블록 접근법이 채택되었고, 이에 의해 침묵 돌연변이가 표적 부위와 점 돌연변이 사이의 코돈에서 수행되어 목적하는 SNP를 생략할 재조합 사상의 가능성을 감소시킨다(도 15). 추가로, 이중성 블록의 삽입은 PCR에 의해 후보 클론을 동정하기 위한 신규 프라이머 결합 부위를 제공한다.
- [0255] 원리의 증명으로서, 돌연변이 유발을 시행한 효모 세포(에스. 세레비지애)의 돌연변이 대립유전자는 상기 접근법을 사용하여 상응하는 야생형 대립유전자로의 치환을 위해 표적화했다. 돌연변이 균주는 실시예 1에 기재된 바와 같이 중간 강도 FBA1 프로모터의 조절하에 Cas9를 구성적으로 발현시키기 위해 제조했다. 이어서, Cas9p-발현 균주는 TRS31, CUE5, ECM38, PGD1, SMC6, NT01 및 DGA1 개방 판독 프레임 각각을 표적화하는 공여체 DNA (단일 삽입 사상의 경우, 한번에 하나) 및 각각 선형 NatA 마커 플라스미드 골격과 중복 상동성을 포함하여 겹 수복을 통해 환상 플라스미드의 생체내 조립을 가능하게 하는 각 유전자좌를 표적화하는 gRNA 작제물로 형질전환시켰다. 세포를 형질전환시키고, 실시예 1에서 기재된 바와 같이 배양한 다음, 7개 유전자좌 각각에 점 돌연변이(복귀 대립유전자)의 도입에 대해 평가했다. 후보 콜로니 및 모 네가티브 대조군(c)은 이중성 블록 및 인접 서열에 대해 콜로니 PCR에 의해 검정하고, 선택된 포지티브 콜로니는 삽입 유전자좌에 걸쳐 큰 PCR 생성물을 서열분석하여 확인했다.

표 7

대립유전자 교환의 cPCR 검정을 위한 프라이머 서열

프라이머 명칭	설명	서열	서열번호
TRS31-US-F	TRS31 유전자 좌 US FOR	GTGCATTGGCTCGAGTTGCTG	서열번호 37
TRS31-DS-R	TRS31 유전자 좌 DS REV	GGGAAGTTATCTACTATCATATA TTCATTGTCACG	서열번호 38
TRS31- het-R	이종성 블록 프라이머	GAAAAGTAGAGATTCAGAATAG ATCCTTGAC	서열번호 39
CUE5-US-F	CUE5 유전자 좌 US FOR	GGAAGGTATCAAGGATTCTTCTC TCC	서열번호 40
CUE5- DS-R	CUE5 유전자 좌 DS REV	GAGGTGGCACATCTTCATCATCT TC	서열번호 41
CUE5- het-R	이종성 블록 프라이머	CCAATAACTCATCCTGCTCCAAT TGT	서열번호 42
ECM38-US-F	ECM38 유전자 좌 US FOR	CAGACGCTGCAGTAACACAAGC	서열번호 43
ECM38-DS- R	ECM38 유전자 좌 DS REV	CTGAAGTGGGCAGTCCATGC	서열번호 44
ECM38- het- R	이종성 블록 프라이머	CAGTGATCTGGATCGTAGAAGGG C	서열번호 45
PGD1-US-F	PGD-1 유전자 좌 US FOR	CCAAGAGCATGCCACGGTTG	서열번호 46
PGD1-DS-R	PGD-1 유전자 좌 DS REV	GAGTTCCCATAGTACTACCGC	서열번호 47

[0256]

PGD1- het-R	이종성 블록 프라이머	GCAGACCTTATCTCTGTCTCG	서열번호 48
SMC6-US-F	SMC6 유전자 좌 US FOR	GAGCTACTTTCTACTGACTGCGC	서열번호 49
SMC6-DS-R	SMC6 유전자 좌 DS REV	GCGCTTCAATAGTAGTACCATCA GATG	서열번호 50
SMC6- het-R	이종성 블록 프라이머	GCCGTTCTCTGATCTCAAAGAGA AT	서열번호 51
NT01-US-F	NT01 유전자 좌 US FOR	CTCAGTATGACATGGATGAACAG GATG	서열번호 52
NT01-DS-R	NT01 유전자 좌 DS REV	GGTACCTCCTGTAAGCTCCCTTTT C	서열번호 53
NT01- het-R	이종성 블록 프라이머	GACTGAGACGTTCTGGACTCCTT C	서열번호 54
DGA1-US-F	DGA1 유전자 좌 US FOR	CTTAACCAAGCACGACAGTGGTC	서열번호 55
DGA1-DS-R	DGA1 유전자 좌 DS REV	GATTCCTAGCGCCACCAAC	서열번호 56
DGA1- het-R	이종성 블록 프라이머	CCTCTCCGGTGGCTGGTGATCTG	서열번호 57
ADH2-US-F	ADH2 유전자 좌 US FOR	CGAGACTGATCTCCTCTGCCGGA AC	서열번호 58
ADH2-DS-R	ADH2 유전자 좌 DS REV	GAATACTTACCACCGAGCGAG	서열번호 59
ADH2- het-R	이종성 블록 프라이머	GCATGTAAGTCTGTATGACATAC TCCTG	서열번호 60
SIN4-US-F	SIN4 유전자 좌 US FOR	CAAACGTCCTAAATGACCCATCG TTG	서열번호 61
SIN4-DS-R	SIN4 유전자 좌 DS REV	CAACTTCGGGTTTTGTGTGGTT AG	서열번호 62
SIN4- het-R	이종성 블록 프라이머	CAATGGCAATTACCGTAGTTGA AACCG	서열번호 63
CYS4-US-F	CYS4 유전자 좌 US FOR	CTCCAGAATCACATATTGGTGTT GC	서열번호 64
CYS4-DS-R	CYS4 유전자 좌 DS REV	CCATCTTAGTAACGATATGGATT GGTTTC	서열번호 65
CYS4-het-R	이종성 블록 프라이머	CTGATGGAGTCAGGAAAGATGGC	서열번호 66

[0257]

[0258]

TRS31, CUE5, ECM38, PGD1, SMC6, NT01, DGA1, ADH2, SIN4 및 CYS4 유전자좌의 각각을 각각 표적화하는 공여체 DNA의 서열; 각 유전자좌의 표적 서열 및 각 유전자좌를 표적화하는 gRNA 작제물은 본원에서 서열번호 67-96로서 제공된다.

[0259]

도 16에 제시된 바와 같이, 높은 비율의 이종성 블록 삽입은 각각의 유전자좌에서 관찰되었고(36.4% 내지 90.9% 범위), 목적하는 돌연변이에 걸친 PCR 단편의 서열분석은 목적하는 대립인자를 함유한 대부분의 클론을 나타냈다.

[0260]

점 돌연변이의 다중 도입의 용이성을 결정하기 위해, 3개 유전자좌(ECM38, PGD1 및 ADH2)는 다중 gRNA를 위한 최적 전달 모드를 사용하여 이종성 블록 삽입(대립인자 치환)을 위해 동시에 표적화했다. 도 17에 제시된 바와 같이, 높은 비율의 삼중 이종성 블록 삽입이 PCR 분석에 의해 관찰되었다(90.9 내지 100%). 보다 고차의 다중 삽입이 용이한지를 결정하기 위해, 5개 유전자좌(ADH2, PGD1, ECM38, SIN4 및 CYS4)를 유사한 방식으로 동시에 표적화했다. 도 18에 제시된 바와 같이, 동시 오중 이종성 블록 삽입은 2/11 콜로니(18.2%)에서 cPCR 검정에 의해 확인되었다.

[0261]

원리의 제2 증명으로서, 11개의 상이한 돌연변이 대립인자를 천연 에스. 세레비지에 내로 도입하고, 생성되는 균주는 이들 SNP에 의해 부여된 표현형을 시험했다. 검사된 대립인자 중에서 하나의 대립인자는 보다 신속한 침강(ACE2 S372*)을 부여하는 산업적 발효와 관련되었고[참조: Oud, Guadalupe-Medina et al., *Proc Natl Acad Sci U S A* 110(45): E4223-4231, 2013], 세포 분열 및 분비 경로에 필수적인 유전자에서 일련의 온도 감응성 대립인자(SEC1 G443E, SEC6 L633P, MYO2 E511K, CDC28 R283Q)[참조: Lorincz and Reed, *Mol Cell Biol* 6(11):

4099-4103, 1986; Roumanie, Wu et al. *J Cell Biol* 170(4): 583-594, 2005] 및 또 다른 쌍은 개선된 고온 성장과 관련되었다(NCS2 H71L 및 END3 S258N)[참조: Sinha, David et al., *Genetics* 180(3): 1661-1670, 2008; Yang, Foulquie-Moreno et al., *PLoS Genet* 9(8): e1003693, 2013]. 또한, 상승된 에탄올 농도에 대한 내성과 연관된 일련의 5개 대립인자가 시험되었다(SPT15 F177S, SPT15 Y195H, SPT15 K218R, PRO1 D154N 및 PUT1 결실)[참조: Takagi, Takaoka et al., *Appl Environ Microbiol* 71(12): 8656-8662, 2005; Alper, Moxley et al., *Science* 314(5805): 1565-1568, 2006].

[0262] 높은 비율의 이중성 블록 삽입(>90%)은 대부분의 개개 대립인자의 도입에 대해 관찰되었고(도 22), 목적하는 돌연변이에 걸친 PCR 단편의 후속 서열분석은 이들 변화를 확인했다. 온도 감응성 돌연변이는 이들의 의도된 표현형을 확인하기 위해 허용 및 제한 온도에서 검정했다. ACE2의 절단에 의해 유발된 분열 동안 세포의 불완전한 분리는 세포의 현저한 응집과 함께 명시야 현미경에 의해 확인되었다(도 23A). CDC28, MYO2 및 SEC1의 온도-감응성 대립인자는 예상된 바와 같이 37°C에서 성장하지 못했고(도 23B), CDC28 대립인자 균주는 성장의 G1 단계에서 중단했다(도 23C). SEC1 및 SEC6 돌연변이에서 제한 온도에서 분비 결함을 입증하기 위해, 세포외낭 복합체 성분 SEC3은 분지 작용의 리포터로서 기능하도록 이의 내인성 유전자좌(또한 CRISPR 사용)에서 카복시-말단 GFP-태그했다. SEC3은 야생형 세포에서 배아에 정상적으로 국지화되지만, 이의 국지화는 분비 돌연변이 둘다에서 명백하게 파괴된다(도 23D).

[0263] 다수의 경우에, 표현형은 다중 대립인자의 상승작용으로부터 기인하지만, 이러한 균주의 조작은 보다 시간 소비적이고, 종종 시험되지 않는다. 최적화된 다중 방법은 이러한 문제에 적용되었다. 천연 CENPK2는 고온 성장(MKT1 D30G)를 위해 하나의 대립인자를 포함하고[참조: Yang, Foulquie-Moreno et al. 2013], 2개의 추가 돌연변이는 NCS2 및 END3에 도입되었다. 소정 온도 범위에서 밤새 성장시키는 경우, 개개 대립인자의 어느 것도 효과를 갖지 않았다. 그러나, 단일 단계로 삽입된 두 추가 대립인자를 함유하는 균주는 42.7°C 이하의 온도에서 생존했다(도 23E).

[0264] 에탄올 내성 대립인자는 또한 상승작용 효과를 부여했다. 야생형 CENPK2는 이 실험에서 최대 17.5% 에탄올에 내성이 있었다(도 23F). SPT15 중의 돌연변이는 최대 20% 에탄올에 대해 내성을 증가시켰다(도 23F). 이들 대립인자의 상호작용을 검사하기 위해, 3개 유전자좌에 걸쳐 5개 표적화된 변화를 천연 균주에 동시에 도입했다(SPT15, PRO1 D154N에서 3개 돌연변이 및 PUT1의 결실). SPT15 중의 3개 돌연변이는 3개 대립인자를 함유하는 유전자의 약 150bp를 절제하기 위해 2개 gRNA를 사용함으로써 단일 공여체 DNA 상에 도입했다. 생성되는 클론의 27%는 콜로니 PCR에 의해 평가하고 서열분석에 의해 확인한 바와 같이 모두 5개 변형을 함유했다(도 22). 수득되는 균주는 최고 에탄올 내성, 최대 22.5% 에탄올 내성을 가졌다(도 23F). 이들 결과가 입증하는 바와 같이, 다중 CRISPR은 유발 대립인자의 조합에 대한 기술의 신속한 평가를 가능하게 한다.

[0265] 이들 결과는 정확한 점 돌연변이 또는 회귀가 본원에 제공된 CRISPR-매개된 게놈 삽입을 위한 최적화된 방법 및 조성물을 사용하여 높은 유효성 및 높은 다중도로 달성될 수 있음을 입증한다.

[0266] 6.5 실시예 5: 이배체 세포의 바이-대립인자 조작

[0267] 본 실시예는 이배체 효모 균주에서 동시 바이-대립인자 (bi-allelic) 삽입을 위한 실시예 1에 기재된 바와 같은 최적 프로토콜(모드 4: 마커 벡터 골격 내로 gRNA 카세트(들)의 생체내 HR-매개된 도입)의 적용을 입증한다.

[0268] 이배체 산업적 SC 균주는 맵핑되지 않지만 발효를 위한 유리한 특성을 갖는 고도로 헤테로접합성이다. 그러나, 이들 균주는 표준 방법에 의해 조작하는 것이 곤란하고, 유전자를 결실시키거나 바이-대립인자 조작을 도입하기 위해 2개의 연속 삽입 단계 및 상이한 마커를 필요로 한다. 따라서, 에스 세레비니아에(*S. cerevisiae*)의 CAT-1 및 PE-2 이배체 산업용 균주에서 GAL80 유전자좌의 CRISPR-매개된 바이-대립인자 결실의 효율은 실시예 1의 최적화 프로토콜을 사용하여 시험했다. Gal80 유전자좌를 표적화하는 공여체 및 GRNA 서열은 실시예 3에 기재되어 있다.

[0269] 도 19에 제시된 바와 같이, Gal80 유전자좌를 표적화하고 동시-형질전환된 선형 NatA 마커 벡터 골격에 상동성 말단을 갖는 공여체 DNA 및 선형 gRNA로 형질전환된 Cas9-발현 균주의 cPCR은 CAT-1 이배체 세포에서 100%의 바이-대립인자 공여체 삽입율(도 19B) 및 PE-2 이배체 세포에서 90%의 삽입율(도 19C)을 나타냈다. 이 비율은 반수체 CENPK2 균주에서 동일한 결실이 수득되는 비율에 필적한다(100%; 도 19A). 이들 결과는 이배체 숙주 세포를 조작하기 위해 본원에 제공된 CRISPR-매개된 게놈 삽입을 위한 최적화 방법 및 조성물의 유효성을 입증한다.

[0270] 6.6 실시예 6: 완전 생합성 경로의 다중 삽입

[0271] 본 실시예는 전체 생합성 경로의 천연 효모 균주 내로 동시 삽입을 위한 실시예 1에 기재된 최적화 프로토콜(모

드 4: 마커 벡터 골격 내로 gRNA 카세트(들)의 생체내 HR-매개된 도입)의 적용을 입증한다.

[0272] 전형적으로, 조작 대사 경로는, 에스. 세레비지에 등의 취급이 용이한 숙주에서도, 시간 소비적이다. 이러한 시간은, 유전자 카세트의 삽입이 약물 선별가능한 마커의 임의의 삽입을 필요로 하지 않고서 병렬로 수행되는 경우에 크게 개선될 것이다. 따라서, CRISPR-매개된 삽입을 위한 최적화 프로토콜은 에스. 세레비지에 내로, 파르네센의 생성을 위한 기능적 경로를 코딩하는(미국 특허 제8,415,136호 참조), 대략 30kb의 총 DNA의 12개 유전자 카세트의 동시 삽입에 적용했다. 유전자 카세트는 미국 특허 제8,221,982호 및 제8,332,160호에 기재된 폴리뉴클레오티드 조립 방법을 사용하여 디자인 및 클로닝했다. 경로는 12개 유전자의 삽입을 위한 3개 공여체 작제물로 나누었다: 아세틸-CoA 티올라제를 코딩하는 ERG10; HMG-CoA 신타제를 코딩하는 ERG13; HMG-CoA 리덕타제를 코딩하는 tHMG1(2카피); 메발로네이트 키나제를 코딩하는 ERG12; 포스포메발로네이트 키나제를 코딩하는 ERG8; 메발로네이트 피로포스페이트 데카복실라제를 코딩하는 ERG19; 이소펜테닐 포스페이트 이소머라제를 코딩하는 IDI1; 아르테미시아 안누아(*Artemisia annua*)의 파르네센 신타제(2카피); 파르네실 포스페이트 신타제를 코딩하는 ERG20; 및 전사 조절인자 GAL4. 3개 작제물은 천연 CENPK2 에스 세레비지에에서 Gal80, HO 및 BUD9 유전자좌 내로의 삽입을 위해 표적화했고, 배지 강도 FBA1 프로모터(서열번호 3)의 조절하에 Cas9는 GRE3 유전자좌로 표적화했다. 동시에, 모든 3개 작제물의 마커-레스 삽입은 실시예 1에 기재된 최적화 캡 수복 방법을 사용하여 시험했고, 클론은 각 공여체 작제물 내의 삽입 표적 유전자좌의 5' 측면 및 내부 링커 서열에 결합하는 cPCR 프라이머 쌍에 의해 검정했다.

[0273] 도 20에 제시된 바와 같이, 스크리닝된 47개 중에서, 11개 클론(23.4%)은 전체 파르네센 경로를 구성하는 30kb의 삽입에 포지티브였다(클론 22, 24, 29-32, 41-43, 45 및 46). 48번째 클론은 모 네가티브 대조군이였다. 모든 삼중 포지티브 후보는 또한 각 표적 유전자좌의 3' 측면에서 cPCR에 의해 후속적으로 검정되었고, 두 측면에서 삽입에 포지티브인 것으로 확인되었다. 이어서, 11개 cPCR 포지티브 클론은 배지 슈크로즈 플레이트 모델 검정에서 파르네센 생성에 대해 시험했다. 도 21에 제시된 바와 같이, 모든 11개 클론은 0.1 내지 대략 1.5g/L 양 범위의 양으로 파르네센을 생성했다. 함께 취하면, 이들 결과는 신속한 다중 대사 조작을 위해 본원에 제공된 CRISPR-매개된 게놈 삽입을 위한 최적 방법 및 조성물의 유효성을 입증한다. 다중 조작을 위한 최적화 프로토콜은 복합 경로의 조작을 위해 시간을 현저히 단축시키기 위해 적용될 수 있다. 예를 들면, 에스 세레비지에 균주에서 3개 점 돌연변이의 동시 도입은 대략 6주의 작업(각 대립인자를 도입하는데 1주, 각 마커를 재순환시키는데 1주 총 3회 사이클)을 필요로 하는 반면, 본원에 제공된 최적화된 방법을 사용하는 경우에 1주를 필요로 한다. 본원에 입증된 초보적 파르네센 경로의 도입 시간은 또한 병렬로 표적화된 유전자좌의 수를 갖는 시간 절약 규모의 양으로서 6배로 압축된다.

[0274] 개념의 추가 증거로서, 뮌콘산에 대한 신규 경로를 코딩하는 3개 유전자좌에 걸쳐 분포된 24kb의 DNA를 함유하는 11개 유전자 카세트의 다중 삽입은 반수체 CENPK2에서 시도되었다(도 24). 뮌콘산은 나일론-6,6, 폴리우레탄 및 폴리에틸렌 테레프탈레이트(PET)를 포함하는 생물플라스틱의 생성을 위한 큰 잠재성을 갖는 전구체 분자이다. 현재, 뮌콘산은 유기 합성을 통해 석유 유래 원료로부터 수득되지만, 재생 자원이 바람직하다. 뮌콘산의 생합성은 방향족 아미노산(쉬키메이트) 생합성 경로의 과발현에 의해 달성된다. 이전에, 뮌콘산의 고도의 생성(36.8g/L, 공급 배치 발효)은 이. 콜라이에서 달성되었다[참조: Niu, Draths et al., *Biotechnol Prog* 18(2): 201-211, 2002]. 그러나, 에스. 세레비지에를사용한 보다 낮은 pH 발효는 상기 프로세스의 하류 처리 및 산업화를 촉진시킬 것이다. 원칙 노력의 증명에서, 141mg/L 이하의 역가가 조작된 에스. 세레비지에 균주에서 진탕 플라스크 실험으로 관찰되었다[참조: Curran, Leavitt et al., *Metab Eng* 15: 55-66, 2013]. 에스. 세레비지에에서 이러한 초기 시험과 대조적으로, 본 실험은 천연 ARO1 유전자좌 보다는 이. 콜라이 쉬키메이트 경로 유전자 AROF, AROB 및 AROD를 사용했다. 조작된 경로는 도 24A에 제시되어 있다.

[0275] 삽입을 위해, 상기 경로는 CENPK에서 GAL80, HO 및 ARO1 유전자좌를 표적화하는 3개 분리된 작제물(생체내에서 상동성 재조합에 의한 재구성을 위해 내부 중첩)로 분할했다(도 24B). HO는 중성 유전자좌로서 선택되었고, GAL80은 갈라토스 오페론의 글루코즈 억제제를 제거하기 위해 선택되었고, ARO1 유전자좌는 조작된 경로를 통한 유동을 강제하기 위해 결실되었다. ARO1 결실은 또한 방향족 아미노산에 대해 균주를 영양요구성으로 되게 하여, 바이오매스 생성 단계(농후 배지) 및 생성 단계(최소 배지) 사이의 단순한 스위치 메카니즘을 생성한다. 동시에, 모든 3개 작제물의 마커-레스 삽입은 최적화 캡 수복 방법을 사용하여 시험했고, 클론은 PCR에 의해 검정했으며, 이는 삼중 삽입의 4.2% 비율을 나타냈다(n=48). 이는 이들 3개 작제물의 삽입이 9개 재조합 사상(유전자좌당 2개의 인접한 및 하나의 내부 사상)을 필요로 하는 것이 주목된다. 관찰된 비율은 다중 결실에 대해 관찰된 것보다 낮지만, 완전 생합성 경로의 도입은 피트니스 결함을 부여할 것으로 예상되고, 이는 적절하게 삽입된 균주의 회복을 제한할 수 있다.

- [0276] 뮤콘산 및 중간체의 생성은 HPLC에 의한 분석과 함께 96웰 진탕-플레이트 검정으로 시험했다. 1단계 삽입된 균주는 PCA의 높은 역가(약 3g/L)를 나타냈고, 이는 PCA를 카테콜로 전환시키는 AroY에서 병목을 나타낸다(도 24A). 하류 경로의 기능성을 확인하기 위해, 최대 1g/L 카테콜을 직접 생성 배지 웰에 공급하고, 시스-트랜스 뮤콘산으로 카테콜의 정량적 전환은 조작된 균주에서 관찰되었지만, 모 균주에서 관찰되지 않았으며, AroY에서 경로 디자인으로 단일 결함을 확실하게 확인한다.
- [0277] 2차 산업적 관련 효모에서 최적화된 갭 수복 방법의 유효성을 시험하기 위해, 시험은 케이. 락티스(*K. lactis*)에서 6개 유전자를 포함하는 뮤콘산 경로의 보다 컴팩트 버전을 삽입하기 위해 수행했다. 상기 경로는 DIT1, ADH1 및 NHT80 유전자좌를 표적화하는 3개 삽입 작제물로 분할했다. 천연 케이. 락티스 균주(ATCC 8585)는 Cas9를 GAL80 유전자좌에 삽입하고 YKU80을 결실시켜 비-상동성 말단 결합의 효과를 최소화시킴으로써 제조했다 [참조: Kooistra, Hooykaas et al., Yeast 21(9): 781-792, 2004; Wesolowski-Louvel, FEMS Yeast Res 11(6): 509-513, 2011]. 모든 3개 작제물의 마커-레스 삽입은 동일한 갭-수복 방법을 사용하여 pKD1 안정성 요소를 함유하는 플라스미드로 1단계로 달성했다 [참조: Chen, Wesolowski-Louvel et al., J Basic Microbiol 28(4): 211-220, 1988]. 삼중 삽입은 PCR(n=48)에 의해 검정된 바와 같이 2.1%의 비율로 발생했다. CENPK 결과와 유사하게, 높은 역가의 PCA(1g/L)는 관찰되었지만, 뮤콘산 생성은 관찰되지 않았다(도 24E). 카테콜 공급 실험은 ARO1 기능에서 동일한 결함을 확인했다. ARO1은 이러한 케이. 락티스 균주에서 결실되지 않았고, 이러한 불일치는 관찰된 보다 낮은 역가의 PCA를 설명할 수 있음이 주목된다. 그럼에도 불구하고, 이들 결과는 2개의 산업적 관련 효모 균주에서 프로토타입 뮤콘산 생성, 및 1개월 이내에 제한 효소, 신속한 디자인 반복을 촉진시키고 2개 잠재적 숙주의 샘플링을 가능하게 하는 작업 흐름을 동정하는 능력을 입증한다.
- [0278] **6.7 실시예 7: 포유동물 세포에서 개선된 삽입**
- [0279] 본 실시예는 포유동물 숙주 세포에서 개선된 삽입율을 달성하기 위해 실시예 1에서 기재된 최적화된 프로토콜 (모드 4: 마커 벡터 골격 내로 gRNA 카세트(들)의 생체내 HR-매개된 도입)의 적용을 입증한다.
- [0280] 에스. 세레비지에에 대해 상기 기재된 gRNA를 위한 갭 수복 전달 방법이 포유동물 세포에서 삽입율을 또한 개선시키는지를 시험하기 위해, 일련의 시약은 HEK-293T 세포에서 형질감염 시험을 위해 생성했다. 개관에 있어서, 세포는, AAVS1 유전자좌를 표적화하는 gRNA 카세트를 함유하는 단편, 및 후기 진단 목적(갭 수복 조건)을 위해 EcoRI 부위에 인접하는 상류 및 하류 상동성을 포함하는 상동성 재조합에 의해 유전자좌의 수복을 위한 공여체 DNA 단편과 함께, CD4 에피토프 태그의 5' 부분에 2A-링커를 통해 융합된 Cas9 발현 카세트를 함유하는 선형화 플라스미드 골격으로 동시-형질감염시켰다. 이러한 형질감염은 밀폐된 플라스미드에 함유된 Cas9-2A-CD4 발현 카세트 및 gRNA 카세트와의 대조군 반응과 비교했다(포지티브 대조군). 또한, gRNA를 함유하지 않는 플라스미드는, 공여체 DNA의 상동성 삽입이 CRISPR-Cas9의 부재하에 측정가능한 속도로 발생했는지를 평가하기 위해 네가티브 대조군으로 사용했다. 형질감염 후, CD4+ 세포(형질감염된 세포)는 항체-결합 자기 비드를 사용하여 분리했고, 세포를 용출시키고 게놈 DNA 제조에 사용했다. 삽입 부위를 포함하는 영역의 PCR을 수행하고, EcoRI를 사용하여 PCR 생성물을 소화시켜, 표적 부위에 공여체 DNA를 삽입한 세포의 분획을 측정했다.
- [0281] **재료 및 방법**
- [0282] Cas9 뉴클레아제 및 연관 gRNA의 발현. CD4 농후 키트(A21175)와 함께 LifeTech/GeneArt CRISPR 뉴클레아제를 사용했다. 제조업자의 지시에 따라, AAVS1 영역(AAVS1, 말리 등으로부터 T2 gRNA) 내부에 서열을 코딩하는 이중-가닥 올리고뉴클레오티드(올리고 CUT1216 및 CUT1217를 어닐링시켜 제조함)을 제공된 선형 벡터에 연결시켜 pAM3473(서열번호 98)을 생성했다. 플라스미드는 맥시-준비했다(Qiagen).
- [0283] 갭 수복을 시험하기에 적합한 pAM3473 버전의 생성. pAM3473 플라스미드를 Bst1107I 및 NheI로 소화시켜 CD4 OFR의 일부인 전체 gRNA 카세트를 제거했다. 골격은 CIP(알칼리 포스파타제) 처리하고, 겔 정제했다. 고유한 ClaI 및 XmaI 부위를 함유하는 다중 클로닝 부위(MCS) 이중-가닥 올리고는 선형 벡터에 대한 혼화성 오버행으로 CUT1214(서열번호 103) 및 CUT1215(서열번호 104) 올리고를 어닐링시켜 제조했다. 이중-가닥 올리고를 정제된 골격에 연결시켜 pAM3472(서열번호 97)를 생성했다. 플라스미드는 맥시-준비했다(Qiagen). 사용 전에, 플라스미드는 ClaI 및 NheI로 소화시켜 선형화하고, 에탄올 침전에 의해 정제/농축시켰다.
- [0284] CD4 에피토프 단독(및 gRNA 비함유)을 함유하는 대조군 플라스미드의 생성. pAM3473(서열번호 98)은 Bst1107I 및 PacI으로 소화시키고, 골격은 CIP 처리하고 겔 정제했다. gRNA 카세트-레스 골격을 재-환상화하도록 디자인된 이중-가닥 올리고는 CUT1254(서열번호 113) 및 CUT1255(서열번호 114) 올리고를 어닐링시켜 제조하고, 벡터 골격에 연결시켜 pAM15068(서열번호 102; 이전에 "A2"로 공지됨)을 생성했다.

[0285] 선형화된 pAM3472 내에서 AAVS1 gRNA 카세트의 겹 수복을 위한 단편의 제조. 프라이머 CUT 1220(서열번호 107) 및 CUT1221(서열번호 108)을 사용하여 pAM3473으로부터 2850bp 단편을 증폭시켰다. 생성물은 RYSE09 수용체 벡터 내로 겹 수복(이. 콜라이)에 의해 서브-클로닝시키고, 작제물은 pAM3515(서열번호 100)을 제조하기 위해 서열분석에 의해 확인했다. 형질전환 전에, 선형 단편은 인접 RYSE0(서열번호 117) 및 RYSE19(서열번호 118) 프라이머를 사용하여 퓨전(Phusion) PCR 증폭에 의해 제조하고, PCR 생성물은 앰푸어 비드(Axygen)을 사용하여 정제했다. 형질감염 전에, 선형 단편은 인접 RYSE0 및 RYSE19 프라이머를 사용하여 퓨전 PCR 증폭에 의해 제조하고, PCR 생성물은 앰푸어 비드(Axygen)을 사용하여 정제했다.

[0286] 선형화된 pAM3472 내에서 CD4-단독 대조군 단편의 겹 수복을 위한 단편 제조. 프라이머 CUT1220(서열번호 107) 및 CUT1252(서열번호 111)은, 주형으로서 pAM3473을 사용하여 CD4 ORF의 3' 말단을 함유하는 상류 단편을 증폭시키기 위해 퓨전 PCR 반응에 사용하고, 프라이머 CUT1253(서열번호 112) 및 CUT1221(서열번호 108)은 gRNA 카세트 하류의 인접 상동성을 함유하는 하류 단편을 증폭시키기 위해 사용했다. 이들 2개 단편을 겔 정제하고, 2차 융합 PCR 반응에 사용하고, 프라이머 RYSE0 및 RYSE19는 약 2kb의 증폭에 사용했다. 생성물은 RYSE09 수용체 벡터 내로 겹 수복(이. 콜라이)에 의해 서브-클로닝하고, 작제물은 pAM3516(서열번호 101)을 제조하기 위해 서열분석에 의해 확인했다. 형질감염 전에, 선형 단편은 인접 RYSE0 및 RYSE19 프라이머를 사용하여 퓨전 PCR 증폭에 의해 이러한 주형으로부터 제조하고, PCR 생성물은 앰푸어 비드(Axygen)을 사용하여 정제했다.

[0287] AAVS1 표적 유전자좌에서 EcoRI 부위의 도입을 위한 공여체 DNA의 제조. 프라이머 CUT1226(서열번호 119) 및 CUT1223(서열번호 120)은, 퓨전 폴리머라제를 사용하여 인간 게놈 DNA(HEK-293 세포로부터 유래)로부터 이의 3' 말단에 합성 EcoRI 부위를 함유하는 약 570bp 상류 단편을 증폭시키기 위해 사용했다. 프라이머 CUT1224(서열번호 109) 및 CUT1227(서열번호 110)은 동일한 인간 게놈 주형으로부터 이의 5' 말단에 EcoRI 부위를 함유하는 약 540bp 하류 단편을 증폭시키기 위해 사용했다. 단편은 인접 RYSE0 및 RYSE19 프라이머와 함께 퓨전 폴리머라제를 사용하여 융합 PCR에 의해 조립하고, 약 1100bp 단편을 겹 수복(이. 콜라이)에 의해 선형 RYSE09 벡터 내로 서브-클로닝시켰다. 작제물은 pAM3514(서열번호 99)를 제조하기 위해 서열분석에 의해 확인했다. 형질감염 전에, 선형 단편은 인접 RYSE0 및 RYSE19 프라이머를 사용하여 퓨전 PCR 증폭에 의해 이러한 주형으로부터 제조하고, PCR 생성물은 앰푸어 비드(Axygen)을 사용하여 정제했다.

[0288] 형질감염 실험. 70% 컨플루언트 부착 293T 세포는 제조업자의 지시에 따라 리포펙타민 3000을 사용하여 DNA로 형질감염시켰다(LF300에 대해 1.5배 DNA 비율). 표 8은 각 형질감염에 사용된 DNA 작제물 및 DNA 양을 제공한다(이중으로 수행됨).

표 8

	벡터		선형 단편				
	선형	밀폐됨	gRNA 포함 유	선형	전체 겹	CD4 겹	ug 전체 DNA
형질감염	pAM3472	pAM3473	"A2"	pAM3514	pAM3515	pAM3516	
2			10	5			15
3	10						10
4	10					5	15
6		10		5			15
8	10			5	5		20

[0290] 48시간 후, 세포는 TrypIE 시약(LifeTech)를 사용하여 수확하고, CD4+ 세포는 Dyanbeads CD4 포지티브 단리 키트(LifeTech)를 사용하여 정제했다. 결합된 세포는 제조업자의 지시에 따라 비드로부터 용출시키고, 게놈 DNA는 제조업자의 지시에 따라 PrepGem 조직 키트(Zygem)를 사용하여 제조했다.

[0291] EcoRI 부위의 삽입을 위한 RFLP 검정. RFLP 검정은, 의도된 게놈 삽입의 문맥에서 공여체 DNA만이 생성물을 생성하도록, "외부" 프라이머와 함께 EcoRI 삽입 부위를 포함하는 프라이머 세트(CUT1297(서열번호 116) 및 CUT1294(서열번호 115))와 함께 퓨전 폴리머라제를 사용하여 증폭시킨 PCR 단편(920bp) 상에서 수행했다. 단편은 앰푸어 비드(Axygen)를 사용하여 정제했고, EcoRI로 소화시켰다. 상동성 제조법에 의해 삽입된 EcoRI 부위를 갖는 PCR 생성물은 348bp 및 572bp 단편을 제공했다. 삽입된 EcoRI 부위를 갖는 주형의 분획은 하기 식을 사용하여 밀도계(Image J)에 의해 계산했다: 소화 밴드 밀도(348bp + 572bp 밀도)/전체 밀도(348bp + 572bp + 920bp 밀도).

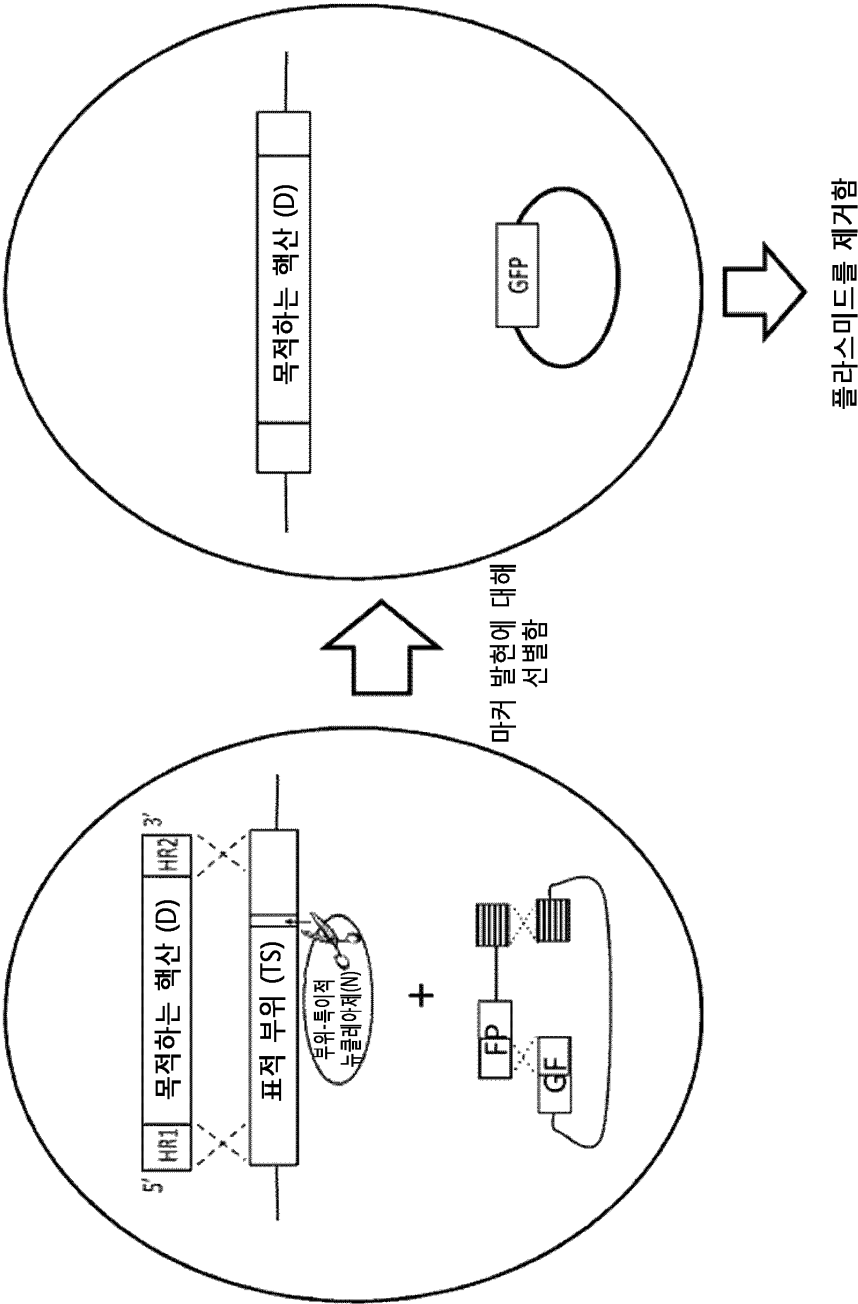
[0292] **결과**

[0293] 겹 수복에 대한 요건을 부과하는 것이 상동성 삽입의 비율을 증가시킬 수 있는지를 시험하기 위해, 본 발명자들은 플라스미드 및 선형 DNA의 몇몇 상이한 조합물로 형질감염시킨 HEK-293T 세포 중의 EcoRI 부위 공여체 DNA 삽입 비율을 비교했다(표 8 및 도 25). EcoRI 공여체 DNA가 CRISPR-Cas9에 의한 표적화된 절단의 부재하에 일부 측정가능한 수준으로 AAVS1 유전자좌에서 삽입할 수 있는지를 평가하기 위해, 세포는 Cas9-2A-CD4 ORF를 함유하지만 gRNA 카세트 및 선형 공여체 DNA(pAM3514 PCR 생성물)를 함유하지 않는 플라스미드 pAM15068로 형질감염시켰다. 형질감염된 세포는 dynabeads를 사용하여 정제하고, 게놈 주형을 제조하고, 삽입 부위에 걸친 PCR 생성물의 소화는 어떠한 소화 생성물도 제공하지 않았으며, 이는 EcoRI 부위 삽입의 비율이 CRISPR-Cas9의 부재하에 측정 불가능함을 나타낸다(도 25, 형질감염 2). 완전 CD4 ORF를 결여하는 선형 pAM3472 플라스미드가 이의 자신에게 CD4+ 표현형을 부여할 수 없는 것을 확인하기 위해, 형질감염을 이러한 선형 단편만으로 수행했다. 이들 정제된 세포로부터 제조된 주형으로부터 어떠한 PCR 생성물도 수득되지 않았고, 이는 PCR 반응을 위한 주형으로서 작용하기 위해 형질감염된 세포와 dynabead와의 불충분한 연관성이 있음을 나타낸다(도 25, 형질감염 3). 겹 수복이 CD4 ORF를 재구성할 수 있음을 확인하기 위해, 선형화된 pAM3472 플라스미드를 CD4 겹 수복 단편(pAM3516 PCR 생성물)로 동시-형질감염시키고, 이들 형질감염으로부터 정제된 세포로부터의 주형은 920bp 밴드를 제공했지만, 어떠한 gRNA도 존재하지 않기 때문에, 어떠한 소화 생성물도 제공하지 않았다(도 25, 형질감염 4). 이어서, AAVS1 gRNA를 포함하는 형질감염을 검사했다. AAVS1 gRNA를 갖는 CRISPR-Cas9 시스템의 기능을 위한 기준선을 확립하기 위해, 본 발명자들은 Cas9, CD4 및 gRNA(pAM3473) 및 선형 공여체 DNA 작제물(pAM3514 PCR 생성물)에 대한 발현 카세트를 함유하는 밀폐된 플라스미드를 동시-형질감염시켰다. PCR 생성물의 EcoRI 소화는 572 및 348bp의 탈락 생성물을 나타냈고, 이는 전체 PCR 생성물의 분획의 소화를 나타낸다(도 25, 형질감염 6). Image J 소프트웨어를 사용한 밴트 밀도의 정량화는 22.5%의 전체 주형이 EcoRI 부위를 함유하는 것을 나타냈다. 겹 수복이 이러한 비율을 개선시킬 수 있는지를 평가하기 위해, 본 발명자들은 CD4의 결실 부분 및 gRNA 카세트(pAM3515 PCR 생성물)를 함유하는 선형화된 pAM3472 벡터 및 겹 수복 단편으로 밀폐된 벡터를 치환시켰다(도 25, 형질감염 8). 밀도측정 프로세스를 반복하면, 본 발명자들은 47.5%의 전체 주형이 EcoRI 부위를 함유하는 것을 관찰했다. 이는 밀폐된 플라스미드(형질감염 6)에 대해 관찰된 비율에 비해 2.1% 배수의 개선을 나타내고, 따라서 게놈 삽입을 위한 개선된 겹 수복 방법의 유효성을 포유동물 세포에서 확인한다.

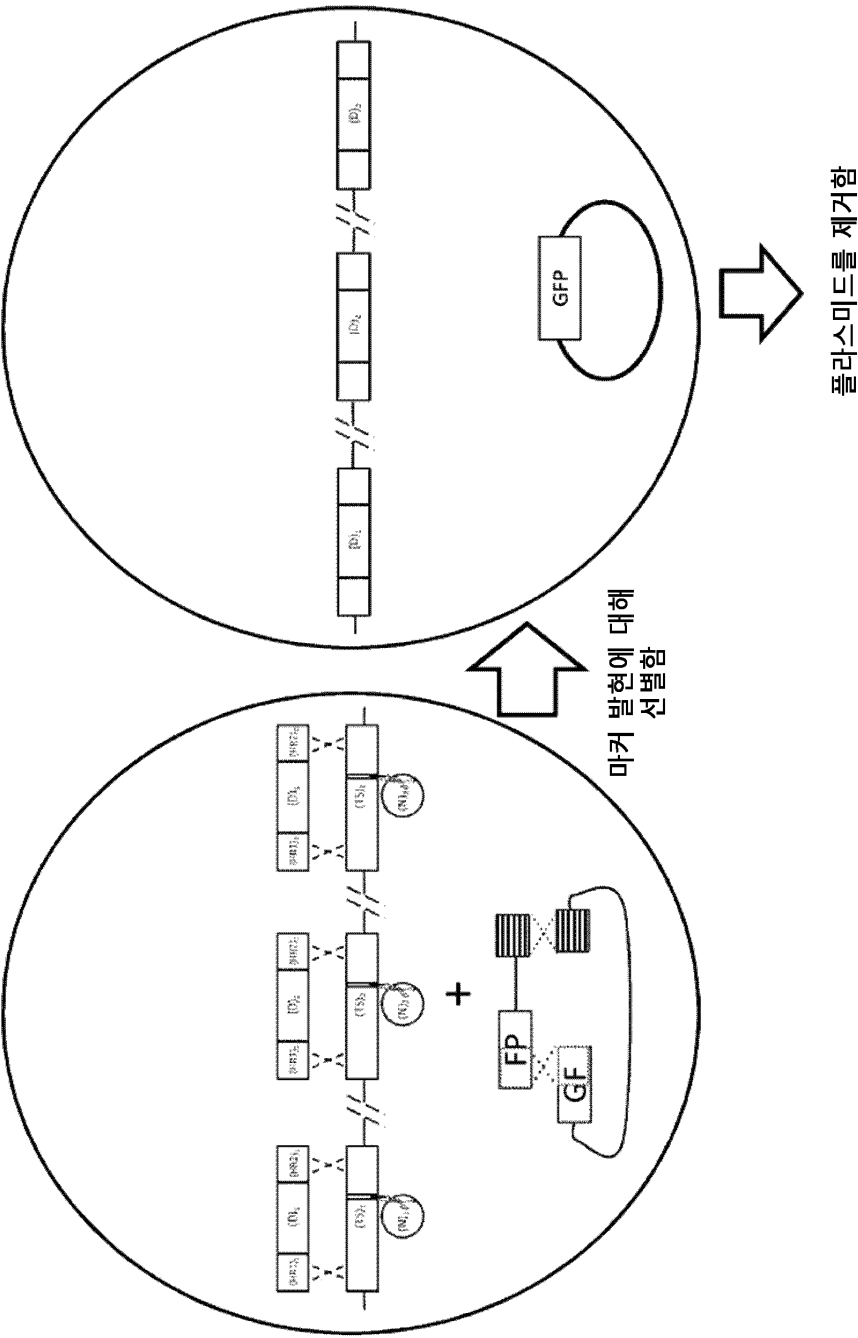
[0294] 본 출원에서 언급한 모든 공보, 특허 및 특허 출원은 각각의 개별 공보 또는 특허 출원이 참조로서 구체적이면서 개별적으로 나타나 있는 것처럼, 본원에 참조로서 포함된다. 전술한 발명이 이해의 명확성을 위해 예시 및 실시예로서 좀더 상세히 기술되었지만, 당해 기술분야의 숙련가는, 특정 변화 및 변형이 첨부된 청구범위의 사상 또는 범위를 벗어나지 않으면서 이루어질 수 있음을 본 발명의 교시에 비추어 용이하게 알 것이다.

도면

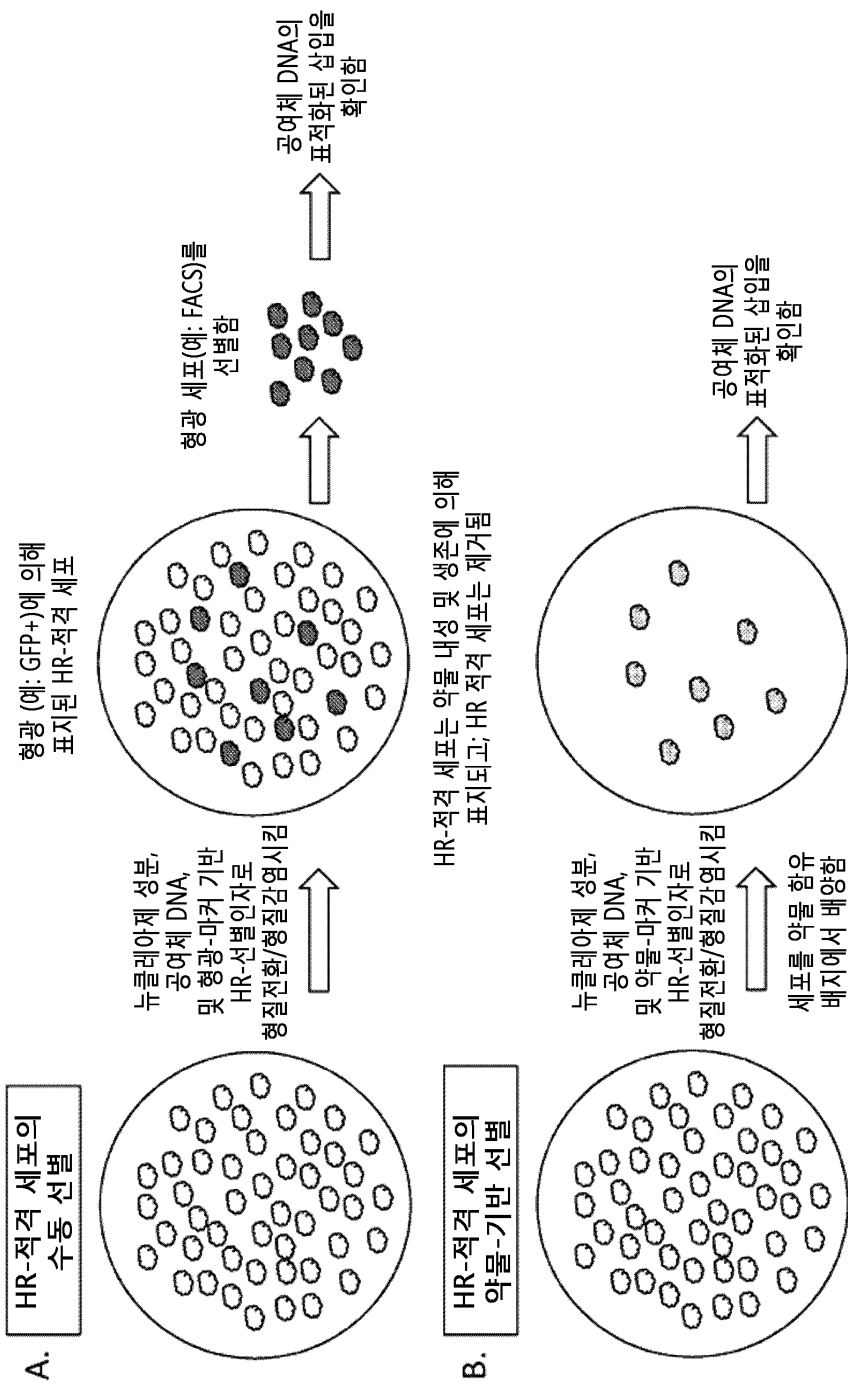
도면1



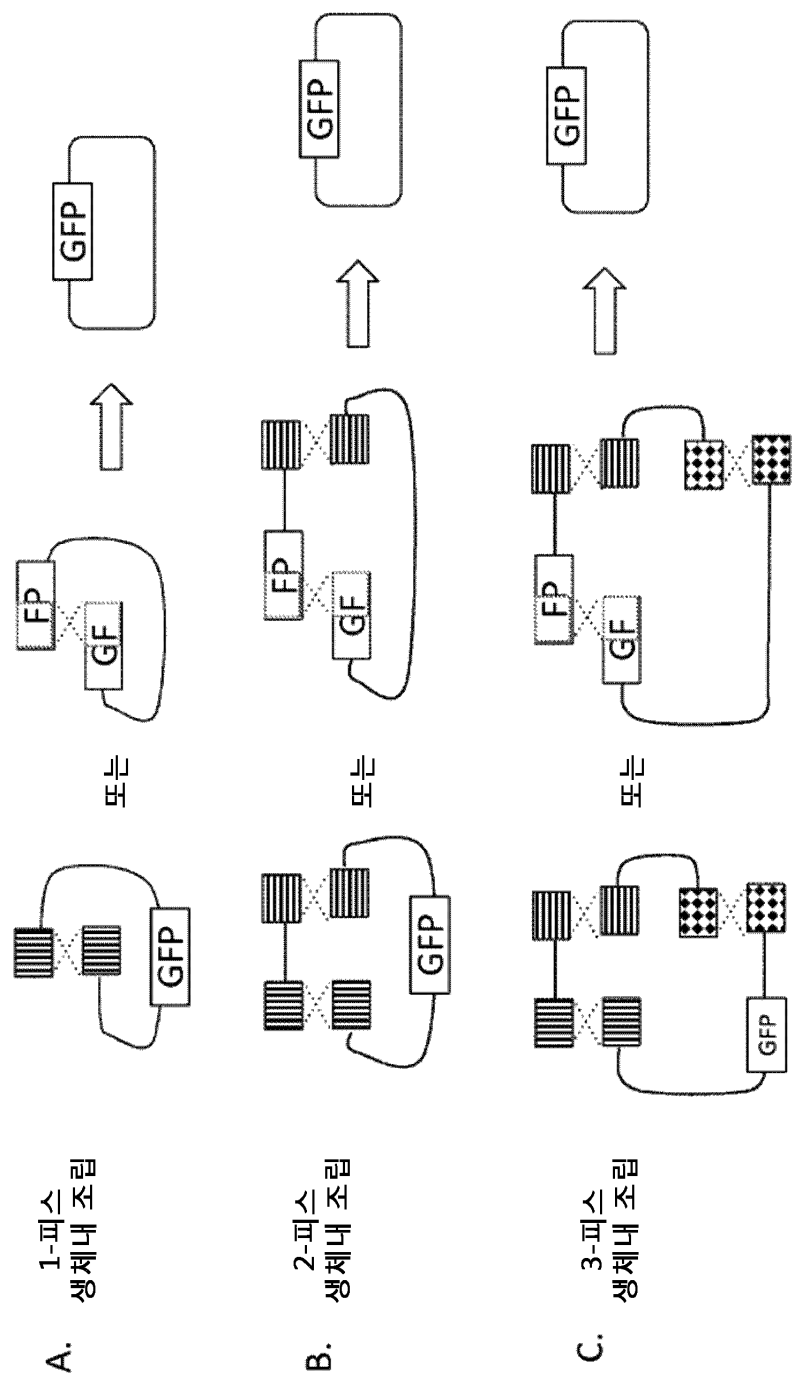
도면2



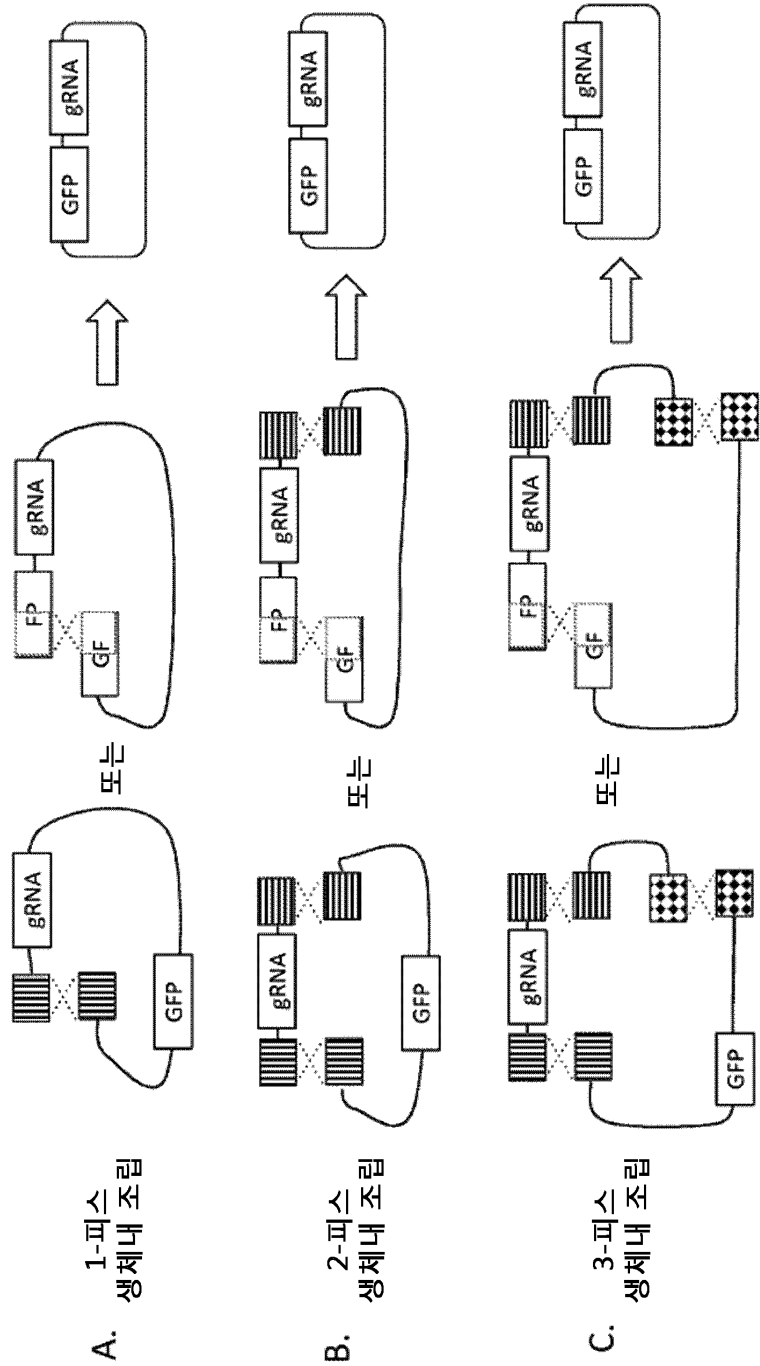
도면3



도면4

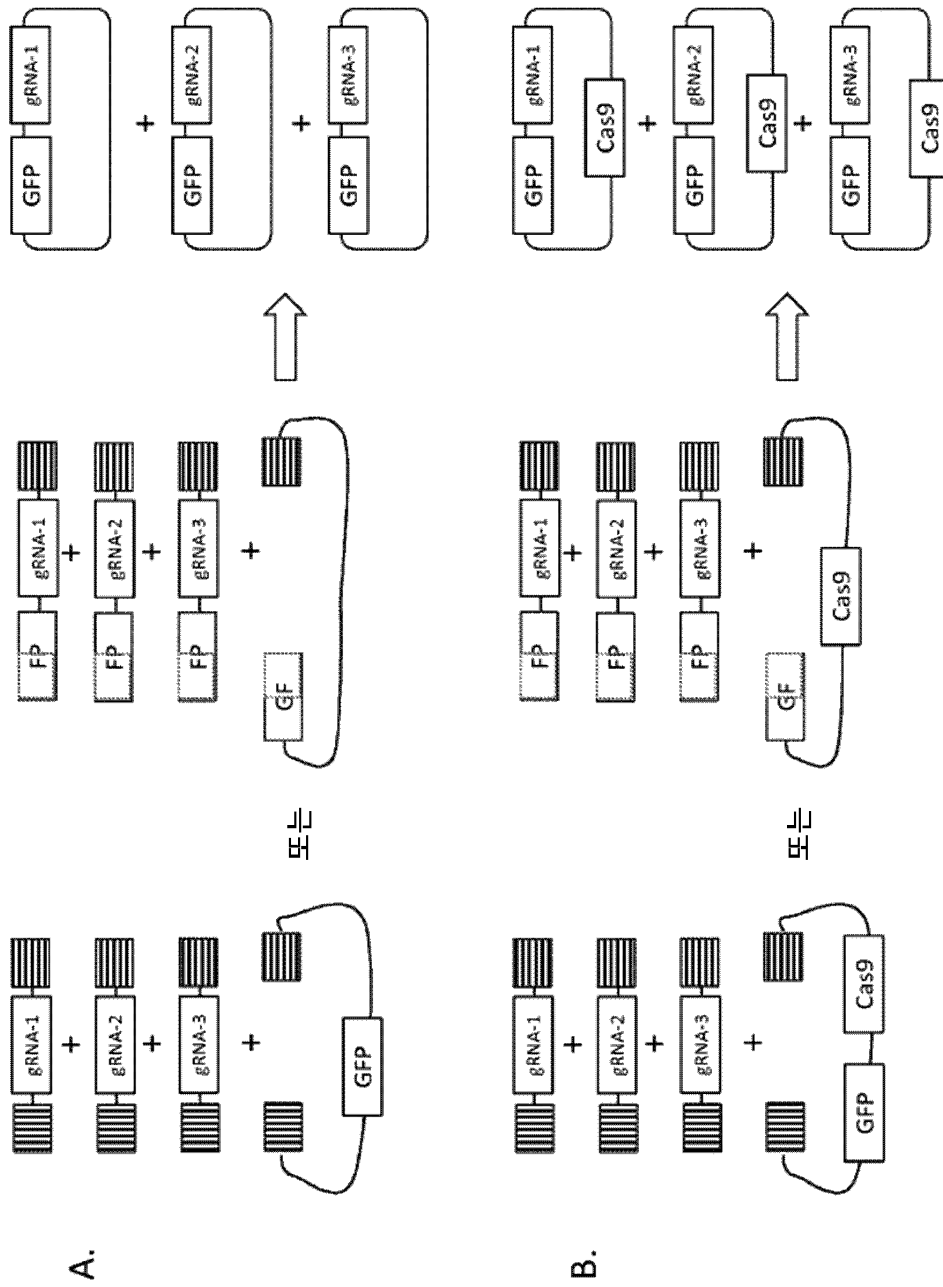


도면5

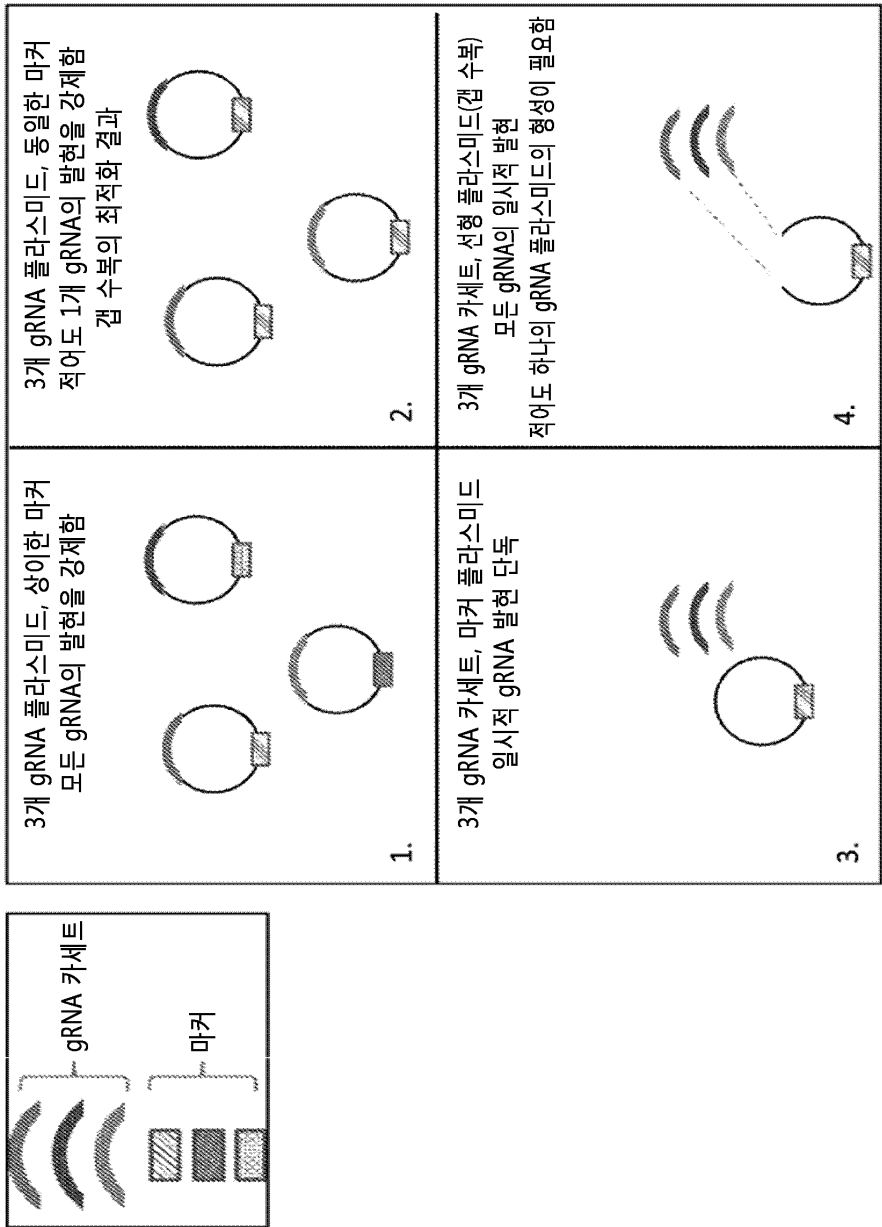


도면6

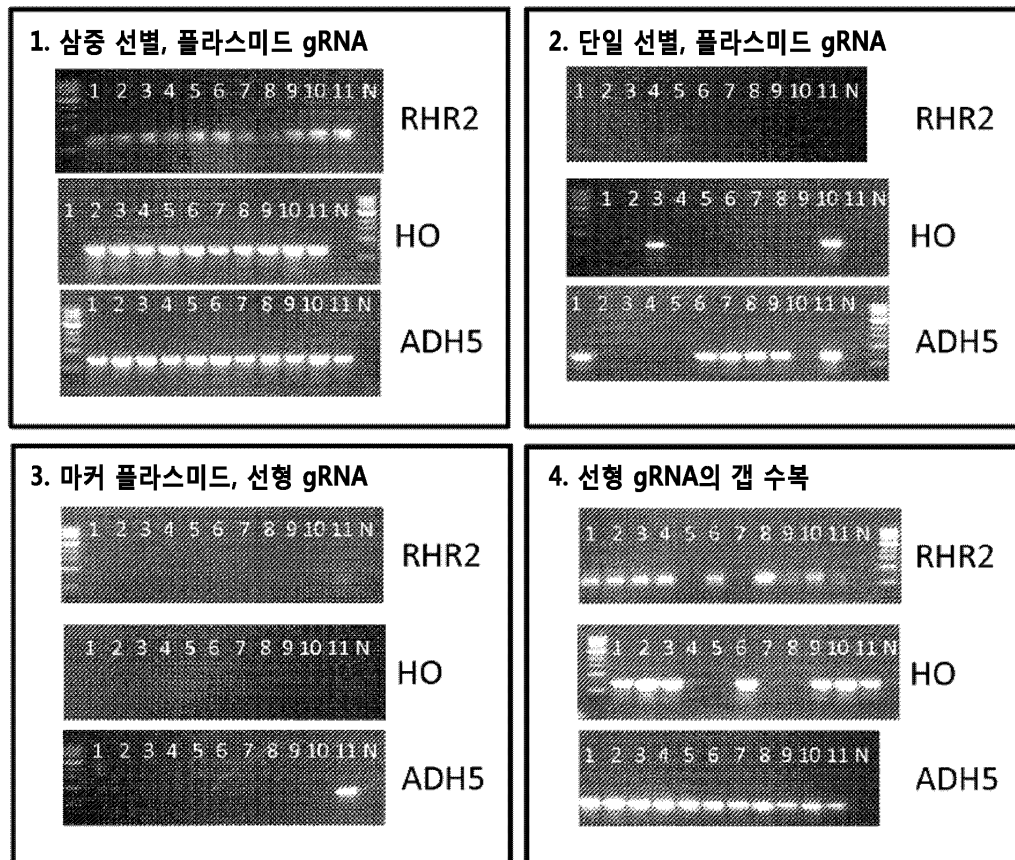
CRISPR-매개된 다중 삽입을 위한 2-피스 생체내 조립



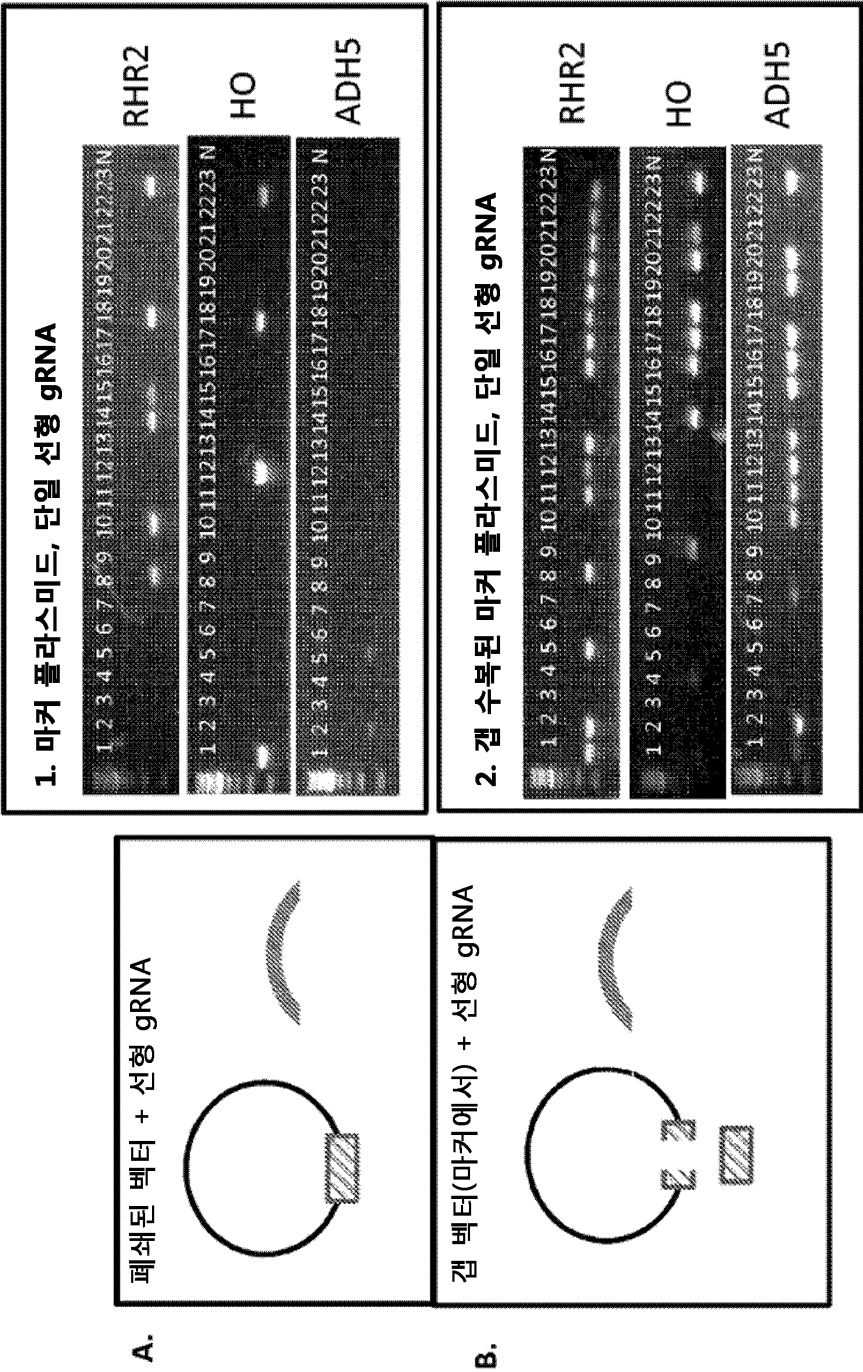
도면7



도면8



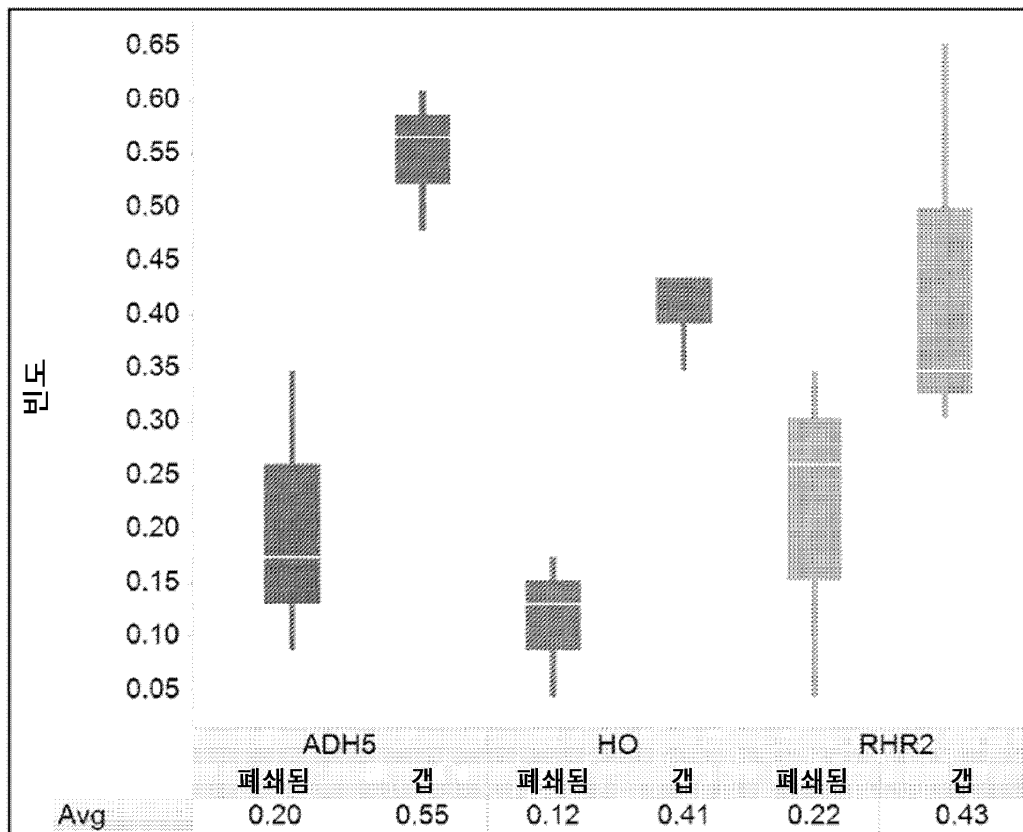
도면9



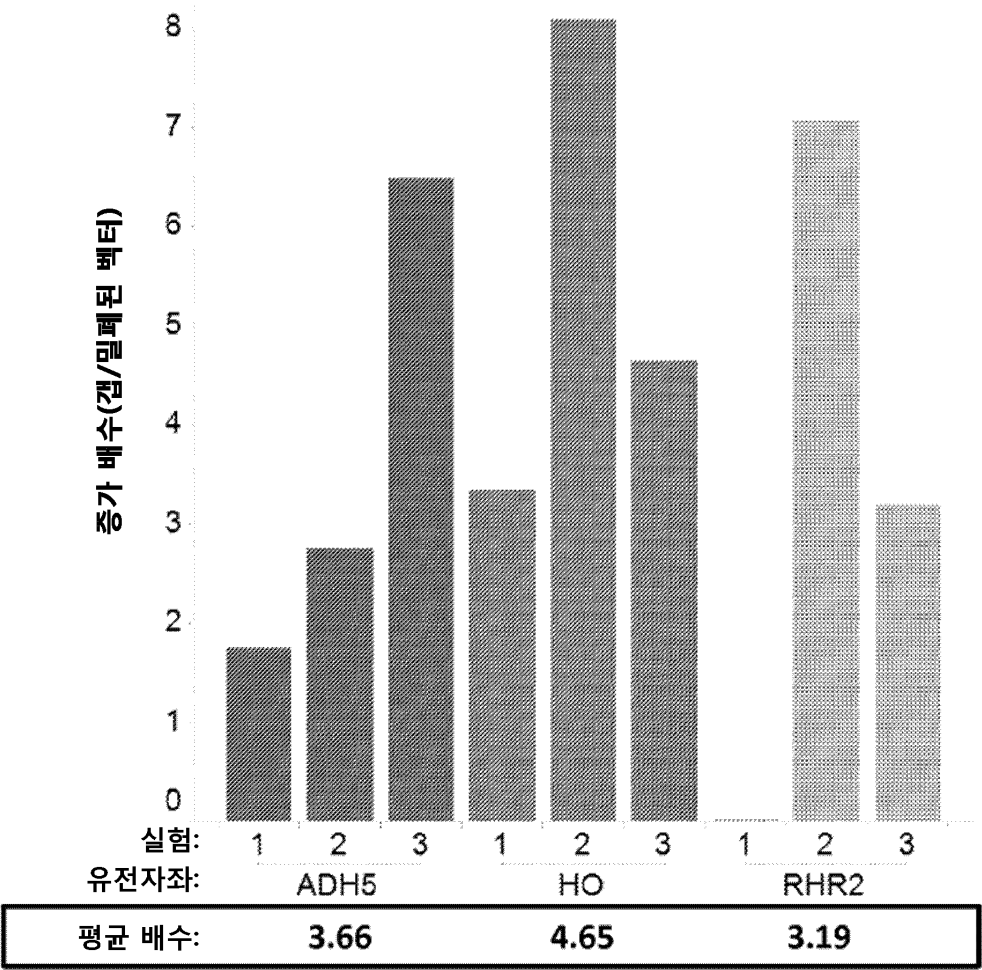
도면10

		1st Expt		2nd Expt		3rd Expt	
		# 포지티브/ # 스크리닝됨	증가 배수	# 포지티브/ # 스크리닝됨	증가 배수	# 포지티브/ # 스크리닝됨	증가 배수
gRNA	마커 벡터						증가 평균 배수
RHR2	폐쇄됨	34.80%		4.30%		26.10%	
RHR2	갭 수복 벡터	34.80%	0	30.40%	7.07	65.20%	2.50
HO	폐쇄됨	13.00%		4.30%		17.40%	
HO	갭 수복 벡터	43.50%	3.35	34.80%	8.09	43.50%	2.50
ADH5	폐쇄됨	34.80%		17.40%		8.70%	
ADH5	갭 수복 벡터	60.90%	1.75	47.80%	2.75	56.50%	6.49
							3.66

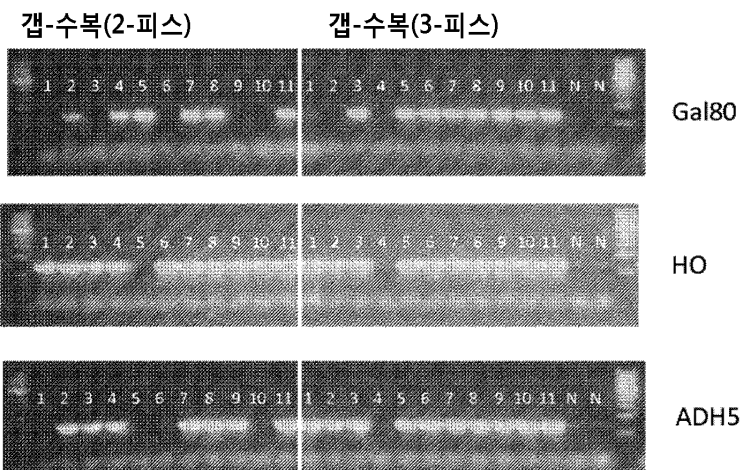
도면11



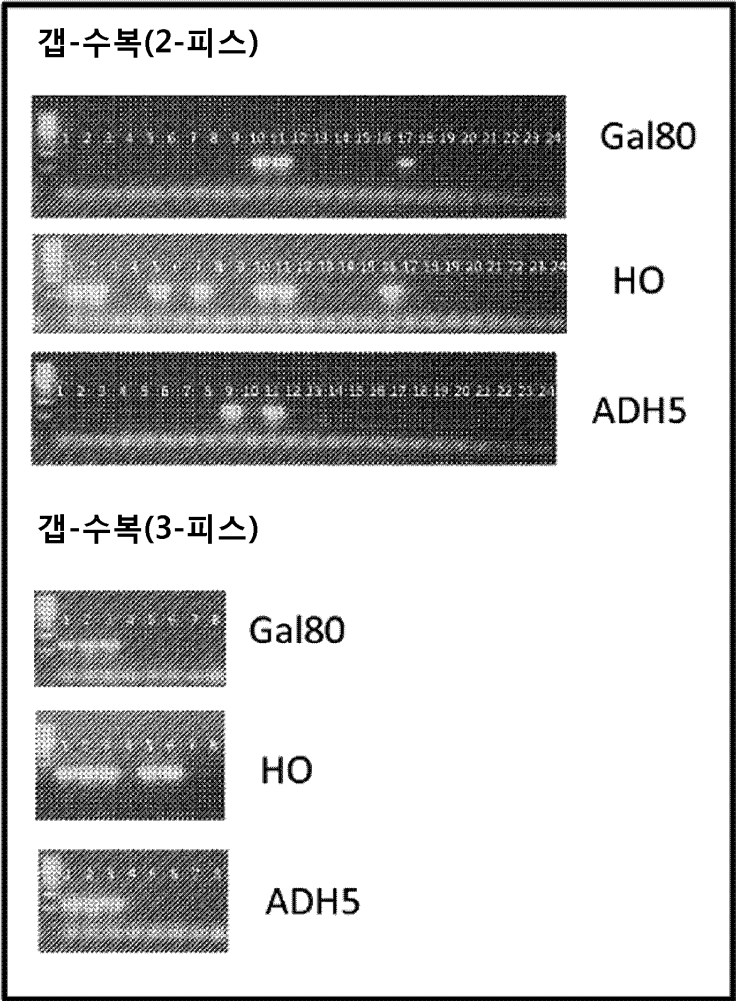
도면12



도면13



도면14

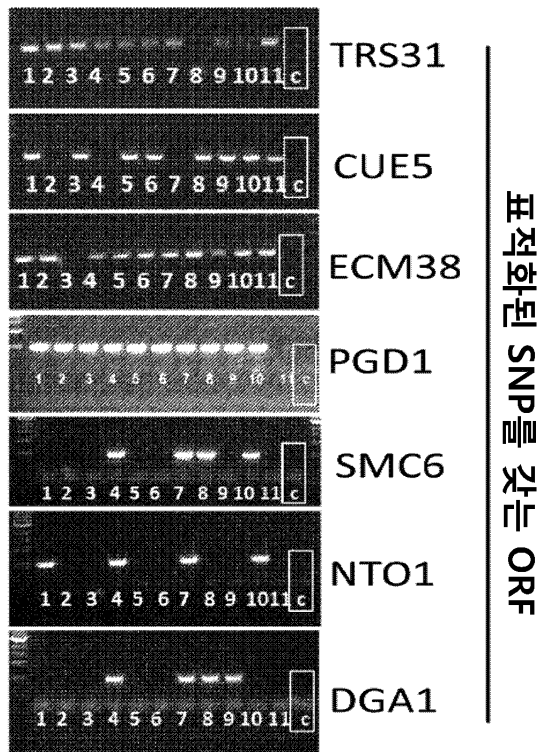


설명	삼중 삽입 비율	N
2-피스 갭 수복	4%	24
2-피스 갭 수복	37.50%	8

도면15

원래 게놈 서열	절단 부위 PAM 표적화된 아미노산 ... GAC GCT CGA TCG ATA GCT AGG AGC TAC CTG CGA GCT AGC TAT CGA TCC TCG ATG ... D A R S I A R S Y
공여체 DNA에 대한 올리고	60-mer 1: ... GAC GCT CGA TCA CGA GCA AGA AGT 60-mer 2: GA GCT AGT GCT CGT TCT TCA ATG ...
dsDNA 공여체 단편을 제조하기 위해 올리고를 어닐링하고 증식함	... GAC GCT CGA TCA CGA GCA AGA AGT TAC CAG CGA GCT AGT GCT CGT TCT TCA ATG ...
변형된 게놈 서열	이중성 블록 I 내지 R 대립인자 변화 ... GAC GCT CGA TC A GC AG AG TAC CTG CGA GCT AGT GCT CGT TCT TCA ATG ... D A R S R A R S Y

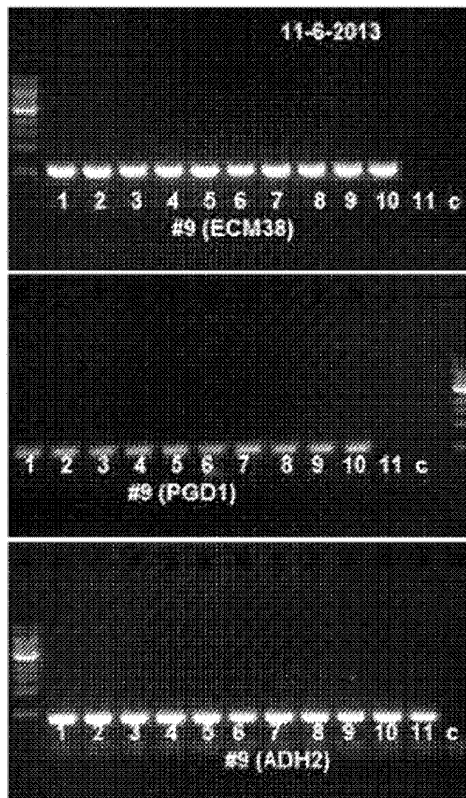
도면16



이종성 블록 삽입 비율

SNP를 갖는 ORF	cPCR 포지티브
TRS31	81.8%
CUE5	72.7%
ECM38	90.9%
PGD1	90.9%
SMC6	36.4%
NTO1	36.4%
DGA1	36.4%

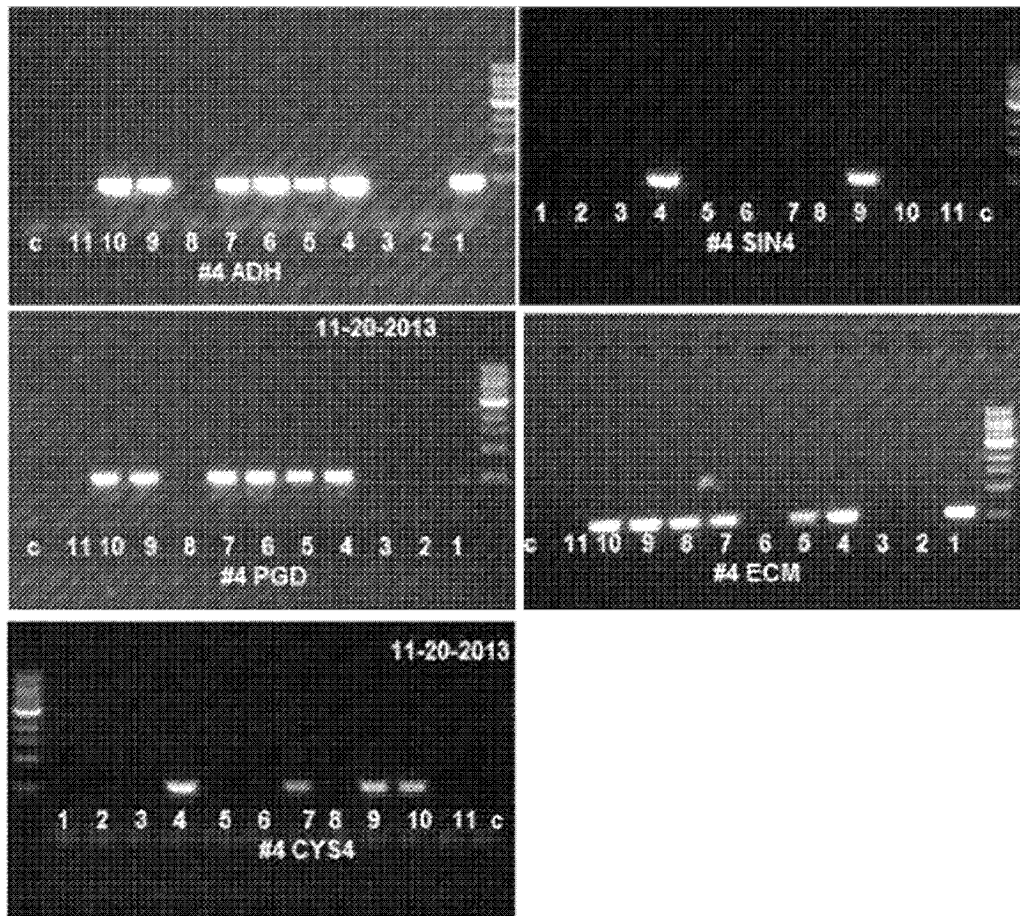
도면17



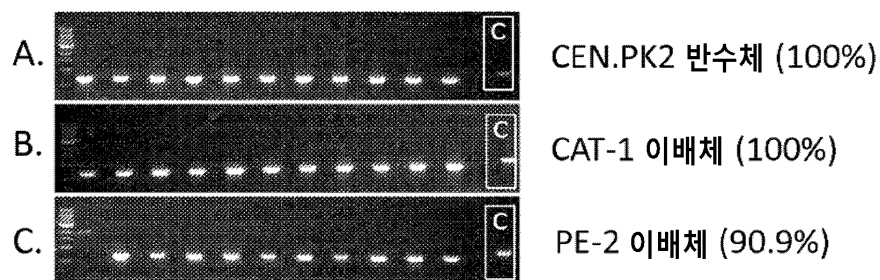
이종성 블록 삽입 비율

SNP를 갖는 ORF	cPCR 포지티브
ECM38	90.9%
PGD1	90.9%
ADH2	100%
삼중	90.9%

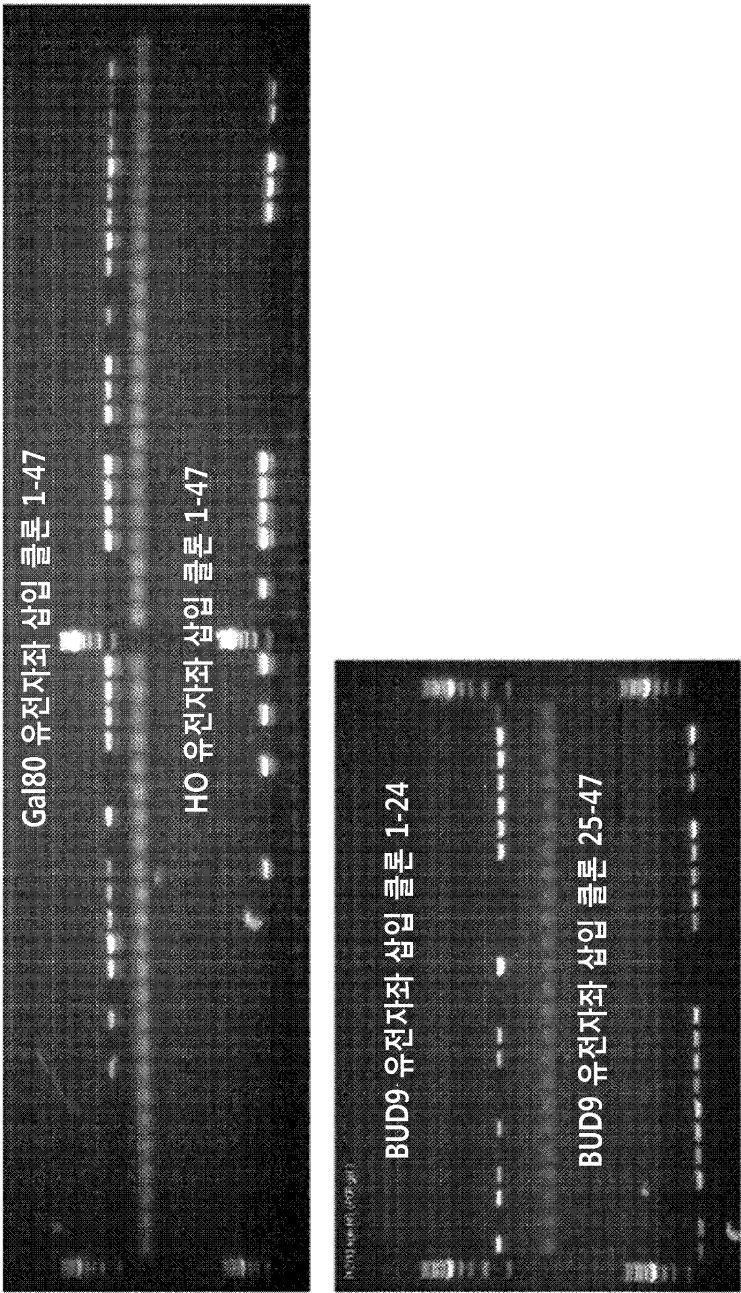
도면18



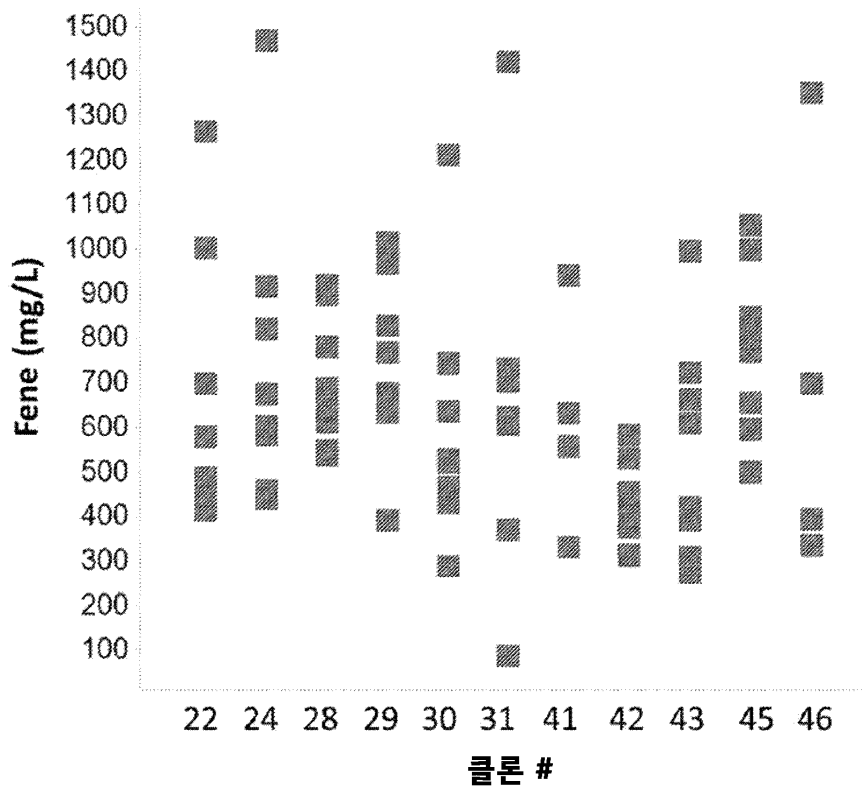
도면19



도면20



도면21

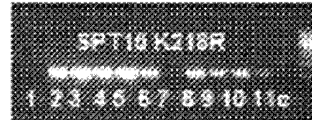


도면22

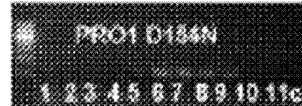
SPT15 F177S (단일 교환, 10/11)



SPT15 K218R (단일 교환, 9/11)



PRO1 D154N (단일 교환, 5/11)



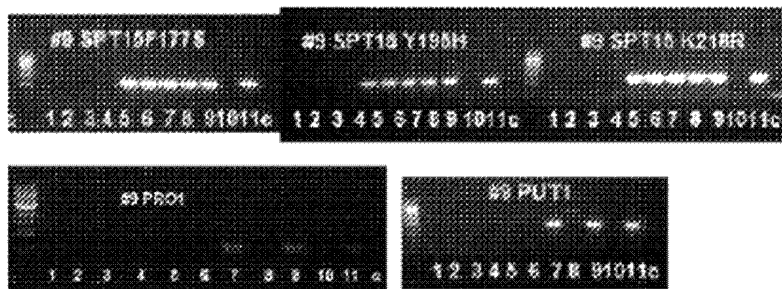
PRO1, PUT1 (이중 교환, 1/11)



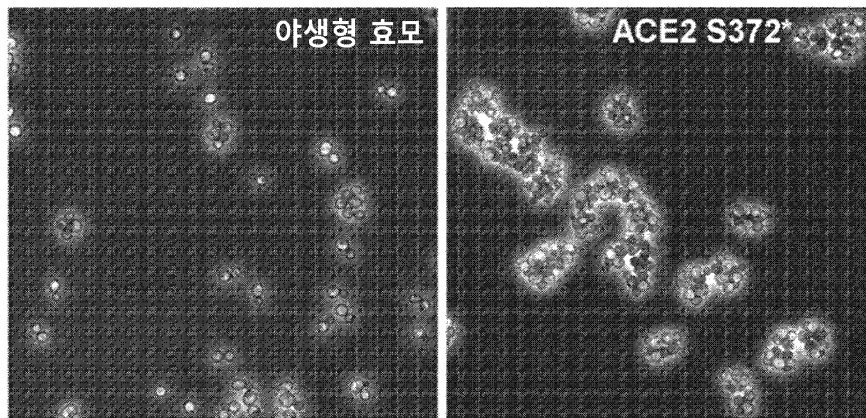
SPT15 삼중 돌연변이
(2 gRNA, 1 공여체, 1/11)



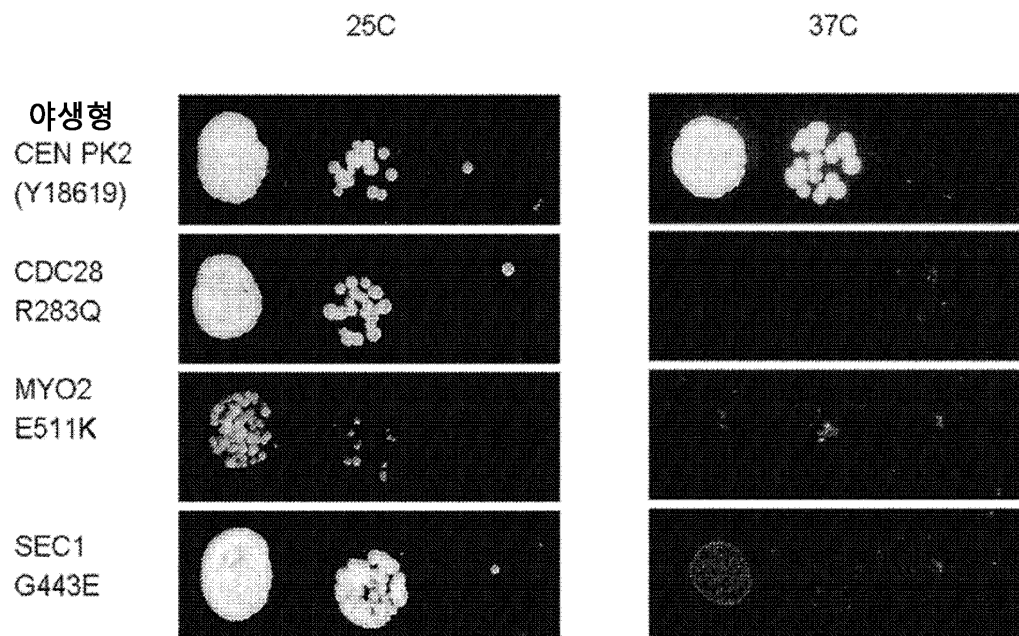
SPT15 삼중, PRO1, PUT1 (4gRNA, 3 공여체, 3/11)



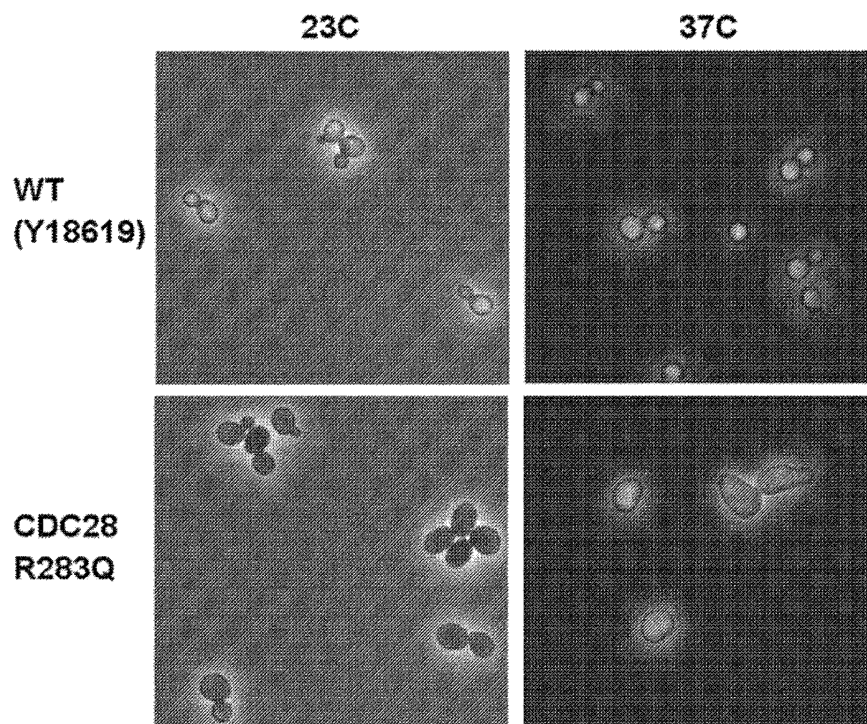
도면23a



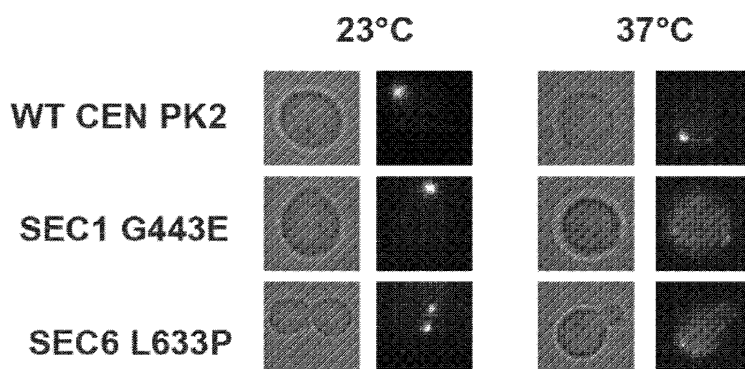
도면23b



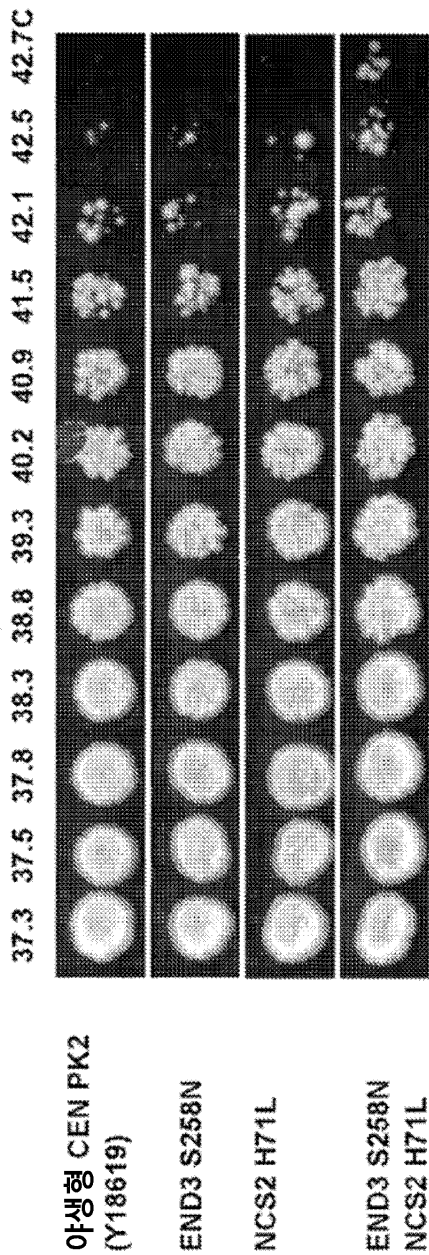
도면23c



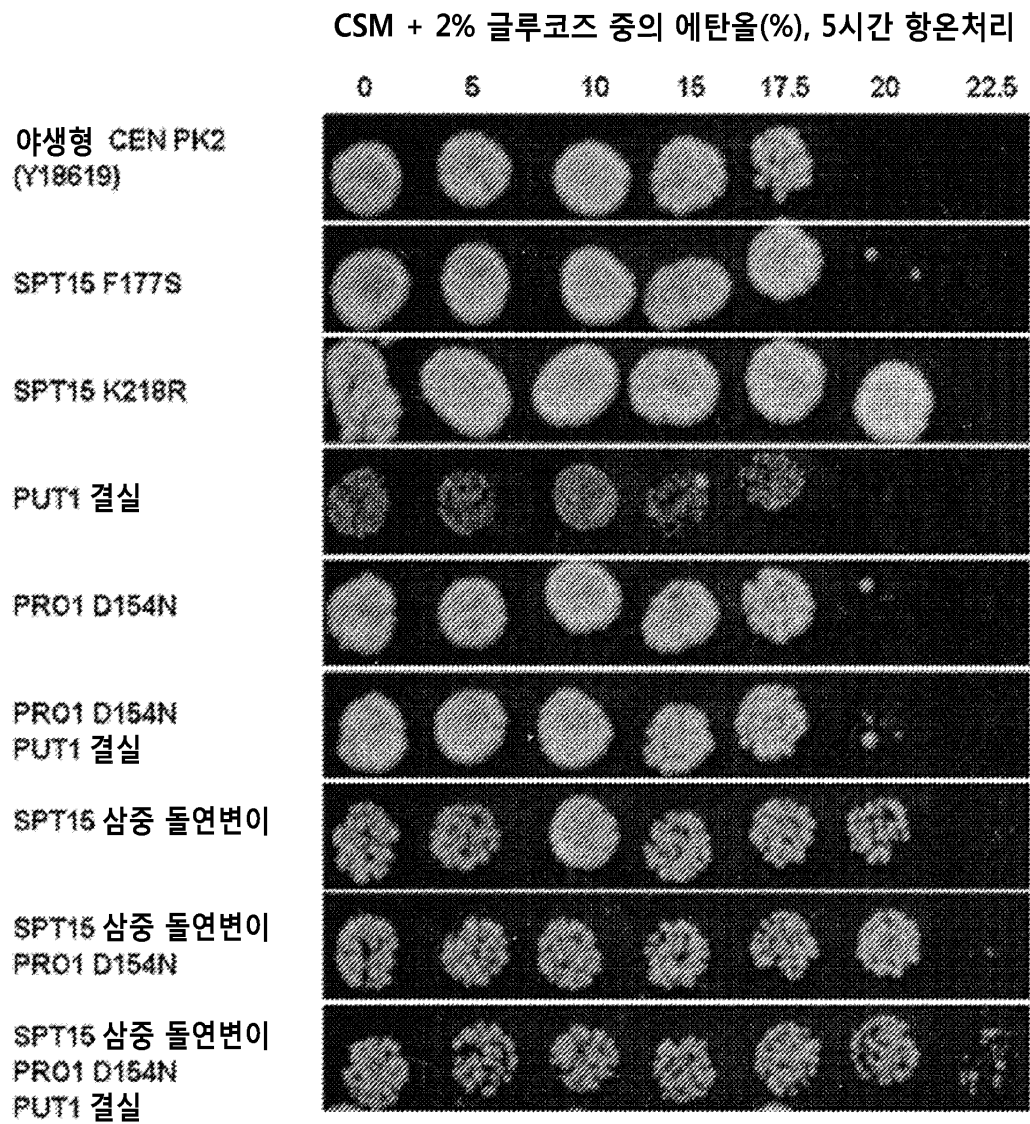
도면23d



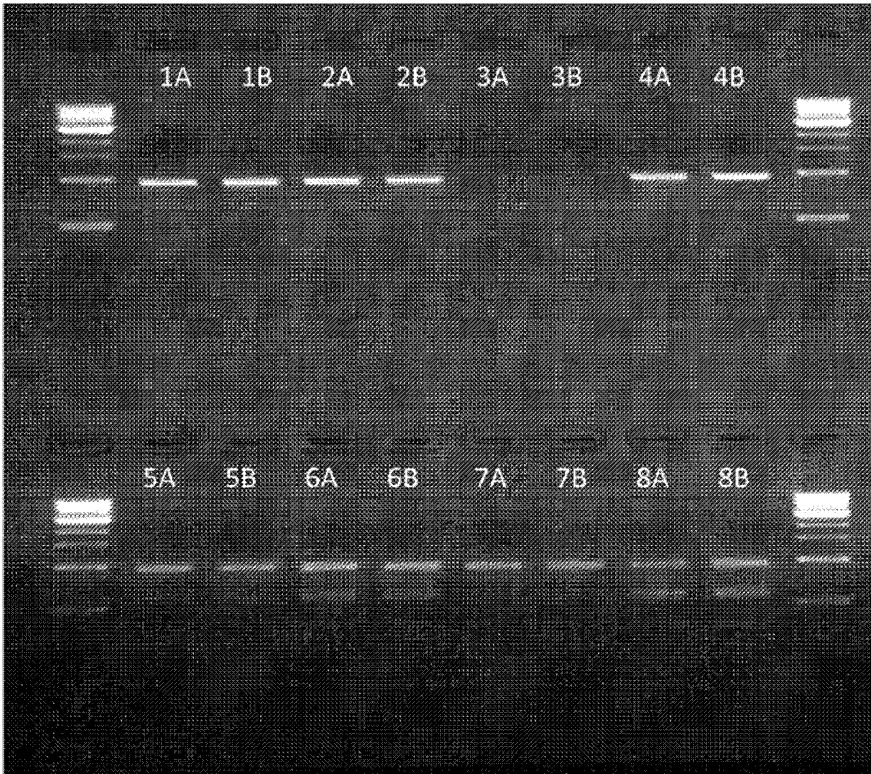
도면23e



도면23f



도면25



서열 목록

SEQUENCE LISTING

<110> AMYRIS, INC.

<120> METHODS FOR GENOMIC INTEGRATION

<130> IPA160571-US

<150> US 61/918,625

<151> 2013-12-19

<150> US 61/937,444

<151> 2014-02-07

<160> 120

<170> PatentIn version 3.5

<210> 1

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic: Homing endonuclease conserved sequence motif

<400> 1

Leu Ala Gly Leu Ile Asp Ala Asp Gly

1 5

<210> 2

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic: Homing endonuclease conserved sequence motif

<400> 2

Gly Ile Tyr Tyr Ile Gly

1 5

<210> 3

<211> 7804

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: Cas9 Genomic Expression Construct

<400> 3

gacggcacgg ccacgcgttt aaaccgccga ccagagattt tgcattccag tattcatcaa 60

tgatgaattc gtagacgcag atactgtaaa tgccgtgttc atcaagaaat gggcgcat 120

ctacaagaag ttttgatatt ttttgtaact gtaatttcac tcatgcacaa gaaaaaaaaa 180

actggattaa aaggagagccc aaggaaaact cctcagcata tatttagaag tctcctcagc 240

atatagtgtt ttgttttctt tacacattca ctgtttaata aaacttttat aatatttcat 300

tatcggaact ctagattcta tacttgtttc ccaattgttg ctggtagtaa acgtatacgt 360

cataaaaggg aaaagccaca tgcggaagaa ttttatggaa aaaaaaaaaa cctcgaagt 420

actacttcta gggggcctat caagtaaat actcctggta cactgaagta tataagggat 480

atagaagcaa atagtgtca gtgcaatcct tcaagacgct cgtccaacgc cggcggacct 540

cgaatcctta catcacacce aatccccac aagtgatccc ccacacacca tagcttcaaa 600

atgtttctac tcctttttta ctcttcaga ttttctcgga ctccgcgcac cgccgtacca 660

cttcaaaaca cccaagcaca gcatactaaa tttccctctt ttcttctctt aggggtgtgt 720

taattaccgc tactaaaggt ttggaaaaga aaaaagagac cgcctcggtt ctttttcttc 780

gtcgaaaaag gcaataaaaa tttttatcac gtttcttttt cttgaaaatt ttttttttg 840

atTTTTTctt ctttcgatga cctccatttg atatttaagt taataaacgg tcttcaattt 900

ctcaagtttc agtttcattt ttcttgttct attacaactt tttttacttc ttgtcatta 960

gaaagaaagc atagcaaacc tcccgcgacc tccaaatcg aactaccttc acaatggata 1020

agaaatactc tatcggtttg gatattggta ctaactccgt tggttgggcc gttatcactg	1080
atgaatacaa ggttccatct aagaagttca aagttttggg taacactgat agacactcta	1140
tcaagaagaa cttgattggg gctttgttat ttgactctgg tgaaaccgct gaggctaccc	1200
gtttaaaaag aactgctaga cgtagataca cccgtcgtaa aaacagaatc tgttatttgc	1260
aagagatctt ctccaacgaa atggctaagg ttgacgactc ttttttccat agattagaag	1320
aatctttctt agttgaagaa gataagaagc acgaacgtca tccaatcttc ggtaacattg	1380
tcgacgaagt tgcttaccat gaaaagtacc caactatcta tcacttgaga aagaaattgg	1440
ttgattctac tgacaaagcc gacttgagat tgatctactt ggctttagct catatgatca	1500
aattccgtgg tcatttttta attgaaggig atttgaaccc agacaactct gacgttgata	1560
aattgttcat ccaattgggt caaacctata accaattgtt tgaagaaaac ccaattaacg	1620
cttctgggtg tgatgctaag gctatcttgt ctgctagatt gtctaaatct agaagattgg	1680
aaaacttaat tgctcaattg ccaggtgaaa aaaaaaacgg tttgttcggt aatttgattg	1740
ctttatcctt gggtttgacc ccaaatttca agtccaactt tgatttggct gaagatgcca	1800
agttgcaatt gtctaaggat acttacgatg atgatttaga taacttattg gctcaaattg	1860
gtgatcaata cgctgatttg tttttagctg ccaagaattt gtccgacgcc attttgttgt	1920
ctgacatctt gagagtcaac actgaaatta ccaaggcccc tttgtctgct tctatgatta	1980
agagatatga cgaacaccac caagacttga ccttgttgaa ggctttgggt agacaacaat	2040
tacctgaaaa gtataaggaa atttttttcg accaatctaa gaacggttac gctggttaca	2100
ttgacggtag tgccctctca gaagaattct acaaatctat caaaccaatc ttggaaaaga	2160
tggacggtag tgaagaattg ttagttaaat tgaacagaga agacttgttg agaaaacaaa	2220
gaacctttga caacggttcc attcctcacc aaatccactt gggtagtga cacgctatct	2280
tgagaagaca agaagatttc taccattct taaaggacaa ccgtgaaaag attgaaaaga	2340
ttttgacctt cagaattcca tactacgtcg gtcctttggc tcgtggtaac tccagattcg	2400
cctggatgac tagaaagtc gaagaaacta ttactccatg gaacttcgaa gaagtcgttg	2460
acaagggtgc ttctgctcaa tcctttatcg aaagaatgac caacttcgac aaaaacttgc	2520
caaacgaaaa agtcttgcca aagcactctt tgttgtatga atactttact gtttataatg	2580
aattgactaa agttaagtac gttactgaag gtatgagaaa accagctttt ttatctgggtg	2640
aacaaaaaaa agctatcgtc gatttgttgt tcaaaactaa ccgtaaagtt accgtcaagc	2700
aattgaagga agattacttc aagaagattg aatgttttga ctccgtcgaa atctccggtg	2760

ttgaagacag attcaatgct tctttgggia cttaccacga cttgttgaaa attatcaagg	2820
acaaggattt cttagataac gaagaaaacg aagacatttt ggaagatatt gtcttgactt	2880
tgactttgtt cgaagataga gaaatgattg aagaaaagatt gaagacttat gctcatttgt	2940
tcgacgataa ggiccatgaag caattaaaaga gaagacgtta cactgggttg ggtagattgt	3000
ctagaaaatt gattaacggt atccgtgata aacaatctgg taagaccatc ttggatttct	3060
taaagtctga tggttttgcc aacagaaaact tcatgcaatt gatccacgac gactctttga	3120
ctttcaagga ggacattcaa aaggctcaag ttcttggtca aggtgactct ttgcatgaac	3180
acattgccaa cttggctggt tctccagcta ttaagaaggg tatcttgcaa actgttaagg	3240
ttgttgatga attagttaag gtcatgggia gacacaagcc agaaaacatc gtcacgaaa	3300
tggctagaga aaaccaaact actcaaaagg gtcaaaagaa ttctagagaa agaatgaaga	3360
gaattgagga aggtattaag gaattagggt cccaaatttt gaaggaacat ccagtcgaaa	3420
acactcaatt gcaaaacgaa aaattgtact tgtactactt acaaaacggt agagatatgt	3480
atgtcgacca agagtggac atcaacagat tgtccgacta cgatgttgat cacatcgttc	3540
cacaatcctt cttaaaggac gactctatcg acaacaaggt cttaccaga tccgacaaaa	3600
acagaggtaa gtctgacaac gttccatccg aagaagttgt taaaaagatg aagaactact	3660
ggagacaatt gttgaacgcc aaattgatca ctcaaagaaa gttcgataat ttgaccaagg	3720
ctgaaagagg tggtttgtct gaattggata aggctggttt tattaaaaga caattggttg	3780
agactagaca aatcaccaag catgtcgctc aaattttaga ttccagaatg aacactaaat	3840
acgacgaaaa cgataagtta attagagaag ttaaggttat taccttgaag tctaagttgg	3900
tttctgattt cagaaaggac ttccaatttt acaaggtcag agaaattaac aactaccatc	3960
acgctcatga tgcttacttg aacgccgttg ttggtaccgc ttgtattaaa aagtacccaa	4020
agttggaatc cgaatttgct tacggtgact acaaggctca cgatgtcaga aaaatgatcg	4080
ctaagtcga acaagagatt ggtaaggcta ctccaagta cttcttttac tctaactca	4140
tgaacttttt caagactgaa atcacttttag ctaacggtga aattcgtaag agaccattga	4200
ttgaaaccaa cggtgagact ggtgaaatcg ttgggataa gggtcgtgat ttcgctactg	4260
ttagaaaaggt cttatctatg ccacaagtta acatcgtaaa gaaaaccgaa gttcaactg	4320
gtggtttttc taaggaatct atcttgccaa aaagaaactc tgataaattg attgctagaa	4380
agaaggattg ggacccaaag aagtacggtg gtttcgattc cccaaccgtc gcttactccg	4440
tcttggttgt cgctaaagtt gaaaagggtta agtccaagaa attgaagtct gttaaaggaat	4500
tgttgggtat cactatcatg gaaagatctt ccttcgaaaa gaacccaatc gattttttag	4560
aggccaaggg ttataaggaa gttaaaaagg acttaattat taagttgcca aagtactctt	4620

tggtcgaatt agaaaacggt agaaaaagaa tgttggcctc tgctgggtgag ttgcaaaaag	4680
gtaacgaatt ggccttgcca tctaagtatg ttaacttttt gtacttggcc tctcattacg	4740
agaagttgaa gggttcccca gaagataacg aacaaaagca attgttcgtc gaacaacaca	4800
aacattactt ggatgaaatt atcgaacaaa tctccgagtt ttccaaacgt gttatcttgg	4860
ctgacgccaa tttggataag gttttgtctg cttataataa gcatagagat aagccaatta	4920
gagaacaagc cgagaacatc attcacttgt tcactttgac taatttaggt gctccagctg	4980
ccttcaaata tttcgacacc accattgata gaaagagata cacctccact aaggaagtct	5040
tggatgccac cttgattcac caatctatca ctggtttgta cgaaactaga atcgatttgt	5100
ctcaattagg tggtgattcc cgtgccgacc caaagaagaa gagaaaggtc taaacaggcc	5160
ccttttctt tgcgatatc atgtaattag ttatgtcacg cttacattca cgccctcccc	5220
ccacatccgc tctaaccgaa aaggaaggag ttagacaacc tgaagtctag gtccctattt	5280
atTTTTTat agttatgtta gtattaagaa cgttatttat atttcaaatt tttctttttt	5340
ttctgtacaa acgcgtgtac gcatgtaaca ttatactgaa aaccttgctt gagaaggttt	5400
tggcatcccc gcgtgcttgg cgggccgttt aatcagcgcc cagagactag cactgaatga	5460
tcaacgggta gttcacagca tgcacgagcg caacgctcac aatgacagtc tggacatcaa	5520
tagtcacact acagaagggtg atctctcaac ttcagcagac catagcgtgt aataaatgca	5580
taattatTTT tctctaaaaa aaactcagct gaaattttat ataagtacta cattttatat	5640
acataattaca tactgaacaa taagcgcggt tgacatttta atttcgaag accgcgaatc	5700
cttacatcac acccagtcce ccaatagttc cccacacac catgcttcaa aaacgcactg	5760
tactcctttt tactcttccg gattttctcg gactctccgc atcgccgcac gagccaagcc	5820
acaccacac acctcatacc atgtttcccc tctttgactc tttcgtgcgg ctccattacc	5880
cgcataaac tgtataaaaag taacaaaaga ctatttcgtt tctttttctt tgtcggaaaa	5940
ggcaaaaaaa aaaattttta tcacatttct tttcttgaa aatttttttt gggatttttt	6000
ctctttcgat gacctccat tgatatTTaa gTTaataaaa ggtctcccg tttccaagtt	6060
ttaatttggt cctcttggtt agtcattctt cttctcagca ttggtcaatt agaaagagag	6120
catagcaaac tgatctaagt tttaattacc atatgaaaa gcctgaactc accgcgacgt	6180
ctgtcgagaa gtttctgac gaaaagtctg acagcgtctc cgacctgatg cagctctcgg	6240
agggcgaaga atctcgtgct ttcagcttcg atgtaggagg gcgtggatat gtctgcggg	6300
taaatagctg cgccgatggt ttctacaaag atcgttatgt ttatcggcac tttgcatcgg	6360
ccgcgtcccc gattccggaa gtgcttgaca ttggggaatt cagcgagagc ctgacctatt	6420
gcatctcccc ccgtgcacag ggtgtcacgt tgcaagacct gcctgaaacc gaactgcccc	6480

ctgtttctgca gccggtcgcg gaggccatgg atcgatcgc tgcggccgat cttagccaga 6540

cgagcgggtt cgcccattc ggaccgcaag gaatcggta atacactaca tggcgtgatt 6600

tcatatgcgc gattgctgat ccccatgtgt atcactggca aactgtgatg gacgacaccg 6660

tcagtgcgtc cgtcgcgcag gctctcgatg agctgatgct ttgggccgag gactgccccg 6720

aagtccggca cctcgtgcac gcggatttcg gctccaacaa tgtcctgacg gacaatggcc 6780

gcataacagc ggtcattgac tggagcgagg cgatgttcgg ggattcccaa tacgaggtcg 6840

ccaacatctt ctcttgagg ccgtggttgg ctgtatgga gcagcagacg cgctacttcg 6900

agcggaggca tccggagctt gcaggatcgc cgcggctccg ggcttatatg ctccgcattg 6960

gtcttgacca acitctatcag agcttgggtg acggcaattt cgatgatgca gcttgggcgc 7020

agggtcgatg cgacgcaatc gtccgatccg gagccgggac tgcggggcgt acacaaatcg 7080

cccgcagaag cgcggccgctc tggaccgatg gctgtgtaga agtactcgcc gatagtggaa 7140

accgacgccc cagcactcgt ccgagggcaa aggaataggg aaattgataa gacttttcta 7200

gttgcatatc ttttatattt aaatcttatt tattagttaa tttttgtaa tttatcctta 7260

tatagtctgg ttattctaaa atatcatttc agtatctaaa atagtctttt tttttttga 7320

gttagatttt tatgggggag agttgaagtg ttgaattttc ccactttgct tcgggattgt 7380

gggtcattct gtcgataact gatatcacat catcaataga acctcttaga tgcacgagcg 7440

caacgtcac aattaatcag cgcccagaga ctagcactga atgatcaacg ggtagttcac 7500

acaggtcgcg cggcgttggg cgagcgtcca gccagtaaaa tccatactca acgacgatat 7560

gaacaaattt cctcattcc gatgctgtat atgtgtataa atttttacat gctcttctgt 7620

ttagacacag aacagcttta aataaaatgt tggatatact tttctgcct gtggtgtcat 7680

ccacgctttt aattcatctc ttgtatggtt gacaatttgg ctatttttta acagaacca 7740

acggtaatg aaattaaaag ggaaacgagt gggggcggcg gtttaaacgc gtggccgtgc 7800

cgtc 7804

<210> 4

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: RHR2 genomic target Sequence

<400> 4

acctctggta cccgtgaca 19

<210> 5

<211> 19
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: H0 genomic target Sequence
 <400> 5
 ccggcttgat cgactcaga 19
 <210> 6
 <211> 19
 <
 212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: ADH5 genomic target sequence
 <400> 6
 gggtcattgg tatcgatgg 19
 <210> 7
 <211> 1026
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: gRNA construct targeting RHR2 locus
 <400> 7
 ccttataaat caaaagaata gaccgagata gggttgagtg ttgttccagt ttggaacaag 60
 agtccactat taaagaacgt ggactccaac gtcaaagggc gaaaaaccgt ctatcagggc 120
 gatggcccac tacgtgaacc atcacccata tcaagttttt tggggtcgag gtgccgtaaa 180
 gcactaaatc ggaaccctaa agggagcccc cgatttagag cttgacgggg aaagcctctt 240
 tgaaaagata atgtatgatt atgctttcac tcataattat acagaaactt gatgttttct 300
 ttcgagtata tacaaggatga ttacatgtac gtttgaagta caactctaga tttttagtagt 360
 ccctcttggg ctagcggtaa aggtgcgcat tttttcacac cctacaatgt tctgttcaaa 420
 agatttttgt caaacgctgt agaagtgaat gttggtgcgc atgtttcggc gttcgaaact 480
 tctccgcagt gaaagataaa tgatcgccgc caatgatgtc aagcagtttt agagctagaa 540
 atagcaagtt aaaataaggc tagtccgtta tcaacttgaa aaagtggcac cgagtcggtg 600
 gtgctttttt tgttttttat gtctcagctt ttgttccctt tagtgagggt taattgcgcg 660
 cttggcgtaa tcatggatcat agctgtttcc tgtgtgaaat tgttatccgc tcacaattcc 720
 acacaacata ggagccggaa gcataaagtg taaagcctgg ggtgcctaata gtagtgaggt 780

actcacatta attgcgttgc gctcactgcc cgttttccag tcgggaaacc tgtcgtgcca 840
gctgcattaa tgaatcgccc aacgcgcggg gagaggcggg ttgcgtattg ggcgctcttc 900
cgcttcctcg ctactgact cgctgcgctc ggtcgttcgg ctgcggcgag cggatcagc 960
tcactcaaag gcggaatac gggttatccac agaatacagg gataacgcag gaaagaacat 1020

gtgagc 1026

<210> 8

<211> 1080

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: gRNA construct targeting HO locus

<400> 8

gacggcacgg ccacgcgttt aaaccgcctt ataatcaaa agaatagacc gagatagggt 60
tgagtgttgt tccagtttgg aacaagagtc cactattaaa gaacgtggac tccaacgtca 120
aagggcgaaa aaccgtctat cagggcgatg gccactacg tgaacctca ccctaataa 180
gttttttggg gtcgagggtc cgtaaagcac taaatcgga ccctaaagg agccccgat 240

ttagagcttg acggggaaag cctctttgaa aagataatgt atgattatgc tttcactcat 300
atttatacag aaacttgatg ttttctttcg agtatataca aggtgattac atgtacgttt 360
gaagtacaac tctagatgtt gtagtgcctt cttgggctag cggtaaagg gcgcattttt 420
tcacacccta caatgttctg ttcaaaagat ttgggtcaaa cgctgtagaa gtgaaagttg 480
gtgcgcatgt ttcgcgctc gaaacttctc cgcagtgaat gataaatgat cgccggcttg 540
atcgactcag agtttttaga ctagaatag caagttaaaa taaggctagt ccgttatcaa 600
cttgaaaaag tggcaccgag tcgggtggtc tttttttgtt ttttatgtct cagcttttgt 660

tccctttagt gagggttaat tgcgcgcttg gcgtaatcat ggtcatagct gtttctgtg 720
tgaaattgtt atccgctcac aattccacac aacataggag ccggaagcat aaagtgtaaa 780
gcctgggggtg cctaattgag gaggttaact acattaattg cgttgcgctc actgcccgct 840
ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa tcggccaacg gcgggggaga 900
ggcggtttgc gtattggcg ctcttccgct tcctcgctca ctgactcgct gcgctcggtc 960
gttcggctgc ggcgagcgg atcagctcac tcaaaggcgg taatacgggt atccacagaa 1020
tcaggggata acgcaggaaa gaacatgtga gccggtgttt aaacccagc gcctggcggg 1080

<210> 9

<211> 1080

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: gRNA construct targeting ADH5 locus

<400> 9

gacggcacgg ccacgcgttt aaaccgcctt ataaatcaaa agaatagacc gagatagggt	60
tgagtgttgt tccagtttgg aacaagagtc cactattaaa gaacgtggac tccaacgtca	120
aagggcgaaa aaccgtctat cagggcgaig gccactacg tgaacatca ccctaataca	180
gttttttggg gtcgaggigc cgtaaagcac taaatcggaa ccctaaaggg agccccgat	240
ttagagcttg acggggaaa cctctttgaa aagataatgt atgattatgc tttactcat	300
attatacag aaacttgatg tttcttttcg agtatataca aggtgattac atgtacgttt	360
gaagtacaac tctagatttt gtagtgcctt ctgggctag cggtaaaggt gcgcattttt	420
tcacacccta caatgttctg ttcaaaagat ttgggtcaaa cgctgtagaa gtgaaagtgt	480
gtgcgcatgt ttcggcgttc gaaacttctc cgcagtgaat gataaatgat cggggtcatt	540
ggatcgtatg ggttttagag ctagaatag caagttaaaa taaggctagt ccgttatcaa	600
cttgaanaag tggcaccgag tcggtggtgc tttttttgtt ttttatgtct cagcttttgt	660
tccctttagt gagggttaat tgcgcgcttg gcgtaatcat ggtcatagct gtttcctgtg	720
tgaaattgtt atccgtcac aattccacac aacataggag ccggaagcat aaagtgtaaa	780
gcctgggggtg cctaagtatg gagtaactc acattaattg cgttgcgctc actgcccgt	840
ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcggggaga	900
ggcggtttgc gtattgggcg ctcttccgct tctcgtca ctgactcgtc gcgctcggtc	960
gttcggctgc ggcgagcgtt atcagctcac tcaaaggcgg taatacgtt atccacagaa	1020
tcaggggata acgcaggaaa gaacatgtga gccggtgttt aaacccacgc gcctggcggg	1080

<210> 10

<211> 5994

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: pRS4XX vector backbone

<400> 10

tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat gcagctccc gagacggtca	60
cagcttgtct gtaagcggat gccgggagca gacaagccc tcagggcgag tcagcgggtg	120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc	180
aaatacacat catcgtccta caagttcatc aaagtgttgg acagacaact ataccagcat	240

ggatctcttg tatcggttct tttctccgc tctctcgaa taacaatgaa cactgggtca 300
 atcatagcct acacaggiga acagagtagc gtttatacag ggttatacag gtgattccta 360

 cgcaaaaaat ttttcatttc taaaaaaaaa aagaaaaatt tttctttcca acgctagaag 420
 gaaaagaaaa atctaattaa attgatttgg tgattttctg agagtccct tttcatata 480
 tcgaattttg aatataaaag gagatcgaaa aaatttttct attcaatctg ttttctggtt 540
 ttatttgata gtttttttgt gtattattat tatggattag tactggttta tatgggtttt 600
 tctgtataac ttctttttat tttagtttgi ttaatcttat tttgagttac attatagttc 660
 cctaactgca agagaagtaa cattaaaaat gaccactctt gacgacacgg cttaccggtta 720
 ccgcaccagt gtccccgggg acgccgaggc catcgaggca ctggatgggt ccttcaccac 780

 cgacaccgtc ttccgcgtca ccgccaccgg ggacggcttc accctgcggg aggtgccggt 840
 ggaccgcccc ctgaccaagg tgttccccga cgacgaatcg gacgacgaat cggacgccgg 900
 ggaggacggc gaccggact cccggacgtt cgtcgcgtac ggggacgacg gcgacctggc 960
 gggcttctgt gtcgtctcgt actccggctg gaaccgccgg ctgaccgtcg aggacatcga 1020
 ggtcgccccg gagcaccggg ggcacggggt cgggcgcgcg ttgatggggc tcgcgacgga 1080
 gttcgccccg gagcggggcg ccgggcacct ctggctggag gtcaccaacg tcaacgcacc 1140
 ggcgatccac gcgtaccggc ggatgggggt caccctctgc ggcttgaca ccgccctgta 1200

 cgacggcacc gcctcggacg gcgagcaggc gctctacatg agcatgccct gcccctgagt 1260
 ttaacttgat actactagat ttttctctt catttataaa atttttggtt ataattgaag 1320
 ctttagaagt atgaaaaaat ctttttttt cattctttgc aacaaaaata agaagcttct 1380
 tttattcatt gaaatgatga atataaacct aacaaaagaa aaagactcga atatcaaca 1440
 ttaaaaaaaaa ataaaagagg ttatctgttt tcccathtag ttggagtttg cattttctaa 1500
 tagatagaac tctcaattaa tgtggattta gtttctctgt tcgttttttt ttgttttgtt 1560
 ctactgtat ttacatttct atttagtatt tagttattca tataatctta acttgcggtg 1620

 tgaaataccg cacagatcgc taaggagaaa ataccgcac aggaattgt aagcgttaat 1680
 attttgtaa aattcgcgtt aaatttttgt taaatcagct cattttttaa ccaataggcc 1740
 gaaatcggca aaatccctta taaatcaaaa gaatagaccg agatagggtt gagtgttgtt 1800
 ccagtttggg acaagagtc actattaaag aacgtggact ccaacgtcaa agggcgaaaa 1860
 accgtctatc agggcgatgg ccactacgt gaaccatcac cctaatcaag ttttttgggg 1920
 tcgaggtgcc gtaaagcact aaatcggaac cctaaaggga gccccgatt tagagcttga 1980

cggggaaagc ctctttgaaa agataatgia tgattatgct ttcactcata tttatacaga 2040

 aacttgatgt tttctttcga gtatatacaa ggtgattaca tgtacgtttg aagtacaact 2100
 ctagattttg tagtgccttc ttgggctagc ggtaaagggt cgcatTTTTT cacaccctac 2160
 aatgttctgt tcaaaagatt ttggicaaac gctgtagaag tgaaagtTgg tgcgcatgtt 2220
 tcggcgttcg aaacttctcc gcagtgaag ataaatgac gatacgttct ctatggagga 2280
 gttttagagc tagaaatagc aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt 2340
 ggcaccgagt cggTggTgct tttttgttt tttatgtctc agcttttgtt ccttttagtg 2400
 agggTtaatt gcgcgctTgg cgtaatcatg gtcatactg tttcctgtgt gaaattgtta 2460

 tccgctcaca attccacaca acataggagc cggaagcata aagtgtaaag cctggggTgc 2520
 ctaatgagtg aggtaaacta cattaattgc gttgcgtca ctgcccgctt tccagtcggg 2580
 aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag gcggtttgcg 2640
 tattgggcgc tcttccgctt cctcgtcac tgactcgtg cgctcggtcg ttcggctgcg 2700
 gcgagcggta tcagctcact caaaggcggT aatacggTta tccacagaat caggggataa 2760
 cgcaggaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc 2820
 gttgctggcg tttttcata ggctccgccc cctgacgag catcacaaaa atcgacgtc 2880

 aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgTttc cccctggaag 2940
 ctccctcgtg cgctctctg ttcgaccct gccgcttacc ggatacctgt cgcctttct 3000
 ccttcggga agcgtggcg tttctcatag ctacgctgt aggtatctca gttcggtgta 3060
 ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc gttcagcccg accgctgcgc 3120
 cttatccggt aactatcgtc ttgagtcaa cccggtTaa cagacttat cgccactggc 3180
 agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagtTctt 3240
 gaagtggTgg cctaactacg gctacactag aaggacagta tttggtatct gcgctctgct 3300

 gaagccagtt accttcggaa aaagagtTgg tagctcttga tccggcaaac aaaccaccgc 3360
 tggtagcggT ggtTTTTTg tttgcaagca gcagattacg cgcagaaaaa aaggatctca 3420
 agaagatcct ttgatcttt ctacggggtc tgacgctcag tggaacgaaa actcacgtta 3480
 agggatTTTg gtcatgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa 3540
 atgaagTTTT aaatcaatct aaagtatata tgagtaaact tggTctgaca gttaccaatg 3600
 cttaatcagt gaggcaccta tctcagcgat ctgtctatTT cgttcaccca tagttgcctg 3660
 actccccgtc gtgtagataa ctacgatacg ggagggtTta ccatctggcc ccagtgtgc 3720

aatgataccg cgagacccac gctcaccggc tccagattta tcagcaataa accagccagc 3780
cggaagggcc gagcgagaa gtggtcctgc aactttatcc gcctccatcc agtctattaa 3840
ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca acgttggtgc 3900
cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggg atggcttcat tcagctccgg 3960
ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag cggtttagctc 4020
cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac tcatgttat 4080
ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt ctgtgactgg 4140

tgagtactca accaagtcac tctgagaata gtgtatgcgg cgaccgagtt gctcttgccc 4200
ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc tcatcattgg 4260
aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat ccagttcgat 4320
gtaacccact cgtgcaccca actgatcttc agcatctttt actttacca gcgtttctgg 4380
gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataaggcgca cacggaaatg 4440
ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg gttattgtct 4500
catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg ttccgcgcac 4560

atttccccga aaagtgccac ctgaacgaag catctgtgct tcattttgta gaacaaaaat 4620
gcaacgcgag agcgctaatt tttcaacaa agaactcgag ctgcattttt acagaacaga 4680
aatgcaacgc gaaagcgcta ttttaccac gaagaatctg tgcttcattt ttgtaaaaca 4740
aaaatgcaac gcgagagcgc taatttttca aacaaagaat ctgagctgca tttttacaga 4800
acagaaatgc aacgcgagag cgctatttta ccaacaaaga atctatactt cttttttgtt 4860
ctacaaaaat gcatcccgag agcgctattt ttctaacaaa gcatcttaga ttactttttt 4920
tctcctttgt gcgctctata atgcagtctc ttgataactt tttgcactgt aggtccgtta 4980

aggttagaag aaggctactt tgggtgtctat tttctcttcc ataaaaaag cctgactcca 5040
cttcccgctt ttactgatta ctacgaagc tgcgggtgca ttttttcaag ataaaggcat 5100
ccccgattat attctatacc gatgtggatt gcgcatactt tgtgaacaga aagtgatagc 5160
gttgatgatt cttcattggg cagaaaatta tgaacggttt cttctatttt gtctctatat 5220
actacgtata ggaaatgttt acattttcgt attgttttcg attcactcta tgaatagttc 5280
ttactacaat ttttttgtct aaagagtaat actagagata aacataaaaa atgtagaggt 5340
cgagtttaga tgcaagtcca aggagcgaaa ggtggatggg taggttatat agggatatag 5400

cacagagata tatagcaaag agatactttt gagcaatgtt tgtggaagcg gtattcgcaa 5460
tattttagta gctcgttaca gtccggtgcg tttttggttt ttgaaagtg cgtcttcaga 5520
gcgcttttgg ttttcaaaag cgctctgaag ttctatact ttctagagaa taggaacttc 5580

ggaataggaa cticaaagcg ttccgaaaa cgagcgcttc cgaaaatgca acgcgagctg 5640

cgcacataca gtcactgtt cagtcgcac ctatatctgc gtgttgctg tatatatata 5700

tacatgagaa gaacggcata gtgcgtgttt atgcttaaat gcgtacttat atgcgtctat 5760

ttatgtagga tgaaggttag tctagtacct cctgtgatat tatcccatc catgcggggt 5820

atcgtatgct tccttcagca ctaccttta gctgttctat atgctgccac tcctcaattg 5880

gattagtctc atccttcaat gctatcattt cctttgatat tggatcatat taagaaacca 5940

ttattatcat gacattaacc tataaaaata ggcgtatcac gaggcccttt cgtc 5994

<210> 11

<211> 6039

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: pAM1114 - closed plasmid bearing a NatA
(Nourseothricin acetyltransferase from Streptomyces noursei)
selectable marker

<400> 11

tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca 60

cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120

ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180

aaatacacat catcgtccta caagttcatc aaagtgttgg acagacaact ataccagcat 240

ggatctcttg tatcggttct tttctccgc tctctcgaa taacaatgaa cactgggtca 300

atcatagcct acacaggtag acagagtagc gtttatacag ggtttatacg gtgattccta 360

cggcaaaaat ttttcatttc taaaaaaaaa aagaaaaatt tttctttcca acgctagaag 420

gaaaagaaaa atctaattaa attgatttgg tgattttctg agagttccct ttttcatata 480

tcgaattttg aatataaaag gagatcgaag aaatttttct attcaatctg ttttctggtt 540

ttatttgata gttttttgt gtattattat tatggattag tactggttta tatgggtttt 600

tctgtataac ttctttttat tttagtttgi ttaatcttat tttgagttac attatagtac 660

cctaactgca agagaagtaa cattaaaaat gaccactctt gacgacacgg cttaccggta 720

ccgcaccagt gtcccggggg acgccgaggc catcgaggca ctggatgggt ccttaccac 780

cgacaccgtc ttcccggtca ccgccaccgg ggacggcttc accctgcggg aggtgccggt 840

ggaccgccct ctgaccaagg tgttccccga cgacgaatcg gacgacgaat cggacgccgg 900

ggaggacggc gaccgggact cccggacgtt cgtcgcgtac ggggacgacg gcgacctggc 960

gggcttcgtg gtcgtctcgt actccggctg gaaccgccgg ctgaccgtcg aggacatcga 1020
 ggtcgccccg gagcaccggg ggcacggggt cgggcgcgcg ttgatggggc tcgcgacgga 1080
 gttcgccccg gagcggggcg ccgggcacct ctggctggag gtcaccaacg tcaacgcacc 1140
 ggcgatccac gcgtaccggc ggatgggggt caccctctgc ggcctggaca cgcacctgta 1200
 cgacggcacc gcctcggacg gcgagcaggc gctctacatg agcatgccct gcccctgagt 1260
 ttaacttgat actactagat tttttctctt catttataaa atttttggtt ataattgaag 1320

 ctttagaagt atgaaaaaat cttttttttt cattctttgc aaccaaata agaagcttct 1380
 tttattcatt gaaatgatga atataaacct aacaaaagaa aaagactcga atatcaaaca 1440
 ttaaaaaaaaa ataaaagagg ttatctgttt tcccathtag ttggagtttg cattttctaa 1500
 tagatagaac tctcaattaa tgtggattta gtttctctgt tcgttttttt ttgttttggt 1560
 ctactgtat ttacatttct atttagtatt tagttattca tataatctta acttgcggtg 1620
 tgaaataccg cacagatcgc taaggagaaa ataccgcac aggaaattgt aagcgtaaat 1680
 attttgtaaa aattcgcgtt aaatttttgi taaatcagct cattttttaa ccaataggcc 1740

 gaaatcgga aaatccctta taaatcaaaa gaatagaccg agatagggtt gagtgttggt 1800
 ccagtttga acaagagtc actattaaag aacgtggact ccaacgtcaa agggcgaaaa 1860
 accgtctatc agggcgatgg cccactacgt gaaccatcac cctaatacag ttttttgggg 1920
 tcgaggtgcc gtaaagcact aaatcggaac cctaaaggga gccccgatt tagagcttga 1980
 cggggaaagc cggcgaaagt ggcgagaaa gaagggaaga aagcgaaagg agcgggcgct 2040
 agggcgctgg caagtgtagc ggtcacgctg cgcgtaacca ccacaccgc cgcgcttaat 2100
 gcgccgtac agggcgctc gcgccattcg ccattcaggc tgcgcaactg ttgggaaggg 2160

 cgatcgggtc gggcctcttc gctattacgc cagctggcga aaggggatg tgctgcaagg 2220
 cgattaagtt gggtaacgcc agggttttcc cagtcacgac gttgtaaaac gacggccagt 2280
 gagcgcgct aatagactc actataggc gaattgggt cggggcccc cctcgaggtc 2340
 gacggtatcg ataagcttga tatcgaattc ctgcagcccg ggggatccac tagttctaga 2400
 gcggcgcca ccgcggtgga gctccagctt ttgttccct tagtgagggt taattgcgcg 2460
 cttggcgtaa tcatggtcat agctgtttcc tgtgtgaaat tgttatccgc tcacaattcc 2520
 acacaacata ggagccggaa gcataaagt taaagcctgg ggtgcctaata gagttaggta 2580

 actcacatta atigcgttgc gctcactgcc cgctttccag tcgggaaacc tgcgtgccca 2640
 gctgcattaa tgaatcgcc aacgcgcggg gagaggcggg ttgcgtattg ggcgctcttc 2700
 cgcttctcgc ctactgact cgctgcgctc ggtcgttcgg ctgcggcgag cggtatcagc 2760
 tcactcaaag gcggtaatc ggttatccac agaatcagg gataacgcag gaaagaacat 2820

gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc tggcgttttt	2880
ccataggtctc cgccccctg acgagcatca caaaaatcga cgctcaagtc agaggtggcg	2940
aaaccgcaca ggactataaa gataccaggc gtttccccct ggaagctccc tcgtgcgtc	3000
tctgttccg accctgccgc ttaccggata cctgtccgcc tttctccctt cgggaagcgt	3060
ggcgctttct catagctcac gctgtaggta tctcagttcg gtgtaggtcg ttcgctccaa	3120
gctgggctgt gtgcacgaac ccccgttca gcccgaccgc tgcgccttat ccgtaacta	3180
tcgtcttag tccaaccgg taagacacga cttatcgcca ctggcagcag cactggtaa	3240
caggattagc agagcgaggt atgtaggcgg tgctacagag ttcttgaagt ggtggcctaa	3300
ctacggctac actagaagga cagtatttgg tatctgcgt ctgctgaagc cagttacctt	3360
cggaaaaaga gttagtagct cttgatccgg caaacaacc accgctggta gcggtggttt	3420
ttttgtttgc aagcagcaga ttacgcgcag aaaaaagga tctcaagaag atcctttgat	3480
cttttctacg ggttctgacg ctacgtggaa cgaaaactca cgtaaggga ttttggtcat	3540
gagattatca aaaaggatct tcacctagat ctttttaaat taaaaatgaa gttttaatc	3600
aatctaaagt atatatgagt aaacttggtc tgacagttac caatgcttaa tcagtgagc	3660
acctatctca gcgatctgtc tatttcgttc atccatagtt gcctgactcc ccgtcgtgta	3720
gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga taccgcgaga	3780
cccacgtca ccggctccag atttatcagc aataaaccag ccagccggaa gggccgagcg	3840
cagaagtggc cctgcaactt tatccgctc catccagtct attaattgtt gccgggaagc	3900
tagagtaagt agttgccag ttaatagttt gcgcaacgtt gttgccattg ctacaggcat	3960
cgtaggtgtca cgctcgtctg ttggatggc ttatttcagc tccggttccc aacgatcaag	4020
gcgagttaca tgatcccca tgttgtgcaa aaaagcggtt agtccttcg gtctccgat	4080
cgttgtcaga agtaagtgg ccgcagtgtt atcactcatg gttatggcag cactgcataa	4140
ttctcttact gtcatgcat ccgtaagatg cttttctgtg actggtgagt actcaacaa	4200
gtcattctga gaatagtga tgcggcgacc gagttgctct tgcccggcgt caatacggga	4260
taataccgcg ccacatagca gaactttaa agtgcctatc attggaaaac gttcttcggg	4320
gcgaaaactc tcaaggatct taccgtgtt gagatccagt tcgatgtaac cactcgtgc	4380
acccaactga tcttcagcat cttttacttt caccagcgtt tctgggtgag caaaaacagg	4440
aaggcaaaat gccgcaaaaa agggataaag ggcgacacgg aaatgttgaa tactcatact	4500
cttccttttt caatattatt gaagcattta tcagggttat tgtctcatga gcggatacat	4560
atttgaatgt atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt	4620

gccacctgaa cgaagcatct gtgcttcatt ttgtagaaca aaaatgcaac gcgagagcgc 4680

taatTTTTca aacaaagaat ctgagctgca tttttacaga acagaaatgc aacgcgaaag 4740

cgctatTTTta ccaacgaaga atctgtgctt catTTTTgtA aaacaaaaat gcaacgcgag 4800

agcgctaatt ttTcaaacaa agaatctgag ctgcattttt acagaacaga aatgcaacgc 4860

gagagcgcta tttTaccAAC aaagaatcta tacttctttt ttgttctaca aaaatgcatc 4920

ccgagagcgc tatTTTTcta acaaagcatc ttagattact tttttctcc tttgtgcgct 4980

ctataatgca gtctcttgat aactTTTTgc actgtaggtc cgTTaaggTt agaagaaggc 5040

tactttggTg tctatTTTct cttccataaa aaaagcctga ctccacttcc cgcgtttact 5100

gattactagc gaagctgcgg gtgcattttt tcaagataaa ggcatccccg attatattct 5160

ataccgatgt ggattgcgca tactttgtga acagaaagtG atagcgttga tgattcttca 5220

ttggtcagaa aattatgaac ggtttcttct attttGctc tatatactac gtataggaaa 5280

tgtttacatt ttcgtattgt tttcgattca ctctatgaat agttcttact acaattttt 5340

tgtctaaaga gtaatactag agataaacat aaaaaatgta gaggtcagat ttagatgcaa 5400

gttcaaggag cgaaggtgg atgggtaggt tatataggga tatagcacag agatatatag 5460

caaagagata cttttgagca atgtttgtgg aagcggTatt cgcaatatTT tagtagctcg 5520

ttacagtccg gtgcgttttt ggttttttga aagtgcgtct tcagagcgct ttTggttttc 5580

aaaagcgctc tgaagttcct atactttcta gagaatagga acttcggaat aggaacttca 5640

aagcgtttcc gaaaacgagc gcttccgaaa atgcaacgcg agctgcgcac atacagctca 5700

ctgttcacgt cgcacctata tctgcgtgtt gctgtatat atatatacat gagaagaacg 5760

gcatagtgcg tgtttatgct taaatgcgta cttatatgcg tctatTTatg taggatgaaa 5820

ggtagctag tacctctgt gatattatcc cattccatgc ggggtatcgt atgcttcctt 5880

cagcactacc ctttagctgt tctatatgt gccactctc aattggatta gtctcatcct 5940

tcaatgctat catttcctt gatattggat catactaaga aaccattatt atcatgacat 6000

taacctataa aaataggcgt atcacgagc cttttcgtc 6039

<210> 12

<211> 5662

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: pAM114 vector linearized using DRPE308/309 (the linear cassette was co-transformed with vector linearized by PCR inside of the NatA marker)

<400> 12

atggggttca cctctgcgg cctggacacc gccctgtacg acggcaccgc ctccgacggc	60
gagcaggcgc tctacatgag catgccctgc ccctgagttt aacttgatac tactagattt	120
tttctcttca ttataaaat ttttggttat aattgaagct ttagaagtat gaaaaaatcc	180
ttttttttca ttctttgcaa ccaaaataag aagcttcttt tattcattga aatgatgaat	240
ataaacctaa caaaagaaaa agactcgaat atcaaacatt aaaaaaaaaat aaaagaggtt	300
atctgttttc ccatttagtt ggagtttgca ttttctaata gatagaactc tcaattaatg	360
tggatttagt ttctctgttc gttttttttt gttttgttct cactgtattt acatttctat	420
ttagtattta gttattcata taatcttaac ttgcggtgtg aaataccgca cagatgcgta	480
aggagaaaaat accgcatcag gaaattgtaa gcgttaatat ttgttataaa ttcgcgttaa	540
atttttgtta aatcagctca tttttaacc aataggccga aatcggaata atcccttata	600
aatcaaaaga atagaccgag atagggttga gtgttgttcc agtttggaac aagagtcac	660
tattaaagaa cgtggactcc aacgtcaaag ggcgaaaaac cgtctatcag ggcgatggcc	720
cactacgtga accatcaccc taatcaagtt ttttggggtc gaggtgccgt aaagcactaa	780
atcggaaccc taaagggagc ccccgattta gagcttgacg gggaaagccg gcgaacgtgg	840
cgagaaagga aggggaagaa gcgaaaggag cgggcgctag ggcgctggca agtgtagcgg	900
tcacgctgcg cgtaaccacc acaccgccg cgcttaatgc gccgctacag ggcgctcgc	960
gccattcgcc attcaggctg cgcaactgtt gggaaggcg atcgggtcgg gcctcttcgc	1020
tattacgcca gctggcgaaa gggggatgtg ctgcaaggcg attaatgttg gtaacgccag	1080
ggttttccca gtcacgacgt tgtaaaacga cgccagtgta gcgcgcgtaa tacgactcac	1140
tataggcgca attgggtacc gggccccccc tcgaggtcga cggatcgcg aagcttgata	1200
tcgaattcct gcagcccggtt ggatccacta gttctagagc ggccgccacc gcggtggagc	1260
tccagctttt gttcccttta gtgagggtta attgcgcgt tggcgtaatc atggcatag	1320
ctgtttctg tgtgaaattg ttatccgctc acaattccac acaacatagg agccggaagc	1380
ataaagtgt aagcctgggg tgcctaatga gtgaggtaac tcacattaat tgcgttcgc	1440
tcactgccc ctttccagtc gggaacctg tcgtgccagc tgcattaatg aatcgccaa	1500
cgcgcgggga gaggcggtt cggtattggg cgtcttccg cttctcgt cactgactcg	1560
ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc actcaaagc ggtaaacgg	1620
ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag	1680
gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg cccccctgac	1740

gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga 1800

taccaggcgt ttccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt 1860

accggatacc tgtccgcttt tctcccttcg ggaagcgtgg cgctttctca tagctcacgc 1920

tgtaggtatc tcagttcggg gtaggtcgtt cgctccaagc tgggctgtgt gcacgaaccc 1980

cccgttcagc ccgaccgtg cgccttatcc ggtaactatc gtcttgagtc caacccggtg 2040

agacacgact tatcgccact ggcagcagcc actggtaaca ggattagcag agcgaggat 2100

gtaggcgggt ctacagagtt cttgaagtgg tggcctaact acggctacac tagaaggaca 2160

gtatttggtg tctgcgctct gctgaagcca gttaccttcg gaaaaagagt tggtagctct 2220

tgatccggca aacaaaccac cgctggtagc ggtgggtttt ttgtttgcaa gcagcagatt 2280

acgcgcagaa aaaaaggatc tcaagaagat cctttgatct tttctacggg gtctgacgct 2340

cagtggaaacg aaaactcacg ttaagggatt ttggtcatga gattatcaa aaggatcttc 2400

acctagatcc ttttaaatta aaaatgaagt tttaaatcaa tctaaagtat atatgagtaa 2460

acttggctg acagttacca atgcttaatc agtgaggcac ctatctcagc gatctgtcta 2520

tttcgttcat ccatagtgc ctgactcccc gtctgttaga taactacgat acgggagggc 2580

ttaccatctg gccccagtcg tgcaatgata ccgcgagacc cacgctcacc ggctccagat 2640

ttatcagcaa taaaccagcc agccggaagg gccgagcgca gaagtgggcc tgcaacttta 2700

tccgcctcca tccagtctat taattgttgc cgggaagcta gagtaagtag ttcgccagtt 2760

aatagtttgc gcaacgttgt tgccattgct acaggcatcg tgggtgcacg ctctctgttt 2820

ggatatggctt cattcagctc cggttcccaa cgatcaaggc gagttacatg atcccccatg 2880

ttgtgcaaaa aagcggttag ctctctcggg cctccgatcg ttgtcagaag taagtggcc 2940

gcagtgttat cactcatggt tatggcagca ctgcataatt ctcttactgt catgccatcc 3000

gtaagatgct tttctgtgac tggtagtac tcaaccaagt cattctgaga atagtgtatg 3060

cggcgaccga gtgtctcttg cccggcgtca ataccgggata ataccgcgc acatagcaga 3120

actttaaaag tgtctcatcat tggaaaacgt tcttcggggc gaaaactctc aaggatctta 3180

ccgctgttga gatccagttc gatgtaacct actcgtgcac ccaactgatc ttcagcatct 3240

tttactttca ccagcgtttc tgggtgagca aaaacaggaa ggcaaaatgc cgcaaaaaag 3300

ggaataaggc cgacacggaa atgttgaata ctcatactct tctttttca atattattga 3360

agcatttatc agggttattg tctcatgagc ggatacatat ttgaatgtat ttagaaaaat 3420

aaacaaatag gggttccgcg cacatttccc cgaagaatgc cactgaacg aagcatctgt 3480

gcttcatttt gtagaacaaa aatgcaacgc gagagcgcta atttttcaaa caaagaatct 3540
gagctgcatt ttacagaac agaaatgcaa cgcgaaagcg ctattttacc aacgaagaat 3600
ctgtgcttca tttttgtaaa acaaaaatgc aacgcgagag cgctaatttt tcaaacaag 3660
aatctgagct gcatTTTTac agaacagaaa tgcaacgcga gagcgctatt ttaccaacaa 3720
agaatctata ctctTTTTtt gtctacaaa aatgcatccc gagagcgcta tttttctaac 3780
aaagcatctt agattacttt tttctctctt tgtgcgctct ataatgcagt ctcttgataa 3840
ctttttgcac tgiaggctcg ttaaggtttag aagaaggcta ctttgggtgc tattttctct 3900

tccataaaaa aagcctgact ccacttcccg cgtttactga ttactagcga agctgcgggt 3960
gcattttttc aagataaagg catccccgat tatattctat accgatgtgg attgcgcata 4020
ctttgtgaac agaaagtgat agcgttgatg attcttcatt ggtcagaaaa ttatgaacgg 4080
ttctctctat ttgtctctta tatactacgt ataggaaatg ttacatttt cgtattgttt 4140
tcgattcact ctatgaatag ttcttactac aatTTTTttg tctaaagagt aatactagag 4200
ataaacataa aaaatgtaga ggtcagttt agatgcaagt tcaaggagcg aaaggtggat 4260
gggtaggtta tatagggata tagcacagag atatatagca aagagatact tttgagcaat 4320

gtttgtggaa gcggtattcg caatatttta gtagctcgtt acagtccggt gcgtttttgg 4380
ttttttgaaa gtgcgtcttc agagcgcttt tggttttcaa aagcgctctg aagttcctat 4440
actttctaga gaataggaac ttcggaatag gaacttcaaa gcgtttccga aaacgagcgc 4500
ttccgaaaat gcaacgcgag ctgcgcacat acagctcact gttcacgtcg cacctatctc 4560
tgcgtgttgc ctgtatatat atatacatga gaagaacggc atagtgcgtg tttatgctta 4620
aatgcgtact tatatgcgtc tatttatgta ggatgaaagg tagtctagta cctcctgtga 4680
tattatccca ttccatgcgg ggtatcgtat gcttcttca gcactaccct ttagctgttc 4740

tatatgctgc cactctcaa ttggattagt ctcatcttc aatgctatca tttcctttga 4800
tattggatca tactaagaaa ccattattat catgacatta acctataaaa ataggcgtat 4860
cacgaggccc tttcgtctcg cgcgtttcgg tgatgacggt gaaaacctct gacacatgca 4920
gtccccggag acggtcacag cttgtctgta agcggatgcc gggagcagac aagcccgtca 4980
gggcgcgtca gcgggtgttg gcgggtgtcg gggctggctt aactatgcgg catcagagca 5040
gattgtactg agagtcaaaa tacacatcat cgtctacaa gttcatcaaa gtgttgagca 5100
gacaactata ccagcatgga tctcttgat cggttctttt ctcccgtct ctgcgcaataa 5160

caatgaacac tgggtcaatc atagcctaca caggtgaaca gagtagcgtt tatacagggt 5220
ttatacgggt attcctacgg caaaaatttt tcatctctaa aaaaaaaag aaaaattttt 5280
ctttccaacg ctagaaggaa aagaaaaatc taattaaatt gatttgggtga tttctgaga 5340

gttccctttt tcatatcgc aattttgaat ataaaaggag atcgaaaaaa tttttctatt 5400
 caatctgttt tctggtttta tttgatagtt tttttgtgta ttattattat ggattagtag 5460
 tggtttatat gggtttttct gtataacttc tttttatttt agtttgttta atcttatttt 5520
 gagttacatt atagtccct aactgcaaga gaagtaacat taaaaatgac cactcttgac 5580

gacacggctt accggtaccg caccagtgtc ccgggggacg ccgaggccat cgaggcactg 5640
 gatgggtcct tcaccaccga ca 5662

<210> 13

<211> 651

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: DRPE310/311 PCR fragment from pAM1114 (a complementary overlapping NatA ORF fragment)

<400> 13

cttattttga gttacattat agttccctaa ctgcaagaga agtaacatta aaaatgacca 60
 ctcttgacga cacggcttac cggtaaccga ccagtgtccc gggggacgcc gaggccatcg 120
 aggcactgga tgggtccttc accaccgaca ccgtcttcgc cgtcacgcc accggggacg 180

gcttcaccct gggggaggtg ccggtggacc cgccctgac caagtggtc cccgacgacg 240
 aatcggacga cgaatcggac gccggggagg acggcgaccc ggactcccgg acgttcgtcg 300
 cgtacgggga cgacggcgac ctggcgggct tcgtgggtcgt ctcgtactcc ggctggaacc 360
 gccggctgac cgtcaggac atcgaggtcg ccccgagca ccgggggcac ggggtcgggc 420
 gcgcgttgat ggggctcgcg acggagttcg cccgcgagcg gggcgccggg cacctctggc 480
 tggaggtcac caacgtcaac gcaccggcga tccacgcgta ccggcggatg gggttcacc 540
 tctgcgcct ggacaccgcc ctgtacgacg gcaccgcctc ggacggcgag caggcgtct 600

acatgagcat gccctgcccc tgagttaac ttgatactac tagattttt c 651

<210> 14

<211> 1024

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: Donor DNA for integration at RHR2 locus

<400> 14

ggaaatccgt atcatcttct cgatacacg aaccgcgtg cgcctggtaa attgcaggat 60
 tctcattgtc cggttttctt tatgggaata atcatcatca ccattatcac tgttactctt 120

gcatcatca tcaataacat aatTTTTTTT acgctgtttg atgatggat gtgcttttgt 180
 tgttccttac tcaccttttc ctttgtgtct ttttaattttg accattttga ccattttgac 240

ctttgatgat gtgtgagttc ctcttttctt tttttctttt cttttttcct ttttttttct 300
 tttcttactc tgtaatacac tttcttttct tttctttcat attgtcgtct tgttcatttt 360
 cgttcaattg ataagtata taaatctttc ctaagtatct cttgattgcc atttttttct 420
 ttccaagttt ccttggtatg aaacgtttca atgttttaaa atatatcaga acaacaaaag 480
 caaatataca aaccatcgca cgctcgcca acgccggcgg acctttttct tttatttttt 540
 tgataaaact actacgctaa aaataaaata aaaatgtatg atttccttcc atttcgacc 600
 aattgtataa ttttatactt gcatgactta ataataaat ataatactta taaaatacga 660

atagaaaaat ttaaaccgat gtaatgcac cttttctttg ttgtcttcgg atgatctgcc 720
 gtgacaggtg gttcgcgcaa atcaaaactgg attagagaat ttaacacaga aataaaaaag 780
 gaagattcaa tcttcgtttt tgttttataat ctactataa aagtgttttt ttttagtacg 840
 acgagaagtt agagtaatta taaaaggaat gcttatttaa atttatttct tagacttctt 900
 ttcagacttc ttagcagcct cagtttgttc cttaacgacc ttcttaacaa tcttttgttc 960
 ctcaatcaag aaagctctga cgattctttc cttgacacag ttggcacatc tggaccacc 1020
 gtaa 1024

<210> 15

<211> 1024

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: Donor DNA for integration at HO locus

<400> 15

cgcaagtctt gtttctatgc ctttctctta gtaattcacg aaataaacct atggtttacg 60
 aaatgatcca cgaaaatcat gttattatct acatcaacat atcgcgaaaa ttcattgtcat 120
 gtccacatta acatcattgc agagcaacaa ttcattttca tagagaaatt tgctactatc 180
 acccactagt actaccattg gtacctacta ctttgaattg tactaccgct gggcgttatt 240
 aggtgtgaaa ccacgaaaag ttcaccataa ctctgaataa agtcgcggaa aaaagtaaac 300

agctattgct acicaaatga ggtttgcaga agcttgttga agcatgatga agcgttctaa 360
 acgcactatt catcattaaa tatttaaagc tcataaaatt gtattcaatt cctattctaa 420
 atggctttta tttctattac aactattagc tctaaatcca tatctcata agcagcaatc 480
 aattctatct atactttaaa cgctcgcca acgccggcgg acctaatgtg tatattagtt 540

taaaaagtgtg tatgtaataa aagtaaaatt taatatatttg gatgaaaaaa accattttta 600
gactttttct taactagaat gctggagtag aaatacgcca tctcaagata caaaaagcgt 660
taccggcact gatttgtttc aaccagtata tagattatta ttgggtcttg atcaactttc 720

ctcagacata tcagtaacag ttatcaagct aaatatattac gcgaaagaaa aacaaatatt 780
ttaattgtga tacttgtgaa ttttatttta ttaaggatac aaagttaaga gaaaacaaaa 840
tttatataca atataagtaa tattcatata tatgtgatga atgcagtctt aacgagaaga 900
catggccttg gtgacaactc tcttcaaacc aacttcagcc tttctcaatt catcagcaga 960
tgggtcttcg atttgcaaag cagccaaagc atcggacaaa gcagcttcaa tcttggactt 1020
ggaa 1024

<210> 16

<211> 1024

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: Donor DNA for integration at ADH5 locus

<400> 16

taattttaca agtagtgtct tcatgacgga ttcatagtct atccaagcgt ttgcccaaaa 60
ttttgcagta aatttaatgt tttctgtata ggataatggg gcgccattca agtcccgca 120
aaaatgactg atgtctacag gacaggggcg caatatatgt tctctgacat tgcacctttt 180
gaatatatca tgtgtttgtt ctgctatctg cttgtagaag ggtacgctaa cagagccggc 240
ggcattgtaa aaagtctcc tttcgcgga ggatgagtca aaaagcatgt gacaatgaaa 300
taatcaaatt gtgacatctg ctgacgcggg atcggttcctt cgtattgtct agattgtaat 360

ctatataaca tactacgaat ataaaagagg gactacaaga ttttctagc gcaaactact 420
gctttactgt ctcaaatgt ctctgattgg aagataccta agaaaattat ttaactacat 480
atctacaaaa tcaaagcatc cgctcgtcca acgccggcgg accttctttt gtaacgaatt 540
tgatgaatat atttttactt tttatataag ctattttgta gatattgact ttttacgatt 600
tatttgaac aatgagaatt actccatttc tgaacttcag taaatagcga gtgattctgt 660
actttgcgag aaccggtgga catttggtat ttgccttac aagaacaacc tatacaaagc 720
tttcaatata taattctttg taatccattg ttttacgaga catataatgt gatatataga 780

tgaactttac gtataaaatg atatatattaa aactagcaac tgcgtgcgta agacaaactg 840
aaataggcca tttacggaaa agaaatttaa taatgtcgac tggaaactga aaccaggagg 900
agtagaaatt ggtaaattg attagctaaa atttactcgt tgtggacaga gtttgagcca 960

agcgggaatgt ttccaaggct ttctttgttt cgaaggcag ctctggctcc tgcccctatg 1020

agaa 1024

<210> 17

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: RHR2 upstream primer

<400> 17

gggtgcgaag taccaccag tttctttttc atctct 36

<210> 18

<211> 38

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: HO upstream primer

<400> 18

acgtgtgtgt ctcatggaaa ttgatgcagt tgaagaca 38

<210> 19

<211> 33

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: ADH5 upstream primer

<400> 19

ggcgttatat ccaaacattt cagacagaag att 33

<210> 20

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: Primer for deletion linker

<400> 20

aggtccgccg gcgttgacg agcg 24

<210> 21

<211> 1078

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: Donor DNA for integration at Gal80 locus

<400> 21

gacggcacgg ccacgcgttt aaaccgcccga gatggaatcc cttccataga gagaaggagc	60
aagcaactga cccaatatgt actgccactg gacctgaaga catgcaacaa agtgcaagca	120
tagtggggcc ttcttccaat gctaatacgg tcactgccac tgctgctacg gaaaaccaac	180
ctaaaggtat taacttcttc actataagaa aatcacacga gcgcccggac gatgtctctg	240
tttaaatggc gcaagttttc cgctttgtaa tatatattta tacccttttc ttctctcccc	300
tgcaatataa tagtttaatt ctaatatata taatatccta tattttcttc atttacggc	360
gcactctcgc ccgaacgacc tcaaaatgtc tgctacattc ataataacca aaagctcata	420
actttttttt ttgaacctga atatatatac atcacatgtc actgctggtc ctgcccggac	480
agcgtataca atctcgatag ttggtttccc gttctttcca ctcccgctcg ctgctccaac	540
gccggcggac ctaagcatct tgccctgtgc ttggcccca gtgcagcgaa cgttataaaa	600
acgaatactg agtatataac tatgtaaaac aaccatatca tttcttgttc tgaactttgt	660
ttacctaact agttttaaat ttcccttttt cgtgcatgcg ggtgttctta tttattagca	720
tactacattt gaaatatcaa atttccttag tagaaaagt agagaagggt cactgacaca	780
aaaaataaaa tgctacgtat aactgtcaaa actttgcagc agcgggcac cttccatcat	840
agcttcaaac atattagcgt tctgatctt cataccgtg ctcaaatga tcaacaaac	900
tgttattgcc aagaaataaa cgcaaggctg cttcaaaaa ctgatccatt agatcctcat	960
atcaagcttc ctcatagaac gcccaattac aataagcatg ttttgctgtt atcacgggt	1020
gataggtttg ctcaaccatg gaaggtagca cgggtgttaa accccagcgc ctggcggg	1078

<210> 22

<211> 1069

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: Donor DNA for integration at H0 locus

<400> 22

gacggcacgg ccacgcgttt aaaccgcccga caagtctgt ttctatgcct ttctcttagt	60
aattcacgaa ataaacctat ggtttacgaa atgatccacg aaaatcatgt tattatttac	120
atcaacatat cgcgaaaatt catgtcatgt ccacattaac atcattgcag agcaacaatt	180
cattttcata gagaaatttg ctactatcac ccactagtag taccattggt acctactact	240

ttgaattgta ctaccgctgg gcgttattag gtgtgaaacc acgaaaagtt caccataact 300

tcgaataaag tcgcggaaaa aagtaaacag ctattgctac tcaaatgagg ttgcagaag 360

cttgttgaag catgatgaag cgttctaaac gcactattca tcattaaata tttaaagctc 420

ataaaattgt attcaattcc tattctaaat ggcttttatt tctattacaa ctattagctc 480

taaatccata tcttcataag cagcaatcaa ttctcgctcg tccaacgccg gcggacctaa 540

tgtgtatatt agtttaaaaa gttgtatgta ataaaagtaa aatttaatat ttggatgaa 600

aaaaaccatt tttagacttt ttcttaacta gaatgctgga gtagaaatac gccatctcaa 660

gatacaaaaa gcgttaccgg cactgatttg tttcaaccag tatatagatt attattgggt 720

cttgatcaac tticctcaga catatcagta acagttatca agctaaatat ttacgcgaaa 780

gaaaaacaaa tattttaatt gtgatacttg tgaattttat tttattaagg atacaaagtt 840

aagagaaaac aaaatttata tacaatataa gtaatatcca tatatatgtg atgaatgcag 900

tcttaacgag aagacatggc cttggtgaca actctcttca aaccaacttc agcctttctc 960

aattcatcag cagatgggtc ttcgatttgc aaagcagcca aagcatcgga caaagcagct 1020

tcaatcttgg acttggaaacc tcggtgttta aaccccagcg cctggcggg 1069

<210> 23

<211> 1053

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: Donor DNA for integration at ADH5 locus

<400> 23

gacggcacgg ccacgcgttt aaaccgcaa ttttacaagt agtgtcttca tgacggattc 60

atagtctatc caagcgtttg cccaaaattt tgcagtaaata ttaattgttt ctgtatagga 120

taatggtgcg ccattcaagt cccgcgaaaa atgactgatg tctacaggac aggggcgcaa 180

tatatgttct ctgacattgc accttttgaa tatatcatgt gtttgttctg ctatctgctt 240

gtagaagggt acgctaacag agccggcggc attgtaaaaa gttctccttt cgcggaagga 300

tgagtcaaaa agcatgtgac aatgaaataa tcaaattgtg acatctgctg acgcgggatc 360

gttcccttctg attgtctaga ttgtaactta tataacatac tacgaatata aaagagggac 420

tacaagatat ttctagcgca aactactgct ttactgtctc acaatgtctc tgattggaag 480

atacctaaga aaattattta actacatatc taaaaaatca aagcatccgc tcgtccaacg 540

ccggcggacc ttcttttgta acgaatttga tgaatatatt ttacttttt atataagcta 600

ttttgtagat attgactttt tacgatttat ttgtaacaat gagaattact ccatttctga 660

acttcagtaa atagcgagtg attctgtact ttgcgagaac cggtaggacat ttggtatttt 720
gccttacaag aacaacctat acaaacgttt caatatctaa ttcittgtaa tccattgttt 780

tacgagacat ataattgtgat atatatagatga actttacgta taaaatgata tatttaaaac 840
tagcaactgc gtgcgtaaga caaactgaaa taggccatit acggaagaaga aatttaataa 900
tgtcgactgg aaactgaaac caggaggagt agaaatttgt taaattgatt agctaaaatt 960
tactcgttgt ggacagagtt tgagccaagc ggaatgtttt caaggctttc ttgtttcga 1020
agggccggtg tttaaaccac agcgctggc ggg 1053

<210> 24
<211> 1026
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic: gRNA construct targeting Gal80 locus

<400> 24

ccttataaat caaagaata gaccgagata gggttgagtg ttgttccagt ttggaacaag 60
agtccactat taaagaacgt ggactccaac gtcaaggaggc gaaaaaccgt ctatcagggc 120
gatggccac tacgtgaacc atcacctaa tcaagttttt tggggtcgag gtgccgtaaa 180
gcactaaatc ggaaccctaa agggagcccc cgatttagag cttgacgggg aaagcctctt 240
tgaaaagata atgtatgatt atgctttcac tcatattat acagaaactt gatgttttct 300
ttcgagtata tacaaggta ttacatgtac gtttgaagta caactctaga tttttagtg 360
ccctcttggg ctacgggtaa aggtgcgcat tttttcacac cctacaatgt tctgttcaaa 420

agatttttgt caaacgctgt agaagtgaag gttggtgcgc atgtttcggc gttcgaaact 480
tctccgcagt gaaagataaa tgatcgtaag gctgctgctg aacgtgtttt agagctagaa 540
atagcaagtt aaaataaggc tagtccgtta tcaacttgaa aaagtggcac cgagtcggtg 600
gtgctttttt tgttttttat gtctcagctt ttgttccctt tagtgagggt taattgcgcg 660
cttggcgtaa tcatggcat agctgtttcc tgtgtgaaat tgttatccgc tcacaattcc 720
acacaacata ggagccggaa gcataaagt taaagcctgg ggtgcctaata gagtgaggta 780
actcacatta attgcgttgc gctcactgcc cgctttccag tcgggaaacc tgcgtgcc 840

gctgcattaa tgaatcgccc aacgcgcggg gagaggcggg ttgcgtattg ggcgccttc 900
cgcttctcg ctactgact cgctgcgctc ggtcgttcgg ctgcggcgag cggtatcagc 960
tactcaaag gcgtaatac ggttatccac agaatacagg gataacgcag gaaagaacat 1020
gtgagc 1026

<210> 25
 <211> 1026
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: gRNA construct targeting H0 locus
 <400> 25
 cttataaat caaaagaata gaccgagata gggttgagtg ttgttccagt ttggaacaag 60
 agtcactat taaagaacgt ggactccaac gtcaaagggc gaaaaaccgt ctatcagggc 120
 gatggccac tacgtgaacc atcacctaa tcaagttttt tggggtcgag gtgccgtaaa 180
 gcactaaatc ggaaccctaa agggagcccc cgatttagag cttgacgggg aaagcctctt 240
 tgaaaagata atgtatgatt atgctttcac tcatatttat acagaaactt gatgttttct 300
 ttcgagtata tacaaggatg ttacatgtac gtttgaagta caactctaga ttttgiagt 360
 ccctcttggg ctacgggtaa aggtgcgcgt ttttcacac cctacaatgt tctgttcaaa 420
 agattttggg caaacgctgt agaagtgaat gttggtgcgc atgtttcggc gtgcgaaact 480
 tctccgcagt gaaagataaa tgatcgccgg cttgatcgac tcagagtttt agagctagaa 540
 atagcaagtt aaaataagcg tagtccgtta tcaacttgaa aaagtggcac cgagtcggtg 600
 gtgctttttt tgttttttat gtctcagctt ttgttccctt tagtgagggt taattgcgcg 660
 cttggcgtaa tcatggatcat agctgtttcc tgtgtgaaat tgttatccgc tcacaattcc 720
 acacaacata ggagccggaa gcataaagtg taaagcctgg ggtgcctaata gagtgaggta 780
 actcacatta attgcgttgc gctcactgcc cgttttccag tcgggaaacc tgcgtgccca 840
 gctgcattaa tgaatcgccc aacgcgcggg gagaggcggg ttgcgtattg ggcgctcttc 900
 cgcttctctg ctactgact cgctgcgctc ggtcggtcgg ctgcggcgag cggtatcagc 960
 tcactcaaag gcggtaatat ggttatccac agaatcaggg gataacgcag gaaagaacat 1020
 gtgagc 1026
 <210> 26
 <211> 1026
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: gRNA construct targeting ADH5 locus
 <400> 26
 cttataaat caaaagaata gaccgagata gggttgagtg ttgttccagt ttggaacaag 60

agtcactat taaagaacgt ggactccaac gtcaaagggc gaaaaaccgt ctatcagggc 120

gatggccac tacgtgaacc atcacctaa tcaagttttt tggggtcgag gtgccgtaaa 180

gcactaaatc ggaaccctaa agggagcccc cgatttagag cttgacgggg aaagcctctt 240

tgaagata atgtatgatt atgctttcac tcataattat acagaaactt gatgttttct 300

ttcgagtata tacaaggta ttacatgtac gtttgaagta caactctaga tttttagtg 360

ccctcttggg ctacgggtaa aggtgcgcgt tttttcacac cctacaatgt tctgttcaaa 420

agattttgtt caaacgctgt agaagtgaat gttggtgcgc atgtttcggc gttcgaaact 480

tctccgcagt gaaagataaa tgatcggggg catgtgtatc gatgggtttt agagctagaa 540

atagcaagtt aaaataaggc tagtccgtta tcaacttgaa aaagtggcac cgagtcgggtg 600

gtgctttttt tgttttttat gtctcagctt ttgttccctt tagtgagggt taattgcgcg 660

cttggcgtaa tcatgggtcat agctgtttcc tgtgtgaaat tgttatccgc tcacaattcc 720

acacaacata ggagccggaa gcataaagt taaagcctgg ggtgcctaata gagtgaggta 780

actcacatta attgcgttgc gctcactgcc cgctttccag tcgggaaacc tgcgtgccca 840

gctgcattaa tgaatcgccc aacgcgcggg gagaggcggg ttgcgtattg ggcgctcttc 900

cgcttctcgc ctactgact cgctgcgctc ggtcgttcgg ctgcggcgag cggtatcagc 960

tactcaaag gcggtaatat ggttatccac agaatcaggg gataacgcag gaaagaacat 1020

gtgagc 1026

<210> 27

<211> 5975

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: Vector backbone for 2-piece assembly

<400> 27

gttttagagc tagaaatagc aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt 60

ggcaccgagt cgggtgtgct tttttgttt tttatgtctc agcttttgtt cccttttagtg 120

agggttaatt gcgcgcttgg cgtaatcatg gtcatactg tttcctgtgt gaaattgtta 180

tccgctcaca attccacaca acataggagc cggaagcata aagtgtaaag cctgggggtgc 240

ctaagagtg aggttaactca cattaattgc gttgcgctca ctgcccgctt tccagtcggg 300

aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag gcggtttgcg 360

tattgggcgc tcttccgctt cctcgtctac tgactcgtg cgctcggtcg ttcggctgcg 420

gcgagcggta tcagctcact caaagcggtt aatcgggtta tccacagaat caggggataa 480

cgaggaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc 540
gttgctggcg tttttccata ggctccgcc cctgacgag catcacaaaa atcgacgctc 600

aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag 660
ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt ccgcctttct 720
cccttcggga agcgtggcg tttctcatag ctacgctgt aggtatctca gttcgggtga 780
ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc gttcagcccg accgctgcgc 840
cttatccggt aactatcgtc ttgagtcga cccggtgaaga cagacttat cgccactggc 900
agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagttctt 960
gaagtggtag cctaactacg gctacactag aaggacagta tttggtatct gcgctctgct 1020

gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc 1080
tggtagcgtt ggtttttttg ttgcaagca gcagattacg cgcagaaaaa aaggatctca 1140
agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa actcacgtta 1200
agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa 1260
atgaagtttt aaatcaatct aaagtatata tgagtaaact tggcttgaca gttaccaatg 1320
ctaatcagt gaggcaccta tctcagcgt ctgtctatct cgttcaccca tagttgcctg 1380
actccccgtc gtgtagataa ctacgatac ggagggctta ccatctggcc ccagtctgc 1440

aatgataccg cgagaccac gctcaccggc tccagattta tcagcaataa accagccagc 1500
cggaagggcc gagcgagaa gtggctcctg aactttatcc gcctccatcc agtctattaa 1560
ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca acgtttgtgc 1620
cattgctaca ggcatcgtgg tgtcacgctc gtcgttttgt atggcttcat tcagctccgg 1680
ttcccaacga tcaaggcgag ttacatgat ccccatgttg tgcaaaaaag cggtagctc 1740
cttcggctct ccgacgttg tcagaagtaa gttggccgca gtgttatcac tcatggttat 1800
ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt ctgtgactgg 1860

tgagtactca accaagtcac tctgagaata gtgtatgcgg cgaccgagtt gctcttgccc 1920
ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc tcatcattgg 1980
aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat ccagttcgat 2040
gtaacccact cgtgcaccca actgatcttc agcatctttt actttacca gcgtttctgg 2100
gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataaggcgca cacggaaatg 2160
ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg gttattgtct 2220

catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg ttccgcgcac	2280
atttccccga aaagtgccac ctgaacgaag catctgtgct tcattttgta gaacaaaaat	2340
gcaacgcgag agcgctaatt ttccaacaa agaactctgag ctgcatTTTT acagaacaga	2400
aatgcaacgc gaaagcgcta ttttaccac gaagaatctg tgcttcattt ttgtaaaaca	2460
aaaatgcaac gcgagagcgc taatttttca aacaaagaat ctgagctgca tttttacaga	2520
acagaaatgc aacgcgagag cgctatTTTA ccaacaaaga atctatactt cttttttgtt	2580
ctacaaaaat gcatcccgag agcgctattt ttctaacaaa gcatcttaga ttactTTTT	2640
tctcctttgt gcgctctata atgcagtctc ttgataactt tttgcactgt aggtccgtta	2700
aggttagaag aaggctactt tgggtgtctat tttctcttcc ataaaaaaag cctgactcca	2760
cttcccgctt ttactgatta ctacggaagc tgcgggtgca ttttttcaag ataaaggcat	2820
ccccgattat attctatacc gatgtggatt gcgcatactt tgtgaacaga aagtgatagc	2880
gttgatgatt cticatttggc cagaaaatta tgaacggttt cttctatTTT gtctctatat	2940
actacgtata ggaaatgttt acattttcgt attgttttcg attcactcta tgaatagttc	3000
ttactacaat ttttttgtct aaagagtaat actagagata aacataaaaa atgtagaggt	3060
cgagtttaga tgcaagtcca aggagcgaaa ggtggatggg taggttatat agggatatag	3120
cacagagata tatagcaaag agatactTTT gagcaatgtt tgtggaagcg gtattcgcaa	3180
tatttttagta gctcgttaca gtccggtgcg tttttggttt tttgaaagtg cgtcttcaga	3240
gcgcttttgg ttttcaaaa cgctctgaag ttctatact ttctagagaa taggaacttc	3300
ggaataggaa cttcaaagcg tttccgaaaa cgagcgcttc cgaaaatgca acgcgagctg	3360
cgcacataca gctcactgtt cacgtcgcac ctatatctgc gtgttgccctg tatatatata	3420
tacatgagaa gaacggcata gtgcgtgttt atgcttaaata gcgtacttat atgcgtctat	3480
ttatgtagga tgaaaggtag tctagtacct cctgtgatat tatcccattc catgcggggt	3540
atcgtatgct tccttcagca ctacccttta gctgttctat atgctgccac tcctcaattg	3600
gattagtctc atccttcaat gctatcattt cctttgatat tggatcatat taagaaacca	3660
ttattatcat gacattaacc tataaaaaata ggcgatcac gaggcccttt cgtctcgcg	3720
gtttcggtga tgacggtgaa aacctctgac acatgcagct cccggagacg gtcacagctt	3780
gtctgtaagc ggatgccggg agcagacaag cccgtcaggg cgctcagcg ggtgttggcg	3840
ggtgtcgggg ctggcttaac tatgcggcat cagagcagat tgtactgaga gtgcaatac	3900
acatcatcgt cctacaagtt catcaaagtg ttggacagac aactatacca gcatggatct	3960

ctgttatcgg ttcttttctc ccgtctctc gcaataacaa tgaacactgg gtcaatcata 4020
 gcctacacag gtgaacagag tagcgtttat acagggttta tacggtgatt cctacggcaa 4080
 aaatttttca ttcttaaaaa aaaaaagaaa aatttttctt tccaacgcta gaaggaaaag 4140
 aaaaatctaa ttaaattgat ttggigattt tctgagagtt cccittttca tatatcgaat 4200
 ttggaatata aaaggagatc gaaaaaatit ttctattcaa tctgttttct ggttttatit 4260
 gatagttttt ttgtgtatta ttattatgga ttagtactgg ttatatggg tttttctgta 4320
 taacttcttt ttattttagt ttgtttaatc ttattttgag ttacattata gtccctaac 4380

 tgcaagagaa gtaacattaa aaatgaccac tcttgacgac acggcttacc ggtaccgcac 4440
 cagtgtcccg ggggacgccg aggccatcga ggcactggat gggtccttca ccaccgacac 4500
 cgtcttccgc gtcaccgcca ccggggacgg cttcaccttg cgggaggtgc cgggtggaccc 4560
 gccccgacc aaggtgttcc ccgacgacga atcggacgac gaatcggacg ccggggagga 4620
 cggcgacccg gactcccga cgttcgtcgc gtacggggac gacggcgacc tggcgggctt 4680
 cgtggtcgtc tcttactccg gctggaaccg ccggtgacc gtcgaggaca tcgaggtcgc 4740
 cccggagcac cgggggcacg gggtcgggcg cgcgttgatg gggctcgcga cggagttcgc 4800

 ccgcgagcgg ggcgccgggc acctctggct ggaggtcacc aacgtcaacg caccggcgat 4860
 ccacgcgtac cggcggatgg gggtcacct ctgcggcctg gacaccgcc tgtacgacgg 4920
 caccgcctcg gacggcgagc aggcgctcta catgagcatg cctgccctt gagtttaact 4980
 tgatactact agattttttc tcttcattta taaaattttt gggtataatt gaagctttag 5040
 aagtatgaaa aaatcctttt ttttcattct ttgcaaccaa aataagaagc ttcttttatt 5100
 cattgaaatg atgaatataa acctaacaaa agaaaaagac tcgaatatca aacattaaaa 5160
 aaaaataaaa gaggttatct gttttcccat ttagttggag ttgcatitit ctaatagata 5220

 gaactctcaa ttaatgtgga tttagtttct ctgttcgttt ttttttgttt tgttctcact 5280
 gtatttacct ttctatttag tatttagtta ttcatataat cttaacttgc ggtgtgaaat 5340
 accgcacaga tgcgtaagga gaaaataccg catcaggaaa ttgtaagcgt taatatitg 5400
 ttaaaattcg cgtaaaattt ttgttaaatc agtcatitit ttaaccaata ggccgaaatc 5460
 ggcaaaatcc ctataaatc aaaagaatag accgagatag ggttgagtgt tgttccagtt 5520
 tggacaaga gtccactatt aaagaacgtg gactccaacg tcaaaggcg aaaaaccgtc 5580
 tatcaggcg atggccact acgtgaacca tcaccctaata caagttitit ggggtcgagg 5640

 tgccgtaaag cactaaatcg gaaccctaaa gggagcccc gatttagagc ttgacgggga 5700
 aagcctcttt gaaaagataa tgtatgatta tgttttact catatitata cagaaacttg 5760
 atgttttctt tcgagtatat acaaggtgat tacatgtacg ttggaagtac aactctagat 5820

ttttagtagtc cctcttgggc tagcggtaaa ggtgcgcatt tttcacacc ctacaatgtt 5880
ctgttcaaaa gattttgggc aaacgtgta gaagtgaag ttggtgcga tgtttcggcg 5940
ttcgaaactt ctccgcagtg aaagataaat gatcg 5975

<210> 28

<211> 1594

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: Vector backbone fragment 1 for 3-piece assembly

<400> 28

gcaagagaag taacattaaa aatgaccact cttgacgaca cggcttaccg gtaccgcacc 60
agtgtcccgg gggacgccga ggccatcgag gcactggatg ggtccttcac caccgacacc 120
gtcttccgcg tcaccgccac cggggacggc ttcaccctgc gggaggtgcc ggtggacccg 180
ccctgacca aggtgttccc cgacgacgaa tcggacgacg aatcggacgc cggggaggac 240
ggcgaccggg actcccggac gttcgtcgcg tacggggacg acggcgacct ggccgggttc 300
gtggtcgtct cgtactccgg ctggaaccgc cggctgaccg tcgaggacat cgaggtcgcc 360

ccggagcacc gggggcacgg ggtcgggcgc gcgttgatgg ggctcgcgac ggagttcgcc 420
cgcgagcggg gcgccgggca cctctggctg gaggtcacca acgtcaacgc accggcgatc 480
cacgcgtacc ggccgatggg gttcacctc tgcggcctgg acaccgccct gtacgacggc 540
accgcctcgg acggcgagca ggccgtctac atgagcatgc cctgcccctg agtttaactt 600
gatactacta gattttttct cttcatTTat aaaatttttg gttataattg aagctttaga 660
agtatgaaaa aatccttttt tttcattctt tgcaacaaaa ataagaagct tcttttattc 720
attgaaatga tgaatataaa cctaacaaaa gaaaaagact cgaatatcaa acattaaaaa 780

aaaataaaag aggttatctg ttttccatt tagttggagt ttgcattttc taatagatag 840
aactctcaat taatgtggat ttagtttctc tgttcgtttt tttttgtttt gttctcactg 900
tatttacatt tctatttagt atttagttat tcatataatc ttaacttgcg gtgtgaaata 960
ccgcacagat gcgtaaggag aaaataccgc atcaggaaat tgtaagcggt aatattttgt 1020
taaaattcgc gttaaatttt tgtaaataca gtcatttttt taaccaatag gccgaaatcg 1080
gcaaaatccc ttataaatca aaagaataga ccgagatagg gttgagtgtt gttccagttt 1140
ggaacaagag tccactatta aagaacgtgg actccaacgt caaagggcga aaaaccgtct 1200

atcagggcga tgcccacta cgtgaacat caccctaate aagttttttg gggtcgaggt 1260
gccgtaaagc actaaatcgg aaccctaaag ggagcccccg atttagagct tgacggggaa 1320

agcctctttg aaaagataat gtatgattat gctttcactc atatttatac agaaacttga 1380
 tgttttcttt cgagtatata caaggtgatt acatgtacgt ttgaagtaca actctagatt 1440
 ttgtagtgcc ctcttgggct agcggtaaag gtgcgcattt tttcacaccc tacaatgttc 1500
 tgttcaaaag attttggica aacgctgtag aagtgaaggt tggcgcgcgt gtttcggcgt 1560
 tcgaaacttc tccgcagtga aagataaatg atcg 1594

<210> 29

<211> 4428

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: Vector backbone fragment 2 for 3-piece assembly

<400> 29

gttttagagc tagaaatagc aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt 60
 ggcaccgagt cggcgggtgct tttttgttt tttatgtctc agcttttgtt cccttttagtg 120
 agggtttaatt gcgcgcttgg cgtaatcatg gtcatactgt tttcctgtgt gaaattgtta 180
 tccgctcaca attccacaca acataggagc cggaagcata aagtgtaaag cctgggggtgc 240
 ctaatgagtg aggttaactca cattaattgc gttgcgtca ctgcccgtt tccagtcggg 300

aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag gcggtttgcg 360
 tattgggcgc tcttccgctt cctcgtcac tgactcgtg cgctcggtcg ttcggctgcg 420
 gcgagcggta tcagctcact caaaggcggc aatacgttta tccacagaat caggggataa 480
 cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc 540
 gttgtcggcg tttttcata ggctccgcc cctgacgag catcacaaaa atcgacgtc 600
 aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag 660
 ctccctcgtg cgctctcctg ttcgaccct gccgcttacc ggatacctgt ccgcctttct 720

cccttcggga agcgtggcg tttctcatag ctacgctgt aggtatctca gttcggtgta 780
 ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc gttcagcccg accgctgcg 840
 ctatccggt aactatcgtc ttgagtcaa cccggtgaaga cagacttat cgccactggc 900
 agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagttctt 960
 gaagtgggtg cctaactacg gctacactag aaggacagta tttggtatct gcgctctgct 1020
 gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc 1080
 tggtagcggg ggtttttttg tttgcaagca gcagattacg cgcagaaaaa aaggatctca 1140

agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa actcacgtta 1200

agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa 1260
 atgaagtttt aaatcaatct aaagtatata tgagtaaact tggctcgaca gttaccaatg 1320
 cttaatcagt gaggcacctt tctcagcgat ctgtctatct cgttcaccca tagttgcctg 1380
 actccccgtc gtgtagataa ctacgatacg ggagggcctt ccatctggcc ccagtgtctc 1440
 aatgataccg cgagaccac gctcaccggc tccagattta tcagcaataa accagccagc 1500
 cggaagggcc gagcgagaa gtggctcctg aactttatcc gcctccatcc agtctattaa 1560

 ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcga acgttgttgc 1620
 cattgtctaca ggcatcgttg tgtcacgctc gtcgttttgg atggcttcat tcagctccgg 1680
 ttcccaacga tcaaggcgag ttacatgac ccccatgttg tgcaaaaaag cggttagctc 1740
 ctccggctct ccgatcgttg tcagaagtaa gtggccgca gtgttatcac tcatggttat 1800
 ggagcactg cataattctc ttactgtcat gccatccgta agatgctttt ctgtgactgg 1860
 tgagtactca accaagtcac tctgagaata gtgtatgcgg cgaccgagtt gctcttgccc 1920
 ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc tcatcattgg 1980

 aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat ccagttcgat 2040
 glaaccact cgtgcacca actgatcttc agcatctttt actttacca gcgtttctgg 2100
 gtgagcaaaa acaggaaggc aaaatgccg aaaaaaggga ataaggcgca cacggaaatg 2160
 ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg gttattgtct 2220
 catgagcgga tacatatctg aatgtattta gaaaaataa caaatagggg ttccgcgcac 2280
 atttccccga aaagtccac ctgaacgaag catctgtgct tcattttgta gaacaaaaat 2340
 gcaacgcgag agcgctaatt tttcaacaa agaactctgag ctgcattttt acagaacaga 2400

 aatgcaacgc gaaagcgcta ttttaccac gaagaatctg tgcttcattt ttgtaaaaca 2460
 aaaatgcaac gcgagagcgc taatttttca aacaaagaat ctgagctgca tttttacaga 2520
 acagaaatgc aacgcgagag cgctatttta ccaacaaaga atctatactt cttttttgtt 2580
 ctacaaaaat gcatcccgag agcgctatct ttttaacaaa gcatcttaga ttactttttt 2640
 tctcctttgt gcgtctata atgcagtctc ttgataactt tttgcactgt aggtccgtta 2700
 aggttagaag aaggctactt tgggtgtctat tttctctcc ataaaaaag cctgactcca 2760
 cttcccgctt ttactgatta ctacgaagc tgcgggtgca ttttttcaag ataaaggcat 2820

 ccccgattat attctatacc gatgtggatt gcgcatactt tgtgaacaga aagtgatagc 2880
 gttgatgatt cttcatttgt cagaaaatta tgaacggttt cttctatttt gtctctatat 2940
 actacgtata ggaaatgttt acattttcgt attgttttcg attcactcta tgaatagttc 3000
 ttactacaat ttttttgtct aaagagtaat actagagata aacataaaaa atgtagaggt 3060

cgagtttaga tgcaagttca aggagcgaaa ggtggatggg taggttatat agggatatag 3120
cacagagata tatagcaaag agatactttt gagcaatggt tgtggaagcg gtattcgcaa 3180
tatttttagta gctcgttaca gtccggtgcg tttttggttt ttgaaagtg cgtcttcaga 3240

gcgcttttgg ttttcaaaag cgctctgaag ttctatact ttctagagaa taggaacttc 3300
ggaataggaa cttcaaagcg ttccgaaaa cgagcgcttc cgaaaatgca acgcgagctg 3360
cgcacataca gtcactgtt cacgtcgac ctatatctgc gtgttgctg tatatatata 3420
tacatgagaa gaacggcata gtgcgtgttt atgcttaaat gcgtacttat atgcgtctat 3480
ttatgtagga tgaaggttag tctagtacct cctgtgatat tatccattc catgcggggt 3540
atcgtatgct tccttcagca ctacccttta gctgttctat atgctgccac tcctcaattg 3600
gattagtctc atccttcaat gctatcattt cctttgatat tggatcatat taagaaacca 3660

ttattatcat gacattaacc tataaaaaata ggcgatcac gaggcccttt cgtctcgcg 3720
gtttcggtag tgacggtgaa aacctctgac acatgcagct cccggagacg gtcacagctt 3780
gtctgtaagc ggatgccggg agcagacaag cccgtcaggg cgctcagcg ggtgttgcg 3840
ggtgtcgggg ctggcttaac tatgcggcat cagagcagat tgtactgaga gtgcaaatac 3900
acatcatcgt cctacaagtt catcaaagt ttggacagac aactatacca gcatggatct 3960
cttgatcgg ttttttctc ccgctctctc gcaataacaa tgaacactgg gtcaatcata 4020
gcctacacag gtgaacagag tagcgtttat acagggttta tacggtgatt cctacggcaa 4080

aaatttttca ttctaaaaa aaaaaagaaa aatttttctt tccaacgcta gaaggaaaag 4140
aaaaatctaa ttaaatgat ttggtgattt tctgagagtt ccttttttca tatatcgaat 4200
tttgaatata aaaggagatc gaaaaaattt ttctattcaa tctgttttct ggttttattt 4260
gatagttttt ttgtgtatta ttattatgga ttagtactgg tttatatggg tttttctgta 4320
taacttcttt ttattttagt ttgtttaatc ttattttgag ttacattata gttccctaac 4380
tgcaagagaa gtaacattaa aaatgaccac tcttgacgac acggctta 4428

<210> 30

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: Gal80 genomic target sequence

<400> 30

taaggtgct gctgaacgt

19

<210> 31

<211> 19
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: H0 genomic target sequence
 <400> 31
 ccggcttgat cgactcaga 19
 <210> 32
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: ADH5 genomic target sequence
 <400> 32
 gggtcattgg tatcgatgg 19

 <210> 33
 <211> 38
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: Gal80 upstream primer
 <400> 33
 caaacggccg cctctgccat ggcaaagaat gctttcca 38
 <210> 34
 <211> 38
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: H0 upstream primer
 <400> 34
 acgtgtgtgt ctcatggaaa ttgatgcagt tgaagaca 38
 <210> 35
 <211> 33
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: ADH5 upstream primer
 <400> 35

ggcggttatat ccaaacattt cagacagaag att 33

<210> 36

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: Primer for deletion linker

<400> 36

aggtccgccg gcgttgacg agcg 24

<210> 37

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: TRS31 upstream primer

<400> 37

gtgcatttgg ctcgagttgc tg 22

<210> 38

<211>

> 35

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: TRS31 downstream primer

<400> 38

gggaagttat ctactatcat atattcattg tcacg 35

<210> 39

<211> 31

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: TRS31 heterology block primer

<400> 39

gaaaagtaga gattcagaat agatccttga c 31

<210> 40

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: CUE5 upstream primer

<400

> 40

ggaaggtatc aaggattctt ctctcc 26

<210> 41

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: CUE5 downstream primer

<400> 41

gaggtggcac atcttcatca tcttc 25

<210> 42

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: CUE5 heterology block primer

<400> 42

ccaataactc atcctgctcc aattgt 26

<210> 43

<211>

> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: ECM38 upstream primer

<400> 43

cagacgtgc agtaacacaa gc 22

<210> 44

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: ECM38 downstream primer

<400> 44

ctgaagtggg cagttccatg c 21

<210> 45

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: ECM38 heterology block primer

<400>

> 45

cagtgatctg gatcgtagaa gggc 24

<210> 46

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: PGD-1 upstream primer

<400> 46

ccaagagcat gccacggttg 20

<210> 47

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: PGD-1 downstream primer

<400> 47

gagttcccat agtactaccg c 21

<210> 48

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: PGD-1 heterology block primer

<400> 48

gcagacctta tctcttgtct cg 22

<210> 49

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: SMC6 upstream primer

<400> 49

gagctacttt cactgactgc gc 22

<210> 50

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: SMC6 downstream primer

<400> 50

gcgcttcaat agtagtacca tcagatg

27

<210> 51

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: SMC6 heterology block primer

<400> 51

gccgttctct gatctcaaag agaag

25

<210> 52

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: NT01 upstream primer

<400> 52

ctcagtatga catggatgaa caggatg

27

<210> 53

<211> 25

<

212> DNA

<213> Artificial Sequence

<220><223> Synthetic: NT01 downstream primer

<400> 53

ggtacctcct gtaagctccc ttttc

25

<210> 54

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: NT01 heterology block primer

<400> 54

gactgagacg ttctggactc cttc 24

<210> 55

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: DGA1 upstream primer

<400> 55

cttaaccaag cagcacgtg gtc 23

<210> 56

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: DGA1 downstream primer

<400> 56

gattccctag cgccaccaac 20

<210> 57

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: DGA1 heterology block primer

<400> 57

cctctccggt ggctggtgat ctg 23

<210> 58

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: ADH2 upstream primer

<400> 58

cgagactgat ctctctgcc ggaac 25

<210> 59

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: ADH2 downstream primer
 <400> 59
 gaatacttca ccaccgagcg ag 22
 <210> 60
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: ADH2 heterology block primer
 <400> 60
 gcatgtaagt ctgtatgaca tactcctg 28
 <210> 61

 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: SIN4 upstream primer
 <400> 61
 caaacgtcct aaatgaccca tcgttg 26
 <210> 62
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: SIN4 downstream primer
 <400> 62
 caacttcggg ttttgttggt ggtag 26
 <210> 63
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: SIN4 heterology block primer
 <
 400> 63
 caatggcaat ttaccgtagt tgaaaccg 28
 <210> 64
 <211> 25

<212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: CYS4 upstream primer
 <400> 64
 ctccagaatc acatattggt gttgc 25
 <210> 65
 <211> 29
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: CYS4 downstream primer
 <400> 65
 ccatcttagt aacgatatgg attggtttc 29
 <210> 66
 <211> 23

 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: CYS4 heterology block primer
 <400> 66
 ctgatggagt caggaaagat ggc 23
 <210> 67
 <211> 100
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: TRS31 donor DNA
 <400> 67
 aagcaagcga taacaagcga agcttcaaca acatacatc cgtaaggat ctattctgaa 60
 tctctacttt tcaaaaggca agaagcgtct ttatcagcaa 100
 <210> 68
 <211> 22
 <212> DNA
 <213> Artificial Sequence

 <220><223> Synthetic: TRS31 locus target sequence
 <400> 68

gactttgagt atattctaga gg 22

<210> 69

<211> 1080

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: TRS31 gRNA construct

<400> 69

gacggcacgg ccacgcgttt aaaccgcctt ataaatcaaa agaatagacc gagatagggt 60

tgagtgttgt tccagtttgg aacaagagtc cactattaaa gaacgtggac tccaacgtca 120

aagggcgaaa aaccgtctat cagggcgatg gccactacg tgaacatca ccctaatcaa 180

gttttttggg gtcgaggtgc cgtaaagcac taaatcggaa ccctaaaggg agcccccgat 240

ttagagcttg acggggaaag cctctttgaa aagataatgt atgattatgc tticactcat 300

atttatacag aaacttgatg ttttctttcg agtatataca aggtgattac atgtacgttt 360

gaagtacaac tctagatttt gtagtgcctt ctgggctag cggtaaaggt gcgcattttt 420

tcacacccta caatgttctg ttcaaaagat ttgggtcaaa cgctgtagaa gtgaaagttg 480

gtgcgcatgt ttcggcgttc gaaacttctc cgagtgaaa gataaatgat cggactttga 540

gtatatctta ggtttttagag ctagaatag caagttaaaa taaggctagt ccgttatcaa 600

cttgaaaaag tggcaccgag tcggtgggtgc tttttttgtt ttttatgtct cagcttttgt 660

tccttttagt gagggttaat tgccgcgttg gcgtaatcat ggtcatagct gtttctgtg 720

tgaaattgtt atccgctcac aattccacac aacataggag ccggaagcat aaagtgtaaa 780

gcctggggtg cctaatgagt gaggtaacac acattaattg cgttgcgctc actgcccgt 840

ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa tcggccaacg gcgggggaga 900

ggcggtttgc gtattgggcg ctcttccgct tcctcgctca ctgactcgct gcgctcggtc 960

gttcgctgc ggcgagcgtt atcagctcac tcaaaggcgg taatacggtt atccacagaa 1020

tcaggggata acgcaggaaa gaacatgtga gccggtgttt aaacccagc gcctggcggg 1080

<210> 70

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: CUE5 donor DNA

<400> 70

gtaagaaaga acccagaggc gccggaaga agacgacaaa cacaattgga gcaggatgag 60

ttattggcaa ggcaattaga tgagcaattc aacagctcac	100
<210> 71	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic: CUE5 locus target sequence	
<400> 71	
aagtttgtcg tcttcttgcc gg	22
<210> 72	
<211> 1080	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic: CUE5 gRNA construct	
<400> 72	
gacggcacgg ccacgcgttt aaaccgcctt ataaatcaaa agaatagacc gagatagggt	60
tgagtgttgt tccagtttgg aacaagagtc cactattaaa gaacgtggac tccaacgtca	120
aagggcgaaa aaccgtctat cagggcgatg gccactacg tgaaccatca ccctaataca	180
gttttttggg gtcgaggatg cgtaaagcac taaatcggaa ccctaaaggg agccccgat	240
ttagagcttg acggggaaa cctctttgaa aagataatgt atgattatgc tttcactcat	300
atttatacag aaacttgatg ttttctttcg agtatataca aggtgattac atgtacgttt	360
gaagtacaac tctagatttt gtagtgcctt cttgggctag cggtaaaggt gcgcattttt	420
tcacacccta caatgttctg ttcaaaagat tttgggtcaaa cgctgtagaa gtgaaagttg	480
gtgcgcatgt ttcggcgctc gaaacttctc cgcagtgaat gataaatgat cgaagtttgt	540
cgtcttcttg cgttttagag ctagaatag caagttaaaa taaggctagt ccgttatcaa	600
cttgaaaaag tggcaccgag tcgggtggtgc tttttttgtt ttttatgtct cagcttttgt	660
tccctttagt gagggttaat tgcgcgcttg gcgtaatcat ggtcatagct gtttctgtg	720
tgaaattgtt atccgctcac aattccacac aacataggag ccggaagcat aaagtgtaaa	780
gcctgggggtg cctaattagt gaggttaact acattaattg cgttgcgctc actgcccgt	840
ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcggggaga	900
ggcggtttgc gtattgggag ctcttccgct tcctcgtcga ctgactcgtc gcgtcggtc	960
gttcggctgc ggcgagcgg atcagctcac tcaaaggcgg taatacggtt atccacagaa	1020
tcaggggata acgcaggaaa gaacatgtga gccggtgttt aaaccccagc gcctggcggg	1080

<210> 73
 <211> 100
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: ECM38 donor DNA
 <400> 73
 tctaacaatgc taatggaact ggctaaaaat ggttccgtgg cgcccttcta cgatccagat 60

 cactggattg caaaatccat gattgatact gtcgcaaaat 100
 <210> 74
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: ECM38 locus target sequence
 <400> 74
 tgatcgaggt cataaaatgg gg 22
 <210> 75
 <211> 1080
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: ECM38 gRNA construct
 <400> 75
 gacggcacgg ccacgcgttt aaaccgcctt ataaatcaaa agaatagacc gagatagggt 60
 tgagtgttgt tccagtttgg aacaagagtc cactattaaa gaacgtggac tccaacgtca 120

 aagggcgaaa aaccgtctat cagggcgatg gccactacg tgaacatca ccctaatcaa 180
 gttttttggg gtcgagggtc cgtaaagcac taaatcggaa ccctaaaggg agcccccgat 240
 ttagagcttg acggggaaag cctctttgaa aagataatgt atgattatgc ttctactcat 300
 atttatacag aaacttgatg ttttctttcg agtatataca aggtgattac atgtacgttt 360
 gaagtacaac tctagatttt gtagtgccct cttgggctag cggtaaaggt gcgcattttt 420
 tcacacccta caatgttctg ttcaaaagat ttgggtcaaa cgctgtagaa gtgaaagtgtg 480
 gtgcgcgatgt ttcggcgctc gaaacttttc cgcagtgaag gataaatgat cgtgatcgag 540

 gtcataaaat ggtttttagag ctagaatag caagttaaaa taaggctagt ccgttatcaa 600
 cttgaaaaag tggcaccgag tcggtggtgc tttttttgtt ttttatgtct cagcttttgt 660

tccctttagt gagggttaat tgcgcgcttg gcgtaatcat ggtcatagct gtttcctgtg 720
 tgaaattgtt atccgctcac aattccacac aacataggag ccggaagcat aaagtgtaaa 780
 gcctggggtg cctaagtagt gaggtaacac acattaattg cgttgcgctc actgcccgt 840
 ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcggggaga 900
 ggcggtttgc gtattgggcg ctcttccgct tcctcgtca ctgactcgct gcgctcggtc 960

gttcggctgc ggcgagcggg atcagctcac tcaaaggcgg taatacggtt atccacagaa 1020
 tcaggggata acgcaggaaa gaacatgtga gccggtgttt aaaccccagc gcctggcggg 1080

<210> 76
 <211> 100
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: PGD-1 donor DNA
 <400> 76
 ggcgtcaagc tagatgatit acaagtata ttggctaaaa atgaaaacga gacaagagat 60
 aaggtctgca agcagatcaa cgaagcacgc gatgaaattc 100

<210> 77
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: PGD-1 locus target sequence

<400> 77
 aagaacgaga atgaaaccag gg 22

<210> 78
 <211> 1080
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: PGD-1 gRNA construct
 <400> 78
 gacggcacgg ccacgcgttt aaaccgcctt ataaatcaaa agaataagacc gagatagggt 60
 tgagtgttgt tccagtttgg aacaagagtc cactattaaa gaacgtggac tccaacgtca 120
 aagggcgaaa aaccgtctat caggcgcatg gccactacg tgaacatca ccctaatcaa 180
 gttttttggg gtcgagggtc cgtaaagcac taaatcgga ccctaaaggg agccccgat 240

ttagagcttg acggggaaag cctctttgaa aagataatgt atgattatgc tticactcat 300
 atttatacag aaacttgatg ttttctttcg agtataatac aggtgattac atgtacgttt 360
 gaagtacaac tctagatttt gtagtgcctt cttgggctag cggtaaaggt gcgcattttt 420
 tcacacccta caatgttctg ttcaaaagat ttgggtcaaa cgctgtagaa gtgaaagtgt 480
 gtgcgcatgt ttcggcgctc gaaacttctc cgcagtgaac gataaatgat cgaagaacga 540
 gaatgaaacc agtttttagag ctagaatag caagttaaaa taaggctagt ccgttatcaa 600
 cttgaaaaag tggcaccgag tcggtggtgc tttttttgtt ttttatgtct cagcttttgt 660

tccttttagt gagggttaat tgcgcgcttg gcgtaatcat ggtcatagct gtttctgtg 720
 tgaaattgtt atccgctcac aattccacac aacataggag ccggaagcat aaagtgtaaa 780
 gcctgggggtg cctaattgag gaggttaactc acattaattg cggtgcgctc actgcccgt 840
 ttccagtcgg gaaacctgct gtgccagctg cattaatgaa tcggccaacg gcgggggaga 900
 ggcggtttgc gtattgggct ctttccgct tcctcgctca ctgactcgct gcgctcggtc 960
 gttcggtgc ggcgagcgg atcagctcac tcaaaggcgg taatacgggt atccacagaa 1020
 tcaggggata acgcaggaaa gaacatgtga gccggtgtt aaaccccagc gcctggcggg 1080

<210> 79

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: SMC6 donor DNA

<400> 79

cgaaattata gtagaaagga taattaagag agatgggccc gcttcattct ctttgagatc 60
 agagaacggc aaggaaatta gcaataagaa aaaggatatt 100

<210> 80

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: SMC6 locus target sequence

<400> 80

gcttctttta gattaaggctc gg 22

<210> 81

<211> 1080

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: SMC6 gRNA construct

<400> 81

gacggcacgg ccacgcgttt aaaccgcctt ataaatcaaa agaatagacc gagatagggt	60
tgagtgttgt tccagtttgg aacaagagtc cactattaaa gaacgtggac tccaacgtca	120
aagggcgaaa aaccgtctat cagggcgatg gccactacg tgaacatca ccctaatcaa	180
gttttttggg gtcgagggtc cgtaaagcac taaatcggaa ccctaaaggg agccccgat	240
ttagagcttg acggggaaag cctctttgaa aagataatgt atgattatgc ttctactcat	300
atttatacag aaacttgatg ttttctttcg agtatataca aggtgattac atgtacgttt	360
gaagtacaac tctagatttt gtagtgcctt cttgggctag cggtaaaggt gcgcattttt	420
tcacacccta caatgttctg ttcaaaagat ttgggtcaaa cgctgtagaa gtgaaagttg	480
gtgcgcatgt ttcggcgttc gaaacttctc cgcagtgaag gataaatgat cggcttcttt	540
tagattaagg tgtttttagg ctagaatag caagttaaaa taaggctagt ccgttatcaa	600
cttgaanaag tggcaccgag tcgggtggtc tttttttgtt ttttatgtct cagcttttgt	660
tccctttagt gagggttaat tgcgcgcttg gcgtaatcat ggcatagct gtttctgtg	720
tgaaattgtt atccgctcac aattccacac aacataggag ccggaagcat aaagtgtaaa	780
gcctgggggtg cctaatgagt gaggtaaact acattaattg cgttgcgctc actgcccgct	840
ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcggggaga	900
ggcgggtttgc gtattgggag ctcttccgct tcctcgctca ctgactcgct gcgctcggtc	960
gttcggctgc ggcgagcggg atcagctcac tcaaaggcgg taatacggtt atccacagaa	1020
tcaggggata acgcaggaaa gaacatgtga gccggtgttt aaacccagc gcctggcggg	1080

<210> 82

<211> 99

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: NT01 donor DNA

<400> 82

atacttttca aacttacatt atatggaacc tatagaagga gtccagaacg tctcagttctc	60
taggtggaaa cttaactgct acatctgtaa aaagaagat	99

<210> 83

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: NT01 locus target sequence

<400> 83

aaatgttagc gtatcacgat gg 22

<210> 84

<211> 1080

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: NT01 gRNA construct

<400> 84

gacggcacgg ccacgcgttt aaaccgcctt ataaatcaaa agaatagacc gagatagggt 60

tgagtgttgt tccagtttgg aacaagagtc cactattaaa gaacgtggac tccaacgtca 120

aaggcgcaaa aaccgtctat caggcgcatg gccactacg tgaacatca ccctaatcaa 180

gttttttggg gtcgagggtc cgtaaagcac taaatcggaa ccctaaaggg agccccgat 240

ttagagcttg acggggaaag cctctttgaa aagataatgt atgattatgc ttctactcat 300

atttatacag aaacttgatg ttttctttcg agtatataca aggtgattac atgtacgttt 360

gaagtacaac tctagatttt gtagtgccct ctgggctag cggtaaaggt gcgcattttt 420

tcacacccta caatgttctg ttcaaaagat ttgggtcaaa cgctgtagaa gtgaaagttg 480

gtgcgcatgt ttcggcgttc gaaacttctc gcgagtgaat gataaatgat cgaaatgtta 540

gcgtatcacg agtttttagag ctagaaatag caagttaaaa taaggctagt ccgttatcaa 600

cttgaaaaag tggcaccgag tcggtggtgc tttttttgtt ttttatgtct cagcttttgt 660

tccctttagt gagggttaat tgcgcgcttg gcgtaatcat ggtcatagct gtttctgtg 720

tgaaattgtt atccgctcac aattccacac aacataggag ccggaagcat aaagtgtaaa 780

gcctgggggt cctaatgagt gaggttaact acattaattg cgttgcgctc actgcccgt 840

ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa tcggccaacg gcgggggaga 900

ggcggtttgc gtattgggcg ctcttcgctt tcctcgctca ctgactcgct gcgctcggtc 960

gttcggtctg ggcgagcgg atcagctcac tcaaaggcgg taatacgggt atccacagaa 1020

tcaggggata acgcaggaaa gaacatgtga gccggtgttt aaacccagc gcctggcggg 1080

<210> 85

<211> 99

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: DGA1 donor DNA

<400> 85

gggttcttgc tattccatat atgatttatt ttttttttga cagatcacca gccaccggag 60

aggttggtcaa tcgatactct ctctcgatttc gttcattgc 99

<210> 86

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: DGA1 locus target sequence

<400> 86

tcgataggtc tcttgcact gg 22

<210> 87

<211> 1080

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: DGA1 gRNA construct

<400> 87

gacggcacgg ccacgcgttt aaaccgcctt ataaatcaaa agaataagacc gagatagggt 60

tgagtgttgt tccagtttgg aacaagagtc cactattaaa gaacgtggac tccaacgtca 120

aagggcgaaa aaccgtctat cagggcgatg gccactacg tgaacatca ccctaataca 180

gttttttggg gtcgagggtgc cgtaaagcac taaatcggaa ccctaaaggg agcccccgat 240

ttagagcttg acggggaaaag cctctttgaa aagataatgt atgattatgc tttcactcat 300

atttatacag aaacttgatg ttttctttcg agtatataca aggtgattac atgtacgttt 360

gaagtacaac tctagatttt gtagtgcctt ctgggctag cggtaaaggt gcgcattttt 420

tcacacccta caatgttctg ttcaaaagat tttggtcaaa cgctgtagaa gtgaaagtgt 480

gtgcgcatgt ttcggcggtc gaaacttctc cgcagtgaat gataaatgat cgtcgatagg 540

tctctgcaa cgtttttagag ctagaatatag caagttaaaa taaggctagt ccgttatcaa 600

cttgaanaag tggcaccgag tcggtgggtgc tttttttgtt ttttatgtct cagcttttgt 660

tccctttagt gagggttaat tgcgcgcttg gcgtaatcat ggtcatagct gtttcctgtg 720

tgaaattggt atccgctcac aattccacac aacataggag ccggaagcat aaagtgtaaa 780

gcctggggtg cctaattgagt gaggttaact acattaattg cgttgcgctc actgcccgt 840

ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcggggaga 900

ggcgggtttgc gtattgggcg ctcttcgct tctcgcctca ctgactcgct gcgctcggtc 960
gttcggctgc ggcgagcggt atcagctcac tcaaaggcgg taatacggtt atccacagaa 1020
tcaggggata acgcaggaaa gaacatgtga gccggtgttt aaaccccagc gcctggcggg 1080

<210> 88

<211> 929

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: ADH2 donor DNA

<400> 88

ccgggcatct ccaacttata agttggagaa ataagagaat ttcagattga gagaatgaaa 60
aaaaaaaaa aaaaaaaggc agaggagagc ataaaaatgg gggttcacttt ttggtaaagc 120
tatagcatgc ctatcacata taaatagagt gccagtagcg acttttttca cactcgaaat 180
actcttacta ctgctctctt gttgttttta tcacttcttg tttcttcttg gtaaatagaa 240
tatcaagcta caaaaagcat acaatcaact atcaactatt aactatatcg taatacacia 300
tgtctattcc agaaactcaa aaagccattt tcttctacga atccaacggc aagttggagc 360

ataaggatat ccagttcca aagccaaagc ccaacgaatt gttaatcaac gtcaaatatt 420
caggagtatg tcatacagac ttacatgcat ggcatggtga ctggccattg ccaactaagt 480
taccattagt tgggtgggac gaaggtgctg gggtcgttgt tgccatgggt gaaaacgtta 540
agggtctgaa gatcggtgac tacgccggtt ttaaatggtt gaacagttca tgtatggcct 600
gcgaatactg tgaattgggt aacgaatcca actgtctca cgctgacttg tctggttaca 660
cccatgacgg ttctttccaa caatacgtca ctgctgacgc ggttcaagcc gtcacattc 720
ctcaaggtag tgacttggtt gaagtcgcc ccatcttggt tgctggtatc accgtataca 780

aggctttgaa gtctgccaac ttgagagcag gccactgggt ggccatttct ggtgctgctg 840
gtggtctagg ttctttggtt gttcaatatg ctaaggcgat gggttacaga gtcttaggta 900
ttgatggtgg tccaggaaag gaagaattg 929

<210> 89

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: ADH2 locus target sequence

<400> 89

taatcaacgt caaatacttt gg 22

<210> 90

<211> 1080

<212> DNA

<213> Artificial Sequence

<220><223

> Synthetic: ADH2 gRNA construct

<400> 90

```
gacggcacgg ccacgcgttt aaaccgcctt ataaatcaaa agaatagacc gagatagggt      60
tgagtgttgt tccagtttgg aacaagagtc cactattaaa gaacgtggac tccaacgtca      120
aagggcgaaa aaccgtctat cagggcgatg gccactacg tgaacatca ccctaatcaa      180
gttttttggg gtcgagggtc cgtaaagcac taaatcggaa ccctaaaggg agccccgat      240
ttagagcttg acggggaaag cctctttgaa aagataatgt atgattatgc tttcactcat      300
atttatacag aaacttgatg ttttctttcg agtatataca aggtgattac atgtacgttt      360
gaagtacaac tctagatttt gtagtgcctt cttgggctag cggtaaaggt gcgcattttt      420
```

```
tcacacccta caatgttctg ttcaaaagat ttgggtcaaa cgctgtagaa gtgaaagttg      480
gtgcgcatgt ttcggcgctc gaaacttctc cgcagtgaag gataaatgat cgtaatcaac      540
gtcaaatact tgtttttagag ctagaatag caagttaaaa taaggctagt ccgttatcaa      600
cttgaaaaag tggcaccgag tcggtgggtc tttttttgtt ttttatgtct cagcttttgt      660
tccctttagt gagggttaat tgcgcgcttg gcgtaatcat ggtcatagct gtttctgtg      720
tgaaattgtt atccgctcac aattccacac aacataggag ccggaagcat aaagtgtaaa      780
gcctgggggtg cctaattagt gaggtaactc acattaattg cgttgcgctc actgcccgt      840
```

```
ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcggggaga      900
ggcgggtttg gtattgggag ctcttccgct tcctcgtcga ctgactcgtc gcgctcggtc      960
gttcggctgc ggcgagcggg atcagctcac tcaaaggcgg taatcgggtt atccacagaa      1020
tcaggggata acgcaggaaa gaacatgtga gccggtgttt aaacccacgc gcctggcggg      1080
```

<210> 91

<211> 979

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: SIN4 donor DNA

<400> 91

```
gcactacttt agatccaacg gatgatgaag ggcatgtgtt gaagttagaa aacctgcatg      60
```

ttgtgtctaa atcatcaatt gagaaagatc cctccccaga aattttagtt ttgtataatg	120
tttgcgatac atcaaaatca ttggtgaaaa gatatagact agctccaaca caactttcag	180
ctgagtatTTT ggtgatattg aaaccagatt taaatatcga tcgaaataat tctacgaatc	240
aaatatTTCCA atcacgccgt tacaaccttc gtcgccacag tgatatcgTT ctcgacaaaa	300
aagtcacctt gataactTct gagatgttcg atggattTgt ttcattttac tttgaagatg	360
gtaccatcga atcctacaat caaaatgatt ggaagttgga gactgaacga cttataggtc	420
agagccaatt aggaaaattc aaaaatatta ttgcatcccc actaagcgcc ggTTTcaact	480
acggtaaatt gccattgcc ccatcggttg agtggatgaa ggTTTcacca tctatgtgtg	540
gggttatcgt taaacagtat aataaaaaat ggccacagtt ttatgctgct gttcaaaaaa	600
actacgcaga tcttgagaag gattctataa atgctacagc tttggcattt ggatacgtga	660
aaagtttaca taagcaaata tctgccgaag acttgaccat cgctgcgaag acgcacatac	720
tcagaatttc attcttggat agaaaaaggg ccaaagaatt tattacgact ttgttgaaaa	780
gcctatactc gtttttcaac atttctctg atgcgcctaa agagattatg gataaaataa	840
taacaagtag gccacttcaa aaaattatgc tattgcagct ggaacttggc agttgttttt	900
cgcaagaaaa tattgaagaa atggcgagag ttatactcta cttgaaaaat gttttgtttg	960
cgttcaatgg cgtggcgag	979
<210> 92	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic: SIN4 locus target sequence	
<400> 92	
actcaaccga tggaggtaaa gg	22
<210> 93	
<211> 1080	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic: SIN4 gRNA construct	
<400> 93	
gacggcacgg ccacgcgttt aaaccgcctt ataaatcaaa agaatagacc gagatagggt	60
tgagtgttgt tccagtttgg aacaagagtc cactattaaa gaacgtggac tccaacgtca	120

aaggcgcaaa aaccgtctat caggcgcatg gccactacg tgaaccatca ccctaataca 180
gttttttggg gtcgagggtc cgtaaagcac taaatcggaa ccctaaaggg agccccgat 240
ttagagcttg acggggaaag cctctttgaa aagataatgt atgattatgc tttcactcat 300
attatacag aaacttgatg ttttcttcg agtataatac aggtgattac atgtacgttt 360
gaagtacaac tctagatttt gtagtgcctt ctgggctag cggtaaaggt gcgcattttt 420
tcacacccta caatgttctg ttcaaaagat ttgggtcaaa cgctgtagaa gtgaaagtgt 480
gtgcgcgtgt ttcggcgttc gaaacttctc cgcagtgaag gataaatgat cgactcaacc 540

gatggaggta agtttttagag ctagaatag caagttaaaa taaggctagt ccgttatcaa 600
cttgaaaaag tggcaccgag tcggtgggtc tttttttgtt ttttatgtct cagcttttgt 660
tccctttagt gagggttaat tgcgcgttg gcgtaatcat ggtcatagct gtttcctgtg 720
tgaaattgtt atccgctcac aattccacac aacataggag ccggaagcat aaagtgtaaa 780
gcctgggggtg cctaattgagt gaggttaact acattaattg cgttgcgctc actgcccgtt 840
ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcggggaga 900
ggcggtttgc gtattggggtg ctcttcctgt tcctcgtcga ctgactcgtt gcgctcgttc 960

gttcggctgc ggcgagcgtt atcagctcac tcaaaggcgg taatacggtt atccacagaa 1020
tcaggggata acgcaggaaa gaacatgtga gccggtgttt aaacccacg gcctggcggg 1080

<210> 94

<211> 1038

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: CYS4 donor DNA

<400> 94

gttggaaaaa gagattcctg gtgcgttat acttgaccaa tataacaata tgatgaaccc 60
agaagctcat tacttttgta ctggctcgca aatccaaaga cagctagaag acttgaattt 120
atttgataat ctacgcgtg ttgttctggt tgcgtgact ggtgggacta ttagcggat 180
ttccaagtac ttgaaagaac agaatgataa gatccaaatc gttggtgctg acccattcgg 240

ttcaatttta gcccaacctg aaaacttgaa taagactgat atcactgact acaaagtga 300
gggtattggt taigattttg ttcttcaggt ttggacaga aaattaattg atgttttgta 360
taagacagac gacaagcctt ctttcaaata cgccagacaa ttgatttcta acgaaggtgt 420
cttggtgggt ggttcttcgt gttctgcctt cactgcggtt gtgaaatact gtgaagacca 480
ccctgaactg actgaagatg atgtcatcgt cgccatcttt cctgactcca tcaggtcgta 540
cctaaccaaa ttcgtcgatg acgaatggtt gaaaaagaac aatttgtggg atgatgacgt 600

gttggcccggt ttgactctt caaagctgga ggcttcgacg acaaaatagc ctgatgtgtt 660

tggtaacgct actgtaaagg atcttcactt gaaaccggtt gtttccgtta aggaaaccgc 720

taaggctact gatgttatca agatattaaa agacaatggc tttagaccaat tgcctgtgtt 780

gactgaagac ggcaagtgtg ctggtttagt tactctctct gagcttctaa gaaaactatc 840

aatcaataat tcaacaacg acaacactat aaagggtaaa tacttggact tcaagaaatt 900

aaacaatttc aatgatgttt cctcttaca cgaataaaa tccggtaga agaagtttat 960

taaattcgat gaaaactcaa agctatctga ctggaatcgt tctttgaaa aaaactcatc 1020

tgccgttatc actgatgg 1038

<210> 95

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: CYS4 locus target sequence

<400> 95

ggtacgacct gatggaatct gg 22

<210> 96

<211> 1080

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: CYS4 gRNA construct

<400> 96

gacggcacgg ccacgcgttt aaaccgcctt ataaatcaaa agaataagacc gagatagggt 60

tgagtgttgt tccagtttgg aacaagagtc cactattaaa gaacgtggac tccaacgtca 120

aagggcgaaa aaccgtctat cagggcgatg gccactacg tgaacatca ccctaatcaa 180

gttttttggg gtcgaggtag cgtaaagcac taaatcggaa ccctaaagg agccccgat 240

ttagagcttg acggggaaa cctctttgaa aagataatgt atgattatgc ttctactcat 300

atttatacag aaacttgatg ttttctttcg agtatataca aggtgattac atgtacgttt 360

gaagtacaac tctagatttt gtagtgccct cttgggctag cggtaaagg gcgcattttt 420

tcacacccta caatgttctg ttcaaaagat ttgggtcaaa cgctgtagaa gtgaaagtgt 480

gtgcgcgatg ttcggcgctt gaaacttctc cgcagtgaat gataaatgat cgggtacgac 540

ctgatggaat cgttttagag ctagaatag caagttaaaa taaggctagt ccgttatcaa 600

cttagaaaag tggcaccgag tcggtgggtgc tttttttgtt ttttatgtct cagcttttgt 660
 tcccttttagt gagggttaat tgcgcgcttg gcgtaatcat ggtcatagct gtttcctgtg 720
 tgaaattgtt atccgctcac aattccacac aacataggag ccggaagcat aaagtgtaaa 780
 gcctgggggtg cctaattagt gaggttaact acattaattg cgttgcgctc actgcccgt 840
 ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcggggaga 900
 ggcggtttgc gtattggcg cttttccgt tctcgtcctc ctgactcgt gcgctcggtc 960
 gtccggtgc ggcgagcgg atcagctcac tcaaaggcgg taatcgggt atccacagaa 1020

 tcaggggata acgcaggaaa gaacatgtga gccggtgttt aaaccccagc gcctggcggg 1080
 <210> 97
 <211> 8045
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: pAM3472
 <400> 97

 gttagcggtt ttgcgtgct tcgcgatgia cgggccagat atacgcgttg acattgatta 60
 ttgactagtt attaatagta atcaattacg gggtcattag ttcatagccc atatatggag 120
 ttccgcttta cataacttac ggtaaatggc ccgcctggct gaccgcccac cgacccccgc 180
 ccattgacgt caataatgac gtatgttccc atagtaacgc caataggac tttcattga 240
 cgtcaatggg tggagtattt acggtaaact gccacttgg cagtacatca agtgtatcat 300

 atgccaagta cgccccctat tgacgtcaat gacggtaaat ggcccgcctg gcattatgcc 360
 cagtacatga cttatggga ctttctact tggcagtaca tctacgtatt agtcacgct 420
 attaccatgg tgatgcggtt ttggcagtac atcaatgggc gtggatagcg gtttgactca 480
 cggggatttc caagtctcca cccattgac gtcaatggga gtttgttttg gcacaaaat 540
 caacgggact ttccaaaat tcgtaacaac tccgccccat tgacgcaaat gggcggtagg 600
 cgtgtacggt gggaggtcta tataagcaga gctcgtttag tgaaccgtca gatcgctgg 660
 agacgccatc cagctgttt tgacctccat agaagacacc gggaccgatc cagcctccgg 720

 agcggccgcc accatgggca agcccatccc taacccctg ttggggctgg acagcaccgc 780
 tcccaaaaag aaaaggaagg tgggcattca cggcgtgcct gcggccgaca aaaagtacag 840
 catcgccctt gatatcgga ccaatagcgt gggtgggccc gttatcacag acgaatacaa 900
 ggtaccagc aagaagtta aggtgctggg gaatacagac aggcactcta tcaagaaaaa 960
 cttatcggg gctctgctgt ttgactcagg cgagaccgcc gaggccacca ggttgaagag 1020
 gaccgcaagg cgaaggtaca cccggaggaa gaacaggatc tgctatctgc aggagatctt 1080

cagcaacgag atggccaagg tggacgacag cttcttcac aggctggagg agagcttcct	1140
tgtcaggag gataagaagc acgaacgaca ccccatcttc ggcaacatag tcgacgaggt	1200
cgcttatcac gagaagtacc ccacatcta ccacctgca aagaaattgg tggatagcac	1260
cgataaagcc gacttgcgac ttatctactt ggctctggcg cacatgatta agttcagggg	1320
ccacttctg atcgagggcg accttaacct cgacaacagt gacgtagaca aattgttcat	1380
ccagcttgta cagacctata accagctgtt cgaggaaaac cctattaacg ccagcggggt	1440
ggatgcgaag gccatactta gcgccaggct gagcaaaagc aggcgcttgg agaacctgat	1500
agcccagctg cccggtgaaa agaagaacgg cctcttcggt aatctgattg ccctgagcct	1560
gggcctgacc cccaacttca agagcaactt cgacctggca gaagatgcca agctgcagtt	1620
gagtaaggac acctatgacg acgacttga caatctgctc gcccaaatcg gcgaccagta	1680
cgctgacctg ttctcgcgg ccaagaacct ttctgacgca atctgctta gcgatatcct	1740
tagggatgaac acagagatca ccaaggcccc cctgagcgcc agcatgatca agaggtaaga	1800
cgagcaccat caggacctga ccttctgaa ggccctgggtg aggcagcaac tgcccagaaa	1860
gtacaaggag atctttttcg accagagcaa gaacggctac gccggctaca tcgacggcgg	1920
agccagccaa gaggagtctt acaagttcat caagcccatc ctggagaaga tggatggcac	1980
cgaggagctg ctggtgaagc tgaacaggga agatttgcct cggaagcaga ggacctttga	2040
caacggtagc atccccacc agatccacct gggcgagctg cacgcaatac tgaggcgaca	2100
ggaggatttc tacccttcc tcaaggacaa tagggagaaa atcgaaaaga ttctgacctt	2160
caggatcccc tactacgtgg gccctcttgc caggggcaac agccgattcg ctgggatgac	2220
aagaaagagc gaggagacca tcaccccttg gaacttcgag gaagtgggtg acaaaggagc	2280
aagcgcgcag tctttcatcg aacggatgac caatttcgac aaaaacctgc ctaacgagaa	2340
ggtgctgccc aagcacagcc tgctttacga gtacttcacc gtgtacaac agctcaccaa	2400
ggtgaaatat gtgaccgagg gcatgcgaaa acccgcttct ctgagcggcg agcagaagaa	2460
ggccatcgtg gacctgctgt tcaagaccaa caggaagggtg accgtgaagc agctgaagga	2520
ggactacttc aagaagatcg agtgctttga tagcgtggaa ataagcggcg tggaggacag	2580
gttcaacgcc agcctgggca cctaccacga ctgtttgaag ataatcaaag acaaggattt	2640
cctggataat gaggagaacg aggatatact cgaggacatc gtgctgactt tgacctgtt	2700
tgaggaccga gagatgattg aagaaaggct caaaacctac gccacctgt tcgacgacaa	2760
agtgatgaaa caactgaaga gacgaagata caccggtctg gccagactgt ccaggaagct	2820

catcaacggc attagggaca agcagagcgg caagaccatc ctggatttcc tgaagtcga	2880
cggcttcgcc aaccgaaact tcatgcagct gattcacgat gacagcttga ctttcaagga	2940
ggacatccag aaggcccagg ttagcggcca gggcgactcc ctgcacgaac atattgcaaa	3000
cctggcaggc tcccctgcga tcaagaaggg catactgcag accgttaagg ttgtggacga	3060
attgttcaag gtcatgggca ggcacaagcc cgaaaacata gttatagaga tggccagaga	3120
gaaccagacc acccaaaaagg gccagaagaa cagccgggag cgcatgaaaa ggatcgagga	3180
gggtatcaag gaactcggaa gccagatcct caaagagcac cccgtggaga ataccagct	3240
ccagaacgag aagctgtacc tgtactacct gcagaacggc agggacatgt acgttgacca	3300
ggagttagac atcaacaggc tttagactia tgacgtggat cacatagtgc cccagagctt	3360
tcttaaagac gatagcatcg acaacaaggt cctgaccgc tccgacaaaa acaggggcaa	3420
aagcgacaac gtccaacg aagaggttgt taaaaagatg aagaactact ggaggcaact	3480
gtcaacgcg aaattgatca cccagagaaa gttcgataac ctgaccaagg ccgagagggg	3540
cggactctcc gaacttgaca aagcgggctt cataaagagg cagctggtcg agaccgaca	3600
gatcacgaag cacgtggccc aaatcctcga cagcagaatg aataccaagt acgatgagaa	3660
tgacaaactc atcagggag tgaaagtgat taccctgaag agcaagttgg tgtccgactt	3720
tcgcaaagat ttccagttct acaaggtgag ggagatcaac aactaccacc atgccacga	3780
cgcatacctg aacgcctgg tcggcaccgc cctgattaag aagtatcaa agctggagtc	3840
cgaatttgct tacggcgact acaaagttta cgatgtgagg aagatgatcg ctaagagcga	3900
acaggagatc ggcaaggcca ccgctaagta tttcttctac agcaacatca tgaacttttt	3960
caagaccgag atcacacttg ccaacggcga aatcaggaag aggccgctta tcgagaccaa	4020
cggtagagacc ggcgagatcg tgtgggacaa gggcagggac ttcgccaccg tgaggaaagt	4080
cctgagcatg ccccaggtga atattgtgaa aaaaactgag gtgcagacag gcggctttag	4140
caaggaatcc atcctgccca agaggaacag cgacaagctg atcgcccga agaaggactg	4200
ggaccctaag aagtatggag gcttcgacag cccaccgta gcctacagcg tgctgttgt	4260
cgcaaggta gagaagggga agagcaagaa actgaagagc gtgaaggagc tgctcggcat	4320
aaccatcatg gagaggtcca gctttgagaa gaacccatt gactttttgg aagccaaggg	4380
ctacaaagag gtcaaaaagg acctgatcat caaactcccc aagtactccc tgtttgaatt	4440
ggagaacggc agaaagagga tgctggcgag cgctggggaa ctgcaaaagg gcaacgaact	4500
ggcgctgccc agcaagtacg tgaattttct gtacctggcg tcccactacg aaaagctgaa	4560
aggcagcccc gaggacaacg agcagaagca gctgttcgtg gagcagcaca agcattacct	4620
ggacgagata atcgagcaaa tcagcgagtt cagcaagagg gtgattctgg ccgacgcgaa	4680

cctggataag gtccctcagcg cctacaacaa gcaccgagac aaacccatca gggagcaggc 4740
 cgagaatatc atacacctgt tcacctgac aaatctgggc gcacctgcgg cattcaaata 4800
 ctctgatacc accatcgaca ggaaaaggta cactagcact aaggaggtgc tggatgccac 4860
 ctctgatccac cagtccatta ccggcctgtg tgagaccagg atcgacctga gccagcttgg 4920

 aggcgactct agggcggacc caaaaaagaa aaggaaggtg gaattctcta gaggcagtgg 4980
 agagggcaga ggaagtctgc taacatgcgg tgacgtcgag gagaatcctg gcccaatgaa 5040
 ccggggagtc ctttttaggc acttgcttct ggtgctgcaa ctggcgctcc tcccagcagc 5100
 cactcaggga aagaaagtgg tgctgggcaa aaaaggggat acagtggaac tgacctgtac 5160
 agcttcccag aagaagagca tacaattcca ctggaaaaac tccaaccaga taaagattct 5220
 gggaaatcag ggctccttct taactaaagg tccatccaag ctgaatgac gcgctgactc 5280
 aagaagaagc ctttgggacc aaggaaactt cccctgac atcaagaatc ttaagataga 5340

 agactcagat acttacatct gtgaagtgga ggaccagaag gaggaggtgc aattgctagt 5400
 gttcggattg actgccaaact ctgacacca cctgcttcag gggcagagcc tgacctgac 5460
 ctgggagagc ccccttggtg ttagccctc agtgcaatgt aggagtccaa ggggtaaaaa 5520
 catacagggg gggaagacc tctccgtgtc tcagctggag ctccaggata gtggcacctg 5580
 gacatgcact gtcttcaga accagaagaa ggtggagtgc aaaatagaca tcgtgtgtgt 5640
 agcttgataa atcgattagc ccgggtaccg tcgacctta gctagagctt ggcgtaatca 5700
 tggatcatagc tgtttctgt gtgaaattgt tatccgtca caattccaca caacatagc 5760

 gccggaagca taaagtgtaa agcctggggt gcctaagtag tgagctaact cacattaatt 5820
 gcgttgcgct cactgccgc tttccagtgc ggaaacctgt cgtgccagct gcattaatga 5880
 atcgccaac gcgcggggag aggcgggttg cgtattgggc gctcttcgc ttcctcgctc 5940
 actgactcgc tgcgctcggt cgttcggtg cggcgagcgg tatcagctca ctcaaaggcg 6000
 gtaatacgtt tatccacaga atcaggggat aacgcaggaa agaactatgt agcaaaaggc 6060
 cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca taggtccgc 6120
 cccctgacg agcatcaca aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga 6180

 ctataaagat accaggcgtt tccccctgga agctccctcg tgcgctctcc tgttccgacc 6240
 ctgccgtta ccggatact gtccgcttt ctccttcgg gaagcgtggc gctttctcat 6300
 agctcacgtt gtaggtatct cagttcgtg taggtcgttc gctccaagct gggctgtgtg 6360
 cacgaacccc ccgttcagcc cgaccgtgc gccttatccg gtaactatcg tcttgagtc 6420
 aacccggtaa gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga 6480
 gcgaggtatg taggcggtgc tacagagttc ttgaagtggg ggcctaacta cggtacact 6540

agaagaacag tatttggat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt 6600

ggtagctctt gatccggcaa acaaaccacc gctggtagcg gtggtttttt tgtttgcaag 6660

cagcagatta cgcgagaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg 6720

tctgacgctc agtggaaacga aaactcacgt taagggattt tggatcatgag attatcaaaa 6780

aggatcttca cctagatcct tttaaattaa aaatgaagtt ttaaataaat cttaaagtata 6840

tatgagtaaa cttggcttga cagttaccaa tgcctaatca gtgaggcacc tatctcagcg 6900

atctgtctat ttcgttcac catagttgcc tgactccccg tcgtgtagat aactacgata 6960

cgggagggct taccatctgg cccagtgct gcaatgatac cgcgagaccc acgtcaccg 7020

gctccagatt tatcagcaat aaaccagcca gccggaaggg ccgagcgcag aagtggtcct 7080

gcaactttat ccgctccat ccagctctat aattgttgc gggaagctag agtaagtagt 7140

tcgccagtta atagtttgcg caacgttgtt gccattgcta caggcatcgt ggtgtcacgc 7200

tcgtcgtttg gtatggcttc attcagctcc ggttccaac gatcaaggcg agttacatga 7260

tccccatgt tgtgcaaaaa agcggttagc tccttcggtc ctccgatcgt tgtcagaagt 7320

aagttggcgg cagtgttacc actcatggtt atggcagcac tgcataatc tcttactgtc 7380

atgccatccg taagatgctt ttctgtgact ggtgagtact caaccaagtc attctgagaa 7440

tagtgtatgc ggcgaccgag ttgctcttgc ccggcgctcaa tacgggataa taccgcgcca 7500

catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt cttcggggcg aaaactctca 7560

aggatcttac cgtgtttgag atccagttcg atgtaacca ctcgtgcacc caactgatct 7620

tcagcatctt ttactttcac cagcgtttct gggtagagcaa aaacaggaag gcaaaatgcc 7680

gcaaaaaagg gaataagggc gacacgaaa tgttgaatac tcatactctt ctttttcaa 7740

tattattgaa gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt 7800

tagaaaaata acaaatagg ggttcgcgc acatttcccc gaaaagtgcc acctgacgtc 7860

gacggatcgg gagatctccc gatccctat ggtgcactct cagtacaatc tgctctgatg 7920

ccgatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgt gagtagtgcg 7980

cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc 8040

ttagg 8045

<210> 98

<211> 9823

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: pAM3473

<400> 98

gttagcgctt ttgcgctgct tcgcgatgta cgggccagat atacgcgttg acattgatta	60
ttgactagtt attaatagta atcaattacg gggtcatttag ttcatagccc atatatggag	120
ttccgcttta cataacttac ggtaaatggc ccgcctggct gaccgcccac cgacccccgc	180
ccattgacgt caataatgac gtatgttccc atagtaacgc caataggacgac ttccattga	240
cgtcaatggg tggagtattt acggtaaaact gccacttgg cagtacatca agtgtatcat	300
atgccaagta cgcacctat tgacgtcaat gacggtaaat ggccgcctg gcattatgcc	360
cagtacatga ctttatggga ctttctact tggcagtaca tctacgtatt agtcatcgct	420
attaccatgg tgatgcggtt ttggcagtac atcaatgggc gtggatagcg gtttgactca	480
cggggatttc caagtctcca cccattgac gtcaatggga gtttgttttg gcacaaaat	540
caacgggact ttccaaaatg tcgtaacaac tccgccccat tgacgcaaata gggcggtagg	600
cgtgtacggt gggagggtcta tataagcaga gctcgtttag tgaaccgtca gatcgctgg	660
agacgccatc cacgtgttt tgacctccat agaagacacc gggaccgatc cagcctccgg	720
agcggccgcc accatgggca agccatccc taacccccctg ttggggctgg acagcaccgc	780
tcccaaaaag aaaaggaagg tgggcattca cggcgtgcct gcggccgaca aaaagtacag	840
catcgccctt gatatcgga ccaatagcgt gggtgggccc gttatcacag acgaatacaa	900
ggtaccagc aagaagttca aggtgctggg gaatacagac aggcactcta tcaagaaaaa	960
ccttatcggg gctctgctgt ttgactcagg cgagaccgcc gagggcacca ggttgaagag	1020
gaccgcaagg cgaaggtaca cccggaggaa gaacaggatc tgctatctgc aggagatctt	1080
cagcaacgag atggccaagg tggacgacag cttcttcac aggcctggagg agagcttcct	1140
tgctgaggag gataagaagc acgaacgaca cccatcttc ggcaacatag tcgacgaggt	1200
cgcttatcac gagaagtacc ccacctcta ccacctgca aagaaattgg tggatagcac	1260
cgataaaggc gacttgcgac ttatctactt ggctctggcg cacatgatta agttcagggg	1320
ccacttctg atcgagggcg accttaacct cgacaacagt gacgtagaca aattgttcat	1380
ccagcttgta cagacctata accagctgtt cgaggaaaac cctattaacg ccagcggggt	1440
ggatgcgaag gccatactta gcgccaggct gagcaaaagc aggcgcttgg agaacctgat	1500
agcccagctg cccggtgaaa agaagaacgg cctcttcggt aatctgattg ccctgagcct	1560
gggcctgacc cccaacttca agagcaactt cgacctggca gaagatgcca agctgcagtt	1620
gagtaaggac acctatgacg acgacttga caatctgctc gcccaaatcg gcgaccagta	1680
cgctgacctg ttctcgcgg ccaagaacct ttctgacgca atctgttta gcgatatcct	1740

tagggtgaac acagagatca ccaaggcccc cctgagcgcc agcatgatca agaggtacga	1800
cgagcaccat caggacctga cccctctgaa ggccctgggtg aggcagcaac tgcccagaaa	1860
gtacaaggag atctttttcg accagagcaa gaacggctac gccggctaca tcgacggcgg	1920
agccagccaa gaggagtctt acaagttcat caagcccatc ctggagaaga tggatggcac	1980
cgaggagctg ctggtgaagc tgaacaggga agatttgcct cggaagcaga ggacctttga	2040
caacggtagc atccccacc agatccacct gggcgagctg cacgcaatac tgaggcgaca	2100
ggaggatttc tacccttcc tcaaggacaa tagggagaaa atcgaaaaga ttctgacctt	2160
caggatcccc tactacgtgg gccctcttgc caggggcaac agccgattcg cttggatgac	2220
aagaaagagc gaggagacca tcccccttg gaacttcgag gaagtgggtg acaaaggagc	2280
aagcgcgcag tctttcatcg aacggatgac caatttcgac aaaaacctgc ctaacgagaa	2340
ggtgctgccc aagcacagcc tgctttacga gtacttcacc gtgtacaac agctcaccaa	2400
ggtgaaatat gtgaccgagg gcatgcgaaa acccgcttcc ctgagcggcg agcagaagaa	2460
ggccatcgtg gacctgctgt tcaagaccaa caggaagggt accgtgaagc agctgaagga	2520
ggactacttc aagaagatcg agtgctttga tagcgtggaa ataagcggcg tggaggacag	2580
gttcaacgcc agcctgggca cctaccacga cttgttgaag ataataaag acaaggattt	2640
cctggataat gaggagaacg aggatatact cgaggacatc gtgctgactt tgacctgtt	2700
tgaggaccga gagatgattg aagaaaggct caaacctac gccacactgt tcgacgacaa	2760
agtgtgaaa caactgaaga gacgaagata caccggctgg gccagactgt ccaggaagct	2820
catcaacggc attagggaca agcagagcgg caagaccatc ctggatttcc tgaagtccga	2880
cggcttcgcc aaccgaaact tcatgcagct gattcacgat gacagcttga ctttaagga	2940
ggacatccag aaggcccagg ttagcggcca gggcgactcc ctgcacgaac atattgcaaa	3000
cctggcaggc tcccctgcga tcaagaaggg catactgcag accgttaagg ttgtggacga	3060
attggtcaag gtcatgggca ggcacaagcc cgaaaacata gttatagaga tggccagaga	3120
gaaccagacc acccaaaagg gccagaagaa cagccgggag cgcatgaaaa ggatcgagga	3180
gggtatcaag gaactcggaa gccagatcct caaagagcac cccgtggaga ataccagct	3240
ccagaacgag aagctgtacc tgtactacct gcagaacggc agggacatgt acgttgacca	3300
ggagtggac atcaacagge ttctcagacta tgactggat cacatagtgc cccagagctt	3360
ttctaaagac gatagcatcg acaacaaggt cctgaccgc tccgacaaaa acaggggcaa	3420
aagcgacaac gtccaagcg aagaggtggt taaaaagatg aagaactact ggaggcaact	3480

gctcaacgcg aaattgatca cccagagaaa gtctgataac ctgaccaagg ccgagagggg 3540
cggactctcc gaacttgaca aagcgggctt cataaagagg cagctggtcg agacccgaca 3600
gatcacgaag cacttgcccc aaatcctcga cagcagaatg aataccaagt acgatgagaa 3660
tgacaaaactc atcagggaag tgaaagtgat taccttgaag agcaagttgg tgtccgactt 3720
tcgcaaagat ttccagtctc acaaggtgag ggagatcaac aactaccacc atgccacga 3780
cgcatacctg aacgccgtgg tcggcaccgc cctgattaag aagtatcaa agctggagtc 3840
cgaatttgtc tacggcgact acaaagttaa cgatgtgagg aagatgatcg ctaagagcga 3900

acaggagatc ggcaaggcca ccgctaagta tttcttctac agcaacatca tgaacttttt 3960
caagaccgag atcacacttg ccaacggcga aatcaggaag aggcgcgtta tcgagaccaa 4020
cgggtgagacc ggcgagatcg tgtgggacaa gggcaggacg ttcgccaccg tgaggaaagt 4080
cctgagcatg ccccagtgta atattgtgaa aaaaactgag gtgcagacag gcggctttag 4140
caaggaatcc atcctgcccc agaggaacag cgacaagctg atcgcccgga agaaggactg 4200
ggaccctaag aagtatggag gcttcgacag cccaccgta gcctacagcg tgctggtggt 4260
cgcaaggta gagaagggga agagcaagaa actgaagagc gtgaaggagc tgctcggcat 4320

aaccatcatg gagaggtcca gctttgagaa gaacccatt gactttttgg aagccaaggg 4380
ctacaaagag gtcaaaaagg acctgatcat caaactcccc aagtactccc tgtttgaatt 4440
ggagaacggc agaaagagga tgctggcgag cgctggggaa ctgcaaaagg gcaacgaact 4500
ggcgtgcccc agcaagtacg tgaattttct gtacctggcg tcccactacg aaaagctgaa 4560
aggcagcccc gaggacaacg agcagaagca gctgttcgtg gagcagcaca agcattacct 4620
ggacgagata atcgagcaaa tcagcgagtt cagcaagagg gtgattctgg ccgacgcgaa 4680
cctggataag gtcttcagcg cctacaacaa gcaccgagac aaacccatca gggagcaggc 4740

cgagaatata atacacctgt tcacctgac aaatctgggc gcacctgcgg cattcaaata 4800
cttcgatacc accatcgaca ggaaaaggta cactagcact aaggaggtgc tggatgccac 4860
cttgatccac cagtcatta ccggcctgta tgagaccagg atcgacctga gccagcttgg 4920
aggcgactct agggcggacc caaaaaagaa aaggaagggtg gaattctcta gaggcagtgg 4980
agagggcaga ggaagtctgc taacatgcgg tgacgtcgag gagaatcctg gcccaatgaa 5040
ccggggagtc ctttttaggc acttgcttct ggtgctgcaa ctggcgctcc tcccagcagc 5100
cactcaggga aagaaagtgg tgctgggcaa aaaaggggat acagtggaac tgacctgtac 5160

agcttcccag aagaagagca tacaattcca ctggaaaaac tccaaccaga taaagattct 5220
gggaaatcag ggctccttct taactaaagg tccatccaag ctgaatgatc gcgctgactc 5280
aagaagaagc ctttgggacc aaggaaactt cccctgatc atcaagaatc ttaagataga 5340

agactcagat acttacatct gtgaagtgga ggaccagaag gaggaggtgc aattgctagt 5400
 gttcggattg actgccaact ctgacacca cctgcttcag gggcagagcc tgacctgac 5460
 cttggagagc ccccttggtg gtagccctc agtgcaatgt aggagtccaa ggggtaaaaa 5520
 catacagggg gggaagacct tctccgtgtc tcagctggag ctccaggata gtggcacctg 5580

 gacatgcact gtcttcaga accagaagaa ggtggagttc aaaatagaca tcgtggtgct 5640
 agctttccag aaggcctcca gcatagtcta taagaaagag gggaacagg tggagtctc 5700
 ctccctcgc gcctttacag ttgaaaagct gacgggcagt ggcgagctgt ggtggcaggc 5760
 ggagagggtc tctctctcca agtcttggat cacttttgac ctgaagaaca aggaagtgtc 5820
 tgtaaacagg gttaccagg accctaagct ccagatgggc aagaagctcc cgctccacct 5880
 caccctgcc caggccttgc ctcatatgc tggctctgga aacctcacc tggcccttga 5940
 agcgaataa ggaaagtgc atcaggaagt gaacctggtg gtgatgagag ccactcagct 6000

 ccagaaaaat ttgacctgtg aggtgtgggg acccacctcc cctaagctga tgctgagctt 6060
 gaaactggag aacaaggagg caaaggtctc gaagcgggag aaggcgggtg ggggtgtgaa 6120
 ccctgaggcg gggatgtggc agtgtctgct gactgactcg ggacaggtcc tgctggaatc 6180
 caacatcaag gtcttgcaca catggtcgac ccggtgcag ccaatggccc tgattgtgct 6240
 ggggggcgtc gccggcctcc tgcctttcat tgggctaggc atcttcttct gtgtcaggtg 6300
 ccggcacacc ggttagtaat gagttaaac gggggaggct aactgaaaca cggaaggaga 6360
 caataccgga aggaaccgcg gctatgacgg caataaaaag acagaataaa acgcacgggt 6420

 gtgggtcgt ttgtcataa acgcggggtt cggctccagg gctggcactc tgtcgatacc 6480
 ccaccgagac cccattgggg ccaatcgcc cgcgtttctt ccttttccc accccacccc 6540
 ccaagtctcg gtgaaggccc agggctcgca gccaacgtcg gggcggcagg ccctgccata 6600
 gcagatctgc gcagctgggg ctctaggggg tatccccacg cgcctgtag cggcgcat 6660
 agcgcggcgg gtgtggtggt tacgcgcagc gtgaccgcta cacttgccag cgccttagcg 6720
 cccgtctctt tcgtttctt ccttctctt ctgccacgt tcgccggtt tccccgtcaa 6780
 gctctaaatc gggggctccc tttaggggtc cgatttagtg ctttacggca cctcgacccc 6840

 aaaaaacttg attagggtga tggttcacgt agtgggcat cgcctgata gacggttttt 6900
 cgcccttga cgttggagtc cagttcttt aatagtggac tcttgttcca aactggaaca 6960
 acactcaacc ctatctcgtt ctattctttt gatttataag ggattttgcc gatttcggcc 7020
 tattggttaa aaaatgagct gatttaacaa aaatttaacg cgaattaatt aaggtcgggc 7080
 aggaagaggg cctatttccc atgattctt catatttgca tatacgatac aaggctgtta 7140
 gagagataat tagaattaat ttgactgtaa acacaaagat attagtacaa aatacgtgac 7200

gtagaaagta ataatttctt gggtagtttg cagttttaaa attatgtttt aaaatggact 7260

atcatatgct taccgtaact tgaagattt tcgatttctt ggctttatat atcttgtgga 7320

aaggacgaaa caccgggggc cactagggac aggatgtttt agagctagaa atagcaagtt 7380

aaaataaggc tagtccgta tcaacttgaa aaagtggcac cgagtcggtg cttttttcta 7440

gtataccgtc gacctctagc tagagcttgg cgtaatcatg gtcatactgt tttcctgtgt 7500

gaaattgtta tccgctcaca attccacaca acatacgagc cggaagcata aagtgtaaag 7560

cctgggggtg ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgtt 7620

tccagtcggg aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag 7680

gcggtttgcg tattgggcgc tcttccgtt cctcgctcac tgactcgctg cgctcggctc 7740

ttcggtcgcg gcgagcggt tcaactcact caaaggcggg aatacggta tccacagaat 7800

caggggataa cgcaggaaa aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta 7860

aaaaggccgc gttgctggcg tttttcata ggctccgcc ccctgacgag catcacaaaa 7920

atcgacgtc aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc 7980

cccctggaag ctccctctg cgctctctg ttcgaccct gccgcttacc ggatactgt 8040

ccgcctttct cccttcggga agcgtggcg tttctcatag ctacgctgt aggtatctca 8100

gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc gttcagcccc 8160

accgctgcgc cttaaccggt aactatctc ttgagtccaa cccgtaaga cacgacttat 8220

cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta 8280

cagagtctt gaagtggtag cctaactacg gctacactag aagaacagta ttgggtatct 8340

gcgctctgct gaagccagtt accttcgga aaagagttgg tagctcttga tccggcaaac 8400

aaaccaccgc tggtagcgtt ggtttttttg ttgcaagca gcagattacg cgcagaaaaa 8460

aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa 8520

actcagtta agggattttg gtcatagat tatcaaaaag gatcttcacc tagatccttt 8580

taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact tggctgaca 8640

gttaccaatg cttaatcagt gaggcacct tctcagcgat ctgtctatct cgttcatcca 8700

tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggtta ccatctggcc 8760

ccagtgtgc aatgataccg cgagaccac gtcacccgc tccagattta tcagcaataa 8820

accagccagc cggaagggcc gagcgagaa gtggtcctgc aactttatcc gcctccatcc 8880

agtctattaa ttgttgcgg gaagctagag taagtagttc gccagttaat agtttgcgca 8940

acgtttgtgc cattgctaca ggcatcgtgg tgcacgctc gtcgtttggt atggcttcat 9000
 tcagtcctgg ttcccaacga tcaaggcgag ttacatgac ccccatgttg tgcaaaaaag 9060
 cggtagctc cttcggctct ccgacgttg tcagaagtaa gttggccgca gtgttatcac 9120
 tcatggttat ggagcactg cataattctc ttactgtcat gccatccgta agatgctttt 9180
 ctgtgactgg tgagtactca accaagtcac tctgagaata gtgtatgcgg cgaccgagtt 9240
 gctcttgccc gggtcaata cgggataata ccgcccaca tagcagaact ttaaaagtgc 9300
 tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat 9360

ccagttcgat gtaaccact cgtgcacca actgatcttc agcatctttt actttacca 9420
 gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataaggcgga 9480
 cacggaaatg ttgaatactc atactcttc ttttcaata ttattgaagc atttatcagg 9540
 gttattgtct catgagcgga tacatatctg aatgtattta gaaaaataaa caaatagggg 9600
 ttccgcgcac atttccccga aaagtccac ctgacgtcga cggatcgga gatctccga 9660
 tcccctatgg tgcactctca gtacaatctg ctctgatgcc gcatagttaa gccagtatct 9720
 gctccctgct tgtgtgttg aggtcgctga gtagtgcgcg agcaaaattt aagctacaac 9780

aaggcaaggc ttgaccgaca attgcatgaa gaatctgctt agg 9823

<210> 99
 <211> 3296
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: pAM3514
 <400> 99

ggtcgttcgg ctgcggcgag cggatcagc tcaactcaaag gcgtaatac ggttatccac 60
 agaatcaggg gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa 120
 ccgtaaaaag gccgcgttgc tggcgttttt ccatagctc cccccctg acgagcatca 180
 caaaaatcga cgtcaagtc agaggtggcg aaaccgaca ggactataaa gataccaggc 240
 gtttccccct ggaagctccc tcgtgcgctc tctgttccg accctgccgc ttaccgata 300

cctgtccgcc tttctccctt cgggaagcgt ggcgctttct catagctcac gctgtaggta 360
 tctcagttcg gtgtaggtcg ttgcgtccaa gctgggctgt gtgcacgaac cccccgttca 420
 gcccgaccgc tgcgccttat ccgtaacta tcgtcttaag tccaaccgg taagacacga 480
 cttatcgcca ctggcagcag ccaactggtta caggattagc agagcgaggt atgtaggcgg 540
 tgctacagag ttcttgaagt ggtggcctaa ctacggctac actagaagaa cagtatttgg 600
 tatctgcgct ctgctgaagc cagttacctt cggaaaaaga gttgtagct ctgatccgg 660

caaacaacc accgctggta gcggtggttt tttgtttgc aagcagcaga ttacgcgcag 720

 aaaaaagga tctcaagaag atcctttgat cttttctacg gggtctgacg ctccagtggaa 780
 cgaaaactca cgtaaggga ttttggatcat gagattatca aaaaggatct tcacctagat 840
 ccttttaaat taaaaatgaa gttttaaatc aatctaaagt atatatgagt aaactggtcg 900
 catgcttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct atttcgttca 960
 tccatagtgt cctgactgcc cgctgtgtag ataactacga tacgggaggg cttaccatct 1020
 ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggtccaga tttatcagca 1080
 ataaaccagc cagccggaag ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc 1140

 atccagtcta ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtttg 1200
 cgcaacgttg ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggatatggt 1260
 tcattcagct ccggttccca acgatcaagg cgagttacat gatccccat gttgtgcaaa 1320
 aaagcggtta gtccttcgg tectccgac gtgtgcagaa gtaagtggc cgagtgta 1380
 tcactcatgg ttatggcagc actgcataat tctcttactg tcatgccac cgtaagatgc 1440
 ttttctgtga ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg 1500
 agttgctctt gcccgcgctc aatacgggat aataccgcgc cacatagcag aactttaaaa 1560

 gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg 1620
 agatccagtt cgatgtaacc cactcgtgca cccaactgat cttcagcatc tttactttc 1680
 accagcgttt ctgggtgagc aaaaacagga aggcataatg ccgcaaaaaa gggaataagg 1740
 gcgacacgga aatgttgaat actcatcaat tgcctttttc aatattattg aagcatttat 1800
 cagggttatt gtctcatgag cggttacata ttgaatgta tttagaaaaa taaacaata 1860
 ggggttcgc gcacatttc ccgaaaagt ccacctgacg tctaagaaac cattattatc 1920
 atgacattaa cctataaaaa taggcgtatc acgagccct ttcattctgc gcgtttcgg 1980

 gatgacggtg aaaacctctg acacatgcag ctcccggaga cagtcacagc ttgtctgtaa 2040
 gcggtatgcc ggagcagaca agcccgtcag ggcgctcag cgggtgttg cgggtgtcgg 2100
 ggctggtaaa acgacggcca gtattaaccc tactaaagg gaactcgagg ctcttcagac 2160
 ggcacggcca cgcgtttaaa ccgcagttt tccacacgga cccccctc ctcaccacag 2220
 ccctgccagg acggggctgg ctactggcct tatctcacag gtaaaactga cgcacggagg 2280
 aacaatataa attggggact agaaaggtga agagccaaag ttagaactca ggaccaactt 2340
 attctgattt tgtttttcca aactgcttct cctcttggga agtgtaagga agctgcagca 2400

ccaggatcag tgaacgcac cagacggccg cgtcagagca gctcaggttc tgggagaggg 2460
tagcgacagg tggccactga gaaccgggca ggtcacgcat ccccccttc cctcccaccc 2520
cctgccaagc tctccctccc aggatcctct ctggctccat cgtaagcaaa ccttagaggt 2580
tctggcaagg agagagatgg ctccaggaaa tgggggtgtg tcaccagata aggaatctgc 2640
ctaacaggag gtgggggtta gacccaatat caggagacta ggaaggagga ggcctaagga 2700
tggggctttt ctgtcagaat tactgtggg gtggagggga cagataaaag taccagaac 2760
cagagccaca ttaaccggcc ctgggaatat aaggtggtcc cagctcgggg acacaggatc 2820

cctggaggca gcaaacatgc tgtctgaag tggacatagg ggcccgggtt ggaggaagaa 2880
gactagctga gctctcgac ccctggaaga tgccatgaca gggggctgga agagctagca 2940
cagactagag aggtaagggg ggtaggggag ctgccc aaat gaaaggagtg agaggtagcc 3000
cgaatccaca ggagaacggg gtgtccaggc aaagaaagca agaggatgga gaggtggcta 3060
aagccaggga gacggggtagc tttgggggtt tccagaaaaa cggtagatgat gcaggcctac 3120
aagaagggga ggcgggacgc aaggagaca tccgtcggag aaggccatcc taagaaacga 3180
gagatggcac agggcccaga aggagaagga aaagggaacg gtgtttaaac cccagcgcct 3240

ggcgggtgaa gagcgagctc ccgctgagca ataactagcg tcatagctgt ttcctg 3296

<210> 100
<211> 5068
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic: pAM3515
<400> 100

ggtcgttcgg ctgcggcgag cggatcagc tactcaaag gcgtaatac gttatccac 60
agaatcaggg gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa 120
ccgtaaaaag gccgcgttgc tggcggtttt ccatagctc cggccccctg acgagcatca 180
caaaaatcga cgtcaagtc agaggtggcg aaaccgcaca ggactataaa gataccaggc 240
gtttccccct ggaagctccc tcgtgcgctc tctgttccg accctgccgc ttaccgata 300

cctgtccgcc tttctccctt cgggaagcgt ggcgctttct catagctcac gctgtaggta 360
tctcagttcg gtgtaggtcg ttcgtccaa gctgggctgt gtgcacgaac cccccgttca 420
gcccgaaccg tgcgccttat ccgtaacta tcgtcttaag tccaaccggg taagacacga 480
cttatcgcca ctggcagcag ccaactggtaa caggattagc agagcgaggt atgtaggcgg 540
tgctacagag ttcttgaagt ggtggcctaa ctacggctac actagaagaa cagtatttgg 600
tatctgcgct ctgctgaagc cagttacctt cggaaaaaga gttggtagct cttgatccgg 660

caaacaacc accgctggta gcggtggttt tttgtttgc aagcagcaga ttacgcgcag 720

 aaaaaagga tctcaagaag atcctttgat cttttctacg gggtctgacg ctccagtggaa 780
 cgaaaactca cgtaaggga ttttggatcat gagattatca aaaaggatct tcacctagat 840
 ccttttaaat taaaaatgaa gttttaaatc aatctaaagt atatatgagt aaactggtcg 900
 catgcttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct atttcgttca 960
 tccatagtgt cctgactgcc cgctgtgtag ataactacga tacgggaggg cttaccatct 1020
 ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggtccaga tttatcagca 1080
 ataaaccagc cagccggaag ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc 1140

 atccagtcta ttaattgttg ccgggaagct agagtaagta gtccgaccagt taatagtttg 1200
 cgcaacgttg ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggatatggt 1260
 tcattcagct ccggttccca acgatcaagg cgagttacat gatccccat gttgtgcaaa 1320
 aaagcggtta gtccttcgg tctccgcatc gttgtcagaa gtaagtggc cgagtgta 1380
 tcactcatgg ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc 1440
 ttttctgtga ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg 1500
 agttgtctct gcccgcgctc aatacgggat aataccgctc cacatagcag aactttaaaa 1560

 gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg 1620
 agatccagtt cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc 1680
 accagcgttt ctgggtgagc aaaaacagga aggcataatg ccgcaaaaaa gggaataagg 1740
 gcgacacgga aatgttgaat actcatcaat tgcctttttc aatattattg aagcatttat 1800
 cagggttatt gtctcatgag cggttacata ttgaatgta tttagaaaaa taaacaata 1860
 ggggttcgc gcacatttc ccgaaaagt ccacctgacg tctaagaaac cattattatc 1920
 atgacattaa cctataaaaa taggcgtatc acgagccct ttcattctgc gcgtttcgg 1980

 gatgacggtg aaaacctctg acacatgcag ctcccggaga cagtcacagc ttgtctgtaa 2040
 gcggtatgcc ggagcagaca agcccgctcag ggcgctcag cgggtgttgg cgggtgtcgg 2100
 ggctggtaaa acgacggcca gtattaaccc tctaataagg gaactcgagg ctcttcagac 2160
 ggcacggcca cgcgtttaaa ccgcataca gtggaactga cctgtacagc tcccagaag 2220
 aagagcatac aattccactg gaaaaactcc aaccagataa agattctggg aatcagggc 2280
 tccttcttaa ctaaaggctc atccaagctg aatgatcgcg ctgactcaag aagaagcctt 2340
 tgggaccaag gaaacttccc cctgatcatc aagaatctta agatagaaga ctccagatac 2400

tacatctgtg aagtggagga ccagaaggag gaggtgcaat tgctagtgtt cggattgact	2460
gccaaactctg acaccacct gcttcagggg cagagcctga ccctgacctt ggagagcccc	2520
cctggtagta gcccctcagt gcaatgtagg agtccaaggg gtaaaaacat acaggggggg	2580
aagacctctt ccgtgtctca gctggagctc caggatagtg gcacctggac atgcactgtc	2640
ttgcagaacc agaagaaggt ggagttcaaa atagacatcg tgggtctagc ttccagaag	2700
gcctccagca tagtctataa gaaagagggg gaacaggtgg agttctcctt cccactcgcc	2760
tttacagttg aaaagctgac gggcagtggc gagctgtggt ggcaggcgga gagggcttcc	2820
tcctccaagt cttggatcac ctttgacctg aagaacaagg aagtgtctgt aaaacgggtt	2880
accaggacc ctaagctcca gatgggcaag aagctccgc tccacctcac cctgccccag	2940
gccttgctc agtatgctgg ctctggaaac ctaccctgg cccttgaagc gaaaacagga	3000
aagttgcatc aggaagtga cctggtggtg atgagagcca ctgagctcca gaaaaatttg	3060
acctgtgagg tgtggggacc cacctcccct aagctgatgc tgagcttgaa actggagaac	3120
aaggaggcaa aggtctcgaa gcgggagaag gcggtgtggg tgctgaacc tgaggcgggg	3180
atgtggcagt gtctgctgag tgactcggga caggctctgc tggaatccaa catcaaggtt	3240
ctgccacat ggtcgacccc ggtgcagcca atggccctga ttgtgctggg gggcgtcgcc	3300
ggcctcctgc ttttcattgg gctaggcatc ttcttctgtg tcagggtccg gcacaccgtt	3360
tagtaatgag tttaaacggg ggaggctaac tgaaacacgg aaggagacaa taccggaagg	3420
aaccgcgct atgacggcaa taaaagaca gaataaacg cacgggtgtt gggtcgtttg	3480
ttcataaac cggggttcgg tcccagggtt ggcactctgt cgataccca ccgagacccc	3540
attggggcca atacgccgc gtttcttctt tttccccacc ccacccccca agttcgggtg	3600
aaggcccagg gctcgagcc aacgtcgggg cggcaggccc tgccatagca gatctgcgca	3660
gctggggctc tagggggtat cccacgcgc cctgtagcgg cgcattaagc gcggcgggtg	3720
tgggtggttac gcgcagcgtg accgtacac ttgccagcgc cctagcgccc gtcctttcg	3780
ctttcttccc ttcttttctc gccacgttcg ccggctttcc ccgtcaagct ctaaatcggg	3840
ggctcccttt agggttccga tttagtgtt tacggcacct cgaccccaaa aaacttgatt	3900
aggggtgatgg ttacgtagt gggccatgc cctgatagac ggtttttcgc cttttgacgt	3960
tggagtccac gttctttaat agtggactct tgttccaaac tggaacaaca ctcaacccta	4020
tctcggctta ttcttttgat ttataaggga ttttgccgat ttcggcctat tggttaaaaa	4080
atgagctgat ttaacaaaaa tttaacgcga attaatatt gtcgggcagg aagaggcct	4140
atttcccatg attccttcat atttgcatat acgatacaag gctgttagag agataattag	4200
aattaatttg actgtaaca caaagatat agtacaaat acgtgacgta gaaagtaata	4260

atttcttggg tagtttgag ttttaaaatt atgttttaaa atggactatc atatgcttac 4320
 cgtaacttga aagtatttcg atttcttggc ttatatatc ttgtggaaag gacgaaacac 4380
 cgggggccac tagggacagg atgttttaga gctagaaata gcaagttaaa ataaggctag 4440
 tccgttatca acttgaaaaa gtggcaccga gtcggtgctt ttttctagta taccgtcgac 4500

 ctctagctag agcttggcgt aatcatggtc atagctgttt cctgtgtgaa attgttatcc 4560
 gtcacaatt ccacacaaca tacgagccgg aagcataaag tgtaaagcct ggggtgccta 4620
 atgagtgagc taactcacat taattgcgtt gcgctcactg cccgctttcc agtcgggaaa 4680
 cctgtcgtgc cagctgcatt aatgaatcgg ccaacgcgcg gggagaggcg gtttgcgtat 4740
 tgggcgctct tccgcttccg cgctcactga ctgcgtgcgc tcggtcgttc ggctgcggcg 4800
 agcggtatca gctcactcaa aggcggtaat acggttatcc acagaatcag gggataacgc 4860
 aggaaagaac atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt 4920

 gctggcgttt ttccataggc tccgcccccc tgacgagcat cacaaaaatc gacgctcaag 4980
 tcagaggtgg cgggtgttaa accccagcgc ctggcgggtg aagagcgagc tcccgtgag 5040
 caataactag cgtcatagct gtttccgtg 5068

 <210> 101
 <211> 4224
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: pAM3516
 <400> 101

 ggtcgttcgg ctgcggcgag cggatcagc tactcaaag gcgtaatac ggttatccac 60
 agaatcaggg gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa 120
 ccgtaaaaag gccgcgttgc tggcgttttt ccataggctc cgccccctg acgagcatca 180

 caaaaatcga cgtcaagtc agaggtggcg aaacccgaca ggactataaa gataccaggc 240
 gtttccccct ggaagctccc tcgtgcgtc tctgttccg acctgccgc ttaccgata 300
 cctgtccgcc ttctccctt cgggaagcgt ggcgctttct catagctcac gctgtaggta 360
 tctcagttcg gtgtaggtcg ttcgtccaa gctgggctgt gtgcacgaac cccccgttca 420
 gcccgaccgc tgcgccttat ccgtaacta tcgtcttaag tccaacccgg taagacacga 480
 ctatcgcca ctggcagcag ccactggtaa caggattagc agagcgaggt atgtaggcgg 540
 tgctacagag ttcttgaagt ggtggcctaa ctacggctac actagaagaa cagtatttgg 600

 tatctgcgct ctgctgaagc cagttacctt cggaaaaaga gttggtagct cttgatccgg 660

caaacaacc accgctggta gcgggtggtt tttgtttgc aagcagcaga ttacgcgcag 720
 aaaaaagga tctcaagaag atcctttgat cttttctacg gggtctgacg ctcagtggaa 780
 cgaaaactca cgtaaggga ttttggcat gagattatca aaaaggatct tcacctagat 840
 ccttttaaat taaaaatgaa gttttaaatc aatctaaagt atatatgagt aaactggtcg 900
 catgcttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct atttcgttca 960
 tccatagtgt cctgactgcc cgctgtgtag ataactacga tacgggaggg cttaccatct 1020

 ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggtccaga tttatcagca 1080
 ataaaccagc cagccggaag ggccgagcgc agaagtggc ctgcaacttt atccgcctcc 1140
 atccagtcta ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtgtg 1200
 cgcaacgttg ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggatggct 1260
 tcattcagct ccggttccca acgatcaagg cgagttacat gatccccat gttgtgcaaa 1320
 aaagcggtta gtccttcgg tcctccgac gttgtcagaa gtaagtggc cgcagtgtta 1380
 tcactcatgg ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc 1440

 ttttctgtga ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg 1500
 agttgtctct gcccggcgtc aatacgggat aataccgcgc cacatagcag aactttaaaa 1560
 gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg 1620
 agatccagtt cgatgtaacc cactcgtgca ccaactgat cttcagcatc tttactttc 1680
 accagcgttt ctgggtgagc aaaaacagga aggcaaatg ccgcaaaaaa gggaataagg 1740
 gcgacacgga aatgttgaat actcatcaat tgcctttttc aatattattg aagcatttat 1800
 cagggttatt gtctcatgag cggttacata ttgaaatgta tttagaaaaa taaacaaata 1860

 ggggttccgc gcacatttcc ccgaaaagt ccacctgacg tctaagaaac cattattatc 1920
 atgacattaa cctataaaaa taggcgtatc acgaggccct ttcattctgc gcgtttcgg 1980
 gatgacggtg aaaacctctg acacatgcag ctcccggaga cagtcacagc ttgtctgtaa 2040
 gcggatgccg ggagcagaca agcccgtcag ggcgctcag cgggtgttgg cgggtgtcgg 2100
 ggctggtaaa acgacggcca gtattaacc tactaaagg gaactcgagg ctcttcagac 2160
 ggcacggcca cgcgtttaaa ccgcataca gtggaactga cctglacagc ttcccagaag 2220
 aagagcatac aattccactg gaaaaactcc aaccagataa agattctggg aatcagggc 2280

 tccttcttaa ctaaaggccc atccaagctg aatgatcgcg ctgactcaag aagaagcctt 2340
 tgggaccaag gaaacttccc cctgatcatc aagaatctta agatagaaga ctcagatact 2400
 tacatctgtg aagtggagga ccagaaggag gaggtgcaat tgctagtgtt cggattgact 2460
 gccactctg acaccacct gcttcagggg cagagcctga cctgacctt ggagagcccc 2520

cctggtagta gccctcagtg gcaatgtagg agtccaaggg gtaaaaacat acaggggggg 2580
aagacctctt ccgtgtctca gctggagctc caggatagtg gcacctggac atgcactgtc 2640
ttgcagaacc agaagaaggt ggagttcaaa atagacatcg tgggtgctagc ttccagaag 2700

gcctccagca tagtctataa gaaagagggg gaacaggtgg agttctcctt cccactcgcc 2760
tttacagttg aaaagctgac gggcagtgga gagctgtggt ggcaggcgga gagggttcc 2820
tcctccaagt cttgatcac ctttgacctg aagaacaagg aagtgtctgt aaaacgggtt 2880
accaggacc ctaagctcca gatgggcaag aagctccgc tccacctcac cctgccccag 2940
gccttgctc agtatgtgg ctctggaaac ctccacctgg cccttgaagc gaaaacagga 3000
aagttgcatc aggaagtga cctgggtgtg atgagagcca ctgagctcca gaaaaattg 3060
acctgtgagg tgtggggacc cactccctt aagctgatgc tgagcttgaa actggagaac 3120

aaggaggcaa aggtctcgaa gcgggagaag gcggtgtggg tgctgaacct tgaggcgagg 3180
atgtggcagt gtctgtgag tgactcgga caggtctgc tggaatcaa catcaaggtt 3240
ctgccacat ggtcgacccc ggtgcagcca atggccctga ttgtgtggg gggcgtcgcc 3300
ggcctctgc ttttcattgg gctaggcatc ttcttctgtg tcaggtgccg gcacaccgtt 3360
tagtaatgag tttaaacggg ggaggtaac tgaaacacgg aaggagaca taccggaagg 3420
aaccgcgct atgacggcaa taaaagaca gaataaacg cacgggtgtt gggtcgtttg 3480
ttcataaacg cggggttcgg tcccagggtt ggcaactctg cgataccca ccgagacccc 3540

attggggcca atacgccgc gtttcttctt tttcccacc ccacccccca agttcgggtg 3600
aagggccagg gctcgcagcc aacgtcgagg cggcaggccc tgccatagcc gtcgacctt 3660
agctagagct tggcgtaatc atggcatag ctgtttcttg tgtgaaattg ttatccgctc 3720
acaattccac acaacatag agccggaagc ataaagtga aagcctgggg tgcctaatga 3780
gtgagetaac tcacattaat tgcgttgccg tcaactgccg ctttcagtc gggaaacctg 3840
tcgtgccagc tgcattaatg aatcgccaa cgcgcgggga gaggcggtt gcgtattggg 3900
cgctcttcg cttctctgct cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg 3960

glatcagctc actcaaagc ggtaatacgg ttatccacag aatcagggga taacgcagga 4020
aagaacatgt gagcaaaagg ccagcaaaag gccagggaacc gtaaaaagc cgcgttgctg 4080
gcgtttttcc ataggctccg cccccctgac gagcatcaca aaaatcgac ctcaagtcag 4140
aggtggcggg gtttaaaccc cagcgcttgg cgggtgaaga gcgagctccc gctgagcaat 4200
aactagcgtc atagctgttt cctg 4224

<210> 102

<211> 9467

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: pAM15068

<400> 102

```

gttaggcgtt ttgcgctgct tcgcgatgta cgggccagat atacgcgttg acattgatta      60

ttgactagtt attaatagta atcaattacg gggtcattag ttcatagccc atatatggag      120
ttccgcgtta cataacttac ggtaaatggc ccgcctggct gaccgcccaa cgacccccgc      180
ccattgacgt caataatgac gtatgttccc atagtaacgc caataggagc ttccattga      240
cgtcaatggg tggagtatit acggtaaact gccacttgg cagtacatca agtgtatcat      300
atgccaagta cgccccctat tgacgtcaat gacggtaaata ggccgcctg gcattatgcc      360
cagtacatga cttatggga ctttctact tggcagtaca tctacgtatt agtcacgct      420
attaccatgg tgatgcggtt ttggcagtac atcaatgggc gtggatagcg gtttgactca      480

cggggatttc caagtctcca cccattgac gtcaatggga gtttgttttg gcacaaaat      540
caacgggact ttccaaaatg tcgtaacaac tccgccccat tgacgcaaata gggcggtagg      600
cgtgtacggt gggaggctta tataagcaga gctcgttttag tgaaccgtca gatcgcttg      660
agacgccatc cacgtgtttt tgacctccat agaagacacc gggaccgatc cagcctccgg      720
agcggccgcc accatgggca agccatccc taacccccctg ttggggctgg acagcaccgc      780
tcccaaaaag aaaaggaagg tgggcattca cggcgtgcct gcggccgaca aaaagtacag      840
catcgccctt gatatcgga ccaatagcgt gggctgggcc gttatcacag acgaatacaa      900

ggtaccagc aagaagtta aggtgctggg gaatacagac aggcactcta tcaagaaaa      960
ccttatcggg gctctgctgt ttgactcagg cgagaccgcc gaggccacca ggttgaagag      1020
gaccgcaagg cgaaggtaca cccggaggaa gaacaggatc tgctatctgc aggagatctt      1080
cagcaacgag atggccaagg tggacgacag ctctctccac aggcctggagg agagcttctt      1140
tgtcgaggag gataagaagc acgaacgaca ccccatcttc ggcaacatag tcgacgaggt      1200
cgcttatcac gagaagtacc ccacatcta ccacctgca aagaatttg tggatagcac      1260
cgataaagcc gacttgcgac ttatctactt ggctctggcg cacatgatta agttcagggg      1320

ccacttctg atcgaggcg accttaacc cgacaacagt gacgtagaca aattgttcat      1380
ccagcttgta cagacctata accagctgtt cgaggaaaac cctattaacg ccagcggggt      1440
ggatgcgaag gccatactta gcgccaggct gagcaaaagc aggcgcttgg agaacctgat      1500
agcccagctg cccggtgaaa agaagaacgg cctcttcggt aatctgattg ccctgagcct      1560
gggcctgacc cccaactta agagcaactt cgacctggca gaagatgcca agctgcagtt      1620

```

gagtaaggac acctatgacg acgacttgga caatctgctc gcccaaatcg gcgaccagta	1680
cgctgacctg ttctctgccg ccaagaacct ttctgacgca atcctgctta gcgatatcct	1740
tagggtgaac acagagatca ccaaggcccc cctgagcgcc agcatgatca agaggtagca	1800
cgagcaccat caggacctga cccctctgaa ggccctggtg aggcagcaac tgcccagaaa	1860
gtacaaggag atctttttcg accagagcaa gaacggctac gccggctaca tcgacggcgg	1920
agccagccaa gaggagtctt acaagtcat caagcccatc ctggagaaga tggatggcac	1980
cgaggagctg ctggtgaagc tgaacaggga agatttctc cggaagcaga ggacctttga	2040
caacggtagc atccccacc agatccacct ggccgagctg cacgcaatac tgaggcgaca	2100
ggaggatttc tacccttcc tcaaggacaa tagggagaaa atcgaaaaga ttctgacctt	2160
caggatcccc tactacgtgg gccctcttgc caggggcaac agccgattcg cttggtatgac	2220
aagaaagagc gaggagacca tccccctg gaacttcgag gaagtggagg acaaaggagc	2280
aagcgcgcag tctttcatcg aacggatgac caatttcgac aaaaacctgc ctaacgagaa	2340
ggtgctgccc aagcacagcc tgctttacga gtacttcacc gtgtacaacg agctcaccaa	2400
ggtgaaatat gtgaccgagg gcatcgaaa acccgcttcc ctgagcggcg agcagaagaa	2460
ggccatcgct gacctgctgt tcaagaccaa caggaagggt accgtgaagc agctgaagga	2520
ggactacttc aagaagatcg agtgctttga tagcgtggaa ataagcggcg tggaggacag	2580
gttcaacgcc agcctgggca cctaccacga cttgttgaag ataataaag acaaggattt	2640
cctggataat gaggagaacg aggatatact cgaggacatc gtgctgactt tgacctgtt	2700
tgaggaccga gagatgattg aagaaaggct caaaacctac gccacctgt tcgacgacaa	2760
agtgatgaaa caactgaaga gacgaagata caccggctgg ggacagactgt ccaggaaagt	2820
catcaacggc attaggagca agcagagcgg caagaccatc ctggatttcc tgaagtccga	2880
cggcttcgcc aaccgaaact tcatgcagct gattcacgat gacagcttga cttcaagga	2940
ggacatccag aaggcccagg ttagcggcca ggccgactcc ctgcacgaac atattgcaaa	3000
cctggcaggc tcccctgcga tcaagaaggg catactgcag accgttaagg ttgtggacga	3060
attggtcaag gtcatgggca ggcacaagcc cgaaaacata gttatagaga tggccagaga	3120
gaaccagacc acccaaaagg gccagaagaa cagccgggag cgcatgaaaa ggatcgagga	3180
gggtatcaag gaactcgga gccagatcct caaagagcac cccgtggaga ataccagct	3240
ccagaacgag aagctgtacc tgtactacct gcagaacggc agggacatgt acgttgacca	3300
ggagttggac atcaacaggc ttccagacta tgacgtggat cacatagtgc cccagagctt	3360

tcttaaagac gatagcatcg acaacaaggt cctgacccgc tccgacaaaa acaggggcaa	3420
aagcgacaac gtgccaagcg aagaggtggt taaaaagatg aagaactact ggaggcaact	3480
gctcaacgcg aaattgatca cccagagaaa gttcgataac ctgaccaagg ccgagagggg	3540
cggactctcc gaacttgaca aagcgggcctt cataaagagg cagctggtcg agacccgaca	3600
gatcacgaag cacgtggccc aaatcctcga cagcagaatg aataccaagt acgatgagaa	3660
tgacaaactc atcaggggaag tgaaagtgat taccctgaag agcaagttgg tgtccgactt	3720
tcgcaaagat ttccagttct acaaggtgag ggagatcaac aactaccacc atgccacga	3780
cgcatacctg aacgccgtgg tcggcaccgc cctgattaag aagtatccaa agctggagtc	3840
cgaatttgtc tacggcgact acaaagttaa cgatgtgagg aagatgatcg ctaagagcga	3900
acaggagatc ggcaaggcca ccgctaagta tttcttctac agcaacatca tgaacttttt	3960
caagaccgag atcacacttg ccaacggcga aatcaggaag aggccgctta tcgagaccaa	4020
cgggtgagacc ggcgagatcg tgtgggacaa gggcaggacgac ttcgccaccg tgaggaaagt	4080
cctgagcatg ccccaggtga atattgtgaa aaaaactgag gtgcagacag gcggctttag	4140
caaggaatcc atcctgcccc agaggaacag cgacaagctg atcgcccgga agaaggactg	4200
ggaccctaag aaglatggag gcttcgacag cccaccgta gcctacagcg tgctggtggt	4260
cgcgaaggta gagaagggga agagcaagaa actgaagagc gtgaaggagc tgctcggcat	4320
aaccatcatg gagaggtcca gctttgagaa gaacccatt gactttttgg aagccaaggg	4380
ctacaaagag gtcaaaaagg acctgatcat caaactcccc aagtactccc tgtttgaatt	4440
ggagaacggc agaaagagga tgctggcgag cgtgggggaa ctgcaaaagg gcaacgaact	4500
ggcgctgccc agcaagtiac tgaattttct gtacctggcg tcccactacg aaaagctgaa	4560
aggcagcccc gaggacaacg agcagaagca gctgttcgtg gagcagcaca agcattacct	4620
ggacgagata atcgagcaaa tcagcgagtt cagcaagagg gtgattcttg ccgacgcgaa	4680
cctggataag gtcctcagcg cctacaacaa gcaccgagac aaacccatca gggagcaggc	4740
cgagaatatc atacacctgt tcacctgac aaatctgggc gcacctgcgg cattcaaata	4800
cttcgatacc accatcgaca ggaaaaggta cactagcact aaggaggtgc tggatgccac	4860
cttgatccac cagtcatta ccggcctgta tgagaccagg atcgacctga gccagcttgg	4920
aggcgactct agggcggacc caaaaaagaa aaggaaggtg gaattctcta gaggcagtgg	4980
agagggcaga ggaagtctgc taacatgcgg tgacgtcgag gagaatcctg gcccattgaa	5040
ccggggagtc ctttttaggc acttgcttct ggtgctgcaa ctggcgctcc tcccagcagc	5100

cactcaggga aagaaagtgg tgctgggcaa aaaaggggat acagtggaac tgacctgtac 5160
 agcttcccag aagaagagca tacaattcca ctggaaaaac tccaaccaga taaagattct 5220
 gggaaatcag ggctccttct taactaaagg tccatccaag ctgaatgatc gcgctgactc 5280
 aagaagaagc ctttgggacc aaggaaactt ccccttgatc atcaagaatc ttaagataga 5340
 agactcagat acttacaict gtgaagtgga ggaccagaag gaggaggtgc aattgctagt 5400
 gttcggattg actgccaact ctgacacca cctgcttcag gggcagagcc tgacctgac 5460
 cttggagagc ccccttggtg gtagccctc agtgcaatgt aggagtcaa ggggtaaaaa 5520

 catacagggg gggaagacc tctccgtgtc tcagctggag ctccaggata gtggcacctg 5580
 gacatgcact gtcttcaga accagaagaa ggtggagtgc aaaatagaca tcgtggtgct 5640
 agctttccag aaggcctcca gcatagtcta taagaaagag ggggaacagg tggagtctc 5700
 ctccactc gcctttacag ttgaaaagct gacgggcagt ggcgagctgt ggtggcaggc 5760
 ggagagggct tctcctcca agtcttggat cacctttgac ctgaagaaca aggaagtgtc 5820
 tgtaaacagg gttaccagg accctaagct ccagatgggc aagaagctcc cgctccacct 5880
 caccctgcc caggccttgc ctacgtatgc tggctctgga aacctaccc tggcccttga 5940

 agcgaataca ggaaagtgc atcaggaagt gaacctggtg gtgatgagag ccactcagct 6000
 ccagaaaaat ttgacctgtg aggtgtgggg acccacctcc cctaagctga tgctgagctt 6060
 gaaactggag aacaaggagg caaaggtctc gaagcgggag aaggcgggtg ggggtgctgaa 6120
 ccctgaggcg gggatgtggc agtgtctgct gactgactcg ggacaggtcc tgctggaatc 6180
 caacatcaag gttctgccc catggtcgac cccggtgcag ccaatggccc tgattgtgct 6240
 ggggggcgtc gccggcctcc tgcttttcat tgggctaggc atcttcttct gtgtcaggtg 6300
 ccggcacacc ggttagtaat gagttaaac gggggaggct aactgaaaca cggaaggaga 6360

 caataccgga aggaaccgc gctatgacgg caataaaaag acagaataaa acgcacgggt 6420
 gttgggtcgt ttgtcataa acgcgggggt cggctccagg gctggcactc tgtcgatacc 6480
 ccaccgagac ccattgggg ccaatacgcc cgcgtttctt cttttcccc accccacccc 6540
 ccaagttcgg gtgaaggccc agggctcgca gccaacgtcg gggcggcagg cctgccata 6600
 gcagatctgc gcagctgggg ctctaggggg tatccccacg cgccctgtag cggcgatta 6660
 agcgcggcgg gtgtgggtgt tacgcgcagc gtgaccgcta cacttgccag cgccctagcg 6720
 cccgtcctt tcgttttctt ccttccctt ctcgccacgt tcgcgggtt tccccgtcaa 6780

 gctctaaatc gggggctccc tttaggggtc cgatttagtg ctttacggca cctcgacccc 6840
 aaaaaacttg attagggtga tggttcacgt agtgggcat cgccctgata gacgggtttt 6900
 cgcccttga cgttggagtc cacttcttt aatagtgac tcttgttcca aactggaaca 6960

acactcaacc ctatctcggt ctattctttt gatttataag ggattttgcc gatttcggcc 7020
 tattggttaa aaaaatgagct gatttaacaa aaatttaacg cgaattaatt aagcgccgc 7080
 gtgcatatac cgtcgacctc tagctagagc ttggcgtaat catggtcata gctgtttcct 7140
 gtgtgaaatt gttatccgct cacaattcca cacaacatac gagccggaag cataaagtgt 7200

 aaagcctggg gtgcctaata agtgagctaa ctcacattaa ttgcgttgcg ctactgccc 7260
 gctttccagt cgggaaacct gtcgtgccag ctgcattaat gaatcgcca acgcgcgggg 7320
 agaggcgggt tgcgtattgg gcgctcttcc gcttctcgc tcaactgactc gctgcgctcg 7380
 gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg cggtataacg gttatccaca 7440
 gaatcagggg ataacgcagg aaagaacatg tgagcaaaag gccagcaaaa ggccaggaac 7500
 cgtaaaaagg ccgcgttgct ggcgttttc cataggctcc gccccctga cgagcatcac 7560
 aaaaatcgac gctcaagtca gaggtggcga aaccgcagag gactataaag ataccaggcg 7620

 tttccccctg gaagctccct cgtgcgctct cctgttccga cctgccgct taccggatac 7680
 ctgtccgct tttcccttc gggaagcgtg gcgctttctc atagctcacg ctgtaggat 7740
 ctcaattcgg ttaggtcgt tgcctccaag ctggcgctgtg tgcacgaacc ccccgctcag 7800
 cccgaccgct gcgccttacc cggtaactat cgtcttgagt ccaaccgggt aagacacgac 7860
 ttatcgccac tggcagcagc cactggtaac aggattagca gagcgaggta ttaggcgggt 7920
 gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaagaac agtatttggt 7980
 atctgcgctc tgcgaagcc agttacctc ggaaaaagag ttggtagctc ttgatccggc 8040

 aaacaaacca ccgtggttag cgggtggtttt tttgtttgca agcagcagat tacgcgcaga 8100
 aaaaaggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc tcagtgaac 8160
 gaaaactcac gtttaaggat tttggtcatg agattatcaa aaagatctt cacctagatc 8220
 cttttaaat aaaaatgaag ttttaaatca atctaaagta tataatgagta aacttggctc 8280
 gacagtacc aatgcttaat cagtaggca cctatctcag cgatctgtct atttcgttca 8340
 tccatagttg cctgactccc cgctgtgtag ataactacga tacgggaggg cttaccatct 8400
 ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggtccaga tttatcagca 8460

 ataaaccagc cagccggaag ggccgagcgc agaagtggc ctgcaacttt atccgcctcc 8520
 atccagtcta ttaattgtg ccgggaagct agagtaagta gttcgccagt taatagtttg 8580
 cgcaacgttg ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggatggct 8640
 tcattcagct ccggttccca acgatcaagg cgagttacat gatccccat gttgtgcaaa 8700
 aaagcggtta gtccttcgg tcctccgac gttgtcagaa gtaagtggc cgcagtgtta 8760
 tcaactatgg ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc 8820

ttttctgtga ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg 8880

agttgtctctt gcccgcgctc aatacgggat aataccgcgc cacatagcag aactttaaaa 8940

gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg 9000

agatccagtt cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc 9060

accagcgttt ctgggtgagc aaaaacagga aggcaaatg ccgcaaaaaa gggaataagg 9120

gcgacacgga aatgttgaat actcatactc ttcctttttc aatattattg aagcatttat 9180

cagggttatt gtctcatgag cggatacata ttgaaatgta tttagaaaaa taaacaata 9240

ggggttcgcg gcacatttcc ccgaaaagtg ccacctgacg tcgacggatc gggagatctc 9300

ccgatccctt atggtgcact ctcaagtaca tctgctctga tgccgcatag ttaagccagt 9360

atctgtctcc tgcttgtgtg ttggaggctg ctgagtagtg cgcgagcaaa atttaagcta 9420

caacaaggca aggcttgacc gacaattgca tgaagaatct gcttagg 9467

<210> 103

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: CUT1214

<400> 103

ctagcttgat aaatcgatta gcccggg 27

<210> 104

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: CUT1215

<400> 104

ccccgggctaa tcgatttatac aag 23

<210> 105

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: CUT1216

<400> 105

ggggccacta gggacaggat gtttt 25

<210> 106
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: CUT1217
 <400> 106
 atcctgtccc tagtggcccc cggtg 25
 <210> 107
 <211> 53
 <212> DNA
 <213>
 Artificial Sequence
 <220><223> Synthetic: CUT1220
 <400> 107
 gacggcacgg ccacgcgttt aaaccgccat acagtggaac tgacctgtac agc 53
 <210> 108
 <211> 49
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: CUT1221
 <400> 108
 cccgccaggc gctggggttt aaacaccgcc acctctgact tgagcgtcg 49
 <210> 109
 <211> 40
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: CUT1224
 <400> 109
 ggggcttttc tgtcagaatt cactgtgggg tggaggggac 40
 <210> 110
 <211> 52
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic: CUT1227

<400> 110
cccgccaggc gctggggttt aaacaccgtt cccttttctt tctccttctg gg 52
<210> 111
<211> 38
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic: CUT1252
<400> 111
gctctagcta gaggtcgacg gctatggcag ggcctgcc 38
<210> 112
<211> 43
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic: CUT1253
<400> 112
gcggcaggcc ctgccatagc cgctgacctc tagctagagc ttg 43
<210> 113
<211> 18
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic: CUT1254
<400> 113
taagcggccg ctgcagta 18
<210> 114
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic: CUT1255
<400> 114
tactgcagcg gccgcttaat 20
<210> 115
<211> 20
<212> DNA
<213> Artificial Sequence

<220><223> Synthetic: CUT1294

<400> 115

cctggettta gccacctctc

20

<210> 116

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: CUT1297

<400> 116

cctgggtgaac acctaggacg

20

<210> 117

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: RYSE0

<400> 117

gacggcacgg ccacgcgttt aaaccgcc

28

<210> 118

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: RYSE19

<400> 118

cccgccaggc gctggggttt aaacaccg

28

<210> 119

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: CUT1226

<400> 119

gacggcacgg ccacgcgttt aaaccgccag tttccacac ggacaccc

48

<210> 120

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: CUT1223

<400> 120

gaattctgac agaaaagccc catccttagg

30