(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년12월11일
(11) 등록번호 10-2741381
(24) 등록일자 2024년12월06일

(51) 국제특허분류(Int. Cl.)
C12N 15/10 (2017.01)
(52) CPC특허분류
C12N 15/1058 (2013.01)
C12N 15/1079 (2013.01)
(21) 출원번호 10-2019-7038679
(22) 출원일자(국제) 2018년06월06일
심사청구일자 2021년06월04일
(85) 번역문제출일자 2019년12월27일
(65) 공개번호 10-2020-0015606
(43) 공개일자 2020년02월12일
(86) 국제출원번호 PCT/US2018/036352
(87) 국제공개번호 WO 2018/226893
국제공개일자 2018년12월13일
(30) 우선권주장
62/515,934 2017년06월06일 미국(US)
(56) 선행기술조사문헌
Metabolic Engineering. Vol. 19, 98-106(2013)
(뒷면에 계속)

(73) 특허권자
지머젠 인코포레이티드
미국 94608 캘리포니아 에머리빌 치론 웨이 5300
(72) 발명자
메이슨 벤자민
미국 94608 캘리포니아 에머리빌 스위트 105 호턴 스트리트 5980 지머젠 인코포레이티드 내
고라노프 알렉시
미국 94608 캘리포니아 에머리빌 스위트 105 호턴 스트리트 5980 지머젠 인코포레이티드 내
(뒷면에 계속)
(74) 대리인
특허법인(유한)케이비케이

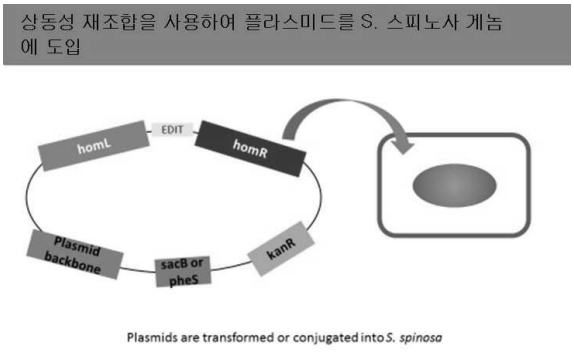
전체 청구항 수 : 총 13 항 심사관 : 최지다

(54) 발명의 명칭 **사카로폴리스포라 스피노사를 개량하기 위한 고 처리량(HTP) 게놈 공학 플랫폼**

(57) 요 약

본 발명은 연산으로 구동되며 분자 생물학, 자동화 및 고급 기계 학습 프로토콜을 통합하는 사카로폴리스포라 종을 위한 HTP 미생물 게놈 공학 플랫폼을 제공한다. 이러한 통합 플랫폼은 특히 과학적 통찰력과 반복적인 패턴 인식에서 유래된 HTP 유전자 디자인 라이브러리를 생성하기 위한 HTP 분자 도구 세트의 모음을 사용한다.

대 표 도 - 도49a



(72) 발명자

켈리 피터

미국 94608 캘리포니아 에머리빌 스위트 105 호턴 스트리트 5980 지머젠 인코포레이티드 내

김 영년

미국 94608 캘리포니아 에머리빌 스위트 105 호턴 스트리트 5980 지머젠 인코포레이티드 내

모디 시탈

미국 94608 캘리포니아 에머리빌 스위트 105 호턴 스트리트 5980 지머젠 인코포레이티드 내

파수마르티 니할

미국 94608 캘리포니아 에머리빌 스위트 105 호턴 스트리트 5980 지머젠 인코포레이티드 내

미짓 벤자민

미국 94608 캘리포니아 에머리빌 스위트 105 호턴 스트리트 5980 지머젠 인코포레이티드 내

엔야트 피터

미국 94608 캘리포니아 에머리빌 스위트 105 호턴 스트리트 5980 지머젠 인코포레이티드 내

(56) 선행기술조사문헌

KR1020010083053 A

Applied biochemistry and biotechnology. Vol. 159, 655-663(2009.01.09.)*

Canadian journal of microbiology. Vol. 55, No. 9, 1070-1075(2009)

Applied microbiology and biotechnology. Vol. 99, 10141-10149(2015)

*는 심사관에 의하여 인용된 문헌

명 세 서

청구범위

청구항 1

다음 단계를 포함하여 원하는 표현형을 획득하기 위해 사카로폴리스포라 종(*Saccharopolyspora sp.*) 미생물을 진화시키는 게놈 공학의 고 처리량(HTP) 방법:

a) 동일한 게놈 균주 배경을 갖는 초기 복수의 사카로폴리스포라 미생물의 게놈을 교란시켜, 독특한 유전자 변이를 갖는 개별 사카로폴리스포라 균주를 포함하는 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리를 생성하는 단계로, 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리는 프로모터 스왑 미생물 균주 라이브러리를 포함하고, 여기서 프로모터 스왑 미생물 균주 라이브러리는 서열번호 1 내지 69 및 172 내지 175로부터 선택된 뉴클레오티드 서열을 갖는 적어도 하나의 프로모터를 포함하는 것인 단계;

b) 원하는 표현형에 대해 단계 a)의 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리의 개별 사카로폴리스포라 균주를 선별 및 선택하는 단계;

c) 선행 단계에서 선별된 적어도 2개의 개별 사카로폴리스포라 균주에 존재하는 유전자 변이로부터 선택된 유전자 변이의 독특한 조합을 각각 포함하는 후속 복수의 사카로폴리스포라 미생물을 제공하여, 후속 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리를 생성하는 단계;

d) 원하는 표현형에 대해 후속 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리의 개별 사카로폴리스포라 균주를 선별 및 선택하는 단계; 및

e) 사카로폴리스포라 미생물이 원하는 표현형을 획득할 때까지 단계 c)-d)를 1회 이상 반복하는 단계, 여기서 각각의 후속 반복은 선행 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리의 적어도 2개의 개별 사카로폴리스포라 균주 중에서 선택된 유전자 변이의 조합인 독특한 유전자 변이를 보유하는 개별 사카로폴리스포라 균주를 포함하는 새로운 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리를 생성한다.

청구항 2

제 1 항에 있어서,

단계 c)에서 유전자 변이가 조합되기 전에 유전자 변이를 함유하는 유전자의 기능 및/또는 신원이 고려되지 않는 것인 게놈 공학의 HTP 방법.

청구항 3

제 1 항에 있어서,

단계 c)에서 조합될 적어도 하나의 유전자 변이가 코딩 DNA 모듈의 반복 세그먼트를 함유하는 게놈 영역에 존재하지 않는 것인 게놈 공학의 HTP 방법.

청구항 4

제 1 항에 있어서,

단계 c)에서 유전자 변이의 독특한 조합을 각각 포함하는 후속 복수의 사카로폴리스포라 미생물은 다음에 의해 생성되는 것인 게놈 공학의 HTP 방법:

1) 플라스미드를 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리에 속하는 개별 사카로폴리스포라 균주에 도입하는 단계로, 여기서 플라스미드는 선택 마커, 카운터 선택 마커, 기본 사카로폴리스포라 균주의 게놈 유전자좌(locus)와 상동성을 갖는 DNA 단편, 및 플라스미드 백본 서열을 포함하며, 여기서 DNA 단편은 또한 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리에 속하는 다른 개별 사카로폴리스포라 균주로부터 유래된 유전자 변이를 갖는다;

2) 게놈에서 선택 마커의 존재에 기초하여 통합 이벤트를 갖는 사카로폴리스포라 균주를 선택하는 단계; 및

3) 카운터 선택 마커 유전자의 부재에 기초하여 루프 아웃된 플라스미드 백본을 갖는 사카로폴리스포라 균주를 선택하는 단계.

청구항 5

제 4 항에 있어서,

플라스미드가 온도 민감성 레플리콘을 포함하지 않는 것인 게놈 공학의 HTP 방법.

청구항 6

제 4 항에 있어서,

선택하는 단계 3)이 통합된 플라스미드의 복제없이 수행되는 것인 게놈 공학의 HTP 방법.

청구항 7

제 1 항에 있어서,

후속 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리는:

a) 초기 HTP 유전자 디자인 미생물 균주 라이브러리의 완전한 조합(full combinatorial) 사카로폴리스포라 균주 라이브러리 또는 이의 서브세트;

b) 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리의 유전자 변이로부터 유래된 완전한 조합 사카로폴리스포라 균주 라이브러리의 서브세트;

c) 선행 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리의 유전자 변이로부터 유래된 완전한 조합 미생물 균주 라이브러리 또는 이의 서브세트; 또는

d) 선행 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리의 유전자 변이로부터 유래된 완전한 조합 사카로폴리스포라 균주 라이브러리의 서브세트인 게놈 공학의 HTP 방법.

청구항 8

제 1 항에 있어서,

단계 a)에서 초기 복수의 사카로폴리스포라 미생물이:

a) 생산 사카로폴리스포라 균주로부터 유래된 독특한 유전자 변이; 또는

b) S_1Gen_1으로 표시된 생산 균주 미생물 및 S_nGen_n으로 표시된 이로부터 유래된 임의의 수의 후속 미생물 세대를 포함하며, 여기서 n은 세대 번호에 해당하는 것인 게놈 공학의 HTP 방법.

청구항 9

제 1 항에 있어서,

단계 c)는 원형질체 융합 기술을 사용하여 유전자 변이를 신속하게 통합시키는 것을 포함하는 것인 게놈 공학의 HTP 방법.

청구항 10

삭제

청구항 11

제 1 항의 방법에 의해 생성된 사카로폴리스포라 균주 라이브러리로, 여기서 라이브러리의 각각의 사카로폴리스포라 균주는:

숙주 세포의 내인성 유전자에 작동 가능하게 연결된 프로모터로, 여기서 프로모터는 내인성 유전자와 이종이며, 프로모터는 서열번호 1-69 및 172-175로 이루어진 그룹으로부터 선택된 서열을 갖는 것인 프로모터

를 포함하는 것인 사카로폴리스포라 균주 라이브러리.

청구항 12

제 1 항에 있어서,

초기 사카로폴리스포라 미생물은 생산 사카로폴리스포라 균주이며, 획득하고자 하는 원하는 표현형은 개량된 표현형 성능이고, 여기서:

a) 단계 c)-d)는 후속 라이브러리의 사카로폴리스포라 균주의 표현형 성능이 생산 사카로폴리스포라 균주의 표현형 성능과 비교하여 측정된 표현형 변수에서 적어도 10% 증가 또는 적어도 1배 증가를 나타낼 때까지 반복된다;

b) 개량된 표현형 성능이 관심 생성물의 부피 생산성, 관심 생성물의 비 생산성, 관심 생성물의 수율, 관심 생성물의 역가 및 이의 조합으로 이루어진 그룹으로부터 선택된다; 또는

c) 개량된 표현형 성능이 관심 생성물의 증가된 또는 보다 효율적인 생산이며, 상기 관심 생성물은 소분자, 효소, 펩타이드, 아미노산, 유기산, 합성 화합물, 연료, 알코올, 1차 세포외 대사산물, 2차 세포외 대사산물, 세포내 성분 분자 및 이들의 조합으로 이루어진 그룹으로부터 선택되는 것인 게놈 공학의 HTP 방법.

청구항 13

제 12 항에 있어서,

관심 생성물은 스피노신, 스피노사드, 스피네토람, 제니스테인, 콜린 옥시다제, 쿠마미딘 화합물, 에리스로마이신, 이버멕틴 아글리콘, HMG-CoA 환원효소 억제제, 카복실산 이성질체, 알파-메틸 메티오닌, 티알리신, 알파-케토부티레이트, 아스파테이트 하이드록사메이트, 아자세린, 5-플루오로인돌, 베타-하이드록시노르발린, 세룰레닌, 퓨린, 피리미딘 및 이의 유사체로 이루어진 그룹으로부터 선택되는 것인 게놈 공학의 HTP 방법.

청구항 14

제 13 항에 있어서,

스피노신은 스피노신 A, 스피노신 D, 스피노신 J, 스피노신 L 또는 이의 조합인 게놈 공학의 HTP 방법.

청구항 15

삭제

청구항 16

삭제

청구항 17

삭제

청구항 18

삭제

청구항 19

삭제

청구항 20

삭제

청구항 21

삭제

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

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

청구항 147

삭제

청구항 148

삭제

청구항 149

삭제

청구항 150

삭제

청구항 151

삭제

청구항 152

삭제

청구항 153

삭제

청구항 154

삭제

청구항 155

삭제

청구항 156

삭제

청구항 157

삭제

청구항 158

삭제

청구항 159

삭제

청구항 160

삭제

청구항 161

삭제

청구항 162

삭제

청구항 163

삭제

청구항 164

삭제

청구항 165

삭제

청구항 166

삭제

청구항 167

삭제

청구항 168

삭제

청구항 169

삭제

청구항 170

삭제

청구항 171

삭제

청구항 172

삭제

청구항 173

삭제

청구항 174

삭제

청구항 175

삭제

청구항 176

삭제

청구항 177

삭제

청구항 178

삭제

청구항 179

삭제

청구항 180

삭제

청구항 181

삭제

청구항 182

삭제

청구항 183

삭제

청구항 184

삭제

청구항 185

삭제

청구항 186

삭제

청구항 187

삭제

청구항 188

삭제

청구항 189

삭제

청구항 190

삭제

청구항 191

삭제

청구항 192

삭제

청구항 193

삭제

청구항 194

삭제

청구항 195

삭제

청구항 196

삭제

청구항 197

삭제

청구항 198

삭제

청구항 199

삭제

청구항 200

삭제

청구항 201

삭제

청구항 202

삭제

청구항 203

삭제

청구항 204

삭제

청구항 205

삭제

청구항 206

삭제

청구항 207

삭제

청구항 208

삭제

청구항 209

삭제

청구항 210

삭제

청구항 211

삭제

청구항 212

삭제

청구항 213

삭제

청구항 214

삭제

청구항 215

삭제

발명의 설명

기 술 분 야

[0001] 본 발명은 고 처리량(HTP) 미생물 게놈 공학에 관한 것이다. 개시된 HTP 게놈 공학 플랫폼은 연산으로 구동되며 분자 생물학, 자동화 및 고급 기계 학습 프로토콜을 통합한다. 이 통합 플랫폼은 일련의 HTP 분자 도구 세트를 사용하여 특히 과학적 통찰력과 반복적인 패턴 인식에서 유래된 HTP 유전자 디자인 라이브러리를 생성한다. 특히, 교시된 플랫폼은 지금까지 다루기 어려운 미생물 종에서 HTP 미생물 게놈 공학을 수행할 수 있다.

배 경 기 술

[0002] 인간은 수천 년 동안 미생물 세포 생합성 경로의 힘을 이용하여 관심있는 제품을 생산하는데, 이런 제품의 가장 오래된 예는 알코올, 식초, 치즈 및 요구르트를 포함한다. 이런 제품은 오늘날에도 여전히 많은 수요가 있으며 미생물에 의해 생산 가능한 제품의 레퍼토리가 계속 증가하고 있다. 유전 공학 기술의 도래는 과학자들이 다양한 생물체로 새로운 생합성 경로를 디자인하고 프로그램하여 광범위한 산업, 의료 및 소비자 제품을 생산하게 하였다. 사실은, 미생물 세포 배양은 이제 소분자, 항생제, 백신, 살충제, 효소, 연료 및 산업 화학물질에 이르는 제품을 생산하는데 사용된다.

[0003] 현대의 산업 미생물에 의해 생산되는 다수의 제품을 고려하면, 소정의 미생물이 표적 제품을 생산할 수 있는 속도 및 효율을 향상시키기 위해 엔지니어가 엄청난 압력을 받고 있는다는 것은 놀랄 일이 아니다.

[0004] 관련된 미생물을 "개량"함으로써 생물학적-기반 산업 공정의 경제를 개량하기 위해 다양한 접근법이 사용되어왔다. 예를 들어, 많은 제약 및 화학 산업은 미생물 배양물의 부모 균주가 화학물질 또는 자외선에 대한 노출을 통해 지속적으로 돌연변이되고 뒤이어 생산성, 수율 및 역가와 같은 성능 개량에 대해 선별되는 미생물 균주 개량 프로그램에 의존한다. 이런 돌연변이유발 과정은 균주가 제품 성능을 적절하게 증가시킬 때까지 광범위하게 반복된다. 다음의 "개량된" 균주는 상업적 생산에 이용된다.

[0005] 상기에서 언급한 바와 같이, 돌연변이유발을 통한 개량된 산업 미생물 균주의 동정은시간 소모적이고 비효율적이다. 이 과정은, 본질적으로, 계획성이 없으며 제품 생산에 바람직한 결과를 가진 돌연변이 발생에 의존한다.

[0006] 전통적인 미생물 균주 개량 프로그램은 비효율적일뿐만 아니라, 이 공정은 또한 높은 수준의 유해한 돌연변이유발성 부하를 갖는 산업용 균주를 유도할 수 있다. 이러한 유형의 프로그램에 적용되는 산업용 균주에서 돌연변이의 축적은 중요해질수 있으며 결과적으로 성능 개량율에 궁극적인 정체를 초래할 수 있다.

[0007] 이는 특히 많은 연구자들이 "다루기 힘든" 것으로 여기는 미생물, 즉 전통적인 균주 공학 도구를 이용할 수 없거나 또는 간단히 작동하지 않는 유기체에 대한 문제이다. 일단 이러한 그룹인 사카로폴리스포라 종(*saccharopolyspora spp.*)은 조작하기 어렵기로 악명이 높은 유기체이다. 이는 광범위한 연구가 수행되고 게놈 공학 도구가 쉽게 이용가능한 모델 시스템 미생물과 비교하여, 사카로폴리스포라 종에 대한 많은 중요한 도구들이 아직 만들어지고, 테스트되고, 및/또는 개량되지 않았기 때문이다.

[0008] 따라서, 사카로폴리스포라 종은 생산 목적으로 미생물을 개량하려는 연구자들에게 독특한 도전을 제시한다. 이러한 도전은 사카로폴리스포라 종의 게놈 공학 분야를 방해하고, 연구자들이 이 미생물 시스템의 모든 잠재력을 활용하지 못하도록 막았다.

발명의 내용

해결하려는 과제

[0009] 따라서, 종래의 균주 개량 프로그램에 내재된 상기한 단점을 겪지 않고 유익한 돌연변이를 발견하고 통합하는 공정을 크게 가속화시키는 산업용 미생물을 조작하는 새로운 방법에 대한 기술이 절실히 필요하다.

[0010] 또한, 미생물 균주 개량 분야에서 현재 사용되는 낡고 해로운 공정에 의해 개발된 산업용 균주를 "재건"시키는

방법이 절실히 필요하다.

[0011] 또한, 당업계는 전통적으로 다루기 어려운 미생물 종에서 HTP 게놈 공학 공정을 수행할 수 있는 도구 및 공정에 필사적이다. 현재 HTP 게놈 공학 공정이 이용 가능하지 않은 미생물 종의 속은 사카로폴리스포라 종이다.

과제의 해결 수단

[0012] 본 발명은 전통적인 미생물 균주 개량 프로그램과 관련된 많은 문제점을 겪지 않는 고 처리량(HTP) 미생물 게놈 공학 플랫폼을 제공한다.

[0013] 또한, 본 발명에 교시된 HTP 플랫폼은 수십 년의 랜덤 돌연변이유발에 기초한 균주 개량 프로그램을 통해 비 유익한 돌연변이를 축적하는 산업용 미생물을 재건시킬 수 있다.

[0014] 본 발명에 기술된 HTP 플랫폼은 연구자들이 전통적으로 다루기 어려운 미생물 유기체에서 HTP 게놈 공학을 수행할 수 있게 하는 새로운 미생물 공학 도구 및 공정을 제공한다. 예를 들어, 교시된 플랫폼은 사카로폴리스포라 종에서 HTP 게놈 공학을 가능하게하는 최초의 플랫폼이다. 지금까지 이 유기체 그룹은 HTP 게놈 공학에 적합하지 않았다. 결과적으로, 개시된 플랫폼은 이 유기체 시스템에서 게놈 공학 분야를 혁신할 것이다.

[0015] 개시된 HTP 게놈 공학 플랫폼은 연산적으로 구동되고 분자 생물학, 자동화 및 진보된 기계 학습 프로토콜을 통합한다. 이런 통합 플랫폼은 특히 과학적 통찰력과 반복적인 패턴 인식에서 유래된 HTP 유전자 디자인 라이브러리를 제작하기 위해 HTP 분자 도구 세트를 사용한다.

[0016] 교시된 HTP 유전자 디자인 라이브러리는 미생물에서 테스트하기 위한 특정 게놈 변경의 라이브러리를 제공함으로써 게놈 공학 과정의 동인으로서 기능한다. 특정 라이브러리 또는 라이브러리의 조합을 이용하여 조작된 미생물은 최종 결과, 예를 들어, 관심있는 제품의 생산을 위한 HTP 방식으로 효율적으로 선별된다. 미생물에서 테스트하기 위한 특정 게놈 변경을 정의하기 위해 HTP 유전자 디자인 라이브러리를 이용하고 이어서 변경을 제공하는 숙주 미생물 게놈을 선별하는 이런 방법은 효율적이고 반복적인 방식으로 수행된다. 일부 양태에서, 게놈 공학 캠페인의 반복 주기 또는 "라운드"는 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100회 이상의 반복/사이클/라운드일 수 있다.

[0017] 따라서, 일부 양태에서, 본 발명은 HTP 유전 공학의 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 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, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, 700, 725, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975, 1000 이상의 "라운드"(예를 들어, SNP 스왑, PRO 스왑, STOP 스왑 또는 이의 조합의 라운드)를 실행하는 방법을 교시한다.

[0018] 일부 실시태양에서, 본 발명은 각각의 후속적인 HTP 유전 공학 라운드가 이전 라운드의 유전 공학에서 확인된 유전적 변이를 기초로 하는 선형 접근법을 교시한다. 다른 실시태양에서, 본 발명은 각각의 후속적인 HTP 유전 공학 라운드가 미리 실행된 분석 및 별도의 HTP 유전 공학 브랜치를 포함하는 유전 공학의 임의의 이전 라운드에서 확인된 유전자 변이를 기초로 하는 비선형 접근법을 교시한다.

[0019] 이런 반복 사이클로부터의 데이터는 대규모 데이터 분석 및 패턴 인식을 가능하게 하며, 이는 통합 플랫폼에 의해 이용되어 HTP 유전자 디자인 라이브러리 구현의 후속 라운드를 통지한다. 결과적으로, 교시된 플랫폼에서 이용된 HTP 유전자 디자인 라이브러리는 대규모 데이터 패턴 인식 알고리즘의 이점을 누리고 미생물 공학의 각각의 반복적인 라운드를 통해 더욱 유익해지는 매우 역동적인 도구이다.

[0020] 일부 실시태양에서, 본 발명의 유전자 디자인 라이브러리는 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 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, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, 700, 725, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975, 1000회 이상의 개별적인 유전자 변화(예를 들어, PRO 스왑 라이브러리에서 프로모터:유전자 조합의 적어도x번호를 포함한다).

[0021] 일부 실시태양에서, 본 발명은 미생물 균주에 대한 HTP 균주 개량 방법의 적용을 설명하는 예시적인 실시예 및 본문을 제공한다. 일부 실시태양에서, 본 발명의 균주 개량 방법은 임의의 숙주 세포에 적용가능하다.

[0022] 일부 실시태양에서, 본 발명은 다음 단계를 포함하여 원하는 표현형을 획득하기 위해 미생물을 진화시키는 게놈 공학의 고 처리량(HTP) 방법을 교시한다: a)

[0023] 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리로서 교란된 게놈을 갖는 초기 복수의 사카로폴리스포라 미생물의 게놈을 수득하는 단계, 여기서 복수의 사카로폴리스포라 미생물은 동일한 게놈 균주 배경을 가지며, 이에 의해 초기 HTP 유전자 디자인을 생성하고, 사카로폴리스포라 균주 라이브러리는 독특한 유전자 변이를 갖는 개별 사카로폴리스포라 균주를 포함한다; b) 원하는 표현형에 대해 초기 HTP 유전자 디자인 미생물 균주 라이브러리의 개별 미생물 균주를 선별 및 선택하는 단계; c) 선행 단계에서 선별된 적어도 2개의 개별 미생물 균주에 존재하는 유전자 변이로부터 선택된 유전자 변이의 독특한 조합을 각각 포함하는 후속 복수의 미생물을 제공하여 후속 HTP 유전자 디자인 미생물 균주 라이브러리를 생성하는 단계; d) 원하는 표현형에 대해 후속 HTP 유전자 디자인 미생물 균주 라이브러리의 개별 미생물 균주를 선별 및 선택하는 단계; e) 미생물이 원하는 표현형을 획득할 때까지 선형 또는 비선형 방식으로 단계 c)-d)를 1회 이상 반복하는 단계, 여기서 각각의 후속 반복은 선행 HTP 유전자 디자인 미생물 균주 라이브러리의 적어도 2개의 개별 미생물 균주 중에서 선택된 유전자 변이의 조합인 독특한 유전자 변이를 보유하는 개별 미생물 균주를 포함하는 새로운 HTP 유전자 디자인 미생물 균주 라이브러리를 생성한다.

[0024] 유전자 변이가 조합될 때, 유전자 변이를 함유하는 유전자의 기능 및/또는 신원은 고려되거나 고려되지 않을 수 있다. 일부 실시태양에서, 유전자 변이를 함유하는 유전자의 기능 및/또는 신원은 고려되지 않는다. 예를 들어, 동일한 유전자 또는 유사한 기능/구조를 갖는 유전자의 유전자 변이가 조합을 위해 선택된다. 일부 실시태양에서, 유전자 변이를 함유하는 유전자의 기능 및/또는 신원은 유전자 변이가 조합되기 전에 고려되지 않는다. 어느 경우에나, 이후의 선별 및 선택 단계는 원하는 표현형, 예를 들어 관심 생성물의 개량된 생산을 갖는 조작된 사카로폴리스포라 균주를 확인하기 위해 수행될 수 있다.

[0025] 일부 실시태양에서, 유전자 변이는 관심 생성물의 직접적인 합성 또는 대사와 관련된 하나 이상의 유전자좌, 또는 합성 또는 대사의 조절과 관련된 유전자좌이다. 일부 실시태양에서, 유전자 변이는 관심 생성물의 직접적인 합성 또는 대사와 관련이 없고 합성 또는 대사의 조절과 관련이 없는 하나 이상의 유전자좌에 존재한다. 일부 실시태양에서, 유전자 변이는 바람직한 이들의 기능 또는 특정 게놈 조합 구조에 대한 어떠한 특정 가설 없이 조합에 대해 랜덤으로 선택된다. 예를 들어, 일부 실시태양에서, 조합의 목적은 DNA 모듈의 반복 세그먼트, 예를 들어, 폴리케타이드 또는 비-리보솜 펩타이드를 코딩하는 유전자의 세그먼트를 함유하는 게놈 영역에서 DNA 모듈을 대체하는 것이 아니다.

[0026] 일부 실시태양에서, 상이한 공급원으로부터의 유전자 변이가 조합되는 상기 방법의 단계 (c)에서, 다양한 기술이 사용될 수 있다. 일부 실시태양에서, 상동성 재조합 플라스미드 시스템이 사용된다. 일부 실시태양에서, 단계 (c)에서 유전자 변이의 독특한 조합을 각각 포함하는 사카로폴리스포라 미생물은 다음에 의해 제조된다: 1) 플라스미드를 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리에 속하는 개별 사카로폴리스포라 균주에 도입하는 단계, 여기서 플라스미드는 (i) 선택 마커, (ii) 카운터 선택 마커, (iii) 기본 사카로폴리스포라 균주의 게놈 유전자좌와 상동성을 갖는 DNA 단편, 및 플라스미드 백본 서열을 포함하며, 여기서 DNA 단편은 또한 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리에 속하는 다른 개별 사카로폴리스포라 균주로부터 유래된 유전자 변이를 가진다; 2) 게놈에서 선택 마커의 존재에 기초하여 통합 이벤트를 갖는 사카로폴리스포라 균주를 선택하는 단계; 3) 카운터 선택 마커 유전자의 부재에 기초하여 루프 아웃된 플라스미드 백본을 갖는 사카로폴리스포라 균주를 선택하는 단계.

[0027] 일부 실시태양에서, 본 개시의 방법은 게놈 모듈성의 이러한 영역에서 표적화된 게놈 편집을 수행할 수 있을 뿐만 아니라, 임의의 게놈 상황에서 게놈에 걸쳐 표적화된 게놈 편집을 가능하게 한다. 결과적으로, 본 개시의 표적화된 게놈 편집은 임의의 영역에서 S. 스피노사 게놈을 편집할 수 있으며, 모듈성을 갖는 영역에서의 편집에만 국한되지 않는다.

[0028] 일부 실시태양에서, 플라스미드는 온도 민감성을 포함하지 않는다.

[0029] 일부 실시태양에서, 선택 단계 3)은 통합된 플라스미드의 복제없이 수행된다.

[0030] 일부 실시태양에서, 본 발명은 초기 HTP 유전자 디자인 미생물 균주 라이브러리가 프로모터 스왑 미생물 균주 라이브러리, SNP 스왑 미생물 균주 라이브러리, 개시/중지 코돈 미생물 균주 라이브러리, 최적화된 서열 미생물

균주 라이브러리, 터미네이터 스왑 미생물 균주 라이브러리, 트랜스포존 돌연변이유발 다양성 라이브러리, 리보솜 결합 부위 미생물 균주 라이브러리, 항-대사산물/발효 생성물 저항성 미생물 라이브러리, 또는 이의 임의의 조합으로 이루어진 그룹으로부터 선택된 적어도 하나인 것을 교시한다. 일부 실시태양에서, 상기 미생물 라이브러리는 사카로폴리스포라 종 라이브러리이다.

[0031] 일부 실시태양에서, 본 발명은 유전자 변이의 독특한 조합을 각각 포함하는 후속 복수의 미생물을 제조하는 방법을 교시하며, 결합된 유전자 변이의 각각은 초기 HTP 유전자 디자인 미생물 균주 라이브러리 또는 선행 단계의 HTP 유전자 디자인 미생물 균주 라이브러리로부터 유도된다.

[0032] 일부 실시태양에서, 후속 복수의 미생물에서의 유전자 변이의 조합은 초기 HTP 유전자 디자인 미생물 균주 라이브러리 또는 선행 단계의 HTP 유전자 디자인 미생물 균주 라이브러리에서 유전자 변이의 가능한 모든 조합의 서브세트를 포함할 것이다.

[0033] 일부 실시태양에서, 본 발명은 후속 HTP 유전자 디자인 미생물 균주 라이브러리가 초기 HTP 유전자 디자인 미생물 균주 라이브러리 또는 선행 단계의 HTP 유전자 디자인 미생물 균주 라이브러리에서 유전자 변이로부터 유래된 완전한 조합 미생물 균주 라이브러리라는 것을 교시한다.

[0034] 예를 들어, 이전의 HTP 유전자 디자인 미생물 균주 라이브러리가 유전자 변이 A, B, C 및 D만을 갖는다면, 상기 변이의 부분 조합은 각각 AB, AC 또는 AD의 독특한 유전자 변이 조합(돌연변이가 나타나는 순서는 중요하지 않음)을 포함하는 3개 미생물을 포함하는 후속 HTP 유전자 디자인 미생물 균주 라이브러리를 포함할 수 있다. 선형 단계의 HTP 유전자 디자인 라이브러리의 유전자 변이로부터 유래된 완전한 조합 미생물 균주 라이브러리는 각각 AB, AC, AD, BC, BD 또는 CD의 독특한 유전자 변이 조합을 포함하는 6개의 미생물을 포함할 것이다.

[0035] 일부 실시태양에서, 본 발명의 방법은 랜덤 돌연변이유발, 표적 서열 삽입, 표적 서열 결실, 표적 서열 교체, 트랜스포존 돌연변이유발 또는 이의 임의의 조합으로 이루어진 그룹으로부터 선택된 적어도 하나의 방법을 이용하여 게놈을 교란시키는 것을 교시한다.

[0036] 본 발명의 방법의 일부 실시태양에서, 초기 복수의 미생물은 산업용 생산 균주 미생물로부터 유래된 독특한 유전자 변이를 포함한다. 일부 실시태양에서, 미생물은 사카로폴리스포라 종이다.

[0037] 본 발명의 방법의 일부 실시태양에서, 초기 복수의 미생물은 S_1Gen_1으로 표시된 산업용 생산 균주 미생물 및 S_nGen_n으로 표시된 이로부터 유래된 임의의 수의 후속 미생물 세대를 포함한다. 일부 실시태양에서, 미생물은 사카로폴리스포라 종이다.

[0038] 일부 실시태양에서, 본 발명은 다음 단계를 포함하는 SNP 스왑 미생물 균주 라이브러리를 생성하는 방법을 교시한다: a) 기준 미생물 균주 및 제 2 미생물 균주를 제공하는 단계, 여기서 제 2 미생물 균주는 기준 미생물 균주에 존재하지 않는 단일 뉴클레오티드 다형성, DNA 삽입 및 DNA 결실로부터 선택된 복수의 확인된 유전자 변이를 포함한다; b) 기준 미생물 균주 또는 제 2 미생물의 게놈을 교란시켜 상기 복수의 개별 미생물 균주의 각각의 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 미생물 균주를 포함하는 초기 SNP 스왑 미생물 균주 라이브러리를 생성하는 단계, 상기 각각의 독특한 유전자 변이는 기준 미생물 균주와 제 2 미생물 균주 사이의 복수의 확인된 유전자 변이로부터 선택된 단일 유전자 변이에 상응한다. 일부 실시태양에서, 미생물 균주는 사카로폴리스포라 균주이다.

[0039] SNP 스왑 라이브러리의 일부 실시태양에서, 기준 미생물 균주의 게놈은 제 2 미생물 균주에서 발견되는 확인된 단일 뉴클레오티드 다형성, DNA 삽입 또는 DNA 결실의 하나 이상을 첨가하기 위해 교란된다.

[0040] 본 발명의 SNP 스왑 라이브러리 방법의 일부 실시태양에서, 제 2 미생물 균주의 게놈은 기준 미생물 균주에서 발견되지 않은 확인된 단일 뉴클레오티드 다형성, DNA 삽입 또는 DNA 결실의 하나 이상을 제거하기 위해 교란된다.

[0041] 일부 실시태양에서, SNP 스왑 라이브러리의 유전자 변이는 기준 미생물 균주와 제 2 미생물 균주 사이에서 모든 확인된 유전자 변이의 서브세트를 포함할 것이다.

[0042] 일부 실시태양에서, SNP 스왑 라이브러리의 유전자 변이는 기준 미생물 균주와 제 2 미생물 균주 사이에서 모든 확인된 유전자 변이를 포함할 것이다.

[0043] 일부 실시태양에서, 본 발명은 다음 단계를 포함하는 산업용 미생물 균주의 표현형 성능을 재건 및 개량시키는 방법을 교시한다: a) 부모 계통의 미생물 균주 및 이로부터 유래된 산업용 미생물 균주를 제공하는 단계, 여기

서 산업용 미생물 균주는 부모 계통 미생물 균주에 존재하지 않는 단일 뉴클레오티드 다형성, DNA 삽입 및 DNA 결실로부터 선택된 복수의 확인된 유전자 변이를 포함한다; b) 부모 계통 미생물 균주 또는 산업용 미생물 균주의 게놈을 교란시켜 상기 복수의 개별 미생물 균주의 각각의 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 미생물 균주를 포함하는 초기 SNP 스왑 미생물 균주 라이브러리를 생성하는 단계, 여기서 상기 독특한 유전자 변이의 각각은 부모 계통 미생물 균주와 산업용 미생물 균주 사이의 복수의 확인된 유전자 변이로부터 선택된 단일 유전자 변이에 해당한다; c) 기준 미생물 균주에 대한 표현형 성능 개량을 위한 초기 SNP 스왑 미생물 균주 라이브러리의 개별 미생물 균주를 선택 및 선택하여 상기 미생물 균주에 표현형 성능 개량을 부여하는 독특한 유전자 변이를 확인하는 단계; d) 선행 단계에서 선별된 적어도 2개의 개별 미생물 균주에 존재하는 유전자 변이로부터 선택된 유전자 변이의 독특한 조합을 각각 포함하는 후속 복수의 미생물을 제공하여 후속 SNP 스왑 미생물 균주 라이브러리를 생성하는 단계; e) 기준 미생물 균주에 대한 표현형 성능 개량을 위한 후속 SNP 스왑 미생물 균주 라이브러리의 개별 미생물 균주 선택 및 선택하여 상기 미생물 균주에 추가 표현형 성능 개량을 부여하는 유전자 변이의 독특한 조합을 확인하는 단계; 및 f) 미생물 균주가 산업용 미생물 균주의 표현형 성능에 비해 원하는 수준의 개량된 표현형 성능을 나타낼 때까지 선형 또는 비선형 방식으로 단계 d)-e)를 1회 이상 반복하는 단계, 각각의 후속 반복은 선행 SNP 스왑 미생물 균주 라이브러리의 적어도 2개의 개별 미생물 균주 중에서 선택된 유전자 변이의 조합인 독특한 유전자 변이를 보유하는 개별 미생물 균주를 포함하는 새로운 SNP 스왑 미생물 균주 라이브러리를 생성한다. 일부 실시태양에서, 미생물 균주는 사카로폴리스포라 균주이다.

[0044] 일부 실시태양에서, 본 발명은 산업용 미생물 균주의 표현형 성능을 재건 및 개량시키는 방법을 교시하며, 여기서 부모 계통 미생물 균주의 게놈은 산업용 미생물 균주에서 발견되는 확인된 단일 뉴클레오티드 다형성, DNA 삽입 또는 DNA 결실의 하나 이상을 첨가하기 위해 교란된다. 일부 실시태양에서, 미생물 균주는 사카로폴리스포라 균주이다.

[0045] 일부 실시태양에서, 본 발명은 산업용 미생물 균주의 표현형 성능을 재건 및 개량시키는 방법을 교시하며, 여기서 산업 미생물 균주의 게놈은 부모 계통 미생물 균주에서 발견되지 않는 확인된 단일 뉴클레오티드 다형성, DNA 삽입 또는 DNA 결실의 하나 이상을 제거하기 위해 교란된다. 일부 실시태양에서, 미생물 균주는 사카로폴리스포라 균주이다.

[0046] 일부 실시태양에서, 본 발명은 프로모터 스왑 미생물 균주 라이브러리를 생성하는 방법을 교시하며, 이 방법은 다음 단계를 포함한다: a) 기본 미생물 균주에 내인성인 복수의 표적 유전자 및 프로모터 래더를 제공하는 단계, 여기서 상기 프로모터 래더는 기본 미생물 균주에서 상이한 발현 프로파일을 나타내는 복수의 프로모터를 포함한다; b) 기본 미생물 균주의 게놈을 조작하여 상기 복수의 개별 미생물 균주의 각각의 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 미생물 균주를 포함하는 초기 프로모터 스왑 미생물 균주 라이브러리를 생성하는 단계, 여기서 상기 독특한 유전자 변이의 각각은 기본 미생물 균주에 내인성인 표적 유전자의 하나에 작동 가능하게 연결된 프로모터 래더로부터의 프로모터의 하나를 포함한다. 일부 실시태양에서, 미생물 균주는 사카로폴리스포라 균주이다. 일부 실시태양에서, 프로모터 래더는 서열번호 1 내지 서열번호 69의 서열을 갖는 프로모터 또는 이의 조합을 포함한다.

[0047] 일부 실시태양에서, 본 발명은 원하는 표현형을 획득하기 위해 미생물을 진화시키는 게놈 공학의 프로모터 스왑 방법을 교시하며, 이 방법은 다음 단계를 포함한다: a) 기본 미생물 균주에 내인성인 복수의 표적 유전자 및 프로모터 래더를 제공하는 단계, 여기서 상기 프로모터 래더는 상기 기본 미생물 균주에서 상이한 발현 프로파일을 나타내는 복수의 프로모터를 포함한다; b) 기본 미생물 균주의 게놈을 조작하여 상기 복수의 개별 미생물 균주의 각각의 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 미생물 균주를 포함하는 초기 프로모터 스왑 미생물 균주 라이브러리를 생성하는 단계, 상기 독특한 유전자 변이의 각각은 기본 미생물 균주에 내인성인 표적 유전자의 하나에 작동 가능하게 연결된 프로모터 래더로부터의 프로모터의 하나를 포함한다; c) 원하는 표현형에 대한 초기 프로모터 스왑 미생물 균주 라이브러리의 개별 미생물 균주를 선택 및 선택하는 단계; d) 선행 단계에서 선별된 적어도 2개의 개별 미생물 균주에 존재하는 유전자 변이로부터 선택된 유전자 변이의 독특한 조합을 각각 포함하는 후속 복수의 미생물을 제공하여 후속 프로모터 스왑 미생물 균주 라이브러리를 생성하는 단계; e) 원하는 표현형에 대한 후속 프로모터 스왑 미생물 균주 라이브러리의 개별 미생물 균주를 선택 및 선택하는 단계; f) 미생물이 원하는 표현형을 획득할 때까지 선형 또는 비선형 방식으로 1회 이상 단계 d)-e)를 반복하는 단계, 여기서 각각의 후속 반복은 선행 프로모터 스왑 미생물 균주 라이브러리의 적어도 2개의 개별 미생물 균주 중에서 선택된 유전자 변이의 조합인 유전자 변이를 제공하는 개별 미생물 균주를 포함하는 새로운 프로모터 스왑 미생물 균주 라이브러리를 생성한다. 일부 실시태양에서, 미생물 균주는 사카로폴리스포

라 균주이다.

[0048] 일부 실시태양에서, 본 발명은 터미네이터 스왑 미생물 균주 라이브러리를 생성하는 방법을 교시하며, 이 방법은 다음 단계를 포함한다: a) 기본 미생물 균주에 내인성인 복수의 표적 유전자 및 터미네이터 래더를 제공하는 단계, 여기서 상기 터미네이터 래더는 기본 미생물 균주에서 상이한 발현 프로파일을 나타내는 복수의 터미네이터를 포함한다; b) 기본 미생물 균주의 게놈을 조작하여 상기 복수의 개별 미생물 균주의 각각의 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 미생물 균주를 포함하는 초기 터미네이터 스왑 미생물 균주 라이브러리를 생성하는 단계, 상기 독특한 유전자 변이의 각각은 터미네이터 래더로부터의 터미네이터의 하나 이상에 작동 가능하게 연결된 기본 미생물 균주에 내인성인 표적 유전자의 하나를 포함한다. 일부 실시태양에서, 미생물 균주는 사카로폴리스포라 균주이다.

[0049] 일부 실시태양에서, 본 발명은 원하는 표현형을 획득하기 위해 미생물을 진화시키는 게놈 공학의 터미네이터 스왑 방법을 교시하며, 이 방법은 다음 단계를 포함한다: a) 기본 미생물 균주에 내인성인 복수의 표적 유전자 및 터미네이터 래더를 제공하는 단계; 여기서 상기 터미네이터 래더는 상기 기본 미생물 균주에서 상이한 발현 프로파일을 나타내는 복수의 터미네이터를 포함한다; b) 기본 미생물 균주의 게놈을 조작하여 상기 복수의 개별 미생물 균주의 각각의 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 미생물 균주를 포함하는 초기 터미네이터 스왑 미생물 균주 라이브러리를 생성하는 단계, 상기 독특한 유전자 변이의 각각은 터미네이터 래더로부터의 터미네이터의 하나 이상에 작동 가능하게 연결된 기본 미생물 균주에 내인성인 표적 유전자의 하나를 포함한다; c) 원하는 표현형에 대한 초기 터미네이터 스왑 미생물 균주 라이브러리의 개별 미생물 균주를 선택 및 선택하는 단계; d) 선행 단계에서 선별된 적어도 2개의 개별 미생물 균주에 존재하는 유전자 변이로부터 선택된 유전자 변이의 독특한 조합을 각각 포함하는 후속 복수의 미생물을 제공하여 후속 터미네이터 스왑 미생물 균주 라이브러리를 생성하는 단계; e) 원하는 표현형에 대한 후속 터미네이터 스왑 미생물 균주 라이브러리의 개별 미생물 균주를 선택 및 선택하는 단계; f) 미생물이 원하는 표현형을 획득할 때까지 선형 또는 비선형 방식으로 1회 이상 단계 d)-e)를 반복하는 단계, 여기서 각각의 후속 반복은 선행 터미네이터 스왑 미생물 균주 라이브러리의 적어도 2개의 개별 미생물 균주 중에서 선택된 유전자 변이의 조합인 유전자 변이를 제공하는 개별 미생물 균주를 포함하는 새로운 터미네이터 스왑 미생물 균주 라이브러리를 생성한다. 일부 실시태양에서, 미생물 균주는 사카로폴리스포라 균주이다. 일부 실시 태양에서, 터미네이터 래더는 서열번호 70 내지 서열번호 80의 서열을 갖는 터미네이터 또는 이의 조합을 포함한다.

[0050] 일부 실시태양에서, 본 발명은 미생물이 원하는 표현형을 획득하도록 진화하도록 게놈 공학의 트랜스포존 돌연변이유발 방법을 교시하며, 상기 방법은 a) 트랜스포사제(transposase) 효소 및 DNA 페이로드 서열을 제공하는 단계를 포함한다. 일부 실시태양에서, 트랜스포사제는 사카로폴리스포라 종에서 작동한다. 일부 실시태양에서, 트랜스포사제는 EZ-Tn5 트랜스포존 시스템으로부터 유래된다. 일부 실시태양에서, DNA 페이로드 서열은 상기 트랜스포사제에 의해 인식될 수 있는 모자이크 요소(ME)에 의해 플랭킹된다. 일부 실시태양에서, DNA 페이로드는 기능상실(LoF) 트랜스포존 또는 기능획득(GoF) 트랜스포존일 수 있다. 일부 실시태양에서, DNA 페이로드는 선택 마커를 포함한다. 일부 실시태양에서, DNA 페이로드는 카운터 선택 마커를 포함한다. 일부 실시태양에서, 카운터 선택 마커는 선택 가능한 마커를 함유하는 DNA 페이로드의 루프-아웃을 용이하게 하기 위해 사용된다. 일부 실시태양에서, GoF 트랜스포존은 GoF 요소를 포함한다. 일부 실시태양에서, GoF 트랜스포존은 프로모터 서열 및/또는 용해도 태그 서열을 포함한다. 일부 실시태양에서, 방법은 b) 트랜스포사제 및 DNA 페이로드 서열을 조합하여 복합체를 형성하는 단계, 및 c) 트랜스포사제-DNA 페이로드 복합체를 미생물 균주로 형질전환시켜, 미생물 균주의 게놈에서 DNA 페이로드 서열의 랜덤 통합을 초래하는 단계를 추가로 포함한다. DNA 페이로드의 랜덤 통합을 포함하는 균주는 초기 트랜스포존 돌연변이유발 다양성 라이브러리를 형성한다. 일부 실시태양에서, 방법은 d) 원하는 표현형에 대한 초기 트랜스포존 돌연변이유발 다양성 라이브러리의 개별 미생물 균주를 선별 및 선택하는 단계를 추가로 포함한다. 일부 실시태양에서, 방법은 e) 각각 유전자 변이의 독특한 조합을 포함하는 후속 복수의 미생물을 제공하는 단계를 추가로 포함하며, 상기 유전자 변이는 선행 단계에서 선별된 적어도 2개의 개별 미생물 균주에 존재하는 유전자 변이로부터 선택되고, 이에 의해 후속 트랜스포존 돌연변이유발 다양성 라이브러리를 생성한다. 일부 실시태양에서, 방법은 f) 원하는 표현형에 대한 후속 트랜스포존 돌연변이유발 다양성 라이브러리의 개별 미생물 균주를 선별 및 선택하는 단계를 추가로 포함한다. 일부 실시태양에서, 상기 방법은 g) 단계 e)-f)를 미생물이 원하는 표현형을 획득할 때까지 선형 또는 비선형 방식으로 1회 이상 반복하는 단계를 추가로 포함하며, 여기서 각각의 후속 반복은 선행 트랜스포존 돌연변이유발 다양성 라이브러리의 적어도 2개의 개별 미생물 균주 중에서 선택된 유전자 변이의 조합인 독특한 유전자 변이를 보유하는 개별 미생물 균주를 포함하는 새로운 트랜스포존 돌연변이유발 다양성 라이브러리를 생성한다. 일부 실시태양에서, 미생물

균주는 사카로폴리스포라 균주이다.

[0051] 일부 실시태양에서, 본 발명은 리보솜 결합 부위(RBS) 스왑 미생물 균주 라이브러리를 생성하는 방법을 교시한다. 일부 실시태양에서, 상기 방법은 다음 단계를 포함한다: a) 기본 미생물 균주에 내인성인 복수의 표적 유전자 및 RBS 래더를 제공하는 단계, 여기서 상기 RBS 래더는 기본 미생물 균주에서 상이한 발현 프로파일을 나타내는 복수의 리보솜 결합 부위를 포함한다; b) 기본 미생물 균주의 게놈을 조작하여 상기 복수의 개별 미생물 균주의 각각의 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 미생물 균주를 포함하는 초기 RBS 미생물 균주 라이브러리를 생성하는 단계, 여기서 상기 독특한 유전자 변이의 각각은 기본 미생물 균주에 내인성인 표적 유전자의 하나에 작동 가능하게 연결된 RBS 래더로부터의 RBS의 하나 이상을 포함한다. 일부 실시태양에서, 미생물 균주는 사카로폴리스포라 균주이다.

[0052] 일부 실시태양에서, 본 발명은 미생물이 원하는 표현형을 수득하도록 진화시키기 위한 게놈 공학의 리보솜 결합 부위(RBS) 스왑 방법을 교시하며, 상기 방법은 다음 단계를 포함한다: a) 기본 미생물 균주에 내인성인 복수의 표적 유전자 및 RBS 래더를 제공하는 단계, 여기서 상기 RBS 래더는 상기 기본 미생물 균주에서 상이한 발현 프로파일을 나타내는 복수의 RBS를 포함한다; b) 기본 미생물 균주의 게놈을 조작하여, 상기 복수의 개별 미생물 균주의 각각의 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 미생물 균주를 포함하는 초기 RBS 라이브러리를 생성하는 단계, 여기서 상기 독특한 유전자 변이의 각각은 기본 미생물 균주에 내인성인 표적 유전자의 하나에 작동 가능하게 연결된 RBS 래더로부터의 RBS의 하나를 포함한다; c) 원하는 표현형에 대한 초기 RBS 라이브러리의 개별 미생물 균주를 선별 및 선택하는 단계; d) 유전자 변이의 독특한 조합을 각각 포함하는 후속 복수의 미생물을 제공하여, 상기 유전자 변이는 선행 단계에서 선별된 적어도 2개의 개별 미생물 균주에 존재하는 유전자 변이로부터 선택되며, 이에 의해 후속 RBS 라이브러리를 생성하는 단계; e) 원하는 표현형에 대한 후속 RBS 라이브러리의 개별 미생물 균주를 선별 및 선택하는 단계; f) 미생물이 원하는 표현형을 획득할 때까지 선형 또는 비선형 방식으로 단계 d)-e)를 1회 이상 반복하는 단계, 여기서 각각의 후속 반복은 선행 RBS 라이브러리의 적어도 2개의 개별 미생물 균주 중에서 선택된 유전자 변형의 조합인 독특한 유전자 변이를 보유하는 개별 미생물 균주를 포함하는 새로운 RBS 라이브러리를 생성한다. 일부 실시태양에서, 미생물 균주는 사카로폴리스포라 균주이다. 일부 실시태양에서, 터미네이터 래더는 서열번호 97 내지 서열번호 127의 서열을 갖는 터미네이터 또는 이의 조합을 포함한다.

[0053] 일부 실시태양에서, 본 발명은 항-대사산물/발효 생성물 저항성 라이브러리를 생성하는 방법을 교시한다. 일부 실시태양에서, 방법은 다음 단계를 포함한다: a) 기준 미생물 균주 및 제 2 미생물 균주를 제공하는 단계, 여기서 제 2 미생물 균주는 복수의 식별 가능한 유전자 변이를 포함하고, 이러한 유전자 변이는 기준 미생물 균주에는 존재하지 않는 단일 뉴클레오티드 다형성, DNA 삽입 및 DNA 결실를 포함하는 임의의 유형일 수 있으나 이에 제한되지 않는다; 및 b) 상기 미생물에 의해 생성된 하나 이상의 소정의 생성물의 존재 하에서 더 저항성인 균주를 선택하는 단계. 일부 실시태양에서, 방법은 c) 선택된 균주의 성능(예를 들어, 균주에서 생성된 하나 이상의 생성물의 수율)을 분석하는 단계 및 HTP 선별에 의한 기준 미생물 균주와 비교하여 개량된 성능을 갖는 균주를 선택하는 단계를 추가로 포함한다. 일부 실시태양에서, 방법은 d) 개량된 성능을 야기하는 돌연변이의 위치 및/또는 서열을 확인하는 단계를 추가로 포함한다. 확인된 개량된 성능을 갖는 이들 선택된 균주는 초기 항-대사산물/발효 생성물 라이브러리를 형성한다. 이러한 라이브러리는 상기 복수의 개별 미생물 균주의 각 균주 내에서 발견되는 독특한 유전자 변이를 갖는 복수의 개별 미생물 균주를 포함하고, 여기서 상기 독특한 유전자 변이의 각각은 복수의 식별 가능한 유전자 변이로부터 선택된 단일 유전자 변이에 상응한다. 일부 실시태양에서, 미생물 균주는 사카로폴리스포라 균주이다. 일부 실시태양에서, 미생물 균주에 의해 생성된 소정의 생성물은 스피노신 합성 경로에 관여하는 임의의 분자, 또는 스피노신의 생성에 영향을 미칠 수있는 임의의 분자이다. 일부 실시태양에서, 소정의 생성물은 스피노신 A, 스피노신 B, 스피노신 C, 스피노신 D, 스피노신 E, 스피노신 F, 스피노신 G, 스피노신 H, 스피노신 I, 스피노신 J, 스피노신 K, 스피노신 L, 스피노신 M, 스피노신 N, 스피노신 O, 스피노신 P, 스피노신 Q, 스피노신 R, 스피노신 S, 스피노신 T, 스피노신 U, 스피노신 V, 스피노신 W, 스피노신 X, 스피노신 Y, 노르류신, 노르발린, 슈도아글리콘(예를 들어, 상이한 스피노신 화합물에 대한 PSA, PSD, PSJ, PSL 등) 및 알파-메틸-메티오닌(aMM)을 포함하나 이에 제한되지 않는다.

[0054] 일부 실시태양에서, 본 발명은 (a) (1) 하나 이상의 배경 미생물 균주에 대한 유전자 변화를 나타내는 입력 및 (2) 상응하는 성능 측정을 포함하는 트레이닝 세트로 채워진 예측 모델에 접근하는 단계; (b) 이런 유전자 변화를 포함하는 후보 미생물 균주에 상응하는 테스트 입력을 유전자 변화를 나타내는 예측 모델에 적용하는 단계; (c) 예측 모델에 적어도 부분적으로 기초하여 후보 미생물 균주의 표현형 성능을 예측하는 단계; (d) 적어도 부분적으로 이들의 예측된 성능에 기초하여 후보 미생물 균주의 제 1 서브세트를 선택하는 단계; (e) 후보 미생물

균주의 제 1 서브세트의 측정된 표현형 성능을 얻는 단계; (f) 적어도 부분적으로 측정된 표현형 성능에 기초하여 후보 미생물 균주의 제 2 서브세트를 선택하는 단계; (g) 예측 모델의 트레이닝 세트에(1) 후보 미생물 균주의 선택된 제 2 서브세트에 상응하는 입력과 함께(2) 후보 미생물 균주의 선택된 제 2 서브세트의 상응하는 측정된 성능을 첨가하는 단계; 및 (h) 적어도 하나의 후보 미생물 균주의 측정된 표현형 성능이 성능 메트릭을 만족시킬 때까지 (b)-(g)를 반복하는 단계에 의해 후보 미생물 균주의 디자인을 반복적으로 개량하는 것을 교시한다. 일부 경우에, 예측 모델에 대한 테스트 입력의 첫 번째 적용 동안, 테스트 입력에 의해 나타낸 유전 자 변화는 하나 이상의 배경 미생물 균주에 대한 유전자 변화를 포함한다; 테스트 입력의 후속 적용 동안, 테스트 입력에 의해 나타낸 유전자 변화는 후보 미생물 균주의 이전에 선택된 제 2 서브세트 내의 후보 미생물 균주에 대한 유전자 변화를 포함한다. 일부 실시태양에서, 미생물 균주는 사카로폴리스포라 균주이다.

[0055] 일부 실시태양에서, 제 1 서브세트의 선택은 상위성 효과(epistatic effect)에 기초할 수 있다. 이것은 제 1 서브세트의 제 1 선택 동안: 하나 이상의 배경 미생물 균주에 대한 유전자 변화를 나타내는 복수의 각각의 입력의 적용에 응답하여 하나 이상의 배경 미생물 균주의 성능 측정 사이의 비유사도를 결정하는 단계; 및 적어도 2개의 후보 미생물 균주에 포함된 유전자 변화의 적용에 응답하여 하나 이상의 배경 미생물 균주의 성능 측정의 비평행성의 정도에 적어도 부분적으로 기초하여 적어도 2개의 후보 미생물 균주를 제 1 서브세트에 포함시키기 위해 선택하는 단계에 의해 성취될 수 있다. 일부 실시태양에서, 미생물 균주는 사카로폴리스포라 균주이다.

[0056] 일부 실시태양에서, 본 발명은 후보 미생물 균주의 반복적 개량에 상위성 효과를 적용하는 것을 교시하고, 상기 방법은 적어도 하나의 미생물 배경 균주에 대해 이루어진 상응하는 유전자 변화에 응답하여 측정된 성능을 나타내는 데이터를 얻는 단계; 적어도 2개의 유전자 변화의 상응하는 반응 성능 측정 사이의 비유사도도에 적어도 부분적으로 기초하여 적어도 2개의 유전자 변화를 선택하는 단계, 여기서 비유사도는 상이한 생물학적 경로를 통해 상응하는 반응성 측정에 영향을 미치는 정도에 관한 것이다; 및 선택된 유전자 변화를 포함한 미생물 배경 균주에 대한 유전자 변화를 디자인하는 단계를 포함한다. 일부 경우에, 적어도 2개의 선택된 유전자 변화가 디자인되는 미생물 배경 균주는 측정된 반응성 성능을 나타내는 데이터가 얻어진 적어도 하나의 미생물 배경 균주와 동일하다. 일부 실시태양에서, 미생물 균주는 사카로폴리스포라 균주이다.

[0057] 일부 실시태양에서, 본 발명은 단일 유형의 유전자 미생물 라이브러리만을 사용하는 HTP 균주 개량 방법을 교시한다. 예를 들어, 일부 실시태양에서, 본 발명은 SNP 스왑 라이브러리만을 사용하는 HTP 균주 개량 방법을 교시한다. 다른 실시태양에서, 본 발명은 PRO 스왑 라이브러리만을 사용하는 HTP 균주 개량 방법을 교시한다. 일부 실시태양에서, 본 발명은 STOP 스왑 라이브러리만을 이용하는 HTP 균주 개량 방법을 교시한다. 일부 실시태양에서, 본 발명은 단지 개시/중지 코돈 스왑 라이브러리를 사용하는 HTP 균주 개량 방법을 교시한다. 일부 실시태양에서, 본 발명은 단지 트랜스포존 돌연변이유발 다양성 라이브러리만을 사용하는 HTP 균주 개량 방법을 교시한다. 일부 실시태양에서, 본 발명은 리보솜 결합 부위 미생물 균주 라이브러리만을 사용하는 HTP 균주 개량 방법을 교시한다. 일부 실시태양에서, 본 발명은 항-대사산물/발효 생성물 저항성 미생물 라이브러리만을 이용하는 HTP 균주 개량 방법을 교시한다. 일부 실시태양에서, 미생물 균주는 사카로폴리스포라 균주이다.

[0058] 다른 실시태양에서, 본 발명은 둘 이상의 유형의 유전자 미생물 라이브러리를 사용하는 HTP 균주 개량 방법을 교시한다. 예를 들어, 일부 실시태양에서, 본 발명은 SNP 스왑 및 PRO 스왑 라이브러리를 조합하는 HTP 균주 개량 방법을 교시한다. 일부 실시태양에서, 본 발명은 SNP 스왑 및 STOP 스왑 라이브러리를 조합하는 HTP 균주 개량 방법을 교시한다. 일부 실시태양에서, 본 발명은 PRO 스왑 및 STOP 스왑 라이브러리를 조합하는 HTP 균주 개량 방법을 교시한다. 일부 실시태양에서, 본 발명은 SNP 스왑 라이브러리를 트랜스포존 돌연변이유발 다양성 라이브러리, 리보솜 결합 부위 미생물 균주 라이브러리 및/또는 항-대사산물 선택/발효 생성물 저항성 미생물 라이브러리와 조합하는 HTP 균주 개량 방법을 교시한다. 일부 실시태양에서, 본 발명은 PRO 스왑 라이브러리를 트랜스포존 돌연변이유발 다양성 라이브러리, 리보솜 결합 부위 미생물 균주 라이브러리 및/또는 항-대사산물 선택/발효 생성물 저항성 미생물 라이브러리와 조합하는 HTP 균주 개량 방법을 교시한다. 일부 실시태양에서, 본 발명은 STOP 스왑 라이브러리를 트랜스포존 돌연변이유발 다양성 라이브러리, 리보솜 결합 부위 미생물 균주 라이브러리 및/또는 항-대사산물 선택/발효 생성물 저항성 미생물 라이브러리와 조합하는 HTP 균주 개량 방법을 교시한다. 일부 실시태양에서, 본 발명은 터미네이터 스왑 라이브러리를 트랜스포존 돌연변이유발 다양성 라이브러리, 리보솜 결합 부위 미생물 균주 라이브러리, 및/또는 항-대사산물 선택/발효 생성물 저항성 미생물 라이브러리와 조합하는 HTP 균주 개량 방법을 교시한다. 일부 실시태양에서, 본 발명은 트랜스포존 돌연변이유발 다양성 라이브러리를 리보솜 결합 부위 미생물 균주 라이브러리 및/또는 항-대사산물 선택/발효 생성물 저항성 미생물 라이브러리와 조합하는 HTP 균주 개량 방법을 교시한다. 일부 실시태양에서, 본 발명은 리보솜 결합 부위 미생물 균주 라이브러리 및 항-대사산물 선택/발효 생성물 저항성 미생물 라이브러리를 조합한 HTP 균주 개량

방법을 교시한다.

[0059] 다른 실시태양에서, 본 발명은 다중 유형의 유전자 미생물 라이브러리를 사용하는 HTP 균주 개량 방법을 교시한다. 일부 실시태양에서, 유전자 미생물 라이브러리는 조합되어 조합 돌연변이(예를 들어, 하나 이상의 유전자에 적용된 프로모터/터미네이터 조합 래더)를 생성한다. 또 다른 실시태양에서, 본 발명의 HTP 균주 개량 방법은 하나 이상의 전통적인 균주 개량 방법과 조합될 수 있다.

[0060] 일부 실시태양에서, 본 발명의 HTP 균주 개량 방법은 개량된 숙주 세포를 초래한다. 즉, 본 발명는 하나 이상의 숙주 세포 특성을 개량시키는 방법을 교시한다. 일부 실시태양에서, 개량된 숙주 세포 특성은 숙주 세포에 의해 생성된 관심 생성물의 부피 생산성, 비 생산성, 수율 또는 역가로 이루어진 그룹으로부터 선택된다. 일부 실시태양에서, 개량된 숙주 세포 특성은 부피 생산성이다. 일부 실시태양에서, 개량된 숙주 세포 특성은 비 생산성이다. 일부 실시태양에서, 개량된 숙주 세포 특성은 수율이다.

[0061] 일부 실시태양에서, 본 발명의 HTP 균주 개량 방법은 HTP 균주 개량 방법을 적용하지 않는 대조군 숙주 세포에 비해 적어도 하나의 숙주 세포 특성에서 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, 150%, 200%, 250%, 300% 이상의 개량(예를 들어, 임의의 범위 및 그 사이의 하부범위를 포함하는 관심 생체분자의 수율 또는 생산성의 X% 개량)을 나타내는 숙주 세포를 초래한다. 일부 실시태양에서, 본 발명의 HTP 균주 개량 방법은 SNP 스왑, PRO 스왑, STOP 스왑, 트랜스포존 돌연변이유발 다양성 라이브러리, 리보솜 결합 부위 미생물 균주 라이브러리, 항-대사산물 선택/발효 생성물 저항성 미생물 라이브러리 및 이의 조합으로 이루어진 그룹으로부터 선택된다.

[0062] 따라서, 일부 실시태양에서, 본 개시의 SNP 스왑 방법은 SNP 스왑 방법을 적용하지 않는 대조군 숙주 세포에 비해 적어도 하나의 숙주 세포 특성에서 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, 150%, 200%, 250%, 300% 이상의 개량(예를 들어, 임의의 범위 및 그 사이의 하부범위를 포함하는 관심 생체분자의 수율 또는 생산성의 X% 개량)을 나타내는 숙주 세포를 초래한다.

[0063] 따라서, 일부 실시태양에서, 본 발명의 PRO 스왑 방법은 PRO 스왑 방법을 적용하지 않는 대조군 숙주 세포에 비해 적어도 하나의 숙주 세포 특성에서 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, 150%, 200%, 250%, 300% 이상의 개량(예를 들어, 임의의 범위 및 그 사이의 하부범위를 포함하는 관심 생체분자의 수율 또는 생산성의 X% 개량)을 나타내는 숙주 세포를 초래한다.

[0064] 일부 실시태양에서, 본 발명의 터미네이터 스왑 방법은 PRO 스왑 방법을 적용하지 않는 대조군 숙주 세포에 비해 적어도 하나의 숙주 세포 특성에서 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59% , 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 대조군에 비해 하나 이상의 숙주 세포 특성에서 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, 150%, 200%, 250%, 300% 이상의 개량(예를 들어, 임의의 범위 및 그 사이의 하부범위를 포함하는 관심 생체분자의 수율 또는 생산성의 X% 개량)을 나타내는 숙주 세포를 초래한다.

[0065] 일부 실시태양에서, 본 발명의 트랜스포존 돌연변이유발 방법은 PRO 스왑 방법을 적용하지 않는 대조군 숙주 세포에 비해 적어도 하나의 숙주 세포 특성에서 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%,

15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59% , 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, 150%, 200%, 250%, 300% 이상의 개량(예를 들어, 임의의 범위 및 그 사이의 하부범위를 포함하는 관심 생체분자의 수율 또는 생산성의 X% 개량)을 나타내는 숙주 세포를 초래한다.

[0066] 일부 실시태양에서, 본 발명의 리보솜 결합 부위 라이브러리를 사용하는 방법은 PRO 스왑 방법을 적용하지 않는 대조군 숙주 세포에 비해 적어도 하나의 숙주 세포 특성에서 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59% , 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, 150%, 200%, 250%, 300% 이상의 개량(예를 들어, 임의의 범위 및 그 사이의 하부범위를 포함하는 관심 생체분자의 수율 또는 생산성의 X% 개량)을 나타내는 숙주 세포를 초래한다. 일부 실시태양에서, 본 발명의 항-대사산물 선택/발효 생성물 저항성 방법은 PRO 스왑 방법을 적용하지 않는 대조군 숙주 세포에 비해 적어도 하나의 숙주 세포 특성에서 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59% , 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, 150%, 200%, 250%, 300% 이상의 개량(예를 들어, 임의의 범위 및 그 사이의 하부범위를 포함하는 관심 생체분자의 수율 또는 생산성의 X% 개량)을 나타내는 숙주 세포를 초래한다.

[0067] 본 발명은 또한 2개 이상의 미생물 균주에서 유전자 변화의 신속한 통합(consolidation) 및 사카로폴리스포라 종에서 유전자 다양성을 생성하는 방법을 제공한다. 일부 실시태양에서, 방법은 원형질체(protoplast) 융합에 기초한다. 일부 실시태양에서, 미생물 균주 중 적어도 하나가 "마킹된" 돌연변이를 함유하는 경우, 방법은 다음 단계를 포함한다: (1) 통합을 위해 조작된 균주의 풀로부터 부모 균주를 선택하는 단계; (2) 통합될 균주로부터 원형질체를 준비하는 단계(예를 들어, 세포벽 제거 등); 및 (3) 관심 균주를 융합시키는 단계; (4) 세포를 회수하는 단계; (5) "마킹된" 돌연변이를 갖는 세포를 선택하는 단계, 및 (6) 다른 부모 균주에 대해 오는 돌연변이의 존재에 대해 성장 세포를 유전자형 분석(genotyping)하는 단계. 선택적으로, 방법은 (7) "마킹된" 돌연변이로부터 플라스미드를 제거하는 단계를 추가로 포함한다. 일부 실시태양에서, 미생물 균주 중 어느 것도 "마킹된" 돌연변이를 함유하지 않는 경우, 방법은 다음 단계를 포함한다: (1) 통합을 위해 조작된 균주의 풀로부터 부모 균주를 선택하는 단계; (2) 통합될 균주로부터 원형질체를 준비하는 단계(예를 들어, 세포벽 제거 등); 및 (3) 관심 균주를 융합시키는 단계; (4) 세포를 회수하는 단계; (5) 제 1 부모 균주에 대해 오는 돌연변이의 존재에 대해 세포를 선택하는 단계, 및 (6) 다른 부모 균주에 대해 오는 돌연변이의 존재에 대해 세포를 선택하는 단계. 일부 실시태양에서, 균주는 제 1 부모 균주 및/또는 다른 부모 균주로부터 오는 돌연변이와 관련된 표현형에 기초하여 선택된다. 일부 실시태양에서, 균주는 유전자형 분석에 기초하여 선택된다. 일부 실시태양에서, 유전자형 분석 단계는 고 처리량 공정으로 수행된다.

[0068] 일부 실시태양에서, 단계 (3)에서, 유용한 (신규) 돌연변이체의 조합을 생성할 가능성을 증가시키기 위해, "마킹된" 돌연변이를 갖는 균주의 더 적은 세포가 사용될 수 있으며, 따라서 이들 "마킹된" 세포가 다른 돌연변이를 가진 세포와 상호작용하고 융합했을 가능성이 증가된다. 일부 실시태양에서, 단계 (4)에서, 한천 오버레이를 사용하지 않고 세포를 삼투적으로 안정화된 매질 상에 플레이팅하여, 공정을 단순화하고 보다 용이한 자동화를 가능하게 한다. 삼투-안정화제는 카운터 선택 마커 유전자(예를 들어, sacB 유전자)를 함유할 수 있는 세포의 성장을 가능하게 하는 것이다. 원형질체 세포는 치료에 매우 민감하고 죽이기 쉽다. 이 단계는 충분한 세포가 회복되는 것을 보증한다. 이 단계가 잘 작동할수록 다운스트림 분석에 더 많은 재료를 사용할 수 있다. 일부 실시태양에서, 단계 (5)에서, 단계는 적절한 항생제를 성장 세포 상에 오버레이함으로써 달성된다. 부모 세포 중 어느 것도 "마킹된" 돌연변이를 보유하지 않는 경우, 균주는 관심 균주를 확인하기 위한 다른 수단에 의해 유전자형 분석될 수 있다. 이 단계는 선택적일 수 있지만, 이는 세포 융합이 일어날 가능성이 가장 높은 세포가 풍부하게 되는 것을 보증한다. 다중 유전자좌를 "마킹"하는 것이 가능하며 이러한 방식으로 관심 조합을 더 빨리 생성할 수 있지만, 이후 "상처없는(scarless)" 균주를 원한다면 다중 플라스미드를 제거해야 할 수 있다. 일부

실시태양에서, 단계 (6)에서, 유전자형에 대한 콜로니의 수는 선택 계획뿐만 아니라 교차(cross)의 복잡성에 의존한다. 일부 실시태양에서, 단계 (7)은 선택적이며 추가 검증 또는 클라이언트 전달에 권장된다. 일부 실시태양에서, 균주에 대한 조작 사이클의 끝에서, 모든 플라스미드 잔재물을 제거할 필요가 있다. 이것이 언제 그리고 얼마나 자주 수행되는지는 사용자의 재량에 달려 있다. 일부 실시태양에서, 카운터 선택 가능한 sacB 유전자의 존재는 이 단계를 간단하게 만든다. 일부 실시태양에서, 균주의 적어도 하나는 "마킹된" 돌연변이를 갖는다. 일부 실시태양에서, 단일 통합 단계 동안 융합된 균주의 수는 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500 이상과 같이 2개 이상일 수 있다. 일부 실시태양에서, 융합을 위한 하나 이상의 균주는 관심 부위에서 선택 마커에 의해 태그될 수 있다.

[0069] 본 발명은 또한 사카로폴리스포라 종에 사용하기 위한 리포터 단백질 및 관련 분석법을 제공한다. 일부 실시태양에서, 리포터 단백질은 대셔(Dasher) GFP(서열번호 81), 파프리카 RFP(서열번호 82) 및 효소 베타-글루쿠로니다제(gusA)(서열번호 83)로 이루어진 그룹으로부터 선택된다. 일부 실시태양에서, 이들 리포터 유전자를 코딩하는 뉴클레오티드 서열은 대장균 또는 사카로폴리스포라 종에 대해 코돈 최적화된다. 일부 실시태양에서, 본 발명의 형광 단백질은 사카로폴리스포라 종에서 관찰된 내인성 형광의 스펙트럼과 겹치지 않는 스펙트럼을 갖는다. 일부 실시태양에서, 리포터 단백질은 사카로폴리스포라 종에서 관심 유전자의 활성을 결정하는데 사용된다. 일부 실시태양에서, 리포터 단백질은 사카로폴리스포라 종에서 관심 프로모터 서열의 강도를 결정하는데 사용된다. 이러한 프로모터는 천연, 합성 또는 이의 조합일 수 있다. 천연 프로모터는 사카로폴리스포라 종에 고유하거나 또는 사카로폴리스포라 종에 이종일 수 있다.

[0070] 일부 실시태양에서, 리포터 단백질은 사카로폴리스포라 종에서 관심 터미네이터 서열의 강도를 결정하는데 사용된다. 일부 실시태양에서, 리포터 단백질은 사카로폴리스포라 종에서 관심 개시 코돈 또는 중지 코돈의 강도를 결정하는데 사용된다. 일부 실시태양에서, 리포터 단백질은 사카로폴리스포라 종에서 관심 리보솜 결합 부위 서열의 강도를 결정하는데 사용된다. 일부 실시태양에서, 리포터 단백질은 사카로폴리스포라 종의 게놈으로부터 서열이 루프 아웃되었는지를 결정하기 위한 마커로서 사용된다.

[0071] 본 발명은 또한 사카로폴리스포라 종에서 유전자 요소의 삽입을 위한 중성 통합 부위(NIS)를 제공한다. 이들 중성 통합 부위는 개별 유전자 또는 다중 유전자 카세트가 사카로폴리스포라 종 균주의 게놈 내에서 안정적이고 효율적으로 통합될 수 있는 유전자좌이다. 이들 부위로의 서열의 통합은 균주의 성장에 영향을 미치지 않거나 제한적이다. 일부 실시태양에서, 중성 통합 부위는 서열번호 132 내지 서열번호 142의 서열을 갖는 유전자좌로 이루어진 그룹으로부터 선택된다. 일부 실시태양에서, 독특한 유전자 서열(즉, 워터마크)이 균주 또는 계통을 표시하기 위해(예를 들어, 독점적인 이유로) NIS에 삽입될 수 있다.

[0072] 일부 실시태양에서, 하나 이상의 유전적 요소는 본 발명에 기재된 사카로폴리스포라 종의 단일 중성 통합 부위에 삽입된다. 일부 실시태양에서, 하나 이상의 유전적 요소는 본 발명에 기재된 사카로폴리스포라 종의 중성 통합 부위의 2, 3, 4, 5, 6, 7, 8, 9, 10 또는 11과 같은 2개 이상의 중성 통합 부위에 삽입된다. 일부 실시태양에서, 중성 통합 부위(들)에 삽입된 유전자 요소(들)를 갖는 사카로폴리스포라 종 균주는 삽입을 갖지 않는 기준 균주와 비교하여 유사한 성장을 갖는다. 일부 실시태양에서, 중성 통합 부위(들)에 삽입된 유전자 요소(들)를 갖는 사카로폴리스포라 종 균주는 삽입이 없는 기준 균주에 비해 성능이 개량된다(예를 들어, 스피노신과 같은 하나 이상의 관심 분자의 수율의 개량). 일부 실시태양에서, 중성 통합 부위(들)에 삽입된 유전자 요소(들)를 갖는 사카로폴리스포라 종 균주는 다양성 라이브러리를 형성하며, 이는 기준 균주에 비해 개량된 성능을 가진 새로운 균주를 생성 및 선택하기 위해 본 발명에 기재된 다른 균주 라이브러리와 추가로 조합될 수 있다. 일부 실시태양에서, 중성 통합 부위(들)에 삽입된 유전자 요소(들)를 갖는 사카로폴리스포라 종 균주는 추가로 돌연변이화되어 원하는 표현형을 갖는 추가의 새로운 균주에 대해 선택될 수 있다.

[0073] 본 발명은 또한 공여자 미생물 세포로부터 사카로폴리스포라 미생물의 수용자 세포로 유전자 물질을 전달하는 방법을 제공한다. 일부 실시태양에서, 방법은 다음 단계를 포함한다: (1) 수용자 세포를 중기-대수(mid-exponential) 상으로 계대배양(subculturing)하는 단계(선택적); (2) 공여자 세포를 중기-대수 상으로 계대배양하는 단계(선택적); (3) 공여자 및 수용자 세포를 결합하는 단계; (4) 컨쥬게이션 매질 상에 공여자 및 수용자 세포 혼합물을 플레이팅하는 단계; (5) 세포를 컨쥬게이션시키기 위해 플레이트를 배양하는 단계; (6) 공여자 세포에 대한 항생제 선택을 적용하는 단계; (7) 비-통합 수용자 세포에 대한 항생제 선택을 적용하는 단계; 및 (8) 통합된 수용자 세포의 성장을 허용하기 위해 플레이트를 추가로 배양하는 단계. 일부 실시태양에서, 공여자 미생물 세포는 대장균 세포이다. 일부 실시태양에서, 수용자 미생물 세포는 사카로폴리스포라 스피노사와 같은

사카로폴리스포라 종 세포이다.

[0074] 일부 실시태양에서, 다음 조건 중 적어도 2개, 3개, 4개, 5개, 6개, 7개 또는 그 이상이 이용된다: (1) 수용자 세포가 세척된다; (2) 공여자 세포 및 수용자 세포는 약 30℃의 온도에서 컨쥬게이션된다; (3) 수용자 세포를 컨쥬게이션하기 전에 적어도 약 48시간 동안 계대배양한다; (4) 컨쥬게이션을 위한 공여자 세포:수용자 세포의 비는 약 1:0.6 내지 1:1.0이다; (5) 공여자 세포에 대한 선택을 위한 항생제 약물은 공여자 세포와 수용자 세포가 혼합된 후 약 15 내지 24시간 후에 혼합물로 전달된다; (6) 수용자 세포에 대한 선택을 위한 항생제 약물은 공여자 세포 및 수용자 세포가 혼합된 후 약 40 내지 48시간 후에 혼합물에 전달된다; (7) 공여자 및 수용자 세포 혼합물로 플레이팅된 컨쥬게이션 매질을 적어도 약 3시간 내지 10시간 동안 건조시킨다; (8) 컨쥬게이션 매질은 적어도 약 3g/L 글루코스를 포함한다; (9) 공여자 세포의 농도는 약 OD600=0.4이다; (10) 수용자 세포의 농도는 약 OD540=13.0이다.

[0075] 일부 실시태양에서, 공여자 세포에 대한 선택을 위한 항생제 약물은 공여자 세포가 민감성인 반면, 수용자 세포는 저항성인 약물이다. 일부 실시태양에서, 수용자 세포에 대한 선택을 위한 항생제 약물은 공여자 세포가 저항성인 반면, 수용자 세포는 민감성인 약물이다.

[0076] 일부 실시태양에서, 공여자 세포에 대한 선택을 위한 항생제는 날리딕산(nalidixic)이고, 농도는 약 50 내지 약 150㎍/ml이다. 일부 실시태양에서, 공여자 세포에 대한 선택을 위한 항생제는 스펙티노마이신이고, 농도는 약 10 내지 약 300㎍/ml이다.

[0077] 일부 실시태양에서, 공여자 세포에 대한 선택을 위한 항생제는 날리딕산이고, 농도는 약 100㎍/ml이다.

[0078] 일부 실시태양에서, 수용자 세포에 대한 선택을 위한 항생제는 아프라마이신(apramycin)이고, 농도는 약 50 내지 약 250㎍/ml이다.

[0079] 일부 실시태양에서, 수용자 세포에 대한 선택을 위한 항생제는 아프라마이신이고, 농도는 약 100㎍/ml이다.

[0080] 일부 실시태양에서, 방법은 고 처리량 공정에서 수행된다. 일부 실시태양에서, 방법은 48-웰 Q-트레이에서 수행된다.

[0081] 일부 실시태양에서, 고 처리량 공정은 자동화된다.

[0082] 일부 실시태양에서, 공여자 세포 및 수용자 세포의 혼합물은 액체 혼합물이고, 액체 혼합물의 충분한 부피가 요동 운동(rocking motion)을 하는 배지 상에 플레이팅되고, 액체 혼합물이 배지의 전체 영역에 걸쳐 분산된다.

[0083] 일부 실시태양에서, 방법은 공여자 세포에 의해 제공된 통합된 DNA를 갖는 수용자 세포의 후속 접종을 위해 효모 핀을 사용하는 콜로니 피킹에 의해 컨쥬게이션 완료체(exconjugant)를 전달하는 자동화된 공정을 포함한다.

[0084] 일부 실시태양에서, 콜로니 피킹은 침지 운동(dipping motion) 또는 교반 운동(stirring motion)에서 수행된다.

[0085] 일부 실시태양에서, 컨쥬게이션 매질은 약 3-10g/L 글루코스를 포함하는 변형된 ISP4 매질이다.

[0086] 일부 실시태양에서, 혼합물에서 공여자 세포 또는 수용자 세포의 총 개수가 약 $5x10^6$ 내지 약 $9X10^6$이다. 일부 실시태양에서, 컨쥬게이션에 사용된 공여자 세포의 농도는 약 OD 0.1 내지 약 OD 0.6이다.

[0087] 일부 실시태양에서, 방법은 다음 조건 중 적어도 2개, 3개, 4개, 5개, 6개 또는 7개를 사용하여 수행된다: (1) 수용자 세포는 컨쥬게이션 전에 세척된다; (2) 공여자 세포 및 수용자 세포는 약 30℃의 온도에서 컨쥬게이션된다; (3) 수용자 세포를 컨쥬게이션하기 전에 적어도 약 48시간 동안 계대배양한다; (4) 컨쥬게이션을 위한 공여자 세포:수용자 세포의 비는 약 1:0.8이다; (5) 공여자 세포에 대한 선택을 위한 항생제 약물은 공여자 세포와 수용자 세포가 혼합된 후 약 20시간 후에 혼합물에 전달된다; (6) 혼합물에서 공여자 세포의 양 또는 수용자 세포의 양은 약 $7x10^6$이다; 및 (7) 컨쥬게이션 매질은 약 6g/L 글루코스를 포함한다.

[0088] 본 발명은 또한 사카로폴리스포라 균주에서 표적화된 게놈 편집 방법을 제공하여, 표적화된 게놈 유전자좌에서 유전자 변이를 함유하는 상처없는 사카로폴리스포라 균주를 생성한다. 일부 실시태양에서, 상기 방법은 a) 플라스미드를 사카로폴리스포라 균주 내로 도입하는 단계를 포함하며, 상기 플라스미드는 다음을 포함한다: (i) 선택 마커, (ii) 카운터 선택 마커, (iii) 표적 유전자좌에서 사카로폴리스포라 게놈으로 통합될 유전자 변이를 함유하는 DNA 단편, 상기 DNA 단편은 원하는 유전자 변이의 측면에 있는(flanking) 표적 게놈 유전자좌에 상동성 암(homology arm)을 가진다, 및 (iv) 플라스미드 백본 서열.

[0089] 일부 실시태양에서, 사카로폴리스포라 균주에서 표적화된 게놈 편집 방법은 b) 초기 상동성 재조합을 겪고 게놈에서 선택 마커의 존재에 기초하여 표적 유전자좌에 통합된 유전자 변이를 갖는 사카로폴리스포라 균주를 선택하는 단계; 및 c) 표적 유전자좌에 통합된 유전자 변이를 갖지만 카운터 선택 마커의 부재에 기초하여 플라스미드 백본을 루프 아웃시키는 추가적인 상동성 재조합을 겪은 사카로폴리스포라 균주를 선택하는 단계를 추가로 포함한다. 일부 실시태양에서, 선택 단계 b) 및 선택 단계 c)는 동시에 수행된다. 일부 실시태양에서, 선택 단계 b) 및 선택 단계 c)는 순차적으로 수행된다. 선택의 결과로, 유전자 변이를 함유하는 DNA 단편은 선택된 사카로폴리스포라 균주의 표적 유전자좌에서 사카로폴리스포라 게놈으로 통합되는 반면, 선택 마커, 카운터 선택 마커 및/또는 플라스미드 백본 서열은 선택된 사카로폴리스포라 균주의 게놈으로부터 "루프 아웃"된다.

[0090] 표적화된 게놈 유전자좌는 사카로폴리스포라 게놈의 임의의 영역을 포함할 수 있다. 일부 실시태양에서, 표적화된 게놈 유전자좌는 코딩 DNA 모듈의 반복 세그먼트를 포함하지 않는 게놈 영역을 포함한다.

[0091] 일부 실시태양에서, 표적화된 게놈 편집을 위한 플라스미드는 온도 민감성 레플리콘을 포함하지 않는다.

[0092] 일부 실시태양에서, 표적화된 게놈 편집을 위한 플라스미드는 복제 기점(origin of replication)을 포함하지 않는다.

[0093] 일부 실시태양에서, 선택 단계 (c)는 통합된 플라스미드의 복제없이 수행된다.

[0094] 일부 실시태양에서, 플라스미드는 단일 상동성 재조합 벡터이다. 일부 실시태양에서, 플라스미드는 이중 상동성 재조합 벡터이다.

[0095] 일부 실시태양에서, 카운터 선택 마커는 sacB 유전자 또는 pheS 유전자이다.

[0096] 일부 실시태양에서, sacB 유전자 또는 pheS 유전자는 사카로폴리스포라 스피노사에 대해 코돈-최적화된다.

[0097] 일부 실시태양에서, sacB 유전자는 서열번호 146의 서열을 포함한다. 일부 실시태양에서, pheS 유전자는 서열번호 147 또는 서열번호 148의 서열을 포함한다.

[0098] 일부 실시태양에서, 플라스미드는 형질전환에 의해 사카로폴리스포라 균주로 도입된다.

[0099] 일부 실시태양에서, 형질전환은 원형질체 형질전환이다.

[0100] 일부 실시태양에서, 플라스미드는 컨쥬게이션에 의해 사카로폴리스포라 균주로 도입되고, 사카로폴리스포라 균주는 수용자 세포이고, 플라스미드를 포함하는 공여자 세포는 플라스미드를 사카로폴리스포라 균주로 전달한다. 일부 실시태양에서, 컨쥬게이션은 플라스미드를 포함하는 대장균 공여자 세포를 기초로 한다. 일부 실시태양에서, 표적 유전자좌는 사카로폴리스포라 균주에서 관심 화합물의 생성과 관련된 유전자좌이다. 일부 실시태양에서, 관심 화합물은 스피노신이다.

[0101] 생성된 사카로폴리스포라 균주는 편집된 게놈이 관심 화합물의 개량된 생산과 같은 하나 이상의 원하는 특성을 가질 수 있다. 일부 실시태양에서, 생성된 사카로폴리스포라 균주는 게놈 편집이 없는 대조군 균주와 비교하여 관심 화합물의 증가된 생산을 가진다.

[0102] 일부 실시태양에서, 방법은 고 처리량 공정으로서 수행된다.

[0103] 전술한 고 처리량(HTP) 방법은 상기 방법의 적어도 하나의 단계를 수행하기 위해 적어도 하나의 자동화 장비(예를 들어, 액체 핸들러 또는 플레이트 핸들러 장치)의 이용을 수반할 수 있다. 본 발명의 HTP 방법은 보다 적은 인력을 사용하여 대규모로 수행될 수 있기 때문에 미생물(예를 들어, 사카로폴리스포라 종)의 게놈 공학의 보다 빠르고 노동 집약적인 방식을 제공한다. 예를 들어, 일부 실시태양에서, 본 발명의 임의의 방법은 48-웰 플레이트, 96-웰 플레이트, 192-웰 플레이트, 384-웰 플레이트 등에서 수행되어, 다수의 균주가, 하나씩이 아니라, 동시에 생성되고 및/또는 테스트된다. 이 방법은 자동화 장비를 사용하지 않는 다른 방법과 비교하여 많은 시간을 절약하다. 일부 실시태양에서, 방법은 본 발명의 방법에서 동일하거나 더 적은 인적 자원이 사용될 때, 자동화 장비가 사용되지 않는 다른 방법에 비해 약 10배, 20배, 30배, 40배, 50배, 60배, 70배, 80배, 90배, 100배, 150배, 200배, 250배, 300배 또는 그 이상 빠르다.

발명의 효과

[0104] 본 발명의 내용 중에 포함되어 있다.

도면의 간단한 설명

[0105] 도 1은 다양성 풀의 변화를 증가시키기 위한 본 발명의 DNA 재조합 방법을 도시한다. 관련 종으로부터의 게놈 영역과 같은 DNA 절편은 물리적 또는 효소적/화학적 수단을 통해 절단될 수 있다. 절단된 DNA 영역은 용융되고 재어닐링되게 하여, 중첩된 유전 영역은 중합효소 신장 반응을 준비한다. 후속 용융/신장 반응은 하나 이상의 출발 서열로부터의 원소를 포함하는 키메릭 DNA로 재조합될 때까지 수행된다.

도 2는 선택된 서열 변형(예를 들어, 스왑을 위한 100개 SNP)을 갖는 새로운 숙주 생물을 생성하기 위한 본 발명의 방법을 개략적으로 도시한다. 간략하게, 이 방법은(1) 원하는 DNA 삽입체는 어셈블리 반응에서 하나 이상의 합성된 올리고를 조합함으로써 디자인되고 생성하고,(2) DNA 삽입체는 형질전환 플라스미드에 클로닝하고,(3) 완성된 플라스미드는 원하는 생산 균주로 옮겨지고, 여기서 이들은 숙주 균주 게놈 속으로 통합된다, 및 (4) 선택 마커 및 기타 원하지 않는 DNA 요소는 숙주 균주에서 루프 아웃된다. 각 DNA 어셈블리 단계는 증폭 및 서열화를 위해 대장균 박테리아 속에 플라스미드를 클로닝하는 것과 같은 추가 품질 관리(QC) 단계를 필요로 할 수 있다.

도 3은 본 발명의 형질전환 플라스미드의 조립 및 숙주 유기체 속으로 이의 통합을 도시한다. 삽입체 DNA는 하나 이상의 합성된 올리고를 어셈블리 반응에서 결합함으로써 생성된다. 원하는 서열을 함유하는 DNA 삽입체는 게놈의 표적 부위와 상동성을 갖는 DNA 영역에 인접해있다. 이런 상동성 영역은 게놈 통합을 용이하게 하고, 일단 통합되면, 후속 단계에서 벡터 백본 DNA를 루핑 아웃(looping-out)하기 위해 디자인된 직접 반복 영역을 형성한다. 조립된 플라스미드는 삽입체 DNA 및 임의적으로, 하나 이상의 선택 마커를 함유한다.

도 4는 숙주 균주로부터 DNA의 선택된 영역을 루핑 아웃하기 위한 절차를 도시한다. 삽입된 DNA 및 숙주 게놈의 직접 반복 영역은 재조합 이벤트에서 "루프 아웃(loop out)"될 수 있다. 선택 마커에 대해 선택된 세포 계수기는 직접 반복 영역에 의해 인접된 루프 DNA의 결실을 함유한다.

도 5는 본 발명의 균주 개량 공정의 한 실시태양을 도시한다. 유전자 디자인(Genetic Design)을 함유하는 숙주 균주 서열은 다양한 균주 배경(Strain Build)에서의 균주 성능 향상에 대해 테스트된다. 유익한 돌연변이를 나타내는 균주를 분석하고(Hit ID 및 분석) 데이터는 추가 분석을 위해 라이브러리에 저장된다(예를 들어, 다른 것들 중에서 SNP 스왑 라이브러리, PRO 스왑 라이브러리 및 이의 조합). 본 발명의 선택 기준은 추가 반복 분석을 위해 하나 이상의 라이브러리로부터의 요소를 조합시키는 예상된 효과에 기초하여 새로운 제안된 숙주 균주 서열을 생성한다.

도 6a 내지 6b는 본 발명의 실시태양의 하나의 DNA 어셈블리, 형질전환 및 균주 선택 단계를 도시한다. 도 6a는 DNA 단편을 구축하고, 상기 DNA 단편을 벡터 속에 클로닝하고, 상기 벡터를 숙주 균주로 형질전환시키고, 카운터 선택을 통해 선택 서열을 루핑 아웃하는 단계를 도시한다. 도 6b는 선택된 숙주 균주의 고 처리량 배양, 선택 및 평가를 위한 단계를 도시한다. 이 도면은 또한 배양 탱크에서 선택된 균주의 배양, 선택 및 평가하는 임의적 단계를 도시한다.

도 7은 본 발명의 자동화 시스템의 한 실시태양을 도시한다. 본 발명은 숙주 유기체를 클로닝, 형질전환, 배양, 선택 및/또는 서열화할 수 있는 다양한 모듈을 갖는 자동화 로봇 시스템의 사용을 교시한다.

도 8은 본 발명의 숙주 균주 개량 프로그램의 실시태양의 개관을 도시한다.

도 9는 대략 840만개의 염기쌍(Galm and Sparks, "Natural product derived insecticides: discovery and development of spinetoram" J. Ind Microbiol Biotechnol. 2015, DOI 10.1007/s10295-015-1710-x로부터 가져옴, 이는 모든 목적을 위해 전체적으로 참조로 포함된다)을 포함하는 사카로폴리스포라 스피노사의 게놈을 나타낸다.

도 10은 코리네박테리움에서 본 발명의 형질전환 실험을 도시한다. 0.5kb 내지 5.0kb 범위의 DNA 삽입체는 미생물 균주의 게놈의 다양한 영역(상대 위치 1-24로 나타냄)에 삽입하기 위해 표적화되었다. 밝은 색은 성공적인 통합을 나타내며 어두운 색은 삽입 실패를 나타낸다.

도 11은 본 발명의 방법에 따른 제 1 라운드 SNP 스와핑 실험을 도시한다. (1) C로부터의 모든 SNP는 개별적으로 및/또는 조합적으로 기본 A 균주에 클로닝될 것이다(A에서 C로 "웨이브 업"). (2) C로부터의 모든 SNP는 상업적 균주 C로부터 개별적으로 및/또는 조합적으로 제거될 것이다(C에서 A로 "웨이브 다운"). (3) B로부터의 모든 SNP는 개별적으로 및/또는 결합하여 기본 A 균주 속으로 클로닝될 것이다. (4) B로부터의 모든 SNP는 개별적으로 및/또는 조합적으로 상업적 균주 B로부터제거될 것이다(B에서 A로 웨이브 다운). (5) C에 고유한 모든 SNP

는 개별적으로 및/또는 조합적으로 상업적 균주 B에 클로닝될 것이다(B에서 C로 웨이브 업). (6) C에 고유한 모든 SNP는 개별적으로 및/또는 조합적으로 상업적 균주 C로부터 제거될 것이다(C에서 B로 웨이브 다운).

도 12a 내지 도 12d는 프로모터 스왑 공정에 이용될 수 있는 스피노신 합성에 관여하는 예시적인 유전자 표적을 도시한다. 도 12a는 다른 게놈 유전자좌에 존재하는 유전자를 포함하는 스피노신 생합성 유전자 클러스터의 그래픽 표현이다. 도 12b는 스피노신 폴리케타이드 스캐폴드의 생합성 어셈블리이다. 도 12c는 최종 스피노신 A 및 D 분자를 형성하기 위한 가교 및 조정(tailoring) 반응을 나타낸다. 도 12d는 3'-O-에틸화 및 5,6-이중 결합 감소를 통한 스피네토람으로의 이후 합성 전환과 함께 스피노신 J의 발효-기반 생산을 나타낸다. 모든 도면은 Galm 및 Sparks. 2015에서 채택되었다.

도 13은 확인된 유전자 표적에 대한 프로모터 스왑 공정을 수행하기 위해 사용되는 예시적인 프로모터 라이브러리를 예시한다. PRO 스왑(즉, 프로모터 스왑) 공정에서 사용된 프로모터는 실시예 4 및 표 1에서 확인할 수 있는 것들이다. 경로 표적의 비 제한적인 예는 좌측 박스에 도시되어 있으며 프로모터 래더의 구성원의 다양한 발현 강도는 중간 박스에 도시되어 있다. 알 수 있는 바와 같이, 프로모터는 강한 것에서 약한 것까지의 "래더"의 발현 강도를 제공한다.

도 14는 프로모터 스와핑 유전자 결과가 표적화되는 특정 유전자에 의존한다는 것을 예시한다.

도 15는 S. 스피노사로 미리 특징 지어진 비-천연 프로모터인 PermE*에 대한 배수 변화로 나타낸 씨드 매질에서 48시간 동안 성장한 프로모터 균주의 평균 형광을 나타내는 예시적인 HTP 프로모터 스와핑 데이터를 도시한다(비-생산 조건). 상대 강도는 대략 50배의 다이나믹 레인지에 걸쳐 있다. 3개의 천연 프로모터는 래더에서 가장 강력한 5개의 프로모터 중 하나이며, P1은 PermE*보다 대략 5배 더 강력하고 다음으로 가장 강력한 프로모터보다 ~2배 더 강력하다. 또한, 합성 프로모터의 상대 강도는 스트렙토마이세스에 대한 문헌에서 보고된 결과와 유사하다. A 및 B는 S. 스피노사의 상이한 균주를 나타낸다. X-축은 상이한 프로모터를 나타내고, Y-축은 형광에 의해 측정될 때 각 프로모터의 상대 강도를 포함한다. 교시된 PRO 스왑 분자 도구는 임의의 관심 화합물의 생산을 최적화 및/또는 증가시키는데 사용될 수 있다. 당업자는 어떻게 표적 유전자를 선택하고, 원하는 화합물의 생산을 암호화한 다음, 교시된 PRO 스왑 절차를 사용하는 방법을 이해할 것이다. 당업자는 본 발명에 제공된 상세한 설명과 함께 본 발명에 교시된 라이신 수율 증가를 예시하는 입증된 데이터가 PRO 스왑 분자 툴이 HTP 게놈 공학에서 광범위하게 적용가능한 진보가 되게 할 수 있다는 것을 쉽게 이해할 것이다.

도 16은 지머젠(Zymergen)의 96-웰 플레이트 모델(생산-관련 조건)에서 성장 된 프로모터 래더 균주(균주 A 및 균주 B)에서 측정된 로그-변형 정규화된 형광의 요약이다. 이들 균주는 숙주 게놈에 통합된 상이한 프로모터>GFP 발현 카세트를 갖는다. 음영 박스는 프로모터 평가의 첫 번째 라운드 동안 평가되고 이후 실험에서 내부 대조군을 나타내는 균주를 나타낸다. 하단 막대는 평균 형광 기준선을 나타낸다.

도 17은 표 8에 기재된 프로모터 P21 및 P1로 조작된 균주에서 개량된 스피노신 J+L 역가를 도시한다. 특히, 7000225635는 strain_B_3g05097에 P1 프로모터를 함유하고; 7000206640은 strain_B_3g00920에서 P21 프로모터를 함유하고; 7000206509는 strain_B_3g02509에 P1 프로모터를 함유하고; 7000206745는 균주 _B_3g07456에 P21 프로모터를 함유하고; 7000206752는 strain_B_3g07766에 P21 프로모터를 함유하고; 7000235481은 strain_B_3g04679에 P21 프로모터를 함유한다. 각각의 균주 ID는 주어진 유전자에서의 프로모터 스왑(위의 유전자형을 가짐)을 나타내고, 따라서 각각의 균주 ID는 특정 균주 유전자형을 지칭한다. 각 점은 고 처리량 분석에서 테스트된 그 균주의 웰 또는 샘플을 나타낸다(즉, 이들은 동일한 균주에서 수집된 모든 개별 데이터 포인트이다). 선택된 프로모터 스왑 균주는 스피노신 생산을 위한 고 처리량 분석에서 테스트될 때 부모 균주(700153593)에 비해 개량을 나타냈다. 선택 가능한 마커, 프로모터-유전자 쌍 및 상동성 영역을 함유하는 플라스미드를 도입하여 중성 부위에서 게놈에 통합시키기 위해 컨쥬게이션을 사용하여 균주를 조작하였다(방법에 대한 보다 자세한 사항은 본 발명의 카운터 선택 가능한 마커 섹션 참조).

도 18은 본 발명에 기술된 방법을 사용하여 코리네박테리움에서 수행된 고려중인 입력 데이터에 대한 상대적인 균주 성능 분포의 예를 도시한다. 그러나, 유사한 절차가 사카로폴리스포라를 위해 맞춤화되었으며 본 발명자들에 의해 성공적으로 수행되었다. 상대적 성능이 0이면 조작된 균주가 인-플레이트 기본 균주에 동일하게 잘 수행되었음을 나타낸다. 여기에 설명된 공정은 제로보다 상당히 높게 수행할 수 있는 균주를 확인하도록 디자인되었다.

도 19는 본 발명의 실시태양의 하나의 DNA 어셈블리 및 형질전환 단계를 도시한다. 흐름도는 DNA 단편을 제작하고, 상기 DNA 단편을 벡터로 클로닝하고, 상기 벡터를 숙주로 형질전환시키고, 카운터 선택을 통해 선택 서열을

반복하는 단계를 도시한다.

도 20은 선택된 숙주 균주의 고 처리량 배양, 선택 및 평가를 위한 단계를 도시한다. 이 도면은 또한 배양 탱크에서 선택된 균주를 배양, 선택 및 평가하는 선택적 단계를 도시한다.

도 21은 본 발명의 프로모터 래더에 따른 조절성 발현의 범위를 나타내는 예시적인 프로모터의 발현 프로파일을 도시한다. 프로모터 A 발현은 박테리아 배양의 유도기에서 최고조인 반면, 프로모터 B 및 C는 각각 대수기 및 정상기에서 최고조이다.

도 22는 본 발명의 프로모터 래더에 따른 조절 발현의 범위를 나타내는 예시적인 프로모터의 발현 프로파일을 도시한다. 프로모터 A 발현은 선택된 기질의 첨가시 즉시 최고조가 되지만, 기질의 농도가 감소함에 따라 감지할 수 없는 수준으로 신속하게 회복된다. 프로모터 B 발현은 선택된 기질의 첨가시 즉시 최고조가 되지만, 기질에서 상응하는 감소와 함께 감지할 수 없는 수준으로 천천히 낮아진다. 프로모터 C 발현은 선택된 기질의 첨가시에 최고조가 되지만, 기질이 소비된 후에도 배양 전체에 걸쳐 높게 발현된 상태로 유지된다.

도 23은 본 발명의 프로모터 래더에 따른 구성적 발현 수준의 범위를 나타내는 예시적인 프로모터의 발현 프로파일을 도시한다. 프로모터 A는 가장 낮은 발현을 나타내고, 각각 프로모터 B 및 C의 발현 수준이 증가된다.

도 24는 균주 개량을 위한 본 발명의 LIMS 시스템의 한 실시태양을 도시한다.

도 25는 본 발명의 LIMS 시스템의 실시태양의 클라우드 컴퓨팅 구현을 도시한다.

도 26은 본 발명의 반복적인 예측 균주 디자인 작업 흐름의 한 실시태양을 도시한다.

도 27은 본 발명의 실시태양에 따른 컴퓨터 시스템의 한 실시태양을 도시한다.

도 28은 본 발명의 한 실시태양에 따른 DNA 어셈블리와 관련된 작업 흐름을 도시한다. 이 과정은 4 단계로 나뉘어진다: 부분 생성, 플라스미드 어셈블리, 플라스미드 QC 및 형질전환을 위한 플라스미드 준비. 부품 생성 동안, 실험실 정보 관리 시스템(LIMS)에 의해 고안된 올리고는 올리고 서열 판매자로부터 주문받고 PCR을 통해 숙주 유기체로부터 표적 서열을 증폭하는데 사용된다. 이러한 PCR 부분은 세척하여 오염물을 제거하고 단편 분석, 관찰된 단편 크기 대 이론적 단편 크기의 인 실리코(in silico) 품질 관리 비교 및 DNA 정량화에 의해 성공에 대해 평가된다. 부분은 어셈블리 벡터와 함께 효모로 형질전환되고 동종 재조합을 통해 플라스미드로 조립된다. 조립된 플라스미드는 효모로부터 분리되고, 후속 조립 품질 제어 및 증폭을 위해 대장균으로 형질전환된다. 플라스미드 조립 품질 관리 동안, 각 플라스미드의 여러 복제물을 분리하고, 롤링 서클 증폭(Rolling Circle Amplification(RCA))을 사용하여 증폭되고, 효소 분해 및 단편 분석에 의해 올바른 조립에 대해 평가된다. QC 과정 동안 확인된 올바르게 조립된 플라스미드는 영구적 인 원료를 생성하기 위해 히트 픽(hit picked)되고 플라스미드 DNA는 표적 숙주 유기체로 변형되기 전에 추출되고 정량화된다.

도 29는 본 발명의 실시태양에 따른, 미생물 균주의 디자인을 위한 돌연변이의 선택에서의 상위성 효과의 고려를 예시하는 흐름도이다.

도 30은 원형질체 융합을 통해 2개의 사카로폴리스포라 종 균주를 통합하기 위한 프로토콜의 예를 도시한다.

도 31a 내지 도 31d는 대셔GFP 및 파프리카RFP 형광 스펙트럼(각각도 31a 및 도 31b) 및 GFP 및 RFP 균주의 혼합된(1:1) 배양물(각각 도 31c 및 도 31d)의 상대 형광의 개략도를 도시한다. 대셔GFP의 형광 여기 및 방출 스펙트럼은 파프리카RFP와 구별되므로, 다른 기자의 큰 간섭없이 두 리포터(하단 패널, 혼합(1:1))를 모두 발현하는 샘플에서 GFP 또는 RFP 형광을 측정할 수 있다. 좌측 하단: ermE*>RFP, ermE*>GFP 균주의 상대 GFP 형광 및 두 균주의 1:1 혼합. RFP 균주에서 ermE*>RFP r균주에서 측정된 것에 비해 GFP 채널에서 형광이 거의 없거나 전혀 없으며, 혼합된 배양물은 예상대로 GFP 균주 단독의 대략 1/2인 신호를 생성한다. 우측 하단: 마찬가지로, RFP 형광에 대한 최적의 파라미터가 사용될 때(오른쪽 위) ermE*>RFP 균주에 대해 강한 형광 신호가 감지되지만, ermE*>GFP 균주에 대한 신호는 거의 또는 전혀 관찰되지 않으며, 1:1 혼합은 다시 ermE*>RFP 균주의 대략 1/2인 형광 신호를 생성한다. 따라서, 형광 리포터 대셔GFP 및 파프리카RFP는 S. 스피노사에서 작동하며 뚜렷한 형광 신호를 갖는다. 대셔GFP의 형광 여기 및 방출 스펙트럼은 파프리카RFP와 구별되므로, GFP 또는 RFP 형광을 다른 리포터로부터의 큰 간섭없이 두 리포터(하단 패널, 혼합(1:1))를 발현하는 샘플에서 측정할 수 있다.

도 32는 기능성 전사 터미네이터 및 비-터미네이터(NoT) 대조군에 대해 예상되는 바이-시스트로닉, 듀얼 리포터

테스트 카세트의 디자인 및 상대 형광을 도시한 개략도를 도시한다. 터미네이터 테스트 카세트는 나란히 배열된 두 개의 형광성 리포터 단백질 - 대셔GFP(GFP) 및 파프리카RFP(RFP)로 구성되어 있다. 이들 리포터의 바이-시스트로닉 발현은 ermE* 프로모터에 의해 유도된다. 다운스트림 리포터(RFP)의 발현은 업스트림 리보솜 결합 부위(RBS)에 의해 가능하다. 비-기능성 터미네이터 서열이 존재하는 경우, RFP 및 GFP의 발현은 터미네이터가 없을 때 관찰 된 것과 유사하다(NoT 대조군). 그러나, 기능성 전사 터미네이터가 GFP와 RFP 유전자 사이에 삽입되면 RFP의 발현이 감쇠된다. 정규화후 GFP에 대한 감쇠(attenuation) 백분율(NoT 대조군의 형광을 사용)은 터미네이터 서열의 강도를 나타낸다.

도 33은 터미네이터 기능 테스트의 결과를 도시한다. 막대는 액체 배양물에서 48시간의 성장 후 S. 스피노사 터미네이터(T1-T12) 또는 터미네이터 없음(NoT) 카세트 균주의 평균(+1sd) 상대 GFP 또는 RFP 형광을 나타낸다. 복제 배양물의 형광을 Tecan Infinite M1000 Pro(Life Sciences) 플레이트 판독기의 96-웰 분석 플레이트에서 측정하였다. 형광을 OD(OD540)로 정규화하고 상대 형광(NoT의 GFP 또는 RFP 형광의 비율로서, 대조군 배양물)으로 보고하였다. NoT에 대한 GFP 형광의 감쇠는 아마도 mRNA의 안정성에 영향을 미침으로써 업스트림 유전자(대셔GFP)의 발현에 대한 터미네이터 서열의 영향을 반영한다. 균주 내에서 GFP에 비해 RFP 형광의 감쇠는 터미네이터의 강도 - 전사를 종결하는 이의 능력을 반영하다. 테스트된 서열 중, T1이 가장 잘 수행되어, GFP에 비해 RFP의 발현이 대략 86% 감소된 반면 GFP 발현의 <30% 감소가 발생하였다. 대조적으로, T2, T4 및 T8은 이들이 RFP의 발현을 감쇠시키지 못하기 때문에 전사 터미네이터로서 기능하지 않는 것으로 보인다. 막대는 평균 +/- 1SD를 나타낸다.

도 34는 각각의 터미네이터 및 2개의 균주 배경에 대한 상대 정규화된 GFP 대 상대 정규화된 RFP 형광의 상관도를 도시한다. 파선은 1:1 상관 관계를 나타낸다. 선 아래의 점은 GFP>RFP인 균주를 나타낸다(RFP 형광의 감쇠를 나타냄). 이 선 아래의 거리(빨간 음영)는 상대 터미네이터 강도를 나타낸다. 밀도 타원들은 90% 신뢰 구간을 나타낸다. 이 도면은 상대 터미네이터 강도를 시각화할 수 있다.

도 35는 gusA 리포터가 S. 스피노사에서 작동하는 것을 도시한다. 막대는 2개의 상이한 부모 균주(A 및 B)에서 생성된 ermE*>gusA 균주로부터 무세포 용해물의 배양 후 405nm에서의 흡광도로 표시된 바와 같이, 평균 gusA 활성(+/- 1 stdev)을 나타낸다. 405nm에서의 흡광도는 4-니트로페닐 β-D-글루쿠로니드 기질에 작용하는 gusA의 효소 활성으로 인한 황색에 비례한다.

도 36은 S. 스피노사의 내인성 형광을 도시한다. 도면은 PBS로 세척한 후 배양 S. 스피노사 세포의 형광 스캔으로 측정한 상대 형광을 나타낸다. 곡선은 350-690nm에서 20nm 간격으로 여기로 인한 형광을 나타낸다. 형광은 500nm 미만에서 비교적 강하지만 여기 파장이 증가함에 따라 감소하다. 대셔GFP 및 파프리카RFP와 관련된 범위에서 내인성 형광은 최소이다. 이들 실험을 위해 대셔GFP는 505nm에서 여기되었고 방출은 525-545nm 사이에서 포착되었다. 이는 ~510nm에서 시작하는 곡선과 가장 비슷하다. 파프리카 RFP는 564nm에서 여기되었고 585-610nm 사이에서 형광이 포착되었다. 이 범위에서 내인성 형광은 거의 관찰되지 않았다.

도 37은 pCM32, pSE101 및 pSE211의 플라스미드 맵을 도시한다. (1) pCM32의 플라스미드 맵(왼쪽) 및 pCM32 제거 효소(excisionase)(xis), 인테그라제(int) 및 부착 부위(attP)를 포함하는 컨쥬게이션 플라스미드. 박스된 부분은 통합을 테스트하기 위해 컨쥬게이션 벡터로 클로닝된 플라스미드 영역을 나타낸다(Chen et al, Applied Microbiology and Biotechnology. PMID 26260388 DOI:10.1007/s00253-015-6871-z); (2) S. 에리트라에아(*erythraea*) 플라스미드 pSE101의 선형 맵. 인테그라제(int) 및 부착 부위(attP)는 맵의 왼쪽 끝에 표시되어 있다(Te Poele et al., (2008) Actinomycete integrative and conjugative elements. Antonie Van Leeuwenhoek 94, 127-143.); (3) S. 에리트라에아 플라스미드 pSE211의 선형 맵. 인테그라제(int) 및 부착 부위(attP)는 맵의 왼쪽 끝에 표시된다(Te Poele et al).

도 38은 S. 스피노사 게놈에 대한 pCM32 부착 부위의 뉴클레오티드 블라스트(Blastn)의 결과를 도시한다. 99% 초과의 동일성(149/150bp)을 가진 부위가 S. 스피노사에서 확인된다.

도 39는 S. 스피노사 게놈에 대한 pSE101 부착 부위의 뉴클레오티드 블라스트(Blastn)의 결과를 도시한다. 코어 76 뉴클레오티드에서 94% 초과의 동일성(104/111bp) 및 100%의 동일성을 갖는 부위가 S. 스피노사에서 확인된다.

도 40은 S. 스피노사 게놈에 대한 pSE211 부착 부위의 뉴클레오티드 블라스트(Blastn)의 결과를 도시한다. 코어 76 뉴클레오티드에서 88% 초과의 동일성(122/138bp) 및 100%의 동일성을 갖는 부위가 S. 스피노사에서

확인된다.

도 41a는 S. 에리트라에아 복제 플라스미드(AICE) pSE101 및 pSE211(Te Poele et al., (2008) Actinomycete integrative and conjugative elements. Antonie Van Leeuwenhoek 94, 127-143.로부터 채택)의 선형 맵을 보여주며, 이는 S. 스피노사에서 사용되기 위한 자기 복제 플라스미드이다. 대각선이 있는 화살표는 DNA 복제에 관여하는 것으로 생각되는 유전자를 나타낸다. 도 41b는 S. 에리트라에아 염색체 복제 기점을 함유하는 예시적인 복제 플라스미드의 개략도를 도시한다. S. 에리트라에아 복제 기점이 S. 스피노사에서 플라스미드의 복제를 유지할 수 있는지 테스트하기 위해, S. 에리트라에아 복제 기점은 카나마이신 저항성 유전자, 대장균 복제 기점(pBR322) 및 전달 기점(OriT)을 함유하는 플라스미드로 클로닝되어 컨쥬게이션에 의해 플라스미드의 전달을 가능하게 한다.

도 42는 플라스미드 디자인, 기능성 평가에 사용된 분석법 및 RBS 라이브러리 선별의 결과의 개략도를 도시한다. 본 발명자는 32개의 통합 플라스미드(31개와 RBS 및 No-RBS 대조군)를 디자인하고 제작하였다. 이들은 각각의 RBS를 ermE* 프로모터와 레반수크라제(sacB)를 코딩하는 유전자 사이의 S. 스피노사 통합 백본으로 상처없이 클로닝함으로써 구축되었다. 생성된 균주를 액체 배양물에서 48시간 동안 성장시키고 연속 희석물을 TSA 및 TSA + 5% 수크로스 옴니 트레이 상에 플레이팅하였다. RBS가 기능적이었다면, 수크로스 상에서 성장할 때 sacB가 독성(성장의 부재)을 초래하는 것으로 발현되었다. RBS를 함유하는 균주의 성장을 양성(sacB RBS를 함유하는 균주) 및 음성(No-RBS) 대조군과 비교함으로써, RBS의 상대 강도를 결정할 수 있었다. 이 분석을 사용하여 19개의 기능 - 16개의 "기능적" 및 3개의 "덜 기능적"인 RBS를 확인하였다. 이들 분석의 결과는 아래의 도 43a 내지 도 43e에 도시되어 있다.

도 43a 내지 도 43e는 수크로스 민감성 분석의 RBS 기능 분석 결과를 도시한다 - S. 스피노사 RBS 루프-인 균주에 대한 TSA + Kan100 대 TSA + Kan100 + 5% 수크로스에서의 성장 비교.

도 44는 S. 스피노사에서 트랜스포존 돌연변이유발에 대한 플라스미드의 선형 맵를 도시한 것이다. 기능상실(Loss-of-Function; LoF) 트랜스포존, 기능획득(Gain-of-Function; GoF) 트랜스포존, 및 GoF(Gain-of-Function) 재활용 가능 트랜스포존이 도시되어 있다.

도 45는 잠재적 중성 통합 부위를 확인하기 위해 사용된 S. 스피노사 게놈에 걸친 평균 유전자 발현의 히트 맵 섹션의 예를 도시한 것이다.

도 46은 생성물(예를 들어, 스피노신 J/L)의 존재가 탱크 내 농도의 1/100에서 S. 스피노사 성장을 억제한다는 것을 보여주는 예를 도시한 것이다.

도 47은 스피노신 J/L의 존재 하에서 부모보다 더 잘 성장하는 단리물(isolates)을 생성한 스피노신 J/L의 존재 하에서 균주의 선택을 도시한다.

도 48a 및 도 48b는 스피노신 J/L(도 48a) 및 aMM(도 48b) 둘 다에서의 선택이 HTP 플레이트 발효 모델에서 부모보다 우수한 성능을 갖는 균주를 생성 하였음을 도시한다.

도 49a 내지 도 49c는 카운터 선택 마커로서 sacB 또는 pheS를 사용하여 상처없는 사카로폴리스포라 스피노사 균주를 생성하는 과정을 도시한다. 도 49a는 상동성 재조합을 사용하여 플라스미드를 S. 스피노사 게놈에 도입하는 것을 보여준다. 도 49b는 양성 선택을 사용하여 단일 크로스오버 통합 이벤트를 선택하는 것을 보여준다. 도 49c는 플라스미드 백본을 잃기 위해 재조합된 균주를 수득하기 위해 음성 선택을 사용하여 상처없는 조작된 균주를 생성하는 것을 보여준다.

도 50은 sacB가 각각의 카운터 선택 제제 수크로스에 대해 S. 스피노사의 민감성을 부여한다는 것을 입증한다. sacB 유전자가 있거나 없는 균주를 5%에서 수크로스 민감성에 대해 시험하였다. 배양 희석 시리즈를 TSA/Kan100 및 TSA 또는 5% 수크로스를 함유하는 TSA/Kan100에 6회 반복하여 스폿팅하였다. 이는 5% 수크로스를 함유한 선택적 매질에서 유전자를 발현하는 균주의 제한적인 성장을 유발한다. 도면에서 "*"는 이 균주가 선택없이 계대배양되었음을 나타낸다.

도 51은 pheS가 균주 A에서 각각의 카운터 선택 제제 4CP에 S. 스피노사의 민감성을 부여한다는 것을 입증한다. 균주 A/PheS(SS) 및 균주 A/Phe(SE)는 2g/L에서 4CP 민감성에 대해 테스트되었다. 배양 희석 시리즈를 6회 반복하여 TSA/Kan100 및 4CP를 함유하는 TSA/Kan100에 스폿팅하였다. SE는 S. 에리트라에아로부터의 pheS 유전자를 나타내고, SS는 S. 스피노사로부터의 pheS 유전자를 나타낸다. 배양 2주 후, PheS 발현 균주 A-유도체는

TSA/Kan100-4CP에서 성장이 억제되지만, TSA/Kan100에서는 영향을 받지 않았다. 이는 PheS(SS) 및 PheS(SE)가 S. 스피노사에서 카운터 선택 마커로 역할할 가능성이 있음을 나타낸다.

도 52는 카운터 선택 마커로서 sacB를 사용하여 HTP에서 조작된 균주의 QC 결과를 보여준다. 62개의 조작된 균주 A 및 14개의 조작된 균주 B가 제조되었다.

도 53은 코리네박테리움에서 수행된 상관 측정을 사용하여 계산된 유사성 매트릭스이다. 그러나, 유사한 절차가 사카로폴리스포라를 위해 맞춤화되었으며 본 발명자들에 의해 성공적으로 수행되고 있다. 매트릭스는 SNP 변형들 사이의 기능적 유사성을 나타낸다. 기능적 유사성이 낮은 SNP의 통합은 기능적 유사성이 높은 SNP의 통합과 달리 균주 성능을 향상시킬 가능성이 더 높을 것으로 예상된다.

도 54a 내지 도 54b는 코리네박테리움에서 수행된 상위성(epistasis) 맵핑 실험의 결과를 도시한다. 그러나, 유사한 절차가 사카로폴리스포라를 위해 맞춤화되었으며 본 발명자들에 의해 성공적으로 수행되고 있다. 기능적 유사성이 낮은 SNP와 PRO 스왑의 조합으로 균주 성능이 향상된다. 도 54a는 모든 SNP/PRO 스왑의 기능적 유사성에 의해 클러스터된 덴드로그램을 도시한다. 도 54b는 생성물 수율에 의해 측정된 통합 SNP의 숙주 균주 성능을 도시한다. 클러스터 거리가 길수록 호스트 균주의 통합 성능이 향상된다.

도 55는 실험 디자인(DOE) 접근법을 사용하여 컨쥬게이션 효율을 개량시키기 위해 고려되는 인자를 도시한다.

도 56a 내지 도 56b는 HTP 포맷의 대장균 S17 + SS015 공여자 세포의 성장을 나타내고(도 56a), HTP 포맷의 대장균 S17 + SS015 공여자 세포를 사용한 컨쥬게이션 실험의 결과를 나타낸다(도 56b).

도 57은 HTP 컨쥬게이션 프로토콜에 기재된 검출을 위한 Qpix 파라미터를 사용하여 식별된 콜로니를 도시한다.

도 58은 HTP 포맷으로 성장한 후 패치로부터 접종된 S. 스피노사 배양물의 성장을 도시한다.

도 59는 DOE-기반 최적화 과정을 통해 완료된 컨쥬게이션 실험의 결과를 도시한다.

도 60은 JMP 파티션 모델링 분석 당 컨쥬게이션 효율에 연루된 것으로 결정된 조건을 도시한다.

도 61은 본 발명에 기술된 바와 같이 SNP 스왑으로 조작된 균주에서 개량된 스피노신 J+L 역가를 도시한다. SNP 스왑(SNPSWP) 균주는 초기(early)(돌연변이 전) 균주 계통과 비교하여 후기(late) 균주에 존재하는 SNP를 확인하고 이들을 후기 균주로부터 제거함으로써 조작되었다(7000153593). 선택된 SNPSWP 균주는 스피노신 생산을 위한 고 처리량 분석에서 테스트될 때 부모 균주(7000153593)에 비해 개량을 나타냈다. 이 경우, 7000153593은 "후기 균주"이자 생성된 SNPSWP의 부모 균주이다. "후기 균주"는 초기 및 후기 계보에 의존하는 SNP 스와핑의 원리로 인해 언급된다.

도 62는 본 발명에 기술된 터미네이터로 조작된 균주에서 개량된 스피노신 J+L 역가를 도시한다. 다수의 유전자 표적 앞에 약 25bp의 표 9에 열거된 터미네이터를 도입함으로써 터미네이터 삽입 균주를 조작하였다. 선택 터미네이터 삽입 균주는 스피노신 생산을 위한 고 처리량 분석에서 테스트될 때 부모 균주(7000153593)에 비해 개량을 나타냈다.

도 63은 본 발명에 기술된 바와 같은 RBS 서열로 조작된 균주에서 개량된 스피노신 J+L 역가를 도시한다. RBS 스왑(RBSSWP) 균주는 코어 생합성 유전자 표적 앞에 약 0 내지 15bp의 표 11에 열거된 RBS를 도입함으로써 조작되었다. 선택 RBSSWP 균주는 스피노신 생산을 위한 고 처리량 분석에서 테스트될 때 부모 균주(7000153593)보다 개량된 것으로 나타났다.

도 64a 내지 도 64c는 다수의 백본이 상이한 구성의 선택 마커 및 발현을 제어하기 위한 유전자 요소(터미네이터 및 프로모터)를 포함하도록 클로닝되어 있으며, 이는 상이한 균주 배경에서 균주 공학 효능을 변경할 수 있다. 일부 경우에, 백본 효능에 대한 게놈 부위의 효과를 시험하기 위해 상이한 통합 부위에서 상동성 암으로 백본을 클로닝하였다. 프로모터 pD1-7, Perm2, 및 Perm8 및 터미네이터 A_T는 이전에 특성화된 프로모터이며; 여기에 나열된 다른 유전자 요소가 이 연구에서 인용되었다.

도 65는 유전자 발현의 녹다운(감쇠 또는 예방)에 대한 터미네이터 라이브러리의 적용을 평가하는데 사용되는 발현 카세트를 도시한 것이다.

도 66a 내지 도 66b는 프로모터 사이에 터미네이터의 삽입 및 GFP의 코딩 서열로 인해 GFP 발현(형광)의 감쇠를 초래한다는 것을 도시한다. 터미네이터 녹다운 GFP 테스트 카세트의 게놈 통합과 함께 균주의 정규화된 GFP 형광(평균 +/- 95% 신뢰 구간)이 도시된다. 도 66a는 강한 프로모터(서열번호 25)와 GFP 사이에 삽입된 T1, T3,

T5, T11 및 T12(서열번호 70, 72, 74, 79 및 80)를 갖는 균주의 발현을 보여준다. "없음"(왼쪽 열)은 터미네이터가 없는 대조군 균주를 나타낸다. 도 66b는 적당히 강한 프로모터(서열번호 33)와 GFP 사이에 삽입된 T1, T3, T5 및 T12(서열번호 70, 72, 74 및 80)를 갖는 균주의 발현을 보여준다. "없음"(왼쪽 열)은 터미네이터가 없는 대조군 균주를 나타낸다. 표준 편차는 일반적으로 다이아몬드 위와 아래에서 관찰되는 수평 대시로 표시된다. 도면의 오른쪽에 있는 원은 모든 쌍의 터키-크레이머(Tukey-Kramer) HSD 테스트를 기반으로 그룹 간(중복되지 않은/교차 원은 서로 크게 다른 그룹을 나타냄) 유의한 차이를 나타낸다.

도 67은 지시된 중성 부위에 통합된 SNP스왑 페이로드를 갖는 균주 B-유래 균주의 생성물 역가(스피노신 J+L)를 도시한다. 부위 1, 2, 3, 4, 6, 9 및 10에 통합된 균주는 유사한 생성물 역가를 가지며 예상 역가(균주 B의 평균 역가; 도면에서 더 높은 막대)와 다르지 않다. 중성 부위 7에서의 통합은 생성물 역가에 부정적인 영향을 미치는 것으로 보인다. 평균 다이아몬드는 그룹 평균과 95% 신뢰 구간을 나타낸다. 표준 편차는 일반적으로 다이아몬드 위와 아래에서 관찰되는 수평 대시로 표시된다. 도면의 오른쪽에 있는 원은 모든 쌍의 터키-크레이머 HSD 테스트를 기반으로 그룹 간(중복되지 않은/교차 원은 서로 크게 다른 그룹을 나타냄) 유의한 차이를 나타낸다.

도 68은 표시된 중성 부위에 통합된 경우 GFP 발현의 비교를 도시한다. 데이터는 지시된 중성 부위에 통합된 GFP 발현 카세트 - GFP(서열번호 81)를 유도하는 강력한 프로모터(서열번호 25) - 를 갖는 WT 및 B-유래 균주의 정규화된 형광을 나타낸다. P1-대조군은 이전에 보고된 중성 부위에 통합된 이 카세트의 형광을 나타낸다. 표현은 대부분의 부위에서 비슷하다. NS7만이 우리가 평가한 다른 중성 부위(NS2, NS3, NS4, NS6 및 NS10)와 크게 다르다. 표준 편차는 일반적으로 다이아몬드 위와 아래에서 관찰되는 수평 대시로 표시된다. 도면의 오른쪽에 있는 원은 모든 쌍의 터키-크레이머 HSD 테스트를 기반으로 그룹 간(중복되지 않은/교차 원은 서로 크게 다른 그룹을 나타냄) 유의한 차이를 나타낸다.

도 69는 항-대사산물 선택에 의해 조작된 균주가 스피노신 생산 성능에 대해 테스트되었음을 나타낸다. 모든 균주는 부모에 대해 스피노신 생산 성능의 감소를 보여 주었다. 이 접근법은 균주를 확인하기 위해 최적화가 필요하다.

발명을 실시하기 위한 구체적인 내용

[0106] 다음의 용어는 당해 기술 분야에서 통상의 지식을 가진 자에 의해 잘 이해되는 것으로 생각되지만, 다음 정의는 본 발명에 개시된 주제의 설명을 용이하게 하기 위해 제시된다.

[0107] 용어 하나("a" 또는 "an")는 그 실체 중 하나 이상을 의미하며, 즉 복수의 지시대상을 의미할 수 있다. 이와 같이, 용어 "하나", "하나 이상" 및 "적어도 하나"라는 본 발명에서 상호 교환적으로 사용된다. 또한, 부정관사에 의한 "한 요소"에 대한 언급은, 내용이 분명하게는 요소의 하나 및 단지 하나가 존재하는 것을 요구하지 않는 한, 하나 이상의 요소가 존재하는 가능성을 배제하지 않는다.

[0108] 본 발명에 사용된 용어 "세포 생물", "미세유기체"또는 "미생물"은 광범위하게 이해되어야 한다. 이 용어들은 상호 교환적으로 사용되며, 두 원핵생물 영역인 박테리아와 고세균뿐만 아니라 특정 진핵생물 균류 및 원생 생물을 포함하나 이에 제한되지 않는다. 일부 실시태양에서, 본 발명은 본 발명에 제공된 목록/표 및 도면의 "미세유기체"또는 "세포유기체"또는 "미생물"을 의미한다. 이런 특성화는 표와 도면의 확인된 분류학적 속과 확인된 분류학적 종뿐만 아니라 상기 표 또는 도면의 임의의 유기체의 다양한 신규하고 새로운 확인된 또는 디자인된 균주를 의미할 수 있다. 동일한 특성화는 실시예에서와 같이, 명세서의 다른 부분에서 이런 용어의 인용에 대해서도 마찬가지이다.

[0109] 용어 "원핵 생물"은 당업계에 공지되어 있으며 핵 또는 다른 세포 기관을 함유하지 않는 세포를 의미한다. 원핵생물은 일반적으로 두 영역, 박테리아와 고세균의 하나로 분류된다. 고세균과 박테리아 영역의 유기체 사이의 명확한 차이는 16S 리보솜 RNA에서 뉴클레오티드 염기 서열의 근본적인 차이에 기초한다.

[0110] 용어 "고세균는 전형적으로 특이한 환경에서 발견되고 세포벽에서 리보솜 단백질의 수 및 뮤라민산의 부족을 포함하는 몇몇 기준에 의해 나머지 원핵 생물과 구별되는 멘도시쿠테스(Mendosicutes) 문의 유기체의 범주를 의미한다. ssrRNA 분석에 기초하여, 고세균은 두 계통발생학적으로 다른 그룹으로 구성된다: 크렌고세균(Crenarchaeota) 및 유리고세균(Euryarchaeota). 생리학을 기초로, 고세균은 세 가지 유형으로 구성될 수 있다: 메테인 생성균(methanogens)(메테인을 생성하는 원핵 생물); 고염성 세균(extreme halophiles)(매우 높은 농도의 염(NaCl)에서 사는 원핵 생물; 및 고온성(초고온성) 세균(extreme(hyper) thermophilus)(초고온에서 사는 원핵 생물). 박테리아와 구별되는 통일된 고세균의 특징(즉, 세포벽, 에스터-연결 막 지질 등에 뮤레인 없음)이

외에, 이런 원핵 생물은 이들의 특정한 서식 환경에 적응시키는 특이한 구조 또는 생화학적 특성을 나타낸다. 크렌고세균은 주로 초고온성 황 의존성 원핵 생물로 이루어지며 유리고세균은 메테인 생성균과 고염성 세균을 함유한다.

[0111] "박테리아" 또는 "진정세균(eubacteria)"는 원핵 생물의 영역을 의미한다. 박테리아는 다음과 같이 적어도 11개의 구별된 그룹을 포함한다: (1) 그람 양성(그람+) 박테리아, 2개위 주요 세부구분이 존재한다: (1) 높은 G+C 그룹(액티노마이세테스, 마이코박테리아, 마이르코콕커스, 기타)(2) 낮은 G+C 그룹(바실러스, 클로스트리디아, 락토바실러스, 스타필로콕키, 스트렙토콕키, 마이코플라스마스);(2) 프로테오박테리아, 예를 들어, 보라색 광합성 + 비 광합성 그람 음성 박테리아(대부분의 "일반적인" 그람 음성 박테리아 포함);(3) 사이아노박테리아, 예를 들면, 산소성 광영양생물;(4) 스피로체테스 및 관련 종;(5) 플랭크토마이세스;(6) 박테로이데스, 플라보박테리아;(7) 클라마이디아;(8) 녹색 황 박테리아;(9) 녹색 비 황 박테리아(또한 혐기성 광영양식물);(10) 방사성 저항성 마이크로콕키 및 동족;(11) 써모토가(Thermotoga) 및 써모시포 써모필레스(Thermosipho thermophiles).

[0112] 용어 "유전자 변형된 숙주 세포", "재조합 숙주 세포"및 및 "재조합 균주"는 본 발명에서 상호 교환적으로 사용되고 본 발명의 클로닝 및 형질전환 방법에 의해 유전자 변형된 숙주 세포를 의미한다. 따라서, 이 용어는 유전자 변경, 변형 또는 조작되어, 숙주 세포가 유래된 자연 발생 유기체와 비교하여 변경, 변형 또는 상이한 유전자형 및/또는 표현형을 나타내는(예를 들어, 유전자 변형이 미생물의 핵산 서열 암호화에 영향을 미칠 때) 숙주 세포(예를 들어, 박테리아, 효모 세포, 곰팡이 세포, CHO, 인간 세포 등)를 포함한다. 일부 실시태양에서, 이 용어는 문제의 특정 재조합 숙주 세포뿐만 아니라 이런 숙주 세포의 자손 또는 잠재적 자손을 의미하는 것으로 이해된다.

[0113] 용어 "야생형 미생물"또는 "야생형 숙주 세포"는 자연에서 발생하는 세포, 즉 유전자 변형되지 않은 세포를 기술한다.

[0114] 용어 "유전자 조작된"은(예를 들어, 삽입, 결실, 돌연변이 또는 핵산의 교체에 의한) 숙주 세포의 게놈의 임의적 조작을 의미할 수 있다.

[0115] 용어 "대조군"또는 "대조군 숙주 세포"는 유전자 변형 또는 실험적 치료의 효과를 측정하기 위한 적절한 비교기 숙주 세포를 의미한다. 일부 실시태양에서, 대조군 숙주 세포는 야생형 세포이다. 다른 실시태양에서, 대조군 숙주 세포는 치료 숙주 세포를 분화시키는 유전자 변형(들)을 제외하고, 유전자 변형된 숙주 세포와 유전적으로 동일하다. 일부 실시태양에서, 본 발명은 대조군 숙주 세포(예를 들어, 균주 개량 프로그램의 기초로 사용된 S_1 균주)로서의 부모 균주의 사용을 교시한다. 다른 실시태양에서, 숙주 세포는 치료 숙주 세포에서 테스트되는 특정 프로모터 또는 SNP가 결여된 유전적으로 동일한 세포일 수 있다.

[0116] 본 발명에 사용된 용어 "생산 균주" 또는 "생산 미생물"은 생산 균주의 성능을 개량시키는 야생형 또는 대조군 숙주 세포 유기체로부터의 하나 이상의 유전적 차이를 포함하는 숙주 세포를 의미한다(예를 들어, 이는 균주를 하나 이상의 화합물의 상업적 생산을 위한 더 나은 후보가 되게 한다). 일부 실시태양에서, 생산 균주는 상업적 생산에 현재 사용되는 균주일 것이다. 일부 실시태양에서, 생산 균주는 균주의 특성을 개량시키기 위해 하나 이상의 라운드의 돌연변이/유전자 조작을 거친 유기체일 것이다.

[0117] 본 발명에 사용된 용어 "대립 유전자(들)"은 유전자의 하나 이상의 대안 형태의 임의의 것을 의미하며, 이의 모두 대립 유전자는 적어도 하나의 형질 또는 특성과 관련된다. 이배체 세포에서, 소정의 유전자의 두 대립 유전자는 한 쌍의 상동 염색체 상에 상응하는 유전자좌를 차지한다.

[0118] 본 발명에 사용된 용어 "유전자좌"(복수 유전자좌)는 예를 들어 유전자 또는 유전자 마커가 발견되는 염색체 상의 특정 장소 또는 장소들 또는 위치를 의미한다.

[0119] 본 발명에 사용된 용어 "유전적으로 연결된"은 교차를 통해 분리하기가 어려워 번식 동안 높은 비율로 공동유전되는 2개 이상의 형질을 의미한다.

[0120] 본 발명에 사용된 바와 같이 "재조합"또는 "재조합 사건"은 염색체 교차 또는 독립된 분류를 의미한다.

[0121] 본 발명에 사용된 바와 같이 용어 "표현형"은 개체의 유전적 구성(즉, 유전자형)과 환경 사이의 상호작용으로부터 기인하는 개별 세포, 세포 배양, 유기체 또는 유기체의 그룹의 관찰 가능한 특성을 의미한다.

[0122] 본 발명에 사용된 바와 같이, 핵산 서열 또는 단백질 서열을 기술할 때 용어 "키메라" 또는 "재조합체"는 적어도 2개의 이종 폴리 뉴클레오티드 또는 2개의 이종 폴리펩티드를 단일 거대분자 속에 연결하거나 적어도 하나의

천연 핵산 또는 단백질 서열의 하나 이상의 요소를 재배열하는 핵산 또는 단백질 서열을 의미한다. 예를 들어, 용어 "재조합체"는 예를 들어, 화학적 합성 또는 유전 공학 기술에 의한 핵산의 분리된 단편의 조작에 의해 서열의 두 개의 분리된 단편의 인공적 조합을 의미할 수 있다.

[0123] 본 발명에 사용된 바와 같이 "합성 뉴클레오티드 서열"또는 "합성 폴리뉴클레오티드 서열"은 자연에서 발생하는 것으로 알려지지 않았거나 자연 발생적이지 않은 뉴클레오티드 서열이다. 일반적으로, 이런 합성 뉴클레오티드 서열은 임의의 다른 자연 발생 뉴클레오티드 서열과 비교할 때 적어도 하나의 뉴클레오티드 차이를 포함할 것이다.

[0124] 본 발명에 사용된 용어 "핵산"은 임의의 길이의 중합체 형태의 리보뉴클레오티드 또는 데옥시리보뉴클레오티드 또는 이의 유사체를 의미한다. 이 용어는 분자의 1 차 구조를 의미하며, 따라서 이중 나선 및 단일 가닥의 DNA 뿐 아니라 이중 및 단일 가닥의 RNA를 포함한다. 또한, 메틸화 및/또는 캡핑된 핵산과 같은 변형된 핵산, 변형된 염기를 함유하는 핵산, 골격 변형 등과 같은 변형 핵산을 포함한다. 용어 "핵산" 및 "뉴클레오티드 서열"은 상화 교환적으로 사용된다.

[0125] 본 발명에 사용된 용어 "유전자"는 생물학적 기능과 관련된 DNA의 임의의 단편을 의미한다. 따라서, 유전자는 암호화 서열 및/또는 그의 발현에 요구되는 조절 서열을 포함하나 이에 제한되지 않는다. 유전자는 또한, 예를 들어, 다른 단백질에 대한 인식 서열을 형성하는 비 발현 DNA 단편을 포함할 수 있다. 유전자는 관심 공급원으로부터의 클로닝 또는 공지되거나 예측된 서열 정보로부터의 합성을 포함하는 다양한 공급원으로부터 얻을 수 있고, 원하는 파라미터를 갖도록 디자인된 서열을 포함할 수 있다.

[0126] 본 발명에 사용된 용어 "상동(homologous)"또는 "동족체(homologue)"또는 "오르쏘로그(ortholog)"는 당업계에 공지되어 있고 공통 조상 또는 가족 구성원을 공유하고 서열 동일성의 정도에 기초하여 결정되는 관련 서열을 의미한다. 용어 "상동성", "상동 기관", "실질적으로 유사" 및 "상응하게 실질적으로"는 본 발명에서 상호 교환적으로 사용된다. 이들은 하나 이상의 뉴클레오티드 염기의 변화가 유전자 발현을 중재하거나 특정 표현형을 생성시키는 핵산 단편의 능력에 영향을 미치지 않는 핵산 단편을 의미한다. 이런 용어는 또한 초기의 변형되지 않은 단편에 비해 생성된 핵산 단편의 기능적 특성을 실질적으로 변화시키지 않는 하나 이상의 뉴클레오티드의 결실 또는 삽입과 같은 본 발명의 핵산 단편의 변형을 의미한다. 따라서, 당업자가 알 수 있는 바와 같이, 본 발명은 특정 예시적인 서열 이상을 포함하는 것으로 이해된다. 이런 용어는 한 종, 아종, 품종, 품종 또는 균주에서 발견된 유전자 및 다른 종, 아종, 품종, 품종 또는 균주에서 상응하는 또는 동등한 유전자 사이의 관계를 기술한다. 본 발명을 위해서, 상동성 서열이 비교된다. "상동 서열"또는 "상동체"또는 "오르쏘로그"는 기능적으로 관련이 있다고 생각되고, 믿거나 알려진다. 기능적 관계는(a) 서열 동일성 및/또는(b) 동일하거나 유사한 생물학적 기능을 포함하나 이에 제한되지 않는 다수의 방식 중 임의의 하나로 표시될 수 있다. 바람직하게는,(a) 및 (b) 모두가 표시된다. 상동성은 Current Protocols in Molecular Biology(F.M. Ausubel et al., eds., 1987) Supplement 30, 섹션 7.718, 표 7.71에서 논의된 바와 같은 당해분야에서 용이하게 이용 가능한 소프트웨어 프로그램을 사용하여 결정될 수있다. 일부 정렬 프로그램은 맥벡터(MacVector)(Oxford Molecular Ltd, Oxford, U.K.), ALIGN 플러스(Plus)(Scientific and Educational Software, Pennsylvania) 및 AlignX(Vector NTI, Invitrogen, Carlsbad, CA)이다. 다른 정렬 프로그램은 기본 매개변수를 사용하여 시퀀처(Sequencher)(Gene Codes, Ann Arbor, Michigan)이다.

[0127] 본 발명에 사용된 용어 "내인성"또는 "내인성 유전자"는 숙주 세포 게놈 내에서 자연적으로 발견되는 위치에서 자연 발생 유전자를 의미한다. 본 발명과 관련하여, 이종성 프로모터를 내인성 유전자에 작동 가능하게 연결시키는 것은 이 유전자가 자연적으로 존재하는 위치에서 기존 유전자 앞에 이종성 프로모터 서열을 유전적으로 삽입하는 것을 의미한다. 본 발명에 기재된 내인성 유전자는 본 발명의 방법 중 어느 하나에 따라 돌연변이된 자연 발생 유전자의 대립 유전자를 포함할 수있다.

[0128] 본 발명에 사용된 용어 "외인성"는 용어 "이종성(heterologous)"과 상호 교환적으로 사용되고, 자연 공급원 이외의 일부 공급원으로부터 유도하는 물질을 의미한다. 예를 들어, 용어 "외인성 단백질" 또는 "외인성 유전자"는 비 자연 공급원 또는 위치의 단백질 또는 유전자 및 생물학적 시스템에 공급된 단배질 또는 유전자를 의미한다.

[0129] 본 발명에 사용된 용어 "뉴클레오티드 변화"는 당업계에서 잘 알려진 바와 같이, 예를 들어 뉴클레오티드 치환, 결실 및/또는 삽입을 의미한다. 예를 들어 돌연변이는 침묵 치환, 추가 또는 결실을 생성하나 암호화된 단백질의 특성 또는 활성 또는 단백질이 어떻게 만들어지는지를 변형하지 않는 변경을 함유한다.

[0130] 본 발명에서 사용된 용어 "단백질 변형"은 당업계에 잘 알려진 바와 같이, 예를 들어 아미노산 치환, 아미노산 변형, 결실 및 또는 삽입을 의미한다.

[0131] 본 발명에 사용된 용어 핵산 또는 폴리펩타이드의 "적어도 일부" 또는 "단편"은 전장 분자를 포함하는 전장 분자의 이런 서열 또는 임의의 더 큰 단편의 최소 크기 특성을 갖는 부분을 의미한다. 본 발명의 폴리뉴클레오티드 단편은 유전자 조절 요소의 생물학적 활성 부분을 암호화할 수있다. 유전자 조절 요소의 생물학적 활성 부분은 유전자 조절 요소를 포함하는 본 발명의 폴리뉴클레오티드 중 하나의 일부를 분리하고 본 발명에 기재된 바와 같은 활성을 평가함으로써 제조될 수 있다. 유사하게, 폴리펩타이드의 일부는 전장 폴리펩타이드까지 이르는 4개 아미노산, 5개 아미노산, 6개 아미노산, 7개 아미노산 등일 수 있다. 사용될 부분의 길이는 특정 용도에 따라 다를 것이다. 하이브리드화 프로브로서 유용한 핵산의 일부는 12개 뉴클레오티드 정도로 짧을 수 있으며; 일부 실시태양에서, 이것은 20개 뉴클레오티드이다. 에피토프로서 유용한 폴리펩티드의 일부는 4개의 아미노산 정도로 짧을 수 있다. 전장 폴리펩타이드의 기능을 수행하는 폴리펩타이드의 일부는 일반적으로 4개 이상의 아미노산보다 길 수 있다.

[0132] 변이체 폴리뉴클레오티드는 또한 DNA 셔플링과 같은 돌연변이 및 재조합 절차로부터 유래된 서열을 포함한다. 그러한 DNA 셔플링을 위한 전략은 당업계에 공지되어 있다. 예를 들어, Stemmer(1994) PNAS 91:10747-10751; Stemmer(1994) Nature 370:389-391; Crameri *et al.*(1997) Nature Biotech. 15:436-438; Moore *et al.*(1997) J. Mol. Biol. 272:336-347; Zhang *et al.*(1997) PNAS 94:4504-4509; Crameri *et al.*(1998) Nature 391:288-291; 및 미국 특허 제5,605,793호 및 제5,837,458호 참조.

[0133] 본 발명에 개시된 폴리뉴클레오티드의 PCR 증폭을 위해, 임의의 관심 유기체로부터 추출된 cDNA 또는 게놈 DNA로부터의 상응하는 DNA 서열을 증폭시키기 위한 PCR 반응에 사용하기 위해 올리고뉴클레오티드 프라이머가 디자인될 수 있다. PCR 프라이머 및 PCR 클로닝을 디자인하기 위한 방법은 당업계에 일반적으로 공지되어 있으며, Sambrook *et al.*(2001) Molecular Cloning: A Laboratory Manual(3rd ed., Cold Spring Harbor Laboratory Press, Plainview, New York). See also Innis *et al.*, eds.(1990) PCR Protocols: A Guide to Methods and Applications(Academic Press, New York); Innis and Gelfand, eds.(1995) PCR Strategies(Academic Press, New York); 및 Innis and Gelfand, eds.(1999) PCR Methods Manual(Academic Press, New York)에 개시된다. PCR의 공지된 방법은 쌍을 이룬 프라이머, 네스티드 프라이머, 단일 특이적 프라이머, 축퇴성 프라이머, 유전자 특이적 프라이머, 벡터 특이적 프라이머, 부분적 불일치 프라이머 등을 사용하는 방법을 포함하나 이에 제한되지 않는다.

[0134] 본 발명에 사용된 용어 "프라이머"는 DNA 중합효소를 부착시키는 증폭 표적에 대한 어닐링을 행할 수 있어, 프라이머 연장 생성물의 합성이 유도되는 조건하에 놓일 때, 즉, 뉴클레오티드 및 DNA 중합효소와 같은 중합화제의 존재하에서 및 적절한 온도 및 pH에서 DNA 합성의 개시점으로서 작용하는 올리고뉴클레오티드를 의미한다. (증폭) 프라이머는 증폭 효율을 최대화하기 위해 바람직하게는 단일 가닥이다. 바람직하게는, 프라이머는 올리고데옥시리보뉴클레오티드이다. 프라이머는 중합화제의 존재 하에서 증량 생성물의 합성을 시작하기에 충분히 길어야한다. 프라이머의 정확한 길이는 프라이머의 온도 및 조성(A/T 대 G/C 함량)을 포함하는 많은 요소에 따라 달라질 것이다. 한 쌍의 양방향성 프라이머는 PCR 증폭과 같은 DNA 증폭 기술 분야에서 일반적으로 사용되는 것과 같은 하나의 순방향 및 역방향 프라이머로 구성된다.

[0135] 본 발명에 사용된 용어 "프로모터"는 암호화 서열 또는 기능성 RNA의 발현을 제어할 수있는 DNA 서열을 의미한다. 일부 실시태양에서, 프로모터 서열은 근위 및 더 원위 업스트림 요소로 구성되고, 후자의 요소는 종종 인핸서로 의미된다. 따라서, "인핸서"는 프로모터 활성을 자극할 수 있는 DNA 서열이며, 프로모터의 선천적인 요소 또는 프로모터의 수준 또는 조직 특이성을 향상시키기 위해 삽입된 이종성 요소일 수 있다. 프로모터는 천연 유전자로부터 완전히 유도되거나 자연계에서 발견되는 다른 프로모터로부터 유래된 상이한 요소로 구성되거나 심지어 합성 DNA 단편을 포함할 수 있다. 상이한 프로모터가 상이한 조직 또는 세포 유형, 또는 상이한 발달 단계 또는 상이한 환경 조건에 대한 반응으로 유전자의 발현을 지시할 수 있음은 당업자에게 이해된다. 또한, 대부분의 경우에, 조절 서열의 정확한 경계가 완전히 정의되지 않았기 때문에, 일부 변이체의 DNA 단편은 동일한 프로모터 활성을 가질 수 있다는 것이 추가로 인식된다.

[0136] 본 발명에서 사용된 용어 "재조합 구조체", "발현 구조체", "키메릭 구조체", "구조체" 및 "재조합 DNA 구조체"는 본 발명에서 상호 교환적으로 사용된다. 재조합 구조체는 천연에서 함께 발견되지 않는 조절 및 암호화 서열과 같은 핵산 단편의 인위적인 조합을 포함한다. 예를 들어, 키메릭 구조체는 상이한 공급원으로부터 유래된 조절 서열 및 암호화 서열, 또는 동일한 공급원으로부터 유도되지만 자연계에서 발견되는 것과 상이한 방식

으로 배열된 조절 서열 및 암호화 서열을 포함할 수 있다. 이러한 구조체는 그 자체로 사용되거나 벡터와 함께 사용될 수 있다. 벡터가 사용되는 경우 벡터의 선택은 당업자에게 주지된 바와 같이 숙주 세포를 형질전환하는데 사용될 방법에 의존한다. 예를 들어, 플라스미드 벡터가 사용될 수 있다. 당업자는 본 발명의 분리된 핵산 단편을 포함하는 숙주 세포를 성공적으로 형질전환, 선택 및 증식시키기 위해 벡터 상에 존재해야 하는 유전자 요소를 잘 알고 있다. 당업자는 또한 상이한 독립적인 형질전환 사건이 발현의 상이한 수준 및 패턴을 유도한다는 것을 인식할 것이며(Jones *et al.*,(1985) EMBO J. 4:2411-2418; De Almeida et al.,(1989) Mol. Gen Genetics 218:78-86), 따라서 다수의 사건은 원하는 발현 수준 및 패턴을 나타내는 라인을 수득하기 위해 선별돼야한다. 이러한 선별은 다른 것들 중에서, DNA의 서던 분석, mRNA 발현의 노던 분석, 단백질 발현의 면역 블로 팅 분석 또는 표현형 분석에 의해 실행될 수 있다. 벡터는 자율적으로 복제하거나 숙주 세포의 염색체에 통합될 수있는 플라스미드, 바이러스, 박테리오파지, 프로-바이러스, 파지미드, 트랜스포존, 인공 염색체 등일 수 있다. 벡터는 또한 자율적으로 복제하지 않는 네이키드 RNA 폴리뉴클레오티드, 네이키드 DNA 폴리뉴클레오티드, 동일한 가닥 내의 DNA 및 RNA 모두로 구성된 폴리뉴클레오티드, 폴리-라이신-컨쥬게이드된 DNA 또는 RNA, 펩타이드-컨쥬게이드된 DNA 또는 RNA, 리포좀-컨쥬게이드된 DNA 등일 수 있다. 본 발명에 사용된 용어 "발현"은 기능적 최종 산물, 예를 들어 mRNA 또는 단백질(전구체 또는 성숙)의 생산을 의미한다.

[0137] "작동 가능하게 연결된"은 추가 폴리뉴클레오티드의 전사를 초래하는 추가의 올리고뉴클레오티드 또는 폴리뉴클레오티드에 의한 본 발명에 따른 프로모터 폴리뉴클레오티드의 순차적 배열을 의미한다.

[0138] 본 발명에서 사용된 용어 "관심 생성물" 또는 "생체분자"는 원료로부터 미생물에 의해 생성된 임의의 생성물을 의미한다. 일부 경우에, 관심 생성물은 소분자, 효소, 펩타이드, 아미노산, 유기산, 합성 화합물, 연료, 알코올 등일 수 있다. 예를 들어, 관심 생성물 또는 생체분자는 임의의 1차 또는 2차 세포외 대사산물일 수 있다. 1차 대사산물은 특히 에탄올, 시트르산, 락트산, 글루탐산, 글루탐산염, 라이신, 스피노신, 스피네토람, 트레오닌, 트립토판 및 다른 아미노산, 비타민, 폴리사카라이드 등일 수 있다. 2차 대사산물은 특히 페니실린과 같은 항생물질 또는 사이클로스포린 A와 같은 면역 억제제, 지베렐린과 같은 식물 호르몬, 로바스타틴과 같은 스타틴 약물, 그리세오풀빈과 같은 살균제 등일 수 있다. 관심 생성물 또는 생체분자는 또한 카탈라아제, 아밀라아제, 펙티나아제, 글루코오스 아이소머라제, 셀룰라아제, 헤미셀룰라아제, 리파아제, 락타아제, 스트렙토 키나아제 및 여러 다른 것을 포함하는 미생물 효소와 같은 미생물에 의해 생성된 임의의 세포내 성분일 수 있다. 세포내 성분은 또한 인슐린, B형 간염 백신, 인터페론, 과립구 콜로니-자극 인자, 스트렙토키나아제 및 기타와 같은 재조합 단백질을 포함할 수 있다.

[0139] 용어 "탄소원"은 일반적으로 세포 성장을 위한 탄소원으로 사용되기에 적합한 물질을 의미한다. 탄소원은 바이오매스 가수분해물, 전분, 수크로오스, 셀룰로오스, 헤미셀룰로오스, 자일로오스 및 리그닌뿐만 아니라 이들 기질의 단량체 성분을 포함하나 이에 제한되지 않는다. 탄소원은 폴리머, 탄수화물, 산, 알코올, 알데하이드, 케톤, 아미노산, 펩타이드 등을 포함하나 이에 제한되지 않는 다양한 형태의 다양한 유기 화합물을 포함할 수 있다. 이들은, 예를 들어, 글루코오스, 덱스트로스(D-글루코오스), 말토오스, 올리고사카라이드, 폴리사카라이드, 포화 또는 불포화 지방산, 숙시네이트, 락테이트, 아세테이트, 에탄올 등, 또는 이의 혼합물과 같은 다양한 모노사카라이드를 포함한다. 광합성 유기체는 광합성의 산물로서 탄소원을 추가로 생산할 수 있다. 일부 실시태양에서, 탄소원은 바이오매스 가수분해물 및 글루코오스로부터 선택될 수 있다.

[0140] 용어 "공급원료"는 미생물 또는 발효 공정에 공급되는 원료 또는 원료의 혼합물로서 다른 생성물이 제조될 수 있는 것으로 정의된다. 예를 들어, 바이오매스 또는 바이오매스에서 유래된 탄소 화합물과 같은 탄소 공급원은 발효 공정에서 관심 생성물(예를 들어, 소분자, 펩타이드, 합성 화합물, 연료, 알코올 등)을 생산하는 미생물의 공급원료이다. 그러나, 공급원료는 탄소 공급원 이외의 영양분을 함 유할 수 있다.

[0141] 용어 "체적 생산성" 또는 "생산 속도"는 단위시간당 매질의 부피당 형성된 생성물의 양으로 정의된다. 체적 생산성은시간당 리터 당 그램(g/L/h)으로 보고될 수 있다.

[0142] 용어 "비 생산성"은 생성물의 형성 속도로 정의된다. 비 생산성은 본 발명에서시간당 세포 건조 중량의 그래당 그램 생성물(g/g CDW/h)의 비 생산성으로서 더 정의된다. 소정의 미생물에 대한 OD_{600}에 대한 CDW의 관계를 사용하여, 비 생산성은시간당 600nm(OD)(g/L/h/OD)에서 배양액의 광학 밀도당 배양 배지당 그램 생성물로 표현될 수 있다.

[0143] 용어 "수율"은 원료의 단위 중량당 얻어진 생성물의 양으로 정의되며 g 기질 당 g 생성물(g/g)로 표현될 수 있다. 수율은 이론적 수율의 백분율로 표현될 수있다. "이론적 수율"은 생성물을 제조하는데 사용된 신진대사 경

로의 화학양론에 따라 결정된 소정량의 기질당 생성될 수 있는 최대 생성물로 정의된다.

[0144] 용어 "역가(titre 또는 titer)"는 용액의 강도 또는 용액 속의 물질의 농도로 정의된다. 예를 들어, 발효액에서 관심 생성물(예를 들어, 소분자, 펩타이드, 합성 화합물, 연료, 알코올 등)의 역가는 발효액 1 리터당 용액 속 관심 제품의 g(g/L)로 기술된다.

[0145] 용어 "총 적가"는 용액 속의 관심 생성물, 적용 가능한 경우 기체상의 관심 생성물, 공정으로부터 제거되고 공정에서 최초 부피 또는 공정에서 작동 부피에 따라 회수된 임의의 관심 생성물을 포함하나 이에 제한되지 않는 공정에서 생산된 모든 관심 생성물의 합으로 정의된다.

[0146] 본 발명에 사용된 용어 "HTP 유전 디자인 라이브러리" 또는 "라이브러리"는 본 발명에 따른 유전자 교란의 집합을 의미한다. 일부 실시태양에서, 본 발명의 라이브러리는 i) 데이터베이스 또는 다른 컴퓨터 파일에서 서열 정보의 집합, ii) 상기한 일련의 유전자 요소를 코딩하는 유전자 구조체의 집합, 또는 iii) 상기 유전자 요소를 포함하는 숙주 세포 균주로서 입증될 수 있다. 일부 실시태양에서, 본 발명의 라이브러리는 개별 요소의 집합(예를 등러, PRO 스왑 라이브러리를 위한 프로모터 집합 또는 STOP 스왑 라이브러리를 위한 터미네이터 집합)을 의미할 수 있다. 다른 실시태양에서, 본 발명의 라이브러리는 프로모터::유전자, 유전자:터미네이터, 또는 심지어 프로모터:유전자:터미네이터의 조합과 같은 유전자 요소의 조합을 의미할 수 있다. 일부 실시태양에서, 본 발명의 라이브러리는 숙주 유기체에서 라이브러리의 각 구성원을 적용하는 효과와 관련된 메타 데이터를 추가로 포함한다. 예를 들어, 본 발명에서 사용된 라이브러리는 특정 종에서 하나 이상의 표현형에 대한 이들 조합의 얻어진 효과와 함께 프로모터::유전자 서열 조합의 집합을 포함할 수 있으며, 따라서 미래 프로모터 스왑에 상기 조합을 사용하는 미래 예측 가치를 개량시킨다.

[0147] 본 발명에서 사용된 용어 "SNP"는 작은 핵 다형성(들)을 의미한다. 일부 실시태양에서, 본 발명의 SNP는 광범위하게 해석되어야하며, 단일 뉴클레오티드 다형성, 서열 삽입, 결실, 역전 및 다른 서열 치환을 포함한다. 본 발명에서 사용된 용어 "비 동의" 또는 비 동의 SNP"는 숙주 세포 단백질에서 암호화 변화를 유도하는 돌연변이를 의미한다. 일부 실시태양에서 본 발명의 SNP는 하나 이상의 유전자의 추가 복제물(copy)(예를 들어, 생합성 효소 유전자를 코딩하는 하나 이상의 폴리뉴클레오티드의 복제물)을 포함한다.

[0148] 게놈 공학의 "고 처리량(HTP)" 방법은 상기 방법의 적어도 하나의 단계를 수행하기 위한 자동화 장비(예를 들어, 액체 핸들러 또는 플레이트 핸들러 장치)의 적어도 하나의 부품의 이용을 필요로 할 수 있다.

[0149] "상처없는 게놈 편집" 또는 "상처없는 유전자 교체"는 원하는 게놈 편집이 달성된 후 임의의 마커 서열 또는 임의의 플라스미드 백본 서열을 종의 게놈에 도입하지 않고 주어진 종의 특정 게놈 서열을 편집하는 방법을 의미한다. 게놈 편집은 게놈의 하나 이상의 핵산의 치환, 결실 및/또는 첨가일 수 있다.

[0150] **균주 개량의 전통적인 방법**

[0151] 균주 개량에 대한 전통적인 접근법은 크게 두 가지 유형의 접근법으로 분류 될 수 있다: 지시된(directed) 균주 조작 및 랜덤 돌연변이유발.

[0152] 균주 개량의 지시된 조작 방법은 특정 유기체의 소수의 유전자 요소의 계획된 교란을 필요로 한다. 이러한 접근법은 전형적으로 특정 생합성 또는 발달 프로그램을 조절하는데 중점을 두고 있으며, 상기 경로에 영향을 미치는 유전적 및 대사적 요인에 대한 사전 지식에 의존한다. 이의 가장 간단한 실시태양에서, 지시된 조작은 하나의 유기체로부터 동일하거나 상이한 종의 다른 유기체로 특성화된 형질(예를 들어, 유전자, 프로모터, 또는 측정 가능한 표현형을 생산할 수 있는 다른 유전 요소)의 전달을 필요로 한다.

[0153] 균주 조작에 대한 랜덤 접근법은 성능 개량을 확인하기 위한 광범위한 선별과 함께 부모 균주의 랜덤 돌연변이 유발을 필요로 한다. 이러한 랜덤 돌연변이를 일으키기 위한 접근법은 자외선 복사에너지 또는 에틸 메테인설포네이트와 같은 돌연변이유발 화학물질에 대한 노출을 포함한다. 랜덤적이며 매우 예측할 수 없음을 통해, 균주 개량에 대한 이런 전통적인 접근법은 보다 지시된 유전자 조작에 비해 몇 가지 장점을 가졌다. 첫째, 많은 산업 유기체는 유전자 및 대사성 레파토리의 관점에서 특성이 좋지 않아서(좋지 않게 유지되어), 불가능한 것은 아니지만 대안적인 직접적인 개량 접근법을 어렵게 만들었다.

[0154] 둘째, 비교적 특성이 규명된 시스템에서도, 산업 성능 개량을 초래하는 유전자형 변화는 예측하기 어렵고, 때때로 알려지고 알려지지 않은 기능의 많은 유전자에서 누적 돌연변이를 필요로 하는 상위성 표현형으로만 나타난다.

[0155] 또한, 수년 동안, 소정의 산업용 유기체에서 직접적인 게놈 돌연변이를 만드는데 필요한 유전자 도구는 이용할

수 없거나 사용이 매우 느리고 및/또는 어려웠다.

[0156] 그러나, 종래의 균주 개량 프로그램의 확장된 적용은 소정의 균주 계보에서 점진적으로 감소된 이득을 가져오고, 궁극적으로는 추가 균주 효율에 대한 소진된 가능성을 야기한다. 유익한 랜덤 돌연변이는 상대적으로 드문 경우이며 큰 선택 풀과 높은 돌연변이율을 필요로 한다. 이것은 필연적으로 "개량된" 균주에서 많은 중립적인 및/또는 해로운(또는 부분적으로 해로운) 돌연변이의 부주의한 누적을 초래하여, 궁극적으로 미래 효율성 향상에 대한 장애물을 형성한다.

[0157] 전통적인 누적 개량 접근법의 다른 한계는 임의의 균주 메트릭에 대한 임의의 특정 돌연변이의 효과에 대한 정보가 거의 또는 전혀 없다는 것이다. 이것은 근본적으로 유익한 돌연변이를 결합 및 통합하거나 중립적이거나 해로운 돌연변이유발성 "수하물(baggage)"을 제거하는 연구자의 능력을 제한한다.

[0158] 돌연변이성 계통 내에서 균주 간의 돌연변이를 랜덤으로 재조합시키는 다른 접근법 및 기술이 존재한다. 예를 들어, DNA 셔플링, 진화 또는 분자 육종으로 때때로 불리는 반복적 서열 재조합에 대한 몇몇 포맷 및 예가 미국 특허출원, Ser. No. 08/198,431, filed Feb. 17, 1994, Serial No. PCT/US95/02126, filed, Feb. 17, 1995, Ser. No. 08/425,684, filed Apr. 18, 1995, Ser. No. 08/537,874, filed Oct. 30, 1995, Ser. No. 08/564,955, filed Nov. 30, 1995, Ser. No. 08/621,859, filed. Mar. 25, 1996, Ser. No. 08/621,430, filed Mar. 25, 1996, Serial No. PCT/US96/05480, filed Apr. 18, 1996, Ser. No. 08/650,400, filed May 20, 1996, Ser. No. 08/675,502, filed Jul. 3, 1996, Ser. No. 08/721, 824, filed Sep. 27, 1996, and Ser. No. 08/722,660 filed Sep. 27, 1996; Stemmer, Science 270:1510(1995); Stemmer *et al.*, Gene164:49-53(1995); Stemmer, Bio/Technology 13:549-553(1995); Stemmer, Proc. Natl. Acad. Sci. U.S.A. 91:10747-10751(1994); Stemmer, Nature370:389-391(1994); Crameri *et al.*, Nature Medicine 2(1):1-3(1996); Crameri *et al.*, Nature Biotechnology 14:315-319(1996)에 기술되며, 이의 각각은 모든 목적을 위해 그 전문이 본 발명에 참조로 포함된다.

[0159] 이들은 돌연변이 균주를 통한 게놈 재조합을 용이하게 하는 원형질 융합 및 전체 게놈 셔플링과 같은 기술을 포함한다. 효모 및 사상균류와 같은 일부 산업용 미생물의 경우, 자연 교배 주기가 쌍을 이루는 방식의 게놈 재조합에도 이용될 수 있다. 이런 방식으로, 해로운 돌연변이는 부모 균주에 의한 돌연변이체를 "백-크로싱"하여 제거될 수 있고 유익한 돌연변이가 강화될 수 있다. 또한, 두 가지 다른 균주 계통의 유익한 돌연변이가 잠재적으로 결합될 수 있어서, 단일 균주 계통을 자체로 돌연변이시킴으로써 얻을 수 있는 것보다 추가적인 개량 가능성을 형성할 수 있다.

[0160] 전통적인 균주 개량 프로그램을 넘어서는 추가적인 개량을 제공하기 위해, 본 발명은 계산적으로 구동되고 분자생물학, 자동화, 데이터 분석 및 기계 학습 프로토콜을 통합하는 독특한 HTP 게놈 공학 플랫폼을 개시한다. 이 통합 플랫폼은 HTP 유전자 디자인 라이브러리를 구축하는데 사용되는 한 벌의 HTP 분자 도구 세트를 사용한다. 이러한 유전자 디자인 라이브러리는 아래에 자세히 설명될 것이다.

[0161] 교시된 HTP 플랫폼 및 이의 독특한 미생물 유전자 디자인 라이브러리는 미생물 균주 개발 및 진화의 패러다임을 근본적으로 변화시킨다. 예를 들어, 산업용 미생물 균주를 개발하는 전통적인 돌연변이유발 기반의 방법은 수년간의 랜덤 돌연변이유발 동안 축적된 중대한 돌연변이유발성 부하에 의해 부담을 진 미생물을 결국 유발할 것이다.

[0162] 이러한 문제점을 해결하는(즉, 이들 미생물에 의해 축적된 유전자 수하물을 제거하는) 능력은 수십 년 동안 미생물 연구자들을 회피하였다. 그러나, 본 발명에 개시된 HTP 플랫폼을 이용하여, 이들 산업용 균주는 "재활"될 수 있고 해로운 유전 자 돌연변이가 확인되고 제거될 수 있다. 부합하게, 유익한 것으로 확인된 유전자 돌연변이는 유지될 수 있고 일부 경우에 개량될 수 있다. 생성된 미생물 균주는 이들의 부모의 균주와 비교하여 우수한 표현형 형질(예를 들어, 관심 화합물의 개량된 생산)을 나타낸다.

[0163] 또한, 본 발명에서 교시된 HTP 플랫폼은 개별 돌연변이가 미생물 균주 성능에 미치는 효과를 확인, 특징화 및 정량화할 수 있다. 이 정보, 즉 소정의 유전자 변화 x가 숙주 세포 표현형 y에 대해 갖는 효과(예를 들어, 화합물 또는 관심 생성물의 생산)는 생성될 수 있고, 이후에 논의되는 미생물 HTP 유전자 디자인 라이브러리에 저장될 수 있다. 즉, 각 유전자 순열에 대한 서열 정보 및 숙주 세포 표현형에 대한 이의 효과는 하나 이상의 데이터베이스에 저장되며, 후속 분(예를 들어, 아래에서 논의되는 바와 같이 상위성 매핑)에 이용가능하다. 본 발명은 또한 유전자 삽입 구조체의 형태로, 또는 상기 유전자 순열을 함유하는 하나 이상의 숙주 세포 생물체의 형태로 귀중한 유전자 순열을 물리적으로 보호/저장하는 방법을 교시한다(예를 들어, 이하에서 논의되는 라이브러

리 참조).

[0164] 이들 HTP 유전 디자인 라이브러리를 정교한 데이터 분석 및 기계 학습 과정과 통합된 반복적 과정에 결합시키면 숙주 세포를 개량하기 위한 극적으로 상이한 방법론이 나타난다. 따라서, 교시된 플랫폼은 숙주 세포 균주를 개발하는 이전에 논의된 전통적인 방법과 근본적으로 상이하다. 교시된 HTP 플랫폼은 이전 방법과 관련된 많은 단점을 겪지 않는다. HTP 분자 도구 세트와 아래 논의된 유래된 유전자 디자인 라이브러리를 참조하면 이러한 장점과 다른 장점이 분명해질 것이다.

[0165] **유전자 디자인 및 미생물 조작: 한 벌의 HTP 분자 도구 및 HTP 유전자 디자인 라이브러리를 이용하는 균주 개량에 대한 체계적인 조합 접근법**

[0166] 상기한 바와 같이, 본 발명은 균주를 통한 유전자 변화의 반복적인 체계적 도입 및 제거를 통해 미생물을 조작하기 위한 신규한 HTP 플랫폼 및 유전자 디자인 전략을 제공한다. 이 플랫폼은 HTP 유전자 디자인 라이브러리의 생성을 가능하게 하고 소정의 숙주 균주 속에 유전자 변이의 효율적 수행을 가능하게 하는 한 벌의 분자 도구에 의해 지원된다.

[0167] 본 발명의 HTP 유전자 디자인 라이브러리는 특정 미생물 균주 배경 속에 도입될 수 있는 가능한 유전자 변형의 공급원으로 작용한다. 이러한 방식으로, HTP 유전자 디자인 라이브러리는 소정의 미생물 균주의 초기 또는 추가 조작에 적용될 수 있는 유전자 다양성의 저장소 또는 유전자 교란의 집합이다. 숙주 균주에 대한 실행을 위한 유전자 디자인을 프로그래밍하는 기술은 본 발명에 참조로 인용된 "Microbial Strain Design System and Methods for Improved Large Scale Production of Engineered Nucleotide Sequences,"이라는 명칭의 계류중인 미국 특허 출원, Serial No. 15/140,296에 기술된다.

[0168] 이 플랫폼에서 이용되는 HTP 분자 도구 세트는 특히 (1) 프로모터 스왑(PRO 스왑), (2) SNP 스왑, (3) 개시/중지 코돈 교환, (4) STOP 스왑, (5) 서열 최적화, (6) 트랜스포존 돌연변이유발 다양성 라이브러리, (7) 리보솜 결합 부위(RBS) 다양성 라이브러리, 및 (8) 항-대사산물 선택/발효 생성물 저항성 라이브러리를 포함할 수 있다. 본 발명의 HTP 방법은 또한 (9) 상위성 매핑 프로토콜을 포함하는 HTP 도구 세트의 통합/조합 사용을 지시하는 방법을 교시한다. 상기한 바와 같이, 이 한 벌의 분자 도구는 단독 또는 조합으로 HTP 유전자 디자인 숙주 세포 라이브러리의 형성을 가능하게 한다.

[0169] 입증된 바와 같이, 상기한 HTP 미생물 조작 플랫폼의 내용에서 상기한 HTP 유전자 디자인 라이브러리의 이용은 유익한 "원인성" 돌연변이 또는 유전자 절편의 확인 및 통합 및 수동적 또는 해로운 돌연변이 또는 유전자 절편의 확인 및 제거를 가능하게 한다. 이 새로운 접근법은 전통적인 랜덤 돌연변이유발 또는 지시된 유전자 조작으로는 달성할 수 없었던 균주 성능의 신속한 개량을 가능하게 한다. 유전자 부담의 제거 또는 유전자 부담이 없는 균주로의 유익한 변화의 통합은 추가적인 개량을 가능하게 할 수 있는 추가 랜덤 돌연변이유발을 위한 새롭고 강력한 출발점을 제공한다.

[0170] 일부 실시태양에서, 본 발명은 직교 유익한 변화가 돌연변이유발성 균주 계통의 다양한 이산 분지를 통해 확인됨에 따라, 이들은 또한 더 나은 실행 균주로 신속하게 통합될 수 있음을 교시한다. 이러한 돌연변이는 지시된 유전 공학 기술로 얻은 개량을 가진 균주와 같이 돌연변이유발 계통의 일부가 아닌 균주로 통합될 수 있다.

[0171] 일부 실시태양에서, 본 발명은 발현 및 비 발현 유전 요소를 포함하는 다수의 이종 게놈 영역에 걸친 돌연변이의 게놈 전체 조합 효과를 분석하고 수집된 정보(예를 들어, 실험 결과)를 사용하여 균주 향상을 일으키는 것으로 예상된 돌연변이 조합을 예측한다는 점에서 공지된 균주 개량 접근법과 상이하다.

[0172] 일부 실시태양에서, 본 발명은 i) 산업용 미생물 및 개시된 발명을 통한 개량에 순종하는 다른 숙주 세포, ii) 다운스트림 분석을 위한 다양성 풀 생성, iii) 고 처리량 선택 및 대형 변이체 풀의 서열화를 위한 방법 및 하드웨어 iv) 게놈 전체 돌연변이의 시너지 효과의 기계 학습 계산 분석 및 예측을 위한 방법 및 하드웨어, 및 v) 고 처리량 균주 조작을 위한 방법.

[0173] 다음 분자 도구 및 라이브러리는 예시적인 미생물의 예에 의해 논의된다. 당업자는 본 발명의 HTP 분자 툴이 진핵생물 세포 및 고차원 생명 형태를 포함하는 임의의 숙주 세포와 양립할 수 있다는 것을 인식할 것이다.

[0174] 미생물 조작 플랫폼에서 이용되는 다양한 HTP 유전자 디자인 라이브러리의 생성을 가능하게 하는 확인된 HTP 분자 도구 세트의 각각은 이제 논의될 것이다.

[0175] ***1. 프로모터 스왑: 프로모터 스왑 미생물 균주 라이브러리의 유도를 위한 분자 도구***

[0176] 일부 실시태양에서, 본 발명은 전체 숙주 균주 표현형에 유익한 효과를 나타내기 위해 최적 발현 특성(예를 들어, 수율 또는 생산성)을 갖는 프로모터를 선별하는 방법을 교시한다.

[0177] 예를 들어, 일부 실시태양에서, 본 발명은 다양한 발현 강도(예를 들어, 아래에서 논의된 프로모터 래더) 또는 뛰어난 조절 특성(예를 들어, 선택된 유전자에 대한 엄격한 조절 제어)을 나타내는 하나 이상의 프로모터를 확인 및/또는 숙주 세포 내에서 하나 이상의 프로모터의 변이체를 생성시키는 방법을 교시한다. 이들 확인된 및/또는 생성된 프로모터의 특정 조합은 프로모터 래더로서 함께 그룹화될 수 있으며, 이는 보다 상세히 이하에서 설명된다.

[0178] 당해 프로모터 래더는 소정의 관심 유전자와 결합된다. 따라서, 프로모터 P_1-P_8(다양한 발현 강도를 나타내는 것으로 확인된 및/또는 생성된 8개의 프로모터를 나타냄)이 있고 프로모터 래더를 미생물에서 관심 단일 유전자와 결합시킨다면(즉, 소정의 표적 유전자에 작동 가능하게 연결된 소정의 프로모터에 의해 미생물을 유전적으로 조작한다), 8개 프로모터의 각 조합의 효과는 조작된 미생물이 표적 유전자와 결합된 특정 프로모터(들)를 제외하고 그렇지 않다면 동일한 유전자 배경을 갖는 것을 고려하여, 각 조합적 노력으로부터 생성되는 각각의 조작된 균주를 특성화함으로써 확인될 수 있다.

[0179] 이 과정을 통해 조작되어 생성된 미생물은 HTP 유전자 디자인 라이브러리를 형성한다.

[0180] HTP 유전자 디자인 라이브러리는 이 과정을 통해 형성되는 실제의 물리적 미생물 균주 집합을 의미할 수 있으며, 각각의 구성원 균주는 그렇지 않다면 동일한 유전자 배경으로 특정 표적 유전자에 작동 가능하게 연결된 소정의 프로모터를 나타내며, 상기 라이브러리는 "프로모터 스왑 미생물 균주 라이브러리"로 불린다.

[0181] 또한, HTP 유전자 디자인 라이브러리는 유전자 교란의 집합-이 경우 소정의 유전자에 작동 가능하게 연결된 소정의 프로모터 x)을 의미할 수 있으며-상기 집합은 "프로모터 스왑 라이브러리"로 불린다.

[0182] 또한, 미생물을 조작하기 위해 표 1의 프로모터를 포함하는 동일한 프로모터 래더를 이용할 수 있으며, 여기서 프로모터의 각각은 상이한 유전자 표적에 작동 가능하게 연결된다. 이 절차의 결과는 관심 표적 유전자에 작동 가능하게 연결된 특정 프로모터를 제외하고, 그렇지 않다면 유전적으로 동일하게 간주되는 미생물이 될 것이다. 이런 미생물은 적절하게 선별되고 특징화되고 다른 HTP 유전자 디자인 라이브러리를 생성한다. HTP 유전자 디자인 라이브러리에서 미생물 균주의 특성화는 관계형 데이터베이스, 객체 지향 데이터베이스 또는 고도로 분산된 NoSQL 데이터베이스를 포함하여 임의의 데이터 저장 구조물에 저장될 수 있는 정보와 데이터를 생성한다. 이 데이터/정보는, 예를 들어, 소정의 유전자 표적에 작동 가능하게 연결된 경우 소정의 프로모터의 효과일 수 있다. 이 데이터/정보는 또한 둘 이상의 본 발명의 프로모터를 소정의 유전자 표적에 작동 가능하게 연결시킴으로써 초래하는 광범위한 세트의 조합적 효과일 수 있다.

[0183] 프로모터 및 표적 유전자의 상기 예는 개념이 다양한 발현 강도 및 임의의 소정 수의 표적 유전자의 제시에 기초하여 함께 그룹화된 임의의 소정의 수의 프로모터와 함께 적용될 수 있기 때문에, 단지 예시적이다. 당업자는 또한 임의의 유전자 표적 앞에서 둘 이상의 프로모터를 작동 가능하게 연결시키는 능력을 인식할 것이다. 따라서, 일부 실시태양에서, 본 발명은 프로모터 래더로부터의 1, 2, 3개 이상의 프로모터가 하나 이상의 유전자에 작동 가능하게 연결된 프로모터 스왑 라이브러리를 교시한다.

[0184] 요약하면, 유기체에서 다양한 유전자의 발현을 유도하는 다양한 프로모터를 이용하면 관심 형질을 최적화하는 강력한 수단이다. 본 발명자들이 개발한 프로모터 스와핑의 분자 도구는 적어도 하나의 조건하에서 적어도 하나의 유전자좌의 발현을 변화시키는 것으로 입증된 프로모터 서열의 래더를 사용한다. 그런 다음 이 래더는 고 처리량 게놈 공학을 사용하여 유기체에서 유전자 그룹에 체계적으로 적용된다. 이 유전자 그룹은 여러 가지 방법의 임의의 하나를 기초로 한 관심 형질에 영향을 줄 가능성을 갖는 것으로 결정된다. 이것은 공지된 기능에 기초한 선택 또는 관심 형질에 대한 영향 또는 이전에 결정된 유익한 유전자 다양성에 기초한 알고리즘 선택을 포함할 수 있다. 일부 실시태양에서, 유전자의 선택은 소정의 숙주 내 모든 유전자를 포함할 수 있다. 다른 실시태양에서, 유전자의 선택은 랜덤으로 선택된 소정의 숙주 내 모든 유전자의 서브세트일 수 있다.

[0185] 유전자에 연결된 프로모터 서열을 함유하는 생물의 생성된 HTP 유전자 디자인 미생물 균주 라이브러리는 고 처리량 선택 모델에서 성능에 대해 평가되고, 증가된 성능을 유도하는 프로모터-유전자 연관이 결정되고, 정보가 데이터베이스에 저장된다. 유전자 교란의 집합(즉, 소정의 유전자 y에 작동 가능하게 연결된 소정의 프로모터 x)은 "프로모터 스왑 라이브러리"를 형성하며, 이는 미생물 공학 처리에 이용될 잠재적인 유전자 교란의 공급원으로서 이용될 수 있다.시간이 지남에 따라, 유전자 교란의 더 큰 세트가 더 큰 다양성의 숙주 세포 배경에 대해 실행되기 때문에, 각 라이브러리는 임의의 관심 배경에 대한 표적화된 변화를 더욱 정확하고 예측되게 디자

인하는데 사용될 수 있는 실험적으로 확인된 데이터의 컬렉션으로서 더욱 강력해진다.

[0186] 유기체에서 유전자의 전사 수준은 유기체 행동에 영향을 미치는 제어의 핵심 포인트이다. 전사는 번역(단백질 발현)과 밀접하게 연관되며, 어떤 단백질은 유기체의 행동을 결정하는 양에서 발현된다. 세포는 수천 가지의 서로 다른 유형의 단백질을 발현하며, 이러한 단백질은 복잡한 복합적인 방식으로 상호작용하여 기능을 생성한다. 단백질 세트의 발현 수준을 체계적으로 변화시킴으로써, 기능이 복잡성으로 인해 예측하기 어렵도록 변형될 수 있다. 일부 변경은 성능을 증가시킬 수 있고 성능을 평가하기 위한 메커니즘과 연관되어, 이 기술은 개량된 기능을 가진 유기체의 생성을 가능하게 한다.

[0187] 소분자 합성 경로의 상황에서, 효소는 기질로 시작하여 관심 소형 분자로 끝나는 직선 또는 가지 사슬에 있는 소형 분자 기질 및 생성물을 통해 상호작용한다. 이러한 상호작용이 순차적으로 연결되기 때문에, 이 시스템은 분산된 제어를 나타내며, 하나의 효소의 발현을 증가시키면 다른 효소가 속도 제한될 때까지만 경로 유속을 증가시킬 수 있다.

[0188] 대사 조절 분석(MCA)은, 실험 데이터 및 제 1 원리로부터, 어느 효소 또는 효소들이 속도를 제한하는지를 결정하는 방법이다. 그러나, MCA는 새로운 속도 제한 효소를 결정하기 위한 각 발현 수준 변경 후에 광범위한 실험이 필요하기 때문에 제한적이다. 프로모터 스와핑은 이러한 상황에서 유리한데, 이는 이 경로에서 각 효소에 대한 프로모터 래더의 적용을 통해, 제한 효소가 발견되고, 속도 제한이 되는 새로운 효소를 찾기 위해 동일한 과정을 수행될 수 있기 때문이다. 또한, 기능에 대한 판독은 관심 소형 분자의 더 나은 생산이기 때문에, 어떤 효소가 제한되는지 결정하는 실험은 생산을 늘리는 공학과 동일하여, 개발시간을 단축시킨다. 일부 실시태양에서, 본 발명은 다중 단위 효소의 개별 서브유닛을 코딩하는 유전자에 대한 PRO 스왑의 적용을 교시한다. 또 다른 실시태양에서, 본 발명은 개개의 효소 또는 전체 생합성 경로를 조절할 책임이 있는 유전자에 PRO 스왑 기술을 적용하는 방법을 교시한다.

[0189] 일부 실시태양에서, 본 발명의 프로모터 스왑 도구는 선택된 유전자 표적의 최적 발현을 확인하는데 사용된다. 일부 실시태양에서, 프로모터 스왑의 목표는 대사 경로 또는 유전자 경로에서 병목을 감소시키기 위해 표적 유전자의 발현을 증가시키는 것일 수 있다. 다른 실시태양에서, 프로모터 스왑의 목표는 상기 표적 유전자의 발현이 요구되지 않을 때 숙주 세포에서 불필요한 에너지 소비를 피하기 위해 표적 유전자의 발현을 감소시키는 것일 수 있다.

[0190] 전사, 수송 또는 신호 전달과 같은 다른 세포 시스템의 상황에서, 다양한 합리적인 방법이 사용되어 어떤 단백질이 발현 변화의 표적이고 그 변화가 무엇이어야하는지를 선험적으로 시험하고 밝혀낼 수 있다. 이러한 합리적인 방법은 성능 개량을 위해 테스트해야하는 교란의 수를 줄이지만, 상당한 비용이 든다. 유전자 결실 연구는 이의 존재가 특정 기능에 필수적인 단백질을 확인하고 중요한 유전자는 과다 발현될 수 있다. 단백질 상호작용의 복잡성으로 인해, 성능을 향상시키는데 종종 효과가 없다. 세포에서 단백질 수준의 함수로서, 첫 번째 원칙으로부터, 전사 또는 신호전달 행동을 기술하려고 시도하는 다양한 유형의 모델이 개발되었다. 이러한 모델은 종종 발현 변화가 상이한 기능이나 개량된 기능으로 이어질 수 있는 표적을 제안한다. 이 모델의 기초가 되는 가설은 단순하고 매개변수는 측정하기가 어려워서, 특히 비-모델 유기체에 대해 이런 모델이 한 예측은 종종 부정확하다. 유전자 삭제와 모델링 모두에 의해, 어떻게 특정 유전자에 영향을 미치는 지를 결정하는데 필요한 실험은 성능을 개량시키는 변화를 만드는 후속 작업과는 상이하다. 프로모터 스와핑은 이러한 어려움을 회피하는데, 이는 특정 교란의 중요성을 강조하는 구성된 균주가 이미 개량된 균주이기 때문이다.

[0191] 따라서, 특정 실시태양에서, 프로모터 스와핑은 다음을 포함하는 다단계 공정이다:

[0192] 1. "래더"로서 작용하는 "x" 프로모터의 세트를 선택하는 단계. 이상적으로 이런 프로모터는 다수의 게놈 유전자좌 전체에서 매우 가변적인 발현을 유도하는 것으로 나타났지만, 유일한 요구조건은 일부 유전자 발현을 교란시키는 것이다.

[0193] 2. 표적화하기 위한 "n" 유전자의 세트를 선택하는 단계. 이 세트는 게놈의 모든 오픈 리딩 프레임(ORF) 또는 ORF의 서브세트일 수 있다. 서브세트는 기능과 관련된 ORF에 대한 주석을 사용하고, 이전에 입증된 유익한 교란(이전 프로모터 스왑 또는 이전 SNP 스왑)에 대한 관계에 의해, 이전에 생성된 교란 사이의 상위성 상호작용에 기초한 알고리즘 선택에 의해, 표적에 대한 유익한 ORF에 관한 가설에 기초한 다른 선택 기준 또는 랜덤 선택을 통해 선택될 수 있다. 다른 실시태양에서, "n" 표적 유전자는 비-암호화 RNA를 포함하는 비-단백질 암호화 유전자를 포함할 수 있다.

[0194] 3. 다음과 같은 유전자 변형을 신속하게- 및 일부 실시태양에서, 병렬-수행하는 고 처리량 균주 조작 단계: 천

연 프로모터가 표적 유전자 n의 앞에 존재하고 이의 서열이 알려질 때, 천연 프로모터를 래더에 있는x프로모터의 각각으로 교체한다. 천연 프로모터가 존재하지 않거나 이의 서열이 알려지지 않을 때, 유전자 n의 앞에 있는 래더에x프로모터의 각각을 삽입한다(예를 들어, 도 13 및 14 참조). 따라서, 일부 실시태양에서, SNP 스왑 라이브러리는 프로모터 삽입 라이브러리일 수 있으며, 프로모터가 없는 유전자 요소 또는 약한 프로모터를 갖는 유전자 요소는 새로 추가된 프로모터로 테스트된다. 프로모터 SWP 라이브러리 변형을 위한 이러한 유전자는 다음을 포함하나 이에 제한되지 않는다: (1) 스피노신과 같은 관심 화합물의 코어 생합성 경로의 유전자; (2) 전구체 합성 또는 풀 이용 가능성의 조절에 직접 관여하는 유전자와 같은 관심 화합물의 전구체 풀 이용 가능성에 관련된 유전자; (3) 보조인자(cofactor) 이용과 관련된 유전자; (4) 전사 조절제로 코딩하는 유전자; (5) 영양 공급원의 수송체를 코딩하는 유전자; 및 (6) 생성물 수출자 등. 이러한 방식으로, 균주의 "라이브러리"(HTP 유전자 디자인 라이브러리라고도 불림)가 구성되며, 라이브러리의 각 구성원은 그렇지 않으면 동일한 유전자 상황에서, n 표적에 작동 가능하게 연결된 x 프로모터의 한 예이다. 상기한 바와 같이, 프로모터의 조합이 삽입되어 라이브러리가 구성되는 조합 가능성의 범위를 확장시킬 수 있다.

[0195] 4. 하나 이상의 메트릭에 대한 성능이 최적화되는 성능을 나타내는 상황에서 균주의 라이브러리의 고 처리량 선택 단계.

[0196] 이 기본 과정은 특히 다음과 같은 방법으로 균주 성능에 추가 개량을 제공하는데 확장될 수 있다: (1) 반복 과정에서 한 번에 하나씩 또는 단일 단계에서 다중 변화로서 단일 균주 배경 속에 다수의 유익한 교란을 통합하는 단계. 다중 교란은 정의된 변경의 특정 세트 또는 부분적으로 랜덤화된 조합 라이브러리의 변화일 수 있다. 예를 들어, 표적 세트가 한 경로의 모든 유전자인 경우, 이전 라이브러리 균주의 개량된 구성원 또는 구성원들 속으로 교란의 라이브러리의 순차적 재생은 어느 유전자가 임의의 소정의 반복에서 속도를 제한하는지와 관계없이 한 경로에서 각 유전자의 발현 수준을 최적화할 수 있다;(2) 각 교란의 상호작용에 기초하여 최적의 교란 세트를 예측하기 위해 그 데이터를 사용하는 알고리즘 속에 라이브러리의 개별 및 조합 생성으로부터 성능 데이터를 제공하는 단계; 및(3) 위의 두 접근법의 조합을 실행하는 단계 (도 13 참조).

[0197] 상기 논의된 분자 도구 또는 기술은 프로모터 스와핑으로서 특징화되지만, 프로모터에 제한되지 않으며 표적 세트의 발현 수준을 체계적으로 변화시키는 다른 서열 변화를 포함할 수 있다. 한 세트의 유전자의 발현 수준을 변화시키는 다른 방법은 다음을 포함할 수 있다: a) 리보좀 결합 부위의 래더(또는 진핵생물에서 코작 서열); b) 각각의 표적의 개시 코돈을 각각의 다른 개시 코돈으로 교체하는 단계 (즉, 아래 논의된 개시/정지 코돈 교환); c) 다양한 mRNA 안정화 또는 불안정화 서열을 전사체의 5' 또는 3' 말단 또는 임의의 다른 위치에 부착하는 단계; d) 단백질 내의 임의의 위치에 다양한 단백질 안정화 또는 불안정화 서열을 부착하는 단계.

[0198] 이 접근법은 산업용 미생물에 의해 본 발명에서 예시되었지만, 원하는 형질이 유전자 돌연변이 집단에서 동정될 수 있는 임의의 유기체에 적용 가능하다. 예를 들어, 이것은 CHO 세포, 효모, 곤충 세포, 조류뿐만 아니라 식물과 같은 다세포 유기체의 성능을 개량시키는데 사용될 수 있다.

[0199] ***2. SNP 스왑:SNP 스왑 미생물 균주 라이브러리 유도를 위한 분자 도구***

[0200] 특정 실시태양에서, SNP 스와핑은 미생물 균주를 개량하기 위한 랜덤 돌연변이유발 접근법이 아니라 오히려 균주 전체에서 개별 소핵 다형질 뉴클레오티드 돌연변이(즉, SNP)의 체계적 도입 또는 제거(즉, "SNP 스와핑")를 필요로 한다.

[0201] 이 과정을 통해 조작된 미생물은 HTP 유전자 디자인 라이브러리를 형성한다.

[0202] HTP 유전자 디자인 라이브러리는 이 과정을 통해 형성된 실제의 물리적 미생물 균주 집합을 의미할 수 있으며, 각각의 구성원 균주는 그렇지 않으면 동일한 유전자 배경에서 소정의 SNP의 존재 또는 부재를 나타내며, 상기 라이브러리는 "SNP 스왑 미생물 균주 라이브러리"로 불린다.

[0203] 또한, HTP 유전자 디자인 라이브러리는 유전자 교란의 집합을 의미할 수 있으며-이 경우 소정의 SNP가 존재하거나 존재하지 않는다-상기 집합은 "SNP 스왑 라이브러리"로 불린다.

[0204] 일부 실시태양에서, SNP 스와핑은 확인된 유익한 성능 효과를 가진 표적 SNP "빌딩 블록"의 최적 조합에 의한 숙주 유기체의 재구성을 필요로 한다. 따라서, 일부 실시태양에서, SNP 스와핑은 다수의 유익한 돌연변이를 반복 과정에서 한 번에 하나씩 또는 단일 단계에서 다중 변화로서 단일 균주 배경 속에 다수의 유익한 교란을 통합하는 단계를 필요로 한다. 다중 변화는 정의된 변화의 특정한 세트 또는 돌연변이의 부분적으로 랜덤화된 라이브러리일 수 있다.

[0205] 다른 실시태양에서, SNP 스와핑은 또한 반복 과정에서 한 번에 하나씩 또는 단일 단계에서 다중 변화로서 균주로부터 유해한 것으로 확인된 다중 돌연변이를 제거하는 단계를 필요로 한다. 다중 변화는 정의된 변화의 특정한 세트 또는 돌연변이의 부분적으로 랜덤화된 라이브러리일 수 있다. 일부 실시태양에서, 본 발명의 SNP 스와핑 방법은 유익한 SNP의 첨가 및 유해 및/또는 중성 돌연변이의 제거를 모두 포함한다.

[0206] SNP 스와핑은 개량된 관심 형질에 대한 돌연변이유발 및 선택에 영향을 받은 균주의 계통에서 유익한 및 유해한 돌연변이 둘 다를 확인하고 이용하기 위한 강력한 도구이다. SNP 스와핑은 돌연변이유발성 계통에서 개별 돌연변이의 영향을 체계적으로 결정하기 위해 고 처리량 게놈 공학 기술을 사용한다. 게놈 서열은 알려진 성능 개량을 갖는 돌연변이유발성 계통의 하나 이상의 세대에 걸쳐 균주에 대해 결정된다. 고 처리량 게놈 공학은 초기 계통 균주에서 개량된 균주로부터 돌연변이를 반복하고, 및/또는 나중 균주의 돌연변이를 초기 균주 서열로 되돌리는데 체계적으로 사용된다. 이들 균주의 성능을 평가하고 개량된 관심 표현형에 대한 각각의 개별 돌연변이의 기여가 결정될 수 있다. 상기한 바와 같이, 이 과정으로부터 초래된 미생물 균주는 분석되고/특성화되어, 숙주 균주 전체에서 미생물 균주 개량을 알릴 수 있는 SNP 스왑 유전자 디자인 라이브러리의 기초를 형성한다.

[0207] 유해한 돌연변이의 제거는 즉각적인 성능 개량을 제공할 수 있고, 돌연변이유발 부담이 없는 균주 배경에 유익한 돌연변이의 통합은 균주 성능을 신속하고 크게 개량시킬 수 있다. SNP 스와핑 과정을 통해 생산된 다양한 미생물 균주는 HTP 유전자 디자인 SNP 스와핑 라이브러리를 형성하는데, 이는 다양한 추가/삭제/또는 통합된 SNP를 포함하지만 그렇지 않으면 동일한 유전자 배경을 가진 미생물 균주이다.

[0208] 이전에 논의된 바와 같이, 성능 개량을 위한 랜덤 돌연변이유발 및 후속 선별은 산업용 균주 개량을 위해 일반적으로 사용되는 기술이며, 대규모 제조에 현재 사용되는 많은 균주가 수년의 기간, 때때로 10년에 걸쳐 반복적으로 이 공정을 사용하여 개발되었다. UV 복사에너지 또는 에틸 메테인설포네이트와 같은 돌연변이유발 화학물질에 대한 노출과 같은 게놈 돌연변이를 생성하는 랜덤 접근법은 다음과 같은 이유로 산업용 균주 개량에 바람직한 방법이었다: 1) 산업용 유기체는 유전적으로 또는 대사적으로 특성이 좋지 않을 수 있으므로 지시된 개량 접근법에 대한 표적 선택을 어렵거나 불가능하게 한다; 2) 비교적 잘 특성화된 시스템에서도, 산업용 성능 개량을 초래하는 변화는 예측하기 어렵고 알려진 기능이 없는 유전자의 교란을 필요로 할 수 있다. 3) 소정의 산업용 유기체에서 지시된 게놈 돌연변이를 만들기 위한 유전자 도구는 사용할 수 없거나 매우 느리거나 및/또는 사용하기 어렵다.

[0209] 그러나, 상기 방법의 상기 이점에도 불구하고, 다수의 공지된 단점이 있다. 유익한 돌연변이는 비교적 드문 사건이며, 고정된 선택 용량으로 이러한 돌연변이를 찾기 위해서는, 돌연변이 속도가 충분히 높아야 한다. 이것이 원치 않는 중성 및 부분적으로 해로운 돌연변이가 유익한 변화와 함께 균주에 통합되는 것을 초래한다.시간이 지남에 따라 이런 '돌연변이유발 부담'은 증가되어, 전반적인 견고함과 성장 속도와 같은 주요 형질이 결핍된 균주를 초래한다. 궁극적으로 '돌연변이유발 부담'은 랜덤 돌연변이유발을 통한 성능의 개량을 얻기가 점차 더 어렵거나 불가능하게 한다. 적절한 도구가 없다면, 균주 계통의 다른 및 유사한 부문에서 발견된 유익한 돌연변이를 통합하는 것은 불가능하다.

[0210] SNP 스와핑은 돌연변이유발 계통 내의 균주를 비교할 때 관찰된 일부 또는 모든 돌연변이를 체계적으로 반복하거나 회귀시킴으로써 이러한 제한을 극복하는 접근법이다. 이 방법으로 유익한('원인성') 돌연변이가 확인되고 통합될 수 있으며 및/또는 해로운 돌연변이가 확인되고 제거될 수 있다. 이것은 추가 랜덤 돌연변이유발 또는 표적화된 유전자 공학에 의해 달성될 수 없는 균주 성능의 신속한 개량을 허용한다.

[0211] 유전자 부담의 제거 또는 유전자 부담이 없는 균주로의 유익한 변화의 통합은 추가 개량을 가능하게 할 수 있는 추가의 랜덤 돌연변이유발을 위한 새롭고 견고한 출발점을 제공한다.

[0212] 또한, 수직 유익한 변화가 돌연변이유발성 균주 계통의 다양하고, 구별된 분지를 통해 확인되므로, 이들 균주는 보다 우수한 실행 균주로 신속하게 통합될 수있다. 이런 돌연변이는 지시된 유전 공학에 의해 얻은 개량을 가진 균주와 같이 돌연변이유발 계통의 일부가 아닌 균주로 통합될 수 있다.

[0213] 돌연변이유발 계통 내에서 균주 간의 돌연변이를 랜덤으로 재조합시키는 다른 접근법 및 기술이 존재한다. 이들은 원형질체 융합 및 돌연변이 균주 전체에서 게놈 재조합을 촉진하는 전체 게놈 셔플링과 같은 기술을 포함한다. 효모 및 사상 균류와 같은 일부 산업용 미생물의 경우, 자연 교배 주기가 쌍을 이루는 게놈 재조합에 이용될 수 있다. 이런 식으로, 해로운 돌연변이는 부모 균주에 의한 "백-크로싱(back-crossing)"에 의해 제거될 수 있고 유익한 돌연변이는 통합될 수 있다.

[0214] 전통적인 접근법은 본 발명에 개시된 SNP 스와핑 방법과 함께 사용되어 랜덤 돌연변이 발견을 균주에 걸쳐 개별

돌연변이의 체계적인 도입 또는 제거와 조합할 수 있다.

[0215] 일부 실시태양에서, 본 발명은 다양성 풀의 유기체들 사이에 존재하는 SNP 서열 다양성을 확인하는 방법을 교시한다. 다양성 풀은 분석을 위해 사용된 소정의 수 n의 미생물일 수 있으며, 상기 미생물의 게놈은 "다양성 풀"을 나타낸다.

[0216] 특정 양태에서, 다양성 풀은 특정 시점(S_1Gen_1)에서 "기준선" 또는 "표준" 유전자 서열을 갖는 본래의 부모 균주(S_1) 및 상기 S_1 균주로부터 유도/성장되고 S_1의 기준선 게놈과 관련하여 다른 게놈($S_{2-n}Gen_{2-n}$)을 갖는 임의의 수의 후속 자손 균주(S_{2-n})일 수 있다.

[0217] 예를 들어, 일부 실시태양에서, 본 발명은 다양성 풀에서 미생물 게놈을 서열화하여 각 균주에 존재하는 SNP를 확인하는 것을 교시한다. 한 실시태양에서, 다양성 풀의 균주는 역사적인 미생물 생산 균주이다. 따라서, 본 발명의 다양성 풀은 예를 들어 산업용 표준 균주 및 전통적인 균주 개량 프로그램을 통해 생산된 하나 이상의 돌연변이 산업용 균주를 포함할 수 있다.

[0218] 일부 실시태양에서, 다양성 풀 내의 SNP는 "표준 균주"를 참조하여 결정된다. 일부 실시태양에서, 표준 균주는 야생형 균주이다. 다른 실시태양에서, 표준 균주는 임의의 돌연변이유발을 받기 전에 원래의 산업용 균주이다. 표준 균주는 실무자에 의해 정의될 수 있으며 원래의 야생형 균주 또는 원래의 산업용 균주일 필요는 없다. 기본 균주는 단순히 "기본", "표준" 또는 원래의 유전자 배경으로 간주 될 수 있는 것의 대표이며, 이로 인해 상기 표준 균주로부터 유도되거나 개발된 후속 균주는 비교될 것이다.

[0219] 다양성 풀 내의 모든 SNPS가 확인되면, 본 발명은 개별적으로 및/또는 그룹으로 SNP의 효과(예를 들어, 관심 표현형의 생성)를 묘사(즉, 정량화 및 특성화)하기 위한 SNP 스와핑 및 선택 방법을 교시한다.

[0220] 일부 실시태양에서, 본 발명의 SNP 스와핑 방법은 돌연변이된 균주(예를 들어, $S_{2-n}Gen_{2-n}$으로부터 균주)에서 확인된 하나 이상의 SNP를 표준 균주(S_1Gen_1) 또는 야생형 균주에 도입하는 단계 ("웨이브 업")를 포함한다.

[0221] 다른 실시태양에서, 본 발명의 SNP 스와핑 방법은 돌연변이된 균주(예를 들어, $S_{2-n}Gen_{2-n}$으로부터 균주)에서 확인된 하나 이상의 SNP를 제거하는 단계 ("웨이브 다운")를 포함한다.

[0222] 일부 실시태양에서, 하나 이상의 SNP 변화(도입 또는 제거)를 포함하는 각각의 생성된 균주는 본 발명의 하나 이상의 기준(예를 들어, 관심 화학물질 또는 생성물의 생산)하에서 배양되고 분석된다. 분석된 각 숙주 균주로부터의 데이터는 특정 SNP 또는 숙주 균주에 존재하는 SNP 그룹과 관련되거나 연관되며 미래 사용을 위해 기록된다. 따라서, 본 발명은 임의의 수의 관심 미생물 유전자 또는 표현형 형질에 대한 소정의 SNP의 효과를 확인할 수 있는 크고 고도로 주석화된 HTP 유전자 디자인 미생물 균주 라이브러리의 생성을 가능하게 한다. 이런 HTP 유전자 디자인 라이브러리에 저장된 정보는 HTP 게놈 공학 플랫폼의 기계 학습 알고리즘에 정보를 제공하고 과정의 미래 반복을 지시하여, 궁극적으로 매우 바람직한 특성/형질을 갖는 진화된 유기체를 유도한다.

[0223] 일부 실시태양에서, 본 발명에 기재된 방법은 정방향 유전학 공정에서 수행될 수 있다. 예를 들어, 일부 실시태양에서, SNP 또는 다른 유형의 유전자 변이를 함유하는 유전자의 기능 및/또는 신원은 알려지지 않았거나, 어느 SNP 또는 다른 유전자 변이가 스왑 또는 조합되는지를 결정할 때 고려되지 않는다. 대신에, 유전자 변이의 조합은 공지된 또는 예측된 유전자 기능을 고려하지 않고 만들어 지지만, 이전 균주 성능의 인간 또는 기계 학습 분석에 의해 영향을 받을 수 있다. 단일 이론에 구애받지 않고, 본 발명자는 기능적으로 불가지론적 선별이 인간의 선입견 및 기대에 의해 제한되지 않기 때문에 효과적이라고 생각한다. 따라서, 일부 실시태양에서, 본 발명의 방법은 유전 공학에 대한 "지능적 디자인" 접근법으로 고려되지 않았을(그리고 심지어 이에 의해 낙담했을지도 모른다) 유전자 변이의 귀중한 조합의 발견을 허용한다.

[0224] 일부 실시태양에서, 본 발명에 기재된 방법은 역방향 유전학 공정에서 수행될 수 있다. 예를 들어, 일부 실시태양에서, SNP 또는 다른 유형의 유전자 변이를 함유하는 유전자의 기능 및/또는 신원은 이미 알려져 있고, 어느 SNP 또는 다른 유전자 변이가 스왑되는지를 결정할 때 고려된다. 예를 들어, 일부 실시태양에서, 관심 화합물(예를 들어, 스피노신)의 합성, 전환 및/또는 분해에 관여하는 유전자의 유전자 변이가 특히 선택되고 조합되며, 왜 이러한 조합이 원하는 표현형을 갖는 개량된 균주를 야기할 수 있는지에 대한 적어도 일부의 가설을 가진다. 이러한 유전자 기능 및/또는 신원 정보는 다음을 포함하나 이에 제한되지 않는다: (1) 스피노신과 같은 관심 화합물의 코어 생합성 경로의 유전자; (2) 전구체 합성 또는 풀 이용 가능성의 조절에 직접 관여하는 유전자와 같은 관심 화합물의 전구체 풀 이용 가능성에 관련된 유전자; (3) 보조인자(cofactor) 이용과 관련된

유전자; (4) 전사 조절제로 코딩하는 유전자; (5) 영양 공급원의 수송체를 코딩하는 유전자; 및 (6) 생성물 수출자 등.

[0225] 일부 실시태양에서, 본 발명에 기재된 방법은 하이브리드 공정에서 수행될 수 있으며, 여기서 적어도 하나의 유전자의 기능 및/또는 신원 또는 유전자 변이가 고려되는 반면, 유전자 변이가 조합될 때 다른 유전자 변이를 함유하는 적어도 하나의 유전자의 기능 및/또는 신원은 고려되지 않는다.

[0226] 특정 유전자는 코딩 DNA 모듈의 반복 세그먼트를 함유한다. 예를 들어, 폴리케타이드 및 비-리보솜 펩타이드는 모듈성을 갖는 것으로 밝혀졌다(US2017/0101659 참조, 전체 내용이 참조로 포함됨). 이러한 단백질의 기능성 단백질 도메인은 반복적인 방식으로 배열되며(모듈 1-모듈 2-모듈 3...) 게놈 상의 DNA의 반복 세그먼트를 초래한다. 일부 실시태양에서, 조합될 적어도 하나의 유전자 변이는 DNA 모듈을 코딩하는 반복 세그먼트를 함유하는 게놈 영역에 있지 않다. 일부 실시태양에서, 유전자 변이의 조합은 이러한 유전자에서 코딩 DNA 모듈의 반복 세그먼트의 치환, 결실 또는 첨가를 포함하지 않는다. 본 발명의 방법은 이러한 게놈 모듈성의 영역에서 표적화된 게놈 편집을 수행할 수 있을 뿐만 아니라, 임의의 게놈 상황에서 게놈에 걸쳐 표적화된 게놈 편집을 가능하게 한다. 결과적으로, 본 발명의 표적화된 게놈 편집은 임의의 영역에서 S. 스피노사 게놈을 편집할 수 있으며, 모듈성을 갖는 영역에서의 편집에만 국한되지 않는다.

[0227] ***3. 개시/정지 코돈 교환: 개시/정지 코돈 미생물 균주 라이브러리 유도를 위한 분자 도구***

[0228] 일부 실시태양에서, 본 발명은 개시 및 정지 코돈 변이체를 스와핑하는 방법을 교시한다. 예를 들어, 에스. 크레비시애(*S. cerevisiae*) 및 포유동물에 대한 전형적인 정지 코돈은 각각 TAA(UAA) 및 TGA(UGA)이다. 단핵구 식물의 전형적인 정지 코돈은 TGA(UGA)인 반면, 곤충과 대장균은 보통 TAA(UAA)를 정지 코돈으로 사용한다(Dalphin et al.(1996) Nucl. Acids Res. 24: 216-218). 다른 실시태양에서, 본 발명은 TAG(UAG) 정지 코돈의 사용을 교시한다.

[0229] 본 발명은 스와핑 개시 코돈을 유사하게 교시한다. 일부 실시태양에서, 본 발명은 대부분의 유기체(특히, 진핵생물)에 의해 이용되는 ATG(AUG) 개시 코돈의 사용을 교시한다. 일부 실시태양에서, 본 발명은 원핵 생물이 ATG(AUG)를 가장 많이 사용하고 이어서 GTG(GUG) 및 TTG(UUG)를 사용하는 것을 교시한다.

[0230] 다른 실시태양에서, 본 발명은 ATG 개시 코돈을 TTG로 교체시키는 것을 교시한다. 일부 실시태양에서, 본 발명은 ATG 개시 코돈을 GTG로 교체시키는 것을 교시한다. 일부 실시태양에서, 본 발명은 GTG 개시 코돈을 ATG로 교체하는 것을 교시한다. 일부 실시태양에서, 본 발명은 GTG 개시 코돈을 TTG로 교체하는 것을 교시한다. 일부 실시태양에서, 본 발명은 TTG 개시 코돈을 ATG로 교체하는 것을 교시한다. 일부 실시태양에서, 본 발명은 TTG 개시 코돈을 GTG로 교체하는 것을 교시한다.

[0231] 다른 실시태양에서, 본 발명은 TAA 정지 코돈을 TAG로 교체하는 것을 교시한다. 일부 실시태양에서, 본 발명은 TAA 정지 코돈을 TGA로 교체하는 것을 교시한다. 일부 실시태양에서, 본 발명은 TGA 정지 코돈을 TAA로 교체하는 것을 교시한다. 일부 실시태양에서, 본 발명은 TGA 정지 코돈을 TAG로 교체하는 것을 교시한다. 일부 실시태양에서, 본 발명은 TAG 정지 코돈을 TAA로 교체하는 것을 교시한다. 일부 실시태양에서, 본 발명은 TAG 정지 코돈을 TGA로 교체시키는 것을 교시한다.

[0232] ***4. 스왑 중지: 최적화된 서열 미생물 균주 라이브러리 유도를 위한 분자 도구***

[0233] 일부 실시태양에서, 본 발명은 세포 유전자 전사의 최적화를 통해 숙주 세포 생산성을 개량시키는 방법을 교시한다. 유전자 전사는 전사 개시(RNAp 모집 및 전사 복합체 형성), 신장(가닥 합성/확장) 및 전사 종결(RNAp 분리 및 종결)을 포함하는 여러 별개의 생물학적 현상의 결과이다.(예를 들어, 프로모터를 변화시키거나 조절 전사 인자를 유도함으로써) 유전자의 전사 조절을 통한 유전자 발현의 제어에 많은 관심이 집중되었지만, 유전자 터미네이터 서열의 조절을 통한 전사의 조절에 대해서는 비교적 적은 노력이 있었다 .

[0234] 전사가 유전자 발현 수준에 영향을 미치는 가장 명백한 방법은 프로모터 또는 인핸서 강도 및 트랜스 활성화 인자의 조합에 의해 조절될 수있는 Pol II 개시 속도를 이용하는 것이다(Kadonaga, JT. 2004 "Regulation of RNA 폴리머라제 II transcription by sequence-특이적 DNA binding factors" Cell. 2004 Jan 23; 116(2):247-57). 진핵생물에서, 연신율은 교차 스플라이싱에 영향을 줌으로써 유전자 발현 양상을 결정할 수있다(Cramer P. et al., 1997 "Functional association between promoter structure and transcript alternative splicing." Proc Natl Acad Sci U S A. 1997 Oct 14; 94(21):11456-60). 유전자에 대한 실패한 종결은 Pol II에 대한 프로모터의 접근성을 감소시킴으로써 다운스트림 유전자의 발현을 손상시킬 수 있다(Greger IH. et al., 2000 "Balancing transcriptional interference 및 initiation on the GAL7 promoter of Saccharomyces

cerevisiae." Proc Natl Acad Sci U S A. 2000 Jul 18; 97(15):8415-20). 전사 간섭으로 알려진 이 과정은 특히 진핵 생물과 관련이 있는데 이는 이들이 종종 밀접하게 이격된 유전자를 갖기 때문이다.

[0235] 종결 서열은 또한 서열이 속하는 유전자의 발현에 영향을 미칠 수 있다. 예를 들어, 연구들은 진핵생물에서 비효율적인 전사 종결은 비접합 pre-mRNA의 축적을 초래한다는 것을 보여준다(West, S., and Proudfoot, N.J., 2009 "Transcriptional Termination Enhances protein Expression in Human Cells" Mol Cell. 2009 Feb 13; 33(3-9); 354-364 참조). 다른 연구들은 또한 3'말단 처리가 비효율적인 종결에 의해 지연될 수 있다는 것을 보여주었다(West, S et al., 2008 "Molecular dissection of mammalian RNA 폴리머라제 II transcriptional termination." Mol Cell. 2008 Mar 14; 29(5):600-10.). 전사 종결은 또한 합성 부위로부터 전사체를 방출함으로써 mRNA 안정성에 영향을 미칠 수 있다.

[0236] ***원핵생물에서 전사의 종결***

[0237] 원핵생물에서, Rho-비의존성 및 Rho-의존성 종결로 불리는 두 주요 메카니즘이 전사 종결을 매개한다. Rho-의존성 종결 신호는 외부의 전사 종결 인자를 필요로하지 않는데, 이는 일련의 우리딘(U) 잔기와 함께 이들 서열로부터 전사된 RNA에서 스템-루프 구조의 형성이 전사 복합체로부터 RNA 사슬의 방출을 촉진하기 때문이다. 한편, Rho 의존성 종결은 Rho라 불리는 전사 종결 인자 및 mRNA 상의 시스 작용 물질을 필요로 한다. Rho에 대한 초기 결합 위치인 Rho 활용(rut) 위치는 높은 실제 터미네이터 서열의 업스트림에서 합성되는 RNA에서 사이티딘/낮은 구아노신 함량 및 비교적 적은 2차 구조를 특징으로 하는 확장된(~70개 뉴클레오티드, 때때로 80-100개 뉴클레오티드) 단일 가닥 영역이다. 폴리머라제 일시 중지 위치와 마주칠 때, 종결이 일어나고 전사체는 Rho의 헬라카제 활성에 의해 방출된다.

[0238] ***터미네이터 스와핑(STOP 스왑)***

[0239] 일부 실시태양에서, 본 발명은 전체 숙주 균주 생산성에 유익한 효과를 생성하기 위한 최적 발현 성질을 갖는 종결 서열("터미네이터")을 선택하는 방법을 교시한다.

[0240] 예를 들어, 일부 실시태양에서, 본 발명은 다양한 발현 강도(예를 들어, 하기 논의되는 터미네이터 래더)를 나타내는 하나 이상의 터미네이터를 확인하고 및/또는 숙주 세포 내에서 하나 이상의 터미네이터의 변이체를 생성하는 방법을 교시한다. 이들 확인된 및/또는 생성된 터미네이터의 특정 조합은 터미네이터 래더로서 함께 그룹화될 수 있으며, 이는 보다 상세히 이하에서 설명된다.

[0241] 당해 터미네이터 래더는 소정의 관심 유전자와 결합된다. 따라서, 터미네이터 T_1-T_8(다양한 발현 강도를 나타내는 것으로 확인된 및/또는 생성된 8개의 터미네이터를 나타냄)이 있고 터미네이터 래더를 숙주 세포에서 관심 단일 유전자와 결합시킨다면(즉, 소정의 표적 유전자에 작동 가능하게 연결된 소정의 터미네이터에 의해 숙주 세포 유전적으로 조작한다), 8개 터미네이터의 각 조합의 효과는 조작된 숙주 세포가 표적 유전자와 결합된 특정 프로모터(들)를 제외하고 그렇지 않다면 동일한 유전자 배경을 갖는 것을 고려하여, 각 조합적 노력으로부터 생성되는 각각의 조작된 균주를 특성화함으로써 확인될 수 있다. 이 과정을 통해 조작되어 생성된 숙주 세포는 HTP 유전자 디자인 라이브러리를 형성한다.

[0242] HTP 유전자 디자인 라이브러리는 이 과정을 통해 형성되는 실제의 물리적 미생물 균주 집합을 의미할 수 있으며, 각각의 구성원 균주는 그렇지 않다면 동일한 유전자 배경으로 특정 표적 유전자에 작동 가능하게 연결된 소정의 터미네이터를 나타내며, 상기 라이브러리는 "터미네이터 스왑 미생물 균주 라이브러리" 또는 "STOP 스왑 미생물 균주 라이브러리"로 불린다.

[0243] 또한, HTP 유전자 디자인 라이브러리는 유전자 교란의 집합을 의미할 수 있으며-이 경우 소정의 유전자 y에 작동 가능하게 연결된 소정의 터미네이터 x-상기 집합은 "STOP 스왑 라이브러리"로 불린다.

[0244] 또한, 미생물을 조작하기 위해 프로모터 T_1-T_8을 포함하는 동일한 터미네이터 래더를 이용할 수 있으며, 8개의 터미네이터 각각은 10개의 상이한 유전자 표적에 작동 가능하게 연결된다. 이 절차의 결과는 관심 표적 유전자에 작동 가능하게 연결된 특정 터미네이터를 제외하고, 그렇지 않으면 유전적으로 동일하게 간주되는 80개의 숙주 세포 균주일 것이다. 이런 80개의 숙주 균주는 적절하게 선별되고 특성화될 수 있으며 다른 HTP 유전자 디자인 라이브러리를 생성할 수 있다. HTP 유전자 디자인 라이브러리에 미생물 균주의 특성화는 관계형 데이터베이스, 객체 지향 데이터베이스 또는 고도로 분산된 NoSQL 데이터베이스를 포함하나 이에 제한되지 않는 임의의 데이터베이스에 저장될 수 있는 정보와 데이터를 생성한다. 이런 데이터/정보는, 예를 들어, 소정의 유전자 표적에 작동 가능하게 연결될 때 소정의 터미네이터(예를 들어, T_1-T_8) 효과를 포함할 수 있다. 이 데이터/정보는 또

한 둘 이상의 프로모터 T_1-T_8을 소정의 유전자 표적에 작동 가능하게 연결시킴으로써 초래되는 보다 광범위한 세트의 조합 효과일 수 있다.

[0245] 상기 8개의 터미네이터 및 10개의 표적 유전자의 예는 단지 예시적인데, 이는 개념이 다양한 발현 강도 및 소정 수의 표적 유전자의 표시에 기초하여 함께 그룹화된 임의의 소정의 수의 프로모터와 함께 적용될 수 있기 때문이다.

[0246] 요약하면, 다양한 터미네이터를 이용하여 유기체에서 다양한 유전자의 발현을 조절하는 것은 관심 형질을 최적화하는 강력한 도구이다. 본 발명자에 의해 개발된 터미네이터 스와핑의 분자 도구는 하나 이상의 조건하에서 적어도 하나의 유전자좌의 발현을 변화시키는 것으로 입증된 터미네이터 서열의 래더를 사용한다. 그런 후에 이 래더는 고 처리량 게놈 기술을 사용하여 유기체에서 한 유전자 그룹에 체계적으로 적용된다. 이 유전자 그룹은 여러 가지 방법 중 하나를 기반으로 관심 형질에 영향을 주는 높은 가능성을 갖는 것으로 결정된다. 이들은 알려진 기능에 기초한 선택 또는 관심 형질에 대한 영향 또는 이전에 결정된 유익한 유전자 다양성에 기초한 알고리즘 선택을 포함할 수 있다.

[0247] 유전자에 연결된 터미네이터 서열을 함유하는 유기체의 생성된 HTP 유전자 디자인 미생물 라이브러리는 고 처리량 선택 모델에서의 성능에 대해 평가되고, 증가된 성능을 유도하는 프로모터-유전자 연관이 결정되고 정보는 데이터베이스에 저장된다. 유전자 교란의 집합(즉 소정의 유전자 y에 연결된 소정이 터미네이터 x)은 "터미네이터 스왑 라이브러리"를 형성하며, 이것은 미생물 조작 처리에 이용될 잠재적인 유전자 변형의 공급원으로 이용될 수 있다.시간이 지남에 따라, 더 큰 세트의 유전자 교란이 더 큰 다양성의 미생물 배경에 대해 실행되기 때문에, 각 라이브러리는 임의의 관심 배경에 대한 표적화된 변화를 더욱 정확하고 예측되게 디자인하는데 사용될 수 있는 실험적으로 확인된 데이터의 컬렉션으로서 더욱 강력해진다. 즉 일부 실시태양에서, 본 발명은 본 발명의 유전자 디자인 라이브러리의 임의의 것과 관련된 메타 데이터 내에 삽입된 이전의 실험 결과에 기초한 숙주 세포 속으로의 하나 이상의 유전자 변화를 도입하는 것을 교시한다.

[0248] 따라서, 특정 실시태양에서, 터미네이터 스와핑은 다음을 포함하는 다단계 공정이다:

[0249] 1. "래더"로서 작용하는 "x" 터미네이터의 세트를 선택하는 단계. 이상적으로 이런 터미네이터는 다수의 게놈 유전자좌 전체에서 매우 가변적인 발현을 유도하는 것으로 나타났지만, 유일한 요구조건은 일부 유전자 발현을 교란시키는 것이다.

[0250] 2. 표적화하기 위한 "n" 유전자의 세트를 선택하는 단계. 이 세트는 게놈의 모든 오픈 리딩 프레임(ORF) 또는 ORF의 서브세트일 수 있다. 서브세트는 기능과 관련된 ORF에 대한 주석을 사용하고, 이전에 입증된 유익한 교란(이전 프로모터 스왑 STOP 스왑 또는 SNP 스왑)에 대한 관계에 의해, 이전에 생성된 교란 사이의 상위성 상호작용에 기초한 알고리즘 선택에 의해, 표적에 대한 유익한 ORF에 관한 가설에 기초한 다른 선택 기준 또는 랜덤 선택을 통해 선택될 수 있다. 다른 실시태양에서, "n" 표적 유전자는 비-암호화 RNA를 포함하는 비-단백질 암호화 유전자를 포함할 수 있다.

[0251] 3. 다음과 같은 유전자 변형을 신속하게- 및 일부 실시태양에서, 병렬-수행하는 고 처리량 균주 조작 단계: 천연 터미네이터가 표적 유전자 n의 3' 말단에 존재하고 이의 서열이 알려질 때, 천연 터미네이터를 래더에 있는x 터미네이터의 각각으로 교체한다. 천연 터미네이터가 존재하지 않거나 이의 서열이 알려지지 않을 때, 유전자 정지 코돈 이후 래더에x터미네이터의 각각을 삽입한다.

[0252] 이러한 방식으로, 균주의 "라이브러리"(HTP 유전자 디자인 라이브러리라고도 불림)가 구성되며, 라이브러리의 각 구성원은 그렇지 않으면 동일한 유전자 상황에서, n 표적에 작동 가능하게 연결된x터미네이터의 한 예이다. 상기한 바와 같이, 터미네이터의 조합이 삽입되어 라이브러리가 구성되는 조합 가능성의 범위를 확장시킬 수 있다.

[0253] 4. 하나 이상의 메트릭에 대한 성능이 최적화되는 성능을 나타내는 상황에서 균주의 라이브러리의 고 처리량 선택 단계.

[0254] 이 기본 과정은 특히 다음과 같은 방법으로 균주 성능에 추가 개량을 제공하는데 확장될 수 있다: (1) 반복 과정에서 한 번에 하나씩 또는 단일 단계에서 다중 변화로서 단일 균주 배경 속에 다수의 유익한 교란을 통합하는 단계. 다중 교란은 정의된 변경의 특정 세트 또는 부분적으로 랜덤화된 조합 라이브러리의 변화일 수 있다. 예를 들어, 표적 세트가 한 경로의 모든 유전자인 경우, 이전 라이브러리 균주의 개량된 구성원 또는 구성원들 속으로 교란의 라이브러리의 순차적 재생은 어느 유전자가 임의의 소정의 반복에서 속도를 제한하는지와 관계없이

한 경로에서 각 유전자의 발현 수준을 최적화할 수 있다; (2) 각 교란의 상호작용에 기초하여 최적의 교란 세트를 예측하기 위해 그 데이터를 사용하는 알고리즘 속에 라이브러리의 개별 및 조합 생성으로부터 성능 데이터를 제공하는 단계; 및 (3) 위의 두 접근법의 조합을 실행하는 단계.

[0255] 이 접근법은 산업용 미생물에 의해 본 발명에서 예시되었지만, 원하는 형질이 유전자 돌연변이 집단에서 동정될 수 있는 임의의 유기체에 적용 가능하다. 예를 들어, 이것은 CHO 세포, 효모, 곤충 세포, 조류뿐만 아니라 식물과 같은 다세포 유기체의 성능을 개량시키는데 사용될 수 있다.

[0256] 일부 실시태양에서, 본 발명에 따른 터미네이터 스왑 라이브러리를 생성하는데 사용될 수 있는 터미네이터 서열 세트가 제공된다. 이러한 터미네이터 서열 세트는 표 3에 기술된 것들 및 예컨대 서열번호 70 내지 서열번호 80과 적어도 70%, 75%, 80%, 85%, 90%, 95%, 99% 또 그 이상의 동일성을 갖는 이의 임의의 기능적 변이체를 포함한다.

[0257] ***5. 트랜스포존 돌연변이유발 다양성 라이브러리: 트랜스포존 돌연변이유발 다양성 라이브러리의 유도를 위한 분자 도구***

[0258] 본 발명에 기술된 특정 도구는 미생물 균주에서 유전자의 기존의 다형체에 관한 것이지만, 미생물 균주의 성능을 향상시키는데 유용한 신규한 돌연변이를 생성하지는 않는다. 본 발명은 숙주 균주의 개량된 특징을 초래하는 것들에 대해 추가로 선별될 수 있는 돌연변이를 랜덤으로 생성하는 트랜스포존 돌연변이유발 시스템을 교시하고, 이는 결국 전체 숙주 균주 표현형에 유리한 효과를 유발한다(예를 들어, 수율 또는 생산성).

[0259] 예를 들어, 일부 실시태양에서, 본 발명은 숙주 세포 내에서 하나 이상의 유전자의 발현 프로파일 범위를 나타내는 숙주 세포 내에서 돌연변이를 생성 및 확인하는 방법을 교시한다. 이 과정에서 생성된 임의의 특정 돌연변이는 트랜스포존 돌연변이유발 다양성 라이브러리로서 함께 그룹화될 수 있으며, 이는 아래에서 보다 상세히 설명된다.

[0260] 이 공정을 통해 조작된 생성 미생물은 HTP 유전자 디자인 라이브러리를 형성한다.

[0261] HTP 유전자 디자인 라이브러리는 이러한 과정을 통해 형성된 실제 물리적 미생물 균주 컬렉션을 지칭할 수 있으며, 각 구성원 균주는 트랜스포존 돌연변이유발에 의해 생성된 주어진 돌연변이를 대표하며, 그렇지 않으면 동일한 유전적 배경에서 상기 라이브러리는 "트랜스포존 돌연변이유발 다양성 라이브러리"로 지칭한다.

[0262] 또한, HTP 유전자 디자인 라이브러리는 유전자 교란의 컬렉션을 지칭할 수 있다 - 이 경우 트랜스포존 돌연변이유발에 의해 생성된 주어진 돌연변이.

[0263] 또한, 트랜스포존 돌연변이유발에 의해 생성된 특정 돌연변이를 제외하고는 유전적으로 동일한 것으로 가정되는 미생물이 또한 제공된다. 이들 미생물은 적절히 선별되고 특성화될 수 있고 또 다른 HTP 유전자 디자인 라이브러리를 생성할 수 있다. HTP 유전자 디자인 라이브러리에서 미생물 균주의 특성은 관계형 데이터베이스, 객체 지향 데이터베이스 또는 고도로 분산된 NoSQL 데이터베이스를 포함하는 모든 데이터 저장 구조에 저장할 수 있는 정보와 데이터를 생성한다. 이 데이터/정보는 예를 들어 숙주 세포 성장에 대한 돌연변이의 영향 또는 숙주 세포에서 분자의 생성일 수 있다. 이 데이터/정보는 또한 둘 이상의 돌연변이로 인한 광범위한 조합 효과일 수 있다.

[0264] 트랜스포존 돌연변이유발에 의해 생성된 상기 언급된 돌연변이의 예는 개념이 다양한 발현 프로파일의 제시 및 임의의 주어진 수의 유전자에 미치는 영향에 기초하여 함께 그룹화된 임의의 주어진 수의 돌연변이로 적용될 수 있기 때문에 단지 예시적이다. 당업자는 또한 트랜스포존 돌연변이유발에 의해 생성된 돌연변이를 임의의 다른 돌연변이와 통합시키는 능력을 인식할 것이다. 따라서, 일부 실시태양에서, 본 발명은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 또는 그 이상의 돌연변이가 통합된 라이브러리를 교시한다.

[0265] 요약하면, 유기체에서 트랜스포존 돌연변이유발에 의해 생성된 다양한 돌연변이를 이용하는 것은 관심 특성을 최적화하기 위한 강력한 도구이다. 본 발명자들에 의해 개발된 트랜스포존 돌연변이유발 다양성 라이브러리의 분자 도구는 다양한 발현 프로파일을 갖는 돌연변이의 집합을 사용한다. 이 컬렉션은 고 처리량 게놈 공학을 사용하여 유기체에 체계적으로 적용된다. 이 돌연변이 그룹은 다수의 방법 중 어느 하나에 기초하여 관심 형질에 영향을 미칠 가능성이 높은 것으로 결정된다. 일부 실시태양에서, 라이브러리는 포화된 수의 돌연변이를 함유한다(예를 들어, 이론에서, 미생물 게놈의 각 유전자는 적어도 한 번 타격(hit)된다). 일부 실시태양에서, 트랜스포존 돌연변이유발 라이브러리에서 돌연변이의 게놈 위치는 결정되지 않으며, 따라서 라이브러리는 미생물의 게

놈에서 랜덤으로 분포된 돌연변이를 함유한다. 일부 실시태양에서, 트랜스포존 돌연변이유발 라이브러리에서의 돌연변이는 관련된 표현형에 기초하여 선택된다. 일부 실시태양에서, 트랜스포존 돌연변이유발 라이브러리에서의 돌연변이가 특징화되고 돌연변이의 게놈 위치가 결정되고, 돌연변이에 의해 파괴된 유전자가 확인된다. 이들은 알려진 기능을 기반으로 한 선택, 또는 관심 형질에 미치는 영향, 또는 이전에 결정된 유익한 유전자 다양성을 기반으로 하는 알고리즘 선택을 포함할 수 있다. 일부 실시태양에서, 돌연변이의 선택은 주어진 숙주에서 모든 유전자를 포함할 수 있다. 다른 실시태양에서, 돌연변이의 선택은 랜덤으로 선택된 주어진 숙주에서 모든 유전자의 서브세트일 수 있다. 다른 실시태양에서, 돌연변이의 선택은 예컨대 사카로폴리스포라 종에서의 스피노신과 같은 주어진 분자의 합성에 관여하는 모든 유전자의 서브세트일 수 있다.

[0266] 트랜스포존 돌연변이유발에 의해 생성된 돌연변이를 함유하는 유기체의 결과 HTP 유전자 디자인 미생물 균주 라이브러리는 고 처리량 선별 모델에서 성능에 대해 평가되고, 성능을 증가시키는 돌연변이가 결정되고 데이터베이스에 저장된 정보가 결정된다. 유전자 교란(즉, 돌연변이)의 컬렉션은 "트랜스포존 돌연변이유발 라이브러리"를 형성하며, 이는 미생물 공학 처리에 사용될 잠재적 유전자 변경의 소스로서 이용될 수 있다. 시간이 지남에 따라, 더 많은 숙주 세포 배경에 대해 더 많은 유전자 교란이 구현됨에 따라, 각 라이브러리는 임의의 관심 배경에 대해 표적으로 하는 변화를 보다 정확하고 예측 가능하게 디자인하는데 사용할 수 있는 실험적으로 확인된 데이터의 컬렉션으로 더욱 강력해진다.

[0267] 일부 실시태양에서, 본 발명의 트랜스포존 돌연변이유발 다양성 라이브러리는 유전자 표적의 최적 발현을 확인하기 위해 사용될 수 있다. 일부 실시태양에서, 목표는 대사 또는 유전자 경로에서 병목 현상을 감소시키기 위해 표적 유전자의 활성을 증가시키는 것일 수 있다. 다른 실시태양에서, 목표는 상기 표적 유전자의 발현이 요구되지 않을 때 숙주 세포에서 불필요한 에너지 소비를 피하기 위해 표적 유전자의 활성을 감소시키는 것일 수 있다.

[0268] 따라서, 특정 실시태양에서, 트랜스포존 돌연변이유발 다양성 라이브러리를 사용하는 방법은 다음을 포함하는 다단계 공정이다:

[0269] 1. 돌연변이유발을 위한 트랜스포존 시스템을 선택하고 주어진 미생물 균주에 시스템을 적용하여 트랜스포존에 의해 야기된 돌연변이를 생성한다. 이상적으로 시스템은 사카로폴리스포라 균주와 같은 선택된 미생물 균주의 게놈에 트랜스포존을 랜덤하게 통합시키는 것으로 나타났다. 이러한 통합은 어떤 식으로든 유전자 발현을 교란시킨다.

[0270] 2. 게놈에 트랜스포존이 통합된 균주를 신속하게 선택하기 위한 고 처리량 균주 공학. 이러한 방식으로, 균주의 "라이브러리"(HTP 유전자 디자인 라이브러리로도 지칭됨)가 구성되며, 여기서 라이브러리의 각 구성원은 트랜스포존 돌연변이를 포함하는 균주, 그렇지 않으면 동일한 유전적 맥락이다. 전술한 바와 같이, 돌연변이의 조합은 통합될 수 있으며, 라이브러리가 구축되는 조합 가능성의 범위를 확장시킨다.

[0271] 3. 하나 이상의 메트릭에 대한 그들의 성능이 최적화된 성능을 나타내는 맥락에서 균주 라이브러리의 고 처리량 선별.

[0272] 이 기본 공정은 특히 다음에 의해 균주 성능의 추가 개량을 제공하도록 확장될 수 있다: (1) 다수의 유익한 교란(돌연변이)을 반복적인 공정에서 한번에 하나씩 또는 단일 단계에서 다수의 변경으로 단일 균주 배경으로 통합. 다수의 교란(돌연변이)은 돌연변이에 의해 변형된 유전자 기능에 관계없이 특정 세트의 정의된 변화 또는 부분적으로 랜덤화된 조합의 변화 라이브러리일 수 있다; (2) 라이브러리의 개별 및 조합 생성으로부터 얻어진 성능 데이터를 각 교란의 상호작용에 기초하여 최적의 교란의 세트를 예측하는데 그 데이터를 사용하는 알고리즘으로 공급; 및 (3) 상기 두 가지 접근법의 조합을 구현.

[0273] 일부 실시태양에서, 트랜스포사제는 사카로폴리스포라 종에서 작동한다. 일부 실시태양에서, 트랜스포사제는 EZ-Tn5 트랜스포존 시스템으로부터 유래된다. 일부 실시태양에서, DNA 페이로드 서열은 상기 트랜스포사제에 의해 인식될 수 있는 모자이크 요소(ME)에 의해 플랭킹된다. 일부 실시태양에서, DNA 페이로드는 기능상실(LoF) 트랜스포존 또는 기능획득(GoF) 트랜스포존일 수 있다.

[0274] 일부 실시태양에서, DNA 페이로드는 선택 마커를 포함한다. 일부 실시태양에서, 본 발명의 트랜스포존 돌연변이유발 공정에 사용될 수 있는 선택 가능한 마커는 다음을 포함하나 이에 제한되지 않는다: 아프라마이신에 대한 저항성을 부여하는 aac(3)IV(서열번호 151), 젠타마이신에 대한 저항성을 부여하는 aacC1(서열번호 152), 네오마이신 B에 대한 저항성을 부여하는 acC8(서열번호 153), 스펙티노마이신/스트렙토마이신에 대한 저항성을 부여하는 aadA(서열번호 154), 블레오마이신에 대한 저항성을 부여하는 ble(서열번호 155), 클로로람페니콜에 대한

저항성을 부여하는 cat(서열번호 156), 에리스로마이신에 대한 저항성을 부여하는 ermE(서열번호 157), 하이그로마이신에 대한 저항성을 부여하는 hyg(서열번호 158), 및 카나마이신에 대한 저항성을 부여하는 neo(서열번호 159). 일부 실시태양에서, 선택 마커는 트랜스포존을 함유하는 사카로폴리스포라 세포를 선별하는데 사용된다.

[0275] 일부 실시태양에서, DNA 페이로드는 카운터 선택 마커를 포함한다. 일부 실시태양에서, 카운터 선택 마커는 선택 가능한 마커를 함유하는 DNA 페이로드의 루프-아웃을 용이하게 하기 위해 사용된다. 일부 실시태양에서, 본 발명의 트랜스포존 돌연변이유발 공정에 사용될 수 있는 카운터 선택 마커는 다음을 포함하나 이에 제한되지 않는다: 서열번호 160(amdSYM), 서열번호 161(tetA), 서열번호 162(lacY), 서열번호 163(sacB), 서열번호 164(pheS, S. 에리트라에아), 서열번호 165(pheS, 코리네박테리움).

[0276] 일부 실시태양에서, 본 발명의 방법은 이러한 게놈 모듈성의 영역에서 표적화된 게놈 편집을 수행할 수 있을 뿐만 아니라, 임의의 게놈 상황에서 게놈에 걸쳐 표적화된 게놈 편집을 가능하게 한다. 결과적으로, 본 발명의 표적화된 게놈 편집은 임의의 영역에서 S. 스피노사 게놈을 편집할 수 있으며, 모듈성을 갖는 영역에서의 편집에만 국한되지 않는다.

[0277] 일부 실시태양에서, GoF 트랜스포존은 GoF 요소를 포함한다. 일부 실시태양에서, GoF 트랜스포존은 프로모터 서열 및/또는 용해도 태그 서열(예를 들어, 서열번호 166)을 포함한다.

[0278] 일부 실시태양에서, 본 발명의 트랜스포존 돌연변이유발 라이브러리는 모든 유전자를 적어도 한 번 타격하는 것에 대해 95% 신뢰를 갖는다. 일부 실시태양에서, 이러한 라이브러리는 유기체에서 유전자 수의 대략 3배인 다수의 단리물을 선별함으로써 얻어진다. ~8000개의 주석이 달린 유전자를 포함하는 S. 스피노사의 경우, ~24,000 멤버의 돌연변이유발 라이브러리 크기가 게놈을 커버할 것으로 예상한다.

[0279] 일부 실시태양에서, 균주의 트랜스포존 돌연변이유발 라이브러리의 고 처리량 선별은 기준 균주와 비교하여 개량된 성능을 갖는 균주 컬렉션을 생성한다. 일부 실시태양에서, 트랜스포존 돌연변이유발로 인한 이러한 수집된 균주에서의 돌연변이는 이러한 수집된 균주의 성능을 향상시켜 강화된(enriched) 관심 표적을 갖는 새로운 균주를 생성하도록 통합된다. 일부 실시태양에서, 강화된 관심 표적을 갖는 이러한 균주는 추가의 지시된 균주 공학을 위해 본 발명의 다른 균주(예를 들어, SNP 스왑 또는 프로모터 스왑 라이브러리에서 개량된 성능을 갖는 균주)와 조합될 수 있다.

[0280] ***6. 리보솜 결합 부위(RBS) 다양성 라이브러리: RBS 미생물 균주 라이브러리의 유도를 위한 분자 도구***

[0281] 일부 실시태양에서, 본 발명은 전체 숙주 균주 표현형(예를 들어, 수율 또는 생산성)에 유리한 효과를 생성하도록 최적의 발현 특성을 갖는 리보솜 결합 부위(RBS)를 선택하는 방법을 교시한다.

[0282] 예를 들어, 일부 실시태양에서, 본 발명은 숙주 세포 내에서 하나 이상의 RBS를 확인 및/또는 숙주 세포 내에서 하나 이상의 RBS의 변이체를 생성하는 방법을 교시하고, 이는 다양한 발현 강도(예를 들어,하기 논의된 RBS 래더) 또는 우수한 조절 특성(예를 들어, 선택된 유전자에 대한 보다 엄격한 조절 제어)을 나타낸다. 이들 식별된 및/또는 생성된 RBS의 특정 조합은 RBS 래더로서 함께 그룹화될 수 있으며, 이는 아래에서 보다 상세히 설명된다.

[0283] 이어서, 일부 실시태양에서 문제의 RBS 래더는 주어진 관심 유전자와 관련된다. 따라서, RBS1 내지 RBS31(범위의 발현 강도, 서열번호 97 내지 서열번호 127을 나타내도록 식별되고 및/또는 생성된 31개의 RBS를 나타냄)을 갖고 RBS 래더를 미생물에서 관심 단일 유전자와 연관시키는 경우(즉, 주어진 표적 유전자에 작동 가능하게 연결된 주어진 RBS를 갖는 미생물을 유전자 조작하는 경우), 조작된 미생물이 표적 유전자와 관련된 특정 RBS(들)를 제외하고는 다른 동일한 유전적 배경을 갖는 것을 고려할 때, 각각의 조합의 노력으로 인한 각각의 조작된 균주를 특성화함으로써 31개의 RBS의 각 조합의 효과를 확인할 수 있다.

[0284] 이 공정을 통해 조작된 생성 미생물은 HTP 유전자 디자인 라이브러리를 형성한다.

[0285] HTP 유전자 디자인 라이브러리는이 공정을 통해 형성된 실제 물리적 미생물 균주 컬렉션을 지칭할 수 있으며, 각 구성원 균주는 다른 표적 유전자에 특정 표적 유전자에 작동 가능하게 연결된 주어진 RBS를 나타내며, 그렇지 않으면 동일한 유전적 배경에서 상기 라이브러리를 "RBS 라이브러리"라고 한다.

[0286] 또한, HTP 유전자 디자인 라이브러리는 유전자 교란의 컬렉션을 지칭할 수 있다 - 이 경우 주어진 RBS x는 주어진 유전자 y에 작동 가능하게 연결된다(및 또한 선택적으로 주어진 프로모터 z에 연결됨).

[0287] 또한, 미생물을 조작하기 위해 표 11의 RBS를 포함하는 동일한 RBS 래더를 이용할 수 있으며, 여기서 각각의

RBS는 상이한 유전자 표적에 작동 가능하게 연결된다. 이 과정의 결과는 관심 대상 유전자에 작동 가능하게 연결된 특정 RBS를 제외하고는 유전자적으로 다른 것으로 가정되는 미생물일 것이다. 이들 미생물은 적절히 선별되고 특성화될 수 있고 또 다른 HTP 유전자 디자인 라이브러리를 생성할 수 있다. HTP 유전자 디자인 라이브러리에서 미생물 균주의 특성은 관계형 데이터베이스, 객체 지향 데이터베이스 또는 고도로 분산된 NoSQL 데이터베이스를 포함하여 모든 데이터 저장 구조에 저장할 수 있는 정보와 데이터를 생성한다. 이 데이터/정보는 예를 들어 주어진 유전자 표적에 작동 가능하게 연결된 경우 주어진 RBS의 효과일 수 있다. 이 데이터/정보는 또한 본 발명의 둘 이상의 RBS를 주어진 유전자 표적에 작동 가능하게 연결함으로써 발생하는 보다 광범위한 조합 효과일 수있다.

[0288] 상기 언급된 RBS 및 표적 유전자의 예는 개념이 다양한 발현 강도 및 임의의 주어진 수의 표적 유전자의 제시에 기초하여 함께 그룹화된 임의의 주어진 수의 RBS와 함께 적용될 수 있기 때문에 단지 예시적이다. 당업자는 또한 임의의 유전자 표적 앞에서 2개 이상의 RBS를 작동 가능하게 연결시키는 능력을 인식할 것이다. 따라서, 일부 실시태양에서, 본 발명은 RBS 래더로부터의 1, 2, 3개 또는 그 이상의 RBS가 하나 이상의 유전자에 작동 가능하게 연결된 RBS 라이브러리를 교시한다.

[0289] 요약하면, 유기체에서 다양한 유전자의 발현을 유도하기 위해 다양한 RBS를 이용하는 것은 관심 특성을 최적화하는 강력한 도구이다. 본 발명자들에 의해 개발된 RBS 라이브러리의 분자 도구는 적어도 하나의 조건 하에서 적어도 하나의 유전자좌의 발현을 변화시키는 것으로 입증된 RBS 서열의 래더를 사용한다. 그런 다음이 래더는 고 처리량 게놈 공학을 사용하여 유기체의 유전자 그룹에 체계적으로 적용된다. 이 유전자 그룹은 다수의 방법 중 하나에 기초하여 관심 형질에 영향을 미칠 가능성이 높은 것으로 결정된다. 이들은 알려진 기능을 기반으로 한 선택, 또는 관심 형질에 미치는 영향, 또는 이전에 결정된 유익한 유전자 다양성을 기반으로 하는 알고리즘 선택을 포함할 수 있다. 일부 실시태양에서, 유전자의 선택은 주어진 숙주에서 모든 유전자를 포함할 수 있다. 다른 실시태양에서, 유전자의 선택은 랜덤으로 선택된 주어진 숙주에서 모든 유전자의 서브세트일 수 있다.

[0290] 이어서, 유전자에 연결된 RBS 서열을 함유하는 유기체의 결과 HTP 유전자 디자인 미생물 균주 라이브러리는 고 처리량 선별 모델에서 성능에 대해 평가되고, 성능을 증가시키는 RBS-유전자 연결이 결정되고 데이터베이스에 저장된 정보가 결정된다. 유전자 교란(즉, 주어진 RBS x는 주어진 유전자 y에 작동 가능하게 연결된다)의 컬렉션은 "RBS 다양성 라이브러리"를 형성하며, 이는 미생물 공학 처리에 사용될 잠재적 유전자 변경의 소스로서 이용될 수 있다. 시간이 지남에 따라, 더 많은 숙주 세포 배경에 대해 더 많은 유전자 교란이 구현됨에 따라, 각 라이브러리는 임의의 관심 배경에 대해 표적으로 하는 변화를 보다 정확하고 예측 가능하게 디자인하는데 사용할 수 있는 실험적으로 확인된 데이터의 모음으로 더욱 강력해진다.

[0291] 유기체에서 유전자의 전사 수준은 유기체 행동에 영향을 미치는 주요 제어 지점이다. 전사는 번역(단백질 발현)과 밀접하게 연관되어 있으며, 어느 단백질이 어떤 양으로 유기체 거동을 결정하는지에 따라 발현된다. 세포는 수천 가지의 다른 유형의 단백질을 발현하며, 이들 단백질은 수많은 복잡한 방식으로 작용하여 기능을 생성한다. 단백질 세트의 발현 수준을 체계적으로 변화시킴으로써, 기능은 복잡성으로 인해 예측하기 어려운 방식으로 변경될 수 있다. 일부 변경은 성능을 향상시킬 수 있으므로 성능을 평가하는 메커니즘과 결합하여 이 기술을 통해 기능이 개량된 유기체를 생성할 수 있다.

[0292] 소분자 합성 경로와 관련하여, 효소는 소분자 기질 및 생성물을 통해 선형 또는 분지형 사슬로 상호작용하며, 기질로 시작하여 관심 소분자로 끝난다. 이들 상호작용이 순차적으로 연결되어 있기 때문에, 이 시스템은 분산된 제어를 나타내며, 한 효소의 발현을 증가시키면 오직 다른 효소가 속도 제한이 될 때까지 경로 플럭스를 증가시킬 수 있다.

[0293] 대사 제어 분석(MCA)은 실험 데이터 및 제 1 원리로부터 어느 효소 또는 효소들이 속도 제한인지를 결정하는 방법이다. 그러나, MCA는 새로운 속도 제한 효소를 결정하기 위해 각각의 발현 수준 변화 후에 광범위한 실험을 필요로 하기 때문에 제한적이다. RBS 라이브러리는 경로에서 각각의 효소에 RBS 래더의 적용을 통해 제한 효소가 발견되고, 후속 라운드에서 동일한 일이 속도 제한이 되는 새로운 효소를 찾기 위해 수행될 수 있기 때문에 이러한 문맥에서 유리하다. 또한, 기능에 대한 판독이 관심 소분자의 보다 나은 생산이기 때문에, 어느 효소가 제한되는지를 결정하는 실험은 생산을 증가시키는 개발과 동일하므로 개발 시간이 단축된다. 일부 실시태양에서, 본 발명은 다중-단위 효소의 개별 서브 유닛을 코딩하는 유전자에 RBS 라이브러리의 적용을 교시한다. 또 다른 실시태양에서, 본 발명은 개별 효소 또는 전체 생합성 경로를 조절하는 유전자에 RBS 라이브러리 기술을 적용하는 방법을 교시한다.

[0294] 일부 실시태양에서, 본 발명의 RBS 라이브러리는 선택된 유전자 표적의 최적 발현을 확인하는데 사용될 수

있다. 일부 실시태양에서, RBS 라이브러리의 목적은 대사 또는 유전자 경로에서 병목 현상을 감소시키기 위해 표적 유전자의 발현을 증가시키는 것일 수 있다. 다른 실시태양에서, RBS 라이브러리의 목적은 상기 표적 유전자의 발현이 요구되지 않을 때 숙주 세포에서 불필요한 에너지 소비를 피하기 위해 표적 유전자의 발현을 감소시키는 것일 수 있다.

[0295] 전사, 수송 또는 신호 전달과 같은 다른 세포 시스템과 관련하여, 다양한 합리적인 방법을 사용하여 어떤 단백질이 발현 변화의 표적인지, 그리고 그 변화가 무엇인지를 우선적으로 시도하고 찾아낼 수 있다. 이 합리적 방법은 성능을 향상시키는 것을 찾기 위해 테스트해야 하는 교란의 수를 줄이지만 상당한 비용이 든다. 유전자 결실 연구는 존재가 특정 기능에 결정적인 단백질을 확인한 후, 중요한 유전자를 과발현할 수 있다. 단백질 상호작용의 복잡성으로 인해, 이는 종종 성능 향상에 비효율적이다. 제 1 원리로부터, 전사 또는 신호 전달 거동을 세포의 단백질 수준의 함수로서 기술하려는 다양한 유형의 모델이 개발되었다. 이러한 모델은 종종 발현 변경으로 인해 기능이 달라지거나 향상될 수 있는 표적을 제안한다. 이러한 모델의 기초가 되는 가정은 단순하고 파라미터를 측정하기가 어렵기 때문에 모델이 아닌 유기체의 경우 예측이 틀린 경우가 많다. 유전자 결실과 모델링 모두에서, 특정 유전자에 영향을 미치는 방법을 결정하는데 필요한 실험은 성능을 향상시키는 변화를 만드는 후속 작업과 다르다. 특정 교란의 중요성을 강조하는 구축된 균주도 이미 개량된 균주이기 때문에 RBS 라이브러리 방법은 이러한 과제를 회피한다.

[0296] 따라서, 특정 실시태양에서, RBS 라이브러리를 사용하는 방법은 다음을 포함하는 다단계 공정이다:

[0297] 1. "래더"로서 작용할 "x" RBS 세트를 선택. 이상적으로, 이들 RBS는 다수의 게놈 유전자좌에 걸쳐 매우 가변적인 발현을 야기하는 것으로 나타났지만, 유일한 필요조건은 이들이 어떤 방식으로든 유전자 발현을 교란시키는 것이다.

[0298] 2. 표적화할 "n"유전자 세트를 선택. 이 세트는 게놈의 모든 오픈 리딩 프레임(ORF)이거나 ORF의 서브세트일 수 있다. 서브세트는 기능과 관련된 ORF에 대한 주석을 사용하여, 이전에 입증된 유익한 교란(이전 RBS 컬렉션 또는 이전 SNP 스왑)과 관련하여, 이전에 생성된 교란 사이의 상위성 상호작용을 기반으로 하는 알고리즘 선택에 의해, 대상에 대한 유익한 ORF에 관한 가설을 기반으로 하는 다른 선택 기준, 또는 랜덤 선택에 의해 선택할 수 있다. 다른 실시태양에서, "n" 표적화된 유전자는 비-코딩 RNA를 포함하는 비-단백질 코딩 유전자를 포함할 수 있다.

[0299] 3. 신속하고, 일부 실시태양에서, 다음 유전자 변형을 병렬로 수행하기 위한 고 처리량 균주 공학: 천연 RBS가 표적 유전자 n 앞에 존재하고 이의 서열이 공지된 경우, 천연 RBS를 래더에 있는 x RBS의 각각으로 대체한다. 천연 RBS가 존재하지 않거나 이의 서열이 알려져 있지 않은 경우, x RBS의 각각을 유전자 n 앞의 래더에 삽입한다. 이러한 방식으로, 균주의 "라이브러리"(HTP 유전자 디자인 라이브러리로도 지칭됨)가 구축되며, 여기서 라이브러리의 각 구성원은 달리 동일한 유전적 맥락에서 n 표적에 작동 가능하게 연결된 x RBS의 예이다. 전술한 바와 같이, RBS의 조합이 삽입될 수 있으며, 라이브러리가 구축되는 조합 가능성의 범위가 확장된다.

[0300] 4. 하나 이상의 매트릭스에 대한 이들의 성능이 최적화되는 성능을 나타내는 상황에서 균주 라이브러리의 고 처리량 선별.

[0301] 이 기본 공정은 특히 다음에 의해 균주 성능의 추가 개량을 제공하도록 확장될 수 있다: (1) 다수의 유익한 교란을 반복적인 공정에서 한번에 하나씩 또는 단일 단계에서 다수의 변경으로 단일 균주 배경으로 통합. 다수의 교란은 특정 세트의 정의된 변화 또는 부분적으로 랜덤화된 조합의 변화 라이브러리일 수 있다. 예를 들어, 표적의 세트가 경로의 모든 유전자인 경우, 이전 균주 라이브러리의 개선된 구성원 또는 구성원으로의 교란 라이브러리의 순차적 재생은 주어진 반복에서 어느 유전자가 속도를 제한하는지에 관계없이 경로에서 각 유전자의 발현 수준을 최적화할 수 있다; (2) 라이브러리의 개별 및 조합 생성으로부터 얻어진 성능 데이터를 각 교란의 상호작용에 기초하여 최적의 교란의 세트를 예측하는데 그 데이터를 사용하는 알고리즘으로 공급; 및 (3) 상기 두 가지 접근법의 조합을 구현.

[0302] 접근법은 본 발명에서 산업용 미생물로 예시되지만, 유전적 돌연변이체 집단에서 원하는 형질이 식별될 수 있는 임의의 유기체에 적용 가능하다. 예를 들어, 이는 식물과 같은 다세포 유기체뿐만 아니라 CHO 세포, 효모, 곤충 세포, 조류의 성능을 향상시키는데 사용될 수 있다.

[0303] 일부 실시태양에서, 본 발명의 RBS 라이브러리는 유전자 다양성의 공급원으로서 사용될 수 있다. 일부 실시태양에서, 사카로폴리스포라 균주에 도입될 때 본 발명의 RBS 래더는 균주의 성능을 개량시킨다. 이러한 개량된 균주는 본 발명의 추가적인 유전자 다양성을 갖는 다른 균주(예를 들어, SNP 스왑 또는 프로모터 스왑 라이브러리

에서 개량된 성능을 갖는 균주)와 추가로 통합되어, 강화된 관심 표적이 있는 새로운 균주를 생성할 수 있다. 일부 실시태양에서, 강화된 관심 표적이 있는 이러한 균주는 추가의 지시된 균주 공학에 사용될 수 있다.

[0304] ***7. 항-대사산물 선택/발효 생성물 저항성 라이브러리: 다형성 미생물 균주 라이브러리의 유도를 위한 분자 도구***

[0305] 미생물에 의한 원하는 화합물의 생산을 향상시키기 위해, 최종 생성물 억제 문제를 극복하는 것이 종종 필요하다. 미생물은 발효 과정의 일부로 다양한 화합물을 생산하다. 때때로 이러한 화합물의 축적은 미생물의 성장과 생리를 심각하게 억제한다. 발효를 개량하고 미생물이 원하는 대사산물을 합성할 수 있는 시간을 연장하기 위해서는 a) 최종 생성물의 잠재적 독성, 및 b) 원하는 최종 생성물의 형성에 필요한 분자 경로의 피드백 억제를 극복해야 한다.

[0306] (a) 일부 실시태양에서, 본 발명은 숙주 세포 내에서 돌연변이를 생성하고 확인하는 방법을 교시하며, 이는 숙주 세포에서 하나 이상의 유전자의 발현 프로파일의 범위, 특히 숙주 세포에서의 대사산물 또는 발효 생성물에 대한 개량된 저항성을 야기하는 돌연변이를 나타내며, 따라서 숙주 세포의 성능을 개량시킨다. 이 과정에서 확인된 임의의 특정 돌연변이는 항-대사산물 선택/발효 생성물 저항성 라이브러리로서 함께 그룹화될 수 있으며, 이는 아래에서 보다 상세히 설명된다.

[0307] 이 공정을 통해 조작된 생성 미생물은 HTP 유전자 디자인 라이브러리를 형성한다.

[0308] HTP 유전자 디자인 라이브러리는 이 과정을 통해 형성된 실제 물리적 미생물 균주 컬렉션을 지칭할 수 있으며, 각 구성원 균주는 공정에서 식별된 주어진 돌연변이를 대표하며, 그렇지 않으면 동일한 유전적 배경에서 상기 라이브러리는 "항-대사산물 선택/발효 생성물 저항성 라이브러리"로 지칭한다.

[0309] 또한, HTP 유전자 디자인 라이브러리는 유전자 교란의 컬렉션을 지칭할 수 있다 - 이 경우 본 발명에 기술된 과정에 의해 생성된 주어진 돌연변이.

[0310] 또한, 주어진 대사산물 또는 발효 생성물에 대한 저항성을 유발하는 특정 돌연변이를 제외하고는 달리 유전적으로 동일한 것으로 추정되는 미생물이 또한 제공된다. 이들 미생물은 적절히 선별되고 특성화될 수 있고 또 다른 HTP 유전자 디자인 라이브러리를 생성할 수 있다. HTP 유전자 디자인 라이브러리에서 미생물 균주의 특성은 관계형 데이터베이스, 객체 지향 데이터베이스 또는 고도로 분산된 NoSQL 데이터베이스를 포함하는 모든 데이터 저장 구조에 저장할 수 있는 정보와 데이터를 생성한다. 이 데이터/정보는 예를 들어 숙주 세포 성장에 대한 돌연변이의 영향 또는 숙주 세포에서 분자의 생성일 수 있다. 이 데이터/정보는 또한 둘 이상의 돌연변이로 인한 광범위한 조합 효과일 수 있다.

[0311] 공정에 의해 생성된 상기 돌연변이의 예는 개념이 다양한 발현 프로파일의 제시 및 임의의 주어진 수의 유전자에 미치는 영향에 기초하여 함께 그룹화된 임의의 주어진 수의 돌연변이로 적용될 수 있기 때문에 단지 예시적이다. 당업자는 본 발명에 기술된 방법에 의해 생성된 돌연변이를 임의의 다른 돌연변이와 통합시키는 능력을 인식할 것이다. 따라서, 일부 실시태양에서, 본 발명은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 또는 그 이상의 돌연변이가 통합된 라이브러리를 교시한다.

[0312] 요약하면, 유기체에서 주어진 대사산물 또는 발효 생성물에 저항성을 유발하는 다양한 돌연변이를 이용하는 것이 관심의 특성을 최적화하는 강력한 도구이다. 분자 도구는 주어진 대사산물 또는 발효 생성물에 대한 저항성 돌연변이 컬렉션을 사용한다. 일부 실시태양에서, 이러한 돌연변이는 균주에서의 개량된 성능, 예컨대 스피노신과 같은 하나 이상의 주어진 분자의 수율 증가 또는 생성을 초래한다. 이 컬렉션은 고 처리량 게놈 공학 사용하여 유기체에 체계적으로 적용된다. 이 돌연변이 그룹은 다수의 방법 중 어느 하나에 기초하여 관심 형질에 영향을 미칠 가능성이 높은 것으로 결정된다. 이들은 알려진 기능을 기반으로 한 선택, 또는 관심 형질에 미치는 영향, 또는 이전에 결정된 유익한 유전적 다양성에 기초한 알고리즘 선택을 포함할 수 있다. 일부 실시태양에서, 돌연변이의 선택은 주어진 숙주에서 모든 유전자를 포함할 수 있다. 다른 실시태양에서, 돌연변이의 선택은 랜덤으로 선택된 주어진 숙주에서 모든 유전자의 서브세트일 수 있다. 다른 실시태양에서, 돌연변이의 선택은 사카로폴리스포라 종의 스피노신과 같은 주어진 분자의 합성에 관여하는 모든 유전자의 서브세트일 수 있다.

[0313] 주어진 대사산물 또는 발효 생성물에 저항성을 유발하는 돌연변이를 함유하는 유기체의 결과 HTP 유전자 디자인 미생물 균주 라이브러리는 고 처리량 선별 모델에서 성능에 대해 평가되고, 성능 증가를 초래하는 돌연변이가 결정되고 데이터베이스에 저장된 정보가 결정된다. 유전자 교란(즉, 돌연변이)의 컬렉션은 "항-대사산물 선택/발효 생성물 저항성 라이브러리"를 형성하며, 이는 미생물 공학 처리에 이용될 수 있는 잠재적 유전자 변경의 소스로서 이용될 수 있다. 시간이 지남에 따라, 더 많은 숙주 세포 배경에 대해 더 많은 유전자 교란이 구현됨

에 따라, 각 라이브러리는 임의의 관심 배경에 대해 표적으로 하는 변화를 보다 정확하고 예측 가능하게 디자인하는데 사용할 수 있는 실험적으로 확인된 데이터의 모음으로 더욱 강력해진다.

[0314] 일부 실시태양에서, 본 발명의 항-대사산물 선택/발효 생성물 저항성 다양성 라이브러리는 유전자 표적의 최적 발현을 확인하기 위해 사용될 수 있다. 일부 실시태양에서, 목표는 대사 또는 유전자 경로에서 병목 현상을 감소시키기 위해 표적 유전자의 활성을 증가시키는 것일 수 있다. 다른 실시태양에서, 목표는 상기 표적 유전자의 발현이 요구되지 않을 때 숙주 세포에서 불필요한 에너지 소비를 피하기 위해 표적 유전자의 활성을 감소시키는 것일 수 있다.

[0315] 따라서, 특정 실시태양에서, 항-대사산물 선택/발효 생성물 저항성 라이브러리를 적용하는 방법은 다음을 포함하는 다단계 공정이다:

[0316] 1. 숙주 균주에서 하나 이상의 주어진 대사산물 또는 발효 생성물에 저항성이 있는 균주를 신속하게 선택하기 위한 고 처리량 균주 공학. 이상적으로, 시스템은 다형체가 주어진 대사산물 또는 발효 생성물의 합성과 관련이 있는지 여부에 관계없이 모든 유형의 다형체를 갖는 균주를 확인하는 것으로 나타났다.

[0317] 2. 실제로 개량된 성능(예를 들어, 주어진 대사산물 또는 발효 생성물의 증가된 수율 또는 생산)을 갖는 균주를 신속하게 선택하기 위한 고 처리량 균주 공학. 이러한 방식으로, 균주의 "라이브러리"(HTP 유전자 디자인 라이브러리로도 지칭됨)가 구축되며, 여기서 라이브러리의 각 구성원은 하나 이상의 유리한 다형체, 그렇지 않으면 동일한 유전적 맥락을 포함하는 균주이다. 전술한 바와 같이, 다형체의 조합은 통합될 수 있으며, 라이브러리가 구축되는 조합 가능성의 범위를 확장시킨다.

[0318] 3. 하나 이상의 메트릭에 대한 그들의 성능이 최적화되는 성능을 나타내는 맥락에서 균주 라이브러리의 고 처리량 선별.

[0319] 일부 실시태양에서, 방법은 또한 세포 성장 억제, 대사산물/발효 생성물의 적절한 농도를 유발하는 바람직한 대사산물/발효 생성물을 선택하는 것과 같이, 상기 기재된 바와 같은 초기 선택 단계 1에 대한 전략을 결정하는 단계를 포함한다.

[0320] 일부 실시태양에서, 본 발명의 항-대사산물 선택/발효 생성물 저항성 라이브러리는 유전자 다양성의 소스로서 사용될 수 있다. 일부 실시태양에서, 본 발명의 방법에 의해 확인된 대사산물 또는 발효 생성물에 대한 개량된 저항성을 야기하는 돌연변이는 균주의 성능을 개량시킨다. 이러한 개량된 균주는 본 발명의 추가적인 유전자 다양성을 갖는 다른 균주(예를 들어, SNP 스왑 또는 프로모터 스왑 라이브러리 또는 트랜스포존 돌연변이유발 라이브러리에서 개량된 성능을 갖는 균주)와 추가로 통합되어, 강화된 관심 표적이 있는 새로운 균주를 생성할 수 있다. 일부 실시태양에서, 강화된 관심 표적이 있는 이러한 균주는 추가의 지시된 균주 공학에 사용될 수 있다.

[0321] ***8. 순서 최적화: 최적화된 서열 미생물 균주 라이브러리 유도를 위한 분자 도구***

[0322] 한 실시태양에서, 제공된 본 발명의 방법은 숙주 유기체에 의해 발현된 하나 이상의 유전자를 최적화하는 코돈을 포함한다. 다양한 숙주에서 발현을 향상시키기 위해 코돈을 최적화하는 방법은 당해 분야에 공지되어 있고, 문헌에 기술된다(그 전체가 본 발명에 참조로 포함된 미국 특허출원 공개공보 제2007/0292918호 참조). 특정 원핵 또는 진핵 숙주(Murray et al.(1989) Nucl. Acids Res.17:477-508 참조)에 의해 선호되는 코돈을 함유하는 최적화된 암호화 서열은, 예를 들어, 번역 속도를 증가시키거나 또는 최적화되지 않은 서열로부터 생산된 전사체와 비교할 때, 더 긴 반감기와 같은 바람직한 특성을 갖는 재조합 RNA 전사체를 생성하도록 제조된다.

[0323] 단백질 발현은 전사, mRNA 프로세싱 및 안정성 및 번역 개시에 영향을 미치는 인자를 포함하는 다수의 인자에 의해 지배된다. 따라서 최적화는 임의의 특정 유전자의 많은 서열 특징의 임의의 것을 처리할 수 있다. 특정 예로서, 드문 코돈 유도 번역 일시 중지는 감소된 단백질 발현을 초래할 수 있다. 드문 코돈 유도 번역 일시 중지는 숙주 유기체에서 거의 사용되지 않는 관심 폴리뉴클레오티드에서 코돈의 존재를 포함하며 이용가능한 tRNA 풀에서의 희소성 때문에 단백질 번역에 부정적인 영향을 미칠 수 있다.

[0324] 대안적인 번역 개시는 또한 감소된 이종 단백질 발현을 초래할 수 있다. 대안적인 번역 개시는 의도하지 않게 리보솜 결합 부위(RBS)로서 기능할 수 있는 모티프를 함유하는 합성 폴리뉴클레오티드 서열을 포함할 수 있다. 이런 위치는 유전자 내부 위치로부터 잘린 단백질의 번역을 시작할 수 있다. 정제 과정에서 제거하기 어려울 수 있는 절단된 단백질을 생산할 가능성을 줄이는 한 가지 방법은 최적화된 폴리뉴클레오티드 서열로부터 추정된 내부 RBS 서열을 제거하는 것을 포함한다.

[0325] 반복체-유래된 폴리머라제의 슬리파지(slippage)는 감소된 이종 단백질 발현을 초래할 수 있다. 반복체-유래된

폴리머라제의 슬리파지는 프레임변이 돌연변이를 초래할 수 있는 DNA 폴리머라제의 슬리파지 또는 스터터링(stuttering)을 유발하는 것으로 밝혀진 뉴클레오티드 서열 반복을 필요로 한다. 이러한 반복체는 또한 RNA 폴리머라제의 슬리파지를 유발할 수 있다. 높은 G+C 함량 편향을 갖는 유기체에서, G 또는 C 뉴클레오티드 반복체로 구성된 더 높은 등급의 반복체가 존재할 수 있다. 따라서, RNA 폴리머라제의 슬리파지를 유도하는 가능성을 줄이는 한 가지 방법은 G 또는 C 뉴클레오티드의 연장된 반복체를 변경하는 것을 포함한다.

[0326] 2차 구조를 간섭하면 또한 감소된 이종 단백질 발현을 초래할 수 있다. 2차 구조는 RBS 서열 또는 개시 코돈을 격리할 수 있고 단백질 발현의 감소와 상관관계가있다. 스템루프 구조는 또한 전사 일시 중지 및 약화에 관여할 수 있다. 최적화된 폴리뉴클레오티드 서열은 RBS 내에 최소 2차 구조 및 뉴클레오티드 서열의 유전자 암호화 영역을 함유하여 개량된 전사 및 번역을 가능하게 한다.

[0327] 예를 들어, 최적화 공정은 숙주에 의해 발현되는 원하는 아미노산 서열을 확인함으로써 시작될 수있다. 아미노산 서열로부터 후보 폴리뉴클레오티드 또는 DNA 서열이 디자인될 수 있다. 합성 DNA 서열의 디자인 동안, 코돈 사용의 빈도는 숙주 발현 유기체의 코돈 사용과 비교될 수 있고 드문 숙주 코돈은 합성 서열로부터 제거될 수있다. 또한, 합성 후보 DNA 서열은 바람직하지 않은 효소 제한 부위를 제거하고 원하는 신호 서열, 링커 또는 비번역 영역을 추가 또는 제거하기 위해 변형될 수 있다. 합성 DNA 서열은 G/C 반복체 및 스템-루프 구조와 같이 번역 과정을 방해할 수 있는 2차 구조의 존재에 대해 분석될 수 있다.

[0328] ***9. 상위성 매핑 - 유익한 유전자 통합을 가능하게 하는 예측 분석 도구***

[0329] 일부 실시태양에서, 본 발명은 유익한 유전자 변형을 예측하고 숙주 세포 내로 결합시키는 상위성 매핑 방법을 교시한다. 유전자 변형은 상기한 HTP 분자 도구 세트(예를 들어, 프로모터 스왑, SNP 스왑, 개시/정지 코돈 교환, 서열 최적화)의 임의의 것에 의해 일어날 수 있고 이러한 유전자 변형의 효과는 유래된 HTP 유전자 디자인 미생물 균주 라이브러리의 특성화로부터 알 수 있다. 따라서, 본 발명에서 사용된 용어 상위성 매핑은 숙주 성능의 증가를 초래할 수 있는유전자 변형의 조합(예를 들어 유익한 SNP 또는 유익한 프로모터/표적 유전자 결합)의 조합을 확인하는 방법을 포함한다.

[0330] 실시태양에서, 본 발명의 상위성 맵핑 방법은 2개의 상이한 작용기로부터의 유익한 돌연변이의 조합이 동일한 작용기로부터의 돌연변이의 조합과 비교하여, 숙주 성능을 개량시키가 더 쉽다는 아이디어에 기초한다. 예를 들어, Costanzo, The Genetic Landscape of a Cell, Science, Vol. 327, Issue 5964, Jan. 22, 2010, pp. 425-431 참조(전문이 본 발명에 참조로 포함).

[0331] 동일한 작용기로부터의 돌연변이는 동일한 메카니즘에 의해 작용할 가능성이 더 높으며, 따라서 전체 숙주 성능에 대해 음성 또는 중성 상위성을 나타내기가 더 쉽다. 대조적으로, 상이한 작용기로부터의 돌연변이는 독립적인 메카니즘에 의해 작동하기 쉽고, 이로 인해 개량된 숙주 성능 및 일부 경우에 상승 효과를 유도할 수 있다.

[0332] 따라서, 일부 실시태양에서, 본 발명은 상이한 작용기에 속하는 것으로 예측되는 SNP를 확인하기 위해 SNP 돌연변이를 분석하는 방법을 교시한다. 일부 실시태양에서, SNP 작용기 유사성은 돌연변이 상호작용 프로파일의 코사인 유사성(상관 계수와 유사, 도 54a 참조)을 계산함으로써 결정된다. 본 발명은 또한 돌연변이 유사성 행렬(도 53 참조) 또는 덴드로그램(도 54a 참조)을 통한 SNP의 비교를 예시한다.

[0333] 따라서, 상위성 매핑 절차는 하나 이상의 유전자 배경에 상기 돌연변이의 효율적이고 효과적인 통합을 위해 하나 이상의 유전자 배경에 적용된 다양성의 유전 돌연변이를 그룹화 및/또는 순위화하는 방법을 제공한다.

[0334] 양태에서, 통합은 표적 생체분자의 생산을 위해 최적화된 신규 균주를 생성하는 목적으로 수행된다. 교시된 상위성 매핑 절차를 통해, 돌연변이의 기능적 그룹화를 확인하는 것이 가능하며 이런 기능적 그룹화는 바람직하지 않은 상위성 효과를 최소화하는 통합 전략을 가능하게 한다.

[0335] 상기한 바와 같이, 산업용 발효에 사용하기 위한 미생물의 최적화는 중요하고 경제, 사회 및 자연계에 대한 광범위하게 관련된 어려운 문제이다. 전통적으로, 미생물 조작은 랜덤 돌연변이유발의 느리고 불확실한 과정을 통해 수행되었다. 이런 접근법은 인위적으로 부과된 선택 압력에 적응할 수 있는 세포의 자연적 진화 능력을 활용한다. 이런 접근법은 유익한 돌연변이의 희귀성, 근본적인 적합함 환경의 견고함에 의해 제한되며 보다 일반적으로 세포 및 분자 생물학의 최신기술을 충분히 활용하지 못한다.

[0336] 현대의 접근법은 기계적 수준에서 세포 기능의 새로운 이해 및 특정 표현형 말단에 대해 표적화된 유전자 조작을 수행하는 새로운 분자 생물학 도구를 활용한다. 실제로, 이런 합리적인 접근법은 생물학의 근본적인 복잡성에 의해 혼란스럽게 된다. 인과관계 매커니즘은, 특히 관찰된 유익한 효과가 있는 두 개 이상의 변화를 결합하

려고 시도할 때 잘 이해되지 않는다. 때로는 순 긍정적인 결과가 예상보다 낮을 수도 있고 경우에 따라 예상보다 높을 수도 있지만, 유전자 변화의 이런 통합은 긍정적인 결과(원하는 표현형 활동의 증가로 측정됨)를 가져온다. 다른 경우, 이런 조합은 순 중성 효과 또는 순 부정적인 효과를 일으킨다. 이 현상은 상위성으로 불리고 미생물 공학(및 일반적으로 유전자 공학)에 대한 근본적인 도전 중 하나이다.

[0337] 상기한 바와 같이, 본 HTP 게놈 공학 플랫폼은 전통적인 미생물 공학 접근법과 관련된 많은 문제점을 해결한다. 현재의 HTP 플랫폼은 자동화 기술을 사용하여 한 번에 수백 또는 수천 개의 유전자 변이를 실행한다. 특정 양태에서, 상기한 합리적인 접근법과는 달리, 개시된 HTP 플랫폼은 전문이 본 발명에 참조로 포함된 Microbial Strain Design System and Methods For Improved Large-Scale Production Of Engineered Nucleotide Sequences 이란 제목의 미국 특허 출원 제15/140,296호에 개시된 바와 같이 수천 개의 돌연변이체의 병렬 건설을 가능하게 하여 관련 게놈 공간의 큰 서브세트를 보다 효과적으로 탐사할 수 있게 한다. 공학적 염기 서열의 개량된 대규모 생산을위한 시스템 및 방법으로서, 본 명세서에서 그 전체가 참고로 인용된다. "모든 것"을 시도함으로써, 현재의 HTP 플랫폼은 제한된 생물학적 이해에 의해 유래된 어려움을 회피한다.

[0338] 그러나, 동시에, 본 HTP 플랫폼은 게놈 공간의 조합 폭발성 크기 및 유전자 상호작용의 복잡성을 고려하여 생성된 데이터 세트를 해석하는 계산 기술의 효과에 의해 근본적으로 제한되는 문제에 직면한다. 원하는 결과를 산출하는 조합의 비 랜덤 선택을 최대화하는 방식으로 광대한 조합 공간의 서브세트를 탐구하는 기술이 필요하다.

[0339] 다소 유사한 HTP 접근법이 효소 최적화의 경우에 효과적이라는 것이 증명되었다. 이 틈새 문제에서, 관심 게놈 서열(약 1000 염기)은 일부 복잡한 물리적 구성을 가진 단백질 사슬을 암호화한다. 정확한 구성은 구성 원자 구성요소 사이의 집단적 전자기 상호작용에 의해 결정된다. 짧은 게놈 서열과 물리적으로 구속된 접힘 문제의 조합은 탐욕 최적화 전략(greedy optimization strategies)에 특히 적합하다. 즉, 모든 잔기에서 서열을 개별적으로 돌연변이시키고, 생성된 돌연변이를 뒤섞어 서열 활성 반응 모델링과 양립가능한 해상도로 로컬 서열 공간을 효과적으로 샘플링하는 것이 가능하다.

[0340] 그러나, 생체 분자에 대한 완전한 게놈 최적화를 위해, 이런 잔기-중심 접근법은 몇 가지 중요한 이유 때문에 불충분하다. 첫째, 생체 분자에 대한 게놈 최적화와 관련된 관련 서열 공간의 기하 급수적인 증가 때문이다. 둘째, 생체분자 합성에서 조절, 발현 및 대사 상호작용의 추가된 복잡성 때문이다. 본 발명자들은 교시 된 상위성 매핑 절차를 통해 이러한 문제점을 해결하였다.

[0341] 상기 돌연변이를 하나 이상의 유전자 배경 속에 더 효율적이고 효과적으로 통합하기 위해서 돌연변이의 집합 사이의 상위성 상호작용을 모델링하는 교시된 방법은 획기적이며 당업계에서 매우 필요하다.

[0342] 상위성 매핑 절차를 기술할 때, 용어 "더 효율적" 및 "더 효과적인"은 특정 표현형 목표와 관련하여 통합 균주 사이의 바람직하지 않은 상위성 상호작용의 회피를 의미한다.

[0343] 상기 과정이 일반적으로 상기에 상세하게 설명되었으므로, 더 구체적인 워크 플로우 예가 이제 설명될 것이다.

[0344] 먼저, M 돌연변이 및 하나 이상의 유전자 배경(예를 들어, 부모 박테리아 균주)의 라이브러리로 시작한다. 라이브러리의 선택이나 유전자 배경의 선택은 여기에 설명된 방법에 특이적이지 않다. 그러나 특정 구현예에서, 돌연변이의 라이브러리는 SNP 스왑 라이브러리, 프로모터 스왑 라이브러리 또는 본 발명에 설명된 임의의 다른 돌연변이 라이브러리를 독점적으로 또는 조합하여 포함할 수 있다.

[0345] 한 구현예에서, 단일 유전자 배경만이 제공된다. 이 경우, 구별된 유전자 배경(미생물 돌연변이)의 집합이 이 단일 배경에서 먼저 생성될 것이다. 이것은 소정의 배경에 돌연변이의 기본 라이브러리(또는 이의 일부 서브세트)를 적용, 예를 들어, 특정 SNP의 HTP 유전 디자인 라이브러리 또는 특정 프로모터의 HTP 유전 디자인 라이브러리의 특정 유전자 배경에 대해 적용함으로써 달성되어 본 발명에 포함된 소정의 HTP 유전자 디자인 라이브러리로부터의 특정 유전자 변형을 제외하고는 동일한 유전자 배경을 가진 미생물 돌연변이 개체군(아마도 100 또는 1,000)을 생성할 수 있다. 아래에서 상세히 설명하는 바와 같이, 이 실시태양은 조합 라이브러리 또는 쌍을 이루는 라이브러리를 유도할 수 있다.

[0346] 또 다른 구현예에서, 구별된 알려진 유전자 배경의 집합이 간단히 주어질 수있다. 아래에서 상세히 설명되는 바와 같이, 이 실시태양은 조합 라이브러리의 서브세트를 유도할 수 있다.

[0347] 한 특정 구현예에서, 이런 배경 사이의 유전자 배경 및 유전자 다양성의 수(돌연변이 또는 서열 편집 거리 등의 수로 측정)는 이 방법의 효과를 최대화하도록 결정된다.

[0348] 유전자 배경은 천연, 고유한 또는 야생형 균주 또는 돌연변이된 조작 균주일 수 있다. N개의 구별된 배경 균주

는 벡터 b로 나타낼 수 있다. 한 실시예에서, 배경 b는 N개의 돌연변이된 배경 균주 $b = m_0b_0 = (m_1b_0, m_2b_0, \ldots m_Nb_0)$를 형성하기 위해 N개의 일차 돌연변이 $m_0 = (m_1, m_2, \ldots m_N)$을 야생형 배경 균주 b_0에 적용하여 형성된 조작된 배경을 나타낼 수 있으며, 여기서, m_ib_0는 배경 균주 b_0에 대한 돌연변이 m_i의 적용을 나타낸다.

[0349] 각각의 경우에(즉, 하나의 제공된 유전자 배경 또는 유전자 배경의 집합), 결과는 N개의 유전적으로 구별된 배경의 집합이다. 관련 표현형은 각 배경에 대해 측정된다.

[0350] 둘째, M 돌연변이 m_1의 집합에서 각 돌연변이는 N개의 배경 균주 b의 집합 내의 각 배경에 적용되어 M × N 돌연변이체의 집합을 형성한다. N개의 배경이 그 자체로 돌연변이 m_0의 기본 세트(위에 설명된 대로)를 적용하여 얻은 구현예에서, 돌연변이체의 생성된 집합은 때로는 조합 라이브러리 또는 쌍을 이루는 라이브러리로 불릴 수 있다. 공지된 배경의 집합이 명시적으로 제공되는 다른 구현예에서, 돌연변이체의 생성된 세트는 조합 라이브러리의 서브세트로 불릴 수 있다. 조작된 배경 벡터의 생성과 유사하게, 실시태양에서, 입력 인터페이스(202)는 돌연변이 벡터 m_1과 배경 벡터 b 및 교차 곱과 같은 특정 연산을 수신한다.

[0351] 상기 조작된 배경 예를 계속하면, MxN 조합 라이브러리를 형성하는 것은 $m_{1x}m_0b_0$에 의해 형성되는 행렬로 표현될 수 있으며, b의 N 배경에 적용된 m_1의 교차 곱 $b = m_0b_0$, m_1에서 각 돌연변이는 b 내의 각 배경 균주에 적용된다. 생성된 MxN 행렬의 각 i번째 행은 m_1 내 i번째 돌연변이의 배경 집합 b 내의 모든 균주에 대한 적용을 나타낸다. 하나의 실시태양에서, $m_1 = m_0$이고 행렬은 출발 균주 b_0에 동일한 돌연변이의 쌍을 이루는 적용을 나타낸다. 이 경우 행렬은 대각선(M = N) 근처에서 대칭이며 대각선은 동일한 돌연변이의 2회 적용을 나타내기 때문에 모든 분석에서 무시될 수 있다.

[0352] 실시태양에서, MxN 행렬을 형성하는 것은 입력 인터페이스(202)에 복합 표현 $m_{1x}m_0b_0$을 입력함으로써 달성될 수 있다. 표현의 구성요소 벡터는 하나 이상의 DNA 명세를 통해 명시적으로 지정된 요소로 직접 입력되거나 인터프리터(204)에 의한 해석 동안 벡터의 검색을 가능하게 하기 위해 라이브러리(206)에 대한 호출로서 입력될 수있다. 인터프리터(204), 실행 엔진(207), 주문 배치 엔진(208) 및 공장(210)을 통해 "Microbial Strain Design System and Methods for Improved Large Scale Production of Engineered Nucleotide Sequences,"이라는 제목의 미국 특허출원 제15/140,296호에 개시된 바와 같이, LIMS 시스템(200)은 입력 표현에 의해 지정된 미생물 균주를 생성한다.

[0353] 셋째, 도 29를 참조하면, 분석 장비(214)는 MxN 조합 라이브러리 행렬(4202) 내의 각각의 돌연변이체에 대한 표현형 반응을 측정한다. 이와 같이, 응답의 집합은 MxN 응답 행렬 R로 해석될 수 있다. R의 각 요소는 $r_{ij} = y(m_i, m_j)$로 나타낼 수 있으며, 여기서 y는 돌연변이 m_i에 의해 돌연변이된 조작된 집합 b 내의 배경 균주 b_j의 응답(성능)을 나타낸다. 단순성과 실용성을 위해, $m_1 = m_0$인 쌍을 이루는 돌연변이를 가정한다. 돌연변이의 세트가 상을 이루는 돌연변이 라이브러리를 나타내는 경우, 생성된 행렬은 유전자 상호작용 행렬 또는 보다 구체적으로 돌연변이 상호작용 행렬로 불릴 수 있다.

[0354] 당업자는, 일부 실시태양에서, 상위성 효과 및 예측 균주 디자인과 관련된 작업이, 예를 들어 분석 장비(214) 또는 인간의 구현에 의한 LIMS 시스템(200)의 자동화 수단을 통해 또는 자동화 및 수동 수단의 조합을 통해 완전히 수행될 수 있다는 것을 인식할 것이다. 작업이 완전히 자동화되지 않으면, 예를 들어, 분석 장비(214)와 같은 LIMS 시스템(200)의 요소는, 예를 들어, 자신의 작업 능력을 통해 결과를 생성하기 보다는 작업의 인간 수행 결과를 수신할 수 있다. 본 발명의 다른 곳에서 기술된 바와 같이, 분석 장비(214)와 같은 LIMS 시스템(200)의 구성요소는 전체적으로 또는 부분적으로 하나 이상의 컴퓨터 시스템에 의해 구현될 수 있다. 일부 실시태양에서, 특히 예측 균주 디자인과 관련된 작업이 자동 및 수동 수단의 조합에 의해 수행되는 경우, 분석 장비(214)는 컴퓨터 하드웨어, 소프트웨어 또는 펌웨어(또는 이의 조합)뿐만 아니라 아래 표 5에 열거된 것과 같은 인간 작업자에 의해 작동된 장비, 예를 들어 "성능 평가" 범주에 나열된 장비를 포함할 수 있다.

[0355] 넷째, 분석 장비(212)는 응답 행렬을 정규화한다. 정규화는 편향을 제거하고 및/또는 이 방법에 특이적인 효과의 관련 부분을 분리할 목적으로 측정된 응답 값을 조정하는 수동 및/또는 자동화 공정으로 구성된다. 도 29와 관련하여, 제 1 단계(4202)는 정규화된 측정 데이터를 얻는 단계를 포함 할 수 있다. 일반적으로, 예측 균주 디자인 및 상위성 매핑에 관한 청구항에서, "성능 측정" 또는 "측정 된 성능" 등의 용어는 일부 방식으로 처리되지 않거나 처리딘 측정 데이터, 예를 들어, 정규화된 데이터를 반영하는 메트릭을 기술하는데 사용될 수 있다.

특정 구현예에서, 정규화는 측정된 응답 값으로부터 이전에 측정된 배경 응답을 뺌으로써 수행될 수 있다. 그 구현예에서, 생성된 응답 요소는 $r_{ij} = y(m_i, m_j) - y(m_j)$로서 형성될 수 있으며, 여기서 $y(m_j)$는 1차 돌연변이 m_j의 부모 균주 b_0에 대한 적용에 의해 유발된 조직된 집합 b 내의 조작된 배경 균주 b_j의 반응이다. 정규화된 응답 행렬의 각 행은 상응하는 돌연변이에 대한 응답 프로파일로 취급된다는 것에 유의한다. 즉, i번째 행은 j = 1 내지 N에 대한 모든 배경 균주 b_j에 적용된 상응하는 돌연변이 m_i의 상대적 효과를 기술한다.

[0356] 쌍을 이루는 돌연변이의 예와 관련하여, 2개의 돌연변이로부터 생성된 균주의 성능/반응은 개별적으로 돌연변이의 각각에 대한 균주의 성능/반응보다 크거나, 작거나 같을 수 있다. 이 효과는 "상위성"으로 알려져 있으며, 일부 실시태양에서, $e_{ij} = y(m_i, m_j) -(y(m_i) + y(m_j)$로 나타낼 수있다. 이런 수학 표현의 변형이 가능하며, 예를 들어 개별 변화가 생물학적으로 어떻게 상화작용하는지에 따라 달라질 수 있다. 상기한 바와 같이, 동일한 작용기로부터의 돌연변이는 동일한 메카니즘에 의해 작동하기 쉽고, 따라서 전체 숙주 성능에 대해 음성 또는 중성 상위성을 나타내기 쉽다. 반대로, 상이한 작용기로부터의 돌연변이는 독립적인 메카니즘에 의해 작동하기 쉽고, 이는 예를 들어 여분의 돌연변이 효과를 감소시킴으로써 개량된 숙주 수행을 유도할 수 있다. 따라서, 유사하지 않은 반응을 일으키는 돌연변이는 유사한 반응을 일으키는 돌연변이보다 부가적인 방식으로 결합하기 쉽다. 이것은 다음 단계에서 유사성 계산을 유도한다.

[0357] 다섯째로, 분석 장비(214)는 응답들 사이의 유사성-쌍을 이루는 돌연변이 실시예에서, 응답 행렬(4204) 내에서 i번째 돌연변이와 j번째(예를 들어, 1차) 돌연변이의 영향 사이의 유사성을 측정한다. R의 i번째 행은 N개의 배경 균주에 대한 i번째 돌연변이 m_i의 성능 영향을 나타내며, 이의 각각은 자체로 상기한 대로 조작된 돌연변이의 결과일 수 있다는 것을 상기한다. 따라서 i번째와 j번째 돌연변이의 영향 사이의 유사성은 유사성 행렬 S를 형성하기 위해 각각 i번째와 j번째 행, ρ_i및 ρ_j 사이의 유사성 s_{ij}로 각각 나타낼 수 있으며, 이의 예는 도 53에 예시된다. 유사성은 교차-상관관계 또는 절대 코사인 유사성, 예를 들어 $s_{ij} = abs(cos(\rho_i, \rho_j)$과 같은 많은 공지된 기술을 사용하여 측정될 수 있다.

[0358] 코사인 유사성과 같은 메트릭에 대한 대안 또는 보완으로서, 응답 프로파일은 클러스터링되어 유사성의 등급을 결정할 수 있다. 클러스터링은 적절한 거리 측정(예를 들어, 유클리드(Euclidean), 해밍(Hamming) 등)과 함께 거리 기반 클러스터링 알고리즘(예를 들어, k- 평균, 계층적 응집 등)의 사용에 의해 수행될 수 있다. 대안적으로, 클러스터링은 적절한 유사성 척도(예를 들어, 코사인, 상관관계 등)를 갖는 유사성 기반 클러스터링 알고리즘(예를 들어, 스펙트럼, 최소-커트 등)을 사용하여 수행될 수 있다. 물론, 거리 측정은 임의의 수의 표준 기능 작업(예를 들어, 대수 함수)을 통해 유사성 측정치로 매핑될 수 있고 그 반대의 경우도 가능하다. 일 구현예에서, 계층적 응집 클러스터링은 절대 코사인 유사성과 함께 사용될 수 있다.(도 54a 참조).

[0359] 클러스터링의 한 예로서, C를 k개의 구별된 클러스터로의 돌연변이 m_i의 클러스터링이라하자. C를 클러스터 멤버십 행렬이라 하며, 여기서 c_{ij}는 돌연변이 i가 클러스터 j에 속하는 정도, 0과 1 사이의 값이다. 돌연변이 i와 j의 사이의 클러스터 기반 유사성은 $C_i x C_j$(C의 i번째 및 j번째 행의 도트 곱)으로 제공된다. 일반적으로, 클러스터 기반의 유사성 행렬은 CC^T(즉, C 곱 C-트랜스포즈)로 제공된다. 하드 클러스터링의 경우(돌연변이가 정확하게 하나의 클러스터에 속한다), 두 돌연변이 사이의 유사성은 동일한 클러스터에 속하면 1이고 그렇지 않은 경우 0이다.

[0360] Costanzo, The Genetic Landscape of a Cell, Science, Vol. 327, Issue 5964, Jan. 22, 2010, pp. 425-431(그 전문이 본 명세서에서 참조로 포함됨)에 기술된 바와 같이, 돌연변이 반응 프로파일의 이런 클러스터링은 세포의 근본적인 기능적 조직의 대강의 매핑에 관한 것이다. 즉, 함께 집단화되는 돌연변이는 근본적인 생물학적 과정이나 대사 경로와 관련되는 경향이 있다. 이런 돌연변이는 본 명발명에서는 "작용기"라고 부른다. 이 방법의 주요 관측은 두 돌연변이가 동일한 생물학적 과정 또는 경로에 의해 작동한다면, 관측된 효과(및 주목할 만하게 관찰된 이점)가 중복될 수 있다는 것이다. 반대로, 두 개의 돌연변이가 구별된 메커니즘에 의해 작동한다면, 유익한 효과가 중복될 가능성이 적을 것이다.

[0361] 여섯째, 상위성 효과에 기초하여, 분석 장비(214)는 비유사 응답을 유도하는 돌연변이의 쌍을 선택하는데, 예를 들어 도 29(4206)에 도시된 바와 같이(예를 들어, 도 53 및 도 54a) 이들의 코사인 유사성 메트릭이 유사성 임계치 아래로 떨어지거나 이들의 반응이 충분히 분리된 클러스터들 내에 있다. 이들의 비유사성에 근거하여, 선택된 쌍의 돌연변이는 유사한 쌍보다 배경 균주로 통합되어야 한다.

[0362] 충분히 비유사한 응답을 유도하는 선택된 쌍의 돌연변이에 기초하여, LIMS 시스템(예를 들어, 인터프리터(204), 실행 엔진(207), 주문 배치기(208) 및 공장(210)의 모든 또는 일부 조합)이 그 선택된 돌연변이(4208)를 가진 미생물 균주를 디자인하는데 사용될 수 있다. 실시태양에서, 이하 및 본 발명의 다른 곳에서 기술된 바와 같이, 상위성 효과는 균주 선택을 가중하거나 여과하기 위해 예측 모델 속에서 형성될 수 있거나 예측 모델과 함께 사용될 수있다.

[0363] 일부 바람직한 예측 모델을 통해 특정 배경 속에 라이브러리로부터의 돌연변이의 집합을 통합함으로써 얻어진 가상의 균주의 성능(즉, 점수)을 추정할 수 있다고 가정한다. 교시된 방법에서 이용된 대표적인 예측 모델은 "Computational Analysis and Prediction of Effects of Genome-Wide Genetic Design Criteria."의 더 큰 섹션에서 발견된 "Predictive Strain Design"란 제목의 이하 섹션에 제공된다.

[0364] 선형 회귀와 같은 예측 균주 디자인 기술을 사용하는 경우, 분석 장비(214)는 단지 충분하게 비유사한 돌연변이를 유지하기 위해 회귀 결과를 필터링하여 낮은 유사성 척도를 가진 돌연변이에 모델을 제한할 수 있다. 대안적으로, 예측 모델은 유사성 행렬로 가중될 수 있다. 예를 들어, 일부 실시태양은 제안된 돌연변이의 상호의존성을 특성화하기 위해 유사성 행렬을 사용하여 가중된 최소 제곱 회귀(weighted least squares regression)를 이용할 수 있다. 한 예로서, 가중치 부여는 회귀 모델에 "커널" 트릭을 적용함으로써 수행될 수 있다.("커널 트릭"이 많은 기계 학습 모델링 접근법에 일반적이기 때문에, 이런 재-가중 전략은 선형 회귀에 제한되지 않는다.)

[0365] 이런 방법은 당업자에게 공지되어 있다. 실시태양에서, 커널은 요소 $1 - w^{*} s_{ij}$를 갖는 행렬이고, 여기서 1은 단위 행렬의 요소이고, w는 0과 1 사이의 실수 값이다. w = 0일 때, 이것은 표준 회귀 모델로 감소한다. 실제로, w의 값은 쌍을 이루는 조합 구조체 및 이의 관련 효과 y(mi, mj)에 대해 평가될 때 예측 모델의 정확성(r2 값 또는 루트 평균 제곱 에러(RMSE))에 연결될 것이다. 하나의 간단한 구현예에서, w는 $w = 1-r^{2}$로 정의된다. 이 경우, 모델이 완전히 예측 가능한 경우, $w = 1-r^{2} = 0$이고 통합은 예측 모델만을 기반으로 하며 통합은 상위성 매핑 절차만 기반으로 한다. 각각의 반복 동안, 정확도를 평가하여 모델 성능이 개량되는지 여부를 결정할 수 있다.

[0366] 본 발명에 기재된 상위성 매핑 절차는 어느 모델이 분석 장비(214)에 의해 사용되는지에 의존하지 않는 것이 명백해야 한다. 이런 예측 모델을 고려하면, 조합 통합을 통해 돌연변이 라이브러리에 접근할 수 있는 모든 가상의 균주를 점수매기고 순위를 매기는 것이 가능하다.

[0367] 일부 실시태양에서, 상위성 효과를 설명하기 위해, 비유사한 돌연변이 반응 프로파일은 예측 모델로부터 각 가상 균주와 연관된 점수 및 순위를 증대하기 위해 분석 모델(214)에 의해 사용될 수 있다. 이 절차는 교체로 점수의 재-가중화로서 생각될 수 있어서 비유사한 응답 프로파일을 가진 후보 균주(예를 들어, 클러스터의 다양성으로부터 얻은 균주)를 지지할 수 있다. 하나의 간단한 구현예에서, 균주는 비유사성 임계값을 만족하지 않거나 동일한 클러스터(적당한 가중치를 가짐)로부터 얻은 구성 돌연변이의 수만큼 감소된 점수를 가질 수 있다. 한 특정 구현예에서, 가상 균주의 성능 평가는(다시 적당한 가중치에 의해) 가상의 균주와 관련된 구성 돌연변이의 모든 쌍과 관련된 유사성 행렬의 항의 합계만큼 감소될 수 있다.가상 균주는 이러한 증가된 점수를 사용하여 순위가 다시 매겨질 수 있다. 실제로, 이런 재-가중 계산은 초기 점수 추정과 함께 수행될 수 있다.

[0368] 결과는 혼동하는 상위성 상호작용을 보다 효과적으로 피하기 위해 증가된 점수 및 순위를 가진 가상 균주의 집합이다. 가상 균주는 이 시점에서 만들어질 수 있거나 후속 분석이나 사용을 위해 다른 계산 방법으로 넘겨질 수 있다.

[0369] 당업자는 본 발명에 기술된 바와 같이 상위성 매핑 및 반복적인 예측 균주 디자인이 단지 쌍을 이루는 돌연변이를 사용하는 것에 제한되지 않고 배경 균주에 더 많은 돌연변이를 동시에 적용하는 것으로 확장될 수 있음을 인식할 것이다. 다른 실시태양에서, 추가적인 돌연변이가 본 발명에 기재된 예측 방법에 따라 선택된 돌연변이를 사용하여 이미 돌연변이된 균주에 순차적으로 적용될 수 있다. 다른 실시태양에서, 상위성 효과는 서로 약간 다른 다수의 균주 배경에 동일한 유전자 변이를 적용하고 돌연변이된 균주 배경 사이에 양성 반응 프로파일에 임의의 유의한 차이가 있음을 나타냄으로써 귀속된다.

[0370] **외인성 DNA를 도입하기 위한 HTP 컨쥬게이팅 컨쥬게이션**

[0371] 본 발명은 또한 공여자 미생물 세포로부터 사카로폴리스포라 미생물의 수용자 세포로 유전자 물질을 전달하는 방법을 제공한다. 공여자 미생물 세포는 대장균 세포를 포함하지만 이에 제한되지 않는 임의의 적합한 공여자

세포일 수 있다. 수용자 미생물 세포는 S. 스피노사 균주와 같은 사카로폴리스포라 종일 수 있다.

[0372] 일반적으로, 상기 방법은 다음 단계를 포함한다: (1) 수용자 세포를 중기-대수(mid-exponential) 상으로 계대배양(subculturing)하는 단계(선택적); (2) 공여자 세포를 중기-대수 상으로 계대배양하는 단계(선택적); (3) 공여자 및 수용자 세포를 결합하는 단계; (4) 컨쥬게이션 매질 상에 공여자 및 수용자 세포 혼합물을 플레이팅하는 단계; (5) 세포를 컨쥬게이션시키기 위해 플레이트를 배양하는 단계; (6) 공여자 세포에 대한 항생제 선택을 적용하는 단계; (7) 비-통합 수용자 세포에 대한 항생제 선택을 적용하는 단계; 및 (8) 통합된 수용자 세포의 성장을 허용하기 위해 플레이트를 추가로 배양하는 단계.

[0373] 본 출원의 발명자들은 S. 스피노사에서 놀랍게도 증가된 외인성 DNA 컨쥬게이션의 빈도를 초래하는 최적화될 수 있는 조건을 발견하였다. 이러한 조건은 다음을 포함하지만 이에 제한되지 않는다: (1) 수용자 세포가 세척된다(예를 들어, 컨쥬게이션 전); (2) 공여자 세포 및 수용자 세포는 상대적으로 낮은 온도에서 컨쥬게이션된다; (3) 수용자 세포를 컨쥬게이션하기 전에 연장된 기간 동안 계대배양한다; (4) 컨쥬게이션을 위한 공여자 세포: 수용자 세포의 적절한 비; (5) 공여자 세포에 대한 선택을 위해 항생제 약물을 컨쥬게이션 혼합물로 전달하는 적절한 타이밍; (6) 수용자 세포에 대한 선택을 위해 항생제 약물을 컨쥬게이션 혼합물로 전달하는 적절한 타이밍; (7) 공여자 및 수용자 세포 혼합물로 플레이팅된 컨쥬게이션 매질을 건조시키는 적절한 타이밍; (8) 고농도의 글루코스; (9) 적절한 농도의 공여자 세포; 및 (10) 적절한 농도의 수용자 세포.

[0374] 일부 실시태양에서, 다음 조건 중 적어도 2개, 3개, 4개, 5개, 6개, 7개 또는 그 이상이 이용되어 컨쥬게이션이 증가된다:

[0375] (1) 수용자 세포가 세척된다;

[0376] (2) 공여자 세포 및 수용자 세포는 약 25℃, 26℃, 27℃, 28℃, 29℃, 30℃, 31℃, 32℃, 33℃, 예를 들어, 30℃의 온도에서 컨쥬게이션된다;

[0377] (3) 수용자 세포는 컨쥬게이션 전 적어도 약 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55시간 동안, 예를 들어, 약 48시간 동안 계대배양된다;

[0378] (4) 컨쥬게이션을 위한 공여자 세포:수용자 세포의 비는 약 1:0.5, 1:0.6, 1:0.7, 1:08, 1:0.9, 1:1.0, 1:1.1, 1:1.2, 1:1.3, 1:1.4, 1:1.5, 1:1.6, 1:1.7, 1:1.8, 1:1.9 또는 1:2.0, 예를 들어, 약 1:0.6 내지 1:1.0이다;

[0379] (5) 공여자 세포에 대한 선택을 위한 항생제 약물은 공여자 세포와 수용자 세포가 혼합된 후 약 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 또는 30시간 후, 예를 들어, 약 24시간 후에 혼합물로 전달된다;

[0380] (6) 수용자 세포에 대한 선택을 위한 항생제 약물은 공여자 세포 및 수용자 세포가 혼합된 후 약 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54 또는 55시간 후, 예를 들어, 약 40 내지 48시간 후에 혼합물에 전달된다;

[0381] (7) 공여자 및 수용자 세포 혼합물로 플레이팅된 컨쥬게이션 매질을 적어도 약 1시간, 2시간, 3시간, 4시간, 5시간, 6시간, 7시간, 8시간, 9시간, 10시간, 11시간, 12시간, 13시간, 14시간 또는 15시간 동안 건조시킨다;

[0382] (8) 컨쥬게이션 매질은 적어도 0.5g/L, 1g/L, 1.5g/L, 2g/L, 2.5g/L, 3g/L, 3.5g/L, 4g/L, 4.5g/L, 5g/L, 5.5g/L, 6g/L, 6.5g/L, 7g/L, 7.5g/L, 8g/L, 8.5g/L, 9g/L, 9.5g/L, 10g/L 또는 그 이상의 글루코스를 포함한다;

[0383] (9) 공여자 세포의 농도는 약 OD600=0.1, 0.15, 0.2, 0.25, 0.30, 0.31, 0.32, 0.33, 0.34, 0.35, 0.36, 0.37, 0.38, 0.39, 0.4, 0.41, 0.42, 0.43, 0.44, 0.45, 0.46, 0.47, 0.48, 0.49, 0.50, 0.51, 0.52, 0.53, 0.54, 0.55, 0.56, 0.57, 0.58, 0.59, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.0이다; 및

[0384] (10) 수용자 세포의 농도는 약 OD540=1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0, 10.5, 11.0, 11.5, 12.0, 12.5, 13.0, 13.5, 14.0, 14.5 또는 15.0이다.

[0385] 일부 실시태양에서, 혼합물 중 공여자 세포 또는 수용자 세포의 총 수는 약 $5x10^6$, $6x10^6$, $7x10^6$, $8x10^6$, 또는 약 $9x10^6$이다.

[0386] 일부 실시태양에서, 공여자 세포는 대장균 세포이고, 공여자 세포에 대한 선택을 위한 항생제는 날리딕산이다. 일부 실시태양에서, 날리딕산의 농도는 약 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 60, 170, 180,

190 또는 200㎍/ml이다.

[0387] 일부 실시태양에서, 수용자 세포에 대한 선택을 위한 항생제는 아프라마이신이고, 농도는 약 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 60, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300㎍/ml이다.

[0388] 본 발명에 기재된 바와 같은 방법은 고 처리량 공정으로 수행될 수 있다. 일부 실시태양에서, 방법은 48-웰 Q-트레이에서 수행된다. 일부 실시태양에서, 고 처리량 공정은 부분적으로 또는 완전히 자동화된다.

[0389] 일부 실시태양에서, 공여자 세포 및 수용자 세포의 혼합물은 액체 혼합물이고, 액체 혼합물의 충분한 부피가 요동 운동(rocking motion)을 하는 배지 상에 플레이팅되고, 액체 혼합물이 배지의 전체 영역에 걸쳐 분산된다.

[0390] 일부 실시태양에서, 방법은 공여자 세포에 의해 제공된 통합된 DNA를 갖는 수용자 세포의 후속 접종을 위해 효모 핀을 사용하는 콜로니 피킹에 의해 컨쥬게이션 완료체(exconjugant)를 전달하는 자동화된 공정을 포함한다. 일부 실시태양에서, 콜로니 피킹은 침지 운동 또는 교반 운동으로 수행된다.

[0391] 일부 실시태양에서, 방법은 다음 조건 중 적어도 2개, 3개, 4개, 5개, 6개 또는 7개를 사용하여 수행된다: (1) 수용자 세포는 컨쥬게이션 전에 세척된다; (2) 공여자 세포 및 수용자 세포는 약 30℃의 온도에서 컨쥬게이션된다; (3) 수용자 세포를 컨쥬게이션하기 전에 적어도 약 48시간 동안 계대배양한다; (4) 컨쥬게이션을 위한 공여자 세포:수용자 세포의 비는 약 1:0.8이다; (5) 공여자 세포에 대한 선택을 위한 항생제 약물은 공여자 세포와 수용자 세포가 혼합된 후 약 20시간 후에 혼합물에 전달된다; (6) 혼합물에서 공여자 세포의 양 또는 수용자 세포의 양은 약 $7x10^6$이다; 및 (7) 컨쥬게이션 매질은 약 6g/L 글루코스를 포함한다.

[0392] **경로 리팩토링(Pathway Refactoring)**

[0393] 본 발명은 경로 리팩토링 방법을 제공한다. 본 발명에 사용된 용어 "경로 리팩토링"은 미생물에서 하나 이상의 완전히 또는 부분적으로 최적인 생합성 경로를 구축하는 과정을 지칭한다. 일부 실시태양에서, 생합성 경로는 스피노신과 같은 하나 이상의 관심 생성물의 합성과 관련된다.

[0394] 경로 리팩토링 방법은 본 발명의 하나 이상의 도구를 이용할 수 있다. 임의의 특정 이론에 구속되지 않고, 경로 리팩토링 방법은 생합성 경로에 직접 관련된 하나 이상의 유전자의 활성, 또는 생합성 경로에 간접적으로 관련된 하나 이상의 유전자(예를 들어, 주어진 관심 생성물의 생합성에 간접적으로 영향을 줄 수 있는 유전자)의 활성을 미세 조정할 수 있다. 일부 실시태양에서, 생합성 경로에 관여하는 하나 이상의 유전자를 미세 조정하기 위해, 본 발명의 방법은 프로모터 래더 라이브러리, RBS 래더 라이브러리, 터미네이터 라이브러리, 정지/개시 코돈 라이브러리 등을 포함하지만 이에 제한되지 않는 본 발명의 하나 이상의 유전자 다양성 라이브러리를 이용하는 것을 포함한다. 일부 실시태양에서, 생합성 경로에 관여하는 하나 이상의 유전자의 활성은 본 발명에 개시된 바와 같은 하나 이상의 유전자 도구에 의해 변형된다. 일부 실시태양에서, 변형된 유전자를 보유하는 균주는 변형이없는 균주와 같은 체크 균주와 비교하여 개량된 성능을 갖는 균주를 확인하기 위해 본 발명에 기재된 바와 같은 하이 스루풋 시스템을 통해 선별될 수 있다.

[0395] 그 결과, 생합성 경로에 관여하는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 그 이상의 유전자가 미세 조정된다. 일부 실시태양에서, 임의의 수의 유전자가 미세 조정된다. 일부 실시태양에서, 미세 조정된 유전자는 동일한 신호 경로 또는 합성 경로에 있다. 일부 실시태양에서, 미세 조정된 유전자는 상이한 신호 경로 또는 합성 경로에 있다. 일부 실시태양에서, 변형이 변형의 성능을 향상시키는 한, 특정 유전자의 활성은 필요에 따라 변형된다. 일부 실시태양에서, 하나 이상의 유전자의 활성은 체크 균주에서의 활성과 비교하여 상향 조절된다. 일부 실시태양에서, 하나 이상의 유전자의 활성은 체크 균주에서의 활성과 비교하여 하향 조절된다. 일부 실시태양에서, 하나 이상의 유전자의 발현 타이밍은 체크 균주에서의 것에 비해 변화된다. 일부 실시태양에서, 하나 이상의 유전자의 발현 위치는 체크 균주에서의 위치와 비교하여 변경된다. 일부 실시태양에서, 속도 결정 단계(RDS) 또는 속도 제한 단계에 관여하는 하나 이상의 유전자의 활성은 체크 균주에서의 것에 비해 변형된다. 일부 실시태양에서, 하나, 둘, 셋, 넷, 다섯, 여섯, 일곱, 여덟, 아홉, 열 또는 그 이상의 변형된 유전자유전자좌가 통합되어 더욱 미세 조정된 생합성 경로를 갖는 균주를 생성한다. 레이트 결정 단계 (RDS) 또는 레이트 제한 단계에 관여하는 하나 이상의 유전자의 활성은 체크 균주와 비교하여 변형된다. 일부 실시태양에서, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 또는 그 이상의 변형된 유전자 유전자좌가 통합되어 더욱 미세 조정된 생합성 경로를 갖는 균주를 생성한다.

[0396] 일부 실시태양에서, 경로 리팩토링 방법은 유전자 물질을 본 발명의 미생물의 게놈에 혼입시키는 것을

포함한다. 일부 실시태양에서, 미생물은 사카로폴리스포라 종, 예를 들어, 사카로폴리스포라 스피노사이며, 유전자 물질은 미생물의 게놈에서의 특정 위치(예를 들어, "랜딩 패드")에 통합된다. 일부 실시태양에서, 특정 위치는 본 발명에 기재된 바와 같은 본 발명의 중성 통합 부위(NIS)로부터 선택된다.

[0397] 일부 실시태양에서, 유전자 물질은 자기 복제 가능 벡터를 통해 본 발명의 미생물 내로 도입된다. 일부 실시태양에서, 미생물은 사카로폴리스포라 종, 예를 들어 사카로폴리스포라 스피노사이며, 유전자 물질이 본 발명에 기술된 바와 같은 본 발명의 자기 복제 플라스미드를 통해 미생물에 도입된다.

[0398] **유전자 디자인에 순응하는 유기체**

[0399] 개시된 HTP 게놈 공학 플랫폼은 산업용 미생물 세포 배양물(예를 들어, 사카로폴리스포라 종)로 예시되지만, 원하는 형질이 유전자 돌연변이 집단에서 확인될 수 있는 임의의 숙주 세포 유기체에 적용 가능하다.

[0400] 따라서, 본 발명에 사용된 용어 "미생물"은 광범위하게 해석되어야 한다. 미생물은 두 가지 원핵생물 영역인 박테리아와 고세균뿐만 아니라 특정 진핵생물 균류와 원생 생물을 포함한다. 그러나, 특정 양태에서, 곤충, 식물 및 동물과 같은 "보다 높은" 진핵생물이 본 발명에 교시된 방법에서 이용될 수 있다.

[0401] 적합한 숙주 세포를 다음을 포함하지만 이에 제한되지 않는다: 사카로폴리스포라 안티마이크로비아(*Saccharopolyspora antimicrobia*), 사카로폴리스포라 카버나에(*Saccharopolyspora cavernae*), 사카로폴리스포라 세부엔시스(*Saccharopolyspora cebuensis*), 사카로폴리스포라 덴드란테마에(*Saccharopolyspora dendranthemae*), 사카로폴리스포라 에리트라에아(*Saccharopolyspora erythraea*), 사카로폴리스포라 플라바(*Saccharopolyspora flava*), 사카로폴리스포라 가르다이엔시스(*Saccharopolyspora ghardaiensis*), 사카로폴리스포라 글로리오사에(*Saccharopolyspora gloriosae*), 사카로폴리스포라 그레고리(*Saccharopolyspora gregorii*), 사카로폴리스포라 할로필(*Saccharopolyspora halophile*), 사카로폴리스포라 할로토레란스(*Saccharopolyspora halotolerans*), 사카로폴리스포라 허슈테(*Saccharopolyspora hirsute*), 사카로폴리스포라 호르데이(*Saccharopolyspora hordei*), 사카로폴리스포라 인디카(*Saccharopolyspora indica*), 사카로폴리스포라 지앙시엔시스(*Saccharopolyspora jiangxiensis*), 사카로폴리스포라 라시살시(*Saccharopolyspora lacisalsi*), 사카로폴리스포라 파탈룽겐시스(*Saccharopolyspora phatthalungensis*), 사카로폴리스포라 퀴지아오진겐시스(*Saccharopolyspora qijiaojingensis*), 사카로폴리스포라 렉티버굴라(*Saccharopolyspora rectivirgula*), 사카로폴리스포라 로세아(*Saccharopolyspora rosea*), 사카로폴리스포라 샨돈겐시스(*Saccharopolyspora shandongensis*), 사카로폴리스포라 스피노사(*Saccharopolyspora spinosa*), 사카로폴리스포라 스피노스포로트리키아(*Saccharopolyspora spinosporotrichia*), 사카로폴리스포라 타베리(*Saccharopolyspora taberi*), 사카로폴리스포라 써모필(*Saccharopolyspora thermophile*), 및 사카로폴리스포라 트립테리기(*Saccharopolyspora tripterygii*).

[0402] 일부 실시태양에서, 숙주 세포는 사카로폴리스포라 인디아네시스(*Saccharopolyspora indianesis*)(ATCC® BAA-2551™), 사카로폴리스포라 에리트라에아(Waksman) 라베다(ATCC® 31772™), 사카로폴리스포라 에리트라에아(Waksman) 라베다(ATCC® 11912™), 사카로폴리스포라 렉티버굴라(Krasil'nikov 및 Agre) 콘-웬디쉬(Korn-Wendisch) 외.(ATCC® 29034™), 사카로폴리스포라 허슈타(*Saccharopolyspora hirsuta*) subsp. 허슈타 라세이(hirsuta Lacey) 및 굿펠로우(Goodfellow)(ATCC® 27875™), NEB#998(ATCC® 98102™), 사카로폴리스포라 허슈타 subsp. 코벤시스(kobensis)(Iwasaki 외.) 라세이(ATCC® 20501™), 사카로폴리스포라 렉티버굴라(Krasil'nikov 및 Agre) 콘-웬디쉬 외.(ATCC® 29035™), 사카로폴리스포라 에리트라에아(Waksman) 라베다(ATCC® 11635D-5™) ATCC® Number: 11635D-5™, 사카로폴리스포라 타베리(Labeda) 콘-웬디쉬 외.(ATCC® 49842™), 사카로폴리스포라 허슈타 subsp. 허슈타 라세이 및 굿펠로우(ATCC® 27876™), 사카로폴리스포라 오란티아카(Saccharopolyspora aurantiaca) 에티엔(Etienne) 외.(ATCC® 51351™), 사카로폴리스포라 그레고리 굿펠로우 외.(ATCC® 51265™), 사카로폴리스포라 에리트라에(Waksman) 라베다(ATCC® 11635™), 사카로폴리스포라 렉티버굴라(Krasil'nikov 및 Agre) 콘-웬디쉬 외.(ATCC® 33515™), 사카로폴리스포라 렉티버굴라(Krasil'nikov 및 Agre) 콘-웬디쉬 외.(ATCC® 15347™), 사카로폴리스포라 스피노사 메르츠(Mertz) 및 야오(Yao)(ATCC® 49460™), 사카로폴리스포라 렉티버굴라(Krasil'nikov 및 Agre) 콘-웬디쉬 외.(ATCC® 21450™), 사카로폴리스포라 호르데이(Saccharopolyspora hordei) 굿펠로우 외.(ATCC® 49856™), 사카로폴리스포라 렉티버굴라(Krasil'nikov 및 Agre) 콘-웬디쉬 외.(ATCC® 29681™), pIJ43 [MCB1023] (ATCC® 39156™), pOJ31 (ATCC® 77416™), 및 사카로폴리스포라 렉티버굴라(21451)로부터 선택된다.

[0403] **유전 디자인 및 HTP 미생물 공학 플랫폼에서 활용을 위한 유전자 다양성 풀 생성**

[0404] 일부 실시태양에 있어서, 본 발명의 방법은 유전자 디자인으로서 특징화된다. 본 발명에 사용된 용어 유전자 디자인은 새로운 우수한 숙주 세포를 디자인하고 생성하기 위해 특정 유전자, 유전자의 일부, 프로모터, 종결 코돈, 5'UTR, 3'UTR, 리보솜 결합 부위, 터미네이터 또는 다른 DNA 서열의 가장 최적의 변이체의 확인 및 선택을 통한 숙주 유기체 게놈의 재구성 또는 변형을 의미한다.

[0405] 일부 실시태양에서, 본 발명의 유전자 디자인 방법의 제 1 단계는 새로운 숙주 게놈이 재구성될 수 있는 복수의 서열 변형을 갖는 초기 유전자 다양성 풀 집단을 얻는 것이다.

[0406] 일부 실시태양에서, 본 발명에 교시된 유전자 디자인 방법의 후속 단계는 상기한 HTP 분자 도구 세트(예를 들어, SNP 스와핑 또는 프로모터 스와핑)의 하나 이상을 사용하여 HTP 유전자 디자인 라이브러리를 구축한 다음, 숙주 세포에서의 시험을위한 특정 게놈 변형의 라이브러리를 제공함으로써 게놈 공학 과정의 동인으로서 작용하는 것이다.

[0407] ***현존하는 야생형 균주로부터의 다양성 풀 사용하기***

[0408] 일부 실시태양에서, 본 발명은 소정의 야생형 집단의 미생물 중에서 존재하는 서열 다양성을 확인하는 방법을 교시한다. 따라서, 다양성 풀은 분석을 위해 사용된 소정의 수 n개의 야생형 미생물일 수 있으며, 상기 미생물의 게놈은 "다양성 풀"을 나타낸다.

[0409] 일부 실시태양에서, 다양성 풀은 상기 야생형 미생물 중 자연적 유전자 변이에 존재하는 현존하는 다양성의 결과일 수 있다. 이 변이는 소정의 숙주 세포의 균주 변이체로부터 얻을 수 있고 미생물이 완전히 다른 종인 결과일 수도 있다. 유전자 변이는 자연 발생 여부와 상관없이 균주의 유전자 서열에서 어떠한 차이도 포함할 수 있다. 일부 실시태양에서, 유전자 변이는 다른 것들 중에서 SNP 스왑, PRO 스왑, 개시/정지 코돈 스왑, STOP 스왑, 트랜스포존 돌연변이유발 다양성 라이브러리, 리보솜 결합 부위 다양성 라이브러리, 항-대사산물 선택/발효 생성물 저항성 라이브러리를 포함할 수 있다.

[0410] ***현존하는 산업용 균주 변이체로부터의 다양성 풀 사용하기***

[0411] 본 발명의 다른 실시태양에서, 다양성 풀은 전통적인 균주 개량 과정 동안 생성된 균주 변이체(예를 들어, 랜덤 돌연변이를 통해 생성되고 수년 동안 개량된 수율을 위해 선택된 하나 이상의 숙주 유기체 균주)이다. 따라서, 일부 실시태양에서, 다양성 풀 또는 숙주 유기체는 역사적 생산 균주의 집합을 포함할 수 있다.

[0412] 다양성 풀은 특정 시점(S_1Gen_1)에서 "기준선" 유전자 서열을 갖는 본래의 부모 균주(S_1) 및 상기 S_1 균주로부터 유도/성장되고 S_1의 기준선 게놈과 관련하여 다른 게놈($S_{2-n}Gen_{2-n}$)을 갖는 임의의 수의 후속 자손 균주(S_{2-n}로 일반화될 수 있는 S_2, S_3, S_4, S_5, 등)일 수 있다.

[0413] 예를 들어, 일부 실시태양에서, 본 발명은 다양성 풀에서 미생물 게놈을 서열화하여 각 균주에 존재하는 SNP를 확인하는 것을 교시한다. 한 실시태양에서, 다양성 풀의 균주는 역사적인 미생물 생산 균주이다. 따라서, 본 발명의 다양성 풀은 예를 들어 산업용 표준 균주 및 전통적인 균주 개량 프로그램을 통해 생산된 하나 이상의 돌연변이 산업용 균주를 포함할 수 있다.

[0414] 다양성 풀 내의 모든 SNPS가 확인되면, 본 발명은 개별적으로 및/또는 그룹으로 SNP의 효과(예를 들어, 관심 표현형의 생성)를 묘사(즉, 정량화 및 특성화)하기 위한 SNP 스와핑 및 선택 방법을 교시한다. 따라서, 상기한 바와 같이, 교시된 플랫폼에서의 초기 단계는 복수의 서열 변이체, 예를 들어 SNP를 가진 초기 유전자 다양성 풀 집단을 얻는 것을 수 있다. 그런 후에, 교시된 플랫폼의 후속 단계는 상기한 HTP 분자 도구 세트(예를 들어, SNP 스와핑 또는 프로모터 스와핑)의 하나 이상을 사용하여 HTP 유전자 디자인 라이브러리를 구축한 다음, 숙주 세포에서의 시험을위한 특정 게놈 변형의 라이브러리를 제공함으로써 게놈 공학 과정의 동인으로서 작용하는 것이다.

[0415] 일부 실시태양에서, 본 발명의 SNP 스와핑 방법은 돌연변이된 균주(예를 들어, $S_{2-n}Gen_{2-n}$으로부터 균주)에서 확인된 하나 이상의 SNP를 표준 균주(S_1Gen_1) 또는 야생형 균주에 도입하는 단계를 포함한다.

[0416] 다른 실시태양에서, 본 발명의 SNP 스와핑 방법은 돌연변이된 균주(예를 들어, $S_{2-n}Gen_{2-n}$으로부터 균주)에서 확인된 하나 이상의 SNP를 제거하는 단계를 포함한다.

[0417] ***돌연변이유발을 통해 다양성 풀 만들기***

[0418] 일부 실시태양에서, 세포의 소정의 다양성 집단에서 관심 돌연변이는 돌연변이유발 화학물질 또는 방사선을 포함하는 돌연변이 균주를 위한 임의의 수단에 의해 인위적으로 생성될 수 있다. 용어 "돌연변이유발"은 본 발명에서 세포질 핵산 물질에서 하나 이상의 유전자 변형을 유도하는 방법을 의미하기 위해 사용된다.

[0419] 용어 "유전자 변형"은 DNA의 임의의 변경을 말한다. 대표적인 유전자 변형은 뉴클레오티드 삽입, 결실, 치환 및 이의 조합을 포함하며, 단일 염기만큼 작거나 수만 염기만큼 클 수 있다. 따라서, 용어 "유전자 변형"은 뉴클레오티드 서열의 역전 및 염색체의 영역을 포함하는 DNA의 위치 또는 배향이 변경되는 다른 염색체 재배열을 포함한다. 염색체 재배열은 염색체 내 재배치 또는 염색체 간 재배치를 포함할 수 있다.

[0420] 한 실시태양에서, 현재 청구된 주제에 사용된 돌연변이유발 방법은 실질적으로 랜덤라서 유전자 돌연변이가 돌연변이유발될 핵산 물질 내의 임의의 이용 가능한 뉴클레오티드 위치에서 발생할 수 있다. 달리 말하면, 한 실시태양에서, 돌연변이유발은 특정 뉴클레오티드 서열에서 발생의 우선성성 또는 증가 빈도를 나타내지 않는다.

[0421] 본 발명의 방법은 자외선, X-레이 방사선, 감마선 방사선, N-에틸-N-니트로소우레아(ENU), 메틸니트로우레아(MNU), 프로카바진(PRC), 트라이에틸렌 멜라민(TMS), 아크릴아마이드 모노머(AA), 클로람부실(CHL), 멜팔란(MLP), 사이클로포스파미드(CPP), 다이에틸 설페이트(DES), 에틸 메테인 설포네이트(EMS), 메틸 메테인 설포네이트(MMS), 6-머캡토퓨린(6-MP), 마이토마이신-C(MMC), N-메틸-N'-니트로-N-니트로소구아니딘(MNNG), $3H_2O$ 및 우레탄(UR)(예를 들어, Rinchik, 1991; Marker *et al.*, 1997; 및 Russell, 1990 참조). 추가적인 돌연변이유발 인자는 iephb.nw.ru/~spirov/hazard/mutagen_lst.에 기재된 것을 포함하여 당업자에게 주지되어 있다.

[0422] 일부 실시태양에서, 본 발명에 기재된 하나 이상의 돌연변이유발 전략을 사용하여 관심 돌연변이를 생성, 선별 및 통합할 수 있다. 일부 실시태양에서, 본 발명에 기재된 유전자 도구는 유전자 다양성을 생성하기 위해 사용될 수 있다. 예를 들어, 프로모터 스왑 방법, SNP 스왑 방법, 개시/중지 코돈 스왑 방법, 터미네이터 스왑 방법, 트랜스포존 돌연변이유발 방법, 리보솜 결합 부위 방법, 항-대사산물 선택/발효 생성물 저항성 방법 또는 이의 임의의 조합은 유전적 다양성을 생성하는 다른 기회로서 이용될 수 있다.

[0423] 용어 "돌연변이유발"은 또한 세포 기능을 변경(예를 들어, 표적화된 돌연변이에 의해) 또는 세포 기능을 조절하여 돌연변이유발의 속도, 품질 또는 범위를 향상시키는 방법을 포함한다. 예를 들어, 세포는 DNA 수리, 돌연변이원 대사, 돌연변이원 감수성, 게놈 안정성 또는 이의 조합에서 기능장애 또는 결핍이 되도록 변형되거나 조절될 수 있다. 따라서 정상적으로 게놈 안정성을 유지하는 유전자 기능의 파괴는 돌연변이유발을 향상시키는데 사용될 수 있다. 파괴의 대표 표적은 DNA 리가아제 I(Bentley *et al*, 2002) 및 카제인 키나아제 I(미국 특허 제 6,060,296 호)를 포함하나 이에 제한되지 않는다.

[0424] 일부 실시태양에서, 부위-특이적 돌연변이유발(예를 들어, 변형기 부위 유도 돌연변이유발 키트(Clontech)와 같은 상업적으로 구입할 수 있는 키트를 사용한 프라이머-유도 돌연변이유발)를 사용하여 핵산 서열 전체에 걸쳐 순서대로 다수의 변화를 일으킨다 본 발명의 절단 효소를 코딩하는 핵산을 생성한다.

[0425] 하나 이상의 돌연변이유발 인자에 노출시 유전자 돌연변이의 빈도는 치료의 투여량 및/또는 반복을 변화시킴으로써 조절될 수 있으며, 특정 용도에 맞게 조절될 수 있다.

[0426] 따라서, 일부 실시태양에서, 본 발명에 사용된 "돌연변이유발"은 본 발명에 기술된 기술의 임의의 것에 의한 오류 경향성 PCR 돌연변이유발, 올리고뉴클레오티드 유도 돌연변이유발, 부위 유발성 돌연변이유발, 트랜스포존 돌연변이유발 및 반복적인 서열 재조합을 포함하는 돌연변이를 유도하기 기술 분야에 공지된 모든 기술을 포함한다.

[0427] ***다양성을 생성하기 위한 단일 유전자좌 돌연변이***

[0428] 일부 실시태양에서, 본 발명은 게놈 DNA의 선택된 부분을 도입, 결실 또는 치환함으로써 세포 집단을 돌연변이하는 것을 교시한다. 따라서, 일부 실시태양에서, 본 발명은 돌연변이를 특정 유전자좌에 대한 돌연변이를 표적화하는 방법을 교시한다. 다른 실시태양에서, 본 발명은 표적 DNA 영역을 선택적으로 편집하기 위해 ZFNs, TALENS 또는 CRISPR과 같은 유전자 편집 기술의 사용을 교시한다.

[0429] 다른 실시태양에서, 본 발명은 숙주 유기체의 외부에서 선택된 DNA 영역을 돌연변이시킨 다음 돌연변이된 서열을 숙주로 유기체 속에 역삽입하는 것을 교시한다. 예를 들어, 일부 실시태양에서, 본 발명은 다양한 발현 성질을 갖는 다양한 프로모터 변이체를 생성하기 위해 천연 또는 합성 프로모터를 돌연변이하는 것을 교시한다(이하 프로모터 래더 참조). 다른 실시태양에서, 본 발명은 ProSAR(Fox *et al*, 2007. 본 발명에 참조로 포함된 "Improving catalytic function by ProSAR-driven enzyme evolution." Nature Biotechnology Vol 25(3) 338-

343)와 같은 단일 유전자 최적화 기술과 양립된다.

[0430] 일부 실시태양에서, DNA의 선택된 영역은 천연 변이체의 유전자 셔플링 또는 합성 올리고체, 플라스미드-플라스미드 재조합, 바이러스 플라스미드 재조합, 바이러스-바이러스 재조합에 의한 셔플링을 통해 시험관 내에서 생성된다. 다른 실시태양에서, 게놈 영역은 에러-유발PCR(error-prone PCR)(예를 들어, 도 1 참조)을 통해 생성된다.

[0431] 일부 실시태양에서, 선택된 유전자 영역에서의 돌연변이 생성은 "재조립 PCR"에 의해 달성된다. 간략하게, 올리고뉴클레오티드 프라이머(올리고)는 관심 핵산 서열의 단편의 PCR 증폭을 위해 합성되어, 올리고뉴클레오티드의 서열은 두 단편의 교차점과 겹쳐진다. 겹쳐진 영역은 전형적으로 약 10 내지 100개 뉴클레오티드 길이이다. 단편의 각각은 이런 프라이머 세트로 증폭된다. 그런 다음 PCR 제품은 재조립 프로토콜에 따라 "재조립"된다. 요약하면, 어셈블리 프로토콜에서, PCR 생성물은 먼저, 예를 들어, 겔 전기영동 또는 크기 배제 크로마토그래피에 의해 프라이머로부터 정제된다. 정제된 생성물은 함께 혼합되고 추가 프라이머가 없이 중합 효소 및 데옥시뉴클레오시드 트라이포스페이트(dNTP's) 및 적절한 버퍼 염의 존재하에서 약 1-10 사이클의 변성, 재어닐링 및 확장이 실시된다("셀프-프라이밍"). 유전자에 인접한 프라이머로 후속 PCR이 사용되어 완전히 재조합되고 뒤섞인 유전자의 수율을 증폭시킨다.

[0432] 본 발명의 일부 실시태양에서, 위에서 논의한 것과 같은 돌연변이된 DNA 영역이 돌연변이체 서열에 대해 농축되어 다중 돌연변이체 스펙트럼, 즉 돌연변이의 가능한 조합이 보다 효율적으로 표본추출된다. 일부 실시태양에서, 돌연변이된 서열은 재조립 반응 이전에 시험관 내에서 친화성-정제 물질을 증폭시키는 바람직한 단계 의해 mutS 단백질 친화성 매트릭스를 통해 확인된다(Wagner *et al.*, Nucleic Acids Res. 23(19):3944-3948(1995); Su *et al.*, Proc. Natl. Acad. Sci.(U.S.A.), 83:5057-5061(1986)). 이 증폭된 물질은 이후 본 발명의 후반부에서 설명하는 바와 같이 조립 또는 재조립 PCR 반응에 투입된다.

[0433] ***프로모터 래더***

[0434] 프로모터는 유전자가 전사되는 속도를 조절하고 다양한 방식으로 전사에 영향을 미칠 수 있다. 예를 들어, 구조성 프로모터는 내부 또는 외부 세포 조건에 관계없이 일정한 비율로 관련 유전자의 전사를 지시하지만, 조절 가능한 프로모터는 내부 및/또는 외부 세포, 예를 들어, 성장 속도, 온도, 특정 환경 화학물질에 대한 반응 등에 따라 유전자가 전사되는 속도를 증가 또는 감소시킨다. 프로모터는 정상적인 세포 컨텍스트로부터 격리될 수 있으며 사실상 모든 유전자의 발현을 조절하도록 조작되어 세포 성장, 생산 수율 및/또는 기타 관심 표현형의 효과적인 변형을 가능하게 한다.

[0435] 일부 실시태양에서, 본 발명은 다운스트림 유전자 디자인 방법에서 사용하기 위한 프로모터 래더 라이브러리를 생산하는 방법을 교시한다. 예를 들어, 일부 실시태양에서, 본 발명은 하나 이상의 프로모터를 확인하고 및/또는 다양한 발현 강도 또는 우수한 조절 특성을 나타내는 숙주 세포 내 하나 이상의 프로모터의 변이체를 생성하는 방법을 교시한다. 이들 확인된 및/또는 생성된 프로모터의 특정 조합은 프로모터 래더로서 함께 그룹화될 수 있으며, 이는 보다 상세히 후술된다.

[0436] 일부 실시태양에서, 본 발명은 프로모터 래더의 사용을 교시한다. 일부 실시태양에서, 본 발명의 프로모터 래더는 연속 범위의 발현 프로파일을 나타내는 프로모터를 포함한다. 예를 들어, 일부 실시태양에서, 프로모터 래더는 자극, 또는 구성성 발현을 통해 다양한 발현 강도를 나타내는 자연, 고유 또는 야생형 프로모터를 확인함으로써 생성된다(예를 들어, 도 13 및 도 21-23). 이러한 확인된 프로모터는 프로모터 래더로서 함께 그룹화될 수 있다.

[0437] 일부 실시태양에서, 프로모터 래더는 상이한 발현 프로파일을 갖는 적어도 2개의 프로모터를 포함한다. 일부 실시태양에서, 프로모터 래더는 상이한 발현 프로파일을 갖는 적어도 3개의 프로모터를 포함한다. 일부 실시태양에서, 프로모터 래더는 상이한 발현 프로파일을 갖는 적어도 4개의 프로모터를 포함한다. 일부 실시태양에서, 프로모터 래더는 상이한 발현 프로파일을 갖는 적어도 5개의 프로모터를 포함한다. 일부 실시태양에서, 프로모터 래더는 상이한 발현 프로파일을 갖는 적어도 6개의 프로모터를 포함한다. 일부 실시태양에서, 프로모터 래더는 상이한 발현 프로파일을 갖는 적어도 7개의 프로모터를 포함한다.

[0438] 다른 실시태양에서, 본 발명은 상이한 조건에 걸쳐 다양한 발현 프로파일을 나타내는 프로모터 래더의 생성을 교시한다. 예를 들어, 일부 실시태양에서, 본 발명은 발효의 상이한 단계에 전체에서 퍼진 발현 피크를 갖는 프로모터의 래더를 생성하는 것을 교시한다(예를 들어, 도 21 참조). 다른 실시태양에서, 본 발명은 특정 자극에 반응하여 상이한 발현 피크 동역학을 갖는 프로모터의 래더를 생성하는 것을 교시한다(도 22 참조). 당업자는

본 개시의 조절 프로모터 래더가 임의의 하나 이상의 조절 프로파일을 나타낼 수 있다는 것을 인식할 것이다.

[0439] 일부 실시태양에서, 본 발명의 프로모터 래더는 연속적인 범위의 반응에 걸쳐 예측 가능한 방식으로 유전자 발현을 교란시키도록 디자인되었다. 일부 실시태양에서, 프로모터 래더의 연속 특성은 균주 개량 프로그램에 추가의 예측력을 제공한다. 예를 들어, 일부 실시태양에서, 선택된 대사 경로의 스와핑 프로모터 또는 종결 서열은 가장 최적의 발현 비율 또는 프로파일을 확인하는 숙주 세포 성능 곡선을 생성할 수 있다; 부적절한 환경에서 불필요한 과발현이나 잘못된 발현을 피하면서 표적 유전자가 더 이상 특정 반응이나 유전자 캐스케이드에 대한 제한 요소가 되지 않는 균주를 생산한다. 일부 실시태양에서, 프로모터 래더는 원하는 프로파일을 나타내는 자연, 고유 또는 야생형 프로모터를 확인함으로써 생성된다. 다른 실시태양에서, 프로모터 래더는 자연 발생 프로모터를 돌연변이시켜 다수의 돌연변이 프로모터 서열을 유도함으로써 생성된다. 이들 돌연변이된 프로모터의 각각은 표적 유전자 발현에 대한 효과에 대해 시험된다. 일부 실시태양에서, 편집된 프로모터는 다양한 조건에 걸쳐 발현 활성에 대해 시험되어 각 프로모터 변이체의 활성이 문서화/특징화/주석처리되고 데이터베이스에 저장된다. 생성된 편집된 프로모터 변이체는 뒤이어 그 발현의 강도에 기초하여 배열된 프로모터 래더로 조직화된다(예를 들어, 상부 근처의 고도로 발현된 변이체 및 하부 근처의 약화된 발현에 의해, 따라서 "래더"라는 용어를 유도한다).

[0440] 일부 실시태양에서, 본 발명은 확인된 자연 발생 프로모터 및 돌연변이된 변이체 프로모터의 조합인 프로모터 래더를 교시한다.

[0441] 일부 실시태양에서, 본 발명은 하기 기준을 모두 만족하는: 1) 구조성 프로모터의 래더를 나타내며; 2) 이상적으로 100개 염기쌍 미만의 짧은 DNA 서열에 의해 암호화될 수 있는 자연, 고유 또는 야생형 프로모터를 확인하는 방법을 교시한다. 일부 실시태양에서, 본 발명의 구조성 프로모터는 두 개의 선택된 성장 조건(전형적으로, 산업용 재배 동안 경험된 조건들 간에 비교됨)에 걸쳐 일정한 유전자 발현을 나타낸다. 일부 실시태양에서, 본 발명의 프로모터는 ~20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000 또는 그 이상의 염기쌍 코어 프로모터로 이루어질 것이다. 일부 실시태양에서, 5' UTR이 존재한다. 일부 실시태양에서, 5' UTR은 약 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 또는 그 이상의 염기쌍 사이의 길이이다.

[0442] 일부 실시태양에서, 상기 확인된 자연 발생 프로모터 서열 중 하나 이상이 유전자 편집을 위해 선택된다. 일부 실시태양에서, 자연 프로모터는 상기 임의의 돌연변이 방법을 통해 편집된다. 다른 실시태양에서, 본 발명의 프로모터는 원하는 서열을 갖는 새로운 프로모터 변이체를 합성함으로써 편집된다.

[0443] 2015년 12월 07일에 출원된 미국 특허 출원 제 62/264,232호, 및 2016년 12월 07일에 출원된 PCT WO2017/100376의 각각의 전체 개시 내용은 모든 목적을 위해 그 전문이 본 발명에 참고로 포함된다

[0444] 본 발명의 프로모터의 비 제한적인 리스트가 하기 표 1에 제공된다.

표 1

[0445] 본 발명의 선택된 프로모터 서열

서열번호	프로모터 명칭
1	P7160
2	P7253
3	P6681
4	P6316
5	P6806
6	P3159
7	P0757
8	P5011
9	P1409
10	P4735
11	P2900
12	P0801
13	P21
14	PA9
15	PA3

16	PB4
17	PB12
18	PB1
19	PC1
20	P72
21	P-C4-1
22	P-A5-19
23	P-C4-14
24	P-D1-7
25	P1
26	P2
27	P3
28	P3v2
29	P4
30	P4v2
31	P5
32	P5v2
33	P6
34	P7
35	P8
36	P9
37	PspnA
38	PspnAv2
39	PspnF
40	PspnG
41	PspnQ
42	PspnQv2
43	P21_mutant
44	P1_core
45	P1(-33)
46	P1+ribswtch
47	P21-P1
48	P1-P21
49	P1765
50	P3747
51	P5078
52	P7419
53	P7156 (P3)
54	P7256
55	P1941
56	P3405 (P8)
57	P3407
58	P2428
59	P0927
60	P0889
61	P0186
62	P3702_v2
63	P7156_v2
64	P7256_v2
65	P1765_v2
66	P7539_v2
67	P7276_v2
68	P0941_v2
69	P0889_v2

[0446] 일부 실시태양에서, 본 발명의 프로모터는 상기 표 1로부터의 프로모터와 적어도 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 82%, 81%, 80%, 79%, 78%, 77%, 76%, 또

는 75% 서열 동일성을 나타낸다.

[0447] ***터미네이터 래더***

[0448] 일부 실시태양에서, 본 발명은 3' 위치에서 하나 이상의 전사 종결 서열을 RNA 암호화 요소의 말단에 제공함으로써 유전적으로 조작된 숙주 균주를 개량시키는 방법을 교시한다. 일부 실시태양에서, 본 발명은 종결 서열의 첨가가 유전적으로 조작된 숙주에서 선택된 유전자의 RNA 전사 효율을 개량시킨다는 것을 교시한다. 다른 실시태양에서, 본 발명은 종결 서열의 첨가가 유전적으로 조작된 숙주에서 선택된 유전자의 RNA 전사 효율을 감소시킨다는 것을 교시한다. 따라서, 일부 실시태양에서, 본 발명의 터미네이터 래더는 다양한 전사 효율을 나타내는 터미네이터 서열(예를 들어, 하나의 약한 터미네이터, 하나의 평균 터미네이터 및 하나의 강한 프로모터)을 나타내는 일련의 터미네이터 서열을 포함한다.

[0449] 전사 종결 서열은 임의의 뉴클레오티드 서열일 수 있으며, 이는 임의의 뉴클레오티드 서열일 수 있으며, 이는 오픈 리딩 프레임을 코딩하는 뉴클레오티드 서열의 전사 다운스트림에 위치될 때, 오픈 리딩 프레임의 전사 종료를 일으킨다. 이런 서열은 당업계에 공지되어 있으며 원핵생물, 진핵생물 또는 파지 기원일 수 있다. 터미네이터 서열의 예로는 PTH-터미네이터, pET-T7 터미네이터, T3-Tφ 터미네이터, pBR322-P4 터미네이터, 수포성 스트라마티스 바이러스 터미네이터, rrnB-T1 터미네이터, rrnC 터미네이터, TTadc 전사 터미네이터 및 Mat α(α 인자) 전사 터미네이터, 고유 α-인자 전사 종결 서열, ADR1 전사 종결 서열, ADH2 전사 종결 서열 및 GAPD 전사 종결 서열과 같은 효모-인식 종결 서열을 포함하나 이에 제한되지 않는다. 전사 종결 서열의 비-제한적인 목록은 http://partsregistry.org/터미네이터s/Catalog에서 구입할 수 있는 iGEM 레지스트리에서 발견될 수 있다.

[0450] 일부 실시태양에서, 전사 종결 서열은 폴리머라제-특이적이거나 비특이적일 수 있지만, 본 실시태양에서 사용하기 위해 선택된 전사 터미네이터는 선택된 프로모터와 '기능적 조합'을 형성해야 하는데, 이는 터미네이터 서열이 프로모터에서 개시되는 RNA 폴리머라제의 유형에 의한 전사를 종결시킬 수 있어야 한다는 것을 의미한다. 예를 들어, 일부 실시태양에서, 본 발명은 진핵생물 RNA pol II 프로모터 및 진핵생물 RNA pol II 터미네이터, T7 프로모터 및 T7 터미네이터, T3 프로모터 및 T3 터미네이터, 효모- 인식 프로모터 및 효모-인식 종결 서열 등은 일반적으로 기능적 조합을 형성할 것임을 교시한다. 사용된 전사 종결 서열의 상동성은 또한 전사가 소정의 프로모터로부터 종결되는 효율에 기초하여 선택될 수 있다. 예를 들어, 이종성 전사 터미네이터 서열은 RNA 암호화 요소의 전사 다운스트림에 제공되어 소정의 프로모터로부터 적어도 60%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 91%, 적어도 92%, 적어도 93%, 적어도 94%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99%의 종결 효율을 달성할 수 있다.

[0451] 일부 실시태양에서, 조작된 발현 구조체로부터의 RNA 전사 효율은 3' 위치에 2개 이상의 헤어핀을 포함하는 2차 구조를 형성하는 핵산 서열을 RNA 암호화 요소의 말단에 제공함으로써 개량될 수 있다. 특정 이론에 얽매이기를 원하지 않으며, 2차 구조는 전사 신장 복합체를 불안정화시키고 폴리머라제가 DNA 주형으로부터 분리되게 하여, 비 기능성 서열의 비생산적인 전사를 최소화하고 목적하는 RNA의 전사를 증가시킨다. 따라서, 2개 이상의 인접한 헤어핀을 포함하는 2차 구조를 형성하는 종료 서열이 제공될 수 있다. 일반적으로, 헤어핀은 자체가 접혀서 암이 단일 가닥 루프에 의해 연결되는 쌍을 이루는 스템 영역을 형성할 수 있는 회문식 뉴클레오티드 서열에 의해 형성될 수 있다. 일부 실시태양에서, 종결 서열은 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 인접한 헤어핀을 포함한다. 일부 실시태양에서, 인접한 헤어핀은 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 또는 15개의 쌍을 이루지 않는 뉴클레오티드에 의해 분리된다. 일부 실시태양에서, 헤어핀 스템은 길이가 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30개 이상의 염기쌍 길이를 포함한다. 특정 실시태양에서, 헤어핀 스템은 12 내지 30개 염기쌍 길이이다. 특정 실시태양에서, 종결 서열은 약 9 내지 25개 염기쌍을 포함하는 스템 영역을 갖는 둘 이상의 중간 크기의 헤어핀을 포함한다. 일부 실시태양에서, 헤어핀은 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 뉴클레오티드의 루프 형성 영역을 포함한다. 일부 실시태양에서, 루프 형성 영역은 4-8개 뉴클레오티드를 포함한다. 특정 이론에 얽매이기를 원하지 않으며, 2차 구조의 안정성은 종단 효율과 관련될 수 있다. 헤어핀의 안정성은 길이, 불일치(mismatches) 또는 이를 함유하는 돌출(bulges)의 수 및 쌍을 이루는 영역의 염기 구성에 의해 결정된다. 구아닌과 시토신 사이의 결합은 3개의 수소 결합을 가지며 2개만 있는 아데닌-티민 결합에 비해 더 안정하다. 헤어핀 형성 회문식 뉴클레오티드 서열의 G/C 함량은 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90% 이상일 수 있다. 일부 실시태양에서, 헤어핀 형성 회문식 뉴클레오티드 서열의 G/C 함량은 적어도 80%이다. 일부 실시태양에서, 종결 서열은 원핵생물, 진핵생물 또는 파지 기원의 하나 이상의 전사 종결 서열로부터 유도된다. 일부 실시태양에서, 일련의 4, 5, 6, 7, 8, 9, 10개 이상의 아데닌(A)을 코딩하는 뉴클레오티드 서열은 종결 서

열에 대해 3'에 제공된다.

[0452] 일부 실시태양에서, 본 발명은 일련의 탠덤 종결 서열의 사용을 교시한다. 일부 실시태양에서, 일련의 2, 3, 4, 5, 6, 7개 이상의 첫 번째 전사 터미네이터 서열은 dsRNA 암호화 요소의 최종 뉴클레오티드에 대해 직접 3' 또는 dsRNA 암호화 요소의 최종 뉴클레오티드에 대해 적어도 1-5, 5-10, 10-15, 15-20, 20-25, 25-30, 30-35, 35-40, 40-45, 45-50, 50-100, 100-150, 150-200 , 200-300, 300-400, 400-500, 500-1,000개 이상의 뉴클레오티드 3'의 거리에 놓일 수 있다. 탠덤 전사 터미네이터 서열 사이의 뉴클레오티드의 수는 변할 수 있으며, 예를 들어, 전사 터미네이터 서열은 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 10-15, 15-20, 20-25, 25-30, 30-35, 35-40, 40-45, 45-50개 이상의 뉴클레오티드에 의해 분리될 수 있다. 일부 실시태양에서, 전사 터미네이터 서열은 구조 예측 알고리즘에 의해 결정된 바와 같이 이들의 예측된 2차 구조에 기초하여 선택될 수 있다. 구조 예측 프로그램은 당업계에 주지되어 있으며, 예를 들어, CLC 메인 워크벤치(Main Workbench)를 포함한다.

[0453] 당업자는 본 발명의 방법이 임의의 종결 서열과 양립할 수 있음을 인식할 것이다. 일부 실시태양에서, 본 발명은 주석된 사카로폴리스포라 종 터미네이터의 사용을 교시한다. 다른 실시태양에서, 본 발명은 http://partsregistry.org/터미네이터s/Catalog에서 구입할 수 있는 iGEM 레지스트리에서 발견된 전사 종결 서열의 사용을 교한다. 본 발명의 전사 터미네이터 서열의 비-제한적인 목록이 하기 표 2에 제공된다.

표 2

[0454] 본 발명의 종결 서열의 비-제한적인 목록.

대장균			
명칭	**설명**	**방향**	**길이**
BBa_B0010	대장균 rrnB로부터의 T1	정방향	80
BBa_B0012	콜리파지T7으로부터의 TE	정방향	41
BBa_B0013	콜리파지 T7으로부터의 TE (+/-)	정방향	47
BBa_B0015	이중 터미네이터 (B0010-B0012)	정방향	129
BBa_B0017	이중 터미네이터 (B0010-B0010)	정방향	168
BBa_B0053	터미네이터 (His)	정방향	72
BBa_B0055	-- 설명 없음 --		78
BBa_B1002	터미네이터 (인공, 소형, %T~=85%)	정방향	34
BBa_B1003	터미네이터 (인공, 소형, %T~=80)	정방향	34
BBa_B1004	터미네이터 (인공, 소형, %T~=55)	정방향	34
BBa_B1005	터미네이터 (인공, 소형, %T~=25%	정방향	34
BBa_B1006	터미네이터 (인공, 대형, %T~>90)	정방향	39
BBa_B1010	터미네이터 (인공, 대형, %T~<10)	정방향	40
BBa_I11013	바이오브릭 부분의 변형 BBa_B0015		129
BBa_I51003	-- 설명 없음 --		110
BBa_J61048	[rnpB-T1] 터미네이터	정방향	113
BBa_K1392970	터미네이터+Tetr 프로모터+T4 엔돌리신		623
BBa_K1486001	아라비노오스 프로모터 + CpxR	정방향	1924
BBa_K1486005	아라비노오스 프로모터 + sfGFP-CpxR [Cterm]	정방향	2668
BBa_K1486009	CxpR & Split IFP1.4 [Nterm + Nterm]	정방향	3726
BBa_K780000	바실러스 서브틸리스를 위한 터미네이터		54
BBa_K864501	T22, P22 후기(late) 터미네이터	정방향	42
BBa_K864600	T0 (21 imm) 전사 터미네이터	정방향	52
BBa_K864601	람다 t1 전사 터미네이터	정방향	
BBa_B0011	LuxICDABEG (+/-)	양방향	46
BBa_B0014	이중 터미네이터 (B0012-B0011)	양방향	95
BBa_B0021	LuxICDABEG (+/-), 역방향d	양방향	46
BBa_B0024	이중 터미네이터 (B0012-B0011), 역방향d	양방향	95
BBa_B0050	터미네이터 (pBR322, +/-)	양방향	33
BBa_B0051	터미네이터 (yciA/tonA, +/-)	양방향	35
BBa_B1001	터미네이터 (인공, 소형, %T~=90)	양방향	34
BBa_B1007	터미네이터 (인공, 대형, %T~=80)	양방향	40
BBa_B1008	터미네이터 (인공, 대형, %T~=70)	양방향	40
BBa_B1009	터미네이터 (인공, 대형, %T~=40%)	양방향	40

BBa_K187025	pAB의 터미네이터, BioBytes 플라스미드		60
BBa_K259006	GFP-터미네이터	양방향	823
BBa_B0020	터미네이터 (역방향 B0010)	역방향	82
BBa_B0022	콜리파지T7으로부터의 TE, 역방향	역방향	41
BBa_B0023	콜리파지 T7으로부터의 TE, 역방향	역방향	47
BBa_B0025	이중 터미네이터 (B0015), 역방향	역방향	129
BBa_B0052	터미네이터 (rrnC)	정방향	41
BBa_B0060	터미네이터 (역방향 B0050)	양방향	33
BBa_B0061	터미네이터 (역방향 B0051)	양방향	35
BBa_B0063	터미네이터 (역방향 B0053)	역방향	72
효모 및 다른 진핵생물			
명칭	**설명**	**방향**	**길이**
BBa_J63002	S. 세레비시애로부터의 ADH1 터미네이터	정방향	225
BBa_K110012	STE2 터미네이터	정방향	123
BBa_K1462070	cyc1		250
BBa_K1486025	ADH1 터미네이터	정방향	188
BBa_K392003	효모 ADH1 터미네이터		129
BBa_K801011	TEF1 효모 터미네이터		507
BBa_K801012	ADH1 효모 터미네이터		349
BBa_Y1015	CycE1		252
BBa_J52016	진핵생물 -- SV40 초기 폴리 A 신호 서열로부터 유도됨	정방향	238
BBa_J63002	S. 세레비시애로부터의 ADH1 터미네이터	정방향	225
BBa_K110012	STE2 터미네이터	정방향	123
BBa_K1159307	콜리플라워 모자이크 바이러스(CaMV)의 35S 터미네이터		217
BBa_K1462070	cyc1		250
BBa_K1484215	노팔린 합성효소 터미네이터		293
BBa_K1486025	ADH1 터미네이터	정방향	188
BBa_K392003	효모 ADH1 터미네이터		129
BBa_K404108	hGH 터미네이터		481
BBa_K404116	hGH_[AAV2]-right-ITR		632
BBa_K678012	SV40 poly A, 포유류 세포를 위한 터미네이터		139
BBa_K678018	hGH poly A, 포유류 세포를 위한 터미네이터		635
BBa_K678019	BGH poly A, 포유류 터미네이터		233
BBa_K678036	아스퍼길루스 니둘란스를 위한 trpC 터미네이터		759
BBa_K678037	T1-motni, 아스퍼길루스 니거를 위한 터미네이터		1006
BBa_K678038	T2-motni, 아스퍼길루스 니거를 위한 터미네이터		990
BBa_K678039	T3-motni, 아스퍼길루스 니거를 위한 터미네이터		889
BBa_K801011	TEF1 효모 터미네이터		507
BBa_K801012	ADH1 효모 터미네이터		349
BBa_Y1015	CycE1		252

[0455] 본 발명의 추가의 터미네이터 서열의 비-제한적 목록이 하기 표 3에 제공된다. 터미네이터 서열 각각은 이종 터미네이터 또는 이종 터미네이터 폴리뉴클레오티드로 지칭될 수 있다.

표 3

[0456] 본 발명의 터미네이터 서열.

ID	관련된 유전자	서열	소스	크기 (bp)
T1	(신장 인자 tu)	CCCGAACCTTCGGGGGCGGGCCCTCTTGCTTTTCAAT (서열번호 70)	*S. 스피노사*	37
T2	(류실 아미노펩티다아제)	CGGGCAATAATACGTGCCCGGACGGTAGTGCGAGCACGAGGTGGGTACG (서열번호 71)	*S. 스피노사*	49

T3	(시토크롬 P450 수산화효소)	AGTTTGTCGAACCGGCGGCGTTCGCCGGcTTTACCTTGCGC (서열번호 72)	*S. 스피노사*	41
T4	(F0F1 ATP 결합효소 서브유닛 베타)	GGTTTCTCGAACCAGTGCTTTGCGTACTGGTTGTCGTTGCAG (서열번호 73)	*S. 스피노사*	42
T5	(FAD-결합 산화환원효소)	CGGAGCCAGAGGGCGCCTGAGTGCCTGTTTTTGATCC (서열번호 74)	*S. 에리트라에아*	37
T6	(포스포리보실기 전이효소)	AAACGCCCCCGGCTCCGGCCGGGGGCgTTTTTGGTTGTG (서열번호 75)	*S. 에리트라에아*	39
T7	(ATP-결합 단백질)	AGACGCAGGAGGTCTCGTGAGGGGCTTTTCCGCGAGC (서열번호 76)	*S. 에리트라에아*	37
T8	SACE_0757 (50s 리보솜 단백질 L32)	CGTGTGACTTGTCCCACTCGGGGTTTTTGTCGCGA (서열번호 77)	*S. 에리트라에아*	35
T9	(tRNA-Arg)	GGATTCGTCCGGCCGAGGCCAATCGGCTTTTCGGGGCCC (서열번호 78)	*S. 에리트라에아*	39
T11	(lsr2)	GCTTTCGTCGGCCGGGAACGCCCTGGTGTTTCTTACCG (서열번호79)	*S. 에리트라에아*	38
T12	(AraC)	TTGGGTGGATTCACCCCTACCGGGTGTTTTTCTCGGCT (서열번호 80)	*S. 에리트라에아*	38
NoT	없음	-	-	0

[0457] 일부 실시태양에서, 본 발명의 터미네이터는 상기 표 3의 터미네이터와 적어도 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 82%, 81%, 80%, 79%, 78%, 77%, 76% 또는 75% 서열 동일성을 나타낸다.

[0458] ***가설-주도 다양성 풀과 언덕 오르기(Hill Climbing)***

[0459] 본 발명은 본 개시의 HTP 게놈 조작 방법은 숙주 세포 성능에서 상당한 이득을 달성하기 위해 사전 유전자 지식을 필요로 하지 않는다는 것을 교시한다. 실제로, 본 발명은 랜덤 돌연변이유발 및 이미 존재하는 숙주 세포 변이체들 사이의 유전자 다양성의 확인(예를 들어, 야생형 숙주 세포와 산업 변이체 사이의 비교)을 포함하는 여러 기능적으로 불가지론적인 접근법을 통해 다양성 풀을 생성하는 방법을 교시한다.

[0460] 그러나, 일부 실시태양에서, 본 발명은 또한 다운스트림 HTP 조작에 사용될 유전자 다양성 돌연변이를 디자인하는 가설-주도적 방법을 교시한다. 즉, 일부 실시태양에서, 본 발명은 선택된 돌연변이의 지시된 디자인을 교시한다. 일부 실시태양에서, 지시된 돌연변이는 본 발명의 조작 라이브러리(예컨대, SNP 스왑, PRO 스왑, STOP 스왑, 트랜스포존 돌연변이유발 다양성 라이브러리, 리보솜 결합 부위 다양성 라이브러리, 항-대사산물 선택/발효생성물 저항성 라이브러리)에 포함된다.

[0461] 일부 실시태양에서, 본 발명은 유전자 주석, 가설된(또는 확인된) 유전자 기능, 또는 게놈 내의 위치에 기초한 지시된 돌연변이의 생성을 교시한다. 본 발명의 다양성 풀은 숙주 세포의 성능이 증가된 문헌에 관련된 특정 대사 경로 또는 유전자 경로에 관여한다고 가설된 유전자의 돌연변이를 포함할 수 있다. 다른 실시태양에서, 본 발명의 다양성 풀은 또한 개량된 숙주 성능과 관련된 오페론에 존재하는 유전자에 대한 돌연변이를 포함할 수 있다. 또 다른 실시태양에서, 본 발명의 다양성 풀은 알고리즘 예측 기능 또는 다른 유전자 주석에 기초한 유전자에 대한 돌연변이를 포함할 수 있다.

[0462] 일부 실시태양에서, 본 발명은 가설 주도 돌연변이의 표적을 우선순위화하기위한 "껍질"에 기초한 접근법을 교시한다. 표적 우선순위화를 위한 껍질 비유는 단지 소수의 1차 유전자가 숙주 세포의 성능(예를 들어, 단일 생체분자의 생산)의 특정 측면의 대부분을 담당한다는 가설에 기초한다. 이들 1차 유전자는 껍질의 코어에 위치되고, 제 2 층에서 2차 효과 유전자, 3차 껍질에서 3차 효과 및 ... 등이 이어진다. 예를 들어, 한 실시태양에서, 껍질의 코어는 선택된 대사 경로 내의 중요한 생합성 효소를 코딩하는 유전자를 포함할 수 있다. 제 2 껍질 상에 위치하는 유전자는 생성물 전환 또는 피드백 신호 전달을 담당하는 생합성 경로 내의 다른 효소를 코딩하는 유전자를 포함할 수 있다. 이러한 예시적인 은유하에서의 제 3층 유전자는 생합성 경로의 발현을 조절하거나 숙주 세포 내에서 일반적인 탄소 플럭스를 조절하는 조절 유전자를 포함할 것이다.

[0463] 본 발명은 또한 모든 확인된 돌연변이로부터 성능 이득을 최적화하기 위한 "언덕 오르기(hill climb)" 방법을 교시한다. 일부 실시태양에서, 본 발명은 HTP 다양성 라이브러리에서 랜덤, 자연 또는 가설-주도 돌연변이는 숙주 세포 성능과 관련된 유전자의 확인을 초래할 수 있음을 교시한다. 예를 들어, 본 방법은 유전자 암호화 서열

상에 또는 그 부근에 위치한 하나 이상의 유익한 SNP를 확인할 수 있다. 이 유전자는 숙주 세포 성능과 관련이 있을 수 있으며, 이의 확인은 유기체의 조합 유전자 돌연변이 공간에서 성능 "언덕"의 발견과 유사할 수 있다.

[0464] 일부 실시태양에서, 본 발명은 SNP 돌연변이로 구현된 확인된 언덕 주변의 조합 공간을 탐색하는 방법을 교시한다. 즉, 일부 실시태양에서, 본 발명은 그 유전자 노드로부터 얻어진 성능 이득(즉, 언덕 오르기)을 최적화하기 위해 확인된 유전자 및 관련 조절 서열의 교란을 교시한다. 따라서, 본 발명의 방법에 따르면, 유전자는 랜덤 돌연변이유발에서 유래된 다양성 라이브러리에서 먼저 확인될 수 있지만, 나중에 동일한 유전자 내 다른 서열의 지시된 돌연변이를 통해 균주 개량 프로그램에서 사용하기 위해 개량될 수 있다.

[0465] 언덕 오르기의 개념은 또한 단일 유전자 서열을 둘러싸는 조합 공간의 탐색을 넘어 확장될 수 있다. 일부 실시태양에서, 특정 유전자의 돌연변이는 숙주 세포 성능에 대한 특정 대사 경로 또는 유전자 경로의 중요성을 나타낼 수 있다. 예를 들어, 일부 실시태양에서, 단일 RNA 분해 유전자에서 돌연변이가 현저한 숙주 성능 증가를 초래한다는 발견은 숙주 유기체로부터 추가적인 성능 이득을 추출하기 위한 수단으로서 관련 RNA 분해 유전자를 돌연변이시키는 기초로서 사용될 수 있다. 당업자는 지시된 유전자 디자인에 대한 상기 껍질 및 언덕 오르기 접근법의 변형을 인식할 것이다. 고 처리량 선별.

[0466] ***세포 배양 및 발효***

[0467] 본 발명의 세포는 임의의 바람직한 생합성 반응 또는 선택을 위해 적절하게 변형된 통상적인 영양 배지에서 배양될 수 있다. 일부 실시태양에서, 본 발명은 프로모터를 활성화시키기 위해 배지를 유도하는 배양을 교시한다. 일부 실시태양에서, 본 발명은 형질전환체(예를 들어, 항생제)의 선택 제제 또는 억제 조건(예를 들어, 높은 에탄올 조건)하에 성장하기에 적합한 유기체의 선택 제제를 포함하는 선택 제제를 갖는 배지를 교시한다. 일부 실시태양에서, 본 발명은 세포 성장을 위해 최적화된 배지에서 세포 배양물을 성장시키는 것을 교시한다. 다른 실시태양에서, 본 발명은 생산 수율에 최적화된 배지에서 세포 배양물을 성장시키는 것을 교시한다. 일부 실시태양에서, 본 발명은 세포 성장을 유도할 수 있는 배지에서 배양 물을 배양하는 것을 교시하며 최종 생성물 생산을 위한 필수 전구체(예를 들어, 에탄올 생산을 위한 높은 수준의 당)를 포함한다.

[0468] 온도, pH 등과 같은 배양 조건은 발현을 위해 선택된 숙주 세포와 함께 사용하기에 적합한 것들이며, 당업자에게 명백할 것이다. 언급한 바와 같이, 박테리아, 식물, 동물(포유류 포함) 및 고세균 기원 세포를 포함하는 많은 세포의 배양 및 생산에 대한 많은 참고 자료가 이용될 수 있다. 예를 들어, Sambrook, Ausubel(all supra), as well as Berger, *Guide to Molecular Cloning Techniques, Methods in Enzymology* volume 152 Academic Press, Inc., San Diego, CA; and Freshney(1994) *Culture of Animal Cells, a Manual of Basic Technique*, third edition, Wiley-Liss, New York and the references cited therein; Doyle and Griffiths(1997) *Mammalian Cell Culture: Essential Techniques* John Wiley and Sons, NY; Humason(1979) *Animal Tissue Techniques*, fourth edition W.H. Freeman and Company; and Ricciardelle *et al.*,(1989) *In Vitro Cell Dev. Biol.* 25:1016-1024, all of which are incorporated herein by reference. For plant cell culture and regeneration, Payne *et al.*(1992) *Plant Cell* and *Tissue Culture in Liquid Systems* John Wiley & Sons, Inc. New York, N.Y.; Gamborg and Phillips(eds)(1995) *Plant Cell, Tissue* and *Organ Culture*; Fundamental Methods Springer Lab Manual, Springer-Verlag(Berlin Heidelberg N.Y.); Jones, ed.(1984) *Plant Gene Transfer and Expression Protocols*, Humana Press, Totowa, N.J. and *Plant Molecular Biology* (1993) R. R. D. Croy, Ed. Bios Scientific Publishers, Oxford, U.K. ISBN 0 12 198370 6을 참조하고, 이의 전문은 참조로 본 발명에 포함된다. 일반적으로 세포 배양 배지는 Atlas and Parks(eds.) *The Handbook of Microbiological Media* (1993) CRC Press, Boca Raton, Fla에서 설명되며, 이는 참조로 본 발명에 포함된다. 세포 배양에 대한 추가 정보는 *Life Science Research Cell Culture Catalogue* from Sigma-Aldrich, Inc(St Louis, Mo.)("Sigma-LSRCCC") 및, 예를 들어, *The Plant Culture Catalogue* and supplement also from Sigma-Aldrich, Inc(St Louis, Mo.)("Sigma-PCCS")와 같은 구입가능한 상업용 논문에서 발견되며, 이의 전부는 참조로 본 발명에 포함된다.

[0469] 사용될 배양 배지는 적절한 방식으로 각각의 균주의 요구를 충족시켜야한다. 다양한 유기체에 대한 배양 배지에 대한 설명은 American Bacteriology Society(Washington D.C., USA, 1981)의 "General Bacteriology Methods for Manual"에 제공된다.

[0470] 또한, 본 발명은 a) 적절한 배지에서 본 발명에 따른 미생물을 배양하여 발효액을 생성하는 단계; b) a)의 발효액 및/또는 미생물의 세포에 관심 생성물을 농축시키는 단계를 포함하여 관심 생성물의 발효성 제조를 위한 방법을 추가로 제공한다.

[0471] 일부 실시태양에서, 본 발명은 생산된 미생물이 원하는 유기-화학적 화합물을 생산하기 위한, 예를 들어 WO 05/021772에 기술된 바와 같이 연속적으로 또는 배치 공정(배치 배양) 또는 페드 배치(fed-batch) 또는 반복된 페드 배치에서 불연속적으로 배양될 수 있음을 교시한다. 알려진 배양 방법에 관한 일반적인 내용은 Chmiel(Bioprozeβtechnik. 1: Einfuhrung in die Bioverfahrenstechnik(Gustav Fischer Verlag, Stuttgart, 1991))에 의한 교과서 또는 Storhas(Bioreaktoren 및 periphere Einrichtungen(Vieweg Verlag, Braunschweig/Wiesbaden, 1994))에 의한 교과서에서 이용할 수 있다.

[0472] 일부 실시태양에서, 본 발명의 세포는 배치 또는 연속 발효 조건하에 성장된다.

[0473] 고전적인 배치 발효는 폐쇄된 시스템이며, 배지의 조성물은 발효의 시작시에 설정되고 발효 동안 인공적인 변형을 겪지 않는다. 배치 시스템의 변형은 본 발명에서 사용되는 페드 배치 발효이다. 이 변형에서, 기질은 발효가 진행됨에 따라 증분으로 첨가된다. 페드 배치 시스템은 이화생성물 억제가 세포의 신진대사를 억제할 가능성이 있고 배지에 제한된 양의 기질을 갖는 것이 바람직한 경우에 유용하다. 배치 및 페드 배치 발효는 일반적이고 당업계에 일반적으로 주지되어 있다.

[0474] 연속 발효는 정의된 발효 배지가 생물 반응기에 연속적으로 첨가되고 동일한 양의 조건 배지가 관심의 목적하는 생체분자 생성물의 가공 및 수확을 위해 동시에 제거되는 시스템이다. 일부 실시태양에서, 연속 발효는 일반적으로 세포가 주로 대수기 성장하는 일정한 고밀도로 배양물을 유지시킨다. 일부 실시태양에서, 연속 발효는 일반적으로 정지기 또는 늦은 대수기/정지기 성장시에 배양물을 유지시킨다. 연속 발효 시스템은 정지기 성장 조건을 유지하기 위해 노력한다.

[0475] 연속 발효 공정을 위한 영양소 및 성장 인자를 조절하는 방법뿐만 아니라 생성물 형성의 속도를 최대화하기 위한 기술은 산업용 미생물학 분야에 잘 알려져 있다.

[0476] 예를 들어, 본 발명의 배양물에 대한 탄소원의 비 제한적인 목록은 당 및 수화물, 예를 들어 글루코오스, 수크로오스, 락토오스, 프룩토오스, 말토오스, 당밀, 사탕무 또는 사탕수수 가공에 의한 수크로오스-함유 용액, 전분, 전분 가수분해물 및 셀룰로오스; 오일 및 지방, 예를 들어, 대두유, 해바라기유, 땅콩유 및 코코넛 지방; 지방산, 예를 들어 팔미트산, 스테아르산 및 리놀레산; 알코올, 예를 들어 글리세롤, 메탄올 및 에탄올; 및 유기산, 예를 들어 아세트산 또는 젖산을 포함한다.

[0477] 본 발명의 배양물을 위한 질소 공급원의 비 제한적인 목록은 펩톤, 효모 추출물, 고기 추출물, 맥아 추출물, 옥수수 침지액, 콩가루 및 우레아와 같은 유기 질소 함유 화합물; 또는 황산 암모늄, 염화 암모늄, 인산 암모늄, 탄산 암모늄 및 질산 암모늄과 같은 무기 화합물을 포함한다. 질소원은 개별적으로 또는 혼합물로서 사용될 수 있다.

[0478] 본 발명의 배양물을 위한 가능한 인 공급원의 비 제한적인 목록은 인산, 인산 이수소 칼륨 또는 인산 수소 이칼륨 또는 상응하는 나트륨 함유 염을 포함한다.

[0479] 배양 배지는 추가로 성장에 필수적인, 나트륨, 칼륨, 마그네슘, 칼슘 및 철과 같은 금속의 염화물 또는 황산염 형태의 염을 포함한다.

[0480] 마지막으로, 아미노산, 예를 들어 호모 세린 및 비타민, 예를 들어 티아민, 바이오틴 또는 판토텐산과 같은 필수 성장 인자가 상기 언급된 물질에 추가로 사용될 수 있다.

[0481] 일부 실시태양에서, 배양물의 pH는 수산화 나트륨, 수산화 칼륨, 암모니아 또는 수성 암모니아; 또는 인산 또는 황산과 같은 산성 화합물을 포함하나 이에 제한되지 않는 임의의 산 또는 염기, 또는 완충 염에 의해 적절한 방식으로 조절될 수 있다. 일부 실시태양에서, pH는 일반적으로 6.0 내지 8.5, 바람직하게는 6.5 내지 8의 값으로 조절된다.

[0482] 일부 실시태양에서, 본 발명의 배양물은 소포제, 예를 들어 지방산 폴리글리콜 에스터를 포함할 수 있다. 일부 실시태양에서, 본 발명의 배양물은 예를 들어 항생제와 같은 적합한 선택 물질을 첨가함으로써 배양물의 플라스미드를 안정화시키도록 변형된다.

[0483] 일부 실시태양에서, 배양은 호기성 조건하에 수행된다. 이러한 조건을 유지하기 위해, 예를 들어 공기와 같은 산소 또는 산소 함유 기체 혼합물이 배양물 내로 도입된다. 과산화수소가 풍부한 액체를 사용하는 것도 가능하다. 발효는, 적절한 경우에, 상승된 압력, 예를 들어 0.03 내지 0.2 MPa의 상승된 압력에서 수행된다. 배양의 온도는 통상 20℃ 내지 45℃, 바람직하게는 25℃ 내지 40℃, 특히 바람직하게는 30℃ 내지 37℃이다. 배치 또는 페드 배치 공정에서, 배양은 바람직하게는 회수하고자 하는 관심의 목적하는 생성물(예를 들어, 유기-화학적 화

합물)의 양이 형성될 때까지 계속된다. 이 목표는 일반적으로 10시간 내지 160시간 이내에 달성될 수 있다. 연속 공정에서, 더 긴 배양시간이 가능하다. 미생물의 활성은 발효 배지 및/또는 상기 미생물의 세포에서 관심 생성물의 농축(축적)을 초래한다.

[0484] 일부 실시태양에서, 배양은 혐기성 조건하에 수행된다.

[0485] ***선별***

[0486] 일부 실시태양에서, 본 발명은 고 처리량 초기 선별을 교시한다. 다른 실시태양에서, 본 발명은 또한 성능 데이터의 견고한 탱크-기반 검증을 교시한다(도 6b 참조).

[0487] 일부 실시태양에서, 고 처리량 선택 공정은 생물 반응기에서의 균주의 성능을 예측하도록 디자인된다. 상기한 바와 같이, 배양 조건은 생물체에 적합하고 생물 반응기 조건을 반영하도록 선택된다. 개별 콜로니를 집어 96웰 플레이트로 옮기고 적당한시간 동안 배양한다. 이어서 세포를 종자 배양 또는 생산 배양을 위해 새로운 96웰 플레이트로 옮긴다. 배양액은 다양한시간 동안 배양되어, 다중 측정이 실행될 수 있다. 다중 측정은 생성물, 바이매스 또는 생물 반응기에서 균주의 성능을 예측하는 다른 특성의 측정을 포함할 수 있다. 고 처리량의 배양 결과는 생물 반응기 성능을 예측하는데 사용된다.

[0488] 일부 실시태양에서, 탱크-기반 성능 검증은 고 처리량 선별에 의해 분리된 균주의 성능을 확인하기 위해 사용된다. 생산성 또는 수율과 같은 관련 균주 성능 특성에 대한 벤치 스케일 발효 반응기를 사용하여 후보 균주를 선별한다.

[0489] ***생성물 회수 및 정량화***

[0490] 관심 생성물의 생산을 위한 선택 방법은 당업자에게 공지되어 있으며 본 명세서 전반에 걸쳐 논의된다. 이러한 방법은 본 발명의 균주를 선별할 때 사용될 수있다.

[0491] 일부 실시태양에서, 본 발명은 비 분비된 세포내 생성물을 생산하도록 디자인된 균주를 개량시키는 방법을 교시한다. 예를 들어, 본 발명은 세포내 효소, 오일, 약제 또는 다른 가치있는 작은 분자 또는 펩타이드를 생성하는 세포 배양물의 견고성, 수율, 효율 또는 전반적인 바람직함을 개량시키는 방법을 교시한다. 비 분비된 세포내 생성물의 회수 또는 분리는 용해 및 본 발명에 기술된 것을 포함하여 당해 분야에 주지된 회수 기술에 의해 달성될 수 있다.

[0492] 예를 들어, 일부 실시태양에서, 본 발명의 세포는 원심분리, 여과, 침전 또는 다른 방법에 의해 수확될 수 있다. 이어서, 수확된 세포는 동결-해동 사이클링, 초음파 처리, 기계적 파쇄 또는 세포 용해제의 사용, 또는 당업자에게 공지된 다른 방법을 포함하는 임의의 편리한 방법으로 파괴된다.

[0493] 관심의 생성된 생성물, 예를 들면, 폴리펩타이드는 당업계에 공지된 다수의 방법 중 임의의 방법에 의해 회수/분리될 수 있으며 선택적으로 정제될 수 있다. 예를 들어, 생성물 폴리 펩타이드는 원심분리, 여과, 추출, 분무 건조, 증발, 크로마토 그래피(예를 들어, 이온 교환, 친화성, 소수성 상호작용, 크로마토포커싱 및 크기 배제) 또는 침전을 포함하나 이에 제한되지 않는 통상적인 절차에 의해 영양 배지로부터 분리될 수 있다. 최종적으로, 고성능 액체 크로마토 그래피(HPLC)가 최종 정제 단계에서 사용될 수 있다.(예를 들어, Purification of intracellular protein as described in Parry *et al.*, 2001, *Biochem. J.*353:117, and Hong *et al.*, 2007, *Appl. Microbiol. Biotechnol.* 73:1331, 참조, 모두는 참조로 본 발명에 포함된다).

[0494] 상기한 참조문헌 이외에, 다양한 정제 방법은, 예를 들어, Sandana (1997) *Bioseparation of Proteins*, Academic Press, Inc.; Bollag *et al.* (1996) *Protein Methods*, 2nd Edition, Wiley-Liss, NY; Walker (1996) *The Protein Protocols Handbook*Humana Press, NJ; Harris and Angal (1990) *Protein Purification Applications: A Practical Approach*, IRL Press at Oxford, Oxford, England; Harris and Angal *Protein Purification Methods: A Practical Approach*, IRL Press at Oxford, Oxford, England; Scopes (1993) *Protein Purification: Principles and Practice* 3rd Edition, Springer Verlag, NY; Janson and Ryden (1998) *Protein Purification: Principles, High Resolution Methods and Applications*, Second Edition, Wiley-VCH, NY; and Walker (1998) *Protein Protocols on CD-ROM*, Humana Press, NJ에 개시된 것들을 포함하여 당업계에 주지되어 있으며, 이의 전부는 본 발명에 참조로 포함된다.

[0495] 일부 실시태양에서, 본 발명은 분비된 생성물을 생산하도록 디자인된 균주를 개량시키는 방법을 교시한다. 예를 들어, 본 발명은 귀중한 작은 분자 또는 펩타이드를 생산하는 세포 배양물의 견고성, 수율, 효율 또는 전반적인

바람직함을 개량시키는 방법을 교시한다.

[0496] 일부 실시태양에서, 면역학적 방법을 사용하여 본 발명의 세포에 의해 생산된 분비된 또는 분비되지 않은 생성물을 검출 및/또는 정제하는데 사용될 수 있다. 한 예시적 접근법에서, 통상적인 방법을 사용하여 생성물 분자(예를 들어, 인슐린 폴리펩타이드 또는 이의 면역원성 단편에 대해)에 대해 생성된 항체는 비드 상에 고정화되고, 엔도글루카나아제가 결합되고 침전되는 조건하에 세포 배양 배지와 혼합된다. 일부 실시태양에서, 본 발명은 효소-결합 면역흡착 분석법(ELISA)의 사용을 교시한다.

[0497] 다른 관련 실시태양에서, 면역크로마토그래피가 미국 특허 제5,002,303호, 미국 특허 제5,591,645호, 미국 특허 제4,855,240호, 미국 특허 제4,435,504호, 미국 특허 제4,980,298호 및 Se-Hwan Paek, *et al.*, "Development of rapid One-Step Immunochromatographic assay, Methods", 22, 53-60, 2000에 개시된 바와 같이 사용되며, 이의 각각은 참조로 본 발명에 포함된다. 일반적인 면역크로마토그래피는 2개의 항체를 사용하여 표본을 검출한다. 제 1 항체는 시험 용액 중에 또는 시험 용액을 떨어뜨린 다공성 막으로 제조된 대략 직사각형 형태의 시험편의 말단 부분에 존재한다. 이 항체는 라텍스 입자 또는 금 콜로이드 입자(이 항체는 이하에서 표지 항체로 불릴 것이다)로 표지한다. 떨어뜨린 시험 용액이 검출될 표본을 포함하는 경우, 표지된 항체는 표본을 인식하여 표본과 결합한다. 표본과 표지된 항체의 복합체는 모세관 현상에 의해 여과지로 만들어지고 표지된 항체를 포함하는 말단의 반대편 말단에 부착된 흡수체를 향하여 흐른다. 이 흐름 동안, 표본과 표지 항체의 복합체가 다공질 막의 중간에 존재하는 제 2 항체(이하에서 태핑 항체로 불릴 것이다)에 의해 인식되어 포획되고, 그 결과, 복합체는 가시광 신호로서 다공성 막 상의 검출부 상에 나타내고 검출된다.

[0498] 일부 실시태양에서, 본 발명의 선택 방법은 측광 검출 기술(흡수, 형광)에 기초한다. 예를 들어, 일부 실시태양에서, 검출은 항체에 결합된 GFP와 같은 형광 단 검출기의 존재에 기초할 수 있다. 다른 실시태양에서, 측광 검출은 세포 배양 물로부터의 원하는 생성물 상의 축적에 기초할 수 있다. 일부 실시태양에서, 생성물은 배양물의 UV 또는 상기 배양물로부터의 추출물을 통해 검출될 수 있다.

[0499] 당업자는 본 발명의 방법이 관심의 임의의 바람직한 생체분자 생성물을 생산하는 숙주 세포와 양립할 수 있음을 인식할 것이다. 하기 표 4는 본 발명의 범위 내에 포함되는 생성물 카테고리, 생체분자 및 숙주 세포의 비 제한적 목록을 제공한다. 이들 실시예는 설명의 목적으로 제공되며, 어떠한 방식으로도 본 개시된 기술의 응용가능성을 제한하려는 것이 아니다.

표 4

[0500] 본 발명의 숙주 세포 및 관심 생성물의 비 제한적 목록.

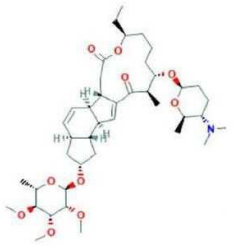
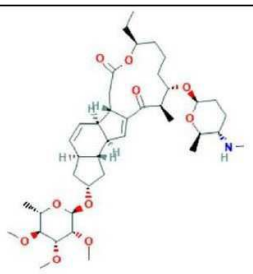
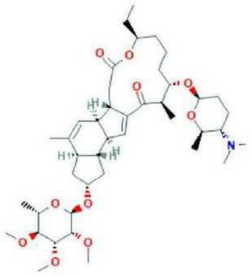
생성물 카테고리	**생성물**	**숙주 카테고리**	**숙주**
아미노산	리신	박테리아	*Corynebacterium glutamicum*
아미노산	메티오닌	박테리아	*Escherichia coli*
아미노산	MSG	박테리아	*Corynebacterium glutamicum*
아미노산	트레오닌	박테리아	*Escherichia coli*
아미노산	트레오닌	박테리아	*Corynebacterium glutamicum*
아미노산	트립토판	박테리아	*Corynebacterium glutamicum*
효소	효소(11)	사상 진균	*Trichoderma reesei*
효소	효소(11)	진균	*Myceliopthora thermophila(C1)*
효소	효소(11)	사상 진균	*Aspergillus oryzae*
효소	효소(11)	사상 진균	*Aspergillus niger*
효소	효소(11)	박테리아	*Bacillus subtilis*
효소	효소(11)	박테리아	*Bacillus licheniformis*
효소	효소(11)	박테리아	*Bacillus clausii*
향신료 & 향료	아가우드	효모	*Saccharomyces cerevisiae*
향신료 & 향료	암브록스	효모	*Saccharomyces cerevisiae*
향신료 & 향료	누트카톤	효모	*Saccharomyces cerevisiae*
향신료 & 향료	파출리 오일	효모	*Saccharomyces cerevisiae*
향신료 & 향료	사프론	효모	*Saccharomyces cerevisiae*
향신료 & 향료	샌달우드 오일	효모	*Saccharomyces cerevisiae*

향신료 & 향료	발렌센	효모	*Saccharomyces cerevisiae*
향신료 & 향료	바닐린	효모	*Saccharomyces cerevisiae*
식품	CoQ10/유비퀴놀1	효모	*Schizosaccharomyces pombe*
식품	오메가 3 지방산	미세조류	*Schizochytrium*
식품	오메가 6 지방산	미세조류	*Schizochytrium*
식품	비타민 B12	박테리아	*Propionibacterium freudenreichii*
식품	비타민 B2	사상 진균	*Ashbya gossypii*
식품	비타민 B2	박테리아	*Bacillus subtilis*
식품	에리쓰리톨	효모-유사 진균	*Torula coralline*
식품	에리쓰리톨	효모-유사 진균	*Pseudozyma tsukubaensis*
식품	에리쓰리톨	효모-유사 진균	*Moniliella pollinis*
식품	스테비올 글리코시드	효모	*Saccharomyces cerevisiae*
하이드로콜로이드	디우탄 검	박테리아	*Sphingomonas sp*
하이드로콜로이드	젤란 검	박테리아	*Sphingomonas elodea*
하이드로콜로이드	잔탄 검	박테리아	*Xanthomonas campestris*
중간체	1,3-PDO	박테리아	*Escherichia coli*
중간체	1,4-BDO	박테리아	*Escherichia coli*
중간체	부타다이엔	박테리아	*Cupriavidus necator*
중간체	n-부탄올	박테리아(절대혐기성 미생물)	*Clostridium acetobutylicum*
유기산	시트르산	사상 진균	*Aspergillus niger*
유기산	시트르산	효모	*Pichia guilliermondii*
유기산	글루콘산	사상 진균	*Aspergillus niger*
유기산	이타콘산	사상 진균	*Aspergillus terreus*
유기산	락트산	박테리아	*Lactobacillus*
유기산	락트산	박테리아	*Geobacillus thermoglucosidasius*
유기산	LCDAs - DDDA	효모	*Candida*
폴리케타이드/Ag	스피노사드	박테리아	*Saccharopolyspora spinosa*
폴리케타이드/Ag	스피네토람	박테리아	*Saccharopolyspora spinosa*
이소플라본	제니스테인	박테리아	*Saccharopolyspora erythraea*
효소	콜린 산화효소	박테리아	*Streptomyces, Thermoactinomyces or Saccharopolyspora*
약학적 조성물	쿠마미딘 화합물	박테리아	*Saccharopolyspora sp.*
선충 애벌레 발달 억제제	이버멕틴 아글리콘	박테리아	*Saccharopolyspora erythraea*
효소 억제제	HMG-CoA 환원효소 억제제	박테리아	*Saccharopolyspora sp.*
유기산	카복실산 이성질체	박테리아	*Saccharopolyspora hirsuta*
항생제	에리스로마이신	박테리아	*Saccharopolyspora erythraea*

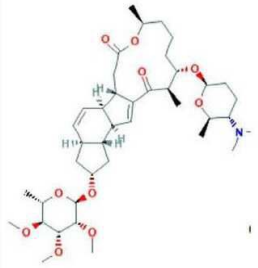
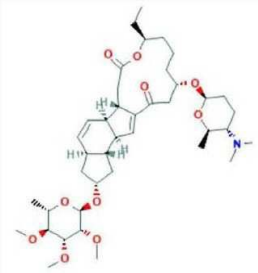
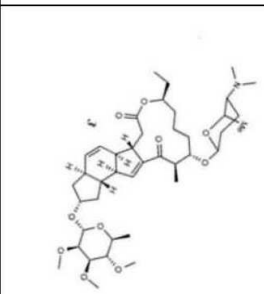
[0501] 일부 실시태양에서, 숙주 세포는 사카로폴리스포라 종이다. 일부 실시태양에서, 사카로폴리스포라 종은 사카로폴리스포라 스피노사 균주이다. 사카로폴리스포라 종에서 생산되는 관심 생성물이 아래 표 4.1에 제공된다.

[0502] [표 4.1]

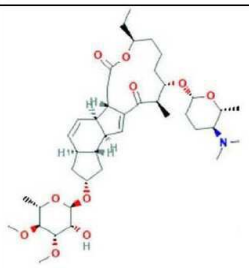
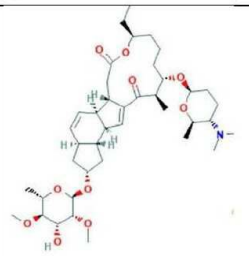
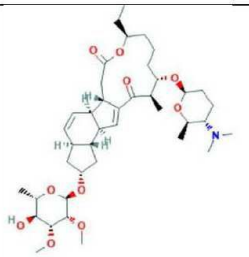
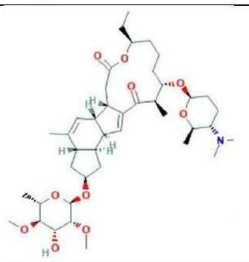
[0503] 본 발명의 사카로폴리스포라 종의 관심 생성물의 비-제한적인 목록.

생성물 명칭	구조
스피노신 A	
스피노신 B	
스피노신 C	4"-디-N-데메틸-스피노신 A
스피노신 D	

[0504]

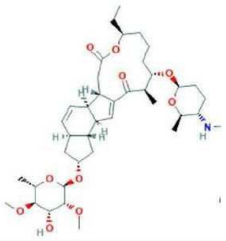
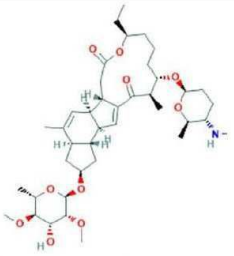
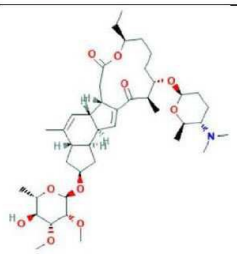
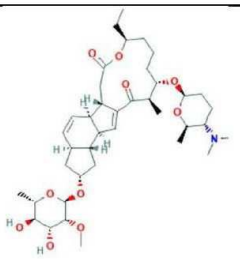
생성물 명칭	구조
스피노신 E	
스피노신 F	
스피노신 G	

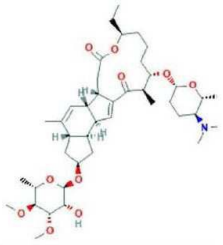
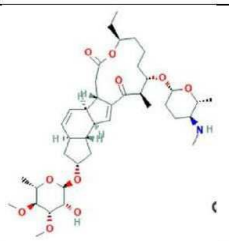
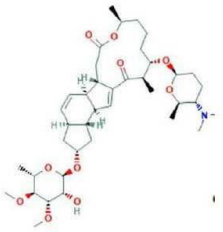
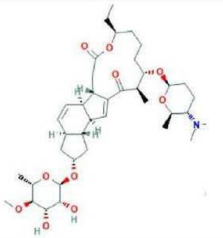
[0505]

생성물 명칭	구조
스피노신 H	
스피노신 I	N/A
스피노신 J	
스피노신 K	
스피노신 L	

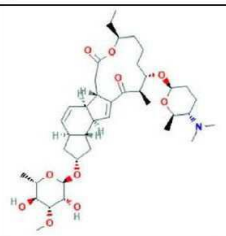
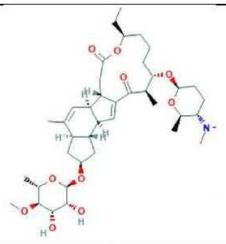
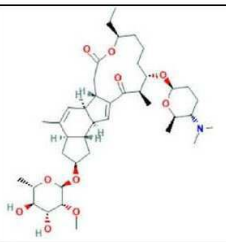
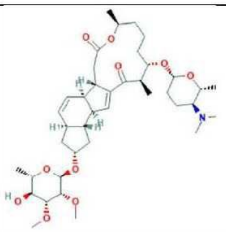
[0506]

[0507]

생성물 명칭	구조
스피노신 M	
스피노신 N	
스피노신 O	
스피노신 P	

생성물 명칭	구조
스피노신 Q	
스피노신 R	
스피노신 S	
스피노신 T	

[0508]

생성물 명칭	구조
스피노신 U	
스피노신 V	
스피노신 W	
스피노신 X	N/A
스피노신 Y	

[0509]

[0510] 스피노신은 2종의 사카로폴리스포라의 발효로부터 생성된 전례없는 화합물의 큰 패밀리이다. 이들의 핵심 구조는 2개의 당류가 첨가된 폴리케타이드 유래 테트라사이클릭 마크로라이드이다. 이들은 작물과 다른 식물에 광범위한 피해를 입히는 여러 상업적으로 중요한 많은 종들에 대해 강력한 살충 활성을 보여준다. 이들은 또한 가축, 반려 동물 및 인간의 중요한 외부 기생충에 대한 활성을 보여준다. S. 스피노사드는 두 가지 주요 발효 인자인 스피노신 A와 D의 정의된 조합이다. 스피노신 A와 스피노신 D는 S. 스피노사에 가장 풍부한 두 가지 발효 성분이다. 구조-활성 관계(SAR)가 광범위하게 연구되어 반합성 2세대 유도체인 스피네토람(Kirst, The Journal of Antibiotics (2010) 63, 101-111)이 개발되었다. 다양한 스피노신 발효로부터 다수의 구조적으로 관련된 화합물이 현재 분리되고 확인되었다. 이들의 구조는 스피노신 A의 아글리콘 또는 당류에서 단일 유형 변화의 몇 가지 일반적인 범주로 분류된다. 일부 인자는 스피노신 A와 비교하여 하나의 추가 또는 하나의 누락된 C-메틸기를 가지며, 이는 폴리케타이드 프레임워크의 형성 동안 적절한 시간에서 아세테이트 및 프로피오네이트의 교환에 의해 생합성적으로 발생한다. 스피노신 D(6-메틸-스피노신 A) 외에, 다른 단일 C-메틸-변형 인자는 스피노신 E(16-데메틸-스피노신 A) 및 스피노신 F(22-데메틸-스피노신 A)를 포함한다. 두 당류의 변형은 스피노신 H(2'-O-데메틸-스피노신 A), 스피노신 J(3'-O-데메틸-스피노신 A), 스피노신 B(4"-N-데메틸-스피노신 A) 및 스피노신 C(4"-디-N-데메틸-스피노신 A)를 포함한다. 다른 구조적 변화는 아미노당인 D-포로사민을 L-오사민(스피노신 G)과 같은 다른 당으로 교체하는 것이다. 최근에, 스피노사드 생합성 경로는 보다 정확도가 명확해졌다: 제 1 형 폴리케타이드 합성효소와 관련이 있는 spnA, spnB, spnC, spnD 및 spnE; 폴리케타이드 합성효소 생성물을 변형시키기 위한 spnF, spnJ, spnL 및 spnM(Kim et al., "Enzyme-catalysed 4+ 2 cycloaddition is a key step in the biosynthesis of spinosyn A". Nature. 2011, 473: 109-112); 람노스 부착 및 메틸화를 위한 spnG, spnH, spnI 및 spnK(Kim eta l., "Biosynthesis of spinosyn in *Saccharopolyspora spinosa*: synthesis of permethylated rhamnose and characterization of the functions of SpnH, SpnI and SpnK." J Am Chem Soc. 2010, 132: 2901-2903); 포로사민 생합성을 위한 spnP, spnO, spnN, spnQ, spnR 및 spnS; 람노스 생합성을 위한 gtt, gdh, epi 및 kre(Madduri et al. "Rhamnose biosynthesis pathway supplies precursors for

primary and secondary metabolism in *Saccharopolyspora spinosa*." J Bacteriol. 2001, 183: 5632-5638) 및 스피노사드 유전자 클러스터 옆에 4개의 유전자 ORF-L16, ORF-R1 및 ORF-R2는 스피노사드 생합성에 영향을 미치지 않는다. 이들 유전자는 본 발명에 기재된 유전 공학 방법의 잠재적 표적 중 하나이다. 스피노신 합성에 관여하는 추가 유전자는 미국 특허 번호 7,626,010, 8,624,009에 기재되어 있으며, 이는 모든 목적을 위해 그 전문이 본 발명에 참조로 포함된다.

[0511] 스피네토람은 화학적으로 변형된 스피노신 J/L 혼합물이다. 혼합물은 3'-O- 에틸-5,6-디하이드로 스피노신 J 및 3'O-에틸 스피노신 L을 2개의 주요 인자로 포함한다. 스피네토람은 스피노사드에 비해 더 넓은 스펙트럼 및 더 강력한 효능을 가지며, 현장에서의 개량된 잔류 활성을 가진다. 스피네토람의 생성은 분자 디자인이 패턴을 인식하기 위해 포유류 뇌의 신경 연결을 모방하는 소프트웨어를 사용하고 제안된 분자 변형의 활성을 추정하는데 사용될 수 있는 인공 신경 네트워크(ANN) 기반 전략의 결과이다. 결과적으로, 람노스 모이어티의 특정 알킬 치환 패턴, 특히 2', 3', 4'-tir-O-에틸 스피노신 A 유사체는 유망한 변형을 나타내는 것으로 밝혀졌다. 또한, 람노스-3'-O-에틸화는 2'- 또는 4'-O-에틸화에 비해 활성 향상에 주요 기여를 나타낸다는 것이 확인되었다. 궁극적으로, 스피네토람이 생성되었다(Sparks et al., 2008, Neural network-based QSAR and insecticide discovery: spinetoram. J Comput Aid Mol Des 22:393-401. doi:10.1007/s10822-008-9205-8).

[0512] 일부 실시태양에서, 관심 생성물은 스피노사드이다. 스피노사드는 S. 스피노사의 발효에 의해 얻어진 천연 제품 패밀리로부터 유래된 새로운 작용-모드 살충 제이다. 스피노신은 20가지가 넘는 자연 형태로 발생하며 실험실에서 200 가지가 넘는 합성 형태(스피노소이드)가 생산되었다(Watson, Gerald (31 May 2001)."Actions of Insecticidal Spinosyns on gama-Aminobutyric Acid Responses for Small-Diameter Cockroach Neurons". Pesticide Biochemistry and Physiology. 71: 20-28, 이의 전체 내용이 참조로 포함됨). 스피노사드는 대략 17:3의 비로 2가지 스피노소이드인, 주요 성분인 스피노신 A 및 스피노신 D(소량 성분)의 혼합물을 함유한다.

[0513] 일부 실시태양에서, 돌연변이 사카로폴리스포라 균주를 선별하는데 사용될 수 있는 분자는 다음을 포함하나 이에 제한되지 않는다: 1) 스피노신 합성 경로에 관련된 분자(예를 들어, 스피노신); 2) SAM/메티오닌 경로에 관련된 분자(예를 들어, 알파-메틸 메티오닌(aMM) 또는 노르류신); 3) 라이신 생산 경로에 관련된 분자(예를 들어, 티알리신 또는 알파-케토부티레이트 및 아스파테이트 하이드록사메이트의 혼합물); 4) 트립토판 경로에 관련된 분자(예를 들어, 아자세린 또는 5-플루오로인돌); 5) 트레오닌 경로에 관련된 분자(예를 들어, 베타-하이드록시노르발린); 6) 아세틸-CoA 생산 경로에 관련된 분자(예를 들어, 세룰레닌); 및 7) 드-노보 또는 회수 퓨린 및 피리미딘 경로에 관련된 분자(예를 들어, 퓨린 또는 피리미딘 유사체).

[0514] 일부 실시태양에서, 선별에 사용되는 스피노신의 농도는 약 10㎍/ml, 20㎍/ml, 30㎍/ml, 40㎍/ml, 50㎍/ml, 60㎍/ml, 70㎍/ml, 80㎍/ml, 90㎍/ml, 100㎍/ml, 200㎍/ml, 300㎍/ml, 400㎍/ml, 500㎍/ml, 600㎍/ml, 700㎍/ml, 800㎍/ml, 900㎍/ml, 1mg/ml, 2mg/ml, 3mg/ml, 4mg/ml, 5mg/ml, 6mg/ml, 7mg/ml, 8mg/ml, 9mg/ml ml, 10mg/ml 또는 그 이상이다.

[0515] 일부 실시태양에서, 선별에 사용되는 aMM의 농도는 약 0.1mM, 0.2mM, 0.3mM, 0.4mM, 0.5mM, 0.6mM, 0.7mM, 0.8mM, 0.9mM, 1mM, 2mM, 3mM, 4mM, 5mM, 6mM, 7mM, 8mM, 9mM, 10mM 또는 그 이상이다.

[0516] 일부 실시태양에서, 선별에 사용되는 분자의 정확한 농도는 사용된 균주에 따라 실험적으로 결정될 수 있다. 일반적으로, 기본 균주는 조작된 균주보다 더 민감하다.

[0517] 사카로폴리스포라 종에 대한 유전자 도구, 자원, 조성물, 방법 및 균주는 미국 특허 번호 6960453, 6270768, 5631155, 5670364, 5554519, 5187088, 5202242, 6616953, 5171740, 6420177, 8624009, 7626010, 5124258, 5362634, 6043064, 4293651, 4389486, 6627427, 5663067, 5081023, 6780633, 6004787, 6365399, 5801032, 8741603, 4328307, 4425430, 7022526, 5234828, 5786181, 5153128, 8841092, 4251511, 9309524, 6437151, 5908764,8911970, 5824513, 6524841, 7198922, 6200813, 9334514, 5496931, 7630836, 5198360, 6710189, 6251636, 7807418, 6780620, 6500960, 및 7459294에서 확인할 수 있으며, 이들 각각은 모든 목적을 위해 그 전문이 본 발명에 참조로 포함된다.

[0518] ***선택 기준 및 목표***

[0519] 본 발명의 방법에 적용되는 선택 기준은 균주 개량 프로그램의 특정 목표에 따라 달라질 것이다. 본 발명은 임의의 프로그램 목표를 충족시키도록 조정될 수 있다. 예를 들어, 일부 실시태양에서, 프로그램 목표는 즉각적인 시간 제한없이 반응의 단일 배치 수율을 최대화하는 것일 수 있다. 다른 실시태양에서, 프로그램 목표는 특정 생성물을 생산하거나 특정 비율의 생성물을 생산하기 위해 생합성 수율을 재조정하는 것일 수 있다. 다른 실시

태양에서, 프로그램 목표는 폴리머의 탄소 사슬을 길게하는 것과 같이 생성물의 화학적 구조를 변형시키는 것일 수 있다. 일부 실시태양에서, 프로그램 목표는 수율, 역가, 생산성, 부산물 제거, 공정 이상현상에 대한 저항성, 최적 성장 온도 및 성장율과 같은 성능 특성을 개량하는 것일 수 있다. 일부 실시태양에서, 프로그램 목표는 미생물에 의해 생성된 관심 생성물의 부피 생산성, 비 생산성, 수율 또는 역가에 의해 측정된 바와 같이 개량된 숙주 성능이다.다른 실시태양에서, 프로그램 목표는 투입량 당 최종 생성물 수율(예를 들어, 수크로오스의 파운드당 생산된 총 에탄올의 양)의 관점에서 상업적 균주의 합성 효율을 최적화하는 것일 수 있다. 다른 실시태양에서, 프로그램 목표는 예를 들어 배치 완료율 또는 연속 배양 시스템에서의 수율로 측정된 합성 속도를 최적화하는 것일 수 있다. 다른 실시태양에서, 프로그램 목표는 특정 파지에 대한 균주 저항성을 증가시키거나, 그렇지 않으면 배양 조건 하에서 균주 생력/강건성을 증가시키는 것일 수 있다.

[0520] 일부 실시태양에서, 균주 개량 프로젝트는 하나 이상의 목표에 종속될 수 있다. 일부 실시태양에서, 변형 프로젝트의 목표는 품질, 신뢰성 또는 전반적인 수익성에 달려있다. 일부 실시태양에서, 본 발명은 상기한 균주 특성의 하나 이상을 가진 관련된 선택된 돌연변이 또는 돌연변이 그룹을 교시한다.

[0521] 당업자는 특정 프로젝트 목표를 달성하기 위해 어떻게 균주 선택 기준을 맞춤하는지를 인식할 것이다. 예를 들어, 반응 포화시 균주의 단일 배치 최대 수율의 선택은 높은 단일 배치 수율을 가진 균주를 확인하는데 적절할 수 있다. 다양한 온도 및 조건에서 수율의 일관성에 기반한 선택은 견고성 및 신뢰성이 증가된 균주를 확인하는데 적절할 수 있다.

[0522] 일부 실시태양에서, 초기 고 처리량 단계 및 탱크-기반 검증에 대한 선택 기준은 동일할 것이다. 다른 실시태양에서, 탱크-기반 선택은 추가 및/또는 상이한 선택 기준하에서 작동할 수 있다. 예를 들어, 일부 실시태양에서, 고 처리량 균주 선택은 단일 배치 반응 완료 수율에 기초할 수 있지만, 탱크-기반 선택은 반응 속도에 대한 수율에 기초한 선택을 포함하도록 확대될 수 있다.

[0523] (a) 일부 실시태양에서, 선택 방법은 사카로폴리스포라 종의 하나 이상의 특정 대사산물 및/또는 하나 이상의 발효 생성물에 저항성인 균주를 선별하는 것을 포함한다. 일부 실시태양에서, 다양한 유전자 다형체를 포함하는 균주 컬렉션은 주어진 분자에 대해 선별된다. 균주의 컬렉션은 본 발명에 기재된 임의의 균주 라이브러리, 또는 이의 조합일 수 있다. 선택이 이루어지는 분자는 균주에 의해 생성된 임의의 최종 생성물, 또는 균주 성장에 영향을 미치는 중간 생성물 또는 최종 생성물의 수율일 수 있다. 예를 들어, 일부 실시태양에서, 분자는 상기 표 4.1의 것들과 같은 관심 스피노신 또는 스피노신의 생성에 영향을 미치는 임의의 분자일 수 있다. 본질적으로, 생성된 하나 이상의 소정의 생성물의 존재하에서 더 저항성인 균주가 선택된다. 일부 실시태양에서, 방법은 c) 선택된 균주의 성능(예를 들어, 균주에서 생성된 하나 이상의 생성물의 수율)을 분석하고 HTP 선별에 의해 기준 미생물 균주와 비교하여 개선된 성능을 갖는 균주를 선별하는 단계를 추가로 포함한다. 일부 실시태양에서, 방법은 d) 개량된 성능을 야기하는 돌연변이의 위치 및/또는 서열을 확인하는 단계를 추가로 포함한다. 초기 항-대사산물/발효 생성물 저항성 라이브러리로부터 개량된 성능을 갖는 이들 선택된 균주. 이러한 라이브러리는 상기 복수의 개별 미생물 균주의 각 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 미생물 균주를 포함하고, 여기서 상기 독특한 유전자 변이 각각은 복수의 식별 가능한 유전자 변이로부터 선택된 단일 유전자 변이에 상응한다. 일부 실시태양에서, 미생물 균주는 사카로폴리스포라 균주이다. 일부 실시태양에서, 미생물 균주에 의해 생성된 소정의 생성물은 스피노신 합성 경로에 관여하는 임의의 분자, 또는 스피노신의 생성에 영향을 줄 수 있는 임의의 분자이다. 일부 실시태양에서, 소정의 생성물은 다음을 포함하나 이에 제한되지 않는다: 스피노신 A, 스피노신 B, 스피노신 C, 스피노신 D, 스피노신 E, 스피노신 F, 스피노신 G, 스피노신 H, 스피노신 I, 스피노신 J, 스피노신 K, 스피노신 L, 스피노신 M, 스피노신 N, 스피노신 O, 스피노신 P, 스피노신 Q, 스피노신 R, 스피노신 S, 스피노신 T, 스피노신 U, 스피노신 V, 스피노신 W, 스피노신 X, 스피노신 Y, 노르류신, 노르발린, 슈도아글리콘(예를 들어, 다른 스피노신 화합물에 대한 PSA, PSD, PSJ, PSL 등) 및/또는 알파-메틸-메티오닌(aMM).

[0524] ***서열화***

[0525] 일부 실시태양에서, 본 발명은 본 발명에 기재된 유기체의 전체-게놈 서열화를 교시한다. 다른 실시태양에서, 본 발명은 또한 본 발명의 방법에 대한 품질 조절제로서의 플라스미드, PCR 생성물 및 다른 올리고의 서열화를 교시한다. 크고 작은 프로젝트를 위한 서열화 방법은 당업자에게 주지되어 있다.

[0526] 일부 실시태양에서, 핵산을 서열화하는 임의의 고 처리량 기술이 본 발명의 방법에 사용될 수 있다. 일부 실시태양에서, 본 발명은 전체 게놈 서열화를 교시한다. 다른 실시태양에서, 본 발명은 유전자 변이를 확인하기위한 앰플리콘 서열화 울트라 딥 서열화를 교시한다. 일부 실시태양에서, 본 발명은 또한 태그맨테이션

(tagmentation)을 포함하는 신규한 라이브러리 제조 방법을 교시한다(WO/2016/073690 참조). DNA 서열화 기술은 표지된 터미네이터 또는 프라이머를 사용하는 고전적인 다이데옥시 서열화 반응(생거 방법) 및 슬라브 또는 모세관에서의 겔 분리; 가역적으로 종결된 표지된 뉴클레오티드를 사용하는 합성에 의한 서열화, 열서열화; 454 서열화; 표지된 올리고 뉴클레오티드 프로브의 라이브러리에 대한 대립유전자 특이적 혼성화; 결찰이 뒤따르는 표지된 클론의 라이브러리에 대한 대립유전자 특이적 혼성화를 사용하는 합성에 의한 서열화; 중합 단계 동안 표지된 뉴클레오티드의 혼입의 실시간 모니터링; 폴로니 서열화; 및 SOLiD 서열화를 포함한다.

[0527] 본 발명의 한 양태에서, 연속적으로 서열화되는 고체 표면상의 개별 분자를 공간적으로 분리하는 단계를 포함하는 서열화의 고 처리량 방법이 사용된다. 이런 고체 표면은 비다공성 표면(솔렉사 서열화, 예를 들어, Bentley et al, Nature, 456: 53-59(2008) or Complete Genomics sequencing, *e.g.* Drmanac et al, Science, 327: 78-81(2010)), 비드-또는 입자 결합 주형을 포함할 수 있는 웰의 어레이(454 서열화, 예를 들어, Margulies et al, Nature, 437: 376-380(2005) or Ion Torrent sequencing, U.S. patent publication 2010/0137143 or 2010/0304982), 미세가공된 막(SMRT 서열화, 예를 들어, Eid et al, Science, 323: 133-138(2009)) 또는 비드 어레이(폴로니 서열화 또는 SOLiD 서열화, 예를 들어, Kim et al, Science, 316: 1481-1414(2007))를 포함할 수 있다.

[0528] 다른 실시태양에서, 본 발명의 방법은 분리된 분자가 고체 표면상에서 공간적으로 분리되기 전 또는 후에 분리된 분자를 증폭시키는 단계를 포함한다. 사전 증폭은 에멀젼 PCR, 또는 롤링 써클 증폭과 같은 에멀젼-기초 증폭을 포함할 수 있다. 또한 벤틀레이 등(상기) 및 제조사 지시(예를 들어, TruSeq™ Sample Preparation Kit and Data Sheet, Illumina, Inc., San Diego, Calif., 2010); 및 추가로 참조로 포함된 다음 참조문헌: 미국 특허 제6,090,592호; 제6,300,070호; 제7,115,400호; 및 EP0972081B1에 기술된 바와 같이 개별 주형 분자가 고체 표면상에서 공간적으로 분리된 후, 브리지 PCR에 의해 평행하게 증촉되어 분리된 클론 집단 또는 클러스터를 형성한 후, 서열화되는 솔렉사-기초 서열화가 교시된다.

[0529] 한 실시태양에서, 고체 표면상에 배치되고 증폭된 개개의 분자는 cm^2당 적어도 10^5 클러스터의 밀도; 또는 cm^2당 적어도 $5x10^5$의 밀도; 또는 cm^2당 적어도 10^6 클러스터의 밀도로 클러스터를 형성한다. 한 실시태양에서, 상대적으로 높은 에러율을 갖는 서열화 화학 반응이 사용된다. 이런 실시태양에서, 이런 화학 반응에 의해 생산된 평균 품질 점수는 서열 판독 길이의 단조 감소 함수이다. 한 실시태양에서, 이런 감소는 서열 판독의 0.5%가 1-75 위치에서 적어도 하나의 오차를 가지며; 서열 판독의 1%가 76-100 위치에서 적어도 하나의 오차를 가지며; 서열 판독의 2%가 101-125 위치에서 적어도 하나의 오차를 갖는 것에 상응한다.

[0530] **게놈-전체 유전자 디자인 기준의 효과의 전산 분석 및 예측**

[0531] 일부 실시태양에서, 본 발명은 소정의 숙주 균주에 포함되는 특정 유전자 변형의 효과를 예측하는 방법을 교시한다. 추가의 양태에서, 본 발명은 숙주가 특정 표현형 형질 또는 균주 매개변수를 갖기 위해 소정의 숙주 균주에 포함되어야 하는 제안된 유전자 변형을 생성하는 방법을 제공한다. 소정의 양태에서, 본 발명은 신규 숙주 균주를 디자인하는데 이용될 수 있는 예측 모델을 제공한다.

[0532] 일부 실시태양에서, 본 발명은 선별의 각 라운드의 수행 결과를 분석하는 방법 및 다음 선택 라운드에서 균주 성능을 향상시키도록 예측된 새로운 제안된 게놈-전체 서열 변형을 생성하는 방법을 교시한다.

[0533] 일부 실시태양에서, 본 발명은 시스템이 이전의 선택 결과에 기초하여 숙주 균주에 대해 제안된 서열 변형을 생성하는 것을 교시한다. 일부 실시태양에서, 본 시스템의 권고는 직전 선택 검사의 결과에 기초한다. 다른 실시태양에서, 본 시스템의 권고는 하나 이상의 선택 검사의 누적 결과에 기초한다.

[0534] 일부 실시태양에서, 본 시스템의 권고는 이전에 개발된 HTP 유전자 디자인 라이브러리에 기초한다. 예를 들어, 일부 실시태양에서, 본 시스템은 이전의 선별로부터의 결과를 저장하고, 그 결과를 동일 또는 상이한 숙주 유기체에서 상이한 프로젝트에 적용하도록 디자인된다.

[0535] 다른 실시태양에서, 본 시스템의 권고는 과학적인 통찰력에 기초한다. 예를 들어, 일부 실시태양에서, 권고는 (주석된 유전자 데이터베이스 및 관련 문헌과 같은 출처로부터의) 유전자의 공지된 특성, 코돈 최적화, 전사 미끄러짐, uORF 또는 다른 가설 주도 서열 및 숙주 최적화에 기초한다.

[0536] 일부 실시태양에서, 시스템 또는 예측 모델에 의해 권장되는 숙주 균주에 대해 제안된 서열 변형은 다음을 포함하는 개시된 분자 도구 세트의 하나 이상을 이용함으로써 수행된다:(1) 프로모터 스왑,(2) SNP 스왑,(3) 코돈 교환 개시/중지,(4) 서열 최적화,(5) 스왑 중지,(5) 상위성 매핑.

[0537] 본 발명에 기술된 HTP 유전자 공학 플랫폼은 임의의 특정 미생물 또는 표현형 형질에 대해 불가지론적이다(예를 들어, 특정 화합물의 생산). 즉, 본 발명에서 교시된 플랫폼 및 방법은 숙주 세포가 임의의 바람직한 표현형 형질을 갖도록 조작하기 위해 임의의 숙주 세포와 함께 사용될 수 있다. 또한, 하나의 새로운 숙주 세포를 생성하기 위해 사용된 소정의 HTP 유전자 공학 공정으로부터 얻은 교훈은, 교시된 방법 동안 발생하는 무수한 공정 매개변수의 저장, 특징화 및 분석의 결과로서, 임의의 수의 다른 숙주 세포에 적용할 수 있다.

[0538] 상위성 맵핑 섹션에서 언급된 바와 같이, HTP 유전자 디자인 라이브러리로부터의 돌연변이의 집합을 일부 바람직한 예측 모델을 통해 특정 배경으로 통합시킴으로써 얻어진 가상 균주의 성능(별칭, 점수)을 추정하는 것이 가능하다. 이런 예측 모델을 감안할 때, 조합 통합을 통해 돌연변이 라이브러리에 접근할 수 있는 모든 가상 균주를 채점하고 순위를 매기는 것이 가능하다. 아래 섹션은 본 HTP 플랫폼에서 사용되는 특정 모델에 대해 간략히 설명한다.

[0539] **예측 균주 디자인**

[0540] 본 발명에 기술 된 것은 유전자 변화 및 균주 성능을 기술하는 방법, 균주에서 변화의 구성에 기초하여 균주 성능을 예측하는 방법, 높은 예측된 성능을 갖는 후보 디자인을 추천하는 방법, 및 2차 고려사항, 예를 들어, 현존하는 균주에 대한 유사성, 상위성 또는 예측의 신뢰를 최적화를 위해 예측을 필터링하는 방법을 포함하는 예측 균주 디자인에 접근법이 본 발명에 기술된다.

[0541] **균주 디자인 모델에 대한 입력**

[0542] 한 실시태양에서, 설명의 용이함을 위해, 입력 데이터는 두 구성요소:(1) 유전자 변화의 세트 및(2) 상대적 균주 성능을 포함할 수 있다. 당업자는 이 모델이 다양한 입력을 고려하면서 과적합에 대한 상반되는 고려사항을 염두하면서 용이하게 확장될 수 있음을 인식할 것이다. 유전자 변화 이외에, 조절될 수 있는 입력 매개변수(독립변수)의 일부는 세포 유형(속, 종, 균주, 계통 발생 특성 등) 및 발효가 세포에 의해 수행되는 동안 공정 매개변수(예를 들어, 환경 조건, 취급 장비, 개조 기술, 등)이다.

[0543] 유전자 변화의 세트는 HTP 유전자 디자인 라이브러리라 불리는 유전자 교란의 이전에 논의된 집합으로부터 유래될 수 있다. 상대적 균주 성능은 임의의 소정의 파라미터 또는 관심 표현형 특성(예를 들어, 화합물, 소분자 또는 관심 생성물의 생산)에 기초하여 평가될 수 있다.

[0544] 세포 유형은 원핵생물 및 진핵생물 시스템, 속, 종, 균주, 조직 배양(vs. 분산 세포) 등과 같은 일반적인 카테고리에서 특정될 수 있다. 조절될 수 있는 공정 파라미터는 온도, 압력, 반응기 구성 및 배지 조성물을 포함할 수 있다. 반응기 구성의 예는 반응기의 부피, 공정이 배치인지 또는 연속적인지 여부, 및 연속적이면 체적 유량 등을 포함한다. 또한, 세포가 존재하는 구조를 특정할 수 있다. 배지 조성물의 예는 전해질, 영양소, 폐기물, 산, pH 등의 농도를 포함한다.

[0545] ***예측 균주 모델을 만드는데 사용되는 초기 선형 회귀 모델에서 사용될 선택된 HTP 유전자 디자인 라이브러리로부터의 유전자 변화 세트***

[0546] 예측 균주 디자인 모델을 생성하기 위해, 동일한 미생물 종의 균주의 유전적 변화가 먼저 선택된다. 각각의 유전자 변화의 이력이 또한 제공된다(예를 들어, 이 균주 계통에서 가장 최근의 변형 - "마지막 변화"). 따라서, 이 균주의 성능을 그의 부모의 성능과 비교하는 것은 "마지막 변화" 돌연변이의 성능에 관한 데이터 포인트를 나타낸다.

[0547] ***제조된 균주 성능 평가***

[0548] 교시된 모델의 목적은 균주에 도입된 유전자 변화의 구성에 기초하여 균주 성능을 예측하는 것이다. 비교를 위한 표준을 구축하기 위해, 균주 성능은 분석 플레이트당 균주당 평균 성능을 먼저 계산함으로써 공통 기준 균주에 대해 상대적으로 계산된다. 그런 후에 상대적인 성능은 동일한 플레이트 내에서 조작된 균주 및 공통 기준 균주 사이의 평균 성능에 차이로 계산된다. 계산을 플레이트 내 비교로 제한하면 고려중인 샘플 모두가 동일한 실험 조건을 얻을 수 있음을 보장한다.

[0549] 도 18은 입력 데이터에 대한 상대 균주 성능의 분포가 고려되는 예를 도시한다. 이는 본 발명에 기재된 방법을 사용하여 코리네박테리움에서 수행되었다. 그러나, 유사한 공정이 사카로폴리스포라를 위해 맞춤화되었으며 발명자들에 의해 성공적으로 수행되고 있다. 0의 상대적 성능은 조작된 균주가 인-플레이트 기본 또는 "참조" 균주에 동일하게 잘 수행되었음을 나타낸다. 0보다 크게 수행할 수 있는 균주를 확인하는 예측 모델의 능력이 중요하다. 더구나, 그리고 더 일반적으로, 임의의 소정의 균주는 일부 기준에 의해 이의 부모보다 우월한지 여부

가 중요하다. 실제로, 기준은 부모 수준 이상의 일부 임계값을 충족하거나 초과하는 생성물 역가일 수 있으며, 원하는 방향으로 부모와 통계적으로 유의미한 차이를 갖는 것은 대신에 또는 추가로 사용될 수도 있다. 기본 또는 "참조" 균주의 역할은 단순히 플레이트 내 또는 플레이트 간의 비교를 위한 추가된 표준화 요소로서 작용하는 것이다.

[0550] 염두에 두어야 할 개념은 부모 균주 및 기준 균주의 차이점이다. 부모 균주는 돌연변이유발의 현재 라운드에 사용된 배경이다. 기준 균주는 특히 플레이트 간의 비교를 용이하게 하기 위해 모든 플레이트에서 사용되는 대조군 균주이며 일반적으로 위에 언급된 "기본 균주"입니다. 그러나 기본 균주(예를 들어, 야생형 또는 산업용 균주는 전체 성능을 벤치마크하는데 사용된다)는 소정의 라운드의 균주 개량에서 돌연변이유발 표적이라는 의미에서 반드시 "기본"이 아니기 때문에, 보다 설명적인 용어는 "기준 균주"이다.

[0551] 요약하면, 기본/기준 균주는 제조된 균주의 성능을 벤치마크하는데 사용되는 반면, 부모 균주는 관련 유전자 배경에서 특이적 유전자 변화의 성능을 벤치마크하는데 사용된다.

[0552] ***선형 회귀 분석에 의한 제조된 균주의 성능 순위 매김***

[0553] 개시된 모델의 목적은 제조된 균주에 도입된 유전자 변화의 구성의 함수로서, 상대적 균주 성능을 기술함으로써, 제조된 균주의 성능을 순위매기기 위한 것이다. 본 발명 전반에 걸쳐 논의된 바와 같이, 다양한 HTP 유전자 디자인 라이브러리는 조작된 균주에 도입되는 가능한 유전자 변화(예를 들어, 유전자 교란/변경)의 레퍼토리를 제공한다. 선형 회귀는 현재 설명된 전형적인 예측 모델의 기초이다.

[0554] 이어서, 유전자 변화 및 상대 성능에 대한 효과가 회귀-기반 모델링을 위해 입력된다. 균주 성능은 균주에 포함된 유전자 변화의 구성의 함수로서 공통 기본 균주에 상대적으로 평가된다.

[0555] ***제조된 균주를 특징화하는 선형 회귀***

[0556] 선형 회귀는 구현 및 해석의 용이함 때문에 기술된 HTP 게놈 공학 플랫폼에 대한 매력적인 방법이다. 얻어진 회귀 계수는 각 유전자 변화의 존재에 기인한 상대적 균주 성능의 평균 증가 또는 감소로 해석될 수 있다.

[0557] 예를 들어, 몇몇 실시태양에서 볼 수 있는 바와 같이, 이 기술은 오리지날 프로모터를 다른 프로모터로 변경하면 평균 대략 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 또는 그 이상의 단위만큼 상대적 균주 성능을 개량시키고, 따라서 어떠한 음성 상위성 상호작용의 부재시에 잠재적으로 매우 바람직한 변화라고 결론 내리게 한다(참고: 입력은 단위가 없는 정규화 값).

[0558] 따라서, 교시된 방법은 선형 회귀 모델을 사용하여 다양한 교시된 라이브러리로부터의 그들의 게놈에 도입된 다양한 유전자 교란을 갖는 제조된 균주를 기술하고/특성화하고 순위를 매긴다.

[0559] ***예측 디자인 모델링***

[0560] 구축된 균주로부터의 데이터를 사용하는 상기한 선형 회귀 모델은 아직 제조되지 않은 균주에 대한 성능 예측을 수행하는데 사용될 수 있다.

[0561] 절차는 다음과 같이 요약될 수 있다: 가능한 모든 형태의 유전자 변화를 인 실리코로 생성한다 → 회귀 모델을 사용하여 상대적인 균주 성능을 예측한다 → 성능에 의해 후보 균주 디자인을 순서화한다. 따라서 회귀 모델을 이용하여 아직 제조되지 않은 균주의 성능을 예측함으로써, 이 방법은 더 적은 실험을 수행하는 동시에 고성능 균주의 생산을 가능하게 한다.

[0562] ***구성 생성***

[0563] 아직 제조되지 않은 균주의 성능을 예측하기 위한 모델을 구축하는 경우, 첫 번째 단계는 일련의 디자인 후보를 생산하는 것이다. 이것은 균주의 유전자 변화의 총 수를 고정한 다음, 가능한 모든 유전자 변화의 조합을 정의함으로써 이루어진다. 예를 들어, 잠재적 유전자 변화/교란의 총 수를 29(예를 들어, 유전자 교란의 영역이 29인 한 29개 가능한 SNP 또는 29개 상이한 프로모터 또는 이의 임의의 조합)으로 설정한 후 29개 잠재적 유전자 변화의 모든 가능한 3-구성원 조합을 디자인하도록 결정할 수 있으며, 이것이 3,654개 후보 균주 디자인을 생성할 것이다.

[0564] 상기한 3,654개 후보 균주에 대한 문맥을 제공하기 위해, n!/((n-r)! * r!)을 사용하여 n개의 가능한 구성원으로부터 크기 r의 비 중복 그룹의 수를 계산할 수 있다고 생각한다. r = 3이면 n = 29는 3,654를 제공한다. 따라서, 29개 잠재적 변화의 모든 가능한 3개 구성원 조합을 디자인하면, 결과는 3,654개 후보 균주이다.

[0565] ***새로운 균주 디자인의 성능 예측***

[0566] 입력으로서 조합 구성으로 위에서 구축된 선형 회귀 분석을 사용하여, 각각의 후보 디자인의 예상된 상대 성능을 예측할 수 있다. 예를 들어, 상위 100개의 예측된 균주 디자인에 대한 변화의 조성은 2차원 맵으로 요약될 수 있으며, 여기서 x-축은 잠재적인 유전적 변화의 풀(29개의 가능한 유전적 변화)을 나열하고, y-축은 순위 순서를 나타낸다. 검은색 셀은 후보 디자인의 특정 변화가 있음을 나타내는데 사용될 수 있는 반면, 흰색 셀은 그 변화가 없음을 나타내는데 사용될 수 있다.

[0567] 예측 정확도는 새로운 관찰이 모델을 반복적으로 재훈련하고 재구성하는데 사용됨에 따라시간이 지남에 따라 증가해야 한다. 본 발명자에 의한 연구 결과는 예측 모델이 반복적으로 재훈련되고 개량될 수 있는 방법을 설명한다. 모델 예측의 품질은 예측된 값과 관찰된 값 사이의 연관성의 강도를 나타내는 상관 계수 또는 평균 모델 오차의 척도인 평균 제곱근 오차를 포함하는 여러 방법을 통해 평가될 수 있다. 모델 평가에 선택된 메트릭을 사용하여, 시스템은 모델이 재훈련되어야 할 때에 대한 규칙을 정의할 수 있다.

[0568] 상기 모델에 대한 몇 가지 설명되지 않은 가정은 다음을 포함한다:(1) 상위성 상호작용이 없다;(2) 예측 모델을 구축하기 위해 사용된 유전자 변화/교란은 유전자 변형의 제안된 조합으로서 모두 동일한 배경에서 만들어진다.

[0569] ***2차 기능에 대한 필터링***

[0570] 상기 설명된 예는 예측된 숙주 세포 성능에 기초한 선형 회귀 예측에 초점을 두었다. 일부 실시태양에서, 본 선형 회귀 방법은 포화 바이오매스, 저항성 또는 다른 측정가능한 숙주 세포 특징과 같은 비-생체분자 인자에도 적용될 수 있다. 따라서, 본 발명의 방법은 또한 제조할 후보자의 우선순위를 정할 때 예측된 성능 이외의 다른 특징을 고려하는 것을 교시한다. 추가적인 관련 데이터가 있다고 가정하면, 비선형 항도 회귀 모델에 포함된다.

[0571] ***기존의 균주와의 밀접성***

[0572] 이미 제조된 것과 유사한 예측된 균주는 상위 예측 후보가 아니더라도시간 및 비용을 절감할 수 있다

[0573] ***변화의 다양성***

[0574] 상기한 모델을 구축할 때, 유전자 변화가 상위성 상화작용의 존재에 의해 진정으로 부가적일 것을 확신할 수 없다(선형 회귀에 의해 가정되고 상기 가정으로 언급됨). 그러므로, 유전자 변화 비유사성의 지식은 긍정적인 가산성의 가능성을 증가시키는데 사용될 수 있다. 예를 들어 상위 균주로부터의 변화가 동일한 대사 경로에 있고 유사한 성능 특성을 갖는 것을 아는 경우, 그 정보는 변화의 비유사한 구성을 가진 다른 상위 균주를 선택하는데 사용될 수 있다. 상위성 매핑에 관한 위의 섹션에서 설명한 바와 같이, 예측된 최고의 유전자 변화는 상당히 비유사한 응답 프로파일을 가진 돌연변이에 대한 선택을 제한하기 위해 필터링될 수 있다. 대안적으로, 선형 회귀는 예측을 가중하기 위해 유사성 매트릭스를 사용하는 가중된 최소 자승 회귀(weighted least squares regression)일 수 있다.

[0575] ***예측된 성능의 다양성***

[0576] 마지막으로, 예측 모델을 검증하고 후속적으로 개량하기 위해, 중간 또는 나쁜 예측된 성능을 가진 균주를 디자인하도록 선택할 수 있다.

[0577] ***반복 균주 디자인 최적화***

[0578] 실시태양에서, 주문 배치 엔진(208)은 공장(210)에 공장 주문을 하여 상위 후보 돌연변이를 포함하는 미생물 균주를 제조한다. 피드백-루프 방식에서, 결과는 분석 장비(214)에 의해 분석되어 어느 미생물이 원하는 표현형 특성을 나타내는지를 결정할 수있다(314). 분석 단계 동안, 변형된 균주 배양물은 평가되어 성능, 즉 산업적 규모로 생산되는 능력을 포함하는 원하는 표현형 특성의 발현을 측정한다. 예를 들어, 분석 단계는, 다른 것들 중에서, 콜로니 건강의 지표로서 미생물 콜로니 성장을 측정하기 위해 플레이트의 이미지 데이터를 사용한다. 분석 장비(214)는 유전자 변화를 표현형 성능과 상관시키고 결과적인 유전자형-표현형 상관 데이터를 라이브러리(206)에 저장될 라이브러리에 저장하여 장래의 미생물 생산을 알려주는데 사용된다.

[0579] 특히, 충분히 높게 측정된 성능을 실제로 초래하는 후보 변화는 상기 표 4와 같은 표에 데이터베이스의 행으로서 추가될 수 있다. 이러한 방식으로, 최고의 성능 돌연변이가 감독된 기계 학습 방식으로 예측 균주 디자인 모델에 추가된다.

[0580] LIMS는 이전의 공장 가동으로부터 개발된 상관관계에 기초하여 디자인/제조/테스트/분석 사이클을 반복한다. 후

속 사이클 동안, 단독으로 또는 인간 조작자와 함께 분석 장비(214)는 상관관계 데이터를 사용하여 유전자 변형을 미세하게 조정하여 더 미세한 정교성(granularity)으로 더 나은 표현형 성능을 달성하도록 입력 인터페이스(202)로 다시 입력하기 위한 기본 균주로서 최고의 후보를 선택할 수 있다. 이러한 방식으로, 본 발명의 실시태양의 실험실 정보 관리 시스템은 품질 개량 피드백 루프를 구현한다.

[0581] 요약하면, 도 26의 흐름도를 참조하면, 반복 예측 균주 디자인 작업 흐름은 다음과 같이 기술될 수 있다:

[0582] ● 입력과 출력 변수의 훈련 세트, 예를 들어, 입력으로 유전자 변화 및 출력으로 성능 특성을 생성한다(3302). 생성은 이전의 유전자 변화 및 이러한 유전자 변화를 포함하는 미생물 균주의 상응하는 측정된 성능에 기초하여 분석 장비(214)에 의해 수행될 수 있다.

[0583] ● 훈련 세트를 기반으로 초기 모델(예를 들어, 선형 회귀 모델)을 개발한다(3304). 이는 분석 장비(214)에 의해 수행될 수 있다.

[0584] ● 디자인 후보 균주를 생성한다(3306)

[0585] ● 한 실시태양에서, 분석 장비(214)는 변화의 조합의 형태로, 배경 균주에 대해 행해지는 유전자 변화의 수를 고정시킬 수 있다. 이런 변화를 나타내기 위해, 분석 장비(214)는 변화의 조합을 나타내는 하나 이상의 DNA 세부사항 표현을 해석자(204)에 제공할 수 있다.(이런 유전자 변화 또는 이런 변화를 포함하는 미생물 균주는 "테스트 입력"으로 불릴 수 있다) 해석자(204)는 하나 이상의 DNA 세부사항을 해석하고, 실행 엔진(207)은 DNA 세부사항을 실행하여 이런 변화에 대한 개별 후보 후보자 디자인 균주를 나타내는 분석된 출력으로 DNA 세부사항을 채운다.

[0586] ● 모델에 기초하여, 분석 장비(214)는 각각의 후보 디자인 균주의 예상 성능을 예측한다(3308).

[0587] ● 분석 장비(214)는 최고의 예측된 성능을 가진 제한된 수, 예를 들어, 100개의 후보 디자인을 선택한다(3310).

[0588] ● 상위성 매핑과 관련하여 본 발명의 다른 곳에서 설명된 바와 같이, 분석 장비(214)는 예를 들어, 상위성 효과에 대한 상위 디자인을 필터링하거나, 예측 모델 속에 상위성을 고려하여 상위성과 같은 2차 효과를 설명할 수 있다.

[0589] ● 주문 배치 엔진(208)에 의해 생성된 공장 주문에 기초하여 필터링된 후보 균주를(공장(210)에서)제조한다(3312).

[0590] ● 분석 장비(214)는 선택된 균주의 실제 성능을 측정하고 우수한 실제 성능을 기초로 제한된 수의 선택된 균주를 선택하고(3314), 디자인 변화와 이의 얻어진 성능을 예측 모델에 추가한다(3316). 선형 회귀 예에서, 디자인 변화 및 이의 관련 성능의 세트를 표 4에 새로운 행으로 추가한다.

[0591] ● 분석 장비(214)는 새로운 디자인 후보 균주의 생성으로 다시 반복하고(3306), 정지 조건이 만족될 때까지 반복을 지속한다. 정지 조건은, 예를 들어, 수율, 성장률 또는 역가와 같은 성능 메트릭을 만족시키는 적어도 하나의 미생물 균주의 측정된 성능을 포함할 수 있다.

[0592] 상기 예에서, 변형 디자인의 반복 최적화는 기계 학습을 구현하기 위해 피드백 및 선형 회귀를 사용한다. 일반적으로, 기계 학습은 제한된 수의 표지된 데이터의 예를 사용하고 알려지지 않은 데이터에 동일한 작업을 실행하여 정보 작업의 수행(예를 들어, 분류 또는 회귀)시에, 매개 변수, 기술 또는 기타 기능과 같은 성능 기준의 최적화로 기술될 수 있다. 상기 선형 회귀 예와 같은 감독된 기계 학습에서, 기계(예를 들어, 컴퓨팅 디바이스)는 예를 들어 훈련 데이터에 의해 나타난 패턴, 카테고리, 통계적 관계 또는 다른 속성을 확인함으로써 학습한다. 그런 다음 학습 결과는 새로운 데이터가 동일한 패턴, 카테고리, 통계적 관계 또는 기타 속성을 나타낼 것인지를 예측하는데 사용된다.

[0593] 본 발명의 실시태양은 훈련 데이터가 이용가능할 때 다른 감독된 기계 학습 기술을 사용할 수 있다. 훈련 데이터가 없는 경우, 실시태양은 감독되지 않은 기계 학습을 채택할 수 있다. 대안적으로, 실시태양은 소량의 표지된 데이터 및 다량의 비표지된 데이터를 사용하는 반-감독된 기계 학습을 채택할 수 있다. 실시태양은 또한 기계 학습 모델의 성능을 최적화하기 위해 가장 관련이 있는 특징의 서브세트를 선택하기 위해 특징 선택을 채택할 수 있다. 선형 회귀에 대한 대안으로서 또는 선형 회귀에 추가로 선택된 기계 학습 접근법의 유형에 따라, 실시태양은, 예를 들어, 로지스틱 회귀, 신경망, 지지 벡터 기계(SVM), 결정 트리, 숨겨진 마르코프 모델, 베이지안 네트워크, 그램 슈미트, 보강-기반 학습, 계층적 클러스터링을 포함하는 클러스터 기반 학습, 유전자 알고

리즘 및 당업계에 공지된 임의의 다른 적절한 학습 기계를 채택할 수 있다. 특히, 실시태양은 로지스틱 회귀를 이용하여 분류 자체와 함께 분류(예를 들어, 상이한 기능 그룹으로 유전자의 분류)의 확률을 제공할 수 있다. 예를 들어, Shevade, A simple and efficient algorithm for gene selection using sparse logistic regression, Bioinformatics, Vol. 19, No. 17 2003, pp. 2246-2253, Leng, et al., Classification using functional data analysis for temporal gene expression data, Bioinformatics, Vol. 22, No. 1, Oxford University Press(2006), pp. 68-76를 참조하고, 이의 전부는 전문이 참조로 본 발명에 포함된다.

[0594] 실시태양은, 특히 심층 신경망(DNN)으로 알려진 형태로 기계 학습 작업을 수행함에 있어 인기가 증가하고 있는 것으로 밝혀진 그래픽 처리 장치(GPU) 가속 아키텍처를 채택할 수 있다. 본 발명의 실시태양은 GPU 기반 딥 러닝 인터페이스: A Performance 및 Power Analysis, NVidia Whitepaper, November 2015, Dahl, et al., Multi-task Neural Networks for QSAR Predictions, Dept. of Computer Science, Univ. of Toronto, June 2014(arXiv:1406.1231 [stat.ML])에 기술된 것과 같은 GPU-기반 기계 학습을 채택할 수 있고, 이의 전부는 전문이 참조로 본 발명에 포함된다. 본 발명의 실시태양에 적용할 수 있는 기계 학습 기술은 다른 참조문헌 중에서, Libbrecht, et al., Machine learning applications in genetics and genomics, Nature Reviews: Genetics, Vol. 16, June 2015, Kashyap, et al., Big Data Analytics in Bioinformatics: A Machine Learning Perspective, Journal of Latex Class Files, Vol. 13, No. 9, Sept. 2014, Prompramote, et al., Machine Learning in Bioinformatics, Chapter 5 of Bioinformatics Technologies, pp. 117-153, Springer Berlin Heidelberg 2005에서 발견될 수 있고, 이의 전부는 전문이 참조로 본 발명에 포함된다.

[0595] **반복적인 예측 균주 디자인: 실시예**

[0596] 다음은 위에서 설명한 반복적인 예측 균주 디자인 작업 흐름의 예시적인 응용을 제공한다.

[0597] 훈련 입력 및 출력 변수의 초기 세트가 준비되었다. 이 세트는 정의된 유전자 구성을 가진 1864개의 고유한 조작 균주로 구성되었다. 각 균주는 5 내지 15개의 조작된 변화를 포함하였다. 총 336개의 고유한 유전자 변화가 훈련에 존재하였다.

[0598] 초기 예측 컴퓨터 모델이 개발되었다. 구현은 일반 선형 모델(4차 다항 커널에 의한 커널 릿지 회귀)을 사용하였다. 구현은 두 가지 뚜렷한 표현형(수율 및 생산성)을 만든다. 아래 도시된 바와 같이, 이들 표현형은 가중치 합계로 결합되어 순위에 대한 단일 점수를 얻는다. 다양한 모델 매개 변수, 예를 들어, 정규화 인자는 지정된 훈련 데이터에 대해 k 배 교차 검증을 통해 조정되었다.

[0599] 상기 구현은 상기 상위성 매핑 섹션에 기술된 바와 같이 상호작용 효과의 어떠한 명백한 분석도 포함하지 않는다. 그러나, 당업자가 이해할 수 있는 바와 같이, 구현된 일반 선형 모델은 암시적으로 커널의 2차, 3차 및 4차 항을 통해 상호작용 효과를 포착할 수 있다.

[0600] 모델은 훈련 세트에 대해 훈련되었다. 훈련 후, 훈련 데이터에 대한 수율 모델의 현저한 품질 적합성이 입증될 수 있다.

[0601] 이후 후보 균주가 생성된다. 이 실시태양은 부모 균주에 대한 새로운 유전자 변화의 도입과 관련된 일련의 제조 제약을 포함한다. 여기서, 후보자는 단순히 변화의 원하는 수의 함수로 간주되지 않는다. 대신에, 분석 장비(214)는, 출발점으로서, 고성능 메트릭("종자 균주")을 갖는 것으로 알려진 이전에 디자인된 균주의 집합을 선택한다. 분석 장비(214)는 각각의 종자 균주에 유전자 변화를 개별적으로 적용한다. 도입된 유전자 변화는 종자 균주에 이미 존재하는 것들을 포함하지 않는다. 다양한 기술적, 생물학적 또는 다른 이유로, 특정 돌연변이가 명시적으로 요구되거나 명시적으로 배재된다.

[0602] 모델에 기초하여, 분석 장비(214)는 후보 균주 디자인의 성능을 예측한다. 분석 장비(214)는 관심있는 두 표현형에 대한 예측된 성능(수율 및 생산성)에 기초하여 후보를 "최상"에서 "최악"으로 순위를 매긴다. 특히, 분석 장비(214)는 가중 합을 사용하여 후보 균주에 점수를 매긴다:

[0603] 점수 = 0.8 * 수율/ 최대(수율) + 0.2 * 생산성/최대(생산성),

[0604] 여기서 수율은 후보 균주에 대한 예측된 수율을 나타내고,

[0605] 최대(수율)는 모든 후보 균주에 대한 최대 수율을 나타내며,

[0606] 생산성은 후보 균주의 생산성을 나타내며,

[0607] 최대(생산성)은 모든 후보 균주에 대한 최대 생산성을 나타낸다.

[0608] 분석 장비(214)는 용량 제약 및 작업 제약 모두를 부과함으로써 후보의 랭킹 된 목록으로부터 최종 세트의 권고를 생성한다. 일부 실시태양에서, 용량 제한은 48개 컴퓨터 생성 후보 디자인 균주와 같은 주어진 수에서 설정될 수 있다.

[0609] 훈련된 모델(위에 설명됨)은 각각의 후보 균주의 예상된 성능(수율 및 생산성에 대해)을 예측하는데 사용될 수 있다. 분석 장비(214)는 위에 제공된 스코어링 함수를 사용하여 후보 균주를 평가할 수 있다. 이후 용량 및 작업 제약이 적용되어 48개 후보 균주의 필터된 세트를 생산할 수 있다.

[0610] 이후 필터된 후보 균주는 순서 배치 엔진(208)에 의해 생성된 공장 순서에 기초하여(공장(210)에서) 제조된다(3312). 순서는 후보 균주에 해당하는 DNA 세부사항에 기초할 수 있다.

[0611] 실제로, 제조 공정은 예측된 실패율을 가지므로 균주의 랜덤으로 세트가 제조되지 않는다.

[0612] 분석 장비(214)는 또한 실제 수율 및 생산성 성능을 측정하는데 사용될 수 있다. 분석 장비(214)는 3개의 기준; 모델 정확성; 균주 성능 개량; 인간 전문가가 생성 한 디자인에 대한 동등성(또는 개량)에 기초하여 모델 및 권장 균주를 평가할 수 있다.

[0613] 수율 및 생산성 표현형은 권장 균주에 대해 측정하고 모델에 의해 예측된 값과 비교할 수 있다.

[0614] 다음으로, 분석 장비(214)는 각각의 권장 균주에 대한 부모 균주로부터의 성능 변화율을 계산한다.

[0615] 예측 정확도는 예측된 값과 관측된 값 간의 연관성을 나타내는 상관 계수 또는 평균 모델 오차의 척도인 평균 제곱근 오차를 포함하는 여러 방법을 통해 평가될 수 있다.

[0616] 여러 라운드의 실험에 걸쳐, 모델 예측은 표류할 수 있고, 새로운 유전자 변화가 예측 정확도를 향상시키기 위해 훈련 입력에 추가될 수 있다. 이 실시예의 경우, 디자인 변경과 이의 결과로 얻은 성능이 예측 모델에 추가되었다(3316).

[0617] **서비스로서의 게놈 디자인 및 공학**

[0618] 본 발명의 실시태양에서, 도 25의 LIMS 시스템 소프트웨어(3210)는 도 25의 클라우드 컴퓨팅 시스템(3202)에서 구현되어, 다수의 사용자가 본 발명의 실시태양에 따라 미생물 균주를 디자인 및 제조할 수 있게 한다. 도 25는 본 발명의 실시태양에 따른 클라우드 컴퓨팅 환경(3204)을 도시한다. 도 25에 도시된 것과 같은 클라이언트 컴퓨터(3206)는 인터넷과 같은 네트워크(3208)를 통해 LIMS 시스템에 접근한다. 실시태양에서, LIMS 시스템 애플리케이션 소프트웨어(3210)는 클라우드 컴퓨팅 시스템(3202)에 상주한다. LIMS 시스템은 도 25에 도시된 유형의 하나 이상의 프로세서를 사용하는 하나 이상의 컴퓨팅 시스템을 사용할 수 있다. 클라우드 컴퓨팅 시스템 자체는 네트워크를 통해 LIMS 시스템 어플리케이션(3210)을 클라이언트 컴퓨터(3206)에 인터페이스하는 네트워크 인터페이스(3212)를 포함한다. 네트워크 인터페이스(3212)는 클라이언트 컴퓨터(3206)에서 클라이언트 애플리케이션이 LIMS 시스템 소프트웨어(3210)에 접근할 수 있게 하는 어플리케이션 프로그래핑 인터페이스(API)를 포함할 수 있다. 특히, API를 통해, 네트워크 컴퓨터(3026)는 입력 인터페이스(202), 해석자(204), 실행 엔진(207), 주문 배치 엔진(208), 공장(210)을 구동하는 소프트웨어뿐만 아니라 테스트 장비(212) 및 분석 장비(214)를 제한없이 포함하는 LIMS 시스템(200)의 구성요소에 접근할 수 있다. 서비스로서 소프트웨어(SaaS) 소프트웨어 모듈(3214)은 클라이언트 컴퓨터(3206)에 서비스로서 LIMS 시스템 소프트웨어(3210)을 제공한다. 클라우드 관리 모듈(3216)은 클라이언트 컴퓨터들(3206)에 의해 LIMS 시스템(3210)에 대한 접근을 관리한다. 클라우드 관리 모듈(3216)은 당업계에 공지된 멀티테넌트 애플리케이션, 가상화 또는 다른 아키텍처를 사용하는 클라우드 아키텍처가 여러 사용자에게 서비스를 제공하는 것을 가능하게 한다.

[0619] **게놈 자동화**

[0620] 본 발명의 방법의 자동화는 다중 테스트 균주 변이체로부터의 표적 생성물의 높은 처리량의 표현형 선택 및 확인을 동시에 가능하게 한다.

[0621] 상기한 게놈 공학 예측 모델링 플랫폼은 수백 및 수천 개의 돌연변이체 균주가 고 처리량 방식으로 구축된다는 사실에 전제된다. 아래에 설명된 로봇 및 컴퓨터 시스템은 이런 고 처리량 공정이 수행될 수있는 구조적 메커니즘입니다.

[0622] 일부 실시태양에서, 본 발명은 숙주 세포 생산성을 개량하거나 산업용 균주를 재활시키는 방법을 교시한다. 이 공정의 일부로서, 본 개시는 DNA를 조립하고, 새로운 균주를 제조하고, 플레이트에서 배양물을 선별하고, 탱크 발효 모델에서 배양 물을 선별하는 방법을 교시한다. 일부 실시태양에서, 본 발명은 새로운 숙주 균주를 생성하

고 테스트하는 상기한 방법의 하나 이상이 자동화 로봇에 의해 보조되는 것을 교시한다.

[0623] 일부 실시태양에서, 본 발명은도 6a-b에 도시된 바와 같이 고 처리량 균주 공학 플랫폼을 교시한다.

[0624] ***HTP 로봇 시스템***

[0625] 일부 실시태양에서, 본 발명의 자동화 방법은 로봇 시스템을 포함한다. 본 발명에 개략된 시스템은 일반적으로 96- 또는 384-웰 미세 역가 플레이트의 사용에 관한 것이지만, 당업자라면 알 수 있듯이, 임의의 수의 상이한 플레이트 또는 구성이 사용될 수 있다. 또한, 본 발명에 개략된 단계의 일부 또는 전부는 자동화될 수 있고, 따라서, 예를 들어, 시스템은 완전히 또는 부분적으로 자동화될 수 있다.

[0626] 일부 실시태양에서, 본 발명의 자동화 시스템은 하나 이상의 작업 모듈을 포함한다. 예를 들어, 일부 실시태양에서, 본 발명의 자동화 시스템은 DNA 합성 모듈, 벡터 클로닝 모듈, 균주 변형 모듈, 선택 모듈 및 서열화 모듈을 포함한다(도 7 참조).

[0627] 당업자가 알 수 있는 바와 같이, 자동화 시스템은 액체 핸드러; 하나 이상의 로봇 암; 마이크로플레이트의 위치 결정을 위한 플레이트 핸들러; 플레이트 씰러(plate sealers), 플레이트 피어서(plate piercers), 비 교차 오염 플레이트상의 웰을 위한 뚜껑을 제거하고 교체하는 자동화 뚜껑 처리기; 일회용 팁에 의해 샘플 배급을 위한 일회용 팁 어셈블리; 샘플 분배를 위한 세척 가능한 팁 어셈블리; 96 웰 로딩 블록; 통합 열 순환기; 냉각된 시약랙; 미세적정 플레이트 피펫 위치(선택적으로 냉각됨); 플레이트와 팁을 위한 스태킹 타워; 마그네틱 비드 가공 스테이션; 여과 시스템; 플레이트 교반기; 바코드 판독기 및 애플리케이터; 및 컴퓨터 시스템을 포함하나 이에 제한되지 않는 매우 다양한 구성요소를 포함할 수 있다.

[0628] 일부 실시태양에서, 본 발명의 로봇 시스템은 유전자 표적화 및 재조합 분야의 공정에서의 모든 단계를 수행하기 위한 고 처리량 피펫팅을 가능하게 하는 자동화된 액체 및 입자 취급을 포함한다. 이것은 흡인, 분배, 혼합, 희석, 세척, 정확한 체적 이동과 같은 액체 및 입자 조작; 피펫 팁의 회수 및 폐기; 단일 샘플 흡인으로 다중 전달을 위한 동일한 부피의 반복적 피펫팅을 포함한다. 이런 조작은 교차 오염이 없는 액체, 입자, 세포 및 유기체 전달이다. 이 장비는 필터, 막 및/또는 도터 플레이트에 대한 마이크로 플레이트 샘플의 자동화 반복, 고밀도 전달, 전체 플레이트 연속 희석 및 고용량 작업을 수행한다.

[0629] 일부 실시태양에서, 본 발명의 맞춤형 자동화 액체 처리 시스템은 TECAN 머신(예를 들어, 맞춤형 TECAN Freedom Evo)이다.

[0630] 일부 실시태양에서, 본 발명의 자동화 시스템은 다중 웰 플레이트, 딥 웰 플레이트, 정사각형 플레이트, 시약 트로프, 테스트 튜브, 미니 튜브, 마이크로퓨즈 튜브, 냉동 튜브, 필터, 마이크로 어레이 칩, 광섬유, 비드, 아가로스 및 아크릴 아마이드 겔을 위한 플랫폼과 양립가능하며 다른 고상 매트릭스 또는 플랫폼은 업그레이드 가능한 모듈식 데크에 수용된다. 일부 실시태양에서, 본 발명의 자동화 시스템은 공급원 및 출력 샘플, 시약, 샘플 및 시약 희석액, 분석 플레이트, 샘플 및 시약 저장조, 피펫 팁 및 활성 팁 세척 스테이션을 배치하기 위한 다중 위치 작업 표면을 위한 적어도 하나의 모듈 데크를 포함한다.

[0631] 일부 실시태양에서, 본 발명의 자동화 시스템은 고 처리량 전기 천공 시스템을 포함한다. 일부 실시태양에서, 고효율 전기 천공 시스템은 96 또는 384-웰 플레이트에서 세포를 형질전환시킬 수 있다. 일부 실시태양에서, 고효율 전기 천공 시스템은 VWR® 고 처리량 전기 천공 시스템, BTX™, Bio-Rad® Gene Pulser MXcell™ 또는 다른 다중-웰 전기 천공 시스템을 포함한다.

[0632] 일부 실시태양에서, 통합된 열 순환기 및/또는 열 조절기는 0℃ 내지 100℃에서 샘플을 배양하는 정확한 온도 제어를 제공하도록 제어된 블록 또는 플랫폼과 같은 열 교환기의 온도를 안정화하는데 사용된다.

[0633] 일부 실시태양에서, 본 발명의 자동화 시스템은 액체, 입자, 세포 또는 다중 세포 유기체를 로봇 공학적으로 조작할 수 있는 단일 또는 다중 자기 프로브, 친 화성 프로브, 리플리케이터 또는 피펫터를 갖는 교환 가능한 머신-헤드(단일 또는 다중 채널)과 양립가능하다. 다중 웰 또는 다중 튜브 자기 분리기 및 여과 스테이션은 단일 또는 다중 샘플 형식으로 액체, 입자, 세포 및 유기체를 조작한다.

[0634] 일부 실시태양에서, 본 발명의 자동화 시스템은 카메라 비전 및/또는 분광계 시스템과 양립가능하다. 따라서, 일부 실시태양에서, 본 발명의 자동화 시스템은 진행중인 세포 배양물에서 색 및 흡수 변화를 검출 및 기록할 수 있다.

[0635] 일부 실시태양에서, 본 발명의 자동화 시스템은 시스템이 다수의 애플리케이션을 수행할 수 있게 하는 다중 하

드웨어 애드온(add-on)으로 융통성 있고 적응할 수 있도록 디자인된다. 소프트웨어 프로그램 모듈은 방법의 생성, 수정 및 실행을 허용한다. 시스템의 진단 모듈은 설정, 장비 정렬 및 모터 작동을 허용한다. 맞춤형 도구, 랩웨어 및 액체 및 입자 전달 패턴은 다양한 응용 프로그램이 프로그램되고 실행되게 한다. 데이터베이스는 방법 및 매개변수 저장을 허용한다. 로봇 및 컴퓨터 인터페이스는 장비 간의 통신을 허용한다.

[0636] 따라서, 일부 실시태양에서, 본 발명은 도 19에 도시된 바와 같이 고 처리량 균주 공학 플랫폼을 교시한다.

[0637] 당업자는 본 발명의 HTP 공학 방법을 수행할 수 있는 다양한 로봇 플랫폼을 인식할 것이다. 아래의 표 5는 도 19에서 기술된 바와 같이 본 발명의 HTP 공학 단계의 각 단계를 수행할 수 있는 과학적 장비의 포괄적 목록을 제공한다.

[0638] [표 5]

[0639] 본 발명의 HTP 공학 방법과 양립가능한 과학 장비의 포괄적 목록.

	장비 형태	실행된 작업(들)	양립가능한 장비 제조사/모델/구성
확득되고 제조된 DNA 조각	액체 처리기	DNA 부분의 PCR 증폭을 위한 프라이머/주형을 히트피킹(Hitpicking)(전 달과 조합)	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, or equivalents
	열 순환기	DNA 부분의 PCR 증폭	Inheco Cycler, ABI 2720, ABI Proflex 384, ABI Veriti, or equivalents
QC DNA 부분	단편 분석기 (모세관 전기영동)	적절한 크기의 PCR 생성물을 확인하기 위한 겔 전기영동	Agilent Bioanalyzer, AATI Fragment Analyzer, or equivalents
	시퀀서(sang er: Beckman)	부분/주형의 서열을 확인	Beckman Ceq-8000, Beckman GenomeLab™, or equivalents
	NGS (next generation sequencing) 장비	부분/주형의 서열을 확인	Illumina MiSeq series sequences, illumina Hi-Seq, Ion torrent, pac bio or other equivalents

[0640]

	장비 형태	실행된 작업(들)	양립가능한 장비 제조사/모델/구성
	나노드랍/플레이트 판독기	DNA 샘플의 농도 평가	Molecular Devices SpectraMax M5, Tecan M1000, or equivalents.
DNA 어셈블리 생성	액체 처리기	클로닝 벡터와 함께 조립을 위한 DNA 부분을 히트피킹(전달과 조합), 조립 반응/공정을 위한 시약의 첨가	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, or equivalents
QC DNA 어셈블리	콜로니 선택기	액체 배지에 콜로니 접종	Scirobotics Pickolo, Molecular Devices QPix 420
	액체 처리기	프라이머/주형을 히트피킹, 샘플 희석	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, or equivalents
	단편 분석기 (모세관 전기영동)	적절한 크기의 조립된 생성물을 확인하기 위한 겔 전기영동	Agilent Bioanalyzer, AATI Fragment Analyzer

[0641]

	장비 형태	실행된 작업(들)	양립가능한 장비 제조사/모델/구성
	시퀀서(sanger: Beckman)	조립된 플라스미드의 서열을 확인	ABI3730 Thermo Fisher, Beckman Ceq-8000, Beckman GenomeLab™, or equivalents
	NGS (next generation sequencing) 장비	조립된 플라스미드의 서열을 확인	Illumina MiSeq series sequences, illumina Hi-Seq, Ion torrent, pac bio or other equivalents
기본 영기 및 DNA 어셈블리 제조	원심분리기	세포 스피닝/펠렛팅	Beckman Avanti floor centrifuge, Hettich Centrifuge
DNA 를 기본 균주로 형질전환	전기 천공기	세포의 전기 천공 형질전환	BTX Gemini X2, BIO-RAD MicroPulser Electroporator
	탄도 형질전환	세포의 탄도 형질전환	BIO-RAD PDS1000
	배양기, 열 순환기	화학적 형질전환/열 충격	Inheco Cycler, ABI 2720, ABI Proflex 384, ABI Veriti, or equivalents
	액체 처리기	DNA, 세포, 버퍼를 혼합	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, or equivalents

[0642]

	장비 형태	실행된 작업(들)	양립가능한 장비 제조사/모델/구성
DNA 를 기본 균주의 게놈에 통합	콜로니 선택기	액체 배지에 콜로니를 접종	Scirobotics Pickolo, Molecular Devices QPix 420
	액체 처리기	세포를 한천에 전달, 배양판으로부터 다른 배양판으로 전달(다른 선택적 배지 속에 접종)	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, or equivalents
	플랫폼 교반기-배양기	미세역가 플레이트 배양물의 교반에 의한 접종	Kuhner Shaker ISF4-X, Infors-ht Multitron Pro
QC 형질전환 균주	콜로니 선택기	액체 배지에 콜로니를 접종	Scirobotics Pickolo, Molecular Devices QPix 420
	액체 처리기	프라이머/주형을 히트피킹, 샘플 희석	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, or equivalents
	열 순환기	균주의 cPCR 확인	Inheco Cycler, ABI 2720, ABI Proflex 384, ABI Veriti, or equivalents

[0643]

	장비 형태	실행된 작업(들)	양립가능한 장비 제조사/모델/구성
	단편 분석기 (모세관 전기영동)	적절한 크기의 cPCR 생성물을 확인하기 위한 겔 전기영동	Infors-ht Multitron Pro, Kuhner Shaker ISF4-X
	시퀀서 (sanger: Beckman)	도입된 변형의 서열 확인	Beckman Ceq-8000, Beckman GenomeLab™, or equivalents
	NGS (next generation sequencing) 장비	도입된 변형의 서열 확인	Illumina MiSeq series sequences, illumina Hi-Seq, Ion torrent, pac bio or other equivalents
통합 QC'd 균주를 선택하고 테스트 플레이트 속에	액체 처리기	배양 플레이트로부터 다른 배양 플레이트에 전달(생산 배지 속에 접종)	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, or equivalents
	콜로니 선택기	액체 배지에 콜로니 접종	Scirobotics Pickolo, Molecular Devices QPix 420
	플랫폼 교반기- 배양기	미세역가 플레이트 배양물의 교반에 의한 접종	Kuhner Shaker ISF4-X, Infors- ht Multitron Pro

[0644]

	장비 형태	실행된 작업(들)	양립가능한 장비 제조사/모델/구성
씨드 플레이트에서 균주 배양	액체 처리기	배양 플레이트로부터 다른 배양 플레이트에 전달(생산 배지 속에 접종)	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, or equivalents
	플랫폼 교반기- 배양기	미세역가 플레이트 배양물의 교반에 의한 접종	Kuhner Shaker ISF4-X, Infors-ht Multitron Pro
	액체 분배기	액체 배양 배지를 미세역가 플레이트에 분배	Well mate (Thermo), Benchcel2R (velocity 11), plateloc (velocity 11)
	미세플레이터 표시기	플레이트에 바코드 부착	Microplate labeler (a2+ cab - agilent), benchcell 6R (velocity11)
균주로부터 생성물 생성	액체 처리기	배양 플레이트로부터 다른 배양 플레이트에 전달(생산 배지 속에 접종)	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, or equivalents
	플랫폼 교반기- 배양기	미세역가 플레이트 배양물의 교반에 의한 접종	Kuhner Shaker ISF4-X, Infors-ht Multitron Pro

[0645]

	장비 형태	실행된 작업(들)	양립가능한 장비 제조사/모델/구성
	액체 분배기	액체 배양 배지를 여러 미세역가 플레이트에 분배하고 플레이트를 밀봉	well mate (Thermo), Benchcel2R (velocity 11), plateloc (velocity 11)
	미세플레이터 표시기	플레이트에 바코드 부착	microplate labeler (a2+ cab - agilent), benchcell 6R (velocity11)
성능 평가	액체 처리기	하류 분석을 위해 배양액 처리	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, or equivalents
	UHPLC, HPLC	전구체 및 표적 화합물의 정량 분석	Agilent 1290 Series UHPLC and 1200 Series HPLC with UV and RI detectors, or equivalent; also any LC/MS
	LC/MS	전구체 및 표적 화합물뿐만 아니라 부생성물 및 분해생성물의 고특이적 분석	Agilent 6490 QQQ and 6550 QTOF coupled to 1290 Series UHPLC

[0646]

	장비 형태	실행된 작업(들)	양립가능한 장비 제조사/모델/구성
	분광 측광기	분광 측광기 기반 분석을 사용하는 상이한 화합물의 정량화	Tecan M1000, spectramax M5, Genesys 10S
플라스크에서 균주 배양	발효기:	교반에 의한 배양	Sartorius, DASGIPs (Eppendorf), BIO-FLOs (Sartorius-stedim). Applikon
	플랫폼 교반기		innova 4900, or any equivalent
균주로부터 생성물 생성	발효기: DASGIPs (Eppendorf), BIO-FLOs (Sartorius-stedim)		
성능 평가	액체 처리기	배양 플레이트로부터 다른 배양 플레이트에 전달(생산 배지 속에 접종)	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, or equivalents
	UHPLC, HPLC	전구체 및 표적 화합물의 정량 분석	Agilent 1290 Series UHPLC and 1200 Series HPLC with UV and RI detectors, or equivalent; also any LC/MS

[0647]

	장비 형태	실행된 작업(들)	양립가능한 장비 제조사/모델/구성
	LC/MS	전구체 및 표적 화합물뿐만 아니라 부생성물 및 분해생성물의 고 특이적 분석	Agilent 6490 QQQ and 6550 QTOF coupled to 1290 Series UHPLC
	유세포 분석기	균주 성능을 특성화(생존가능성 측정)	BD Accuri, Millipore Guava
	분광 측광기	균주 성능을 특성화(바이오매스 측정)	Tecan M1000, Spectramax M5, or other equivalents

[0648]

[0649] ***컴퓨터 시스템 하드웨어***

[0650] 도 27은 본 발명의 실시태양에 따라 비 일시적 컴퓨터 판독 가능 매체(예를 들어, 메모리)에 저장된 프로그램 코드를 실행하는데 사용될 수 있는 컴퓨터 시스템(800)의 한 예를 도시한다. 컴퓨터 시스템은 애플리케이션에 따라 인간 사용자 및/또는 다른 컴퓨터 시스템과 인터페이스하기 위해 사용될 수 있는 입력/출력 서브 시스템(802)을 포함한다. I/O 서브 시스템(802)은 예를 들어, 입력을 위한 키보드, 마우스, 그래픽 사용자 인터페이스, 터치 스크린, 또는 다른 인터페이스, 및 예를 들어 애플리케이션 프로그램 인터페이스(APIs)를 포함하는 출력을 위한 LED 또는 다른 평면 디스플레이를 포함할 수 있다. LIMS 시스템의 구성요소와 같은 본 발명의 실시태양의 다른 요소는 컴퓨터 시스템(800)과 같은 컴퓨터 시스템으로 구현될 수 있다.

[0651] 프로그램 코드는 2차 메모리(810) 또는 메인 메모리(808) 또는 둘 다에서 의 영구 저장장치와 같은 비 일시적인 매체에 저장될 수 있다. 메인 메모리 808)는 랜덤 액세스 메모리(RAM)와 같은 휘발성 메모리 또는 리드 온리 메모리(ROM)와 같은 비 휘발성 메모리뿐만 아니라 명령들 및 데이터에 대한보다 빠른 접근를 위한 상이한 수준의 캐시 메모리를 포함할 수 있다. 2차 메모리는 솔리드 스테이트 드라이브, 하드 디스크 드라이브 또는 광 디스크와 같은 영구 저장장치를 포함할 수 있다. 하나 이상의 프로세서(804)는 하나 이상의 비 일시적인 매체로부터 프로그램 코드를 판독하고 컴퓨터 시스템이 본 실시태양에 의해 수행된 방법을 완성할 수 있도록 코드를 실행한다. 당업자는 프로세서(들)가 소스 코드를 섭취하고 소스 코드를 프로세서(들)(804)의 하드웨어 게이트 레벨에서 이해할 수 있는 머신 코드로 해석 또는 컴파일할 수 있다는 것을 이해할 것이다. 프로세서(들)(804)은 컴퓨팅 집약적인 작업을 처리하기 위한 그래픽 처리 장치(GPU)를 포함할 수 있다. 특히, 기계 학습에서, 하나 이상의 CPU(804)는 하나 이상의 GPU(804)로 대량의 데이터 처리를 오프로드할 수 있다.

[0652] 프로세서(들)(804)는 네트워크 인터페이스 카드, WiFi 트랜시버 등과 같은 하나 이상의 통신 인터페이스(807)를 통해 외부 네트워크와 통신할 수 있다. 버스(805)는 I/O 서브시스템(802), 프로세서(들)(804), 주변 장치(806), 통신 인터페이스(807), 메모리(808) 및 영구 저장장치(810)와 통신가능하게 연결된다. 본 발명의 실시태양은 이런 대표적 아키텍처에 제한되지 않는다. 대안적 실시태양은 상이한 배열 및 유형의 구성요소, 예를 들어 입출력 구성요소 및 메모리 서브시스템을 위한 개별 버스를 사용할 수 있다.

[0653] 당업자는 본 발명의 실시태양의 요소의 일부 또는 전부 및 이들의 수반되는 작업이 컴퓨터 시스템(800)의 하나 이상의 프로세서 및 하나 이상의 메모리를 포함하는 하나 이상의 컴퓨터 시스템에 의해 전체적으로 또는 부분적으로 구현될 수 있다는 것을 이해할 것이다. 특히, LIMS 시스템(200)의 요소 및 본 발명에 기술된 임의의 로봇 및 다른 자동화 시스템 또는 장치는 컴퓨터로 구현될 수 있다. 일부 요소 및 기능은 국소적으로 구현될 수 있고, 다른 것들은 예를 들어, 클라이언트- 서버 방식과 같은, 상이한 서버를 통해 네트워크 도처에 분산된 방식으로 구현될 수 있다. 특히 도 25와 같이, 서버 측 작업은 서비스로서 소프트웨어(SaaS) 방식으로 여러 클라이언트에 이용될 수 있다.

[0654] 본 내용에서 구성요소라는 용어는 소프트웨어, 하드웨어 또는 펌웨어(또는 이의 임의의 조합) 구성요소를 포괄적으로 의미한다. 구성요소는 일반적으로 지정된 입력(들)을 사용하여 유용한 데이터 또는 다른 출력을 생성할 수 있는 기능적 구성요소이다. 구성요소는 자급식일 수 있거나 아닐 수 있다. 애플리케이션 프로그램("애플리케이션"이라고도 함)은 하나 이상의 구성요소를 포함할 수 있으며 구성요소는 하나 이상의 애플리케이션 프로그램을 포함할 수 있다.

[0655] 일부 실시태양은 다른 모듈 또는 애플리케이션 구성요소와 함께 구성요소의 일부, 전부를 포함하거나 하나도 포함하지 않는다. 여전히, 다양한 실시태양은 이들 구성요소의 둘 이상을 단일 모듈에 통합하고 및/또는 이들 구성요소의 하나 이상의 기능의 일부를 다른 구성요소와 연관시킬 수 있다.

[0656] 용어 "메모리"는 정보를 저장하기 위해 사용되는 임의의 장치 또는 메커니즘일 수 있다. 본 발명의 일부 실시태양에 따르면, 메모리는 휘발성 메모리, 비 휘발성 메모리 및 동적 메모리의 임의의 유형을 포함하나 이에 제한되는 것은 아니다. 예를 들어, 메모리는 랜덤 액세스 메모리, 메모리 저장 장치, 광학 메모리 장치, 자기 매체, 플로피 디스크, 자기 테이프, 하드 드라이브, SIMMs, SDRAM, DIMMs, RDRAM, DDR RAM, SODIMMS, 지울 수 있는 프로그램가능한 리드 온리 메모리(EPROMs), 전기적으로 지울 수 있는 프로그램가능한 리드 온리 메모리(EEPROMs), 컴팩트 디스크, DVDs 및/또는 이와 동일한 종류일 수 있다. 일부 실시태양에 따르면, 메모리는 하나 이상의 디스크 드라이브, 플래시 드라이브, 데이터베이스, 로컬 캐시 메모리, 프로세서 캐시 메모리, 관계형 데이터베이스, 플랫 데이터베이스, 서버, 클라우드 기반 플랫폼 및/또는 이와 동일한 종류를 포함할 수 있다. 또한, 당업자는 정보를 저장하기 위한 많은 부가적인 장치 및 기술이 메모리로서 사용될 수 있다는 것을 이해할

것이다.

[0657] 메모리는 프로세서 상의 하나 이상의 애플리케이션 또는 모듈을 실행하기 위한 명령을 저장하는데 사용될 수 있다. 예를 들어, 일부 실시태양에서 메모리는 본 발명에 개시된 하나 이상의 모듈 및/또는 애플리케이션의 기능을 실행하는데 필요한 명령의 전부 또는 일부를 수용하기 위해 사용될 수 있다.

[0658] **유전자 디자인 예측을 기반으로 하는 HTP 미생물 균주 공학: 예시적 작업 흐름**

[0659] 일부 실시태양에서, 본 발명은 본 발명의 전산 분석 시스템의 권고에 기초한 새로운 숙주 유기체의 지시된 공학을 교시한다.

[0660] 일부 실시태양에서, 본 발명은 모든 유전자 디자인 및 클로닝 방법과 양립할 수 있다. 즉, 일부 실시태양에서, 본 발명은 폴리머라제 연쇄 반응, 제한 효소 절단, 결찰, 상동성 재조합, RT PCR 및 당업계에 일반적으로 공지된 다른 기술과 같은 전통적인 클로닝 기술의 사용을 교시하고 예를 들어 참조로 본 발명에 포함된 Sambrook *et al.*(2001) Molecular Cloning: A Laboratory Manual(3rd ed., Cold Spring Harbor Laboratory Press, Plainview, New York에 기술된다.

[0661] 일부 실시태양에서, 복제된 서열은 본 발명에서 교시된 임의의 HTP 유전자 디자인 라이브러리, 예를 들면: 프로모터 스왑 라이브러리로부터의 프로모터, SNP 스왑 라이브러리로부터의 SNP, 개시/중지 코돈 변화 라이브러리로부터의 개시 또는 중지 코돈, STOP 스왑 라이브러리의 터미네이터 또는 서열 최적화 라이브러리로부터의 서열 최적화의 임의의 것으로부터 가능성을 포함할 수 있다.

[0662] 또한, 특정 구조체에 포함되어야 하는 정확한 서열 조합은 상위성 맵핑 기능에 의해 통지될 수 있다.

[0663] 다른 실시태양에서, 복제된 서열은 합리적인 디자인(가설-주도적) 및/또는 과학 출판물과 같은 다른 소스에 기초한 서열에 기초한 서열을 또한 포함할 수 있다.

[0664] 일부 실시태양에서, 본 발명은 i) 맞춤형 SNP 특이적 DNA를 생성하는 단계, ii) SNP 특이적 플라스미드를 조립하는 단계,(iii) SNP 특이적 DNA로 표적 숙주 세포를 형질전환시키는 단계, 및 iv) 모든 선택 마커를 루프아웃하는 단계를 포함하는 지시된 공학의 방법을 교시한다(도 2 참조).

[0665] 도 6a는 DNA의 획득 및 조립, 벡터의 조립, 숙주 세포의 형질전환 및 선택 마커의 제거를 포함하는 본 발명의 균주 공학 방법의 일반적인 작업 흐름을 도시한다.

[0666] ***특정 DNA 올리고 뉴클레오티드 제조***

[0667] 일부 실시태양에서, 본 발명은 숙주 세포 유기체의 DNA 절편을 삽입 및/또는 교체 및/또는 변경 및/또는 삭제하는 것을 교시한다. 일부 양태에서, 본 발명에 교시된 방법은 숙주 유기체의 게놈 속에 포함될 관심 올리고뉴클레오티드(즉, 표적 DNA 절편)를 제조하는 것을 포함한다. 일부 실시태양에서, 본 발명의 표적 DNA 절편은 공지된 주형으로부터의 복제 또는 절단, 돌연변이 또는 DNA 합성을 포함하여 당업계에 공지된 임의의 방법을 통해 얻을 수있다. 일부 실시태양에서, 본 발명은 표적 DNA 서열(예를 들어, GeneArt™, GeneMaker™, GenScript™, Anagen™, Blue Heron™, Entelechon™, GeNOsys, Inc., 또는 Qiagen™)을 생산하기 위한 상업적으로 구입가능한 유전자 합성 생성물과 양립가능하다.

[0668] 일부 실시태양에서, 표적 DNA 절편은 숙주 유기체의 선택된 DNA 영역에 SNP를 포함시키도록(예를 들어, 유익한 SNP를 첨가하여) 디자인된다. 다른 실시태양에서, DNA 절편은 숙주 유기체의 DNA로부터 SNP를 제거하도록(예를 들어, 해로운 또는 중성 SNP를 제거하도록) 디자인된다.

[0669] 일부 실시태양에서, 본 발명의 방법에서 사용된 올리고뉴클레오티드는 당 업계에 공지된 효소적 또는 화학적 합성 방법 중 임의의 방법을 사용하여 합성될 수 있다. 올리고뉴클레오티드는 제어된 공극 유리(CPG), 폴리스티렌 비드 또는 CPG를 함유할 수 있는 열가소성 폴리머로 구성된 막과 같은 고체 지지체 상에서 합성될 수 있다. 올리고 뉴클레오티드는 미세유체역학(Tian *et al.*, Mol. BioSyst., 5, 714-722(2009)) 또는 둘 다의 조합을 제공하는 공지된 기술(Jacobsen *et al.*, 미국 특허 출원 제2011/0172127호 참조)을 사용하여 병렬 마이크로스케일로 어레이 상에서 합성될 수 있다.

[0670] 어레이 상에 또는 미세유체역학을 통한 합성은 보다 낮은 시약 사용을 통해 비용을 감소시킴으로써 통상적인 고체 지지체 합성에 비해 이점을 제공한다. 유전자 합성에 필요한 규모는 작아서, 어레이 또는 미세유체역학을 통해 합성된 올리고 뉴클레오티드 생성물의 규모가 허용가능하다. 그러나, 합성된 올리고뉴클레오티드는 고체 지지체 합성을 사용할 때보다 품질이 떨어진다(이하의 Tian 참조; 또한 Staehler *et al.*, U.S. Pat. App. No.

2010/0216648 참조).

[0671] 퍼옥시 음이온 탈보호를 사용하여 1980년대에 처음으로 기술되었기 때문에 종래의 4단계 포스포아미다이트 화학에서 많은 발전이 이루어져왔다(예를 들어, Sierzchala, *et al. J. Am. Chem. Soc.,* 125, 13427-13441(2003) using peroxy anion deprotection; Hayakawa *et al.*, U.S. Pat. No. 6,040,439 for alternative protecting groups; Azhayev et al, *Tetrahedron* 57, 4977-4986(2001) for universal supports; Kozlov *et al.*, *Nucleosides, Nucleotides,* and *Nucleic Acids,* 24(5-7), 1037-1041(2005) for improved synthesis of longer oligonucleotides through the use of large-pore CPG; and Damha *et al.*, *NAR,* 18, 3813-3821(1990) for improved derivatization 참조)

[0672] 합성 유형에 관계없이, 생성된 올리고뉴클레오티드는 보다 긴 올리고뉴클레오티드를 위한 더 작은 빌딩 블록을 형성할 수 있다. 일부 실시태양에서, 더 작은 올리고뉴클레오티드는 폴리머라제 연쇄 조립(PCA), 리가아제 연쇄 반응(LCR) 및 열역학적으로 균형 잡힌 인사이드-아웃 합성(TBIO)과 같은 당업계에 공지된 프로토콜을 사용하여 함께 결합될 수 있다(Czar *et al.* Trends in Biotechnology, 27, 63-71(2009) 참조). PCA에서, 원하는 긴 제품의 전체 길이에 걸쳐있는 올리고뉴클레오티드는 어닐링되고 여러 사이클(일반적으로 약 55 사이클)에서 연장되어 전장 생성물을 최종적으로 얻는다. LCR은 리가아제 효소를 사용하여 제 3 올리고뉴클레오티드에 모두 어닐링되는 두 올리고뉴클레오티드에 연결시킨다. TBIO 합성은 목적 생성물의 중심에서 시작하여 유전자의 5' 말단에서 포워드 가닥과 상동성이고 유전자의 3' 말단에서 리버스 가닥과 상동성인 중첩 올리고뉴클레오티드를 사용함으로써 점진적으로 양방향으로 연장된다.

[0673] 더 큰 이중 가닥 DNA 단편을 합성하는 다른 방법은 상위-가닥 PCR(TSP)을 통해 더 작은 올리고뉴클레오티드를 조합하는 것이다. 이 방법에서, 복수의 올리고뉴클레오티드는 원하는 생성물의 전체 길이에 걸쳐 있고 인접한 올리고뉴클레오티드(들)에 중첩 영역을 포함한다. 증폭은 범용 포워드 및 리버스 프라이머로 수행될 수 있으며, 증폭의 다중 사이클을 통해 전장 이중 가닥 DNA 생성물이 형성된다. 이런 생성물은 선택적인 오류 보정 및 원하는 이중 가닥 DNA 단편 최종 생성물을 생성하는 추가 증폭을 진행될 수 있다.

[0674] TSP의 한 방법에서, 전장의 원하는 생성물을 형성하도록 결합되어질 더 작은 올리고뉴클레오티드의 세트는 40-200 염기 길이이고 적어도 약 15-20 염기만큼 서로 중첩된다. 실제적인 목적을 위해, 중첩 영역은 올리고뉴클레오티드의 특정 어닐링을 보장할만큼 충분히 길어야 하고 사용된 반응 온도에서 어닐링하기에 충분한 높은 용융 온도(T_m)를 가져야 한다. 중첩은 소정의 올리고뉴클레오티드가 인접한 올리고뉴클레오티드에 의해 완전히 중첩되는 지점까지 연장될 수 있다. 중첩의 양은 최종 생성물의 품질에 영향을 미치지 않는 것으로 보입니다. 어셈블리의 첫 번째 및 마지막 올리고뉴클레오티드 빌딩 블록은 포워드 및 리버스 증폭 프라이머에 대한 결합 위치를 포함해야 한다. 하나의 실시태양에서, 첫 번째 및 마지막 올리고뉴클레오티드의 최종 말단 서열은 보편적 프라이머의 사용을 허용하도록 상보성의 동일한 서열을 포함한다.

[0675] ***맞춤형 플라스미드 조립/복제***

[0676] 일부 실시태양에서, 본 발명은 숙주 유기체의 게놈에 원하는 표적 DNA 부분(예를 들어, 특정 SNP를 포함)을 삽입할 수 있는 벡터를 제조하는 방법을 교시한다. 일부 실시태양에서, 본 발명은 표적 DNA, 상동성 암 및 적어도 하나의 선택 마커를 포함하는 벡터를 복제하는 방법을 교시한다(도 3 참조).

[0677] 일부 실시태양에서, 본 발명은 숙주 유기체로의 형질전환에 적합한 임의의 벡터와 양립가능하다. 일부 실시태양에서, 본 발명은 숙주 세포와 양립가능한 셔틀 벡터의 사용을 교시한다. 한 실시태양에서, 본 발명에 제공된 방법에서 사용하기 위한 셔틀 벡터는 대장균 및/또는 사카로폴리스포라 숙주 세포와 양립가능한 셔틀 벡터이다. 본 발명에 제공된 방법에서 사용하기 위한 셔틀 벡터는 본 발명에 기술된 바와 같은 선택 및/또는 역-선택을 위한 마커를 포함할 수 있다. 표지는 당업계에 공지된 및/또는 본 발명에 제공된 임의의 마커일 수 있다. 셔틀 벡터는 당업계에 공지된 바와 같은 상기 셔틀 벡터의 어셈블리에 유용한 임의의 조절 서열(들) 및/또는 서열을 추가로 포함할 수 있다. 셔틀 벡터는 예를 들어 대장균 또는 C. 글루 타미쿰과 같은 본 발명에 제공된 바와 같은 숙주 세포에서 증식에 필요할 수 있는 임의의 복제 기점을 추가로 포함할 수 있다. 조절 서열은 숙주 세포의 유전자 기구에 의해 사용된 프로모터, 개시, 중지, 신호, 분비 및/또는 종결 서열과 같은 당업계에 공지되거나 본 발명에 제공되는 임의의 조절 서열일 수 있다. 특정 예에서, 표적 DNA는 상업용 벡터(예를 들어, DNA2.0 맞춤형 또는 GATEWAY® 벡터)와 같은 임의의 저장소 또는 카탈로그 생성물로부터 얻을 수 있는 벡터, 구조체 또는 플라스미드에 삽입될 수 있다.

[0678] 일부 실시태양에서, 본 발명의 조립/복제 방법은 다음 조립 전략: i) II형 통상적인 복제, ii) II형 S-매개 또

는 "골든 게이트" 복제(예를 들어, Engler, C., R. Kandzia, and S. Marillonnet. 2008 "A one pot, one step, precision cloning method with high-throughput capability". PLos One 3:e3647; Kotera, I., and T. Nagai. 2008 "A high-throughput and single-tube recombination of crude PCR products using a DNA polymerase inhibitor and type IIS restriction enzyme." J Biotechnol 137:1-7.; Weber, E., R. Gruetzner, S. Werner, C. Engler, and S. Marillonnet. 2011 Assembly of Designer TAL Effectors by Golden Gate Cloning. PloS One 6:e19722 참조), iii) GATEWAY® 재조합, iv) TOPO® 복제, 엑소뉴클레아제 매개 어셈블리(Aslanidis 및 de Jong 1990. "Ligation-independent cloning of PCR products(LIC-PCR)." Nucleic Acids Research, Vol. 18, No. 20 6069), v) 상동성 재조합, vi) 비 상동성 말단 결합, vii) 깁슨 어셈블리(Gibson et al., 2009 "Enzymatic assembly of DNA molecules up to several hundred kilobases" Nature Methods 6, 343-345) 또는 이의 조합의 적어도 하나를 사용할 수 있다. 모듈형 IIS 기반 조립 전략은 PCT 공개공보 WO 2011/154147에 개시되며, 그 내용은 본 발명에 참고로 포함된다.

[0679] 일부 실시태양에서, 본 발명은 하나 이상의 선택 마커를 갖는 벡터의 클로닝을 교시한다. 다양한 선택 마커 유전자는 선택적 압력하에서 원핵생물 세포(예를 들어, 암피실린, 카나마이신, 테트라사이클린, 클로람페니콜, 제오신, 스펙티노마이신/스트렙토마이신) 또는 진핵생물 세포(예컨대, 제네티신, 네오마이신, 하이그로마이신, 푸로마이신, 블라스티시딘, 제오신)에서 선택을 위한 항생제 저항을 코딩하는 것이 당업계에 주지되어 있다. 다른 마커 시스템은 성공적으로 형질유래된 숙주 세포에서 발현된 녹색 또는 적색 형광 단백질과 같은 X-gal 또는 형광 리포터(fluorescent reporter)의 존재하에서 양성 클론을 선택하기 위해 박테리아에서 사용된 주지된 청색/백색 선택 시스템과 같은 원하거나 원하지 않는 세포의 선택 및 확인을 허용한다. 대다수가 원핵생물 시스템에서만 기능하는 선택 마커의 또 다른 부류는 생산자 세포를 죽이는 독성 유전자 생성물을 발현하는 "사멸 유전자"라고도하는 카운터 선택가능한 마커 유전자에 관한 것이다. 이러한 유전자의 예는 sacB, rpsL(strA), tetAR, pheS, thyA, gata-1 또는 ccdB를 포함하며, 그 기능은(Reyrat *et al.* 1998 "Counterselectable Markers: Untapped Tools for Bacterial Genetics and Pathogenesis." Infect Immun. 66(9): 4011-4017)에 설명된다.

[0680] ***카운터 선택 마커***

[0681] 본 발명은 또한 사카로폴리스포라 종의 유전자 조작을 위한 카운터 선택 마커를 제공한다. 일부 실시태양에서, 사카로폴리스포라 종은 사카로폴리스포라 스피노사이다. 일부 실시태양에서, 카운터 선택 마커는 레반수크라제(EC 2.4.1.10)를 코딩하는 sacB(레반수크라제) 유전자, 페닐알라닌 tRNA 합성효소(pheS) 유전자, 또는 이의 조합이다.

[0682] 일부 실시태양에서, sacB 또는 pheS 유전자를 코딩하는 뉴클레오티드 서열은 사카로폴리스포라 스피노사와 같은 사카로폴리스포라 종에 대해 코돈-최적화된다. 일부 실시태양에서, sacB를 코딩하는 뉴클레오티드 서열은 서열번호 146을 포함한다. 일부 실시태양에서, sacB를 코딩하는 뉴클레오티드 서열은 서열번호 146에 대해 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 그 이상의 상동성을 갖는다. 일부 실시태양에서, pheS를 코딩하는 뉴클레오티드 서열은 서열 번호 147 또는 서열 번호 148을 포함한다. 일부 실시태양에서, pheS를 암호화하는 뉴클레오티드 서열은 서열 번호 147 또는 서열 번호 148에 대해 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 그 이상의 상동성을 갖는다.

[0683] 본 발명의 카운터 선택 마커 유전자를 포함하는 사카로폴리스포라 종에 대한 게놈 통합을 위한 플라스미드가 또한 제공된다. 일부 실시태양에서, 플라스미드는 플라스미드 백본, 카운터 선택 마커 유전자에 더하여 양성 선택 마커, 상동성 좌측 암 서열, 상동성 우측 암 서열 및 DNA 페이로드(예를 들어, 통합될 편집된 유전자)를 포함한다. 상동성 좌측 암 우측 암 서열은 표적화된 야생형 유전자좌와 DNA 페이로드 사이에서 상동성 재조합을 가능하게 한다. 일부 실시태양에서, 카운터 선택 마커는 sacB 유전자 또는 pheS 유전자이다.

[0684] 또한, 사카로폴리스포라 종의 돌연변이 균주를 생성하는 방법이 제공된다. 일부 실시태양에서, 방법은 a) 본 발명의 카운터 선택 마커 유전자를 포함하는 플라스미드를 부모 사카로폴리스포라 균주 내로 도입하는 단계를 포함한다. 이는 상동성 재조합 또는 다른 적절한 공정을 사용하여 수행할 수 있다. 일부 실시태양에서, 상기 방법은 b) (예를 들어, 플라스미드에서 양성 선택 마커에 기초하여) 양성 선택을 사용하여 통합 이벤트를 갖는 균주를 선택하는 단계를 추가로 포함한다. 일부 실시태양에서, 방법은 (예를 들어, 카운터 선택 마커 유전자에 기초하여) 음성 선택을 사용하여 루프 아웃된 플라스미드 백본을 갖는 균주를 선택하는 단계를 추가로 포함한다. 일부 실시태양에서, 생성된 사카로폴리스포라 균주는 통합된 DNA가없는 부모 균주에 비해 더 우수한 성능을 갖는다. 일부 실시태양에서, 카운터 선택 마커는 sacB 유전자 또는 pheS 유전자이다.

[0685] 레반수크라제(EC 2.4.1.10)는 화학 반응을 촉매하는 효소이다

[0686] 수크로스 + (2,6-베타-D-프룩토실)n ⇌ {\displaystyle \rightleftharpoons} \rightleftharpoons 글루코스 + (2,6-베타-D-프룩토실)n+1

[0687] 이 효소의 2가지 기질은 수크로스 및 (2,6-베타-D-프룩토실)n인 반면, 2개의 생성물은 글루코오스 및 (2,6-베타-D-프룩토실)n+1이다. 이 효소는 글리코실 트랜스퍼라제 패밀리, 특히 헥소실트랜스퍼라제에 속한다. 이 효소 부류의 체계적인 명칭은 수크로스:2,6-베타-D-프룩탄 6-베타-D-프룩토실트랜스퍼라제이다. 통상적으로 사용되는 다른 명칭은 수크로스 6-프룩토실 트랜스퍼라제, 베타-2,6-프룩토실 트랜스퍼라제 및 β-2,6-프룩탄:D-글루코스 1-프룩토실 트랜스퍼라제를 포함한다.

[0688] ***사카로폴리스포라 균주의 상처없는 표적 게놈 편집***

[0689] 또한, 사카로폴리스포라 스피노스 균주와 같은 사카로폴리스포라 균주에서 표적화된 게놈 편집 방법이 제공된다. 상기 방법은 표적화된 게놈 유전자좌에서 유전자 변이를 함유하는 상처없는 사카로폴리스포라 균주를 초래한다.

[0690] 일부 실시태양에서, 방법은 (a) 게놈 편집 플라스미드를 사카로폴리스포라 균주 내로 도입하는 단계를 포함한다. 상기 게놈 편집 플라스미드는 (1) 선택 마커; (2) 카운터 선택 마커, (3) 게놈으로 도입될 하나 이상의 원하는 유전자 변형이 있는 DNA 단편; 및 (4) 플라스미드 백본 서열을 포함한다. 일부 실시태양에서, 하나 이상의 원하는 유전자 변이를 보유하는 DNA 단편은 표적 유전자좌에서 사카로폴리스포라 게놈에 통합될 하나 이상의 유전자 변이, 및 원하는 유전자 변이의 측면에 있는(flanking) 표적 게놈 유전자 유전자좌에 대한 상동성 암을 포함한다.

[0691] 일부 실시태양에서, 상기 방법은 (b) 초기 상동성 재조합을 겪고 게놈에서 선택 마커의 존재에 기초하여 표적 유전자좌에 통합된 유전자 변이를 갖는 사카로폴리스포라 균주를 선택하는 단계를 추가로 포함한다.

[0692] 일부 실시태양에서, 상기 방법은 (c) 표적 유전자좌에 통합된 유전자 변이를 갖지만 카운터 선택 마커의 부재에 기초하여 플라스미드 백본을 루프 아웃시키는 추가적인 상동성 재조합을 겪은 사카로폴리스포라 균주를 선택하는 단계를 추가로 포함한다. 일부 실시태양에서, 카운터 선택 마커는 본 발명에 기재된 것으로부터 선택된다.

[0693] 일부 실시태양에서, 방법의 단계 (b) 및 단계 (c)는 동일한 매질에서 동시에 수행된다. 일부 실시태양에서, 방법의 단계 (b) 및 단계 (c)는 별도의 매질에서 순차적으로 수행된다.

[0694] 일부 실시태양에서, 표적화된 게놈 유전자 유전자좌는 코딩 DNA 모듈의 반복 세그먼트를 함유하지 않는 게놈 영역을 포함하는 사카로폴리스포라 게놈의 임의의 영역을 포함할 수 있다.

[0695] 일부 실시태양에서, 게놈 편집 플라스미드는 사카로폴리스포라 균주에서 기능성인 온도 민감성 레플리콘을 포함하지 않는다.

[0696] 일부 실시태양에서, 게놈 편집 플라스미드는 사카로폴리스포라 균주 내에서 플라스미드의 자기 복제를 가능하게 하는 복제 기점을 포함하지 않는다.

[0697] 일부 실시태양에서, 선택 단계 (c)는 통합된 플라스미드의 복제없이 수행된다.

[0698] 일부 실시태양에서, 사카로폴리스포라 균주에서의 게놈 편집 플라스미드는 본 발명에 기재된 바와 같은 컨쥬게이션 방법을 사용하여 사카로폴리스포라 균주에 도입된다. 일부 실시태양에서, 게놈 편집 플라스미드를 전달하는 공여자 세포는 대장균 세포이다. 일부 실시태양에서, 수용자 세포는 사카로폴리스포라 스피노사 세포이다. 대안적으로, 일부 실시태양에서, 게놈 편집 플라스미드는 사카로폴리스포라 균주로 직접 형질전환된다.

[0699] 다양한 상동성 재조합 플라스미드가 사용될 수 있다. 일부 실시태양에서, 게놈 편집 플라스미드는 단일 상동성 재조합 벡터이다. 단일 상동성 재조합 플라스미드는 "삽입 카세트"를 포함할 수 있다. 삽입 상동성 재조합 카세트는 단일 상동성 재조합 교차 이벤트를 촉진하는 표적 부위에 충분한 서열 동일성을 공유하는 단일 영역을 포함한다. 특정 실시태양에서, 삽입 카세트는 관심 폴리뉴클레오티드를 추가로 포함한다. 단일 교차 이벤트만 발생하면, 전체 삽입 카세트-및 포함된 플라스미드/벡터-가 표적 부위에 통합된다. 이러한 삽입 카세트는 일반적으로 원형 벡터/플라스미드에 함유된다. 미국 특허 공개 2003/0131370, 2003/0157076, 2003/0188325, 및 2004/0107452, Thomas et al. (1987)*Cell* 51:503-512, 및 Pennington et al.(1991) *Proc. Natl. Acad. Sci. USA* 88:9498-9502를 참조하며, 이들 각각은 그 전문이 본 발명에 참조로 포함된다.

[0700] 일부 실시태양에서, 게놈 편집 플라스미드는 이중 상동성 재조합 벡터이다. 예를 들어, 상동성 재조합 카세트는 "교체 벡터"를 포함한다. 교체 상동성 재조합 카세트는 진핵세포 내 표적 부위의 상응하는 제 1 및 제 2 영역과 충분한 서열 동일성을 갖는 제 1 및 제 2 영역을 포함한다. 이중 상동성 재조합 교차 이벤트가 발생하고, 제 1 및 제 2 영역 내부의 임의의 폴리뉴클레오티드가 표적 부위에 통합된다(즉, 카세트의 제 1 상동성 영역과 표적 부위의 상응하는 제 1 영역 사이의 상 동성 재조합 및 재조합 카세트의 제 2 상동성 영역과 표적 부위의 상응하는 제 2 영역 사이의 상동성 재조합). Yang et al. (2014) *Applied and Environmental Microbiology* 80:3826-3834, Posfai et al. (1999) *Nucleic Acids Research* 27(2):4409-4415; Graf et al. (2011) *Applied and Environmental Microbiology* 77:5549-5552를 참조하며, 이들 각각은 그 전문이 본 발명에 참조로 포함된다.

[0701] ***원형질체화 방법***

[0702] 한 실시태양에서, 본 발명에 제공된 방법 및 시스템은 사상균류 세포로부터의 원형질체의 생성을 이용한다. 원형질체의 제조에 적합한 절차는 예를 들어 EP 238,023 및 Yelton et al.(1984, Proc. Natl. Acad. Sci. USA 81:1470-1474)에 기술된 것을 포함하는 당업계에 공지된 임의의 방법일 수 있다. 한 실시태양에서, 원형질체는 사상균류 세포의 배양물을 하나 이상의 용균 효소 또는 이의 혼합물로 처리함으로써 생성된다. 용균 효소는 베타-글루카나아제 및/또는 폴리갈락투로나제일 수 있다. 한 실시태양에서, 원형질체 생성용 효소 혼합물은 비노테이스트(VinoTaste) 농축물이다. 효소 처리 후, 원형질체는 예를 들어 원심분리와 같은 당업계에 공지된 방법을 사용하여 분리될 수 있다.

[0703] 전 배양 및 실제 원형질화 단계는 원형질체의 수 및 형질전환 효율을 최적화하도록 변형될 수 있다. 예를 들어, 접종물 크기, 접종물 방법, 전 배양 배지, 전 배양시간, 배양 전 온도, 혼합 조건, 세척 완충액 조성물, 희석비, 용균 효소 처리 동안 완충제 조성물 사용된 용균 효소의 형태 및/또는 농도, 용균 효소에 의한 배양시간, 원형질체 세척 절차 및/또는 완충액, 실제 형질전환 동안 원형질체 및/또는 폴리뉴클레오티드 및/또는 형질전환 시약의 농도, 형질전환 동안 물리적 매개변수, 얻어진 형질전환체까지의 형질전환 이후 절차의 변형이 존재할 수 있다.

[0704] 본 발명은 또한 2종 이상의 미생물 균주에서 유전자 변화의 신속한 통합 및 원형질체 융합에 기초하여 사카로폴리스포라 종에서 유전자 다양성을 생성하는 방법을 제공한다. 일부 실시태양에서, 미생물 균주의 적어도 하나가 "마킹된" 돌연변이를 함유하는 경우, 방법은 다음 단계를 포함한다: (1) 통합을 위해 조작된 균주의 풀로부터 부모 균주를 선택하는 단계; (2) 통합될 균주로부터 원형질체를 준비하는 단계(예를 들어, 세포벽 제거 등); 및 (3) 관심 균주를 융합시키는 단계; (4) 세포를 회수하는 단계; (5) "마킹된" 돌연변이를 갖는 세포를 선택하는 단계, 및 (6) 다른 부모 균주에 대해 오는 돌연변이의 존재에 대해 성장 세포를 유전자형 분석(genotyping)하는 단계. 선택적으로, 방법은 (7) "마킹된" 돌연변이로부터 플라스미드를 제거하는 단계를 추가로 포함한다. 일부 실시태양에서, 미생물 균주 중 어느 것도 "마킹된" 돌연변이를 함유하지 않는 경우, 방법은 다음 단계를 포함한다: (1) 통합을 위해 조작된 균주의 풀로부터 부모 균주를 선택하는 단계; (2) 통합될 균주로부터 원형질체를 준비하는 단계(예를 들어, 세포벽 제거 등); 및 (3) 관심 균주를 융합시키는 단계; (4) 세포를 회수하는 단계; (5) 제 1 부모 균주에 대해 오는 돌연변이의 존재에 대해 세포를 선택하는 단계, 및 (6) 다른 부모 균주에 대해 오는 돌연변이의 존재에 대해 세포를 선택하는 단계. 일부 실시태양에서, 균주는 제 1 부모 균주 및/또는 다른 부모 균주로부터 오는 돌연변이와 관련된 표현형에 기초하여 선택된다. 일부 실시태양에서, 균주는 유전자형 분석에 기초하여 선택된다. 일부 실시태양에서, 유전자형 분석 단계는 고 처리량 공정으로 수행된다.

[0705] 본 발명에 기재된 바와 같은 방법은 전통적인 방법과 비교하여 극히 효율적이다. 예를 들어, 사카로폴리스포라 종에서 돌연변이를 조합하는 전통적인 방법은 통합 및 카운터 선택(~45일)을 통해 제 1 돌연변이를 기본 균주로 생성함으로써 돌연변이 균주(예를 들어, Mut1)를 생성한 다음 Mut1 균주를 수용자로 사용하여 다음 돌연변이로 과정을 반복하고, 45일의 조작 공정을 다시 거침으로써 2개의 돌연변이(예를 들어, Mut2)를 갖는 새로운 균주를 생성한다. 그러나, 본 발명의 방법은 동일한 균주에 도달하기 위해 단지 약 14일, 15일, 16일, 17일, 18일, 19일, 20일 또는 21일 미만을 필요로 한다.

[0706] 일부 실시태양에서, 단계 (3)에서, 유용한 (신규) 돌연변이체의 조합을 생성할 가능성을 증가시키기 위해, "마킹된" 돌연변이를 갖는 균주의 더 적은 세포가 사용될 수 있으며, 따라서 이들 "마킹된" 세포가 다른 돌연변이를 가진 세포와 상호작용하고 융합했을 가능성이 증가된다. 일부 실시태양에서, "마킹된" 돌연변이를 갖는 균주의 세포 대 "마킹되지 않은" 돌연변이를 갖는 균주의 세포의 비는 약 1:1.5, 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:15, 1:20, 1:25, 1:30, 1:35, 1:40, 1:45, 1:50, 1:55, 1:60, 1:65, 1:70, 1:75, 1:80, 1:85, 1:90, 1:95, 1:100, 1:200, 1:300, 1:400, 1:500, 1:600, 1:700, 1:800, 1:900, 1:1000 또는 그 이상이

다.

[0707] 일부 실시태양에서, 단계 (4)에서, 한천 오버레이를 사용하지 않고 세포를 삼투적으로 안정화된 매질 상에 플레이팅하여, 공정을 단순화하고 보다 용이한 자동화를 가능하게 한다. 삼투-안정화제는 카운터 선택 마커 유전자(예를 들어, sacB 유전자)를 함유할 수 있는 세포의 성장을 가능하게 하는 것이다. 원형질체 세포는 치료에 매우 민감하고 죽이기 쉽다. 이 단계는 충분한 셀이 회복되는 것을 보증한다. 이 단계가 잘 작동할수록 다운스트림 분석에 더 많은 재료를 사용할 수 있다.

[0708] 일부 실시태양에서, 단계 (5)에서, 단계는 적절한 항생제를 성장 세포 상에 오버레이함으로써 달성된다. 부모 세포 중 어느 것도 "마킹된" 돌연변이를 보유하지 않는 경우, 균주는 관심 균주를 확인하기 위한 다른 수단에 의해 유전자형 분석될 수 있다. 이 단계는 선택적일 수 있지만, 이는 세포 융합이 일어날 가능성이 가장 높은 세포가 풍부하게 되는 것을 보증한다. 다중 유전자좌를 "마킹"하는 것이 가능하며 이러한 방식으로 관심 조합을 더 빨리 생성할 수 있지만, 이후 "상처없는(scarless)" 균주를 원한다면 다중 플라스미드를 제거해야 할 수 있다.

[0709] 일부 실시태양에서, 단계 (6)에서, 유전자형에 대한 콜로니의 수는 선택 계획뿐만 아니라 교차(cross)의 복잡성에 의존한다.

[0710] 일부 실시태양에서, 단계 (7)은 선택적이며 추가 검증 또는 클라이언트 전달에 권장된다. 일부 실시태양에서, 균주에 대한 조작 사이클의 끝에서, 모든 플라스미드 잔재물을 제거할 필요가 있다. 이것이 언제 그리고 얼마나 자주 수행되는지는 사용자의 재량에 달려 있다. 일부 실시태양에서, 카운터 선택 가능한 sacB 유전자의 존재는 이 단계를 간단하게 만든다. 일부 실시태양에서, 균주의 적어도 하나는 "마킹된" 돌연변이를 갖는다. 일부 실시태양에서, 단일 통합 단계 동안 융합된 균주의 수는 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500 또는 그 이상과 같이 2개 이상일 수 있다. 일부 실시태양에서, 융합을 위한 하나 이상의 균주는 관심 부위에서 선택 마커에 의해 태그될 수 있다. 일부 실시태양에서, 부모 균주 중 하나가 "마킹된" 유전자 돌연변이를 포함하는 반면, 다른 부모 균주의 유전자 돌연변이는 마킹되지 않은 경우, 마킹되지 않은 균주 대 마킹된 균주의 비는 약 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, 150:1, 200:1, 250:1, 300:1, 또는 그 이상이다. 일부 실시태양에서, 부모 집단이 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50개 또는 그 이상의 마킹되지 않은 균주를 갖는 경우, 각각 동일한 비율이 사용된다. 일부 실시태양에서, 살아있는 마킹되지 않은 균주 및 죽은 마킹된 균주를 사용하는 경우, 생:사의 비율은 약 1:1 또는 약 1:2 (생:사)이다.

[0711] 본 발명의 방법은 이전에 기술된 방법(Practical Streptomyces Genetics, ISBN 0-7084-0623-8)과 비교하여 중요한 개량을 포함한다. 이러한 개량에는 다음이 포함되지만 이에 국한되지는 않는다:

[0712] ● 원형질체 생성을 위한 초기 원심분리는 보다 짧은시간 5분 대 10분 동안 고속(5000xg 대 100xg)으로 수행된다. 이는 프로토콜을 완료하는데 필요한 시간을 단축시킨다.

[0713] ● 일부 실시태양에서, 변형된 조성물을 갖는 YEME 매질은 sacB 유전자를 갖는 균주의 사용을 수용하기 위해 사용된다. 전형적인 YEME 조성물은 수크로스를 포함하며, 이는 sacB 유전자를 갖는 균주에 의해 허용되지 않는다. 본 발명의 변형된 YEME 매질은 1M 소르비톨을 갖는 수크로스를 대체한다;

[0714] ● 일부 실시태양에서, 균사체를 원형질체로부터 분리하기 위해 소화된 세포를 면모를 통해 통과시키는 여과 단계가 없다. 일부 실시태양에서, 효소 처리 후 남은 균사체가 없으므로 이 단계는 필요하지 않다;

[0715] ● 일부 실시태양에서, 원형질체는 Practical Streptomyces Genetics (ISBN 0-7084-0623-8)에서 권장되는 부피의 약 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:15, 1:20 또는 그 미만으로 재현탁되어 후속 스핀 단계를 제거하고 자동화를 용이하게한다;

[0716] ● 일부 실시태양에서, 융합된 원형질체는 탑-한천이 아닌 R2YE 브로스에서 회수된다. 이는 자동화 및 취급을 크게 단순화시킨다. 한천은 팁을 굳히고 막을 수 있으며 프로토콜 중에 웜을 유지하기 위해 필요하다. 브로스는 이들 합병증을 가지지 않는다. 이러한 변형은 원형질체 생존성을 크게 감소시키지 않는다.

[0717] ● 일부 실시태양에서, 원형질체는 0.5M 소르비톨 및 0.5M 만노스가 보충된 R2YE 매질에서 회수된다. 이 제제는 개발에 시간과 실험이 필요하였다. 본 발명자들은 원래 1M 또는 0.5M에서 소르비톨만을 사용하려고 시도했지만, 원형질체를 안정화시키는데 효과적이지 않았고, 1M 소르비톨의 존재 하에서 세포가 느리게 성장 하였다.

그러나, 본 발명자들은 매질에 소르비톨 및 만노스(각 0.5M)가 보충되면 삼투 안정화 매질로서 더 잘 작용한다는 것을 발견하였다.

[0718] 일부 실시태양에서, 단계 (2)에서, 세포벽은 리소자임 처리에 의해 제거된다. 일부 실시태양에서, 멸균 P-완충액 중 약 1mg/ml, 2mg/ml, 3mg/ml, 4mg/ml, 5mg/ml, 6mg/ml, 7mg/ml, 8mg/ml, 9mg/ml 또는 10mg/ml 리소자임이 사용된다. 일부 실시태양에서, 총 배양시간은 37℃에서 약 70분, 75분, 80분, 85분, 90분, 95분 또는 100분이다. 일부 실시태양에서, 생성된 원형질체는 이들이 물에 의해 용해되는지 여부를 평가함으로써 검증된다. 일부 실시태양에서, 현미경 및 삼투-안정화된 매질에서의 성장에 의해 물 민감성을 결정할 수 있다.

[0719] ***숙주 세포의 형질전환***

[0720] 일부 실시태양에서, 본 발명의 벡터는 형질전환, 형질감염, 형질도입, 바이러스 감염, 유전자 총 또는 Ti 매개 유전자 전달을 포함하는 임의의 다양한 기술을 사용하여 숙주 세포에 도입될 수 있다(Christie , PJ, and Gordon, JE, 2014 "Agrobacterium Ti Plasmids"Microbiol SPectr. 2014; 2(6); 10.1128 참조). 특정 방법은 인산 칼슘 형질감염, DEAE-덱스트란 매개 형질감염, 리포펙션 또는 전기천공을 포함한다(Davis, L., Dibner, M., Battey, I. 1986 "Basic Methods in Molecular Biology"). 형질전환의 다른 방법은 예를 들어, 아세트산리튬 형질전환 및 전기천공을 포함한다. 예를 들어, Gietz *et al.*, Nucleic Acids Res. 27:69-74(1992); Ito *et al.*, J. Bacterol. 153:163-168(1983); 및 Becker and Guarente, Methods in Enzymology 194:182-187(1991) 참조. 일부 실시태양에서, 형질전환된 숙주 세포는 재조합 숙주 균주로 불린다.

[0721] 일부 실시태양에서, 본 발명은 본 발명의 96-웰 플레이트 로봇 플랫폼 및 액체 처리 기계를 사용하는 세포의 고처리량 형질전환을 교시한다.

[0722] 일부 실시태양에서, 본 발명은 상기한 바와 같이 하나 이상의 선택 마커로 형질전환된 세포를 선별하는 것을 교시한다. 이런 한 실시태양에서, 카나마이신 저항성 마커(KanR)를 포함하는 벡터로 형질전환된 세포는 유효량의 카나마이신 항생제를 함유하는 배지 상에 덮인다. 카나마이신-레이스(kanamycin-laced) 배지에서 볼 수 있는 콜로니 형성 단위는 벡터 카세트를 게놈에 통합시킨 것으로 추정된다. 원하는 서열의 삽입은 PCR, 제한 효소 분석 및/또는 관련 삽입 위치의 서열화를 통해 확인될 수 있다.

[0723] ***선택된 서열의 루핑 아웃***

[0724] 일부 실시태양에서, 본 발명은 숙주 유기체로부터 DNA의 선택된 영역을 루핑 아웃하는 방법을 교시한다. 루핑 아웃 방법은 Nakashima et al. 2014 "Bacterial Cellular Engineering by Genome Editing and Gene Silencing." Int. J. Mol. Sci. 15(2), 2773-2793에 기술될 수 있다. 일부 실시태양에서, 본 발명은 양성 형질전환 체로부터 선택 마커를 루핑 아웃하는 것을 교시한다. 루핑 아웃 결실 기술은 당해 분야에 공지되어 있으며(Tear et al. 2014 "Excision of Unstable Artificial Gene-specific inverted Repeats Mediates Scar-Free Gene Deletions in Escherichia coli." Appl. Biochem. Biotech. 175:1858-1867)에 기술된다. 본 발명에서 제공된 방법에 사용된 루핑 아웃 방법은 단일-교차 상동성 재조합 또는 이중-교차 상동성 재조합을 사용하여 수행될 수 있다. 한 실시태양에서, 본 발명에 기술된 바와 같이 선택된 영역을 루핑하는 것은 본 발명에 기술된 바와 같이 단일-교차 상동성 재조합을 사용하는 것을 수반할 수 있다.

[0725] 먼저, 루프 아웃 벡터가(예를 들어 상동성 재조합, CRISPR 또는 다른 유전자 편집 기술을 통해) 숙주 유기체의 게놈 내의 선택된 표적 영역 내로 삽입된다. 한 실시태양에서, 단일-교차 상동성 재조합이 도 3에 도시된 바와 같은 원형 플라스미드 또는 벡터를 루프-인시키기 위해 원형 플라스미드 또는 벡터와 숙주 세포 게놈 사이에 사용된다. 삽입된 벡터는 현존하는 또는 숙주 서열 근처에 도입된 직접 반복체인 서열을 갖도록 디자인되어, 직접 반복체가 루핑 및 결실이 예정된 DNA의 영역에 인접하게 된다. 일단 삽입되면, 루프 아웃 플라스미드 또는 벡터는 선택 영역의 결실을 위해 카운터 선택될 수 있다(예를 들어, 도 4 참조; 선택 유전자에 대한 저항성의 결핍).

[0726] 당업자는 루프 아웃 절차의 설명이 게놈으로부터 원하지 않는 영역을 삭제하는 하나의 예시적인 방법을 나타내는 것을 인식할 것이다. 실제로, 본 발명의 방법은 CRISPR, TALENS, FOK 또는 다른 엔도뉴클레아제를 통한 유전자 편집을 포함하나 이에 제한되지 않는 게놈 결실에 대한 임의의 방법과 양립될 수 있다. 당업자는 상동성 재조합 기술을 통해 게놈의 원하지 않는 영역을 교체할 수 있는 능력을 인식할 것이다.

[0727] **중성 통합 부위**

[0728] 외래 유전자 및 심지어 전체 경로는 종종 섀시(chassis) 유기체로 이식되며, 플라스미드-기반 발현 또는 게놈

통합을 위한 중성 부위의 식별이 요구된다. 게놈 통합이 플라스미드 기반 발현에 비해 더욱 안정적이고 예측 가능하기 때문에, 이는 종종 산업용 미생물 균주에 대해 바람직한 변형 방법이다.

[0729] 이들 중성 통합 부위는 개별 유전자 또는 다중 유전자 카세트가 사카로폴리스포라 종 균주와 같은 미생물 균주의 게놈 내에 안정하고 효율적으로 통합될 수 있는 유전자좌이다. 이들 부위로의 서열의 통합은 균주의 성장에 영향을 미치지 않거나 제한적이다. 본 발명에 사용된 "중성 통합 부위"는 미생물 세포의 염색체에 본래 존재하는 유전자 또는 염색체 유전자좌를 지칭하며, 이의 일반적인 기능은 세포의 성장 또는 특정 생물학적 과정에 대한 모든 기능을 수행하기 위한 세포의 모든 기능 수행 능력에 필요하지 않다. 그 유전자 내에 정상적으로 존재하지 않는 DNA 서열의 통합에 의해 파괴될 때, 파괴된 중성 통합 부위 유전자를 보유하는 세포는 생물학적 과정을 생산적으로 수행할 수 있다.

[0730] 일부 실시태양에서, 본 발명은 S. 스피노사에서 중성 통합 부위(NIS)를 제공한다. 이러한 중성 통합 부위에는 서열번호 132 내지 서열번호 142 중 어느 하나의 서열을 갖는 유전자좌가 포함되나 이에 제한되지는 않는다. 이들 NIS는 모든 사카로폴리스포라 종 중에서 보존적일 수 있다. 따라서, S. 스피노사 이외의 S. 스피노사의 NIS와 상동성을 공유하는 사카로폴리스포라 종에서의 유전자좌도 또한 잠재적 중성 통합 부위이다.

[0731] 이러한 중성 통합 부위는 다수의 도구를 갖는다. 예를 들어, 비교적 큰 크기를 갖는 외인성 DNA 단편이 본 발명에 기재된 단일 중성 통합 부위에 삽입될 수 있다. 이러한 DNA 단편은 숙주 세포의 성장에 영향을 미치지 않으면서 적어도 5kb, 6kb, 7kb, 8kb, 9kb, 10kb, 11kb, 12kb, 13kb, 14kb, 15kb, 16kb, 17kb, 18kb, 19kb, 20kb, 25kb, 30kb, 40kb, 50kb, 60kb, 70kb, 80kb, 90kb, 100kb 또는 그 이상의 크기를 가질 수 있다.

[0732] NIS에 통합될 DNA 단편은 임의의 원하는 서열일 수 있다. 통합될 이러한 DNA 단편은 숙주 세포에 새로운 기능을 제공하거나, 숙주 세포의 기존 기능을 향상시키거나, 숙주 세포 성장에 부정적인 영향을 미칠 수 있는 임의의 인자의 효과를 감소시킬 수 있다. 예를 들어, 중성 통합 부위(들) 내로 삽입된 유전자 요소(들)를 갖는 사카로폴리스포라 종 균주는 삽입을 갖지 않는 기준 균주에 비해 성능이 향상 될 수 있다(예를 들어, 스피노신과 같은 하나 이상의 관심 분자의 수율 향상).

[0733] 일부 실시태양에서, 통합될 DNA 단편은 숙주 세포에 대해 상동성 및/또는 이종성 서열을 포함한다. 일부 실시태양에서, 통합될 DNA 단편은 숙주 세포에서 기능하는 선택된 프로모터를 포함한다. 일부 실시태양에서, 통합될 DNA 단편은 숙주 세포에서 기능하는 선택된 터미네이터 서열을 포함한다. 일부 실시태양에서, 프로모터 및 터미네이터 서열은 본 발명에 기재된 임의의 서열 또는 당해 분야에 공지된 서열일 수 있다.

[0734] 일부 실시태양에서, 통합될 DNA 단편은 통합 DNA 단편을 포함하는 세포를 선택하는데 사용될 수 있는 하나 이상의 선택 마커를 포함한다. 일부 실시태양에서, 통합될 DNA 단편은 카운터 선택 마커를 포함하며, 이는 통합 DNA 단편의 전체 또는 일부의 루프-아웃을 용이하게 하기 위해 사용될 수 있다.

[0735] 일부 실시태양에서, 본 발명에 기술된 바와 같이, 하나 이상의 외인성 유전자가 사카로폴리스포라 종의 NIS에 통합되어 신규한 경로를 확립하는 것과 같이 미생물 종에 신규한 기능을 도입할 수 있다. 일부 실시태양에서, 이러한 신규 경로는 천연 숙주 세포에 존재하지 않는 합성 경로 및/또는 신호 전달 경로이다. 일부 실시태양에서, 통합될 DNA 단편은 인테그라제에 대한 부착 부위를 포함하여 생합성 경로 또는 이의 성분의 후속적이고 효율적인 표적화된 통합을 가능하게 한다. 일부 실시태양에서, 2차 대사산물에 대한 생합성 경로의 일부인 하나 이상의 유전자 산물(들)을 코딩하는 유전자 클러스터의 일부 또는 전체 유전자 클러스터를 포함하는 DNA 단편은 본 발명의 NIS에 통합된다. 2차 대사산물은 종종 초식 및 다른 종간 방어에 대한 식물 방어에 중요한 역할을 한다. 2차 대사산물은 복잡한 미생물 환경에서 영양소와 서식지에 대한 투쟁에서 역할을 할 수 있다. 일부 실시태양에서, 2차 대사산물은 경쟁하는 박테리아, 진균, 효모 또는 다른 유기체에 대한 생물학적 활성을 갖는다. 일부 실시태양에서, 2차 대사산물은 경쟁자의 영양 섭취 효소의 억제제로서 작용하거나, 항균 또는 항진균 활성을 직접적으로 나타낸다. 일부 실시태양에서, 2차 대사산물은 경쟁자의 방어 메커니즘에 대응하고, 또 다른 대사산물은 경쟁자의 공격 메커니즘에 대응한다. 2차 대사산물은 화학적 특성면에서 엄청난 다양성을 나타내는 것으로 잘 알려져 있다. 따라서, 인간은 일부 2차 대사산물을 의약품, 향료 및 레크리에이션 약물로 사용한다. 2차 대사산물은 다음과 같은 범주로 나눌 수 있다: 베타-락탐, 알칼로이드, 테르페노이드, 글리코사이드, 천연 페놀, 페나진, 비페닐 및 디벤즈퓨란과 같은 작은 "소분자"; 폴리케타이드, 복합 글리코사이드, 비리보솜 펩티드 및 상기 3가지의 하이브리드와 같은 큰 모듈형 "분자 공장"에 의해 생성된 큰 "소분자"; 및 비-"소분자" - DNA, RNA, 리보솜, 또는 리보솜 펩티드와 같은 다당류 "고전" 바이오폴리머.

[0736] 일부 실시태양에서, 본 발명의 NIS는 벡터에 통합될 수 있다. "벡터"는 플라스미드, 파지, 박테리아 인공 염색

체(BAC) 또는 코스미드와 같은 레플리콘으로, 다른 DNA 세그먼트(예를 들어, 외래 유전자)가 부착된 세그먼트의 복제를 야기하기 위해 통합될 수 있어, 도입된 서열의 발현을 초래한다. 벡터는 도입된 DNA와 이종이지만 숙주 세포에 의해 인식되고 사용되는 프로모터 및 하나 이상의 제어 요소(예를 들어, 인핸서 요소)를 포함할 수 있다. 일부 실시태양에서, 상기 벡터는 상이한 미생물 종의 게놈에 추가로 혼입되어 상이한 미생물 종에서 NIS를 확립할 수 있다. 예를 들어, 본 발명에 기재된 사카로폴리스포라 스피노사의 NIS는 관련 사카로폴리스포라 종의 게놈에 통합될 수 있다.

[0737] ***인테그라제***

[0738] "인테그라제"라는 효소는 2개의 부착(att) 부위(일반적으로 숙주 염색체에서 tRNA 유전자 내에 위치한 보존된 뉴클레오티드 서열)를 인식하고, 2개의 DNA 분자를 결합시키고 DNA 이중 가닥 파괴를 촉매한다. 재결합 이벤트는 DNA 분자 중 하나의 수용자 세포의 다른 DNA로의 통합을 초래한다(N. D. Grindley, K. L. Whiteson, P. A. Rice, 2006. Annu. Rev. Biochem. 75, 567-605.). 따라서, 인테그라제는 보존된 부위에서의 인식 및 부착을 통해 DNA 페이로드의 표적 통합을 지시할 수 있다.

[0739] 본 발명은 공여자 유기체로부터 숙주 세포 내로 DNA 단편의 표적화된 클로닝 및/또는 전사를 위한 조성물 및 방법을 제공한다. 일부 실시태양에서, 변형될 숙주 세포는 주어진 인테그라제에 의해 인식될 수 있는 att 부위와 동일하거나 상동성을 갖는 서열을 포함한다. 일부 실시태양에서, 변형될 숙주 세포는 주어진 인테그라제에 의해 인식될 수 있는 att 부위와 동일하거나 상동성을 갖는 서열을 포함하지 않는다. 제 2 시나리오에서, att 부위와 동일하거나 att 부위와 상동성을 갖는 서열은 먼저 본 발명에 기재된 NIS와 같은 숙주 세포의 중성 통합 부위에 삽입될 수 있다.

[0740] 일부 실시태양에서, 인테그라제는 사카로폴리스포라 종으로부터 유래된다. 일부 실시태양에서, 인테그라제는 S. 엔도피티카, S. 에리트라에아 또는 S. 스피노사로부터 유래된다. 일부 실시태양에서, 인테그라제는 서열번호 85, 87, 89, 91, 93의 서열 또는 이의 임의의 기능적 변이체를 포함한다.

[0741] 일부 실시태양에서, 인테그라제는 사카로폴리스포라 종으로부터 유래된 att 부위를 인식한다. 일부 실시태양에서, att 부위는 S. 엔도피티카, S. 에리트라에아 또는 S. 스피노사로부터 유래된다. 일부 실시태양에서, 인테그라제 부착 부위는 서열번호 167 내지 171의 서열, 또는 이의 임의의 기능적 변이체를 포함한다.

[0742] 일부 실시태양에서, 숙주 세포의 게놈에 통합될 DNA 단편은 5kb, 6kb, 7kb, 8kb, 9kb, 10kb, 11kb, 12kb, 13kb, 14kb, 15kb, 16kb, 17kb, 18kb, 19kb, 20kb, 25kb, 30kb, 40kb, 50kb, 60kb, 70kb, 80kb, 90kb, 100kb 또는 그 이상의 크기를 갖는다.

[0743] 본 발명은 외인성 DNA를 사카로폴리스포라 종과 같은 숙주 세포의 게놈에 통합시키기 위한 벡터를 제공한다.

[0744] 일부 실시태양에서, 벡터는 제거 효소(xis), 인테그라제(int) 및/또는 부착 부위(attP)를 코딩하는 서열(들)을 포함한다. 일부 실시태양에서, 상기 벡터의 서열(들)은 S. 엔도피티카로부터 유래된다. 일부 실시태양에서, 벡터는 첸(Chen) 외("Characterization of the chromosomal integration of *Saccharopolyspora* plasmid pCM32 and its application to improve production of spinosyn in *Saccharopolyspora spinosa*." Applied Microbiology and Biotechnology. PMID 26260388 DOI: 10.1007/s00253-015-6871-z)에 의해 기술된 것과 같은 pCM32에 기반한다. 일부 실시태양에서, 상기 벡터의 서열(들)은 S. 에리트라에아에서 유래된다. 일부 실시태양에서, 벡터는 테 폴(Te Poele) 외에 의해 기술된 바와 같이 pSE10 및/또는 pSE211에 기초한다("Actinomycete integrative and conjugative elements." Antonie Van Leeuwenhoek 94, 127-143).

[0745] 일부 실시태양에서, 본 발명의 벡터는 사카로폴리스포라 스피노사의 게놈에서 서열을 인식한다. 일부 실시태양에서, 본 발명의 인테그라제에 의해 인식될 수 있는 사카로폴리스포라 스피노사의 게놈에서의 서열은 서열번호 167 내지 171로부터 선택된 서열, 또는 이의 임의의 기능적 변이체를 갖는다. 일부 실시태양에서, S. 엔도피티카 및/또는 S. 에리트라에아로부터 유래된 att 부위는 사카로폴리스포라 스피노사의 게놈에 도입된다. 일부 실시태양에서, S. 엔도피티카 및/또는 S. 에리트라에아로부터 유래된 att 부위는 본 발명에 기재된 것과 같은 사카로폴리스포라 스피노사의 NIS에 도입된다.

[0746] 인테그라제를 사용하기 위한 추가 도구 및 방법은 WO/2001/051639A2, WO/2013/189843A1, WO/2001/087936A2, WO/2001/083803A1, WO/2001/075116A2, 및 미국 특허 번호 6569668에 기재되며, 각각의 전문이 본 발명에 참조로 포함된다.

[0747] ***복제 기점***

[0748] 본 발명은 또한 사카로폴리스포라 스피노사와 같은 사카로폴리스포라 종에 사용될 수 있는 자기 복제 플라스미드 시스템에 대한 복제 기점 및 복제 요소를 제공한다.

[0749] 일부 실시태양에서, 자기 복제의 기점 및 요소는 사카로폴리스포라 종에서 수행될 수 있는 유전자 조작 및 선별의 유형을 향상시킨다. 일부 실시태양에서, 자기 복제의 기점은 S. 에리트라에아(서열번호 94)로부터의 추정 염색체 복제 기점으로부터 유래된다. 일부 실시태양에서, 자기 복제 기점은 S. 에리트라에아로부터의 플라스미드 pSE101 및 pSE211(각각 서열번호 95 및 서열번호 96)을 복제하는데 있어 액티노마이세트 통합 및 컨쥬게이티브 요소(AICE)로부터 유래된다. 일부 실시태양에서, 본 발명의 자기 복제의 기점은 항생제 저항성 마커를 함유하고 자기 복제에 필요한 다른 유전자의 존재 또는 부존재 하에(예를 들어, AICE의 경우) 플라스미드로 조립된다. 조립된 플라스미드는 사카로폴리스포라 종으로 전달될 수 있으며, 자기 복제 플라스미드를 갖는 형질전환체를 선택하기 위해 항생제 선택이 사용될 수 있다.

[0750] 일부 실시태양에서, 본 발명의 자기 복제 기점은 사카로폴리스포라 스피노 사와 같은 사카로폴리스포라 종에 도입될 수 있다. 일부 실시태양에서, 복제 기점을 포함하는 DNA 단편은 5kb, 6kb, 7kb, 8kb, 9kb, 10kb, 11kb, 12kb, 13kb, 14kb, 15kb, 16kb, 17kb, 18kb, 19kb, 20kb, 25kb, 30kb, 40kb, 50kb, 60kb, 70kb, 80kb, 90kb, 100kb 또는 그 이상과 같은 비교적 큰 크기를 갖는다.

[0751] 일부 실시태양에서, 사카로폴리스포라 종으로 도입될 복제 기점을 포함하는 DNA 단편은 숙주 세포에 새로운 기능을 부여하거나, 숙주 세포의 기존 기능을 향상 시키거나, 또는 숙주 세포 성장에 부정적인 영향을 미칠 수 있는 임의의 인자의 효과를 감소시킬 수 있다. 예를 들어, 게놈에 삽입된 유전자 성분(들)을 갖는 사카로폴리스포라 종 균주는 삽입을 갖지 않는 기준 균주에 비해 성능이 개량될 수 있다(예를 들어, 스피노신과 같은 하나 이상의 관심 분자의 수율 향상).

[0752] 일부 실시태양에서, 도입될 복제 기점을 포함하는 DNA 단편은 숙주 세포에 대해 상동성 및/또는 이종성 서열을 포함한다. 일부 실시태양에서, 도입될 복제 기점을 포함하는 DNA 단편은 숙주 세포에서 기능하는 선택된 프로모터를 포함한다. 일부 실시태양에서, 도입될 DNA 단편은 숙주 세포에서 기능하는 선택된 터미네이터 서열을 포함한다. 일부 실시태양에서, 프로모터 및 터미네이터 서열은 본 발명에 기재된 임의의 서열 또는 당해 분야에 공지된 서열일 수 있다.

[0753] 일부 실시태양에서, 도입될 복제 기점을 포함하는 DNA 단편은 하나 이상의 선택 마커를 포함하며, 이는 DNA 단편을 포함하는 세포를 선별하는데 사용될 수 있다. 일부 실시태양에서, 도입될 복제 기점을 포함하는 DNA 단편은 카운터 선택 마커를 포함하며, 이는 DNA 단편의 전체 또는 일부의 루프-아웃을 용이하게 하기 위해 사용될 수 있다.

[0754] 일부 실시태양에서, 하나 이상의 외인성 유전자는 사카로폴리스포라 종으로복제 기점과 함께 도입되어, 새로운 경로를 확립하는 것과 같이 미생물 종에 새로운 기능을 도입할 수 있다. 일부 실시태양에서, 이러한 신규 경로는 천연 숙주 세포에 존재하지 않는 합성 경로 및/또는 신호 전달 경로이다. 일부 실시태양에서, DNA 단편은 2차 대사산물에 대한 생합성 경로의 일부인 하나 이상의 유전자 생성물(들)을 코딩하는 유전자 클러스터의 전체 또는 일부의 유전자 클러스터를 포함한다.

[0755] ***리포터***

[0756] 사카로폴리스포라는 분자 생물학적 도구가 거의 확립되지 않은 상당히 다루기 어려운 숙주의 속이다. 이러한 도구는 공학 도구의 개발 및 공학 노력에 매우 중요하다. 본 발명은 또한 사카로폴리스포라 스피노사와 같은 사카로폴리스포라 종에 대한 리포터 단백질 및 분석을 제공한다. 따라서, 본 발명은 부족했던 리포터 시스템을 제공한다.

[0757] 일부 실시태양에서, 사카로폴리스포라 종에서 기능성인 리포터 단백질이 제공된다. 일부 실시태양에서, 리포터 단백질은 형광 단백질 및 효소 베타-글루쿠로니다제이다. 일부 실시태양에서, 형광 단백질은 녹색 형광 단백질 및 적색 형광 단백질이다. 일부 실시태양에서, 리포터 단백질은 대셔 GFP 및 파프리카 RFP(ATUM, https://www.atum.bio/products/protein-paintbox?exp=2) 및 효소 베타-글루쿠로니다제(gusA)(Jefferson et al. (1986). "Beta-Glucuronidase from Escherichia coli as a gene-fusion marker". Proceedings of the National Academy of Sciences of the United States of America. 83 (22): 8447-51.)이다.

[0758] 일부 실시태양에서, 리포터 단백질을 코딩하는 유전자는 코돈-최적화된다. 일부 실시태양에서, 형광 단백질을 코딩하는 유전자는 대장균에 대해 코돈-최적화된다. 일부 실시태양에서, 형광 단백질을 코딩하는 유전자는 서열

번호 81 또는 서열번호 82의 뉴클레오티드 서열을 갖는다. 일부 실시태양에서, 베타-글루쿠로니다제(gusA)를 코딩하는 유전자는 예를 들어 서열번호 83의 뉴클레오티드 서열을 갖는 S. 스피노사에서의 발현을 위해 코돈-최적화된다.

[0759] 일부 실시태양에서, 형광 단백질을 코딩하는 유전자는 리포터 단백질의 형광 여기 및 방출 스펙트럼을 변경하도록 변형된다.

[0760] 일부 실시태양에서, 2개 이상의 형광 단백질이 단일 사카로폴리스포라 세포에 사용된다. 일부 실시태양에서, 녹색 형광 단백질 및 적색 형광 단백질은 단일 사카로폴리스포라 세포에 사용된다. 일부 실시태양에서, 녹색 형광 리포터 단백질 및 적색 형광 리포터 단백질의 형광 여기 및 방출 스펙트럼은 서로 다르다.

[0761] 일부 실시태양에서, 본 발명의 리포터 단백질은 유전자 발현을 위한 조절 요소의 활성을 결정하기 위해 사용된다. 일부 실시태양에서, 조절 요소는 프로모터, 리보솜 결합 부위, 개시/중지 코돈, 터미네이터, 인핸서, 억제제, 단일 가닥 RNA, 이중 가닥 RNA, 요소 유사체 또는 이의 임의의 조합일 수 있다. 예를 들어, 프로모터가 본 발명의 리포터를 코딩하는 서열에 작동 가능하게 연결되어 미생물 균주로 발현되는 경우, 유전자 발현을 촉진시키는 프로모터의 강도는 형광 신호에 의해 결정될 수 있다. 유사하게, 본 발명의 리포터를 코딩하는 서열이 터미네이터 서열에 작동 가능하게 연결될 때, 유전자 발현을 억제하는 터미네이터의 강도는 형광 신호에 의해 결정될 수 있다. 따라서, 일부 실시태양에서, 리포터는 프로모터, 리보솜 결합 부위, 개시/중지 코돈, 터미네이터, 인핸서, 억제제, 단일 가닥 RNA, 이중 가닥 RNA 및 요소 유사체의 그룹의 강도를 결정하는데 유용하며, 따라서 래더(라이브러리)를 확립한다. 일부 실시태양에서, 본 발명의 리포터 단백질은 선별 도구로서 사용될 수 있다. 예를 들어, 리포터 단백질에 의해 "마킹된" 주어진 표현형을 갖는 균주는 유동 세포 분석법에 의한 것과 같은 리포터 단백질의 존재 또는 부재에 기초하여 또는 여기 스펙트럼 하에서 플레이트상에서의 관찰에 기초하여 분류될 수 있다.

[0762] 일부 실시태양에서, 본 발명의 리포터 단백질은 내인성 또는 외인성 폴리펩티드에 융합되어 사카로폴리스포라 세포에서 발현될 수 있다. 일부 실시태양에서, 리포터 단백질은 사용자가 원하는 임의의 방식으로 사용될 수 있다.

[0763] 일부 실시태양에서, 본 발명의 리포터 단백질을 코딩하는 유전자는 터미네이터 서열에 연결될 수 있다. 일부 실시태양에서, 터미네이터는 서열번호 149의 서열을 갖는다.

[0764] **실시예**

[0765] 하기 실시예는 본 발명의 다양한 실시태양을 설명하기 위해 제공되며 어떠한 방식으로든 본 발명을 제한하려는 것은 아니다. 청구항의 범위에 의해 정의된 바와 같이, 본 발명의 취지 내에 포함되는 변경 및 다른 용도는 당업자에 의해 인식될 것이다.

[0766] 아래에서 단지 독자를 돕기 위해 내용의 간략한 표가 제공된다. 이런 표의 내용의 어떤 것도 본 출원의 실시예 또는 설명의 범위를 제한하지 않는다.

표 6

[0767] 실시예 부분에 대한 내용의 표

실시예 #	제목	간략한 설명
1	사카로폴리스포라의 HTP 형질전환 & SNP 라이브러리 생성의 입증	본 발명의 **고 처리량 유전 공학 방법**의 실시태양을 기술한다.
2	HTP 게놈 공학 - 산업용 미생물 균주를 재건/개량하기 위한 SNP 라이브러리의 실행	본 발명의 **SNP 스왑** 방법을 통해 산업용 유기체를 재건시키는 접근법을 기술한다.
3	HTP 게놈 공학 - 사카로폴리스포라에서 스피노신 생산의 균주 성능을 개량시키는 SNP 스왑 라이브러리의 실행	**스피노신을 생산하는 사카로폴리스포라 균주의 성능을 개량**하기 위한 **SNP 스왑** 기술의 실행을 기술한다. 또한, 선택된 **2차 및 3차 돌연변이 통합**을 개시한다.
4	HTP 게놈 공학 - 산업용 미생물 균주를 개량하기 위한 프로모터 스왑 라이브러리의 실행	본 발명의 **PRO 스왑** 유전자 디자인 라이브러리를 통해 숙주 유기체의 균주 성능을 개량하는 방법을 기술한다.
5	HTP 게놈 공학 - 스피노신 생산을 위한 균주 성능을 개량시키기 위한 PRO 스왑 라이브러리의 실현	**스피노신을 생산하는 사카로폴리스포라 균주의 성능을 개량하기** 위한 **PRO 스왑** 기술의 구현을 기술한다.

6	상위성 매핑 - 유익한 돌연변이 통합을 예측하기 위한 알고리즘 도구	**유익한 유전자 돌연변이 통합을 예측하기 위한** 본 발명의 **자동화 도구/알고리즘**의 실시태양을 기술한다.
7	HTP 게놈 공학 - Pro 스왑 돌연변이 통합 및 다중 인자 조합 테스팅	**다수의 유전자/유전자 디자인 라이브러리 조합의 조합적 통합에 의해 생성된 큰 솔루션 공간을 효과적으로 탐색하는** 본 발명의 HTP 방법의 능력을 설명하고 예시한다.
8	HTP 게놈 공학 - 산업용 숙주 균주를 개량하기 위한 터미네이터 라이브러리의 실행	본 발명의 **STOP 스왑** 유전자 디자인 라이브러리의 응용을 설명하고 예시한다.
9	HTP 게놈 공학 - 사카로폴리스포라에서 유전자 다양성을 생성하기 위한 유전자 변화의 신속한 통합	**스피노신을 생산하는 사카로폴리스포라 균주에서** 유전적 변화의 **신속한 통합** 및 유전적 다양성을 생성하는 방법을 기술한다.
10	HTP 게놈 공학 - 사카로폴리스포라에서 사용하기 위한 리포터 단백질 및 관련 분석	사카로폴리스포라 스피노사에서 **3가지 리포터 유전자의 활용 및 정량적 평가**의 실시태양을 기술한다. 본 발명은 또한 S. 스피노사에서 GusA 발현의 정량적 평가를 가능하게 하는 비색 분석의 최적화 및 적용을 기술한다.
11	HTP 게놈 공학 - 사카로폴리스포라 스피노사에서 표적화되고 효율적인 게놈 통합을위한 인테그라제 기반 시스템	사카로폴리스포라의 게놈에 **유전자 요소의 통합을 위한 통합 기반 시스템**을 기술한다.
12	사카로폴리스포라 스피노사를 위한 자기 복제 플라스미드 시스템을 위한 복제 기점	사카로폴리스포라에서 복제 기능이 있는 **복제 기점 및 복제 요소**를 기술한다.
13	HTP 게놈 공학 - 산업용 숙주 균주 개량을위한 터미네이터 라이브러리 구현	스피노신을 생산하는 사카로폴리스포라 균주의 성능 개량을 위한 **리보솜 결합 부위** 기술의 구현을 기술한다.
14	HTP 게놈 공학 - 사카로폴리스포라의 균주 성능을 개량시키기 위한 트랜스포존 돌연변이유발 라이브러리의 구현	스피노신을 생산하는 사카로폴리스포라 균주의 성능 개량을 위한 트랜스포존 돌연변이유발 기술의 구현을 기술한다.
15	사카로폴리스포라에서 유전자 요소의 삽입을 위한 중성 통합 부위	사카로폴리스포라의 게놈에 서열의 통합을 위한 **중성 통합 부위**의 구현을 기술한다.
16	HTP 게놈 공학 - 사카로폴리스포라의 균주 성능을 개량시키기 위한 항-대사산물 선택/발효 생성물 저항성 라이브러리 구현	스피노신을 생산하는 사카로폴리스포라 균주의 성능 개량을 **항-대사산물/발효 생성물 저항성** 기술의 구현을 기술한다.
17	HTP 게놈 공학 -상처없는 돌연변이 균주를 생성하기 위해 S.스피노사에서 카운터 선택 마커로서 sacB 또는 pheS의 사용	**카운터 선택 마커로 sacB 또는 pheS를 사용하여 상처없는 돌연변이** 사카로폴리스포라 균주를 **생성하는 방법**의 구현에 대해 기술한다.
18	사카로폴리스포라의 HTP 컨쥬게이션 및 사카로폴리스포라로의 내인성 DNA의 도입 시연	본 발명의 **고 처리량 유전 공학** 방법의 실시태양을 기술한다.

[0768] **실시예 1: 사카로폴리스포라의 HTP 형질전환 & SNP 라이브러리 생성의 입증**

[0769] 이 실시예는 본 발명의 HTP 유전자 공학 방법의 실시태양을 예시한다. 숙주 세포는 다양한 크기의 다양한 SNP 서열로 형질전환되며, 모두 게놈의 다른 영역을 표적으로 한다. 결과는 본 발명의 방법이 숙주 세포의 전체 게놈을 통해 임의의 종류의 신속한 유전자 변화를 일으킬 수 있음을 입증한다.

[0770] ***A. 형질전환 벡터의 클로닝***

[0771] 다양한 SNP는 소정의 사카로폴리스포라 균주(예를 들어, 사카로폴리스포라 스피노스 균주)로부터 랜덤으로 선택될 것이고 효모 상동성 재조합 클로닝 기술을 사용하여 사카로폴리스포라 클로닝 벡터로 클로닝되어 "맞춤형 플라스미드 조립/복제" 섹션에 기술되고 도 3에 예시된 것과 같이 각각의 SNP가 직접 반복체 영역에 플랭킹하는 벡터를 조립하였다.

[0772] 이 실시예를 위한 SNP 카세트는 약 0.5Kb, 1Kb, 2Kb 및 5Kb, 또는 임의의 다른 원하는 길이의 상동성 직접 반복체 암 길이의 범위를 포함하도록 디자인될 것이다. 또한, SNP 카세트는 아래에서 보다 상세히 기술하는 바와 같이, 게놈의 다양한 구별된 영역을 표적으로 하는 상동성 재조합을 위해 디자인될 것이다. 코리네박테리움에서 입증된 예시적인 형질전환 실험에 대해서는 도 10을 참조. 그러나, 유사한 공정이 사카로폴리스포라를 위해 맞춤화되었으며 발명자들에 의해 성공적으로 수행되고 있다.

[0773] S. 스피노사 게놈은 크기가 약 8,581,920bp이고(도 9 참조), 약 8,302개의 예측된 코딩 서열(CDS)을 포함한다(Pan et al. (JOURNAL OF BACTERIOLOGY, June 2011, p. 3150-3151, doi:10.1128/JB.00344-11 참조)). 게놈은

임의의 동일한 크기의 유전자 영역으로 나눌 수 있으며, SNP 카세트는 각 영역을 표적으로 하도록 디자인될 것이다.

[0774] 각 DNA 삽입체는 상업적으로 공급된 올리고 및 주형으로서 상기한 숙주 균주 게놈 DNA를 사용하여 상동 영역의 PCR 증폭에 의해 생산될 것이다. 게놈에 도입될 SNP는 올리고 꼬리에서 암호화될 것이다. 효모에서 상동성 재조합을 사용하여 벡터 백본에 PCR 단편을 조립할 것이다.

[0775] 벡터 속으로 각 SNP 및 상동성 암의 클로닝을 도 6a-b, 도 3 및 표 5에 기술된 HTP 공학 작업 흐름에 따라 수행할 것이다.

[0776] **B. *조립된 클론의 대장균으로의 형질전환***

[0777] 벡터는 정확하게 조립된 클론을 확인하고 사카로폴리스포라 형질전환을 위한 벡터 DNA를 증폭시키기 위해 표준 열충격 형질전환 기술을 사용하여 대장균으로 처음에 형질전환시킬 것이다.

[0778] 예를 들어, 형질전환된 대장균을 조립 성공 여부에 대해 테스트할 것이다. 각각의 대장균 형질전환 플레이트로부터 콜로니를 배양하고 PCR을 통해 정확한 조립에 대해 테스트할 것이다. 이 공정을 형질전환 위치의 각각 및 상이한 삽입체 크기에 대해 반복할 것이다. 이 실험의 결과는 각 처리(삽입체 크기 및 게놈 위치)에 대해 테스트될 콜로니 중에서 확인된 정확한 콜로니의 수로 나타낼 것이다.

[0779] **C. *조립된 클론의 사카로폴리스포라로의 형질전환***

[0780] 확인된 클론을 전기천공을 통해 사카로폴리스포라 스피노사 숙주 세포로 형질전환시킬 것이다. 각각의 형질전환에 대해, DNA의 μg당 콜로니 형성 단위(CFU)의 수를 삽입체 크기의 함수로 측정할 것이다. 게놈 통합은 또한 상동성 암 길이의 함수로 분석될 것이다.

[0781] 게놈 통합 효율은 또한 사카로폴리스포라 스피노사 형질전환체에서 표적화된 게놈 위치에 대해 분석할 것이다.

[0782] **D. *선택 마커 루핑 아웃***

[0783] 삽입체 카세트의 성공적인 통합을 갖는 것으로 확인된 사카로폴리스포라의 배양물을 선택 유전자의 루프 아웃 선택에 대항하기 위해 배지에서 배양할 것이다. 이들 결과는 루프 아웃 효율이 0.5kb 내지 5kb의 상동성 암 길이 또는 다른 원하는 길이에 걸쳐 일정하게 유지되는지를 설명할 것이다.

[0784] 루프 아웃 이벤트를 추가로 검증하기 위해, 저항성을 나타내는 콜로니를 배양하고 서열화를 통해 분석할 것이다.

[0785] **실시예 2: HTP 게놈 공학 - 산업용 미생물 균주를 재건/개량하기 위한 SNP 라이브러리의 실행**이 실시예는 본 발명의 HTP 균주 개량 프로그램의 SNP 스왑 라이브러리의 여러 양태를 예시한다. 구체적으로, 이 실시예는 현재 존재하는 산업용 균주를 재건하기 위한 여러 계획된 접근법을 예시한다. 이 실시예는 "기본", "중간" 및 산업용 균주 사이에 존재할 수 있는 여러 유전자 차이에 의해 생성된 표현형 솔루션 공간을 탐색하는 웨이브 업 및 웨이브 다운 접근법을 기술한다.

[0786] **A. 다양성 풀에서 SNP 확인**

[0787] 본 발명의 방법을 사용하는 예시적인 균주 개량 프로그램을 본 발명에서 "C"로 불리는 산업용 생산 미생물 균주에 대해 수행할 것이다. 이 프로그램에 대한 다양성 풀 균주는 A, B 및 C로 표시된다. 균주 A는 임의의 돌연변이유발 이전의 원래 생산 숙주 균주를 나타내었다. 균주 C는 전통적인 균주 개량 프로그램을 통해 수년간 돌연변이유발 및 선택을 거친 현재의 산업용 균주를 나타내었다. 균주 B는 돌연변이유발을 일으킨 "중층" 균주를 나타내며, C 균주의 전신이었다.

[0788] 균주 A, B, 및 C를 서열화하고 이의 게놈을 균주 간의 유전자 차이에 대해 분석할 것이다. 모든 비 유사 SNP를 확인할 것이다. 이 중 특정 SNP는 C에 고유하고, 특정 SNP는 B와 C에 의해 추가로 공유될 것이며, 특정 SNP는 균주 B에 고유할 것이다. 이러한 SNP는 다운스트림 균주 개량 사이클의 다양성 풀로 사용될 것이다.

[0789] **B. SNP 스와핑 분석**

[0790] 실시예 2의 파트 A에서 다양성 풀로부터 확인된 SNP를 분석하여 숙주 세포 성능에 대한 이들의 영향을 결정할 것이다. 균주 성능의 초기 "학습" 라운드는 아래에 설명된 6 단계로 세분화되고 도 11에 도표로 표시된다.

[0791] *첫째*, C로부터의 모든 SNP는 개별적으로 및/또는 조합적으로 기본 A 균주 내로 클로닝될 것이다. 이 형질전환체

의 목적은 유익한 SNP를 확인하는 것이 될 것이다.

[0792] *둘째*, C로부터의 모든 SNP는 상업적 균주 C로부터 개별적으로 및/또는 조합 적으로 제거될 것이다. 이런 형질전환체의 목적은 중립적 및 해로운 SNP를 확인하는 것이 될 것이다. 추가 선택적 단계 3-6은 아래에 기술된다. 2개의 유전자 시점(기본 균주 A 및 산업용 균주 C)으로부터 SNPS를 첨가 및 제거하는 제 1 및 제 2 단계는 "웨이브 업"(기본 균주에 대한 SNP의 첨가, 첫 번째 단계) 및 "웨이브 다운"(산업용 균주로부터 SNP 제거, 두 번째 단계)를 포함하는 "웨이브"로 본 발명에서 불린다. 웨이브 개념은 SNPS의 추가 첨가/제거로 확장된다.

[0793] *셋째*, B로부터의 모든 SNP는 개별적으로 및/또는 조합적으로 기본 A 균주 내로 클로닝될 것이다. 이런 형질전환체의 목적은 유익한 SNP를 확인하는 것이 될 것다. 형질전환체의 일부는 첫 번째 단계에서 생산된 형질전환체에 대한 검증 데이터로서 사용될 것이다.

[0794] *넷째*, B로부터의 모든 SNP는 상업적 균주 B로부터 개별적으로 및/또는 조합 적으로 제거될 것이다. 이 형질전환체의 목적은 중립적 및 해로운 SNP를 확인하는 것이 될 것이다. 형질전환체의 일부는 두 번째 단계에서 생산된 형질전환체에 대한 검증 데이터로서 사용될 것이다.

[0795] *다섯째*, C에 고유한 모든 SNP(즉, B에 존재하지 않음)는 상업용 B 균주 내로 개별적으로 및 /또는 조합적으로 클로닝될 것이다. 이 형질전환체의 목적은 유익한 SNP를 확인하는 것이 될 것이다. 형질전환체의 일부는 첫 번째 및 세 번째 단계에서 생산된 형질전환체에 대한 검증 데이터로서 사용될 것이다.

[0796] *여섯째*, C에 고유한 모든 SNP는 상업용 균주 C로부터 개별적으로 및/또는 조합적으로 제거될 것이다. 이 형질전환체의 목적은 중립적 및 해로운 SNP를 확인하는 것이 될 것이다. 형질전환체의 일부는 두 번째 및 네 번째 단계에서 생산된 형질전환체에 대한 검증 데이터로서 사용될 것이다.

[0797] 이들 각각의 단계로부터 수집된 데이터는 유익한, 중립적인 또는 유해한 것으로 각 SNP를 분류하는데 사용된다.

[0798] 선택적으로, 다른 예에서, 균주 A는 원래의 생산 숙주 균주를 나타내며, 이는 이미 약간의 돌연변이유발을 가질 수 있지만 너무 많지는 않을 수 있다. 균주 C는 현재의 산업용 균주를 나타내며, 이는 전통적인 균주 개량 프로그램을 통해 수년간의 돌연변이유발 및 선택을 거쳤다. 균주 B는 "중간" 균주를 나타내며, 이는 균주 C에 비해 돌연변이유발이 훨씬 적지만 균주 A에 비해 더 많은 돌연변이유발이 있는 오래된 산업용 균주이다. 상기 기재된 것과 유사한 단계를 수행하여 데이터를 생성하고 각 SNP를 분류하는데 사용될 수 있다. 일부 실시태양에서, 각각의 백그라운드 균주에서 모든 SNP를 제조하는 대신에, 특정 SNP 세트를 먼저 선택하고 추가 공학을 위해 우선순위를 정할 수 있는 것으로 이해된다.

[0799] 이 공학적 접근법의 유용성을 입증하는 데이터는 도 61에 도시되어 있다. 변이원성(mutagenic) SNP는 기본 계통과 비교하여 진보된 계통 균주에서 확인되었고, 전술한 공학적 접근법을 사용하여, 이들 SNP는 진보된 계통에서 거의 제거되지 않았다. 폴리케타이드 생산성에 대한 플레이트 분석에서 "SNP스왑" 균주를 부모 균주(진보된 계통 균주)와 비교하여 테스트하였고, 일부 균주는 부모 균주와 비교하여 개량을 나타냈다.

[0800] **C. 유익한 SNP 조합을 결정하기 위한 상위성 매핑의 사용**

[0801] 실시예 2의 파트 B에서 확인된 유익한 SNP는 결합시 숙주 성능을 개량시킬 가능성이 있는 SNP를 확인하기 위해 본 발명의 상위성 매핑 방법을 통해 분석될 것이다.

[0802] 새로운 조작된 균주 변형체는 상위성 매핑 예측에 따라 SNP 조합을 테스트하기 위해 실시예 1의 공학 방법을 사용하여 생성될 것이다. SNPs 통합은 순차적으로 발생할 수 있으며, 또는 다중 가지에 걸쳐 대안으로적 발생할 수 있으므로 하나 이상의 개량된 균주가 유익한 SNP의 서브세트와 함께 존재할 수 있다. 중립적 또는 해로운 SNP 수하물 없이 유익한 SNP의 최적 조합을 함유하는 최종 균주가 생산될 때까지, SNP 통합은 여러 균주 개량 라운드 동안 지속될 것이다.

[0803] **실시예 3: HTP 게놈 공학 - 사카로폴리스포라에서 스피노신 생산의 균주 성능을 개량시키는 SNP 스왑 라이브러리의 실행**

[0804] 이 실시예는 사카로폴리스포라 스피노사에서의 스피노신 생산의 수율 및 생산성 향상을 목적으로 실시예 2의 SNP 스왑 HTP 디자인 균주 개량 프로그램의 일부의 예시적인 실행을 제공한다.

[0805] 이 실시예의 섹션 B는 본 발명의 HTP 균주 개량 프로그램의 돌연변이 통합 단계를 추가로 예시한다. 따라서, 본 실시예는 본 발명의 HTP 균주 개량 방법의 제 1, 제 2 및 제 3 라운드 통합에 대한 실험 결과를 제공한다.

[0806] 제 2 및 제 3 라운드 통합에 대한 돌연변이는 개별 유전자 라이브러리 스왑으로부터 유도된다. 따라서 이런 결과는 다중 가지 병렬 트랙이 수행될 HTP 변형 프로그램에 대한 능력 및 본 발명의 유전자 디자인 라이브러리의 다양한 형태와 관련된 메타 데이터 속에 삽입될 수 있는 유익한 돌연변이의 "메모리"를 예시한다.

[0807] 상기한 바와 같이, 제공된 기본 참고 균주(균주 A) 및 제 2 "조작된" 균주(균주 C)의 게놈을 서열화하고 모든 유전자 차이를 확인하였다. 기본 균주는 돌연변이유발을 거치지 않은 사카로폴리스포라 스피노사 변이체이었다. 조작된 균주는 전통적인 돌연변이 개량 프로그램의 여러 라운드 후 기본 균주로부터 생산된 사카로폴리스포라 스피노사 균주이었다.

[0808] **A. HTP 공학 및 고 처리량 선별**

[0809] 본 발명의 클로닝 및 형질전환 방법에 따라, 확인된 SNP 각각을 개별적으로 다시 기본 균주에 첨가할 것이다. 단일 SNP를 포함하는 각각의 새롭게 생성된 균주를 생성물 역가 성능을 평가하도록 디자인된 소규모 배양물에서 스피노신 수율에 대해 테스트할 것이다. 소규모 배양은 산업용 규모의 배양 배지를 사용하여 수행할 것이다. 생성물 역가는 표준 비색 분석법으로 탄소 배출(즉, 단일 배치 수율을 나타냄)시 광학적으로 측정할 것이다. 반응을 종점으로 진행시키고 Tecan M1000 플레이트 분광 광도계를 사용하여 광학 밀도를 측정하였다.

[0810] **B. 두 번째 라운드 HTP 공학 및 고 처리량 선택 - 선택된 PRO 스왑 히트와 SNP 스왑 라이브러리의 통합**본 발명의 HTP 방법의 장점 중 하나는 각 SNP/프로모터/터미네이터/트랜스포존 돌연변이유발/항-대사산물/개시 코돈의 숙주 세포 표현형에 대한 영향과 관련된 정보와 함께 HTP 유전 디자인 라이브러리를 저장하는 능력이다. 본 발명자들은 사카로폴리스포라 스피노사에서 몇 가지 프로모터 스왑을 확인한 프로모터 스왑 실험을 이전에 수행하였다(예를 들어, 실시예 4 참조).

[0811] 본 발명자들은 (1) 프로모터 스왑(PRO Swap) 라이브러리, (2) SNP 스왑 라이브러리, (3) 개시/중지 코돈 교환 라이브러리, (4) STOP 스왑 라이브러리, (5) 서열 최적화 라이브러리, (6) 트랜스포존 돌연변이유발 다양성 라이브러리, (7) 리보솜 결합 부위(RBS) 다양성 라이브러리, 및 (8) 항-대사산물 선택/발효 생성물 저항성 라이브러리에서의 것들과 같은, 이전에 확인된 유전자 다양성 중 하나를 포함하도록 본 실시예의 기본 균주 A를 변형시킬 것이다. 초기 선별에서 확인된 최고의 유전자 다양성은 새로운 유전자 다양성 미생물 라이브러리를 만들기 위해 이 새로운 기본 균주로 다시 도입될 것이다. 이전 단계에서와 같이, 하나 이상의 유전자 다양성을 포함하는 각각의 새롭게 생성된 균주는 스피노신 수율에 대해 테스트될 것이다. 선택된 후보 균주는 스피노신 생산을 측정함으로써 생산성 대용물(proxy)에 대해 테스트될 것이다.

[0812] 이 제 2 라운드의 SNP 스왑으로부터의 결과는 프로모터 스왑 돌연변이를 포함하는 기본 균주에서 기본 균주 수율 및 스피노신의 생산성을 증가시킬 수 있는 SNP를 식별할 것이다.

[0813] **C. 탱크 배양 검증**

[0814] 상기 HTP 단계 동안 확인된 상위 SNP를 함유하는 균주를 중간 크기 테스트 발효 탱크 속에 배양할 것이다. 간단히 말해서, 각 균주의 작은 배양물을 성장시킬 것이며, 테스트 발효 탱크에서 큰 배양물을 동일한 양의 접종물과 접종하는데 사용할 것이다. 동일한 세포 밀도를 포함하도록 접종물을 표준화할 것이다.

[0815] 생성된 탱크 배양물을 채취하기 전에 정해진 시간 동안 진행시킬 것이다. 수율 및 생산성 측정은 발효를 통해 다양한 지점에서 탱크에서 채취한 시료에서 기질 및 생성물 역가로부터 계산할 것이다. 샘플은 적절한 표준을 사용하여 고압 액체 크로마토 그래피에 의해 특정 소분자 농도에 대해 분석될 것이다.

[0816] **실시예 4: HTP 게놈 공학 - 산업용 미생물 균주를 개량하기 위한 프로모터 스왑 라이브러리의 실행**

[0817] 이전의 실시예는 산업용 균주의 재건을 위한 본 발명의 HTP 균주 개량 프로그램의 능력을 입증하였다. 실시예 2 및 3은 다양한 기본, 중간 및 산업용 균주 내에서 기존 유전자 다양성을 탐구하는 SNP 스왑 기술 및 라이브러리의 실행을 기술하였다

[0818] 이 실시예는 본 발명의 PRO 스왑 기술을 사용하는 HTP 균주 개량 프로그램의 실시태양을 도시한다. 실시예 3과는 달리, 이 실시예는 PRO 스왑 라이브러리 생성을 통한 새로운 돌연변이 생성을 위한 방법을 교시한다.

[0819] **A. 프로모터 스와핑을 위한 표적의 확인**

[0820] 상기한 바와 같이, 프로모터 스와핑은 표적에 대한 "n" 유전자의 세트를 선택하는 단계를 포함하는 다단계 공정이다.

[0821] 본 발명에 기재된 게놈 공학 방법은 프로모터 스와핑을 위해 게놈의 임의의 위치를 표적화할 수 있게 한다. 본 실시예에서, 본 발명자들은 하기 열거된 코어 생합성 경로 유전자를 포함하여 본 발명의 프로모터 래더 방법을 통해 조절할 유전자를 확인하였다. (도 12a 내지 도 12d 참조). 또한, 전구체 풀, 보조인자 이용 가능성, 경쟁 2차 대사산물, 폴리케타이드 샤페론, 2차 대사산물 생성을 위한 주요 전사 조절제 및 시그마 인자, 기질 및 생성물 수송체는 물론 생성물 형성과 알려지지 않은 관계를 갖는 유전자(비-경로 유전자)는 균주 개량을 가능하게 하는 프로모터 스와핑의 모든 후보자이다.

표 7

[0822] S. 스피노사에서 스피노신 생산에 관여하는 잠재적 유전자

스피노신 합성 경로 유전자	유전자 정보 (서열, 기능 등)
spnA	폴리케타이드 합성효소 로딩 로딩 익스텐더 모듈 1 spnA
spnB	폴리케타이드 합성효소 익스텐더 모듈 2 spnB
spnC	폴리케타이드 합성효소 익스텐더 모듈 3-4 spnC
spnD	폴리케타이드 합성효소 익스텐더 모듈 5-7 spnD
spnE	폴리케타이드 합성효소 익스텐더 모듈 8-10 spnE
spnF	메틸트랜스퍼라제-유사 단백질 spnF
spnG	추정 NDP-람노실트랜스퍼라제 spnG
spnH	추정 O-메틸트랜스퍼라제 spnH
spnI	추정 O-메틸트랜스퍼라제 spnI
spnJ	추정 산화환원효소 spnJ
spnK	추정 O-메틸트랜스퍼라제 spnK
spnL	메틸트랜스퍼라제-유사 단백질 spnL
spnM	SpnM
spnN	추정 NDP-헥소오스-3-케토환원효소 spnN
spnO	추정 NDP-헥소오스-2 3-디하이드라타아제 spnO
spnP	추정 NDP-포로사밀트랜스퍼라제 spnP
spnQ	추정 NDP-헥소오스-3 4-디하이드라타아제 spnQ
spnR	추정 아미노트랜스퍼라제 spnR
spnS	추정 N-디메틸트랜스퍼라제 spnS
kre	dTDP-4-디하이드로람노스 환원효소 kre
gdh	dTDP-글루코스 4 6-디하이드라타아제 gdh
epi	dTDP-4-디하이드로람노스 3 5-에피머라제 epi
gtt	글루코스-1-포스페이트 티미딜일트랜스퍼라제 1 gtt
MetK	S-아데노실메티오닌 합성효소 MetK
PFK	피로포스페이트--프룩토스 6-포스페이트 1-ㅍ포스포트랜스퍼라제 PFK
rsmG	리보소말 RNA 스몰 서브유닛 메틸트랜스퍼라제 G rsmG
rpsL	30S 리보소말 단백질 S12 rpsL
gk	글루코키나아제
asb1	안트라닐레이트 합성효소 성분 1 asb1
pntA	NAD(P) 수소전달효소 서브유닛 알파 파트 1 pntA
pntB	NAD(P) 수소전달효소 서브유닛 알파 pntB
mmsd	메틸말로네이트-세미알데하이드 탈수소효소 (아실화)
Acat	아세틸-CoA 아세틸트랜스퍼라제
glcP	글루코피라노스
sucA	옥소글루타레이트 탈수소효소 E1 성분

[0823] **B. 프로모터 래더의 생성**

[0824] 프로모터 스왑 공정의 실행에서 다른 단계는 "래더"로서 작용하는 "x" 프로모터의 한 세트의 선택이다. 이상적으로 이 프로모터는 여러 게놈 유전자좌 전체에서 매우 다양한 발현을 유도하는 것으로 나타났지만, 유일한 요구 사항은 이들이 어떤 방식으로 유전자 발현을 교란한다는 것이다.

[0825] 특정 실시태양에서, 이런 프로모터 래더는 관심 표적 유전자와 관련된 천연, 고유 또는 야생형 프로모터를 확인하고 상기 프로모터를 돌연변이시켜 다중 돌연변이 프로모터 서열을 유도함으로써 생성된다. 이들 돌연변이 프

로모터의 각각은 표적 유전자 발현에 대한 효과에 대해 테스트된다. 일부 실시태양에서, 편집된 프로모터는 다양한 조건에 걸쳐 발현 활성에 대해 테스트되어, 각 프로모터 변이체의 활성이 문서화/특징화/주석 처리되고 데이터베이스에 저장된다. 생성된 편집된 프로모터 변이체는 뒤이어 그 발현의 강도에 기초하여 배열된 "래더"로 조직화된다(예를 들어, 상부 근처의 고도로 발현된 변이체 및 하부 근처의 약화된 발현에 의해, 따라서 "래더"라는 용어를 유도한다).

[0826] 본 예시적인 실시태양에서, 본 발명자는 스피노신 합성 경로에서 표적 유전자에 대한 프로모터 래더:ORF 조합을 생성할 것이다.

[0827] 본 발명의 유전자 공학 노력 및 대사 공학의 주요 목표는 숙주 대사를 변경하고, 생합성 경로를 최적화하며, 목적 생성물의 수율을 향상시키기 위해 경로 유전자를 도입 또는 복제하는 것이다. 성공은 생합성 유전자 클러스터의 내부(온-경로) 및 외부(오프-경로)에서 유전자의 발현을 교란시키고 균형을 잡는 능력에 의존하거나 도입된 비-천연 유전자 또는 유전자의 복제물의 과발현에 의존한다. 본 발명은 우리가 S. 스피노사에서 유전자 발현을 교란시키고 조정할 수 있게 하는 유전자 도구이다.

[0828] 개량된 표현형을 조작하기 위해 다수의 공학 라운드가 종종 요구된다. 이 래더의 유전적 다양성은 반복된 DNA 서열(예를 들어, 오프-표적 재조합을 위한 상동성 영역) 및 전사 희석 효과를 사용하는 것과 관련된 공학적 도전을 회피한다. 이 래더에서 서열 및 서열의 소스 다양성으로 인해, 본 발명은 이러한 과제를 회피한다.

[0829] 다른 보다 일반적인 숙주(모형 유기체; 예를 들어, Siegl et al.(2013, "Design, construction and characterization of a synthetic promoter library for fine-tuned gene expression in actinomycetes." Metab Eng. 19:98-106) and Seghezzi et al.(2011, "The construction of a library of synthetic promoters revealed some specific features of strong Streptomyces promoters." Appl Microbiol Biotechnol. 90(2):615-23. 참조)에 대한 프로모터 래더가 존재하지만, S. 스피노사는 매우 다루기 어려운 숙주이며 이 유기체를 위해 개발된 유전자 도구는 거의 없다. 본 발명은 S. 스피노사에서 개발되고 정량적으로 특징지어진 첫 번째 프로모터 래더를 나타내며, 여기에 설명된 프로모터는 가까운 숙주들에서 예측 가능한 역학을 보여줄 것으로 예상된다.

[0830] 추정 천연 프로모터 서열을 확인하고 선택하기 위해 채택한 접근법은 우리에게 이용 가능한 데이터를 사용하였다. S. 스피노사에 대한 조립 및 주석이 달린 참조 게놈을 사용하여 유전자의 예측된 코딩 서열의 업스트림에 유전자간 영역을 확인하였다. RNAseq 데이터(발효 동안 샘플링되고 2개의 균주에서 발현을 비교하는 복제된 시계열)를 사용하여 강하게 발현된 유전자 및 상이한 시간적 발현 프로파일을 갖는 유전자를 확인하였다. 이어서, 관심 유전자의 업스트림 서열(GOI)을 프로모터-형광 단백질 발현 카세트의 구축을 위해 선택하였다. 발효 관련 종자 배양 조건 및 생산 배양 조건 하에서 성장된 프로모터 래더 균주에서 상대 GFP 형광을 정량화하고 비교함으로써 프로모터 강도를 간접적으로 평가하였다. 잠재적으로 유용한 프로모터는 하기 표 8에 열거되어 있다. 프로모터 평가의 제 1 라운드는 프로모터 강도의 래더를 생성하였다(도 15). 후속 평가를 통해, 우리는 원래 식별된 것보다 훨씬 더 강한 것을 포함하여 추가적인 기능성 프로모터를 확인할 수 있었다(도 16).

표 8

[0831] 프로모터 래더의 요약: 서열 이름, 소스, 특성 및 테스트 상태

서열 번호	이름	관련 유전자	bp	형태	상태
1	P7160	샤페로닌 GroEL	141	천연	테스트됨
2	P7253	신장 인자 Tu	198	천연	테스트됨
3	P6681	F-형 ATPase 서브유닛 델타	232	천연	테스트됨
4	P6316	PspA/IM30 패밀리 단백질	361	천연	테스트됨
5	P6806	2-옥소글루타레이트 디카복실라아제	303	천연	테스트됨
6	P3159	추정 에노일-CoA 하이드라타아제 echA8	253	천연	테스트됨
7	P0757	추정 L-라이신-입실론 아미노트랜스퍼라제	168	천연	테스트됨
8	P5011	가설 단백질	363	천연	테스트됨
9	P1409	NAD-특이적 글루타메이트 탈수소효소	300	천연	테스트됨
10	P4735	류실 아미노펩티다아제 (아미노펩티다아제 T)	322	천연	테스트됨
11	P2900	시토크롬 P450-terp	149	천연	테스트됨
12	P0801	주변세포질 뮤레인 펩타이드-결합 단백질 전구체	199	천연	테스트됨
13	P21	-	41	Siegl 외	테스트됨

14	PA9	-	41	Siegl 외	테스트됨
15	PA3	-	41	Siegl 외	테스트됨
16	PB4	-	41	Siegl 외	테스트됨
17	PB12	-	41	Siegl 외	테스트됨
18	PB1	-	41	Siegl 외	테스트됨
19	PC1	-	41	Siegl 외	테스트됨
20	P72	-	41	Siegl 외	테스트됨
21	P-C4-1	-	44	Seghezzi 외	테스트됨
22	P-A5-19	-	44	Seghezzi 외	테스트됨
23	P-C4-14	-	44	Seghezzi 외	테스트됨
24	P-D1-7	-	44	Seghezzi 외	테스트됨
25	P1	분비 단백질	242	천연	테스트됨
26	P2	가설 단백질	65	천연	테스트됨
27	P3	RNA 폴리머라제 시그마 인자 SigD	201	천연	테스트됨
28	P3v2	RNA 폴리머라제 시그마 인자 SigD	300	천연	
29	P4	항원 Ag88	220	천연	테스트됨
30	P4v2	항원 Ag88	177	천연	
31	P5	DNA-지시된 RNA 폴리머라제 서브유닛 베타	200	천연	테스트됨
32	P5v2	DNA-지시된 RNA 폴리머라제 서브유닛 베타	299	천연	
33	P6	분자 샤페론 GroEL	240	천연	테스트됨
34	P7	UDP-4-아미노-4-데옥시-L-아라비노스--옥소글루타레이트 아미노트랜스퍼라제	242	천연	테스트됨
35	P8	프롤린 라세미화 효소	300	천연	테스트됨
36	P9	페닐옥사졸린 합성효소 MbtB	359	천연	테스트됨
37	PspnA	폴리케타이드 합성효소 로딩 및 확장 모듈 1	433	천연	테스트됨
38	PspnA_v2	폴리케타이드 합성효소 로딩 및 확장 모듈 1	115	천연	테스트됨
39	PspnF	메틸트랜스퍼라제-유사 단백질	496	천연	테스트됨
40	PspnG	추정 NDP-람노실트랜스퍼라제	202	천연	테스트됨
41	PspnQ	추정 NDP-헥소오스-3	379	천연	테스트됨
42	PspnQ_v2	추정 NDP-헥소오스-3	261	천연	테스트됨
49	P1765	글루타민 합성효소 1	332	천연	테스트됨
50	P3747	가설 단백질	300	천연	테스트됨
51	P5078	가설 단백질	247	천연	테스트됨
52	P7419	혐기성 벤조에이트 이화작용 전사 조절제	230	천연	테스트됨
53	P7156	RNA 폴리머라제 시그마 인자 SigD	300	천연	테스트됨
54	P7256	30S 리보솜 단백질 S12	300	천연	테스트됨
55	P1941	응답 조절제 단백질 vraR	298	천연	테스트됨
56	P3405 (P8)	프롤린 라세미화 효소	300	천연	테스트됨
57	P3407	ABC 트랜스포터 아르기닌-결합 단백질 1 전구체	300	천연	테스트됨
58	P2428	아세틸-CoA 합성효소	248	천연	테스트됨
59	P0927	4-하이드록시페닐피루베이트 2산소화효소	250	천연	테스트됨
60	P0889	선형 그라미시딘 탈수소효소 LgrE	250	천연	테스트됨
61	P0186	L,D-트랜스펩티다아제 촉매영역	298	천연	테스트됨
62	P3702_v2	가설 단백질	296	천연	테스트됨
63	P7156_v2	RNA 폴리머라제 시그마 인자 SigD	300	천연	테스트됨
64	P7256_v2	30S 리보솜 단백질 S12	226	천연	테스트됨
65	P1765_v2	글루타민 합성효소 1	191	천연	테스트됨
66	P7539_v2	항원 Ag88	177	천연	테스트됨
67	P7276_v2	DNA-지시된 RNA 폴리머라제 서브유닛 베타	299	천연	테스트됨
68	P0941_v2	가설 단백질	266	천연	테스트됨
69	P0889_v2	선형 그라미시딘 탈수소효소 LgrE	163	천연	테스트됨
43	P21_mutant	-	41	합성	테스트됨
44	P1_core	분비 단백질	40	천연	테스트됨
45	P1(-33)	분비 단백질	209	천연	테스트됨
46	P1+ribswtch	분비 단백질	433	천연	테스트됨

47	P21-P1	-	53	합성	테스트됨
48	P1-P21	-	52	합성	테스트됨
172	Pmut-1	-	41	합성	테스트됨
173	B2	-	41	합성	테스트됨
174	D1	-	41	합성	테스트됨
175	D2	-	41	합성	테스트됨

[0832] 라이브러리에서 프로모터의 발현 강도는 프로모터 서열의 다운스트림에 위치한 형광 리포터 단백질을 사용하는 것을 특징으로 하였다. 프로모터-리포터 서열을 2개의 별개의 실험 균주의 게놈에서 중성 통합 부위에 통합시키고, 형광을 상이한 성장 체제 하에서 측정하여 프로모터 강도에 대한 정량적 메트릭을 제공하였다. 이 프로모터 라이브러리는 S. 스피노사 및 관련 숙주에서 유전자 발현의 조절(시간적 역학의 증가, 감소 또는 변경) 및 개량된 표현형에 대한 조작을 가능하게 한다. 본 발명은 이 숙주의 유전 공학을 위한 몇 가지 용도를 갖는다: 1) PROSWP(Zymergen technology)와 함께 사용하기 위한 것; 2) 천연 유전자의 이종성 또는 복제된 복제물의 과발현을 위한 것; 3) 생합성 또는 관련 유전자의 다중 유전자 통합의 발현 균형을 잡기 위한 것. 선택된 프로모터-유전자 쌍을 조작하면 특정 균주에서 스피노신 생산이 향상된다(도 17).

[0833] 따라서, 본 발명자들은 적어도 다음의 프로모터를 제공하여 프로모터 라이브러리를 형성한다:

[0834] (1) S. 스피노사 게놈으로부터 새로 확인된 천연 프로모터 서열;

[0835] (2) 관련 숙주 유기체에 사용하기 위해 기재된 합성 프로모터 서열(Siegl 등 및 Seghezzi 등);

[0836] (3) 다양한 프로모터 서열의 돌연변이유발 라이브러리; 및

[0837] (4) 프로모터의 조합 재-배열(진행 중)로 구성된 하이브리드 프로모터 서열.

[0838] 이들 프로모터는 상당한 뉴클레오티드 다양성으로 구성되는 동안 다양한 발현 강도를 나타낸다(도 15 및 도 16 참조). 이 프로모터 라이브러리는 S. 스피노사 및 관련 숙주에서 사용될 수 있는 다운스트림 유전자의 발현을 조절하는 DNA 서열 세트를 제공한다. 본 발명에 기술된 라이브러리는, 예를 들어 약 50 내지 100배 동적 범위(도 15 및 도 16 참조)에 걸친 발현 강도의 "래더"를 나타내고, 추가로 뉴클레오티드 다양성의 범위를 보여준다. 이 프로모터 라이브러리는 측정 가능한 표현형을 개량하기 위해 반복적인 조작 라운드를 위해 숙주 게놈의 정확한 튜닝을 위해 함께 사용될 수 있다. 각각의 프로모터 유형, 강도 및 고유 서열은 대사 공학에서 종종 직면하는 미지의 과제를 극복할 수 있는 기회를 제공한다. 이러한 변경은 다음을 포함하지만 이에 제한되지 않는다: (1) 프로모터가 각각의 독특한 상황에서 어떻게 기능할 것인지(그것이 주어진 유전자의 발현에 어떻게 영향을 미치는지)를 정확하게 예측할 수 없음; (2) 주어진 유전자에 최적인 발현 수준; (3) 어떻게 시간적 역학이나 조절이 교란을 성공시키는지를 예측할 수 없음; 및 (4) 균형 잡힌 또는 최적화된 생합성 경로를 초래할 발현 수준. 본 발명에 기재된 프로모터는 스피노신과 같은 S. 스피노사에서 화학물질의 개량된 생산을 위해 균주 유전자형을 부여하기 위한 특정 유전자 표적과 상호작용할 수 있다.

[0839] **C. 래더로부터의 프로모터를 표적 유전자와 연결**

[0840] 프로모터 스왑 공정의 실현에서 다른 단계는 특정 표적 유전자와 관련된 프로모터 래더로부터의 소정의 프로모터를 포함하는 다양한 균주의 HTP 공학이다.

[0841] 고유 프로모터가 표적 유전자 n의 앞에 존재하고 이의 서열이 알려진 경우, 래더에서x프로모터의 각각에 의한 고유 프로모터의 교체가 수행될 수 있다. 고유 프로모터가 존재하지 않거나 이의 서열이 알려지지 않은 경우, 유전자 n 앞에 있는 래더에x프로모터의 각각의 삽입이 실행될 수 있다. 이러한 방식으로, 균주의 라이브러리가 구축되고, 여기서 라이브러리의 각 구성원은 그렇지 않으면 동일한 유전자 상황에서 n 표적에 작동 가능하게 연결된x프로모터의 예이다(예를 들어, 도 13 참조).

[0842] **D. 균주의 HTP 선별**

[0843] 프로모터 스왑 공정에서 최종 단계는 상기한 라이브러리에서 균주의 HTP 선별이다. 유래된 균주의 각각은 그렇지 않으면 동일한 유전자 배경에서, n 표적에 연결된x프로모터의 예를 나타낸다.

[0844] 하나 이상의 메트릭에 대한 성능이 특징화되는 시나리오에서, 각 균주의 HTP 선별을 실행함으로써, 발명자는 소정의 메트릭에 대해 어떤 프로모터/표적 유전자 결합이 가장 유익한 지를 결정할 수 있다(예를 들어, 관심 분자의 생산의 최적화). 도 13 참조.

[0845] 도 17에 도시된 예시적인 실시태양에서, 본 발명자는 스피노신의 생산을 최적화하기 위해 프로모터 스왑 공정을 사용하였다. 상기한 Pro SWAP 방법의 응용은 아래의 실시예 5에 기술된다.

[0846] **실시예 5: HTP 게놈 공학 - 스피노신 생산을 위한 균주 성능을 개량시키기 위한 PRO 스왑 라이브러리의 실현.**

[0847] 아래 섹션은 실시예 4에 기술된 바와 같이, 본 발명의 PRO 스왑 HTP 디자인 균주 개량 프로그램 도구의 예시적 실행을 제공한다. 이 실시예에서, S. 스피노사 균주에 스피노신의 숙주 세포 수율을 증가시키기 위해 본 발명의 PRO 스왑 방법을 적용하였다.

[0848] **A. 프로모터 스왑**

[0849] 프로모터 스왑을 실시예 4에 기재된 바와 같이 수행하였다. 스피노신 생산에서 역할을 한다고 가정된 게놈에 걸친 유전자를 열거된 프로모터 래더를 사용하여 프로모터 스왑에 대한 표적으로 하였다(예를 들어, 도 13). 프로모터 스왑을 위한 이러한 유전자는 (1) 예를 들어, 스피노신과 같은 관심 화합물의 코어 생합성 경로에서의 유전자; (2) 전구체 합성 또는 풀 이용 가능성의 조절에 직접 관여하는 유전자와 같은 관심 화합물의 전구체 풀 이용 가능성에 관련된 유전자; (3) 보조인자(cofactor) 이용과 관련된 유전자; (4) 전사 조절제로 코딩하는 유전자; (5) 영양 공급원의 수송체를 코딩하는 유전자; 및 (6) 생성물 수출자 등.

[0850] **B. HTP 공학 및 고 처리량 선별**

[0851] 프로모터 스왑의 HTP 공학을 실시예 1 및 3에 기술된 바와 같이 수행하였다. 생성된 프로모터 스왑 균주의 HTP 선별을 실시예 3에 기술된 바와 같이 수행하였다. 상이한 기능적 길이(코어 생합성 클러스터로부터 오프-경로에 이르는)에 걸친 다수의 유전자가 프로모터 스왑에 대해 표적화되고, 부모 균주에 비해 개량된 균주 성능을 나타내는 데이터가 도 17에 제시되어 있다.

[0852] 유사하게, 프로모터 스왑은 도 13의 좌측 패널에 기재된 스피노신 생합성 경로로부터 선택된 유전자에 대해 수행될 것이고, 게놈을 가로지르는 유전자는 새로운 개량된 균주를 확인하기 위해 수행될 것이며, 이는 위의 표 8에 기재된 프로모터를 사용한 프로모터 스왑에 대해 표적화될 것이다.

[0853] 시각화될 때, 프로모터 스왑 라이브러리 선별의 결과는 측정되는 성능 메트릭과 가장 밀접하게 상관되는 유전자 표적을 확인하는 역할을 할 것이다.

[0854] 선택된 균주를 작은 플레이트에서 재-배양하고 상기 기재된 바와 같이 스피노신 수율에 대해 테스트할 것이다.

[0855] **실시예 6: 상위성 매핑 - 유익한 돌연변이 통합을 예측하기 위한 알고리즘 도구**

[0856] 이 실시예는 본 발명의 HTP 균주 개량 프로그램의 일부로서 이용되는 예측 모델링 기술의 한 실시태양을 기술한다.(상기한 바와 같은 유전자 디자인 라이브러리의 사용을 통한) 잠재적으로 유익한 돌연변이의 초기 확인 후, 본 발명은 2차, 3차, 4차 및 추가의 후속 라운드의 HTP 균주 개량에서 유익한 돌연변이의 통합 방법을 교시한다. 일부 실시태양에서, 본 발명은 돌연변이 통합이 상기 돌연변이 각각의 개별적인 성능에 기초할 수 있음을 교시한다. 다른 실시태양에서, 본 발명은 둘 이상의 돌연변이가 단일 숙주 세포 내로 통합되는 경우에 부가적 또는 상승 효과를 나타낼 가능성을 예측하는 방법을 교시한다. 아래 실시예는 본 발명의 예측 도구의 실시태양을 도시한다.

[0857] 실시예 3 및 5의 SNP 스왑 및 프로모터 스와핑(PRO 스왑) 라이브러리로부터의 선택된 돌연변이를 분석하여 균주 숙주 성능 개량으로 이어질 가능성이 가장 높은 SNP/PRO 스왑 조합을 확인할 것이다.

[0858] SNP 스와핑 라이브러리 서열을 본 발명의 "상위성 매핑" 섹션에 기재된 바와 같이 코사인 유사성 매트릭스를 사용하여 서로 비교할 것이다. 분석 결과는 각 SNP/PRO 스왑 조합에 대해 기능적 유사성 점수를 산출할 것이다. 모든 SNP/PRO 스왑 사이의 기능적 유사성의 시각적 표현은 도 53의 히트 맵에 도시된다. 생성된 기능적 유사성 점수는 도 54a의 예에서와 유사하게 각 SNP/PRO 스왑의 각각 사이의 유사성 거리를 도시하는 덴드로그램을 개발하는데 사용할 것이다.

[0859] 동일하거나 유사한 작용기로부터의 돌연변이(즉, 높은 기능적 유사성을 갖는 SNP/PRO 스왑)는 동일한 메카니즘에 의해 작동하기 쉽고, 따라서 조합된 경우 전체적인 숙주 성능에 대해 음성 또는 중성의 상위성을 나타낼 가능성이 더 높다. 대조적으로, 상이한 작용기로부터의 돌연변이는 독립적인 메카니즘에 의해 작동하기 쉽고, 따라서 숙주 성능에 유익한 첨가제 또는 조합 효과를 나타낼 가능성이 더 높다.

[0860] 상위성에 대한 생물학적 경로의 효과를 설명하기 위해, 다양한 기능적 유사성을 나타내는 SNP 및 PRO 스왑을 조

합하고 숙주 균주에 대해 테스트할 것이다. 3개의 SNP/PRO 스왑 조합을 실시예 1에 기술된 바와 같이 S. 스피노사의 게놈에 조작할 것이다.

[0861] SNP/PRO 스왑 조합을 함유하는 각각의 숙주 세포의 성능을 실시예 3에 기술된 바와 같이 테스트하고, 대조군 숙주 세포의 성능과 비교할 것이다.

[0862] 따라서, 상위성 매핑 절차는 디자인된 유전자 변화의 효과적인 및/또는 긍정적인 통합을 예측/프로그래밍/통지하는데 유용하다. 상위성 매핑 절차로부터의 분석적 통찰력은 미생물 균주 개발의 후속 라운드를 안내할 수 있는 예측성 규칙 세트의 생성을 허용한다. 상위성 라이브러리로부터 얻은 예측성 통찰력은 미생물 유형 및 표적 분자 유형에 걸쳐 사용될 수 있다.

[0863] **실시예 7: HTP 게놈 공학 - Pro 스왑 돌연변이 통합 및 다중 인자 조합 테스팅**

[0864] 이전의 실시예는 적은 수의 미리 선택된 PRO 스왑 돌연변이를 SNP 스왑 라이브러리와 통합하기 위한 방법을 설명하였다(실시예 3). 다른 실시예는 부가적 또는 상승적 유익한 숙주 세포 성질을 나타낼 가능성이 가장 큰 돌연변이 통합을 선택하는 상위성 방법을 설명하였다(실시예 6). 이 실시예는 다수의 유전자/유전자 디자인 라이브러리 조합(예를 들어, PRO 스왑 라이브러리xSNP 라이브러리 또는 PRO 스왑 라이브러리 내의 조합)의 조합적 통합에 의해 생성된 큰 솔루션 공간을 효과적으로 탐색하는 본 발명의 HTP 방법의 능력을 설명한다.

[0865] 본 발명의 HTP 균주 개량 방법의 이러한 예시적인 적용에서, 실시예 5의 숙주 성능에 긍정적 효과를 갖는 것으로 확인된 프로모터 스왑은 원래의 PRO 스왑 라이브러리와 2차 조합으로 통합될 것이다. PRO 스왑 돌연변이를 통합하는 결정은 각 돌연변이의 수율 또는 생산성에 대한 전반적인 효과와 두 돌연변이의 조합이 첨가물 효과 또는 시너지 효과를 창출 할 가능성에 근거한다.

[0866] **A. PRO 스왑 균주 공학을 위한 통합 라운드**

[0867] 균주를 이전 실시예 1에서 기술한 바와 같이 형질전환시킬 것이다. 간단하게, 원하는 PRO 스왑 돌연변이를 이미 함유하고 있는 균주를 제 2 원하는 PRO 스왑 돌연변이로 다시 형질전환시킬 것이다.

[0868] 단일 및 이중 통합 돌연변이의 솔루션 공간을 탐색하기 위한 HTP 방법은 또한 제 3, 제 4 및 후속 돌연변이 통합에 적용될 수 있다.

[0869] **실시예 8: HTP 게놈 공학 - 산업용 숙주 균주를 개량하기 위한 터미네이터 라이브러리의 실행**

[0870] 본 실시예는 STOP 스왑을 포함하는 추가의 HTP 유전자 디자인 라이브러리에 본 발명의 HTP 방법을 적용한다. 본 실시예는 보다 복잡한 유전자 디자인 라이브러리(예를 들어, 프로모터 및 터미네이터 모두를 포함하는 PRO-STOP 스왑 라이브러리)를 생성하기 위해 기본 유전자 디자인 라이브러리(예를 들어, PRO 스왑, SNP 스왑, STOP 스왑 등)로부터의 요소를 조합하는 본 발명의 능력을 추가로 설명한다. 일부 실시태양에서, 본 발명은 이전에 개시된 유전자 디자인 라이브러리 중 임의의 것을 조합하여 유래된 것들을 포함하는 임의의 및 모든 가능한 유전자 디자인 라이브러리를 교시한다.

[0871] 이 실시예에서, 유전자 발현에 대한 본 발명의 STOP 스왑 방법의 효과를 입증하기 위해 소규모 실험을 수행할 것이다. 본 발명의 터미네이터는 아래 기술된 바와 같이 2개의 고유 S. 스피노사 프로모터의 하나와 쌍을 이루고, 형광 단백질의 발현에 영향을 미치는 능력에 대해 분석될 것이다.

[0872] 조합 유전자 공학 및 대사 경로 리팩토링 접근법은 유전자 발현을 교란시키고 표적 분자의 생산에 영향을 미치거나 원하는 숙주 표현형을 변경시키기 위해 정확한 위치에서 숙주 게놈으로 조합되거나 삽입될 수 있는 DNA 요소의 라이브러리(예를 들어, 프로모터, 리보솜 결합 부위, 전사 터미네이터)에 의존한다. 이들 라이브러리의 개량된 이해 및 정량적 평가/특성화는 유전적 변화의 예측성을 향상시킬 수 있는 기회를 제공하기 때문에 바람직하다. 일반적인 DNA 라이브러리 유형 중에서 전사 터미네이터는 아마도 가장 잘 이해되지 않고 있다. 터미네이터는 (1) 전사를 완료하는데 역할을 하지만, 이들은 또한 (2) mRNA 반감기에 영향을 미치며(Curran et al., 2015, "Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast." ACS Synth. Biol. 4(7): 824-832), 차례로 단백질 발현에 영향을 미친다. 따라서, 터미네이터는 임의의 합성 생물학 도구의 중요한 구성 요소로 간주되어야 한다. 터미네이터 라이브러리 또는 래더의 생성은 다음 두 가지 기준에 따라 터미네이터 성능을 평가하고 수량화하는 메커니즘이 필요하다: (1) 전사를 종결시키는 능력; (2) mRNA 반감기 및 업스트림 유전자의 발현에 영향을 미치는 능력. 본 발명은 S. 스피노사에서 이를 수행하기 위한 강력한 첫 번째 도구를 제공한다.

[0873] 유사한 용액이 존재하고 다른 유기체에서 사용되어왔지만(Chen et al., 2013, "Characterization of 582 natural and synthetic terminators and quantification of their design constraints." Nat. Methods 10, 659-664, and Cambray et al., 2013, "Measurement and modeling of intrinsic transcription terminators." Nucleic Acids Research. 41(9): 5139-5148), 본 발명은 (1) 터미네이터 기능성을 평가하기 위한 제 1 시스템 및 분석법; (2) S. 스피노사에서 개발되고 특성화된 전사 터미네이터 라이브러리를 제공한다.

[0874] 추정 터미네이터를 확인하기 위해, S. 스피노사 및 S. 에리트라에아의 게놈 서열을 핵산 서열에서 rho-비의존 터미네이터의 예측을 위한 온라인 도구에 입력 하였다. 잘 주석이 달린 유전자의 다운스트림(유전자 간 영역에서)에서 발생하는 온라인 도구에 의해 예측된 12개의 터미네이터 서열(4개의 천연 및 8개의 이종 서열, 하기 표 9 참조)이 분석을 위해 선택되었다.

표 9

[0875] 테스트된 추정 터미네이터의 서열, 소스 및 크기.

ID	관련된 유전자	서열	소스	크기 (bp)
T1	(신장 인자 tu)	CCCGAACCTTCGGGGGCGGGCCCTCTTGCTTTTCAAT (서열번호 70)	*S. spinosa*	37
T2	(류실 아미노펩티다아제)	CGGGCAATAATACGTGCCCGGACGGTAGTGCGAGCACGAGGTGGGTACG (서열번호 71)	*S. spinosa*	49
T3	(시토크롬 P450 수산화효소)	AGTTTGTCGAACCGGCGGCGTTCGCCGGcTTTACCTTGCGC (서열번호 72)	*S. spinosa*	41
T4	(FOF1 ATP 합성효소 서브유닛 베타)	GGTTTCTCGAACCAGTGCTTTGCGTACTGGTTGTCGTTGCAG (서열번호 73)	*S. spinosa*	42
T5	(FAD-결합된 산화환원효소)	CGGAGCCAGAGGGCGCCTGAGTGCCTGTTTTTGATCC (서열번호 74)	*S. erythraea*	37
T6	(포스포리보실트랜스퍼라제)	AAACGCCCCCGGCTCCGGCCGGGGGCgTTTTTGGTTGTG (서열번호 75)	*S. erythraea*	39
T7	(ATP-결합 단백질)	AGACGCAGGAGGTCTCGTGAGGGGCTTTTCCGCGAGC (서열번호 76)	*S. erythraea*	37
T8	(50s 리보솜 단백질 L32)	CGTGTGACTTGTCCCACTCGGGGTTTTTGTCGCGA (서열번호 77)	*S. erythraea*	35
T9	(tRNA-Arg)	GGATTCGTCCGGCCGAGGCCAATCGGCTTTTCGGGGCCC (서열번호 78)	*S. erythraea*	39
T11	(lsr2)	GCTTTCGTCGGCCGGGAACGCCCTGGTGTTTCTTACCG (서열번호 79)	*S. erythraea*	38
T12	(AraC)	TTGGGTGGATTCACCCCTACCGGGTGTTTTTCTCGGCT (서열번호 80)	*S. erythraea*	38
NoT	없음	-	-	0

[0876] 이들 추정 터미네이터를 테스트하기 위해, 이중 리포터 디자인 및 분석이 이용되었다. 테스트에 사용된 이중 리포터 디자인 및 분석(실시예 10에 추가로 기술 됨)은 추정 전사 터미네이터 서열의 기능성 및 상대 강도의 신속한 평가를 가능하게 한다. 분석은 추정 전사 터미네이터의 성능을 평가하기 위해 별개의 스펙트럼 시그니처(도 31a-d)와 함께 2개의 형광 리포터 단백질(대셔GFP 및 파프리카RFP; DNA2.0의 IP-free 서열)을 사용한다. 이 시스템은 사용자가 1) 전사를 정지시키고 2) 업스트림 유전자의 발현에 영향을 미치는 능력에 대해 추정 터미네이터를 평가할 수 있게 한다. 이중 형광 리포터 테스트 카세트는 S. 스피노사의 유전 공학에 필요한 추정 터미네이터 서열의 강도의 정량적 평가 및 mRNA 안정성에 대한 영향을 평가하는 메커니즘을 가능하게 한다.

[0877] 이들 성능 기준의 정량적 평가는 ermE* 프로모터에 의해 구동되는 2개의 형광 단백질의 바이-시스트로닉 발현을 이용하는 디자인에 의해 가능하다(Bibb et al., 1985, "Cloning and analysis of the promoter region of the erythromycin resistance gene (ermE) of Streptomyces erythraeus." Gene. 38(1-3):215-226). 각각의 추정 터미네이터 서열을 두 리포터(GFP의 다운스트림 및 RBS 및 RFP의 업스트림) 사이에서 클로닝하였다. 다운스트림 리포터(RFP)의 발현(형광)은 양성 대조군(리포터 사이에 터미네이터 서열이 없는 동일한 폴리시스트로닉 카세트; NoT; 도 33 참조)의 GFP 및 RFP 형광을 사용하여 정규화후 업스트림 리포터(GFP)의 것과 비교하여 결정되었다. 이 시스템은 터미네이터 라이브러리의 정량적 평가를 위한 강력한 메커니즘을 제공하고 S. 스피노사의 유전 공학에 사용하기 위한 추정 터미네이터 서열의 성능을 확인하고 특성화하는 도구를 제공한다. 시스템의 강도는 뚜렷한 형광 스펙트럼을 갖는 두 개의 형광 리포터를 적용하는데 있다(도 31a-d). 리포터는 다른 리포터와

의 스펙트럼 간섭 없이 넓은 동적 범위(~50x)에서 형광의 정량화(각 리포터의 단백질 발현)를 허용하므로 복잡한 신호 보정(중복 형광 신호의 분리)이 필요 없다. 각 리포터의 표현은 독립적으로 측정할 수 있다. 이 값들은 다른 리포터에 대한 형광(RFU) 및 터미네이터 없이 대조군 균주로부터 생성된 형광을 비교함으로써 각각의 리포터의 발현에 기여하는 유전적 요소의 성능을 평가할 수 있게 한다. 두 리포터 사이의 추정 터미네이터 서열을 스왑하는 동안 다른 모든 요소를 일정하게 유지함으로써 간접적으로 다음을 평가할 수 있다: (1) 상이한 터미네이터가 존재하는 경우, 업스트림 리포터(GFP)의 상대적인 형광을 비교함으로써 mRNA 안정성에 대한 터미네이터의 영향; (2) 터미네이터 없이 대조군 균주의 형광에 의한 정규화 후 다운스트림 리포터(RFP)의 상대적 형광을 업스트림 리포터(GFP)의 형광과 비교함으로써 전사를 중지시키는 터미네이터의 능력. 이 시스템을 통해 (1) 기능성 터미네이터 및 (2) mRNA 안정성에 영향을 미치거나 그 특성을 증진시키는 특성이 다른 터미네이터를 식별할 수 있다.

[0878] 후보 터미네이터의 핵산 서열을 테스트 카세트에 클로닝하고 공지된 중성 통합 부위에서 S. 스피노사 게놈에 통합시켰다. 생성된 균주를 액체 배지(씨드 매질)에서 48시간 동안 성장시키고, PBS로 세척하고, 플레이트 판독기를 사용하여 형광(GFP 및 RFP)을 측정하였다. 형광은 OD_{540} 흡광도에서 정규화되었다.

[0879] 분석에 기초하여, S. 스피노사에서 다양한 기능 또는 강도(다운스트림 유전자의 전사를 중단시키거나 업스트림 유전자의 전사를 약화시킬 수 있는 능력)를 갖는 11개의 전사 터미네이터 서열(S. 에리트라에아로부터의 4개의 천연 및 7개의 이종성 서열)의 라이브러리. 이들 서열의 길이는 35-49개 뉴클레오티드 길이이며, 공학 디자인에 쉽게 통합될 수 있다(표 3; 표 8; 도 32 및 도 33). 결과는 엔지니어에게 더 크고 다양한 솔루션 공간과 표적 유전자의 발현을 교란시키고 조작할 수 있는 기회를 제공하는 강도와 mRNA 안정성에 영향을 미치는 다양한 터미네이터 라이브러리이다(도 34).

[0880] 본 발명의 전사 터미네이터 서열의 라이브러리는 S. 스피노사의 유전 공학에 필요한 도구를 제공한다. 전사 터미네이터 서열은 몇 가지 공학적 적용을 갖는다: (1) 의도하지 않은 결과의 업스트림 조절로부터 보호하기 위한 프로모터 또는 유전자 통합의 절연체로서; (2) 유전자 삽입을 위한 전사 터미네이터로서; 및 (3) mRNA 안정성에 대한 그들의 영향을 통해 또는 프로모터와 번역 개시 부위 사이에 유전자의 코딩 서열의 업스트림에 삽입함으로써 발현 및 균형 경로를 조정하기 위해. 이 후자의 적용은 다운스트림 유전자의 발현을 넉다운시키거나 효과적으로 방지할 수 있다.

[0881] 유전자 발현을 녹다운하거나 제거하기 위한 이 터미네이터 라이브러리의 적용을 평가하기 위해, 개별 터미네이터(우리의 터미네이터 라이브러리의 서브세트: 서열번호 70, 72, 74, 79 및 80)를 2개의 상이한 프로모터(서열번호 25 및 33) 중 하나와 형광 리포터(서열번호 81) 사이에 삽입함으로써 이 적용을 테스트하였다(도 65). 그런 다음 이 테스트 카세트를 균주 A에 통합하고 결과 균주의 GFP 발현을 사용하여 GFP 발현의 감쇠에 대한 터미네이터 삽입의 영향을 평가하였다(도 66a-b). 도 66a는 강한 프로모터(서열번호 25)와 GFP 사이에 삽입된 T1, T3, T5, T11 및 T12(서열번호 70, 72, 74, 79 및 80)를 갖는 균주의 발현을 보여준다. "없음(None)"(왼쪽 열)은 터미네이터가 없는 대조군 균주을 나타낸다. 도 66b는 적당히 강한 프로모터(서열번호 33)와 GFP 사이에 삽입된 T1, T3, T5 및 T12(서열번호 70, 72, 74 및 80)를 갖는 균주의 발현을 보여준다. "없음"(왼쪽 열)은 터미네이터가 없는 대조군 균주을 나타낸다. 표준 편차는 일반적으로 다이아몬드 위와 아래에서 관찰되는 수평 대시로 표시된다. 도면의 오른쪽에 있는 원은 모든 쌍의 터키-크레이머(Tukey-Kramer) HSD 테스트를 기반으로 그룹 간의(중복되지 않은/교차 원은 서로 현저하게 다른 그룹을 나타냄) 유의한 차이를 나타낸다.

[0882] 이 공학적 접근법의 유용성을 입증하는 데이터는 도 62에 도시되어 있다. 터미네이터는 유전자 발현을 변형시키기 위해 다수의 표적화된 유전자의 업스트림에 삽입되었고, 이들 조작된 균주는 폴리케타이드 생산성에 대한 플레이트 분석에서 부모 균주와 비교하여 테스트되었다. 몇몇 "터미네이터 삽입" 균주는 부모 균주와 비교하여 개량을 나타냈다. 일부 실시태양에서, 터미네이터 삽입(서열, 터미네이터-유전자 조합, 또는 균주)의 수집은 "터미네이터 삽입 미생물 라이브러리"로 지칭된다.

[0883] **실시예 9: 사카로폴리스포라에서 유전자 다양성을 생성하기 위한 유전자 변화의 신속한 통합**

[0884] 이 실시예는 사카로폴리스포라 스피노스에서 유전자 변화의 신속한 통합 및 유전자 다양성을 생성하는 방법을 설명한다. S. 스피노사 균주의 공학은 주로 유기체에 대한 느린 성장과 유전자 도구의 부족으로 인해서 긴 과정이다. 이 문제는 생산 균주에서 더 악화되며, 이는 성장률이 감소하고 견고성이 떨어질 가능성이 높다. 예를 들어, 본 발명 이전에 S. 스피노사를 조작하는 방법은 컨쥬게이션에 의해 외래 DNA를 도입하는 것이다(Matsushima et al, 1994. Gene, 146 39-45). 이 과정은 전달된 플라스미드를 숙주 DNA로 단일 크로스 오버하는 것에 기초한다. 외래 DNA를 도입하고 관심 균주를 선택하는 과정은 약 14-21일이 소요된다. 조작이 "상처 없이"되어야 하는

경우, 돌연변이를 전달하는데 사용되는 플라스미드 요소(예를 들어, 플라스미드 백본)는 초기 통합 후에 "페이로드(pay-load)"만 남기고 제거해야 한다. "페이로드"는 단일 뉴클레오티드 다형성(SNP), 유전자 프로모터의 변화, 리보솜 결합 부위의 변화, 유전자 터미네이터의 변화, 다중 유전자 카세트, 약 1-10000bp 크기를 갖는 임의의 유전자 요소, 임의의 크기의 결실일 수 있는 바람직한 돌연변이이다. 전달 플라스미드 요소를 제거하면 조작 공정에 ~20일이 더 추가된다. 일부 경우에, 중성 부위에서 완전한 유전자 통합의 경우와 같이, 전달 플라스미드 요소의 제거가 즉시 요구되지는 않는다. 이러한 경우에, 플라스미드 및 플라스미드-코딩된 선택 가능한(kanR) 및 카운터 선택 가능한(sacB) 마커는 숙주 염색체에서 유지되고, 돌연변이는 "마킹된" 것으로 간주된다. 이러한 돌연변이를 조합하는 전통적인 방법은 통합 및 카운터 선택(~45일)을 통해 제 1 돌연변이를 기본 균주로 생성하여 돌연변이 균주(예를 들어, Mut1)를 생성한 후 수령자로서 Mut1 균주를 사용하며 45일의 조작45 일의 엔지니어링 과정을 다시 수행하여 과정을 다시 수행하는 다음 돌연변이를 이용하여 과정을 반복함으로써 2개의 돌연변이(예를 들어, Mut2)를 갖는 새로운 균주를 생성하는 것이다. 3번째 돌연변이를 추가하려면 최소 45일이 더 소요된다.

[0885] 본 발명은 유전자 변화의 신속한 통합을 통해 숙주 세포의 균주 개량 프로그램을 가속화하는 새로운 방법을 교시한다. 조작 시간을 단축하기 위해, 본 발명자들은 합리적으로 조작된 돌연변이의 신속한 통합 방법을 디자인(기존 방법에 대해 개량됨)하였다. 새로운 방법은 이전에 조작된 균주 및/또는 "좋은" 돌연변이(들)를 갖는 균주와 같은 선택된 균주의 원형질체 융합에 기초한다.

[0886] ***A. 일반적인 방법***

[0887] 돌연변이를 통합하기 위한 예시적인 공정이 도 30에 입증되어 있다. 출발점으로서, 관심 돌연변이를 함유하는 게놈을 갖는 부모 균주가 생성되고 선택되었다.

[0888] 일부 실시태양에서, 이러한 돌연변이 중 하나가 마킹되는 것이 바람직하다. 일단 균주가 생성되고 테스트되면, 최고의 돌연변이는 본 발명에 요약된 과정을 사용하여 신속하게 통합될 수 있다. 간략하게, 원형질체는 관심 균주로부터 생성된 후, 상이한 비율로 함께 혼합되며, "마킹된" 균주는 마킹되지 않은 균주에 비해 훨씬 더 낮은 농도로 사용된다. 융합 후, 생성된 균주는 공정을 위해 변형된 매질에서 회수되고, "마킹된" 균주에 대해 선택이 적용되어 "마킹된" 돌연변이를 받지 않은 임의의 세포를 사멸시킨다. HTP 균주 QC는 이렇게 선택된 균주에 어떤 다른 혼합 돌연변이가 존재하는지 신속하게 결정할 수 있다. 대부분의 균주는 다른 돌연변이 중 적어도 하나 및 경우에 따라 하나 이상의 돌연변이를 포함하고 있다.

[0889] 이 과정은 일반적으로 균주를 생성하는데 7-10일이 걸리며, 단일 통합 반응은 혼합된 균주의 수에 따라 여러 가지 다른 유전자형을 초래할 수 있다. 예를 들어, 균주 M1, S1, S2 및 P1의 4방향 융합은 4개의 희귀 단일 돌연변이 및 10개의 상이한 조합을 초래할 수 있다: M1 S1; M1 S2; M1 P1; S1 S2; S1 P1; S2 P1; M1 S1 S2; M1 S2 P1; S1 S2 P1; M1 S1 S2 P1. 또한, M1에서의 마킹된 돌연변이의 선택이 적용되면 S1 S2; S1 P1; S2P1; S1 S2 P1 유형은 손실될 것이다.

[0890] 예를 들어, 본 발명에 기술된 방법은 다음 단계를 포함할 수 있다:

[0891] (1) 조작된 균주 풀에서 부모 균주를 선택하면 선택된 균주가 통합된다. 일부 실시태양에서, 적어도 하나의 균주는 "마킹된" 돌연변이를 갖는다. 부모에게 사용되는 흥미로운 균주는 후속 단계에서 유용한 균주가 생성될 가능성을 크게 증가시킨다.

[0892] (2) 통합될 균주로부터 원형질체(예를 들어,세포벽 제거 등)를 준비한다. 세포는 삼투적으로 안정화된 매질 및 완충액에서 성장시킬 필요가 있으며, 이는 완충액 및 매질이 종래 기술과 상이하다.

[0893] (3) 관심 균주 융합. 일부 실시태양에서, 돌연변이체의 유용한(신규) 조합을 생성할 확률을 증가시키기 위해, "마킹된" 돌연변이를 갖는 균주의 더 적은 세포가 사용될 수 있으며, 따라서 이들 "마킹된" 세포가 상이한 돌연변이를 운반하는 세포와 상호작용하고 융합될 가능성이 증가된다. 이것은 세포가 서로 융합되고 통합되는 단계이다. 이 단계에서 사용되는 균주의 정확한 비율은 특정 조합을 얻을 가능성에 영향을 미칠 것이다.

[0894] (4) 세포의 회복. 일부 실시태양에서, 세포는 한천 오버레이를 사용하지 않고 삼투적으로 안정화된 매질 상에 플레이팅되어, 공정을 단순화하고 보다 용이한 자동화를 가능하게 한다. 삼투-안정제는 카운터 선택 마커 유전자(예를 들어, sacB 유전자)를 함유할 수 있는 세포의 성장을 가능하게 하는 정도이다. 원형질체 세포는 치료에 매우 민감하고 죽이기 쉽다. 이 단계는 충분한 세포가 복구되도록 한다. 이 단계가 잘 작동할수록 다운스트림 분석에 더 많은 재료를 사용할 수 있다.

[0895] (5) "표시된" 돌연변이를 갖는 세포의 선택. 이는 성장하는 세포에 적절한 항생제를 뿌려서 달성된다. 부모 세포 중 어느 것도 "마킹된" 돌연변이를 보유하지 않는 경우, 균주는 관심 균주를 확인하기 위한 다른 수단에 의해 유전자형 분석될 수 있다. 이 단계는 선택적일 수 있지만, 이는 세포 융합이 일어날 가능성이 가장 높은 세포가 풍부하게 되는 것을 보증한다. 다중 유전자좌를 "마킹"하는 것이 가능하며 이러한 방식으로 관심 조합을 더 빨리 생성할 수 있지만, 이후 "상처없는(scarless)" 균주를 원한다면 다중 플라스미드를 제거해야 할 수 있다.

[0896] (6) 다른 부모 균주로부터 오는 돌연변이의 존재에 대해 성장 세포의 유전자형 분석. 이 단계는 통합해야할 다른 돌연변이의 존재를 찾는다. 유전자형에 대한 콜로니의 수는 선택 계획뿐만 아니라 교차(cross)의 복잡성에 의존한다.

[0897] (7) (선택적) "마킹된" 돌연변이로부터 플라스미드의 제거. 이것은 선택 사항이며 추가 확인 또는 클라이언트 전달에 권장된다. 일부 실시태양에서, 균주에 대한 조작 사이클의 끝에서, 모든 플라스미드 잔재물을 제거할 필요가 있다. 이것이 언제 그리고 얼마나 자주 수행되는지는 사용자의 재량에 달려 있다. 일부 실시태양에서, 카운터 선택 가능한 sacB 유전자의 존재는 이 단계를 간단하게 만든다.

[0898] 생성된 균주는 원하는 관심 표현형에 대해 테스트될 수 있다. 게놈에 유전적으로 매우 가까운 돌연변이는 통합하기가 더 어려울 것이다. 성공적으로 통합될 가능성을 높이기 위해 통합을 위해 어떤 돌연변이가 선택되는지 아는 것이 신중해야한다. 또한, 여기에 설명된 단계 2, 3, 4는 성공에 필수적이며 건너뛰거나 제대로 실행되지 않으면 프로토콜의 결과가 된다.

[0899] 일부 실시태양에서, 돌연변이 중 어느 것도 "마킹"되지 않는다. 예를 들어, 돌연변이와 유전적으로 연결된 마커는 없다. 각각 독특한 마킹되지 않은 돌연변이를 함유하는 총 N(N≥3)의 상이한 균주가 통합될 때, 본 발명의 방법은 재귀 셔플 링 이벤트를 통해 서클 시간의 감소를 제공하고 상이한 게놈 사이의 재조합 기회를 최대화한다. 이 경우, 상기 방법은 다음 단계를 포함한다: (1) 조작된 균주 풀로부터 부모 균주를 선택, 이후 선택된 균주가 통합될 것이다; (2) 통합될 균주로부터 원형질체를 준비(예를 들어, 세포벽 제거 등); (3) 관심 균주 융합. 이 단계에서, 세포는 함께 융합되고 통합이 일어난다; (4) 세포 회수; (5) 관심 돌연변이 중 적어도 하나를 보유하는 세포의 선택. 이는 유전자형 분석 또는 관심 돌연변이를 확인하기 위한 다른 적절한 수단에 의해 수행될 수 있다; (6) 다른 부모 균주에 대해 오는 추가적인 하나 이상의 관심 돌연변이가 있는 세포의 선택.

[0900] 원형질체를 생성하는 방법은 Kieser et al. (Practical Streptomyces Genetics, John Innes Center, ISBN0708406238)에 기술되어 있으나, 이에 제한되는 것은 아니다.

[0901] **B. 결과**

[0902] 하나의 실험에서, 1개의 마킹된 균주 및 3개의 마킹되지 않은 균주는 각각 표시된 유전자좌와 상이한 거리에서 SNP 돌연변이를 보유한다. 융합된 원형질체는 항생제의 존재하에 선택되어 모든 마킹되지 않은 균주를 죽였다. 그런 다음 유전자 교환을 확인하기 위해 각 SNP의 유전자좌를 시퀀싱할 것이다. 특정 이론에 구애받지 않고, 유전자좌가 잘 분리되어 있으면 교환이 더 빈번할 수 있다.

[0903] 다른 실험에서, 상이한 균주로부터 유래된 융합된 원형질체를 제조하기 위해, 1%의 마킹된 균주 및 99%의 마킹되지 않은 균주가 혼합되어 선택될 것이다. 상대적인 스피노신 생산은 통합된 돌연변이를 갖는 선택된 균주에서 테스트될 것이고, 부모 균주(마킹된 및 마킹되지 않은 부모 균주 모두)와 비교될 것이다. 결과는 다양성이 발생했음을 나타낼 것이다: 일부 균주는 부모 모두에 비해 성능이 좋을 것이며, 일부 균주는 성능이 저하되거나 동등하게 나타날 것이다.

[0904] 세번째 예에서, 셔플링에 의해 생성된 표현형 다양성이 관찰되고 도시될 것이다. "마킹된" 부모로부터 마커를 보유한 세포만이 이 배지에서 성장할 것이다. 콜로니 형태(굵은-불투명한 색상 및 포자화(백색) 세포)와 콜로니 크기(대형 및 소형)에서 관찰된 차이는 셔플링 이벤트를 나타낸다. 세포는 R2YE Sorb/Man 매질에서 회수될 sacB 마커와 같은 카운터 선택 마커를 함유한다.

[0905] **실시예 10: 사카로폴리스포라 스피노사에서 사용하기 위한 리포터 단백질 및 관련 분석**

[0906] S. 스피노사는 공학 도구의 개발 및 이 유기체에 대한 공학 노력을 지원하는데 필요한 분자 생물학 도구가 거의 없는 매우 다루기 힘든 숙주이다. 리포터 단백질은 이 유기체에 대해 부족한 중요한 도구를 나타낸다.

[0907] 본 발명자들의 유전 공학 노력 및 보다 크게 대사 공학의 목적은 원하는 생성물의 수율을 향상시키기 위해 숙주 대사를 변경하고, 생합성 경로를 최적화하며, 경로 유전자를 도입 또는 복제하는 것이다. 성공은 생합성 유전자

클러스터의 내부(온-경로) 및 외부(오프-경로)에서 유전자의 발현을 교란시키고 균형을 잡는 능력에 의존하거나 도입된 비-천연 유전자 또는 유전자의 복제물의 과발현에 의존한다. 이러한 노력에는 공학 디자인에 사용될 수 있는 라이브러리 유전자(DNA) 요소(예를 들어, 프로모터, 리보솜 결합 부위, 전사 터미네이터)의 개발 및 특성화가 필요하다. 이들의 발현을 평가하기 위한 리포터 단백질 및 분석은 이들 라이브러리의 특성화에 필수적이다.

[0908] 이 실시예에서, 본 발명은 사카로폴리스포라 스피노사에서 3개의 리포터 유전자의 입증 및 정량적 평가를 제공한다. 여기에 설명된 세 가지 리포터 유전자에는 두 개의 형광 리포터 단백질((Dasher GFP and Paprika RFP; ATUM, https://www.atum.bio/products/protein-paintbox?exp=2) 및 효소 베타-글루쿠로니다제(gusA)가 포함된다(Jefferson et al. (1986). "Beta-Glucuronidase from Escherichia coli as a gene-fusion marker". Proceedings of the National Academy of Sciences of the United States of America. 83 (22): 8447-51.). 본 발명은 이러한 마커가 S. 스피노사에서 분자 도구로서 성공적으로 사용된 것을 처음으로 나타낸다. 본 발명은 또한 S. 스피노사에서 GusA 발현의 정량적 평가를 가능하게 하는 비색 분석의 최적화 및 적용을 기술한다.

[0909] 대서GFP(ATUM) 및 파프리카 RFP(ATUM)를 코딩하는 뉴클레오티드 서열을 대장균에 대해 코돈-최적화 하였다. (서열번호 81 및 서열번호 82). 베타-글루쿠로니다제(gusA)를 코딩하는 뉴클레오티드 서열은 S. 스피노사(서열번호 83)에 대해 코돈 최적화되었다.

[0910] 리포터 유전자를 테스트하기 위해, ermE* 프로모터(서열번호 149)를 리포터 코딩 서열 앞에서 클로닝하고, 생성된 제작물을 S. 스피노사 게놈의 공지된 중성 부위에 통합시켰다. 생성된 균주를 액체 배양물(성장 매질)에서 48시간 동안 성장시켰다. 배양물의 분취액을 PBS로 세척한 다음, 다음 중 하나를 수행하였다 (1) Tecan Infinite M1000 Pro(Life Sciences) 플레이트 판독기를 사용하여 96-웰 플레이트에서 복제 배양물의 분취액에 대해 형광 측정을 수행; (2) 락토바실러스 종에 대한 변형된 OpenWetWare 프로토콜(http://www.openwetware.org/wiki/Beta-glucuronidase_protocols)에 따라, 4-니트로페닐 β-D-글루쿠로니드의 존재하에 37℃에서 인큐베이션한 후 무-세포 추출물의 흡광도(OD_{405})를 측정.

[0911] 리포터의 형광 대서GFP 및 파프리카RFP는 이러한 리포터를 함유하도록 조작된 S. 스피노사 균주에서 측정되었다. 결과는 두 리포터 모두 S. 스피노사에서 작동하며 뚜렷한 형광 특성을 가지고 있음을 보여준다(도 31a-d 참조). 리포터 대서GFP 및 파프리카RFP를 코딩하는 뉴클레오티드 서열이 대장균에 대해 최적화되었지만 S. 스피노사에서 단백질이 발현되었기 때문에 이는 예상치 못한 결과이다. 우리가 다른 리포터 유전자를 선택한 경우에는 그렇지 않았을 수 있다. 또한, 선택된 형광 단백질은 S. 스피노사에서 관찰된 내인성 형광의 스펙트럼과 겹치지 않는 스펙트럼을 가졌다(도 36).

[0912] S. 스피노사에서 최적화된 베타-글루쿠로니다제(gusA)의 GusA 활성을 락토 바실러스 종(Jefferson et al. (1986). "Beta-Glucuronidase from Escherichia coli as a gene-fusion marker". Proceedings of the National Academy of Sciences of the United States of America. 83 (22): 8447-51.)에 사용하기 위해 개발된 비색 4-니트로페닐 β-D-글루쿠로니드 분석을 사용하여 측정하였다. 결과는 락토바실러스 종에 사용하기 위해 개발된 4-니트로페닐 β-D-글루쿠로니드 분석(세포 용해 및 효소 반응 포함)이 또한 S. 스피노사에서도 작동함을 나타낸다(도 35).

[0913] GusA 분석 프로토콜은 다음과 같이 간단히 기술된다:

[0914] 1. OD600이 0.6과 1.0 사이가 될 때까지 배양한다

[0915] 2. 다음을 추가함으로써 10mL의 GUS 완충액(샘플 10개 측정)을 준비한다:

[0916] ● 5mL의 인산나트륨 완충액(pH=7)

[0917] ● 3mL H20

[0918] ● 염화칼륨 용액 1mL

[0919] ● 황산마그네슘 용액 1mL

[0920] ● 35μL β-머캅토에탄올

[0921] ● 20mg 리소자임

[0922] 3. 1분 동안 원심분리하여 1.5ml의 배양물을 펠렛화한다.

[0923] 4. 다음을 포함하는 1ml 100mM 인산나트륨 완충액에 재현탁시킨다:

[0924] ● 0.1M 염화칼륨 용액

[0925] ● 10mM 황산 마그네슘 용액

[0926] ● 1M Na2CO3

[0927] ● 4-니트로페닐 β-D-글루쿠로니드(4-NPG) 저장 용액(50mM 인산나트륨 완충액(pH=7) 중 10mg/ml) 오직 1mL만 생성!!!

[0928] ● β-머캅토에탄올

[0929] ● 10% 트리톤 X-100(수중)

[0930] 5. 원심분리에 의해 다시 펠렛화한다.

[0931] 6. 750μL GUS 완충액에 재현탁한다.

[0932] 7. 짧게 와류(Vortex)하여 혼합한다.

[0933] 8. 37℃ 수조에서 30분 동안 배양한다.

[0934] 9. 10% 트리톤-X 8ul을 첨가한다.

[0935] 10. 짧게 와류하고 얼음에서 5분간 배양한다.

[0936] 11. 80ul의 4-NPG 용액을 넣고 타이머를 시작한다.

[0937] 12. 37℃ 수조에서 배양한다.

[0938] 13. 색상이 투명한 노란색(10분에서 30분 사이)인 경우 300μL 1M Na2CO3를 추가하여 반응을 중지시킨다.

[0939] 14. 시간을 기록한다.

[0940] 15. 최고 속도로 1분 동안 반응물을 원심분리한다.

[0941] 16. 상청액의 OD405를 측정한다.

[0942] 본 발명은 이러한 라이브러리의 정량적 평가를 가능하게 하지만 다른 잠재적 인 응용(예를 들어, 바이오 센서의 개발 및 콜로니의 선별, 및 유전자-편집 기술의 입증을 위한 마커 및 표적)을 갖는다. 여기에 설명된 세 리포터는 S. 스피노사에 사용하기 위해 개발된 최초의 리포터 유전자와 정량 분석이다. 또한 다른 생물학적 시스템에서 일반적인 리포터가 되는 이점이 있으므로 이미 탐지에 최적화된 기존의 방법과 도구를 사용할 수 있다.

[0943] **실시예 11: HTP 게놈 공학 - 사카로폴리스포라 스피노사에서 표적화되고 효율적인 게놈 통합을 위한 인테그라제 기반 시스템**

[0944] 외인성 DNA의 통합은 균주 성능을 개량하기 위한 효과적인 방법이지만, 이는 S. 스피노사, 특히 큰 DNA 조각(>10kb)에 대해 매우 비효율적이다. S. 스피노사와 같은 숙주에서 생합성 경로를 복제하고 리팩토링하는 능력은 대사 공학 노력에 중요하지만, 이러한 경로의 크기는 이러한 노력을 제한한다.

[0945] 본 실시예는 S. 스피노사의 게놈 내로 유전자 요소의 통합을 위한 인테그라제-기반 시스템을 기술한다. 인테그라제는 보존된 부위(att 부위; 전형적으로 숙주 염색체에서 tRNA 유전자 내에 위치한 보존된 뉴클레오티드 서열)에서의 인식 및 부착을 통한 DNA 페이로드의 표적화된 통합을 지시한다. 본 발명에 의해 기술된 인테그라제-기반 시스템은 수십 킬로베이스 크기의 유전자 페이로드의 전달을 가능하게하여, 이종 유기체로부터의 외인성 DNA의 효율적인 도입 또는 S. 스피노사로부터의 천연 유전자의 복제를 가능하게 할 것으로 예상한다. 우리는 다음 선택된 인테그라제 중 하나 또는 여러 개가 게놈의 특정 부위에 DNA를 효율적으로 도입할 수 있음을 보여줄 수 있을 것으로 기대한다.

표 10

[0946] S. 스피노사의 게놈으로의 유전자 요소 통합을 위한 인테그라제

인테그라제	기원	서열
pCM32 [1]	*S. endophytica*	pCM32integrase+*attP* (서열번호 84) atgccgcgtaagaaccgcgatgaaggcacccgggcgcccaacggcgcgagcagcatctacaagggcaaagacg gctactggcacggccgcgtctggatgggcaccaaggacgacggcagtgaggaccgtcgccacaggtcagcgaa gagcgaaacagagctcctcaataaggttcgcaagctcgaacgggagcgggacagcggcaaggtgcagaagcct ggccgcgcctggaccgtcgagaaatggcttacgcactgggtggagaacatcgccgctcccaccgtgcggccga ccacgatggtcggctaccgcgcctcggtgtataagcatctgatccccggcgtgggcaagcaccggatcgacag gttgcagccggaacacctcgaaaagctctacgccaagatgcagcgcgatggactcaaggccgcgacagcgcac ctcgcgcaccggacggtgcgggtcgcgctgaacgaggccaagaagcgacgtcacatcaccgagaacccggcca atatcgcgaagccgcccagggtggacgaggaggagattgtcccgttcacggtggatgaagcccgccggatcct cgcagcagctgcggagacgcggaacggcgctcgctttgtcatcgcgctgacccttggcctgcgcaggggtgaa gcactcgggttgaagtggtcggatctctcgatcacctggaagcacggatgccggaaggggagcgcgtgccggg tgggtcgccgagccgagcagtgcggcgagcgtcgcggcagcggcacgctcgtcatccggcgcgcgattcagca gcaggtttggcagcacggttgctcagaggacaagccgtgcgaccaccgctacggcgctcactgcccgcgccgg catagcggcggtgtggtcgtgaccgatgtgaagtccagggcgggtcggcgaaccgtgggccttccgcacccgg tggtggaagcgctcgaagagcaccgcgcccgccagcggacagagcgggagaaggcgcgcaacgagtgggacga cgccgattgggtcttcacgaacaggtggggtcgcccggttcatccgaccgttgactacgacgcctggaaggca ctgctcagggcagcgaacgtgcgcaacgcgcggttgcacgacgcacgccacaccgcggcgacgatgttgctgg tgttgaaggttccgctgcctgcggtcatggaaatcatgggctggtcggaagcctctatggccaagcgctacat gcacgtgccgcacgagctcgtgaccgcgatcgcggaccaggtgggtgacctggtgtggcccgtcccagagacc gaggaggaggcgccaccgcctgaggaggagtgggcgctggacgccaaccaggtggcggcgatccggaagctgg ccggagctctcccgccgcagttgcgggagcagttcgaggcgctgctgcccggcgacgacgaggacgacggccc gacttcgggagtggtcatccctgcgtaaccagtgcggccagaacccggcctaacggggcctactgagacgaaa actgagactggacat**gcgagaggcccggaagcgagatcgcttccgggcctctgacctgcggaggatacgggat tcgaacccgtgagggctattaacccaacacgatttccaattccgatggcgcgagtgccaggggtagctgaac gtgccttttgcctggtcag**tggcactacggcaacatcaggtgtggcttgatccgtgcgcgt >pCM32integrase_protein (서열번호 85) MPRKNRDEGTRAPNGASSIYKGKDGYWHGRVWMGTKDDGSEDRRHRSAKSETELLNKVRKLERERDSGKVQKP GRAWTVEKWLTHWVENIAAPTVRPTTMVGYRASVYKHLIPGVGKHRIDRLQPEHLEKLYAKMQRDGLKAATAH LAHRTVRVALNEAKKRRHITENPANIAKPPRVDEEEIVPFTVDEARRILAAAAETRNGARFVIALTLGLRRGE ALGLKWSDLSITWKHGCRKGSACRVGRRAEQCGERRGSGTLVIRRAIQQQVWQHGCSEDKPCDHRYGAHCPRR HSGGVVVTDVKSRAGRRTVGLPHPVVEALEEHRARQRTEREKARNEWDDADWVFTNRWGRPVHPTVDYDAWKA LLRAANVRNARLHDARHTAATMLLVLKVPLPAVMEIMGWSEASMAKRYMHVPHELVTAIADQVGDLVWPVPET EEEAPPPEEEWALDANQVAAIRKLAGALPPQLREQFEALLPGDDEDDGPTSGVVIPA* >attP site in pCM32(서열번호167) Gcgagaggcccggaagcgagatcgcttccgggcctctgacctgcggaggatacgggattcgaacccgtgaggg ctattaacccaacacgatttccaattccgatggcgcgagtgccaggggtagctgaacgtgcctttgcctgg tcag

pSE101[2]	*S. erythraea*	pSE101integrase+attP (서열번호 86) atgccccgcaaacgccgcccagaaggcacccgagcccccaacggcgccagcagcatctactacagcgagacgg acggctactggcacgggcgcgtcacgatgggcgtccgcgacgacggcaagcccgaccgtcgccacgtccaagc caagaccgagaccgaggtcatcgataaggtccgcaagctcgaacgtgaccgggatagcggcaacgcgcggaag cctggtcgcgcgtggacagtcgagaagtggctgactcactgggtcgagaacatcgcggtgcactccgttcggt acaagacgcttcagggctaccgaacggcggtctacaagcacctgatccccggtatcggcgcgcaccggatgga ccgcatcgagccggagcacttcgagcggttctacgccaggatgcaggccgccggcgccagtgcagggaccgca catcaggtgcaccggactgccaaaacggcattcaacgaatacttccggcggcagcgcatcaccgggaacccca tcgccttcgtgaaagcgccgcgcgtcgaggaaaaggaagtggagccgttcacgccgcaggaagccaagagcat catcacggccgcgctcaagcggcgcaacggcgtgcgatacgtcgtcgccttggctctcggttgtcgccaaggc gaagccctggggttcaagtgggaccgcctcgaccgcgggaaccggctttaccgcgtacggcaggcattgcagc ggcaggcttggcaacacggatgcgacgacccgcacgcctgcggagcacgacttcatcgggtggcgtgcccgga caactgcacccagcatcgcaaccgcaagagctgcattcgcgacgagaagggccaccaccgtccgtgcccgccg aactgcaccaggcacgcgagcagttgcccgcagcggcacggtggtgggctcgtcgaggtcgacgtgaagtcga aggctggtcgccggagcttcgttctgccagatgaggtcttcgatctgctgatgcgccacgagcaggcgcagca gcgggagcgcaagcacgccggtagcgagtggcaggagggggctgggtcttcacccagcccaacggccggccg atcgatccgcggcgcgactggggtgagtggaaggacatcttgggggaggcaggtgttcgggatgctcggctgc acgacgcgcgccacactgcggcgacggtcctcatgctgctccgcgttccagaccgggccgtccaggatcacat gggctggtcctcgatccggatgaaggagcgctacatgcacgtcaccgaggaactgcgacgagagatcgccgat cagctcaacgggtacttctgggacgtcaactgagacggaaagtgagacgaaaagcgcctggtcagggacctgt cgacggcgtttccgctggtagtt**tcggagccgctgaggggactcgaacccctgaccgtccgcttacaa**ggcgg gcgctctaccaactgagctacagcggcgtgcgctacgtcgcgcgcgaacatcgtaagcgtccacc >pSE101integrase_protein (서열번호 87) MPRKRRPEGTRAPNGASSIYYSETDGYWHGRVTMGVRDDGKPDRRHVQAKTETEVIDKVRKLERDRDSGNARK PGRAWTVEKWLTHWVENIAVHSVRYKTLQGYRTAVYKHLIPGIGAHRMDRIEPEHFERFYARMQAAGASAGTA HQVHRTAKTAFNEYFRRQRITGNPIAFVKAPRVEEKEVEPFTPQEAKSIITAALKRRNGVRYVVALALGCRQG EALGFKWDRLDRGNRLYRVRQALQRQAWQHGCDDPHACGARLHRVACPDNCTQHRNRKSCIRDEKGHHRPCPP NCTRHASSCPQRHGGGLVEVDVKSKAGRRSFVLPDEVFDLLMRHEQAQQRERKHAGSEWQEGGWVFTQPNGRP IDPRRDWGEWKDILGEAGVRDARLHDARHTAATVLMLLRVPDRAVQDHMGWSSIRMKERYMHVTEELRREIAD QLNGYFWDVN* >attP site in pSE101(서열번호168) Tcggagccgctgaggggactcgaacccctgaccgtccgcttacaa

pSE211[3] 및 [4]	*S. erythraea*	>pSE211integrase+*attP* (서열번호 88) acgtcacccaactcgccgccacgctcgcctcgctcgcggccctgctcgccgaacagcagcccgccccggaacc cgagcccgaaccggccgcccgcaggctgcccaaccgcgtgctgctcacggtcgaggaagcggccaagcaactg gggctcggcaggaccaagacctacgcgctggtggcgtctggcgagatcgaatctgtccggatcggtcggctca ggcgcatcccgcgcaccgccatcgacgactacgccgcccgactcatcgcccagcagagcgccgcctgaaggga accactatggaacaaaagcgcacccgaaaccccaacggtcgatcgacgatctacctcgggaacgacggctact ggcacggccgcgtcaccatgggcatcggcgacgacggcaagcctgaccggcgccacgtcaagcgcaaggacaa ggacgaagttgtcgaggaggtcggcaagctcgaacgggagcgggactccggcaacgtccgcaagaagggccag ccgtggacagtcgagcggtggctgacgcactgggtggagagcatcgcgccgctgacctgccggtacaagacca tgcggggctaccagacggccgtgtacaagcacctcatccccggtttgggcgcgcacaggctcgatcggatcca gaaccatccggagtacttcgagaagttctacctgcgaatgatcgagtcgggactgaagccggcgacggctcac caggtacaccgcacggcgcgaacggctttcggcgaggcgtacaagcggggacgcatccagaggaacccggttt cgatcgcaaaggcacctcgggtggaagaggaggaggtcgaaccgcttgaggtcgaggacatgcagctggtcat caaggccgccctggaacgccgaaacggcgtccgctacgtcatcgcactggctctcggaactcggcagggcgaa tcgctcgcgctgaagtggccgcggctgaaccggcagaagcgcacgctgcggatcaccaaggcactccaacgtc agacgtggaagcacgggtgctctgacccgcatcggtgcggcgcgacctaccacaagaccgagccgtgcaaggc ggcctgcaagcggcacacgcgagcttgtccgccgccatgcccgccagcttgcaccgaacacgcccggtggtgc ccgcagcgaaccggtggcgggctggtcgaggtcgacgtcaagtcgagggctggacgacggaccgtgacgctgc ccgaccaactgttcgacttgatcctcaagcacgaaaagcttcaggggccgaacgggagctcgcgggcacgga gtggcacgacggcgagtggatgttcacccagcccaacggcaagccgatcgatccacgtcaggacctcgacgag tggaaagcaatccttgttgaagccggagtccgcgaggcgcggctacatgacgcacggcacaccgccgcgactg tgctgttggtcctcggagtgcccgaccgggtcgtgatggagctgatgggctggtcgtccgtcaccatgaagca gcggtacatgcacgtcatcgactccgtccggaacgacgtagcggaccgcctgaacacctacttctggggcacc aactgagacccagactgagacccaaaacgcccccgtcgagatcgacgggggcgtttt**ggcagctcttggtggt ggccaggggcggggtcgaaccgccgaccttccgctt**ttcaggcggacgctcgtaccaactgagctacctggcc gttcgcgcccggctcaaagccgaaccgctgtggcgacccagacgggactcgaacccgcgacctccgccgtgac agggcggcgcgctaaccaactgcgccactgggccatgttctgttgttgcgtacccccaacgggattcgaaccc gcgctaccgccttgaaagggcggcgtcctaggccgctagacgatgggggcttggccgattcggaaccgacccg gcctcgcctccaaccggctttccctttcggggcgccccgttgggagcagtgaaagcttacgacacacccccca gcgccccacaacgggggggtccccaaacctcacgagcccccgcgcggcccacgcccgccggtcacgtcggtcg ccaccatatgccatctgaccagccttttccatcgcctatcctcagtcggcccact >pSE211integrase_pro (서열번호 89) MEQKRTRNPNGRSTIYLGNDGYWHGRVTMGIGDDGKPDRRHVKRKDKDEVVEEVGKLERERDSGNVRKKGQPW TVERWLTHWVESIAPLTCRYKTMRGYQTAVYKHLIPGLGAHRLDRIQNHPEYFEKFYLRMIESGLKPATAHQV HRTARTAFGEAYKRGRIQRNPVSIAKAPRVEEEEVEPLEVEDMQLVIKAALERRNGVRYVIALALGTRQGESL ALKWPRLNRQKRTLRITKALQRQTWKHGCSDPHRCGATYHKTEPCKAACKRHTRACPPPCPPACTEHARWCPQ RTGGGLVEVDVKSRAGRRTVTLPDQLFDLILKHEKLQGAERELAGTEWHDGEWMFTQPNGKPIDPRQDLDEWK AILVEAGVREARLHDARHTAATVLLVLGVPDRVVMELMGWSSVTMKQRYMHVIDSVRNDVADRLNTYFWGTN* >attP site in pSE211(서열번호169) ggcagctcttggtggtggccaggggcggggtcgaaccgccgaccttccgctt

pSE101 homolog	*S. spinosa*	>SS101_homolog(3g00449)_CDS+*attP* (서열번호 90)atgccacgcaaacgccgcccggaaggcacccgggcacccaacggagccagcagcatctacctcggcaagg acggctactggcacggccgcgtcaccgtcggagttcgcgacgacggtaagcccgaccgccctcacgtccaggc caagaccgaggccgaagtcatcgacaaggtgcgcaagctcgaacgcgatcgcgatgcggggaaggtgcgaaag cctggccgggcctggaccgtcgagaagtggcttacgcactgggtcgagaacatcgccgcgccatccgtccgtt acaagacccttcagggctaccgcacggcggtgtacaagcacttgatccccggcatcggcgcgcaccggatcga ccgaattgaaccggagcacttcgagaagctctacgcgaagatgcaggaatccggcgcgaaagcgggaaccgcg caccaggtgcaccgcaccgctcgggccgcctttaacgaagccttccggcgtcggcacctcaccgaaagcccgg tgcggttcgtgaaagcgccgaaggtcgaagaagaggaagtcgagcccttcacgccgaaggaagcccagcagat cattacggccgcgctcaatcgtcgaaacggcgtgcgattcgtgatcgctctcgcactgggctgccgccagggt gaagcgctgggcttcaagtgggaacggctcgaccgggaaaacaggctctaccacgttcggagggcgcttcagc gtcaagcctggcaacacggctgtgaagatccgcacaactgcggtgcgaggttccaccgggttgcttgcgccga gaactgcaagcggcaccgcaatcggaagaactgcattcgcaacgagaagggacacgctcgaccgtgcccgccg aactgcgaccgacacgccagcagctgcccgaaacggcacggcggaggcctgcgcgaggtggatgtgaagtcga aggctggccgccggcggttcgttcttcctgacgagatcttcgacctgctcatgcggcatgaggaagtccagcg gcacgaacgggttcacgccggtaccgagtggcaggagggcggctggatcttcacgcagcccaacggcaggccg atcgatccgcgccgcgattggggcgagtggaaggagatcctcgcggaggccggtgttcgggatgcccggctgc acgacgcgcggcacaccgcagcgacggtgctcatgctgctccgtgttccggaccgggccgttcaggaccacat gggatggtcgtcgatccggatgaaagagcggtacatgcacgtcaccgaggaactgcgccgcgagatcgccgat cagctgaatgggtatttctggaaccccaactgagaccgaaagtgagacggatcgcgcctggtcaccgggtggg caggcgcgtttccgctggtacgg**tcggagccgctgaggggactcgaacccctgaccgtccgcttacaaggc**gg gcgctctaccaactgagctacagcggcatgcacttcgtcgtgcggggacatcgtaagcggcgat >SS101_homolog(3g00449)_protein (서열번호 91) MPRKRRPEGTRAPNGASSIYLGKDGYWHGRVTVGVRDDGKPDRPHVQAKTEAEVIDKVRKLERDRDAGKVRKP GRAWTVEKWLTHWVENIAAPSVRYKTLQGYRTAVYKHLIPGIGAHRIDRIEPEHFEKLYAKMQESGAKAGTAH QVHRTARAAFNEAFRRRHLTESPVRFVKAPKVEEEEVEPFTPKEAQQIITAALNRRNGVRFVIALALGCRQGE ALGFKWERLDRENRLYHVRRALQRQAWQHGCEDPHNCGARFHRVACAENCKRHRNRKNCIRNEKGHARPCPPN CDRHASSCPKRHGGGLREVDVKSKAGRRRFVLPDEIFDLLMRHEEVQRHERVHAGTEWQEGGWIFTQPNGRPI DPRRDWGEWKEILAEAGVRDARLHDARHTAATVLMLLRVPDRAVQDHMGWSSIRMKERYMHVTEELRREIADQ LNGYFWNPN* >attP site in pSE101 homolog (서열번호170) tcggagccgctgaggggactcgaacccctgaccgtccgcttacaaggc

pSE211 homolog	*S. spinosa*	>SS211_homolog(3g00347)_CDS+*attP* (서열번호 92) atgccacgcaagcgccgcccggaaggcacccgggcacccaacggagccagcagcatctacctcggaaacgacg gctactggcacggccgcgtcacgatgggaacccgtgacgacggccgccccgaccgacggcatgtccagggcaa gaccgaggccgaagtcatagacaaagtgcgcaagctcgaacgcgaccgcgacgccggacggatgcgcaagcct ggccgggcctggaccgtcgagaagtggctgatgcactggctggagcacattgcgaagccatcggtccggccga aaaccgtcgcccggtatcggacttccgtcgagcaatacctgattcctggtctcggtgcgcaccgcatcgaccg cttgcagccggagaacattgagaagctgtacgcaaaattgctcgctcgcgggttggcgccgtccactgtgcac catgttcaccggactctgcgcgtcgctttcaacgaggcgttcaagcgggaacacatcacgaaaaacccggtcc tcgttgcgaaagcgccgaagctggtcgaaccggagatcgagccgttcaccgtggccgaagcacaacgaattct cgatgttgcacggacacggcggaatggtgctcggttcgcactcgcgctcgcgctgggaatgcgccagggcgaa gctctcggactcaagtggtccgacctgcgaatcacctggcaccacgggtgcgcatccggactcaccgaagaac agcaggcggccatcgaaatgctcgcgaaggtcgatccgcagcgatggaagcggcctgacgattccgggtgcgg attcaaggacgtggaggactgcccgcaggctcacccggccgcgacactgaacattcggcgcgcattgcagcgc cacacctggcaacacgggtgcggtgacaaaccgacgtgcggcaagaaacggggcgcggactgcccgcagcgtc atggcggcggcttggccatcgtcccggtgaagtcgagggcggggacgcgctcgatcagcgtgcctgagccgct gattcatgcgttgctcgatcacgacgaggcgcaggatgaggaacggcacttggcccggaacctgtggcacgac gatggatggatgttcgctcagcccaacgggaaggcgacggacccgagggccgactatggcgaatggcgcgagc tgctggacgccgcgaaggttcggccggcgcggctgcacgacgcgcggcacaccgccgcgacgatgttgctggt tctcaaggtcgcaccacgggcaatcatggacgtgatgggctggtcggaggcgtcgatgctgacccgctacgtc cacgtgccggacgagatcaagcagggcatcgcgggccaggtcggcggactgctgtggaaggactggcagcagc ccgacgacggcccagacgacgaggacggcggcaccgccgggcaccctgtcccggcctgacgtgcccactgcca gaggaggcgtttgagccggaaactgagccggaacgacaccaggcgctttccgtgtccacggaaagcgcctggt gagagc**ggagccgcctaagggaatcgaacccttgacctacgcattacgagtgcgtcgctctagccgactgagc taaggcggc**gttgcacggccaagtgtagcgggccggacctcgccgtcgttcatggccccgact >SS211_homolog (3g00347)_protein (서열번호 93) MPRKRRPEGTRAPNGASSIYLGNDGYWHGRVTMGTRDDGRPDRRHVQGKTEAEVIDKVRKLERDRDAGRMRKP GRAWTVEKWLMHWLEHIAKPSVRPKTVARYRTSVEQYLIPGLGAHRIDRLQPENIEKLYAKLLARGLAPSTVH HVHRTLRVAFNEAFKREHITKNPVLVAKAPKLVEPEIEPFTVAEAQRILDVARTRRNGARFALALALGMRQGE ALGLKWSDLRITWHHGCASGLTEEQQAAIEMLAKVDPQRWKRPDDSGCGFKDVEDCPQAHPAATLNIRRALQR HTWQHGCGDKPTCGKKRGADCPQRHGGGLAIVPVKSRAGTRSISVPEPLIHALLDHDEAQDEERHLARNLWHD DGWMFAQPNGKATDPRADYGEWRELLDAAKVRPARLHDARHTAATMLLVLKVAPRAIMDVMGWSEASMLTRYV HVPDEIKQGIAGQVGGLLWKDWQQPDDGPDDEDGGTAGHPVPA* >attP site in pSE211 homolog (서열번호171) ggagccgcctaagggaatcgaacccttgacctacgcattacgagtgcgtcgctctagccgactgagctaaggc ggc

[1] Chen J, Xia H, Dang F, Xu Q, Li W, Qin Z. Characterization of the chromosomal integration of Saccharopolyspora plasmid pCM32 and its application to improve production of spinosyn in *Saccharopolyspora spinosa*. Applied Microbiology and Biotechnology. PMID 26260388 DOI: 10.1007/s00253-015-6871-z

[2] Brown DP, Chiang SJ, Tuan JS, Katz L (1988a) Site-specific integration in *Saccharopolyspora erythraea* and multisite integration in Streptomyces lividans of actinomycete plasmid pSE101. J Bacteriol 170:2287-2295

[3] Brown,D.P., Idler, K.B. and Katz, L. (1990) Characterization of the genetic elements required for site-specific integration of plasmid pSE211 in *Saccharopolyspora erythraea*. J. Bacteriol., 172, 1877-1888.

[4] Te Poele E. M., Bolhuis H., Dijkhuizen L. (2008) Actinomycete integrative and conjugative elements. Antonie Van Leeuwenhoek 94, 127-143.

[0947] pCM32 인테그라제는 S. 스피노사에서 작동하는 것으로 나타났다(Chen *et al.*, "Characterization of the chromosomal integration of Saccharopolyspora plasmid pCM32 and its application to improve production of spinosyn in *Saccharopolyspora spinosa*. Applied Microbiology and Biotechnology." PMID 26260388 DOI: 10.1007/s00253-015-6871-z). pCM32 부착 부위와 99% 동일성을 갖는 부착 부위가 S. 스피노사 게놈에서 발견되기 때문에 이는 놀라운 일이 아니다(도 38). Chen 등은 개량된 스피노신 역가를 갖는 균주를 초래하는 2개의 유전자의 표적화된 통합을 달성하였다(전체가 참고로 인용된 특허 출원 CN105087507A 참조).

[0948] pSE101 및 pSE211 인테그라제 및 이들의 부착 부위가 설명되었다. pSE10 및 pSE211 부착 부위 둘 다의 코어는 S. 스피노사(각각 도 39 및 도 40)에서 발견된다. 이 통합 시스템은 테스트되었으며 작동하지 않았다. 본 발명자는 수정된 시스템 및 다른 통합 시스템을 테스트할 것이다.

[0949] pCM32, pSE101 및 pSE211을 사용하여 S. 스피노사로 서열의 통합을 위한 벡터는 도 37에 기술되어 있다. 유사하게, S. 스피노사의 pSE101 상동체(homolog) 또는 pSE101 상동체를 사용하는 벡터도 구축될 수 있다. 이들 벡터는 외인성 DNA를 S. 스피노사의 게놈에 통합시키는 능력을 조사하기 위해 테스트될 것이다.

[0950] 본 발명에 기재된 방법에 의해 생성된 통합 외인성 DNA를 함유하는 S. 스피노사 균주는 사카로폴리스포라 스피노사에서 균주 성능을 개량하기 위한 기초로서 사용될 수 있다. 예를 들어, 이러한 균주는 HTP 시스템에서 상기 예에 기술된 SNP 스왑 라이브러리, 프로모터 스왑 라이브러리 및/또는 터미네이터 라이브러리와 조합되어 스피노신과 같은 원하는 생성물의 개량된 생산을 갖는 새로운 S. 스피노사 균주를 생성할 수 있다.

[0951] 표 10에 기술된 통합 시스템이 테스트되었고 작동하지 않았다. 본 발명자는 수정된 시스템 및 다른 통합 시스템을 테스트할 것이다.

[0952] **실시예 12. 사카로폴리스포라 스피노사에 대한 자기 복제 플라스미드 시스템의 복제 기점**

[0953] 본 실시예에서, 복제 기점 및 복제 요소(예를 들어, 플라스미드 복제에 필요한 효소를 코딩하는 유전자)가 제공된다. 이러한 유전적 요소는 S. 스피노사에서 복제 기능을 제공할 수 있으며, 따라서 S. 스피노사에 대한 자기 복제 플라스미드 시스템의 구축을 가능하게 할 수 있다. 자기 복제 플라스미드 시스템은 이 숙주에서 수행될 수 있는 유전 공학 및 선별의 유형을 향상시킬 것이다.

[0954] S. 스피노사에 현재 부족한 하나의 중요한 분자 유전자 도구는 자기 복제 플라스미드 시스템이다. 플라스미드 시스템은 여러 가지 방법으로 S. 스피노사의 조작 능력을 확장시킬 것이다. 예를 들어, 이는 (1) 대사 공학 디자인를 테스트하기 위해 상동성 재조합에 의한 성공적인 통합의 필요성을 제거할 수 있다(예를 들어, 유전자 복제 또는 이종 효소가 플라스미드 시스템을 사용하여 도입되어 숙주 표현형에 대한 영향을 결정할 수 있음); (2) 라이브러리(유전자, 프로모터, 터미네이터 또는 리보솜 결합 부위)의 보다 신속한 선별을 가능하게 한다; (3) 이는 사용자가 플라스미드 시스템의 제어하에 CRISPR 시스템 구성요소를 도입할 수 있게 하여 CRISPR 기반 게놈 편집을 용이하게 할 것이다.

[0955] S. 에리트라에아(Vara et al., 1989, "Cloning of genes governing the deoxysugar portion of the erythromycin biosynthesis pathway in *Saccharopolyspora erythraea* (Streptomyces erythraeus". J. Bacteriol. 171, 5872-5881.) 및 pIJ101에서 광범위하게 사용되는 자기 복제 플라스미드 pWHM4를 포함하는 밀접하게 관련된 종으로부터의 다른 플라스미드, 스트렙토마이세스 리비단스(Streptomyces lividans)(Kieser et al., 1982, "pIJ101, a multi-copy broad host-range Streptomyces plasmid: functional analysis and development of DNA cloning vectors." Mol Gen Genet 185:223-228)로부터의 다중-복사 광범위 숙주-범위 플라스미드가 S. 스피노사에서 사용하기 위해 조사된 바 있지만 우리의 지식에 성공적으로 사용되지 않았다.

[0956] 일부 실시태양에서, 복제 기점 소스는 S. 에리트라에아에서 발견된 추정 염색체 복제 기점 및 S. 에리트라에아의 플라스미드 pSE10 및 pSE211에서 액티노마이세트 통합 및 공액 요소(AICE)를 포함한다(Te Poele *et al.*, (2008) Actinomycete integrative and conjugative elements. Antonie Van Leeuwenhoek 94, 127-143.), 도 41a 참조. 액티노마이세트 통합 및 공액 요소(AICES)는 사카로폴리스포라 종을 포함하여 액티노마이세트에서 공통적인 이동 유전자 요소이다. 이들 요소는 게놈에 통합되어 있거나 자율적인 자기 복제 플라스미드로서 발견될 수 있다.

[0957] 이들 추정 복제 기점을 테스트하기 위해, 항생제 저항성 마커를 함유하는 플라스미드 및 자기 복제에 필요한 추정 복제 기점 +/- 다른 유전자(예를 들어, AICE의 경우)를 조립하였다. 조립된 플라스미드를 S. 스피노사로 전달하고 플라스미드를 갖는 형질전환체를 선별하기 위해 항생제 선택을 사용하였다. 플라스미드의 유지 및 안정성을 확인하기 위해 PCR을 사용하였다. 예시적인 플라스미드가 도 41b에 도시되어 있다. 이러한 추정 복제 기점이 테스트되었으며 작동하지 않았다. 본 발명자는 수정된 디자인 및 다른 추정 복제 기점을 테스트할 것이다.

[0958] **실시예 13. HTP 게놈 공학 - 사카로폴리스포라에서 스피노신 생산에서의 균주 성능을 향상시키기 위한 리보솜 결합 부위(RBS) 라이브러리의 구현**

[0959] 이전의 실시예는 산업용 균주를 재건시키기 위한 본 발명의 HTP 균주 개량 프로그램의 힘을 입증하였다. 실시예 2 및 3은 다양한 염기, 중간체 및 산업용 균주 내에서 기존의 유전적 다양성을 탐구하는 SNP 스왑 기술 및 라이브러리의 구현을 설명하였다.

[0960] 이 실시예는 본 발명의 리보솜 결합 부위 라이브러리 기술을 사용하는 HTP 균주 개량 프로그램의 실시태양을 예시한다.

[0961] **A. RBS 라이브러리를 적용하기 위한 표적의 식별**

[0962] RBS 라이브러리를 적용하는 것은 표적에 대한 "n" 유전자 세트를 선택하는 단계를 포함하는 다단계 공정이다.

[0963] 본 발명자들은 본 발명의 프로모터 래더 방법을 통해 조절할 수 있는 잠재적 경로 유전자 그룹을 확인하였다. (실시예 4 및 도 12a 내지 도 12d 참조).

[0964] **B. RBS 라이브러리 생성**

[0965] 본 발명의 유전자 공학 노력 및 대사 공학의 주요 목표는 숙주 대사를 변경하고, 생합성 경로를 최적화하며, 목적 생성물의 수율을 향상시키기 위해 경로 유전자를 도입 또는 복제하는 것이다. 성공은 생합성 유전자 클러스터의 내부(온-경로) 및 외부(오프-경로)에서 유전자의 발현을 교란시키고 균형을 잡는 능력에 의존하거나 도입된 비-천연 유전자 또는 유전자의 복제물의 과발현에 의존한다. S. 스피노사에는 특징적인 RBS를 포함하여 이용 가능한 유전자 도구가 제한되어 있다. 본 발명은 S. 스피노사에서 단백질 발현의 통합 및 조정을 위한 다중 유전자 폴리시스트로닉 오페론의 디자인를 가능하게 하는 유전 공학 도구이다.

[0966] 리보솜 결합 부위(RBS)는 리보솜을 동원하고 단백질의 번역을 개시하는 mRNA 전사체에서 개시 코돈의 업스트림에 위치한 짧은 뉴클레오티드 서열이다. 따라서, 이들은 번역 및 단백질 발현의 중요한 조절자이다. 그러나, RBS는 또한 유전자의 프로모터 또는 코딩 영역인 5'UTR에서 인근 뉴클레오티드와 상호작용하여 전사 및/또는 번역 속도에 영향을 줄 수 있다. 이러한 상호작용 및 결과적인 2차 구조를 통해, 리보솜 결합 부위는 유전자의 발현을 "조정"할 수 있다.

[0967] RBS 라이브러리는 합성 생물학 도구의 공통 구성요소이며 다양한 유기체를 위해 개발되었다. 또한, 관심 유전자와 유리하게 상호작용할 합성 RBS를 예측하기 위한 도구가 개발되었다(Salis *et al.*, "Automated design of synthetic ribosome binding sites to control protein expression." Nat Biotechnol. 2009; 27:946-950. doi: 10.1038/nbt.1568.). 그러나, 이것은 S. 스피노사에 대해 기술되고 특징 지어진 최초의 이러한 라이브러리 및 최초의 천연 RBS이다.

[0968] 추정 천연 RBS를 확인하기 위해, 폴리시스트로닉 오페론에서 유전자 사이의 START 코돈 또는 유전자 간 영역으로부터 업스트림의 뉴클레오티드 서열을 선택하였다. RBS는 문헌(Luo *et al.*, "Comparative proteomic analysis of *Saccharopolyspora spinosa* SP06081 and PR2 strains reveals the differentially expressed proteins correlated with the increase of spinosad yield." Proteome SCI. 2011, 9: 1-12)으로부터의 프로테오믹 데이터에 기초하여 고도로 발현될 것으로 예상되는 유전자, 또는 스피노신 생산과 관련된 유전자에 대해 선택되었다. 예측 분석 시점에서 PATRIC 데이터베이스(https://www.patricbrc.org/)에서 사용 가능한 주석을 기반으로 하였다. 기능성에 대한 척도를 구성하는 선택적 매질에서 카운터 선택 가능한 마커(sacB) 수준의 성장을 사용하여 RBS를 분석하였다.

[0969] 이 실시예에서, 본 발명자들은 S. 스피노사 및 관련 숙주에서 사용하기 위해 다양한 정도의 번역 활성을 갖는 19개의 리보솜 결합 부위(RBS)의 라이브러리를 생성하였다. 라이브러리는 이전에 특성화되지 않은 S. 스피노사에 고유한 상이한 숙주 및 서열에서 이전에 기재된 합성 서열로 구성된다:

표 11

[0970] RBS 서열, 소스, 크기 및 상대 기능의 요약

	유전자	RBS 서열	(bp)	기능
RBS1	PermE*	aggaggtcccat (서열번호 97)	12	+
RBS2	spnA (폴리케타이드 합성효소 로딩 & 확장 모듈1)	ccaggaatcggaggggcagtaccga (서열번호 98)	25	++
RBS3	spnC (폴리케타이드 합성효소 확장 모듈 3-4)	gcaacttcctggagggaaacgccac (서열번호99)	25	++
RBS4	spnO (추정 NDP-헥소오스-2,3-디하이드라타아제)	tcgtcacggcagtgagggattgggc (서열번호 100)	25	+
RBS5	gdh (gdh (dTDP-글루코스 4,6-디하이드라타아제))	cgaaatcccggcgaggaagggcgcg (서열번호 101)	25	++
RBS6	linker_A (알데하이드 탈수소효소, AldA)	cgcctcggccccttcaggaggagacag(서열번호 102)	28	++

RBS7	linker_B (아세토락테이트 합성효소)	ctccagacgcccacgcaaggagaccc (서열번호 103)	26	++
RBS8	linker_C	actagtaaggaggtccaac(서열번호 104)	19	++
RBS9	linker_D	aagaggtatatatta(SE Q ID NO. 105)	15	-
RBS10	gtt (글루코스-1-포스페이트 티미딜일트랜스퍼라제 1)	ccaccgctggaggtatccgg(서열번호 106)	20	++
RBS11	TDH (글리세르알데하이드-3-포스페이트 탈수소효소)	aggagagatcggc(서열번호107)	13	+
RBS12	BioBrick_1	aaagaggagaaa(서열번호 108)	12	++
RBS13	BioBrick_2	attaaagaggagaaa(서열번호 109)	15	++
RBS14	GroES (분자 샤페론 GroES)	agaaggtggaggtcacacc(서열번호 110)	19	++
RBS15	GroEL (분자 샤페론 GroEL)	aagggctgttggaatc(서열번호 111)	16	-
RBS16	IF-1 (번역 개시 인자 IF-1)	attgaggtcgagggtcgg(서열번호 112)	18	-
RBS17	XNR_1700 (주변세포질 뮤레인 펩타이드-결합 단백질 전구체)	ggcggtgaatgatccgccgcgc(서열번호 113)	22	++
RBS18	S20 (30s 리보솜 단백질 S20)	gacgaggaagaggcgccaca(서열번호 114)	20	++
RBS19	S12 (리보솜 단백질 S12)	acgttacgctcgtcgc(서열번호 115)	15	NA
RBS20	S12 (리보솜 단백질 S12)	gggacgttacgctcgtcgc(서열번호 116)	19	++
RBS21	DnaK (Hsp 70)	tcgtgacctcggtgctgaaca(서열번호 117)	21	-
RBS22	신장 인자 Tu	aggaggaacaatcca(서열번호 118)	15	NA
RBS23	F0F1 ATP 합성효소 서브유닛 베타	ccgcaggaagtgagtgac(서열번호 119)	18	NA
RBS24	분자 샤페론 DnaK	cgtgacctcggtgctgaaca(서열번호 120)	20	-
RBS25	파지 충격 단백질 A, PspA	aattcccggggatctacc(서열번호 121)	18	-
RBS26	2-옥소글루타레이트 디카복실라아제	cgaggcgaacgcagcc(서열번호 122)	17	-
RBS27	5-메틸테트라하이드로프테로일트리글루타메이트 호모시스테인 메틸트랜스퍼라제	gcgaaggagagcccccc(서열번호 123)	16	++
RBS28	50S 리보솜 단백질 L7/L12	ccgaaaggaacgccgac(서열번호 124)	17	++
RBS29	DNA-지시된 RNA 폴리머라제 서브유닛 알파	gaggaaaggaaaacgaa(서열번호 125)	17	++
RBS30	30S 리보솜 단백질 S5	gaacggaagggacgcctg(서열번호 126)	18	NA
RBS31	DnaK (6929)	cggcgggtcggagaggagtgc(서열번호 127)	21	-
RBS32	Negative_1 (ermE 단독)	-	0	

참고: 라이브러리는 26개의 천연 RBS(관련 유전자 ID가 있는 것)를 포함한다. 나머지 5개의 서열은 합성 또는 이종 공급원으로부터 유래된다.

(-) = 기능하지 않음; (+) = 덜 기능성; (++) = 기능성

"NA"는 데이터가 없는 RBS를 나타낸다

[0971] 따라서, 본 발명은 다중 유전자 폴리시스트로닉 통합에서 유전자 사이의 스페이서로서 요구되는 기능성 RBS 서열의 다양한 라이브러리를 제공한다. 이들 RBS의 강도의 서열 다양성 및 변동은 이들을 사용하여 프로모터와 유전자 사이에 상이한 RBS를 삽입함으로써 유전자의 발현을 상하로 조절할 기회를 제공한다.

[0972] **C. 라이브러리로부터의 RBS를 표적 유전자와 연관시키기**

[0973] RBS 라이브러리의 구현에서 다른 단계는 특정 표적 유전자와 관련된 RBS 라이브러리로부터 주어진 RBS를 포함하

는 다양한 균주의 HTP 조작이다.

[0974] 천연 RBS가 표적 유전자 n 앞에 존재하고 이의 서열이 공지된 경우, 천연 RBS를 라이브러리의 각 RBS로 교체할 수 있다. 천연 RBS가 존재하지 않거나 이의 서열이 알려지지 않은 경우, 유전자 n 앞의 라이브러리에서 각각의 RBS의 삽입이 수행될 수 있다. 이러한 방식으로, 균주의 라이브러리가 구성되며, 여기서 라이브러리의 각 구성원은 동일한 유전적 맥락에서 n 표적에 작동 가능하게 연결된 RBS의 예이다.

[0975] **D. 균주의 HTP 선별**

[0976] RBS 라이브러리를 적용하는 최종 단계는 상기 라이브러리에서 균주의 HTP 선별이다. 유래된 균주는 각각 동일한 유전자 배경에서 n 표적에 연결된 RBS의 예를 나타낸다.

[0977] 하나 이상의 메트릭에 대한 그들의 성능이 특성화되는 시나리오에서, 각각의 균주의 HTP 선별을 구현함으로써, 본 발명자들은 주어진 메트릭에 대해 어떤 RBS/표적 유전자 회합이 가장 유리한지를 결정할 수 있을 것이다(예를 들어, 관심 분자 생산의 최적화).

[0978] 이 공학적 접근법의 유용성을 입증하는 데이터는 도 63에 도시되어 있다. 리보솜 결합 부위는 번역 효율을 변경하기 위해 다수의 표적화된 유전자의 업스트림에 삽입되었고, 이들 조작된 균주는 폴리케타이드 생산성에 대한 플레이트 분석에서 부모 균주와 비교하여 테스트되었다. 몇몇 "RBS 스왑" 균주는 부모 균주와 비교하여 개량을 나타내었다.

[0979] **실시예 14 - HTP 게놈 공학 - 사카로폴리스포라에서 균주 성능을 개량시키기 위한 트랜스포존 돌연변이유발 라이브러리의 구현.**

[0980] 이 실시예는 S. 스피노사에서 생체 내 트랜스포존 돌연변이유발에 의해 균주 라이브러리를 생성하는 방법을 기술한다. 개량된 표현형을 나타내는 균주(예를 들어, 스피노신과 같은 특정 화합물의 역가)를 확인하기 위해 결과 라이브러리를 선별할 수 있다. 균주는 순환 공학의 라운드에서 또는 균주 성능에 기여하는 유전자형을 해독하기 위해 추가로 사용될 수 있다. 라이브러리의 균주는 또한 상기 실시예 3에서 사용된 SNP 스왑 라이브러리와 유사하게, 하나 이상의 원하는 화합물의 생성이 증가된 개량된 균주의 생성을 위해 상이한 유전자 교란(들)을 갖는 다른 균주와의 통합에 사용될 수 있다.

[0981] 따라서, 본 발명은 S. 스피노사에서 EZ-Tn5 트랜스포좀 시스템(Epicenter Bio)을 사용하여 트랜스포존 돌연변이유발 미생물 균주 라이브러리를 생성하는 방법을 기술한다. 트랜스포사제 효소는 먼저 모자이크 요소(ME) 서열에 의해 측면에 있는 DNA 페이로드 서열과 복합체화 될 수 있고, 생성된 단백질-DNA 복합체는 세포로 형질전환될 수 있다. 이것은 DNA 페이로드를 유기체의 게놈 DNA에 랜덤하게 통합시킨다. 도입할 페이로드에 따라 LoF(Loss-of-Function) 라이브러리 또는 GoF(Gain-of-Function) 라이브러리를 생성할 수 있다.

[0982] 기능상실(LoF) 트랜스포존 라이브러리 - 페이로드의 서열은 다양한 표현형 반응을 이끌어 내기 위해 변화될 수 있다. 기능상실(LoF) 라이브러리의 기초 사례에서, 이 페이로드는 성공적인 트랜스포존 통합 이벤트의 선택을 가능하게 해주는 마커를 포함한다.

[0983] 랜덤 기능상실 돌연변이는 Tn5 트랜스포사제 시스템(EZ-Tn5; EpiCentre®)을 사용하여 미생물에서 생체 내에서 이루어질 수 있다. EZ-Tn5 트랜스포사제 시스템은 안정적이고 전기 천공에 의해 살아있는 미생물에 도입될 수 있다. 세포 내에서, 트랜스포존 시스템은 숙주 세포에서 Mg2+에 의해 활성화되고 트랜스포존은 숙주의 게놈 DNA에 랜덤으로 삽입된다.

[0984] 기능획득(GoF) 트랜스포존 라이브러리 - GoF 라이브러리를 생성하기 위해, 프로모터 요소, 용해도 태그(이 경우, 기능획득 용해도 태그 트랜스포존이라고 함) 및/또는 선택 가능한 마커를 함유하는 페이로드의 일부의 루프-아웃을 용이하게 하여 일련의 트랜스포존 돌연변이유발(이 경우, 기능획득 재활용 가능 트랜스포존이라고 함)을 가능하게 하기 위한 카운터 선택 가능 마커와 같은 추가 특징을 통합함으로써, 유전자 페이로드의 보다 복잡한 화신이 기초 사례에 구축된다. 이러한 구현을 통해 다양한 라이브러리를 만들어 호스트 표현형을 향상시킬 수 있다.

[0985] 본 발명의 트랜스포존에 대한 비 제한적인 예시적인 제작물이 도 44에 도시되어 있으며, 대표적인 기능상실(LoF) 트랜스포존, 기능획득(GoF) 트랜스포존, 재활용 가능 트랜스포존, 및 기능획득 용해도 태그 트랜스포존의 서열이 각각 서열번호 128, 서열번호 129, 서열번호 130 및 서열번호 131로 제공된다. 이들 트랜스포존은 트랜스포사제와 복합체를 형성하여 세포로 형질전환될 수 있다. 생성된 세포는 DNA 페이로드의 랜덤 통합을 가지므로 트랜스포존 돌연변이유발 미생물 균주 라이브러리를 형성할 것이다. 라이브러리는 본 발명에 기술된 HTP 공

정에 따라 추가로 선별될 수 있고 표현형 개량에 대해 평가될 수 있다. 트랜스포존 통합으로 인해 원하는 표현형을 갖는 균주를 본 발명에 기재된 임의의 방법에 따라 추가 특성화 및 추가 공학을 위해 분리될 수 있다.

[0986] 예를 들어, LoF 트랜스포존 라이브러리 및 GoF 트랜스포존 라이브러리는 부모 균주에 대해 선별될 수 있고, 성능 데이터(스피 오신의 역가)가 분석될 수 있다. 이 라이브러리에서 생성된 새로운 균주 중 일부는 상위 균주에 비해 성능이 향상된다.

[0987] 본 발명에 기술된 방법은 2가지 주요 문제를 해결한다. 첫째, 잘 연구된 유기체에서도 게놈 환경의 많은 부분이 잘 이해되지 않은 채 남아 있다. 또한 유전자 요소가 예기치 않은 방식으로 상호작용할 수 있음을 알았다. 이를 위해, 본 발명은 표현형 교란의 유발을 위한 효과적인 유전 공학 방법을 제공한다. 둘째, 성장이 느리거나 유전적으로 재발성인 유기체-특히 큰 게놈을 가진 것들-는 모든 가능한 유전자 표적에 대해 표적화된 유전자 교란을 수행하는데 시간이나 비용이 많이들 수 있다. 본 발명은 교란된 게놈을 갖는 균주를 생성하는 효과적인 방법을 제공하며, 이는 균주에서 원하는 화합물을 생산하는데 있어서 개량된 성능을 야기할 수 있다. 따라서, 본 발명은 이러한 문제를 해결하고, 생체 내 트랜스포존 돌연변이유발을 사용하여 숙주 유기체의 유전자 요소를 쉽고 랜덤하게 조절하는 방법에 의해 수행된다. 이러한 방식으로, 상이한 돌연변이(기능획득 및 기능상실)를 보유하는 균주 라이브러리를 매우 신속하게 만들 수 있고, 숙주의 표현형을 추가로 개량시키기 위해 새로운 유전자 표적을 내포할 수 있다.

[0988] **실시예 15. 사카로폴리스포라에서 유전자 요소의 삽입을 위한 중성 통합 부위**

[0989] S. 스피노사에서의 공학 유전자 복제 및 리팩토링된 생합성 경로는 이 숙주에 대해 특징지어진 공지된 중성 통합 부위의 수에 의해 제한된다. S. 스피노사 게놈 내에 여러 개의 중성 부위가 존재할 가능성이 있지만, 현재까지 단 하나의 중성 통합 부위만 특징 지어졌다. 이 특정 부위인 obsA(US20100282624, 전체 내용이 참조로 포함됨)는 이전에 보고된 바 있지만, 추가 부위가 없으면 다수의 일련의 유전적 변화를 만드는 능력이 제한된다. 추가적인 중성 통합 부위는 다수의 조합 유전자 통합을 엔지니어링하고 테스트할 수 있는 능력과 속도를 향상시킬 것이다.

[0990] RNAseq 데이터(발효 동안 샘플링되고 2개의 균주에서 발현을 비교하는 복제 된 시계열)를 사용하여 균주에서 또는 발효 동안 임의의 시점에서 발현이 거의 또는 전혀 없는 다중 유전자 유전자좌를 확인하였다. 주요 근거는 발효 중 또는 균주에서 언제라도 발현되지 않은 유전자는 생산에 필수적이지 않거나 중요하지 않을 가능성이 높다는 것이다(도 45 참조). 따라서 이러한 유전자좌로의 통합은 표현형에 해로운 영향을 미칠 가능성이 없거나 적다. 일단 이러한 부위가 확인되면, 유전자좌는 참조 게놈 내에 위치하고 통합 제작물은 부위 중앙에 단일 염기쌍 돌연변이를 도입하도록 디자인되었다.

[0991] 따라서, 본 발명은 예를 들어, 컨쥬게이션 및 상동성 재조합에 의해 S. 스피노사의 게놈 내에서 개별 유전자 또는 다중 유전자 카세트가 안정적이고 효율적으로 통합될 수 있는 유전자 유전자좌와 같은 중성 통합 부위 세트를 제공한다. 중성 부위로 간주되기 위해 페이로드의 유전자 통합은 성장 및 예측 가능한 발현 수준에 제한적인 영향을 나타낸다. 우리가 확인하고 현재 조사중인 부위에는 게놈 전체에 분산된 11개의 유전자좌가 포함된다. 각 부위는 유전자 페이로드를 통합하는 통합 부위를 생성하여 유전자 엔지니어링 용량을 확장할 가능성이 있다. 이용 가능한 부위의 수는 전체 요인, 조합 유전자 통합 디자인에 포함할 수 있는 요소의 수에 비례하므로 엔지니어링 능력을 향상시킨다. 이들 부위는 아래 표 12에 요약된다.

표 12

[0992] 11개의 추정 중성 통합 부위, 관련 유전자, 도입된 돌연변이 및 통합 효율 - 각 부모 균주에 대한 콜로니-형성 단위(CFU)의 요약.

중성 부위	서열번호	돌연변이	CFU	
			A	B
1	132	C:G	88	47
2	133	T:A	50	33
3	134	C:G	91	17
4	135	C:G	67	28
5	136	G:C	16	0
6	137	G:C	129	18
7	138	C:G	84	32
8	139	T:A	-	-

9	140	A:T	94	25
10	141	C:G	55	41
11	142	A:T	-	-
주석은 우리가 표시된 돌연변이를 도입한 중성 부위의 중심에 있는 유전자를 지칭한다. **CFU**는 분할된(48웰) Q-Tray의 단일 웰에서 계산된 콜로니(ex-conjugants) 수를 나타낸다. * 컨쥬게이션이 진행 중(결과 보류 중)				

[0993] 부위는 발현이 거의 없거나 전혀 없는 다중 유전자좌(전사; mRNA) 내에 위치한다. 이들은 2개의 상이한 균주에서 유전자 발현을 비교하는 시계열의 RNAseq 데이터를 사용하여 확인되었다.

[0994] 통합 효율을 평가하기 위해, 단일 뉴클레오티드 다형성이 각 부위의 중심에 도입되었다. 균주 A 및 B의 각 부위에 대한 컨쥬게이션 효율이 보고되어 있다(표 12).

[0995] 생성된 균주 B-유래된 균주를 균주 B 부모와 비교하여 생성물 역가에 대해 평가하였다(도 67). 지시된 중성 부위에 통합된 SNP스왑 페이로드를 갖는 균주 B-유래된 균주의 생성물 역가(spinosyns J+L)를 분석하였다. 부위 1, 2, 3, 4, 6, 9 및 10에서 통합된 균주는 유사한 생성물 역가를 가지며 예상 역가(즉, 균주 B의 평균 역가; 도면에서 더 높은 막대)와 다르지 않았다. 중성 부위 7에서의 통합은 생성물 역가에 부정적인 영향을 미치는 것으로 보인다.

[0996] 이들 부위를 추가로 평가하고 통합된 페이로드의 발현을 비교하기 위해, 본 발명자들은 균주 A(WT)와 B(도 68)의 각 부위에서의 통합 후 강한 프로모터(서열번호 25)의 제어하에 형광 리포터(서열번호 81)의 발현을 평가하였다. 발현은 대부분의 부위에서 비슷하다. 오직 NS7만이 우리가 평가한 다른 중성 부위(NS2, NS3, NS4, NS6 및 NS10)와 크게 달랐다.

[0997] **실시예 16. HTP 게놈 공학 - 사카로폴리스포라에서 균주 성능 개량을 위한 항-대사산물 선택/발효 생성물 저항성 라이브러리의 구현**

[0998] 이 실시예는 사카로폴리스포라에서 유전자 다양성을 생성하기 위한 항-대사 물질 선택/발효 생성물 저항성 라이브러리를 생성하는 실시태양 및 이러한 라이브러리를 HTP 유전자 공학에 사용하는 방법을 예시한다.

[0999] 미생물에 의한 원하는 화합물의 생산을 향상시키기 위해, 합리적인 공학에 즉시 적용할 수 없는 형태의 분자 조절을 우회하는 것이 종종 필요하다. 예는 다른 경로에 의한 최종 산물 억제를 포함하다. 이 예에서, 우리는 S. 스피노사를 항-대사산물(알파-메틸 메티오닌) 또는 발효 생성물(스피노신 J/L) 및 이들 조건 하에서 성장이 개량된 단리된 콜로니에 노출시켰다. 이러한 콜로니의 고 처리량 선별은 부모 균주와 비교하여 플레이트 모델에서 더 나은 발효 성능을 갖는 단리물을 확인하였으며, 이는 전략이 균주 성능 개량에 잠재적으로 유용함을 나타냈다.

[1000] 미생물은 발효 공정의 일부로서 다양한 화합물을 생성한다. 때때로 이러한 화합물의 축적은 미생물의 성장과 생리를 심각하게 억제한다. 에탄올 생산은 발효 생성물에 의한 성장 억제(독성)의 예이다. 분자 수준에서 경로의 생성물은 종종 폐기물을 최소화하기 위해 생산을 담당하는 효소를 억제할 수 있다. 이것은 미생물의 진화와 생존에 유리하지만, 이러한 피드백 메커니즘은 목표가 특정 경로 및 생성물의 축적을 통해 플럭스를 근본적으로 증가시키는 것인 산업적 발효를 심각하게 방해할 수 있다(Fermentation Microbiology and Biotechnology, Third Edition, ISBN 9781439855799). 발효를 개량하고 미생물이 원하는 대사산물을 합성할 수 있는 시간을 연장하려면 다음을 극복하는 것이 필요하다 a) 최종 생성물의 잠재적 독성, 및 b) 원하는 최종 생성물의 형성에 필요한 분자 경로의 피드백 억제.

[1001] S. 스피노사 성장은 발효 생성물의 존재에 민감하다. (도 46). 우리는 제품에 대한 저항성을 개량하면 균주 생산성이 향상될 수 있다는 가설을 세웠다. 따라서, 발효 생성물을 더 잘 생존시킬 수 있는 균주를 선별하기 위해 하기에 요약된 단계를 수행하였다. 우리는 더 많은 저항성 균주를 분리하였다(도 47). 흥미롭게도 단리물들 중 2개가 스피노신 생산을 위한 플레이트 모델에서 부모보다 스피노신 J/L에 대해 훨씬 더 잘 수행되었다(도 48a). 우리는 또한 스피노신 생산을 위한 플레이트 모델에서 부모보다 대사산물 알파-메틸-메티오닌(aMM)에 대해 훨씬 더 잘 수행된 한 가지 균주를 분리하였다(도 48b).

[1002] 스피노신 생산에는 보조인자로서 NADPH 및 SAM이 필요하였다. 이들 중 하나가 스피노신 형성에 대해 제한요소가 될 수 있고, 각각은 각각의 합성을 담당하는 효소를 억제할 수 있기 때문에, 본 발명자들은 SAM에 의한 피드백 억제를 제거하는 방법을 찾았다. 대장균에서 SAM은 SAM에 대한 전구체의 합성을 담당하는 MetA 단백질을 억제할

수 있다. 대장균의 전형적인 접근법은 피드백 조절에 둔감한 metA 돌연변이체를 선택하는 항-대사산물 알파-메틸-메티오닌(aMM)의 존재 하에서 균주를 성장시키는 것이었다(Ususda and Kurahashi, 2005, Appl Env. Micro, June 2005, p3228-3234). S. 스피노사에는 명확한 metA 동족체가 없지만, S. 스피노사가 aMM에 민감하기 때문에, 우리는 SAM 축적을 증가시키고 더 나은 스피노신 생산을 기대하기 위해 유사한 접근법을 선택하고 저항성 돌연변이체를 선택하였다. 미생물에 의한 원하는 화합물의 생산을 향상시키기 위해, 합리적인 공학에 바로 따르지 않는 형태의 분자 조절을 우회하는 것이 종종 필요하다. 실시예는 다른 경로에 의한 최종 생성물 억제를 포함하다.

[1003] 특히, 본 발명자들은 부모 S. 스피노사 균주를 항-대사산물(예를 들어, 알파-메틸 메티오닌) 또는 발효 생성물(예를 들어, 스피노신 J/L)에 적용하였다. 먼저 선택제에 대한 S. 스피노사의 민감도 및 실험 조건 및 매질을 결정하였다. 사용된 농도 측면에서 적절한 시작점이 없으면 실험이 완전히 실패할 수 있다. 스피노신 J/L 실험의 경우, 성장을 억제했지만 세포를 죽이지 않는 농도를 찾기 위해 몇 주가 걸리고 여러 번 시도되었다. 용액은 스피노신 농도 및 접종에 사용되는 바이오 매스 양의 조정의 조합이었다. aMM은 최소한의 매질을 사용해야하며, 매질을 선택하기 전에 먼저 확인하고 검증해야하였다. 이 단계는 성공적인 선택 전략을 위한 토대를 마련한다. 최소 매질의 구성은 다음과 같다:

[1004] 성분(1L 당):

[1005] ■ 녹말, 용해성 10.0g

[1006] ■ 인산이칼륨 1.0g

[1007] ■ 황산마그네슘, 칠수화물 1.0g

[1008] ■ 염화나트륨 1.0g

[1009] ■ 황산암모늄 2.0g

[1010] ■ 탄산칼슘 2.0g

[1011] ■ 황산제일철, 칠수화물 0.001g

[1012] 일단 선택을 위한 농도가 결정되면, 상기 기재된 조건에서 보다 저항성 단리 물에 대한 선택이 수행되었다. 선택을 위해, 액체(7계대, ~40세대)에서의 선택을 위해 배양물의 다중 계대배양이 필요하였다. 부과된 선택을 만족시키는 독립적인 돌연변이 사건의 가능성을 증가시키기 위해 다수의 독립적인 배양물을 병렬로 유지 하였다. 각 통과의 지속시간 및 빈도는 경험적으로 결정될 수 있다. 선택 전략에 따라 어떤 형질이 선택되는지 결정된다. 좋지 않은 설계는 원하는 산업 조건에서 잘 수행되지 않는 균주의 선택을 초래할 수 있다. 선택의 성공을 향상시키기 위해서는 우수한 정렬 및/또는 완화 전략(2차 선별)이 필요하다. 스피노신 J/L의 존재하에 균주를 선택하는 예가 도 47에 설명되어 있다. 선택된 균주는 스피노신 J/L의 존재 하에서 부모보다 분명히 더 잘 자랐다.

[1013] 선택된 균주를 이들 단리물이 실제로 부모 균주보다 더 저항성이 있음을 입증하기 위해 추가로 확인하였다. 이러한 선택의 유효성 검사는 전략이 작동하고 있음을 나타내는 좋은 지표이며 다음 단계로 진행할 시기의 결정 지점으로 사용될 수 있다.

[1014] 다음으로, 선택된 균주를 HTP 선별에 의해 추가로 분석하여 선택된 특성이 원하는 산업 공정에 유리한지를 결정하였다. 세포는 많은 방식으로 특정 선택 문제를 해결할 수 있기 때문에, 대부분 산업적으로 관심이 없을 수 있으므로, HTP 선별은 추가 특성화되고 통합에 사용될 단리물을 확인하는데 중요한 단계이다. 첫 번째 연구에서 선택된 단리물의 ~2-5%만 관심 대사인 것으로 보인다. HTP 플레이트 발효 모델에서 부모보다 더 우수한 성능을 갖는 선택된 균주의 예는 도 48a(스피노신 J/L) 및 도 48b(aMM)에 도시되어 있다.

[1015] 선택적으로, 선택된 균주에서 개량된 성능을 야기한 돌연변이가 확인될 수 있고 관련 서열이 분리될 수 있다. 이는 이들 돌연변이를 본 발명에 기술된 다른 원하는 균주로 통합하는 것을 촉진할 것이다. 초기 테스트 결과는 도 69에 나와 있다.

[1016] **실시예 17. HTP 게놈 공학 - 상처없는 돌연변이 균주의 생성을 위한 S. 스피노사에서의 카운터 선택 마커로서 sacB 또는 pheS의 사용**

[1017] 이 실시예는 카운터 선택 마커로서 sacB 또는/및 pheS를 사용하여 "상처없는" 돌연변이체 사카로폴리스포라 균

주를 생성하는 실시태양을 예시한다.

[1018] 이전에 당업계에 기술된 US20170101659A1은 온도 민감성 복제 기점 및 선택 마커와 같은 도구를 사용하여 폴리케타이드 합성효소 유전자 유전자좌에서 생산성을 개량하기 위한 폴리케타이드 생성 균주를 논의한다. 본 방법론의 상세한 요건 및 제약(PKS 코딩 영역의 반복적 특성에 의존하는 것을 포함함) 및 당업계의 제한된 추가의 예는 S. 스피노사와 같은 산업상 관련 미생물의 공학적 과제를 설명한다. 그러나, 게놈의 임의의 위치에서 정확한 게놈 편집은 유기체의 표현형 개량의 적용을 포함하여 S. 스피노사 숙주 균주에서 의도된 변형을 만들 수 있는 것이 중요하다. 또한 저항성 마커 재활용은 제한된 저항성 마커가 존재하는 단일 균주에서 유전자 변형을 쌓을 수 있게 하며, 제조 응용(즉, 항생제 저항성이 없는)에서 이들 미생물의 등록을 용이하게 하는데에도 중요하다. 본 실시예에서, 본 발명자들은 상처없고, 마커가 없는 지시된 게놈 편집을 가능하게 하기 위해 게놈 내의 임의의 위치를 표적화하는데 사용되는 상동성 암과 함께 카운터 선택 마커로서 sacB 및/또는 pheS의 사용을 입증한다(도 49a 내지 도 49c 참조).

[1019] sacB 유전자는 수크로스를 레반으로 전환시키는 레반수크라제를 암호화하고, 이는 많은 미생물에 독성이 있는 것으로 알려져 있다(Reyrat et al., "Counterselectable Markers: Untapped Tools for Bacterial Genetics and Pathogenesis", Infect Immun. 1998 Sep; 66(9): 4011-4017, and Jager et al., "Expression of the Bacillus subtilis sacB gene leads to sucrose sensitivity in the gram-positive bacterium Corynebacterium glutamicum but not in Streptomyces lividans.", J Bacteriol. 1992 Aug;174(16):5462-5). 수크로스가 없는 경우, sacB 유전자의 담체는 건강한 방식으로 성장하고, 수크로스의 존재 하에서 sacB 유전자를 상실한 균주만이 생존할 수 있다. 이 개념은 많은 그람 음성 미생물에 많이 사용되었지만 그람 양성 미생물(코리네박테리움 글루타미쿰 및 마이코박테리움 종 제외)은 일반적으로 레반의 영향에 저항성이 있다. 우리는 여기서 sacB 유전자가 S. 스피노사에서 2-3 로그의 수크로스 민감성을 부여한다는 것을 입증한다(도 50). 따라서, 우리의 실험은 sacB가 마커 없는 균주 생성을 위해 S. 스피노사에서 카운터 선택 가능한 마커로서 활용될 수 있음을 나타낸다. sacB 유전자 서열은 S. 스피노사(서열번호 143)에 대해 코돈 최적화되었다.

[1020] pheS 유전자는 박테리아가 4-클로로페닐알라닌(4CP)에 민감하게 만드는 페닐알라닌-tRNA 합성효소의 α 서브유닛을 암호화한다(Miyazaki, "Molecular engineering of a PheS counterselection marker for improved operating efficiency in *Escherichia coli*." Biotechniques. 2015 Feb 1;58(2):86-8.). 4-클로로페닐알라닌이 없는 경우, pheS 유전자의 담체는 4-클로로페닐알라닌의 존재하에 건강한 방식으로 성장하지만, pheS 유전자를 잃은 균주만이 생존할 수 있다. 우리는 여기서 사카로폴리스포라 에리트라에아에서 유래된 pheS 유전자의 돌연변이 된 버전이 S. 스피노사에서 4-클로로페닐알라닌 민감성을 부여하고 따라서 카운터 선택 가능한 마커로서 활용될 수 있음을 보여준다(도 51). 미야자키 2015에 기술된 돌연변이를 갖는 S. 에리트라에아 및 S. 스피노사 둘 다로부터 유래된 pheS 유전자를 테스트하고 기능적인 것을 확인하였다(각각 서열번호 144 및 서열번호 145).

[1021] 균주 조작을 위한 벡터 백본은 배경 균주 특성(예를 들어, 기초 균주 저항/선택 및 카운터 선택 제제에 대한 민감도)에 따라 균주 조작 효율을 변경하기 위해 다수의 구성(도 49a 내지 도 49c)으로 디자인되었다. 이것은 코딩된 마커의 발현을 변경시키기 위해 상이한 프로모터로 발현된 하나 또는 모든 카운터 선택 가능한 유전자를 사용하는 것을 포함한다.

[1022] 이 도구는 본 발명의 HTP 시스템에 적용되어 조작된 상처없는 S. 스피노사 균주를 생성하였고, 품질 관리 결과는 도구의 성공적인 적용을 보여준다(도 52). 따라서, 여기에는 S. 스피노사에서 카운터 선택 마커로서 sacB 및 pheS의 사용 및 이들의 유전자 편집에의 적용이 기재되어 있다. 카운터 선택 가능한 마커, 또는 음성 선택 마커의 미생물 발현은 특정 기질(sacB 및 pheS에 대해 각각 수크로스 및 4-클로로페닐알라닌)에서 제한된 성장을 야기하므로, 카운터 선택 마커를 함유하지 않는 미생물의 선택을 가능하게 한다. sacB와 pheS는 다른 숙주의 문헌에서 카운터 선택 마커로 설명되어 있지만, 우리가 아는 한, 이것은 S. 스피노사에서의 이들의 사용의 첫 번째 특성화이다. 여기에서, 우리는 HTP 게놈 공학을 위한 강력한 도구 인 S. 스피노사에서 표적화된 상처없는 유전자 편집을 수행하기 위해 상동성 재조합을 조합하여 카운터 선택을 사용한다.

[1023] **실시예 18: 사카로폴리스포라의 HTP 컨쥬게이션 및 사카로폴리스포라로의 내인성 DNA 도입의 시연**

[1024] 이 실시예는 본 발명의 HTP 유전자 조작 방법의 실시태양을 예시한다. 특히, 공여자 유기체로서 대장균을 사용하여 사카로폴리스포라(예를 들어, S. 스피노사)의 종간 컨쥬게이션을 위한 고 처리량 공정을 보여준다. 이 과정을 통해 자동화 및 자동화 호환 배양 형식을 사용한 사카로폴리스포라(예를 들어, S. 스피노사)의 유전자 변형이 가능하여 단일 교차 상동성 재조합에 의해 유전 물질을 도입할 수 있다.

[1025] S. 스피노사는 산업적으로 관련된 숙주이고, 본 발명은 이 숙주에서의 게놈 공학을 위한 고도로 병렬화된 노력이 실현될 수 있게 한다. 결과는 본 발명의 방법이 숙주 세포의 전체 게놈에 걸쳐 임의의 종류의 신속한 유전자 변화를 생성하거나 임의의 외인성 DNA를 도입할 수 있음을 입증한다.

[1026] 종간 컨쥬게이션(일명 인터제네릭 컨쥬게이션)은 사카로폴리스포라에서의 유전자 전이, 특히 그의 강력한 제한 장벽을 우회하기 위한 효과적인 메카니즘이다. 그러나, 현재의 컨쥬게이션 방법론은 상대적으로 효율이 낮으며 완료를 위한 수동 공정을 필요로 한다(즉, 한 번에 10회 미만의 수동 조작자에 의해 완료됨). 이 연구의 목표는 S. 스피노사에서 컨쥬게이션 효율을 개량하고 S. 스피노사에서 고 처리량(HTP) 게놈 엔지니어링을 가능하게 하는 컨쥬게이션을 위한 자동화된 프로토콜을 개발하는 것이었다. 이 문제를 해결하기 위해서는 1) HTP 형식의 컨쥬게이션 완료체를 생산하기 위한 컨쥬게이션 효율을 높이고, 2) 배양, 플레이팅 및 콜로니 피킹을 위한 자동화된 프로토콜을 개발해야했다.

[1027] 우리는 배양 접시에서 표준 컨쥬게이션 과정에서 파라미터를 사용하여 HTP 컨쥬게이션을 위한 프로토콜 개발을 시작하였다. 페트리 접시를 사용하는 여러 가지 컨쥬게이션 프로토콜이 이 작업이 시작될 때 개발 중이었으며 추가 개발을 위한 기초로 내부 프로토콜을 선택하였다. 프로토콜이 다른 프로토콜보다 낮은 컨쥬게이션 효율을 가져 왔지만, 이 프로토콜은 수동 처리(예를 들어, 셀 스크래핑)를 필요로하는 어떠한 특별한 단계를 요구하지 않았으며 따라서 자동화하에 가장 적합하였다.

[1028] 본 발명자들은 자동화된 공정을 위한 프로토콜을 동시에 개발하면서 컨쥬게이션 효율을 높이기 위해 통합된 접근법을 선택했다. 우리는 페트리 접시의 컨쥬게이션을 위한 초기 균주 프로토콜을 최적화하기 위해 실험 디자인(Design of Experiment; DOE) 접근 방식을 사용하여 이 공정을 시작했으며, 이는 48-웰 분할 Q-트레이 및 추가 DOE 기반 최적화에서 컨쥬게이션을 수행하기 위한 기초로 사용되었다. 표준 페트리 접시와 비교하여 분할된 Q-트레이는 표면적이 감소하고(페트리 접시에서 8배 감소) 2D 한천 플레이트 형식을 유지하며 자동화 시스템과 잘 연결된다. 48-웰 Q-트레이 형식은 컨쥬게이션의 전체 과정: 공여자 배양, 플레이팅 공여자 및 수용자 세포, 컨쥬게이션 완료체에 대한 항생제 선택, 컨쥬게이션 완료체 콜로니 피킹, 패치 및 배양을 자동화하기 위한 표준 공정 개발의 기초를 제공하였다. 개량된 컨쥬게이션을 위한 실험 인자의 큰 파라미터 공간을 탐색하기 위한 실험 접근법의 디자인을 포함하는 실험 입력이 이하에 제공된다.

[1029] 컨쥬게이션 효율을 개량하기 위해 DOE를 사용하는 초기 실험을 위해, 우리는 최종 선별 디자인 전략(Definitive Screening Design strategy)을 사용하기로 선택하였으며, 이는 일반적으로 다수의 실험 인자를 조합하여 평가하는데 효과적이다. 중요하게, 최종 선별 디자인은 인자 상호작용에도 불구하고 모델을 관리하는 주요 효과를 식별할 수 있으며 정량적 인자의 비선형 효과도 식별할 수 있다. DOE는 최적화 도구이며 컨쥬게이션에 대한 제한된 실험 데이터(즉, 0이 아닌 컨쥬게이션 완료체를 초래한 실험)는 HTP 형식에 적합한 프로토콜을 달성하기 위해 여러 라운드의 최적화가 필요하다고 제안하였다.

[1030] 따라서, 컨쥬게이션 효율을 개량하기 위한 우리의 작업은 3가지 일반적인 단계를 거쳤다. 단계 I에서는 통계적으로 유의한 컨쥬게이션 모델 및 반복을 통한 효율성을 개량을 알려주지 않은 실험 결과를 사용하였다. 단계 II에서, 콜로니를 반복적으로 생성하는 일련의 컨쥬게이션 조건을 식별할 때, 컨쥬게이션 결과를 더욱 향상시킬 새로운 조건을 확인하려고 시도하였다. 단계 III에서는 이러한 조건에 대한 데이터를 사용하여 최적화된 컨쥬게이션을 위한 새로운 실험 조건 세트를 개발 한 다음, 다른 작업자를 통해 생물학적 복제물로 검증하였다.

[1031] DOE- 기반 컨쥬게이션의 최적화를 위해 고려된 인자는 수용자 균주의 배양, 공여자 균주의 배양, 공동 배양 컨쥬게이션 조건 및 컨쥬게이션 완료체의 선택을 포함하는, 도 55에 상세히 설명된 컨쥬게이션 프로토콜의 4개의 주요 부분으로 분류되었다. 이러한 각 요소는 수정/최적화를 위해 고려되었으며 실험 테스트를 위해 우선 순위를 매겼다. JMP 소프트웨어 버전 1 1.2.1을 사용하여 데이터를 분석하였다. 달리 명시되지 않는 한, 결과는 통계적 유의성의 문맥에서 아래에 보고된다.

[1032] 컨쥬게이션을 위한 주요 단계는 다음과 같다:

[1033] 1) 수용자 세포를 중기-대수 상으로 계대배양하는 단계

[1034] 2) 공여자 세포를 중기-대수 상으로 계대배양하는 단계

[1035] 3) 공여자 및 수용자 세포를 결합하는 단계

[1036] 4) 컨쥬게이션 매질 상에 공여자 및 수용자 세포 혼합물을 플레이팅하는 단계

[1037] 5) 세포를 컨쥬게이션시키기 위해 플레이트를 배양하는 단계

[1038] 6) 공여자 세포에 대한 항생제 선택을 적용하는 단계

[1039] 7) 비-통합 수용자 세포에 대한 항생제 선택을 적용하는 단계

[1040] 8) 통합된 수용자 세포(컨쥬게이션 완료체)의 성장을 허용하기 위해 플레이트를 추가로 배양하는 단계

[1041] 우리는 아래 섹션 1에서 컨쥬게이션 효율을 높이고 섹션 2에서 자동화 공정을 개발하기 위한 실험 및 결과를 설명한다.

[1042] **섹션 1: 컨쥬게이션 효율 향상**

[1043] 실험 1.

[1044] *실험 목표*: DOE 접근법을 사용하여 페트리 접시상에서의 컨쥬게이션을 최적화

[1045] *실험 디자인*: 초기 균주 프로토콜을 사용하는 페트리 접시상의 컨쥬게이션은 효율이 낮았으며 Q-트레이 형식으로 전환하면 면적 감소로 인해 효율성이 훨씬 낮아질 것으로 예상하였다. 따라서 우리는 페트리 접시에서 컨쥬게이션 효율을 개량하여 HTP 형식으로 전환된 프로토콜이 성공을 위한 가장 큰 기회를 갖도록 노력하였다. 초기 균주 프로토콜을 최적화하기 위해 DOE 접근 방식을 사용하고 컨쥬게이션에 가장 큰 영향을 미치는 것으로 가정된 다양한 실험 조건을 사용하였다.

[1046] ● 수용자 계대배양 시간: 24-48시간

[1047] ● 날리딕산 농도: 14-50ug/ml

[1048] ● 아프라마이신 농도: 36-100ug/ml

[1049] ● 날리딕산 전달 시간: 2-24시간

[1050] ● 아프라마이신 전달 시간: 16-48시간

[1051] ● 예상되는 공여자 농도: 105-108

[1052] ● 예상되는 수용자 농도: 105-109

[1053] ● 공여자 대 수용자의 비율: 6:1, 1:100

[1054] ● 공여자 스트레스: 항생제 스트레스 없음 또는 공여자 세포는 4ug/ml-8ug/ml 날리딕산으로 1.5시간 동안 처리됨

[1055] *결과:*

[1056] 컨쥬게이션 완료체를 산출한 조건을 표 13에 나타내었다.

[1057] *결과 해석:*

[1058] (1) 조건 3은 가장 많은 양의 컨쥬게이션 완료체를 초래하여 Q-트레이 웰 당 총 6개의 컨쥬게이션 완료체를 산출하였다.

[1059] (2) 실험 데이터의 통계 분석은 컨쥬게이션 효율에 대한 어떠한 단일 파라미터의 유의한 영향을 나타내지 않았다. 그러나, 콜로니를 생성하는 모든 조건은 >24시간으로 나누어진 공여자 및 수용자 항생제 선택 시간을 사용하였다.

[1060] 실험 2.

[1061] *실험 목표:*

[1062] ● 페트리 접시상에서의 컨쥬게이션을 위한 초기 균주 프로토콜이 분할된 Q-트레이상에서의 컨쥬게이션에 사용될 수 있는지를 확인

[1063] ● 공여자 세포에 항생제 스트레스를 가하면 컨쥬게이션 효율이 향상되는지를 테스트

[1064] ● 컨쥬게이션체 완료체 양성 선택을 위한 증가된 아프라마이신 농도가 컨쥬게이션 효율을 개선하는지를 테스트

[1065] *실험 디자인:*

[1066] 1) 두 가지 농도의 수용자 세포를 사용하였다:

[1067] 원래 페트리 접시 프로토콜에 따라 OD=12에서 S. 스피노사 배양액 50ul

[1068] Q-트레이 웰 공간의 공간 감소를 고려하여 OD=12에서 S. 스피노사 배양액 5ul

[1069] 2) 고정된 농도의 공여자 세포를 사용하였다:

[1070] Q-트레이 웰 공간의 감소를 고려하여 원래 페트리 접시 프로토콜보다 10배 적음

[1071] 3) 공여자 스트레스의 영향을 조사하였다:

[1072] 공여자 세포 배양의 절반을 1.5시간 동안 4ug/ml 날리딕산으로 처리하였다.

[1073] 공여자 세포 배양물의 절반은 처리되지 않은 채로 유지되었다

[1074] 4) 컨쥬게이션 완료체의 선택을 위해 2개의 아프라마이신 농도를 사용하였다:

[1075] 최종 농도 62.5ug/ml 한천

[1076] 최종 농도 100ug/ml 한천

[1077] 5) 통계적으로 유의미한 데이터를 제공하기 위해 각 조건을 여러 웰에 걸쳐 반복하였다

[1078] *결과*: 컨쥬게이션 완료체를 산출한 조건을 표 14에 나타내었다.

[1079] *결과 해석*:

[1080] 1) 조건 E는 가장 많은 양의 컨쥬게이션 완료체를 초래하여 Q-트레이 웰 당 총 1.5개의 컨쥬게이션 완료체를 산출하였다.

[1081] 2) 수용체 세포 농도의 감소는 컨쥬게이션 효율을 감소시켰다.

[1082] 3) 아프라마이신 농도 및 공여자 스트레스는 컨쥬게이션 효율에 영향을 미치지 않았다.

[1083] 4) 전반적으로, 이러한 결과는 48-웰 Q-트레이에서 컨쥬게이션이 수행될 수 있음을 보여주었다.

[1084] 실험 3.

[1085] *실험 목표*: 실험 1에서 페트리 접시상에서의 컨쥬게이션에 대한 최적화된 파라미터가 Q-트레이상에서의 컨쥬게이션 효율을 향상시킬 수 있는지 확인.

[1086] *실험 디자인*: 우리는 Q-트레이상에서의 컨쥬게이션을 위한 페트리 접시상에서의 콜로니를 산출하는 각 조건 세트를 테스트하려고 하였다. 그러나, Q-트레이 웰은 페트리 접시보다 면적이 대략 8배 작기 때문에, 단일 Q-트레이 웰상에서의 컨쥬게이션을 위한 두 가지 세포 농도를 테스트하는데 관심이 있었다.

[1087] ● 페트리 접시 실험에 사용된 대략 동일한 총 세포 농도.

[1088] ● 페트리 접시 실험에 사용된 총 세포 농도의 대략 1/8.

[1089] *결과*: 컨쥬게이션 완료체를 산출한 조건을 표 15에 나타내었다.

[1090] *결과 해석*:

[1091] 1) 조건 #8이 가장 많은 양의 컨쥬게이션 완료체를 초래하여 Q-트레이 웰 당 총 3.3개의 컨쥬게이션 완료체가 생성되었다. 참고로, 공여자 농도에 대해 스케일링된 이 조건은 또한 페트리 접시 형식에서 가장 많은 양의 컨쥬게이션 완료체를 초래하였다.

[1092] 2) 전체적으로, 최적화된 페트리 접시 조건은 48-웰 Q-트레이상에서의 향상된 컨쥬게이션을 초래하였다.

[1093] 실험 4.

[1094] *실험 목표*: Q-트레이에서 컨쥬게이션을 위한 DOE를 실행하여 컨쥬게이션에 사용되는 조건을 최적화한다.

[1095] *실험 디자인*: 상기 실험의 결과로부터, 48-웰 Q-트레이상에서 컨쥬게이션이 수행될 수 있음이 명백하였다. 이들 효율이 매우 낮았기 때문에 컨쥬게이션에 가장 큰 영향을 줄 것으로 예상되는 조건을 다양하게 모색하였다.

[1096] o 수용자 계대배양 시간: 24-48시간

[1097] o 날리딕산 농도: 25-100ug/ml

[1098] o 아프라마이신 농도: 50-200ug/ml

[1099] o 예상되는 공여자 농도: 10^5-10^6

[1100] o 예상되는 수용자 농도: 10^5-10^6

[1101] o 공여자 대 수용자의 비율: 3:1, 1:1:, 1:3

[1102] o 공여자 스트레스: 항생제 스트레스 없음 또는 공여자 세포는 4ug/ml 날리딕산 + 4ug/ml 아프라마이신으로 1.5시간 동안 처리됨

[1103] *결과*: 컨쥬게이션 완료체를 산출한 조건을 표 16에 나타내었다.

[1104] *결과 해석*:

[1105] ● 산출된 가장 많은 수의 컨쥬게이션 완료체는 Q-트레이 웰 당 0.7개 컨쥬게이션 완료체였다.

[1106] ● 이러한 낮은 값은 Q-트레이가 완전히 건조되지 않고 배양되었다는 사실에 기인했을 수 있다.

[1107] ● 테스트된 파라미터가 일관되지 않은 실험 조건에 의해 영향을 받는지 이해하려면 추가 DOE를 수행하는 것이 중요할 것이다.

[1108] 실험 5.

[1109] *실험 목표*:

[1110] 1) 국소적인 최적화로서 이 조건 주위의 실험 파라미터를 변화시키면서 실험 2의 조건 #8을 사용하여 DOE를 실행한다.

[1111] 2) 플레이팅을 위해 Tecan 자동 액체 핸들러를 사용하는 것이 수동 플레이팅과 비교하여 컨쥬게이션 효율에 영향을 미치는지 테스트하기 위함(참조: 이 실험까지 자동화 및 수동 액체 처리가 컨쥬게이션을 완료하는데 사용되었지만 자동화된 액체 처리가 수동 플레이팅과 비교하여 더 많거나 더 적은 컨쥬게이션 효율을 초래하였지는 확실하지 않다).

[1112] 3) Q-트레이 건조가 컨쥬게이션에 미치는 영향을 테스트한다.

[1113] *실험 디자인*: 우리는 Q-트레이상에서의 컨쥬게이션을 위한 페트리 접시상에서의 콜로니를 산출하는 각 조건 세트를 테스트하려고 하였다. 그러나, Q-트레이 웰은 페트리 접시보다 면적이 대략 8배 작기 때문에, 단일 Q-트레이 웰상에서의 컨쥬게이션을 위한 두 가지 세포 농도를 테스트하는데 관심이 있었다.

[1114] 1) 페트리 접시 실험에 사용된 대략 동일한 총 세포 농도.

[1115] 2) 페트리 접시 실험에 사용된 총 세포 농도의 대략 1/8.

[1116] *결과*: 컨쥬게이션 완료체를 산출한 조건을 표 17에 나타내었다.

[1117] *결과 해석*:

[1118] ● 조건 12 및 7은 Q-트레이 웰당 가장 많은 양의 컨쥬게이션 완료체를 초래하였으며, 조건 12는 Q-트레이 웰당 총 8.4개의 컨쥬게이션 완료체를 산출하였다.

[1119] ● 아프라마이신 농도(200ug/ml)를 증가시키면 컨쥬게이션 효율이 증가되었다.

[1120] ● 추가로 건조하면 컨쥬게이션 완료체의 총 수가 더 많았지만, 이들 데이터는 통계적으로 유의하지 않았다. 또한, 추가 건조로 인해 플레이트가 갈라지고 너무 얇아져 콜로니 피킹과 같은 다운스트림 공정에 어려움이 있었다.

[1121] ● 자동화된 액체 취급은 수동 플레이팅과 비교하여 컨쥬게이션 효율에 영향을 미치지 않았다.

[1122] ● 실험 계획의 현재 시점에서, 우리는 Q-트레이 웰 당 ≥5개의 콜로니를 생산하는 여러 조건을 확인하였다. 비록 컨쥬게이션을 위한 통계적으로 유의미한 선형 모델을 구축하기 위한 데이터는 없었지만, 이러한 조건은 새로운 인자를 탐구함으로써 추가로 개량될 수 있는 특정 실험 조건을 정확히 찾아내었음을 암시한다.

[1123] 실험 6.

[1124] *실험 목표:*

[1125] 1) 컨쥬게이션 효율을 더욱 향상시키고 컨쥬게이션에 대한 통계 모델을 알기 위한 새로운 실험 인자를 식별

[1126] 2) Q-트레이 매질 구성요소에 대해 DOE를 실행하여 컨쥬게이션을 위한 최적의 매질 조건을 결정

[1127] *실험 디자인*:

[1128] 우리는 다음과 같은 다양한 조건을 선택하였다:

[1129] ● ISP4 분말: 27.8g/L-55.5g/L

[1130] ● 효모 추출물: 0.5g/L-2g/L

[1131] ● 글루코스: 1.5g/L-6g/L

[1132] ● MgCl2: 10mM-40mM

[1133] ● 추가 한천: 0g/L-7.5g/L

[1134] 우리는 이전의 고성능 조건 및 추가적인 새로운 조건을 반영한 실험 조건을 사용하여 이러한 다른 매질 조건의 영향을 테스트하기로 선택하였다:

[1135] ● 실험 5로부터의 조건 #12;

[1136] ● 조건 #8: 더 높은 날리딕산 및 아프라마이신 농도를 사용하여 플레이팅 공정을 용이하게 한다;

[1137] ● 공여자 세포 농도 변동성을 설명하기 위해 #8A로 명명된 조건 #8의 변형된 버전;

[1138] ● 15:1 내지 1:5 사이의 공여자 대 수용자 비율 및 이전 결과를 바탕으로 105-106 사이의 총 예상되는 세포 농도를 변화시켜 네 가지 새로운 조건을 생성하였다.

[1139] *결과*: 컨쥬게이션 완료체를 산출한 조건을 표 18에 나타내었다.

[1140] *결과 해석*:

[1141] ● 글루코스가 높으면 컨쥬게이션 효율이 향상되었다. 다른 모든 매질 성분은 컨쥬게이션 효율에 유의한 영향을 미치지 않는 것으로 판단되었다.

[1142] ● 높은 날리딕산 농도(100ug/ml)는 컨쥬게이션 효율을 높였다.

[1143] ● JMP에 의한 비선형 파티션 모델링은 낮은 아프라마이신 농도(100ug/ml)가 컨쥬게이션 효율을 증가시킬 것으로 예측하였다.

[1144] ● 조건 #12, #8A 및 #8은 순서대로 가장 많은 수의 컨쥬게이션 완료체를 초래하였으며, 조건 #12는 18개의 컨쥬게이션 완료체/Q-트레이 웰을 산출하였다.

[1145] 실험 7.

[1146] *실험 목표*: 최고 성능 조건을 재실행하고 다양한 공여자와 수용자 농도가 이러한 조건의 성능을 향상시킬 수 있는지 테스트

[1147] *실험적 디자인*:

[1148] 1) 기본 조건으로 실험 5의 조건 #7, 실험 5의 조건 #12 및 실험 6의 조건 #8을 선택하였다.

[1149] 2) 실험 전반에 걸쳐 정량화된 변동성을 사용하여 이러한 기본 조건으로부터 공여자 및 수용자 농도를 변경하기로 결정했다. 우리의 프로토콜은 세포 농도에 대한 프록시로 OD를 사용하기 때문에, 실험 사이에 공여자 및 수용자 농도의 고유의 변동성이 존재한다. 우리는 이 변동량(CV)을 계산하고, 공여자 및 수용자 농도를 비례적으로 변경하였다. 우리는 낮은(CV에 의한 비례 감소), 높은(CV에 의한 비례 증가) 및 기준 공여자 및 수용자 농도의 모든 조합으로 컨쥬게이션 실험을 수행하였다.

[1150] *결과*: 컨쥬게이션 완료체를 산출한 조건을 표 19에 나타내었다.

[1151] *결과 해석*:

[1152] 1) 공여자 및 수용자 농도가 낮거나 높으면 컨쥬게이션 효율이 개량되지 않았다.

[1153] 2) 원래 기준 세포 농도를 갖는 조건 #8 및 조건 #12는 Q-트레이 웰 당 ~5개의 컨쥬게이션 완료체를 갖는 가장 많은 수의 컨쥬게이션 완료체를 초래하였다.

[1154] 실험 8

[1155] *실험 목표*:

[1156] 1) 실험 6에 대한 매질 최적화 실험으로부터 조건 #8A 및 #12를 반복하여 새로운 매질 조건이 컨쥬게이션 효율을 개량시켰는지를 검증

[1157] 2) 실험 6(아프라마이신 농도 100ug/ml)에서 JMP 예측을 반영하도록 조정된 조건 #12가 컨쥬게이션 효율을 향상시켰는지를 테스트(이 조건을 #12JA라고 함)

[1158] 3) 조건 7이 표준 매질에서 잘 수행되는 것으로 입증되었기 때문에 새로운 매질 조건을 사용하여 실험 5의 조건 #7을 테스트

[1159] *실험 디자인*: - 우리는 새로운 매질 조건상에서 그리고 비교를 위해 표준 컨쥬게이션 매질 조건상에서 조건 #12, #12JA, #8A 및 #7을 실행하였다.

[1160] *결과*: 컨쥬게이션 완료체를 산출한 조건을 표 20에 나타내었다.

[1161] ● *결과 해석*: 이 실험 디자인에서 가장 많은 수의 컨쥬게이션 완료체는 조건 #12JA에 대한 것으로 40개의 컨쥬게이션 완료체/Q-트레이 웰을 산출하였다.

[1162] ● 새로운 매질 조건이 입증되어 컨쥬게이션 효율을 향상시켰다.

[1163] ● 아프라마이신 농도가 낮을수록 통계적 유의성을 평가하기에 충분한 데이터가 없었음에도 불구하고 더 많은 수의 컨쥬게이션 완료체/Q-트레이 웰이 발생하였다.

[1164] 실험 9

[1165] *실험 목표*: 잘못된 밀도/성장 상태에서 공여자 및 수용자 세포를 사용하여 현재 최적화된 컨쥬게이션 조건에 대한 민감도를 평가한다. 이것은 이들 파라미터의 가변성이 부위마다 발생할 것으로 예상되기 때문에, 컨쥬게이션 프로토콜이 세포의 농도 또는 성장기에 얼마나 민감한지에 대한 지표를 제공할 것이다.

[1166] *실험적 디자인*:

[1167] ● 컨쥬게이션 실험의 기본 조건으로 조건 #8과 #12를 사용하였다.

[1168] ● 공여자 및 수용자 세포에 대해 저밀도, 표준 밀도 및 고밀도의 모든 조합에 대한 컨쥬게이션 실험을 수행하였다.

[1169] ● 낮은 공여자 세포 배양은 OD600=0.2에서 사용되었다.

[1170] ● 표준 공여자 세포 배양은 OD600=0.4에서 사용되었다.

[1171] ● 높은 공여자 세포 배양은 OD600=0.8에서 사용되었다.

[1172] ● OD540=9.6에서 낮은 수용자 세포 배양

[1173] ● OD540=13.0에서 표준 수용자 세포 배양

[1174] ● OD540=14에서 높은 수용자 세포 배양

[1175] ● 새로운 최적화된 매질 조건(실험 6의 실험의 매질 3)을 사용하여 실험을 수행하였다.

[1176] *결과*:

[1177] 1) 공여자 세포를 저밀도로 사용하면 총 컨쥬게이션 완료체의 ~60% 감소가 초래되었다.

[1178] 2) 공여자 세포를 고밀도로 사용하면 총 컨쥬게이션 완료체의 ~50% 감소가 초래되었다.

[1179] 3) 수용자 세포를 저밀도로 사용하면 총 컨쥬게이션 완료체의 ~80% 감소가 초래되었다.

[1180] 4) 수용자 세포를 고밀도로 사용하면 0개의 총 컨쥬게이션체 완료체가 초래되었다.

[1181] ● *결과 해석*: 표준 세포 밀도를 갖는 조건 #12는 40개의 컨쥬게이션 완료체/Q-트레이 웰을 초래하였다.

[1182] ● 잘못된 공여자 및 수용자 세포 농도/성장 단계로 인해 컨쥬게이션 효율이 훨씬 낮아졌으며, 정확한 수용자 배양 조건이 특히 중요하다.

[1183] 실험 10

[1184] *실험 목표*:

[1185] 1) 새로운 작업자의 손에서 최적의 조건을 검증

[1186] 2) 부정확한 밀도/성장 상태에서의 공여자 및 수용자 세포를 사용하여 현재 최적화된 컨쥬게이션 조건에 대한 민감도를 평가

[1187] *실험 디자인*:

[1188] ● 새로운 최적화된 매질 조건을 갖는 조건 #12JA를 사용하였다.

[1189] ● 공여자 및 수용자 세포에 대해 저밀도, 표준 밀도 및 고밀도의 모든 조합에 대한 컨쥬게이션 실험을 수행하였다.

[1190] ● 낮은 공여자 세포 배양은 OD600=0.3에서 사용되었다.

[1191] ● 표준 공여자 세포 배양은 OD600=0.4에서 사용되었다.

[1192] ● 높은 공여자 세포 배양은 OD600=1.0에서 사용되었다.

[1193] ● OD540=4.6에서 낮은 수용자 세포 배양

[1194] ● OD540=8.0에서 표준 수용자 세포 배양

[1195] ● OD540=10.6에서 높은 수용자 세포 배양

[1196] *결과*:

[1197] ● 공여자 세포를 저밀도로 사용하면 총 컨쥬게이션 완료체의 ~100% 증가가 초래되었다.

[1198] ● 공여자 세포를 고밀도로 사용하면 총 컨쥬게이션 완료체의 ~70% 증가가 초래되었다.

[1199] ● 수용자 세포를 저밀도로 사용하면 총 컨쥬게이션 완료체의 ~80% 감소가 초래되었다.

[1200] ● 수용자 세포를 고밀도로 사용하면 총 컨쥬게이션 완료체의 ~80% 감소가 초래되었다.

[1201] *결과 해석*:

[1202] 1) 조건 #12JA가 새로운 작업자에 의해 완료되어 15개의 컨쥬게이션 완료체/Q-트레이 웰이 초래되었다. 이는 이전 결과로부터의 감소였으며, 새로운 작업자가가 처음으로 공정을 시도했기 때문일 수 있다.

[1203] 2) 수용자 세포 농도/성장 단계의 민감도는 실험 9에서 이전 작업자에 의해 결정된 실험 결과와 일치하였다.

[1204] 3) 잘못된 공여자 세포 농도/성장 단계를 사용한 결과는 실험 9의 데이터와 일치하지 않았다. 잘못된 공여자 세포 농도를 사용하면 표준 프로토콜로부터 향상된 컨쥬게이션 효율을 초래하였지만, 이들 데이터는 이전 실험 데이터와 관련하여 결정적인 유의성을 가지지 않았다.

[1205] 4) 수용자 세포의 현미경 검사는 세포 상태를 확인하는데 유용하였다. 후기 대수 세포는 액체 배양물에서 더 단편적으로 나타난다.

표 13

[1206] 표 13. 저 처리량 컨쥬게이션의 최적화에 기반한 실험 디자인의 결과

	수용자 세척 조건	컨쥬게이션 온도 (C)	수용자 계대배양 시간 (hr)	공여자 스트레스 NA (ug/ml)	비율 (D:R)	총 세포	NA 전달 시간	Apra 전달 시간	NA 농도 (ug/ml 한천)	Apra 농도 (ug/ml 한천)	#컨쥬게이션 완료체/페트리 접시
1	세척 없음	30	24	8	0.10	2.74E+06	16	46	28	36	4
2	세척 없음	37	36	4	1.00	3.34E+07	16	41.5	28	54	3
3	세척	30	48	0	42.2	5.00E+06	21	48	28	72	6

4	세척 없음	37	24	0	1.53	1.00E+07	24	48	14	100	2
5	세척 없음	30	24	4	2.45	8.40E+06	16	42	32	68	3
SOP	세척 없음	30	48	0	NA	NA	20	20	14	35	0

표 14

[1207]

표 14: 초기 Q-트레이 컨쥬게이션 실험 2의 결과

조건	수용자 (cfu/ml)	공여자 스트레스 날리딕산 (ug/ml)	아프라마이신 농도 (ug/ml 한천)	# 컨쥬게이션 완료체/Q-트레이 웰
A	1.65×10^5	0	62.5	0.5
B	1.65×10^5	4	62.5	0.5
C	1.65×10^4	0	62.5	0
D	1.65×10^4	4	62.5	0
E	1.65×10^5	0	100	1.5
F	1.65×10^5	4	100	0.5
G	1.65×10^4	0	100	0
H	1.65×10^4	4	100	0

표 15

[1208]

표 15. LTP 컨쥬게이션 조건을 HTP 형식으로 전달한 결과

	수용자 세척 조건	컨쥬게이션 온도	수용자 계대배양 시간 (hrs)	공여자 스트레스 NA (ug/ml)	비율 (D:R)	총 세포 (cfu/ml)	LTP 조건으로부터의 상대적 세포량	NA 전달 시간 (hrs)	Apra 전달 시간 (hrs)	NA 농도 (ug/ml 한천)	Apra 농도 (ug/ml 한천)	#컨쥬게이션 완료체/Q-트레이 웰
1	세척 없음	30	24	8	0.04	2.1E+06	1x	16	46	28	36	0.2
2	세척 없음	37	36	4	1.15	7.3E+06	1x	16	41.5	28	54	0.0
3	세척	30	48	0	1.43	2.1E+06	1x	21	48	28	72	1.5
4	세척 없음	37	24	0	0.06	3.6E+07	1x	24	48	14	100	0.0
5	세척 없음	30	24	4	0.29	2.7E+07	1x	16	42	32	68	0.5
SOP	세척 없음	30	48	0	16.67	2.4E+06	1x	20	20	14	35	0.2
6	세척 없음	30	24	8	0.04	3.5E+05	1/8x	16	46	28	36	0.0
7	세척 없음	37	36	4	1.15	1.2E+06	1/8x	16	41.5	28	54	0.0
8	세척	30	48	0	1.43	3.5E+05	1/8x	21	48	28	72	3.3
9	세척 없음	37	24	0	0.06	6.0E+06	1/8x	24	48	14	100	0.0
10	세척 없음	30	24	4	0.29	4.4E+06	1/8x	16	42	32	68	0.7
SOP (희석)	세척 없음	30	48	0	16.67	4.0E+05	1/8x	20	20	14	35	0.0

표 16

[1209]

표 16: 컨쥬게이션 실험 4로부터의 최고의 조건

수용자 세척 조건	컨쥬게이션 온도 (C)	수용자 계대배양 시간 (hrs)	NA 전달 시간 (hrs)	Apra 전달 시간 (hrs)	NA 농도 (ug/ml 한천)	Apra 농도 (ug/ml 한천)	비율 (D:R)	총 세포 농도 (cfu/ml)	#컨쥬게이션 완료체/ 웰
세척	30	48	21	45	62.5	50	1.1	1.5E+0.6	0.7

표 17

[1210]

표 17: 컨쥬게이션 실험 5로부터의 최고의 조건

	수용자 세척 조건	컨쥬게이션 온도 (C)	수용자 계대배양 시간 (hrs)	NA 전달 시간 (hrs)	Apra 전달 시간 (hrs)	NA 농도 (ug/ml 한천)	Apra 농도 (ug/ml 한천)	비율 (D:R)	총 세포 농도 (cfu/ml)	#컨쥬게이션 완료체/웰
12	세척	30	48	20	42	100	200	14.3	3.0E+0.6	8.4
7	세척	30	24	20	42	62.5	200	0.1	1.5E+0.6	5.4

표 18

[1211]

표 18: 컨쥬게이션 실험 6으로부터의 최고의 조건

	매질	수용자 세척 조건	컨쥬게이션 온도 (C)	수용자 계대배양 시간 (hrs)	NA 전달 시간 (hrs)	Apra 전달 시간 (hrs)	NA 농도 (ug/ml 한천)	Apra 농도 (ug/ml 한천)	비율 (D:R)	총 세포 농도 (cfu/ml)	#컨쥬게이션 완료체/웰
12	3	세척	30	48	18	42	100	200	8.1	1.76E+06	18.0
8A	3	세척	30	48	18	42	50	100	0.8	2.66E+05	7.5
8	8	세척	30	48	18	42	50	100	5.5	9.47E+05	6.5

표 19

[1212]

표 19: 컨쥬게이션 실험 7로부터의 최고의 조건

	매질	수용자 세척 조건	컨쥬게이션 온도 (° C)	수용자 계대배양 시간 (hrs)	NA 전달 시간 (hrs)	Apra 전달 시간 (hrs)	NA 농도 (ug/ml 한천)	Apra 농도 (ug/ml 한천)	비율 (D:R)	총 세포 농도 (cfu/ml)	#컨쥬게이션 완료체/웰
12	표준	세척	30	48	20	42	100	200	0.006	3.87E+08	5.2
8	표준	세척	30	48	20	42	50	100	0.019	6.54E+07	5.5

표 20

[1213]

표 20: 컨쥬게이션 실험 8로부터의 최고의 조건

	매질	수용자 세척 조건	컨쥬게이션 온도 (C)	수용자 계대배양 시간 (hrs)	NA 전달 시간 (hrs)	Apra 전달 시간 (hrs)	NA 농도 (ug/ml 한천)	Apra 농도 (ug/ml 한천)	비율 (D:R)	총 세포 농도 (cfu/ml)	#컨쥬게이션 완료체/웰
12JA	3	세척	30	48	20	42	100	100	1.23	4.19E+06	39.9
12	3	세척	30	48	20	42	100	200	1.23	4.19E+06	19.9
8A	3	세척	30	48	20	42	50	100	0.19	1.11E+06	0.8
7	3	세척	30	24	20	42	62.5	200	0.08	2.02E+06	1.1

[1214] ***섹션 2:*** *자동화 개발*

[1215] 실험 1 1. 고 처리량 공여자 배양(자동화 요소)

[1216] *실험 목표*: 컨쥬게이션을 위한 공여자 세포를 HTP 형식으로 성장시키기

[1217] *실험 디자인*:

[1218] 1) 우리는 96 웰 딥 웰 스퀘어 플레이트(E&K EK-2440-ST)에서 대장균 공여자 배양물의 성장을 테스트했다. 가장 낮은 OD 판독값을 갖는 배양물이 1:100 접종에 상응하도록 하룻밤 동안의 배양물의 OD600에 기초하여 접종 부피를 정규화함으로써 배양물을 접종하였다.

[1219] 2) 성장을 위해 3가지 부피의 LB 매질을 테스트하였다: 250ul, 500ul, 750ul.

[1220] 3) 컨쥬게이션에 대한 HTP 성장의 영향을 평가하기 위해, 이 HTP 포맷으로 성장한 대장균 S17 + SS015를 사용하여 컨쥬게이션을 수행하였다.

[1221] *결과*: 세포 성장 및 컨쥬게이션 데이터가 도 56a-b에 도시되어 있다.

[1222] *결과의 해석*: 테스트된 모드 부피에서 배양물이 강력하게 성장하였다. 또한, 500ul 부피로 배양된 배양물은 차이가 통계적으로 유의하지는 않았지만 가장 많은 수의 컨쥬게이션 완료체를 산출했다. 500ul 부피는 액체 처리가 용이하고 OD 체크를 위한 충분한 부피를 제공하므로 고 처리량 공여자 성장을 위해 선택되었다.

[1223] 실험 12. 컨쥬게이션을 위한 HTP 형식의 세포 및 항생제 플레이팅(자동화 요소)

[1224] *실험 목표*:

[1225] 1) 세포와 항생제를 HTP 형식으로 플레이팅

[1226] 2) 공여자 및 수용자 세포 위에 항생제를 층화하기 위한 다수의 플레이팅 단계가 컨쥬게이션에 필요하기 때문에, 컨쥬게이션 프로토콜 전체에 걸쳐 일관된 플레이팅을 달성

[1227] *실험 디자인: 우리는 48-웰 분할된 Q-트레이에서 세포 및 항생제를 플레이팅하기 위한 3가지 잠재적인 공정을 확인하였다:*

[1228] ● 스팟 플레이팅 - 단일 지점에서 액체 부피를 플레이팅하고 플레이팅된 영역에서 건조되도록 두었다

[1229] ● 비드를 이용한 플레이팅 - 단일 지점에서 액체 부피를 플레이팅한 다음 비드를 사용하여 웰의 전체 영역에 액체를 분산시킨다

[1230] ● Q-트레이 웰에 플러딩(flooding) - 요동 운동으로 액체가 웰 전체 영역에 분산되도록 충분한 양의 액체를 플레이팅

[1231] *결과:*

[1232] ● 스팟 플레이팅으로 인해 세포 플레이팅이 일관되지 않았으며 또한 플레이팅된 셀의 소수성이 이 방법을 사용한 컨쥬게이션 완료체 선택을 위해 항생제를 플레이팅하는 것을 어렵게 만들었다. 스팟팅된 항생제 부피는 플레이팅된 세포의 전체 영역으로 분산되지 않았으며 표면 장력을 수동으로 파괴하지 않으면 퍼질 수 없었다.

[1233] ● 비드를 이용한 플레이팅은 일관된 플레이팅을 초래하였지만 오염의 결과를 나타내었다. 각 웰에 비드가 포함된 Q-트레이를 흔들면 플레이팅액이 뿌려져 가끔 비드가 웰 사이를 가로지른다. 또한, 비드를 이용한 플레이팅은 자동화 시스템과 인터페이스하기 위해 상당한 커스텀화를 필요로 한다.

[1234] ● Q-트레이 웰을 플러딩하면 컨쥬게이션 공정 전반에 걸쳐 일관된 플레이팅이 가능해져서 세포와 항생제가 웰 영역에 균일하게 플레이팅되었다. 그러나, 배양물 및 항생제가 플레이팅된 후 플레이트를 완전히 건조시키기 위해 긴 배양 기간이 필요하였다.

[1235] ● 또한, 액체를 분산시키기 위해 플레이트를 앞뒤로 흔들 수 있도록 자동화된 솔루션을 개발할 필요가 있다.

[1236] *결과 해석:*

[1237] ● 우리의 플레이팅 시도를 기초로, 우리는 Q-트레이 웰에 플러딩하도록 충분한 액체를 플레이팅하는 것이 자동화된 컨쥬게이션에 사용하기에 가장 유망한 공정라는 것을 확인하였다. 이 공정은 일관되고 균일한 플레이팅을 초래하며 자동화된 액체 핸들러와 쉽게 연결될 수 있다.

[1238] ● 액체를 분산시키기 위해 Q-트레이를 흔드는 수동 단계를 극복하기 위해, 우리는 VWR로부터의 3D Rotator Wave를 구입하고 Q-트레이 치수를 수용할 수 있도록 맞춤형 부품으로 플랫폼을 수정하였다. 3D Rotator Wave는 궤도 운동 및 z 평면에서 움직일 수 있기 때문에, 판 흔들림을 수동으로 수행했을 때와 동일한 움직임을 제공하였다.

[1239] 실험 13. 컨쥬게이션 완료체 피킹(자동화 요소)

[1240] *실험 목표:*

[1241] ● 컨쥬게이션 플레이트에서 컨쥬게이션 완료체를 탐지하기 위한 표준 공정 개발

[1242] ● 컨쥬게이션 Q-트레이에서 콜로니를 선택적 한천 옴니 트레이에 패치/스탬프하기

[1243] *실험 디자인*:

[1244] 1) Qpix 420 및 해당 소프트웨어를 사용하여 컨쥬게이션 플레이트에서 S. 스피노사 컨쥬게이션 완료체를 확인하였다.

[1245] 2) 컨쥬게이션 완료체를 탐지하기 위해 다음과 같은 이미징 파라미터를 사용하여 실험하였다:

[1246] ● 임계값

[1247] ● 노출

[1248] ● 획득

[1249] ● 이미지 반전

[1250] ● 배경 제거

[1251] 3) 탐지 피킹 가능 컨쥬게이션 완료체를 포함시키기 위해 다음 기능 선택 파라미터를 사용하여 실험하였다:

[1252] 1) 소형

[1253] 2) 축 비율

[1254] 3) 최소 직경

[1255] 4) 최대 직경

[1256] 5) 최소 근접

[1257] 4) 두 가지 유형의 핀을 가진 피킹 헤드를 사용하여 테스트하였다:

[1258] 1. 효모 피킹 핀(X4377)

[1259] 2. 대장균 피킹 헤드(X4370)

[1260] 5) 크고 견고한 패치를 만들기 위하여 고체 한천 옴니 트레이에 접종하기 위해 두 가지 다른 기능을 사용해서 시도하였다.

[1261] ● 싱글 딥

[1262] ● 젓기

[1263] *결과*:

[1264] ● 우리는 Q-트레이 이미징 과정에서 이미지를 뒤집는 것이 S. 스피노사 컨쥬게이션 완료체를 탐지하는데 매우 유용하다는 것을 확인하였다.

[1265] ● 다중 컨쥬게이션 플레이트를 이미징한 후, S. 스피노사 컨쥬게이션 완료체를 정확하게 확인하기 위해 단일 임계값 및 노출값을 사용할 수 없음을 발견하였다(도 57 참조). 각각의 플레이트(예를 들어, 남은 죽은 공여자 및 수용자 세포)상의 배경의 가변성 때문에, 각각의 플레이트에 대한 임계값 및 노출값을 조정해야했다. 이러한 파라미터에 사용할 값 범위를 식별하였다.

[1266] ● 우리는 대장균 핀을 사용하여 컨쥬게이션 완료체를 옮기는 것이 효과가 없다는 것을 확인하였다 - 이것은 이러한 핀이 후속 접종을 위해 S. 스피노사 세포를 충분히 선택할 수 없기 때문일 수 있다.

[1267] ● 효모 핀을 이용한 콜로니 피킹은 플레이트 접종에 잘 작동하였다. 그러나 피킹 후 피킹 헤드가 옴니 트레이에서 완전히 분리되지 않고 옴니 트레이와 함께 옴니 트레이를 운반하였다. 대상 옴니 트레이를 테이핑한 후 이는 더 이상 문제되지 않았다.

[1268] ● 딥 기능은 접종에 대해 잘 작동하였다. 교반 기능은 또한 유망한 접종 방법으로 보였다. 불행하게도, 교반 기능을 이용하여 접종한 옴니 트레이가 곰팡이 오염을 경험했기 때문에 이러한 결과는 결론적이지는 않았다.

[1269] *결과 해석*: 효모 핀이 장착된 Qpix 픽킹 헤드와 딥 접종 기능을 사용하여 플레이트 변동성을 기반으로 조정할 수 있는 S. 스피노사 컨쥬게이션 완료체 피킹하기 위한 일반적인 파라미터 세트를 설정하였다. 이 프로토콜은 눈에 띄는 대장균 오염이 없는 강력한 S. 스피노사 성장을 초래하였다.

[1270] 실험 14

[1271] *실험 목표*: 배양 및 저장을 위해 옴니 트레이에서 96웰 딥 웰 플레이트로 컨쥬게이션 완료체 패치를 피킹.

[1272] *실험 디자인*:

[1273] 1. 표준 피킹 조건을 사용하여 96웰 딥 웰 스퀘어 플레이트(E&K EK-2440-ST)에서 S. 스피노사 패치의 피킹을 테스트하였다.

[1274] 2. 성장에 대해 3가지 부피의 DAS 매질 2를 테스트했다: 300ul, 400ul, 500ul.

[1275] *결과*: 도 58에서 볼 수 있듯이, 400ul의 배지를 사용하여 접종한 웰만 강력한 성장을 가져왔다.

[1276] *결과 해석*:

[1277] 1. 300ul 및 500ul 접종 부피가 왜 컨쥬게이션 완료체 배양물의 성장을 초래하지 않았는지는 확실하지 않는다. 우리는 이것이 매질 볼륨 자체가 아닌 접종 공정과 관련이 있다고 의심하였다.

[1278] 2. 이 프로토콜은 피킹된 컨쥬게이션 완료체 패치의 강력한 성장을 보장하기 위해 몇몇 추가 검증 및 최적화가 필요하다.

[1279] 요약: 도 59는 DOE 기반 최적화 과정을 통해 완료된 컨쥬게이션 실험의 결과를 요약한다. 최적화 과정에서, 통계 분석에 따르면 컨쥬게이션의 가장 중요한 조건은 약물의 선택 농도와 매질의 글루코스 농도였다(도 60 참조). 실험적 분석은 수용자 배양물 성장 단계가 또한 컨쥬게이션을 위한 중요한 조건임을 추가로 제안하였다. 광학 밀도 판독은 수용자 배양물이 컨쥬게이션을 위한 준비가 되었는지를 결정하는 비교적 양호한 지표인 것으로 보이지만, 보다 최근의 실험은 세포 형태가 또한 세포 상태를 확인하는데 유용하다는 것을 시사한다. 따라서 수용자 배양물은 광학 밀도 외에 정확한 세포 형태를 검사해야하다. 최적화된 컨쥬게이션 프로토콜은 공여자 농도 및 성장 단계에 대해 큰 민감성을 나타내지 않았으므로, 확립된 프로토콜은 공여자 세포 성장의 편차에 민감하지 않을 수 있다. 이것은 HTP 형식의 다양한 균주를 사용하여 작업할 때 편리할 것이다.

[1280] 본 발명은 고 처리량(HTP) 게놈 공학을 가능하게 하는 컨쥬게이션을 위한 자동화된 프로토콜을 수반하여 사카로폴리스포라(예를 들어, S. 스피노사)에서의 컨쥬게이션 효율을 개량시키는 방법을 가능하게 한다. 우리는 처리량이 많은 컨쥬게이션을 위한 프로토콜을 개발하여 전체 평균 24개의 컨쥬게이션 완료체/Q-트레이 웰을 만들었다(두 개별 작업자에 의해 중복하여 실행). 최대 수의 컨쥬게이션 완료체를 산출하는 컨쥬게이션 조건은 다음을 포함한다: 수용자 세포 세척, 30℃에서 컨쥬게이션, 대략 48시간 동안 수용자 균주를 계대배양, 컨쥬게이션 20시간 후 100ug/ml 날리딕산으로 선택, 컨쥬게이션 42시간 후 100㎍/ml 아프라마이신으로 선택, 6g/L 글루코스을 가진 ISP4 변형 매질을 이용, ~$7x10^6$ 세포의 총 수를 가진 ~1:0.8의 공여자 대 수용자 비율.

표 21

[1281] **서열번호 식별자를 갖는 본 발명의 서열**

프로모터 래더			
소스	**명칭**	**설명**	**서열번호**
사카로폴리스포라 스피노사	P7160	샤페로닌 GroEL과 관련된 프로모터 서열	1
사카로폴리스포라 스피노사	P7253	신장 인자 Tu와 관련된 프로모터 서열	2

사카로폴리스포라 스피노사	P6681	F-형 ATPase 서브유닛 델타와 관련된 프로모터 서열	3
사카로폴리스포라 스피노사	P6316	PspA/IM30 패밀리 단백질과 관련된 프로모터 서열	4
사카로폴리스포라 스피노사	P6806	2-옥소글루타레이트 디카복실라아제와 관련된 프로모터 서열	5
사카로폴리스포라 스피노사	P3159	추정 에노일-CoA 하이드라타아제 echA8과 관련된 프로모터 서열	6
사카로폴리스포라 스피노사	P0757	추정 L-라이신-입실론 아미노트랜스퍼라제와 관련된 프로모터 서열	7
사카로폴리스포라 스피노사	P5011	가설 단백질과 관련된 프로모터 서열	8
사카로폴리스포라 스피노사	P1409	NAD-특이적 글루타메이트 탈수소효소와 관련된 프로모터 서열	9
사카로폴리스포라 스피노사	P4735	류실 아미노펩티다아제(아미노펩티다아제 T)와 관련된 프로모터 서열	10
사카로폴리스포라 스피노사	P2900	시토크롬 P450-terp와 관련된 프로모터 서열	11
사카로폴리스포라 스피노사	P0801	주변세포질 뮤레인 펩타이드-결합 단백질 전구체와 관련된 프로모터 서열	12
합성	P21	Siegl 등에 기술된 합성 프로모터	13
합성	PA9	Siegl 등에 기술된 합성 프로모터	14
합성	PA3	Siegl 등에 기술된 합성 프로모터	15
합성	PB4	Siegl 등에 기술된 합성 프로모터	16
합성	PB12	Siegl 등에 기술된 합성 프로모터	17
합성	PB1	Siegl 등에 기술된 합성 프로모터	18
합성	PC1	Siegl 등에 기술된 합성 프로모터	19
합성	P72	Siegl 등에 기술된 합성 프로모터	20
합성	P-C4-1	Seghezzi 등에 기술된 합성 프로모터	21
합성	P-A5-19	Seghezzi 등에 기술된 합성 프로모터	22
합성	P-C4-14	Seghezzi 등에 기술된 합성 프로모터	23
합성	P-D1-7	Seghezzi 등에 기술된 합성 프로모터	24
사카로폴리스포라 스피노사	P1	서열 분비 단백질과 관련된 프로모터	25
사카로폴리스포라 스피노사	P2	서열 가설 단백질과 관련된 프로모터	26
사카로폴리스포라 스피노사	P3	RNA 폴리머라제 시그마 인자 SigD와 관련된 프로모터 서열	27
사카로폴리스포라 스피노사	P3v2	RNA 폴리머라제 시그마 인자 SigD, 버전 2와 관련된 프로모터 서열	28
사카로폴리스포라 스피노사	P4	항원 Ag88과 관련된 프로모터 서열	29
사카로폴리스포라 스피노사	P4v2	항원 Ag88, 버전 2와 관련된 프로모터 서열	30
사카로폴리스포라 스피노사	P5	DNA-지시된 RNA 폴리머라제 서브유닛 베타와 관련된 프로모터 서열	31
사카로폴리스포라 스피노사	P5v2	DNA-지시된 RNA 폴리머라제 서브유닛 베타와 관련된 프로모터 서열	32
사카로폴리스포라 스피노사	P6	분자 샤페론 GroEL과 관련된 프로모터 서열	33
사카로폴리스포라 스피노사	P7	UDP-4-아미노-4-데옥시-L-아라비노스--옥소글루타레이트 아미노트랜스퍼라제와 관련된 프로모터 서열	34
사카로폴리스포라 스피노사	P8	프롤린 라세미화 효소와 관련된 프로모터 서열	35
사카로폴리스포라 스피노사	P9	페닐옥사졸린 합성효소 MbtB와 관련된 프로모터 서열	36
사카로폴리스포라 스피노사	PspnA	폴리케타이드 합성효소 로딩 및 확장 모듈 1과 관련된 프로모터 서열	37

사카로폴리스포라 스피노사	PspnA_v2	폴리케타이드 합성효소 로딩 및 확장 모듈 1, 버전 2와 관련된 프로모터 서열	38
사카로폴리스포라 스피노사	PspnF	메틸트랜스퍼라제-유사 단백질과 관련된 프로모터 서열	39
사카로폴리스포라 스피노사	PspnG	추정 NDP-람노실트랜스퍼라제와 관련된 프로모터 서열	40
사카로폴리스포라 스피노사	PspnQ	추정 NDP-헥소오스-3과 관련된 프로모터 서열	41
사카로폴리스포라 스피노사	PspnQ_v2	추정 NDP-헥소오스-3, 버전 2와 관련된 프로모터 서열	42
합성	P21_mutant	합성프로모터	43
사카로폴리스포라 스피노사	P1_core	분비 단백질과 관련된 프로모터 서열	44
사카로폴리스포라 스피노사	P1(-33)	분비 단백질과 관련된 프로모터 서열	45
사카로폴리스포라 스피노사	P1+ribswtch	분비 단백질과 관련된 프로모터 서열	46
합성	P21-P1	합성프로모터	47
합성	P1-P21	합성프로모터	48
사카로폴리스포라 스피노사	P1765	글루타민 합성효소 1과 관련된 프로모터 서열	49
사카로폴리스포라 스피노사	P3747	가설 단백질과 관련된 프로모터 서열	50
사카로폴리스포라 스피노사	P5078	가설 단백질과 관련된 프로모터 서열	51
사카로폴리스포라 스피노사	P7419	혐기성 벤조에이트 이화작용 전사 조절제와 관련된 프로모터 서열	52
사카로폴리스포라 스피노사	P7156	RNA 폴리머라제 시그마 인자 SigD와 관련된 프로모터 서열	53
사카로폴리스포라 스피노사	P7256	30S 리보솜 단백질 S12와 관련된 프로모터 서열	54
사카로폴리스포라 스피노사	P1941	응답 조절제 단백질 vraR과 관련된 프로모터 서열	55
사카로폴리스포라 스피노사	P3405 (P8)	프롤린 라세미화 효소와 관련된 프로모터 서열	56
사카로폴리스포라 스피노사	P3407	ABC 트랜스포터 아르기닌-결합 단백질 1 전구체와 관련된 프로모터 서열	57
사카로폴리스포라 스피노사	P2428	아세틸-CoA 합성효소와 관련된 프로모터 서열	58
사카로폴리스포라 스피노사	P0927	4-하이드록시페닐피루베이트 2산소화효소와 관련된 프로모터 서열	59
사카로폴리스포라 스피노사	P0889	선형 그라미시딘 탈수소효소 LgrE와 관련된 프로모터 서열	60
사카로폴리스포라 스피노사	P0186	L,D-트랜스펩티다아제 촉매영역과 관련된 프로모터 서열	61
사카로폴리스포라 스피노사	P3702_v2	가설 단백질과 관련된 프로모터 서열	62
사카로폴리스포라 스피노사	P7156_v2	RNA 폴리머라제 시그마 인자 SigD와 관련된 프로모터 서열	63
사카로폴리스포라 스피노사	P7256_v2	30S 리보솜 단백질 S12와 관련된 프로모터 서열	64
사카로폴리스포라 스피노사	P1765_v2	글루타민 합성효소 1과 관련된 프로모터 서열	65
사카로폴리스포라 스피노사	P7539_v2	항원 Ag88과 관련된 프로모터 서열	66
사카로폴리스포라 스피노사	P7276_v2	DNA-지시된 RNA 폴리머라제 서브유닛 베타와 관련된 프로모터 서열	67
사카로폴리스포라 스피노사	P0941_v2	가설 단백질과 관련된 프로모터 서열	68

사카로폴리스포라 스피노사	P0889_v2	선형 그라미시딘 탈수소효소 LgrE와 관련된 프로모터 서열	69
합성	Pmut-1	합성프로모터	172
합성	B2	합성프로모터	173
합성	D1	합성프로모터	174
합성	D2	합성프로모터	175
추정 터미네이터			
소스	**명칭**	**설명**	**서열번호**
사카로폴리스포라 스피노사	T1	신장 인자 tu와 관련된 터미네이터 서열	70
사카로폴리스포라 스피노사	T2	류실 아미노펩티다아제와 관련된 터미네이터 서열	71
사카로폴리스포라 스피노사	T3	시토크롬 P450 수산화효소와 관련된 터미네이터 서열	72
사카로폴리스포라 스피노사	T4	F0F1 ATP 합성효소 서브유닛 베타와 관련된 터미네이터 서열	73
사카로폴리스포라 에리트라에아	T5	FAD-결합된 산화환원효소와 관련된 터미네이터 서열	74
사카로폴리스포라 에리트라에아	T6	포스포리보실트랜스퍼라제와 관련된 터미네이터 서열	75
사카로폴리스포라 에리트라에아	T7	ATP-결합 단백질과 관련된 터미네이터 서열	76
사카로폴리스포라 에리트라에아	T8	50s 리보솜 단백질 L32와 관련된 터미네이터 서열	77
사카로폴리스포라 에리트라에아	T9	tRNA-Arg와 관련된 터미네이터 서열	78
사카로폴리스포라 에리트라에아	T11	lsr2와 관련된 터미네이터 서열	79
사카로폴리스포라 에리트라에아	T12	AraC와 관련된 터미네이터 서열	80
리포터 유전자			
소스	**명칭**	**설명**	**서열번호**
인공 서열	대셔GFP 리포터 유전자	코돈 최적화된 리포터 유전자 대셔GFP	81
인공 서열	파프리카RFP 리포터 유전자	코돈 최적화된 리포터 유전자 파프리카RFP	82
인공 서열	gusA 리포터 유전자	코돈 최적화된 리포터 유전자 gusA	83
인공 서열	대셔GFP	대셔GFP 리포터 단백질	143
인공 서열	파프리카RFP	파프리카RFP 리포터 단백질	144
인공 서열	gusA	gusA 리포터 단백질	145
인공 서열	GFP/RFP용 터미네이터 서열	GFP 또는 RFP 리포터 유전자를 발현시키는데 사용되는 터미네이터 서열	150
인테그라제 관련 서열			
소스	**명칭**	**설명**	**서열번호**
S. 엔도피티카	pCM32 인테그라제 유전자	pCM32 + attP 서열의 인테그라제	84
S. 엔도피티카	pCM32 인테그라제 단백질	플라스미드 pCM32의 인테그라제의 단백질 서열	85
S. 엔도피티카	pCM32의 attP	pCM32의 attP 부위 서열	167
S. 에리트라에아	pSE101 인테그라제 유전자	pSE101 + attP 서열의 인테그라제	86
S. 에리트라에아	pSE101 인테그라제 단백질	플라스미드 pSE101의 인테그라제의 단백질 서열	87
S. 에리트라에아	attP in pSE101	pSE101의 attP 부위 서열	168
S. 에리트라에아	pSE211 인테그라제 유전자	pSE211 + attP 서열의 인테그라제	88

S. 에리트라에아	pSE211 인테그라제 단백질	플라스미드 pSE211의 인테그라제의 단백질 서열	89
S. 에리트라에아	pSE211의 attP	pSE211의 attP 부위 서열	169
S. 스피노사	pSE101 상동체 인테그라제 유전자	*S.스피노사*attP서열의 pSE101 인테그라제 상동체 유전자	90
S. 스피노사	pSE101 상동체 인테그라제 단백질	*S.스피노사*의 pSE101 인테그라제 상동체 유전자의 단백질 서열	91
S. 스피노사	attP pSE101 상동체	pSE101 상동체 제작물의 attP 부위 서열	170
S. 스피노사	pSE211 상동체 인테그라제 유전자	*S.스피노사*attP서열의 pSE211 인테그라제 상동체 유전자	92
S. 스피노사	pSE211 상동체 인테그라제 단백질	*S.스피노사*의 pSE211 인테그라제 상동체 유전자의 단백질 서열	93
S. 스피노사	attP pSE211 상동체	pSE211 상동체 제작물의 attP 부위 서열	171
복제 기점 및 요소			
소스	**명칭**	**설명**	**서열번호**
S. 에리트라에아	S. 에리트라에아의 복제 기점	*S.에리트라에아*로부터의 추정 염색체 복제 기점	94
S. 에리트라에아	pSE101 AICE 요소	복제 플라스미드 pSE101의 액티노마이세트 통합 및 컨쥬게이티브 요소 (AICEs)	95
S. 에리트라에아	pSE211 AICE 요소	복제 플라스미드 pSE211의 액티노마이세트 통합 및 컨쥬게이티브 요소 (AICEs)	96
리보솜 결합 부위(RBS) 서열			
소스	**명칭**	**설명**	**서열번호**
S. 스피노사	RBS1	PermE*와 관련된 RBS	97
S. 스피노사	RBS2	spnA(폴리케타이드 합성효소 로딩 & 확장 모듈1)와 관련된 RBS	98
S. 스피노사	RBS3	spnC(폴리케타이드 합성효소 확장 모듈 3-4)와 관련된 RBS	99
S. 스피노사	RBS4	spnO(추정 NDP-헥소오스-2,3-디하이드라타아제)와 관련된 RBS	100
S. 스피노사	RBS5	gdh(gdh (dTDP-글루코스 4,6-디하이드라타아제))와 관련된 RBS	101
S. 스피노사	RBS6	링커_A(알데하이드 탈수소효소, AldA)와 관련된 RBS	102
S. 스피노사	RBS7	링커_B(아세토락테이트 합성효소)와 관련된 RBS	103
S. 스피노사	RBS8	링커_C와 관련된 RBS	104
S. 스피노사	RBS9	링커_D와 관련된 RBS	105
S. 스피노사	RBS10	gtt(글루코스-1-포스페이트 티미딜트랜스퍼라제 1)와 관련된 RBS	106
S. 스피노사	RBS11	TDH(글리세르알데하이드-3-포스페이트 탈수소효소)와 관련된 RBS	107
S. 스피노사	RBS12	BioBrick_1과 관련된 RBS	108
S. 스피노사	RBS13	BioBrick_2와 관련된 RBS	109
S. 스피노사	RBS14	GroES(분자 샤페론 GroES)와 관련된 RBS	110
S. 스피노사	RBS15	GroEL(분자 샤페론 GroEL)와 관련된 RBS	111
S. 스피노사	RBS16	IF-1(번역 개시 인자 IF-1)와 관련된 RBS	112
S. 스피노사	RBS17	XNR_1700(주변세포질 뮤레인 펩타이드-결합 단백질 전구체)과 관련된 RBS	113
S. 스피노사	RBS18	S20(30s 리보솜 단백질 S20)과 관련된 RBS	114
S. 스피노사	RBS19	S12(리보솜 단백질 S12)와 관련된 RBS	115
S. 스피노사	RBS20	S12(리보솜 단백질 S12)와 관련된 RBS	116
S. 스피노사	RBS21	DnaK(Hsp 70)와 관련된 RBS	117
S. 스피노사	RBS22	신장 인자 Tu와 관련된 RBS	118
S. 스피노사	RBS23	F0F1 ATP 합성효소 서브유닛 베타와 관련된 RBS	119
S. 스피노사	RBS24	분자 샤페론 DnaK와 관련된 RBS	120
S. 스피노사	RBS25	파지 충격 단백질 A, PspA와 관련된 RBS	121

S. 스피노사	RBS26	2-옥소글루타레이트 디카복실라아제와 관련된 RBS	122
S. 스피노사	RBS27	5-메틸테트라하이드로프테로일트리 글루타메이트 호모시스테인 메틸트랜스퍼라제와 관련된 RBS	123
S. 스피노사	RBS28	50S 리보솜 단백질 L7/L12와 관련된 RBS	124
S. 스피노사	RBS29	DNA-지시된 RNA 폴리머라제 서브유닛 알파 와 관련된 RBS	125
S. 스피노사	RBS30	30S 리보솜 단백질 S5와 관련된 RBS	126
S. 스피노사	RBS31	DnaK(6929)와 관련된 RBS	127
트랜스포존 서열			
소스	**명칭**	**설명**	**서열번호**
인공 서열	기능상실(LoF) 트랜스포존	트랜스포존 돌연변이유발 페이로드 서열 LoF	128
인공 서열	기능획득(GoF) 트랜스포존	트랜스포존 돌연변이유발 페이로드 서열 프로모터를 갖는 기능획득	129
인공 서열	기능획득 재활용 가능 트랜스포존	트랜스포존 돌연변이유발 페이로드 서열 카운서 선택 마커를 갖는 기능획득	130
인공 서열	기능획득 용해도 태그 트랜스포존	트랜스포존 돌연변이유발 페이로드 서열 용해도 태그를 갖는 기능획득	131
인공 서열	용해도 태그	트랜스포존 제작물에 포함될 수 있는 GST 용해도 태그 서열	166
중성 부위 서열			
소스	**명칭**	**설명**	**서열번호**
S. 스피노사	SS_NeutralSite_1	*S. 스피노사* 중성 부위 1	132
S. 스피노사	SS_NeutralSite_2	*S. 스피노사* 중성 부위 2	133
S. 스피노사	SS_NeutralSite_3	*S. 스피노사* 중성 부위 3	134
S. 스피노사	SS_NeutralSite_4	*S. 스피노사* 중성 부위 4	135
S. 스피노사	SS_NeutralSite_5	*S. 스피노사* 중성 부위 5	136
S. 스피노사	SS_NeutralSite_6	*S. 스피노사* 중성 부위 6	137
S. 스피노사	SS_NeutralSite_7	*S. 스피노사* 중성 부위 7	138
S. 스피노사	SS_NeutralSite_8	*S. 스피노사* 중성 부위 8	139
S. 스피노사	SS_NeutralSite_9	*S. 스피노사* 중성 부위 9	140
S. 스피노사	SS_NeutralSite_10	*S. 스피노사* 중성 부위 10	141
S. 스피노사	SS_NeutralSite_11	*S. 스피노사* 중성 부위 11	142
선택 및 카운터 선택 마커			
인공 서열	SacB 유전자	sacB 유전자 서열, *S.스피노사*에 대해 코돈 최적화됨	146
인공 서열	돌연변이된 pheS 유전자 (*S.에리트라에아*)	미야자키 2015에 기술된 돌연변이를 갖는 *S. 에리트라에아* 유전자 서열	147
인공 서열	Mutated pheS 유전자 (*S.스피노사*)	미야자키 2015에 기술된 돌연변이를 갖는 *S. 스피노사* 유전자 서열	148
S. 에리트라에아	ermE 프로모터 서열	ermE 선택 유전자의 발현을 유도하는 ermE 프로모터 서열	149
인공 서열	aac(3)IV	아프라마이신에 대한 저항성을 부여하는 aac(3)IV 단백질	151
인공 서열	aacC1	젠타마이신에 대한 저항성을 부여하는 aacC1 단백질	152
인공 서열	aacC8	네오마이신 B에 대한 저항성을 부여하는 aacC8 단백질	153
인공 서열	aadA	스펙티노마이신/스트렙토마이신에 대한 저항성을 부여하는 aadA 단백질	154
인공 서열	ble	블레오마이신에 대한 저항성을 부여하는 ble 단백질	155
인공 서열	cat	클로로람페니콜에 대한 저항성을 부여하는 cat 단백질	156
인공 서열	ermE	에리스로마이신에 대한 저항성을 부여하는 ermE 단백질	157

인공 서열	hyg	하이그로마이신에 대한 저항성을 부여하는 hyg 단백질	158
인공 서열	neo	카나마이신에 대한 저항성을 부여하는 neo 단백질	159
인공 서열	amdSYM	카운터 선택 마커 amdSYM 유전자	160
인공 서열	tetA	카운터 선택 마커 tetA 유전자	161
인공 서열	lacY	카운터 선택 마커 lacY 유전자	162
인공 서열	sacB	카운터 선택 마커 sacB 유전자	163
인공 서열	pheS, S. 에리트라에아	S. 에리트라에아로부터 유래한 카운터 선택 마커 pheS 유전자	164
인공 서열	pheS, 코리네박테리움	코리네박테리움으로부터 유래한 카운터 선택 마커 pheS 유전자	165
인테그라제 부착 부위(att 부위)			
소스	**명칭**	**설명**	**서열번호**
사카로폴리스포라 엔도피티카	-	pCM32의 attP 부위	167
사카로폴리스포라 에리트라에아	-	pSE101의 attP 부위	168
사카로폴리스포라 에리트라에아	-	pSE211의 attP 부위	169
S. 스피노사	-	pSE101 상동체의 attP 부위	170
S. 스피노사	-	pSE211 상동체의 attP 부위	171

[1282] **본 개시의 번호가 매겨진 실시태양들**

[1283] 첨부된 항목들에도 불구하고, 본 발명은 다음의 번호가 매겨진 실시태양들을 제시한다.

[1284] **사카로폴리스포라 종을 진화시키는 고 처리량 게놈 공학**

[1285] 1. 다음 단계를 포함하여 원하는 표현형을 획득하기 위해 사카로폴리스포라 종(*Saccharopolyspora sp.*) 미생물을 진화시키는 게놈 공학의 고 처리량(HTP) 방법:

[1286] a. 동일한 게놈 균주 배경을 갖는 초기 복수의 사카로폴리스포라 미생물의 게놈을 교란시켜, 독특한 유전자 변이를 갖는 개별 사카로폴리스포라 균주를 포함하는 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리를 생성하는 단계;

[1287] b. 원하는 표현형에 대해 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리의 개별 사카로폴리스포라 균주를 선별 및 선택하는 단계;

[1288] c. 선행 단계에서 선별된 적어도 2개의 개별 사카로폴리스포라 균주에 존재하는 유전자 변이로부터 선택된 유전자 변이의 독특한 조합을 각각 포함하는 후속 복수의 사카로폴리스포라 미생물을 제공하여, 후속 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리를 생성하는 단계;

[1289] d. 원하는 표현형에 대해 후속 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리의 개별 사카로폴리스포라 균주를 선별 및 선택하는 단계; 및

[1290] e. 사카로폴리스포라 미생물이 원하는 표현형을 획득할 때까지 선형 또는 비선형 방식으로 단계 c)-d)를 1회 이상 반복하는 단계, 여기서 각각의 후속 반복은 선행 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리의 적어도 2개의 개별 사카로폴리스포라 균주 중에서 선택된 유전자 변이의 조합인 독특한 유전자 변이를 보유하는 개별 사카로폴리스포라 균주를 포함하는 새로운 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리를 생성한다.

[1291] 1.1 다음 단계를 포함하여 원하는 표현형을 획득하기 위해 사카로폴리스포라 종(*Saccharopolyspora sp.*) 미생물을 진화시키는 게놈 공학의 고 처리량(HTP) 방법:

[1292] a. 독특한 유전자 변이를 갖는 개별 사카로폴리스포라 균주를 포함하는 초기 복수의 사카로폴리스포라 미생물을 수득함으로써, 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리를 생성하는 단계;

[1293] b. 원하는 표현형에 대해 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리의 개별 사카로폴리스포라 균주를 선별 및 선택하는 단계;

[1294] c. 선행 단계에서 선별된 적어도 2개의 개별 사카로폴리스포라 균주에 존재하는 유전자 변이로부터 선택된 유전자 변이의 독특한 조합을 각각 포함하는 후속 복수의 사카로폴리스포라 미생물을 제공하여, 후속 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리를 생성하는 단계;

[1295] d. 원하는 표현형에 대해 후속 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리의 개별 사카로폴리스포라 균주를 선별 및 선택하는 단계; 및

[1296] e. 사카로폴리스포라 미생물이 원하는 표현형을 획득할 때까지 선형 또는 비선형 방식으로 단계 c)-d)를 1회 이상 반복하는 단계, 여기서 각각의 후속 반복은 선행 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리의 적어도 2개의 개별 사카로폴리스포라 균주 중에서 선택된 유전자 변이의 조합인 독특한 유전자 변이를 보유하는 개별 사카로폴리스포라 균주를 포함하는 새로운 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리를 생성한다.

[1297] 1.2 제 1.1 항에 있어서, 독특한 유전자 변이를 갖는 개별 사카로폴리스포라 균주를 포함하는 초기 복수의 사카로폴리스포라 미생물이 동일한 게놈 균주 배경을 갖는 초기 복수의 사카로폴리스포라 미생물의 게놈을 교란시킴으로써 생성되는 것인 HTP 방법.

[1298] 2. 제 1 항 내지 1.2 항 중 어느 한 항에 있어서, 단계 b에서 유전자 변이가 조합되기 전에 유전자 변이를 함유하는 유전자의 기능 및/또는 신원이 고려되지 않는 것인 게놈 공학의 HTP 방법.

[1299] 3. 제 1 항 내지 제 2 항 중 어느 한 항에 있어서, 조합될 적어도 하나의 유전자 변이가 코딩 DNA 모듈의 반복 세그먼트를 함유하는 게놈 영역에 존재하지 않는 것인 게놈 공학의 HTP 방법.

[1300] 4. 제 1 항에 있어서, 단계 c에서 유전자 변이의 독특한 조합을 각각 포함하는 후속 복수의 사카로폴리스포라 미생물은 다음에 의해 생성되는 것인 게놈 공학의 HTP 방법:

[1301] 1) 플라스미드를 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리에 속하는 개별 사카로폴리스포라 균주에 도입하는 단계로, 여기서 플라스미드는 선택 마커, 카운터 선택 마커, 기본 사카로폴리스포라 균주의 게놈 유전자좌(locus)와 상동성을 갖는 DNA 단편, 및 플라스미드 백본 서열을 포함하며, 여기서 DNA 단편은 또한 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리에 속하는 다른 개별 사카로폴리스포라 균주로부터 유래된 유전자 변이를 갖는다;

[1302] 2) 게놈에서 선택 마커의 존재에 기초하여 통합 이벤트를 갖는 사카로폴리스포라 균주를 선택하는 단계;

[1303] 3) 카운터 선택 마커 유전자의 부재에 기초하여 루프 아웃된 플라스미드 백본을 갖는 사카로폴리스포라 균주를 선택하는 단계.

[1304] 5. 제 1 항 내지 제 4 항 중 어느 한 항에 있어서, 플라스미드가 온도 민감성 레플리콘을 포함하지 않는 것인 게놈 공학의 HTP 방법.

[1305] 6. 제 1 항 내지 제 5 항 중 어느 한 항에 있어서, 선택하는 단계 3)이 통합된 플라스미드의 복제없이 수행되는 것인 게놈 공학의 HTP 방법.

[1306] 7. 제 1 항 내지 제 6 항 중 어느 한 항에 있어서, 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리는 프로모터 스왑 미생물 균주 라이브러리, SNP 스왑 미생물 균주 라이브러리, 개시/중지 코돈 미생물 균주 라이브러리, 최적화된 서열 미생물 균주 라이브러리, 터미네이터 스왑 미생물 균주 라이브러리, 트랜스포존 돌연변이유발 미생물 균주 다양성 라이브러리, 리보솜 결합 부위 미생물 균주 라이브러리, 항-대사산물/발효 생성물 저항성 라이브러리, 종결 삽입 미생물 균주 라이브러리, 및 이의 임의의 조합으로 이루어진 그룹으로부터 선택된 적어도 하나의 라이브러리를 포함하는 것인 게놈 공학의 HTP 방법.

[1307] 8. 제 1 항 내지 제 7 항 중 어느 한 항에 있어서, 후속 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리는 초기 HTP 유전자 디자인 미생물 균주 라이브러리의 완전한 조합(full combinatorial) 사카로폴리스포라 균주 라이브러리인 게놈 공학의 HTP 방법.

[1308] 9. 제 1 항 내지 제 8 항 중 어느 한 항에 있어서, 후속 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리는 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리의 유전자 변이로부터 유래된 완전한 조합 사카로폴리스포라 균주 라이브러리의 서브세트인 게놈 공학의 HTP 방법.

[1309] 10. 제 1 항 내지 제 9 항 중 어느 한 항에 있어서, 균주 라이브러리의 유전자 변이로부터 유래된 후속 HTP 유

전자 디자인이 선행 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리의 유전자 변이로부터 유래된 완전한 조합 미생물 균주 라이브러리인 게놈 공학의 HTP 방법.

[1310] 11. 제 1 항 내지 제 10 항 중 어느 한 항에 있어서, 후속 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리는 선행 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리의 유전자 변이로부터 유래된 완전한 조합 사카로폴리스포라 균주 라이브러리의 서브세트인 게놈 공학의 HTP 방법.

[1311] 12. 제 1 항 내지 제 11 항 중 어느 한 항에 있어서, 게놈을 교란시키는 단계는 랜덤 돌연변이유발, 표적 서열 삽입, 표적 서열 결실, 표적 서열 교체, 트랜스포존 돌연변이유발, 및 이의 임의의 조합으로 이루어진 그룹으로부터 선택된 적어도 하나의 방법을 사용하는 것을 포함하는 것인 게놈 공학의 HTP 방법.

[1312] 13. 제 1 항 내지 제 12 항 중 어느 한 항에 있어서, 초기 복수의 사카로폴리스포라 미생물이 생산 사카로폴리스포라 균주로부터 유래된 독특한 유전자 변이를 포함하는 것인 게놈 공학의 HTP 방법.

[1313] 14. 제 1 항 내지 제 13 항 중 어느 한 항에 있어서, 초기 복수의 사카로폴리스포라 미생물은 S_1Gen_1으로 표시된 생산 균주 미생물 및 S_nGen_n으로 표시된 이로부터 유래된 임의의 수의 후속 미생물 세대를 포함하는 것인 게놈 공학의 HTP 방법.

[1314] 15. 제 1 항 내지 제 14 항 중 어느 한 항에 있어서, 단계 c는 원형질체 융합 기술을 사용하여 유전자 변이를 신속하게 통합시키는 것을 포함하는 것인 게놈 공학의 HTP 방법.

[1315] 16. 제 1 항 내지 제 15 항 중 어느 한 항에 있어서, 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리 또는 후속 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리가 프로모터 스왑 미생물 균주 라이브러리를 포함하는 것인 게놈 공학의 HTP 방법.

[1316] 17. 제 16 항에 있어서, 프로모터 스왑 미생물 균주 라이브러리가 서열번호 1 내지 69 및 172 내지 175로부터 선택된 뉴클레오티드 서열을 갖는 적어도 하나의 프로모터를 포함하는 것인 게놈 공학의 HTP 방법.

[1317] 18. 제 1 항 내지 제 17 항 중 어느 한 항에 있어서, 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리 또는 후속 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리가 SNP 스왑 미생물 균주 라이브러리를 포함하는 것인 게놈 공학의 HTP 방법.

[1318] 19. 제 1 항 내지 제 18 항 중 어느 한 항에 있어서, 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리 또는 후속 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리가 터미네이터 스왑 미생물 균주 라이브러리를 포함하는 것인 게놈 공학의 HTP 방법.

[1319] 20. 제 19 항에 있어서, 터미네이터 스왑 미생물 균주 라이브러리가 서열번호 70 내지 80으로부터 선택된 뉴클레오티드 서열을 갖는 적어도 하나의 터미네이터를 포함하는 것인 게놈 공학의 HTP 방법.

[1320] 21. 제 1 항 내지 제 20 항 중 어느 한 항에 있어서, 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리 또는 후속 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리가 트랜스포존 돌연변이유발 미생물 균주 다양성 라이브러리를 포함하는 것인 게놈 공학의 HTP 방법.

[1321] 22. 제 21 항에 있어서, 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리 또는 후속 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리가 기능상실(Loss-of-Function; LoF) 트랜스포존 및/또는 기능획득(Gain-of-Function; GoF) 트랜스포존을 포함하는 것인 게놈 공학의 HTP 방법.

[1322] 23. 제 22 항에 있어서, GoF 트랜스포존이 용해도 태그, 프로모터 및/또는 카운터 선택 마커를 포함하는 것인 게놈 공학의 HTP 방법.

[1323] 24. 제 1 항 내지 제 23 항 중 어느 한 항에 있어서, 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리 또는 후속 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리가 리보솜 결합 부위 미생물 균주 라이브러리를 포함하는 것인 게놈 공학의 HTP 방법.

[1324] 25. 제 24 항에 있어서, 리보솜 결합 부위 미생물 균주 라이브러리가 서열번호 97 내지 127로부터 선택된 뉴클레오티드 서열을 갖는 적어도 하나의 리보솜 결합 부위(RBS)를 포함하는 것인 게놈 공학의 HTP 방법.

[1325] 26. 제 1 항 내지 제 25 항 중 어느 한 항에 있어서, 초기 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리 또는 후속 HTP 유전자 디자인 사카로폴리스포라 균주 라이브러리가 항-대사산물/발효 생성물 저항성 라이브

러리를 포함하는 것인 게놈 공학의 HTP 방법.

[1326] 27. 제 26 항에 있어서, 항-대사산물/발효 생성물 저항성 라이브러리는 사카로폴리스포라에서 스피노신 합성에 관련된 분자에 대한 사카로폴리스포라 균주 저항성을 포함하는 것인 게놈 공학의 HTP 방법.

[1327] **SNP 스왑 사카로폴리스포라 균주 라이브러리 생성**

[1328] 28. 다음 단계를 포함하는 SNP 스왑 사카로폴리스포라 균주 라이브러리를 생성하는 방법:

[1329] a. 기준(reference) 사카로폴리스포라 균주 및 제 2 사카로폴리스포라 균주를 제공하는 단계, 여기서 제 2 사카로폴리스포라 균주는 기준 사카로폴리스포라 균주에 존재하지 않는 단일 뉴클레오티드 다형성, DNA 삽입 및 DNA 결실로부터 선택된 복수의 확인된 유전자 변이를 포함한다; 및

[1330] b. 기준 사카로폴리스포라 균주 또는 제 2 사카로폴리스포라의 게놈을 교란시켜 상기 복수의 개별 사카로폴리스포라 균주의 각각의 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 사카로폴리스포라 균주를 포함하는 초기 SNP 스왑 사카로폴리스포라 균주 라이브러리를 생성하는 단계, 여기서 상기 각각의 독특한 유전자 변이는 기준 사카로폴리스포라 균주와 제 2 사카로폴리스포라 균주 사이의 복수의 확인된 유전자 변이로부터 선택된 단일 유전자 변이에 상응한다.

[1331] 29. 제 28 항에 있어서, 기준 사카로폴리스포라 균주의 게놈은 제 2 사카로폴리스포라 균주에서 발견되는 확인된 단일 뉴클레오티드 다형성, DNA 삽입 또는 DNA 결실의 하나 이상을 첨가하기 위해 교란되는 것인 SNP 스왑 사카로폴리스포라 균주 라이브러리를 생성하는 방법.

[1332] 30. 제 28 항 내지 제 29 항 중 어느 한 항에 있어서, 제 2 사카로폴리스포라 균주의 게놈은 기준 사카로폴리스포라 균주에서 발견되지 않은 확인된 단일 뉴클레오티드 다형성, DNA 삽입 또는 DNA 결실의 하나 이상을 제거하기 위해 교란되는 것인 SNP 스왑 사카로폴리스포라 균주 라이브러리를 생성하는 방법.

[1333] 31. 제 28 항 내지 제 30 항 중 어느 한 항에 있어서, 독특한 유전자 변이를 갖는 생성된 복수의 개별 사카로폴리스포라 균주는 기준 사카로폴리스포라 균주와 제 2 사카로폴리스포라 균주 사이의 모든 확인된 유전자 변이의 완전한 조합 라이브러리를 포함하는 것인 SNP 스왑 사카로폴리스포라 균주 라이브러리를 생성하는 방법.

[1334] 32. 제 28 항 내지 제 31 항 중 어느 한 항에 있어서, 독특한 유전자 변이를 갖는 생성된 복수의 개별 사카로폴리스포라 균주는 기준 사카로폴리스포라 균주와 제 2 사카로폴리스포라 균주 사이의 모든 확인된 유전자 변이의 완전한 조합 라이브러리의 서브세트를 포함하는 것인 SNP 스왑 사카로폴리스포라 균주 라이브러리를 생성하는 방법.

[1335] **생산 사카로폴리스포라 균주의 표현형 성능의 재건 및 개량**

[1336] 33. 다음 단계를 포함하는 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법:

[1337] a. 부모 계통의 사카로폴리스포라 균주 및 이로부터 유래된 생산 사카로폴리스포라 균주를 제공하는 단계, 여기서 생산 사카로폴리스포라 균주는 부모 계통 사카로폴리스포라 균주에 존재하지 않는 단일 뉴클레오티드 다형성, DNA 삽입 및 DNA 결실로부터 선택된 복수의 확인된 유전자 변이를 포함한다;

[1338] b. 부모 계통 사카로폴리스포라 균주 또는 생산 사카로폴리스포라 균주의 게놈을 교란시켜 초기 사카로폴리스포라 균주 라이브러리를 생성하는 단계, 여기서 초기 라이브러리의 각각의 균주는 부모 계통의 사카로폴리스포라 균주 및 생산 사카로폴리스포라 균주 사이의 복수의 확인된 유전자 변이로부터의 독특한 유전자 변이를 포함한다;

[1339] c. 기준 사카로폴리스포라 균주에 대한 표현형 성능 개량을 위한 초기 SNP 스왑 사카로폴리스포라 균주 라이브러리의 개별 사카로폴리스포라 균주를 선별 및 선택하여, 표현형 성능 개량을 부여하는 독특한 유전자 변이를 확인하는 단계;

[1340] d. 선행 단계에서 선별된 적어도 2개의 개별 사카로폴리스포라 균주에 존재하는 변이로부터 독특한 유전자 변이의 조합을 각각 포함하는 후속 복수의 미생물을 제공하여, 사카로폴리스포라 균주의 후속 라이브러리를 생성하는 단계;

[1341] e. 기준 사카로폴리스포라 균주에 대한 표현형 성능 개량을 위한 후속 균주 라이브러리의 개별 균주를 선별 및 선택하여, 추가 표현형 성능 개량을 부여하는 유전자 변이의 독특한 조합을 확인하는 단계; 및

[1342] f. 사카로폴리스포라 균주가 생산 사카로폴리스포라 균주의 표현형 성능에 비해 원하는 수준의 개량된 표현형

성능을 나타낼 때까지 선형 또는 비선형 방식으로 단계 d)-e)를 1회 이상 반복하는 단계, 여기서 각각의 후속 반복은 사카로폴리스포라 균주의 새로운 라이브러리를 생성하며, 여기서 새로운 라이브러리의 각각의 균주는 선행 라이브러리의 적어도 2개의 개별 사카로폴리스포라 균주 중에서 선택된 유전자 변이의 조합인 유전자 변이를 포함한다.

[1343] 34. 제 33 항에 있어서, 사카로폴리스포라 균주의 초기 라이브러리는 부모 계통 사카로폴리스포라 균주 및 생산 사카로폴리스포라 사이의 모든 확인된 유전자 변이를 포함하는 완전한 조합 라이브러리인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법.

[1344] 35. 제 33 항 내지 제 34 항 중 어느 한 항에 있어서, 사카로폴리스포라 균주의 초기 라이브러리가 기준 부모 계통 사카로폴리스포라 균주 및 생산 사카로폴리스포라 균주 사이에서 확인된 유전자 변이의 서브세트를 포함하는 완전한 조합 라이브러리의 서브세트인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법.

[1345] 36. 제 33 항 내지 제 35 항 중 어느 한 항에 있어서, 사카로폴리스포라 균주의 후속 라이브러리가 초기 라이브러리의 완전한 조합 라이브러리인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법.

[1346] 37. 제 33 항 내지 제 36 항 중 어느 한 항에 있어서, 사카로폴리스포라 균주의 후속 라이브러리가 초기 라이브러리의 완전한 조합 라이브러리인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법.

[1347] 38. 제 33 항 내지 제 37 항 중 어느 한 항에 있어서, 사카로폴리스포라 균주의 후속 라이브러리가 선행 라이브러리의 완전한 조합 라이브러리인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법.

[1348] 39. 제 33 항 내지 제 38 항 중 어느 한 항에 있어서, 사카로폴리스포라 균주의 후속 라이브러리가 선행 라이브러리의 완전한 조합 라이브러리의 서브세트인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법.

[1349] 40. 제 33 항 내지 제 39 항 중 어느 한 항에 있어서, 부모 계통 사카로폴리스포라 균주의 게놈이 생산 사카로폴리스포라 균주에서 발견되는 확인된 단일 뉴클레오티드 다형성, DNA 삽입 또는 DNA 결실의 하나 이상을 첨가하기 위해 교란되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법.

[1350] 41. 제 33 항 내지 제 40 항 중 어느 한 항에 있어서, 생산 사카로폴리스포라 균주의 게놈이 부모 계통 사카로폴리스포라 균주에서 발견되지 않은 확인된 단일 뉴클레오티드 다형성, DNA 삽입 또는 DNA 결실 중 하나 이상을 제거하기 위해 교란되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법.

[1351] 42. 제 33 항 내지 제 41 항 중 어느 한 항에 있어서, 게놈을 교란시키는 단계는 랜덤 돌연변이유발, 표적 서열 삽입, 표적 서열 결실, 표적 서열 교체 및 이의 조합으로 이루어진 그룹으로부터 선택된 적어도 하나의 방법을 사용하는 것을 포함하는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법.

[1352] 43. 제 33 항 내지 제 41 항 중 어느 한 항에 있어서, 후속 라이브러리의 사카로폴리스포라 균주의 표현형 성능이 생산 사카로폴리스포라 균주의 표현형 성능과 비교하여 측정된 표현형 변수에서 적어도 10% 증가를 나타낼 때까지 단계 d)-e)가 반복되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법.

[1353] 44. 제 33 항 내지 제 43 항 중 어느 한 항에 있어서, 후속 라이브러리의 사카로폴리스포라 균주의 표현형 성능이 생산 사카로폴리스포라 균주의 표현형 성능과 비교하여 측정된 표현형 변수에서 적어도 1배 증가를 나타낼 때까지 단계 d)-e)가 반복되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법.

[1354] 45. 제 33 항 내지 제 44 항 중 어느 한 항에 있어서, 단계 f)의 개량된 표현형 성능이 관심 생성물의 부피 생산성, 관심 생성물의 비 생산성, 관심 생성물의 수율, 관심 생성물의 역가 및 이의 조합으로 이루어진 그룹으로부터 선택되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법.

[1355] 46. 제 33 항 내지 제 45 항 중 어느 한 항에 있어서, 단계 f)의 개량된 표현형 성능이 관심 생성물의 증가된 또는 보다 효율적인 생산이며, 상기 관심 생성물은 소분자, 효소, 펩타이드, 아미노산, 유기산, 합성 화합물, 연료, 알코올, 1차 세포외 대사산물, 2차 세포외 대사산물, 세포내 성분 분자 및 이의 조합으로 이루어진 그룹으로부터 선택되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법.

[1356] 47. 제 46 항에 있어서, 관심 생성물이 스피노신, 스피노사드, 스피네토람, 제니스테인, 콜린 옥시다제, 쿠마미딘 화합물, 에리스로마이신, 이버멕틴 아글리콘, HMG-CoA 환원효소 억제제, 카복실산 이성질체, 알파-메틸 메티오닌, 티알리신, 알파-케토부티레이트, 아스파테이트 하이드록사메이트, 아자세린, 5-플루오로인돌, 베타-하이

드록시노르발린, 세룰레닌, 퓨린, 피리미딘 및 이의 유사체로 이루어진 그룹으로부터 선택되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법.

[1357] 48. 제 46 항에 있어서, 스피노신이 스피노신 A, 스피노신 D, 스피노신 J, 스피노신 L 또는 이의 조합인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법.

[1358] 49. 제 33 항 내지 제 48 항 중 어느 한 항에 있어서, 확인된 유전자 변이가 프로모터 스왑 라이브러리로부터의 인공 프로모터 스왑 유전자 변이를 추가로 포함하는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법.

[1359] 50. 제 33 항 내지 제 49 항 중 어느 한 항에 있어서, 사카로폴리스포라 균주의 초기 라이브러리 또는 사카로폴리스포라 균주의 후속 라이브러리 중 적어도 하나의 미생물 균주의 게놈을 조작하여, 내인성 사카로폴리스포라 표적 유전자에 작동 가능하게 연결된 프로모터 래더로부터의 하나 이상의 프로모터를 포함하는 것을 추가로 포함하는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법.

[1360] 51. 제 33 항 내지 제 50 항 중 어느 한 항에 있어서, 균주 라이브러리는 프로모터 스왑 미생물 균주 라이브러리, SNP 스왑 미생물 균주 라이브러리, 개시/중지 코돈 미생물 균주 라이브러리, 최적화된 서열 미생물 균주 라이브러리, 터미네이터 스왑 미생물 균주 라이브러리, 트랜스포존 돌연변이유발 미생물 균주 다양성 라이브러리, 리보솜 결합 부위 미생물 균주 라이브러리, 항-대사산물/발효 생성물 저항성 라이브러리, 종결 삽입 미생물 균주 라이브러리, 및 이의 임의의 조합으로 이루어진 그룹으로부터 선택된 적어도 하나의 라이브러리를 포함하는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법.

[1361] 52. 제 51 항에 있어서, 균주 라이브러리가 다음으로 이루어진 그룹으로부터 선택된 적어도 하나의 라이브러리를 포함하는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 재건 및 개량시키는 방법:

[1362] 1) 서열번호 1 내지 69로부터 선택된 서열을 갖는 적어도 하나의 프로모터를 포함하는 프로모터 스왑 미생물 균주 라이브러리;

[1363] 2) 서열번호 70 내지 80으로부터 선택된 서열을 갖는 적어도 하나의 터미네이터를 포함하는 터미네이터 스왑 미생물 균주 라이브러리; 및

[1364] 3) 서열번호 97 내지 127로부터 선택된 서열을 갖는 적어도 하나의 RBS를 포함하는 리보솜 결합 부위(RBS) 라이브러리.

[1365] **프로모터 스왑 사카로폴리스포라 균주 라이브러리의 생성 및 생산 사카로폴리스포라 균주의 표현형 성능 개량을 위한 이의 이용**

[1366] 53. 다음 단계를 포함하는 프로모터 스왑 사카로폴리스포라 균주 라이브러리를 생성하는 방법:

[1367] a. 기본 사카로폴리스포라 균주에 내인성인 복수의 표적 유전자 및 프로모터 래더를 제공하는 단계, 여기서 상기 프로모터 래더는 기본 사카로폴리스포라 균주에서 상이한 발현 프로파일을 나타내는 복수의 프로모터를 포함한다; 및

[1368] b. 기본 사카로폴리스포라 균주의 게놈을 조작하여 상기 복수의 개별 사카로폴리스포라 균주의 각각의 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 사카로폴리스포라 균주를 포함하는 초기 프로모터 스왑 사카로폴리스포라 균주 라이브러리를 생성하는 단계, 여기서 상기 독특한 유전자 변이의 각각은 기본 사카로폴리스포라 균주에 내인성인 표적 유전자의 하나에 작동 가능하게 연결된 프로모터 래더로부터의 프로모터의 하나 이상을 포함한다.

[1369] 54. 제 53 항에 있어서, 복수의 프로모터 중 적어도 하나가 서열번호 1 내지 69로부터 선택된 서열을 갖는 프로모터를 포함하는 것인 프로모터 스왑 사카로폴리스포라 균주 라이브러리를 생성하는 방법.

[1370] 55. 다음 단계를 포함하는 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 프로모터 스왑 방법:

[1371] a. 기본 사카로폴리스포라 균주에 내인성인 복수의 표적 유전자 및 프로모터 래더를 제공하는 단계, 여기서 상기 프로모터 래더는 상기 기본 사카로폴리스포라 균주에서 상이한 발현 프로파일을 나타내는 복수의 프로모터를 포함한다;

[1372] b. 기본 사카로폴리스포라 균주의 게놈을 조작하여, 상기 복수의 개별 사카로폴리스포라 균주의 각각의 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 사카로폴리스포라 균주를 포함하는 초기 프로모터 스왑 사

카로폴리스포라 균주 라이브러리를 생성하는 단계, 여기서 상기 독특한 유전자 변이의 각각은 기본 사카로폴리스포라 균주에 내인성인 표적 유전자의 하나에 작동 가능하게 연결된 프로모터 래더로부터의 프로모터의 하나 이상을 포함한다;

[1373] c. 기준 사카로폴리스포라 균주에 대한 표현형 성능 개량을 위한 초기 프로모터 스왑 사카로폴리스포라 균주 라이브러리의 개별 사카로폴리스포라 균주를 선별 및 선택하여, 표현형 성능 개량을 부여하는 독특한 유전자 변이를 확인하는 단계;

[1374] d. 선행 단계에서 선별된 적어도 2개의 개별 사카로폴리스포라 균주에 존재하는 유전자 변이로부터 독특한 유전자 변이의 조합을 각각 포함하는 후속 복수의 사카로폴리스포라 미생물을 제공하여, 후속 프로모터 스왑 사카로폴리스포라 균주 라이브러리를 생성하는 단계;

[1375] e. 기준 대장균(대장균) 균주에 대한 원하는 표현형 성능 개량을 위한 후속 프로모터 스왑 사카로폴리스포라 균주 라이브러리의 개별 사카로폴리스포라 균주를 선별 및 선택하여, 추가 표현형 성능 개량을 부여하는 유전자 변이의 독특한 조합을 확인하는 단계; 및

[1376] f. 사카로폴리스포라 균주가 생산 사카로폴리스포라 균주의 표현형 성능에 비해 원하는 수준의 개량된 표현형 성능을 나타낼 때까지 선형 또는 비선형 방식으로 단계 d)-e)를 1회 이상 반복하는 단계, 여기서 각각의 후속 반복은 사카로폴리스포라 균주의 새로운 프로모터 스왑 사카로폴리스포라 균주 라이브러리를 생성하며, 여기서 새로운 라이브러리의 각각의 균주는 선행 프로모터 스왑 사카로폴리스포라 균주 라이브러리의 적어도 2개의 개별 사카로폴리스포라 균주 중에서 선택된 유전자 변이의 조합인 유전자 변이를 포함한다.

[1377] 56. 제 55 항에 있어서, 후속 프로모터 스왑 사카로폴리스포라 균주 라이브러리는 초기 프로모터 스왑 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 프로모터 스왑 방법.

[1378] 57. 제 55 항 내지 제 56 항 중 어느 한 항에 있어서, 후속 프로모터 스왑 사카로폴리스포라 균주 라이브러리는 초기 프로모터 스왑 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 프로모터 스왑 방법.

[1379] 58. 제 55 항 내지 제 57 항 중 어느 한 항에 있어서, 후속 프로모터 스왑 사카로폴리스포라 균주 라이브러리는 초기 프로모터 스왑 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리의 서브세트인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 프로모터 스왑 방법.

[1380] 59. 제 55 항 내지 제 58 항 중 어느 한 항에 있어서, 후속 프로모터 스왑 사카로폴리스포라 균주 라이브러리는 선행 프로모터 스왑 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 프로모터 스왑 방법.

[1381] 60. 제 55 항 내지 제 59 항 중 어느 한 항에 있어서, 후속 프로모터 스왑 사카로폴리스포라 균주 라이브러리는 선행 프로모터 스왑 사카로폴리스포라 균주 라이브러리의 전체 조합 라이브러리의 서브세트인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 프로모터 스왑 방법.

[1382] 61. 제 55 항 내지 제 60 항 중 어느 한 항에 있어서, 후속 프로모터 스왑 사카로폴리스포라 균주 라이브러리의 사카로폴리스포라 균주의 표현형 성능이 생산 사카로폴리스포라 균주의 표현형 성능과 비교하여 측정된 표현형 변수에서 적어도 10% 증가를 나타낼 때까지 단계 d)-e)가 반복되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 프로모터 스왑 방법.

[1383] 62. 제 55 항 내지 제 61 항 중 어느 한 항에 있어서, 후속 프로모터 스왑 사카로폴리스포라 균주 라이브러리의 사카로폴리스포라 균주의 표현형 성능이 생산 사카로폴리스포라 균주의 표현형 성능과 비교하여 측정된 표현형 변수에서 적어도 1배 증가를 나타낼 때까지 단계 d)-e)가 반복되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 프로모터 스왑 방법.

[1384] 63. 제 55 항 내지 제 62 항 중 어느 한 항에 있어서, 단계 f)의 개량된 표현형 성능이 관심 생성물의 부피 생산성, 관심 생성물의 비 생산성, 관심 생성물의 수율, 관심 생성물의 역가 및 이의 조합으로 이루어진 그룹으로부터 선택되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 프로모터 스왑 방법.

[1385] 64. 제 55 항 내지 제 63 항 중 어느 한 항에 있어서, 단계 f)의 개량된 표현형 성능이 관심 생성물의 증가된 또는 보다 효율적인 생산이며, 상기 관심 생성물은 소분자, 효소, 펩타이드, 아미노산, 유기산, 합성 화합물,

연료, 알코올, 1차 세포외 대사산물, 2차 세포외 대사산물, 세포내 성분 분자 및 이의 조합으로 이루어진 그룹으로부터 선택되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 프로모터 스왑 방법.

[1386] 65. 제 64 항에 있어서, 관심 생성물이 스피노신, 스피노사드, 스피네토람, 제니스테인, 콜린 옥시다제, 쿠마미딘 화합물, 에리스로마이신, 이버멕틴 아글리콘, HMG-CoA 환원효소 억제제, 카복실산 이성질체, 알파-메틸 메티오닌, 티알리신, 알파-케토부티레이트, 아스파테이트 하이드록사메이트, 아자세린, 5-플루오로인돌, 베타-하이드록시노르발린, 세룰레닌, 퓨린, 피리미딘 및 이의 유사체로 이루어진 그룹으로부터 선택되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 프로모터 스왑 방법.

[1387] 66. 제 65 항에 있어서, 스피노신이 스피노신 A, 스피노신 D, 스피노신 J, 스피노신 L 또는 이의 조합인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 프로모터 스왑 방법.

[1388] 67. 제 55 항 내지 제 66 항 중 어느 한 항에 있어서, 프로모터 래더가 서열번호 1 내지 69로부터 선택된 뉴클레오티드 서열을 갖는 적어도 하나의 프로모터를 포함하는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 프로모터 스왑 방법.

[1389] **터미네이터 스왑 사카로폴리스포라 균주 라이브러리의 생성 및 생산 사카로폴리스포라 균주의 표현형 성능 개량을 위한 이의 이용**

[1390] 68. 다음 단계를 포함하는 터미네이터 스왑 사카로폴리스포라 균주 라이브러리를 생성하는 방법:

[1391] a. 기본 사카로폴리스포라 균주에 내인성인 복수의 표적 유전자 및 터미네이터 래더를 제공하는 단계, 여기서 상기 터미네이터 래더는 기본 사카로폴리스포라 균주에서 상이한 발현 프로파일을 나타내는 복수의 터미네이터를 포함한다; 및

[1392] b. 기본 사카로폴리스포라 균주의 게놈을 조작하여 상기 복수의 개별 사카로폴리스포라 균주의 각각의 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 사카로폴리스포라 균주를 포함하는 초기 터미네이터 스왑 사카로폴리스포라 균주 라이브러리를 생성하는 단계, 여기서 상기 독특한 유전자 변이의 각각은 기본 사카로폴리스포라 균주에 내인성인 표적 유전자의 하나에 작동 가능하게 연결된 터미네이터 래더로부터의 터미네이터의 하나 이상을 포함한다.

[1393] 69. 다음 단계를 포함하는 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 터미네이터 스왑 방법:

[1394] a. 기본 사카로폴리스포라 균주에 내인성인 복수의 표적 유전자 및 터미네이터 래더를 제공하는 단계, 여기서 상기 터미네이터 래더는 상기 기본 사카로폴리스포라 균주에서 상이한 발현 프로파일을 나타내는 복수의 터미네이터를 포함한다;

[1395] b. 기본 사카로폴리스포라 균주의 게놈을 조작하여, 상기 복수의 개별 사카로폴리스포라 균주의 각각의 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 사카로폴리스포라 균주를 포함하는 초기 터미네이터 스왑 사카로폴리스포라 균주 라이브러리를 생성하는 단계, 여기서 상기 독특한 유전자 변이의 각각은 기본 사카로폴리스포라 균주에 내인성인 표적 유전자의 하나에 작동 가능하게 연결된 터미네이터 래더로부터의 터미네이터의 하나 이상을 포함한다;

[1396] c. 기준 사카로폴리스포라 균주에 대한 표현형 성능 개량을 위한 초기 터미네이터 스왑 사카로폴리스포라 균주 라이브러리의 개별 사카로폴리스포라 균주를 선별 및 선택하여, 표현형 성능 개량을 부여하는 독특한 유전자 변이를 확인하는 단계;

[1397] d. 선행 단계에서 선별된 적어도 2개의 개별 사카로폴리스포라 균주에 존재하는 유전자 변이로부터 독특한 유전자 변이의 조합을 각각 포함하는 후속 복수의 사카로폴리스포라 미생물을 제공하여, 후속 터미네이터 스왑 사카로폴리스포라 균주 라이브러리를 생성하는 단계;

[1398] e. 기준 사카로폴리스포라 균주에 대한 표현형 성능 개량을 위한 후속 터미네이터 스왑 사카로폴리스포라 균주 라이브러리의 개별 사카로폴리스포라 균주를 선별 및 선택하여, 추가 표현형 성능 개량을 부여하는 유전자 변이의 독특한 조합을 확인하는 단계; 및

[1399] f. 사카로폴리스포라 균주가 생산 사카로폴리스포라 균주의 표현형 성능에 비해 원하는 수준의 개량된 표현형 성능을 나타낼 때까지 선형 또는 비선형 방식으로 단계 d)-e)를 1회 이상 반복하는 단계, 여기서 각각의 후속 반복은 미생물 균주의 새로운 터미네이터 스왑 사카로폴리스포라 균주 라이브러리를 생성하며, 여기서 새로운 라이브러리의 각각의 균주는 선행 라이브러리의 적어도 2개의 개별 사카로폴리스포라 균주 중에서 선택된 유전

자 변이의 조합인 유전자 변이를 포함한다.

[1400] 70. 제 69 항에 있어서, 후속 터미네이터 스왑 사카로폴리스포라 균주 라이브러리는 초기 터미네이터 스왑 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 터미네이터 스왑 방법.

[1401] 71. 제 69 항 내지 제 70 항 중 어느 한 항에 있어서, 후속 터미네이터 스왑 사카로폴리스포라 균주 라이브러리는 초기 터미네이터 스왑 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리의 서브세트인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 터미네이터 스왑 방법.

[1402] 72. 제 69 항 내지 제 71 항 중 어느 한 항에 있어서, 후속 터미네이터 스왑 사카로폴리스포라 균주 라이브러리는 선행 터미네이터 스왑 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 터미네이터 스왑 방법.

[1403] 73. 제 69 항 내지 제 72 항 중 어느 한 항에 있어서, 후속 터미네이터 스왑 사카로폴리스포라 균주 라이브러리는 선행 터미네이터 스왑 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리의 서브세트인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 터미네이터 스왑 방법.

[1404] 74. 제 69 항 내지 제 73 항 중 어느 한 항에 있어서, 후속 터미네이터 스왑 사카로폴리스포라 균주 라이브러리의 사카로폴리스포라 균주의 표현형 성능이 생산 사카로폴리스포라 균주의 표현형 성능과 비교하여 측정된 표현형 변수에서 적어도 10% 증가를 나타낼 때까지 단계 d)-e)가 반복되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 터미네이터 스왑 방법.

[1405] 75. 제 69 항 내지 제 74 항 중 어느 한 항에 있어서, 후속 터미네이터 스왑 사카로폴리스포라 균주 라이브러리의 사카로폴리스포라 균주의 표현형 성능이 생산 사카로폴리스포라 균주의 표현형 성능과 비교하여 측정된 표현형 변수에서 적어도 1배 증가를 나타낼 때까지 단계 d)-e)가 반복되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 터미네이터 스왑 방법.

[1406] 76. 제 69 항 내지 제 75 항 중 어느 한 항에 있어서, 단계 f)의 개량된 표현형 성능이 관심 생성물의 부피 생산성, 관심 생성물의 비 생산성, 관심 생성물의 수율, 관심 생성물의 역가 및 이의 조합으로 이루어진 그룹으로부터 선택되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 터미네이터 스왑 방법.

[1407] 77. 제 69 항 내지 제 76 항 중 어느 한 항에 있어서, 단계 f)의 개량된 표현형 성능이 관심 생성물의 증가된 또는 보다 효율적인 생산이며, 상기 관심 생성물은 소분자, 효소, 펩타이드, 아미노산, 유기산, 합성 화합물, 연료, 알코올, 1차 세포외 대사산물, 2차 세포외 대사산물, 세포내 성분 분자 및 이의 조합으로 이루어진 그룹으로부터 선택되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 터미네이터 스왑 방법.

[1408] 78. 제 69 항 내지 제 77 항 중 어느 한 항에 있어서, 관심 생성물이 스피노신, 스피노사드, 스피네토람, 제니스테인, 콜린 옥시다제, 쿠마미딘 화합물, 에리스로마이신, 이버멕틴 아글리콘, HMG-CoA 환원효소 억제제, 카복실산 이성질체, 알파-메틸 메티오닌, 티알리신, 알파-케토부티레이트, 아스파테이트 하이드록사메이트, 아자세린, 5-플루오로인돌, 베타-하이드록시노르발린, 세룰레닌, 퓨린, 피리미딘 및 이의 유사체로 이루어진 그룹으로부터 선택되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 터미네이터 스왑 방법.

[1409] 79. 제 78 항에 있어서, 스피노신이 스피노신 A, 스피노신 D, 스피노신 J, 스피노신 L 또는 이의 조합인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 터미네이터 스왑 방법.

[1410] 80. 제 69 항 내지 제 79 항 중 어느 한 항에 있어서, 터미네이터 래더가 서열번호 70 내지 80으로부터 선택된 뉴클레오티드 서열을 갖는 적어도 하나의 터미네이터를 포함하는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 터미네이터 스왑 방법.

[1411] **리보솜 결합 부위(RBS) 사카로폴리스포라 균주 라이브러리의 생성 및 생산 사카로폴리스포라 균주의 표현형 성능 개량을 위한 이의 이용**

[1412] 81. 다음 단계를 포함하는 리보솜 결합 부위(RBS) 사카로폴리스포라 균주 라이브러리를 생성하는 방법:

[1413] a. 기본 사카로폴리스포라 균주에 내인성인 복수의 표적 유전자 및 RBS 래더를 제공하는 단계, 여기서 상기 RBS 래더는 기본 사카로폴리스포라 균주에서 상이한 발현 프로파일을 나타내는 복수의 RBS를 포함한다; 및

[1414] b. 기본 사카로폴리스포라 균주의 게놈을 조작하여 상기 복수의 개별 사카로폴리스포라 균주의 각각의 균주 내

에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 사카로폴리스포라 균주를 포함하는 초기 RBS 사카로폴리스포라 균주 라이브러리를 생성하는 단계, 여기서 상기 독특한 유전자 변이의 각각은 기본 사카로폴리스포라 균주에 내인성인 표적 유전자의 하나에 작동 가능하게 연결된 RBS 래더로부터의 RBS의 하나 이상을 포함한다.

[1415] 82. 다음 단계를 포함하는 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 방법:

[1416] a. 기본 사카로폴리스포라 균주에 내인성인 복수의 표적 유전자 및 RBS 래더를 제공하는 단계, 여기서 상기 RBS 래더는 상기 기본 사카로폴리스포라 균주에서 상이한 발현 프로파일을 나타내는 복수의 RBS를 포함한다;

[1417] b. 기본 사카로폴리스포라 균주의 게놈을 조작하여, 상기 복수의 개별 사카로폴리스포라 균주의 각각의 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 사카로폴리스포라 균주를 포함하는 초기 RBS 사카로폴리스포라 균주 라이브러리를 생성하는 단계, 여기서 상기 독특한 유전자 변이의 각각은 기본 사카로폴리스포라 균주에 내인성인 표적 유전자의 하나에 작동 가능하게 연결된 RBS 래더로부터의 RBS의 하나 이상을 포함한다;

[1418] c. 기준 사카로폴리스포라 균주에 대한 표현형 성능 개량을 위한 초기 RBS 사카로폴리스포라 균주 라이브러리의 개별 사카로폴리스포라 균주를 선별 및 선택하여, 표현형 성능 개량을 부여하는 독특한 유전자 변이를 확인하는 단계;

[1419] d. 선행 단계에서 선별된 적어도 2개의 개별 사카로폴리스포라 균주에 존재하는 유전자 변이로부터 독특한 유전자 변이의 조합을 각각 포함하는 후속 복수의 사카로폴리스포라 균주를 제공하여, 후속 RBS 사카로폴리스포라 균주 라이브러리를 생성하는 단계;

[1420] e. 기준 사카로폴리스포라 균주에 대한 표현형 성능 개량을 위한 후속 RBS 사카로폴리스포라 균주 라이브러리의 개별 사카로폴리스포라 균주를 선별 및 선택하여, 추가 표현형 성능 개량을 부여하는 유전자 변이의 독특한 조합을 확인하는 단계; 및

[1421] f. 사카로폴리스포라 균주가 생산 사카로폴리스포라 균주의 표현형 성능에 비해 원하는 수준의 개량된 표현형 성능을 나타낼 때까지 선형 또는 비선형 방식으로 단계 d)-e)를 1회 이상 반복하는 단계, 여기서 각각의 후속 반복은 미생물 균주의 새로운 RBS 사카로폴리스포라 균주 라이브러리를 생성하며, 여기서 새로운 라이브러리의 각각의 균주는 선행 라이브러리의 적어도 2개의 개별 사카로폴리스포라 균주 중에서 선택된 유전자 변이의 조합인 유전자 변이를 포함한다.

[1422] 83. 제 82 항에 있어서, 후속 RBS 사카로폴리스포라 균주 라이브러리는 초기 RBS 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 방법.

[1423] 84. 제 82 항 내지 제 83 항 중 어느 한 항에 있어서, 후속 RBS 사카로폴리스포라 균주 라이브러리는 초기 RBS 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리의 서브세트인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 방법.

[1424] 85. 제 82 항 내지 제 84 항 중 어느 한 항에 있어서, 후속 RBS 사카로폴리스포라 균주 라이브러리는 선행 RBS 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 방법.

[1425] 86. 제 82 항 내지 제 85 항 중 어느 한 항에 있어서, 후속 RBS 사카로폴리스포라 균주 라이브러리는 선행 RBS 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리의 서브세트인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 방법.

[1426] 87. 제 82 항 내지 제 87 항 중 어느 한 항에 있어서, 후속 RBS 사카로폴리스포라 균주 라이브러리의 사카로폴리스포라 균주의 표현형 성능이 생산 사카로폴리스포라 균주의 표현형 성능과 비교하여 측정된 표현형 변수에서 적어도 10% 증가를 나타낼 때까지 단계 d)-e)가 반복되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 방법.

[1427] 88. 제 82 항 내지 제 87 항 중 어느 한 항에 있어서, 후속 RBS 사카로폴리스포라 균주 라이브러리의 사카로폴리스포라 균주의 표현형 성능이 생산 사카로폴리스포라 균주의 표현형 성능과 비교하여 측정된 표현형 변수에서 적어도 1배 증가를 나타낼 때까지 단계 d)-e)가 반복되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 방법.

[1428] 89. 제 82 항 내지 제 88 항 중 어느 한 항에 있어서, 단계 f)의 개량된 표현형 성능이 관심 생성물의 부피 생산성, 관심 생성물의 비 생산성, 관심 생성물의 수율, 관심 생성물의 역가 및 이의 조합으로 이루어진 그룹으로

부터 선택되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 방법.

[1429] 90. 제 82 항 내지 제 89 항 중 어느 한 항에 있어서, 단계 f)의 개량된 표현형 성능이 관심 생성물의 증가된 또는 보다 효율적인 생산이며, 상기 관심 생성물은 소분자, 효소, 펩타이드, 아미노산, 유기산, 합성 화합물, 연료, 알코올, 1차 세포외 대사산물, 2차 세포외 대사산물, 세포내 성분 분자 및 이의 조합으로 이루어진 그룹으로부터 선택되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 방법.

[1430] 91. 제 82 항 내지 제 90 항 중 어느 한 항에 있어서, 관심 생성물이 스피노신, 스피노사드, 스피네토람, 제니스테인, 콜린 옥시다제, 쿠마미딘 화합물, 에리스로마이신, 이버멕틴 아글리콘, HMG-CoA 환원효소 억제제, 카복실산 이성질체, 알파-메틸 메티오닌, 티알리신, 알파-케토부티레이트, 아스파테이트 하이드록사메이트, 아자세린, 5-플루오로인돌, 베타-하이드록시노르발린, 세룰레닌, 퓨린, 피리미딘 및 이의 유사체로 이루어진 그룹으로부터 선택되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 방법.

[1431] 92. 제 91 항에 있어서, 스피노신이 스피노신 A, 스피노신 D, 스피노신 J, 스피노신 L 또는 이의 조합인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 방법.

[1432] 93. 제 82 항 내지 제 92 항 중 어느 한 항에 있어서, RBS 래더가 서열번호 97 내지 127로부터 선택된 뉴클레오티드 서열을 갖는 적어도 하나의 RBS를 포함하는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량하기 위한 방법.

[1433] **트랜스포존 돌연변이유발 사카로폴리스포라 균주 라이브러리의 생성 및 생산 사카로폴리스포라 균주의 표현형 성능 개량을 위한 이의 이용**

[1434] 94. 다음을 포함하는 트랜스포존 돌연변이유발 사카로폴리스포라 균주 다양성 라이브러리를 생성하는 방법:

[1435] a) 트랜스포존을 하나 이상의 기본 사카로폴리스포라 균주의 세포의 집단에 도입하는 단계; 및

[1436] b) 랜덤하게 통합된 트랜스포존을 포함하는 사카로폴리스포라 균주를 선택함으로써, 상기 복수의 개별 사카로폴리스포라 균주의 각각의 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 사카로폴리스포라 균주를 포함하는 초기 사카로폴리스포라 균주 라이브러리를 생성하는 단계, 여기서 상기 각각의 독특한 유전자 변이는 하나 이상의 랜덤하게 통합된 트랜스포존을 포함한다.

[1437] 95. 제 94 항에 있어서, c) 기본 사카로폴리스포라 균주의 표현형 성능과 비교하여 측정된 표현형 변수에서 적어도 1 증가를 나타내는 서브서열 사카로폴리스포라 균주 라이브러리를 선택하는 단계를 추가로 포함하는 것인 방법.

[1438] 96. 제 94 항 내지 제 95 항 중 어느 한 항에 있어서, 트랜스포존이 트랜스포존의 사카로폴리스포라 균주의 게놈 내로의 생체 내(*in vivo*) 전위를 가능하게 하는 트랜스포존 및 트랜스포사제 단백질(transposase 단백질)의 복합체를 사용하여 기본 사카로폴리스포라 균주에 도입되는 것인 방법.

[1439] 97. 제 94 항 내지 제 96 항 중 어느 한 항에 있어서, 트랜스포사제 단백질이 EZ-Tn5 트랜스포좀 시스템으로부터 유래되는 것인 방법.

[1440] 98. 제 94 항 내지 제 97 항 중 어느 한 항에 있어서, 트랜스포존은 기능상실(LoF) 트랜스포존 또는 기능획득(GoF) 트랜스포존인 방법.

[1441] 99. 제 94 항 내지 제 98 항 중 어느 한 항에 있어서, GoF 트랜스포존이 용해도 태그, 프로모터 및/또는 카운터선택 마커를 포함하는 것인 방법.

[1442] 100. 다음 단계를 포함하는 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법:

[1443] a. 트랜스포존 돌연변이유발에 의해 기본 사카로폴리스포라 균주의 게놈을 조작하여, 상기 복수의 개별 사카로폴리스포라 균주의 각각의 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 사카로폴리스포라 균주를 포함하는 초기 트랜스포존 돌연변이유발 사카로폴리스포라 균주 라이브러리를 생성하는 단계, 여기서 상기 독특한 유전자 변이의 각각은 하나 이상의 트랜스포존을 포함한다;

[1444] b. 기준 사카로폴리스포라 균주에 대한 표현형 성능 개량을 위한 초기 트랜스포존 돌연변이유발 사카로폴리스포라 균주 라이브러리의 개별 사카로폴리스포라 균주를 선별 및 선택하여, 표현형 성능 개량을 부여하는 독특한 유전자 변이를 확인하는 단계;

[1445] c. 선행 단계에서 선별된 적어도 2개의 개별 사카로폴리스포라 균주에 존재하는 유전자 변이로부터 독특한 유전

자 변이의 조합을 각각 포함하는 후속 복수의 사카로폴리스포라 균주를 제공하여, 후속 트랜스포존 돌연변이유발 사카로폴리스포라 균주 라이브러리를 생성하는 단계;

[1446] d. 기준 사카로폴리스포라 균주에 대한 표현형 성능 개량을 위한 후속 트랜스포존 돌연변이유발 사카로폴리스포라 균주 라이브러리의 개별 사카로폴리스포라 균주를 선별 및 선택하여, 추가 표현형 성능 개량을 부여하는 유전자 변이의 독특한 조합을 확인하는 단계; 및

[1447] e. 사카로폴리스포라 균주가 생산 사카로폴리스포라 균주의 표현형 성능에 비해 원하는 수준의 개량된 표현형 성능을 나타낼 때까지 선형 또는 비선형 방식으로 단계 c)-d)를 1회 이상 반복하는 단계, 여기서 각각의 후속 반복은 미생물 균주의 새로운 트랜스포존 돌연변이유발 사카로폴리스포라 균주 라이브러리를 생성하며, 여기서 새로운 라이브러리의 각각의 균주는 선행 라이브러리의 적어도 2개의 개별 사카로폴리스포라 균주 중에서 선택된 유전자 변이의 조합인 유전자 변이를 포함한다.

[1448] 101. 제 100 항에 있어서, 후속 트랜스포존 돌연변이유발 사카로폴리스포라 균주 라이브러리는 초기 트랜스포존 돌연변이유발 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1449] 102. 제 100 항 내지 제 101 항 중 어느 한 항에 있어서, 후속 트랜스포존 돌연변이유발 사카로폴리스포라 균주 라이브러리는 초기 트랜스포존 돌연변이유발 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리의 서브세트인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1450] 103. 제 100 항 내지 제 102 항 중 어느 한 항에 있어서, 후속 트랜스포존 돌연변이유발 사카로폴리스포라 균주 라이브러리는 선행 트랜스포존 돌연변이유발 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1451] 104. 제 100 항 내지 제 103 항 중 어느 한 항에 있어서, 후속 트랜스포존 돌연변이유발 사카로폴리스포라 균주 라이브러리는 선행 트랜스포존 돌연변이유발 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리의 서브세트인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1452] 105. 제 100 항 내지 제 104 항 중 어느 한 항에 있어서, 후속 트랜스포존 돌연변이유발 사카로폴리스포라 균주 라이브러리의 사카로폴리스포라 균주의 표현형 성능이 생산 사카로폴리스포라 균주의 표현형 성능과 비교하여 측정된 표현형 변수에서 적어도 10% 증가를 나타낼 때까지 단계 c)-d)가 반복되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1453] 106. 제 100 항 내지 제 105 항 중 어느 한 항에 있어서, 후속 트랜스포존 돌연변이유발 사카로폴리스포라 균주 라이브러리의 사카로폴리스포라 균주의 표현형 성능이 생산 사카로폴리스포라 균주의 표현형 성능과 비교하여 측정된 표현형 변수에서 적어도 1배 증가를 나타낼 때까지 단계 c)-d)가 반복되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1454] 107. 제 100 항 내지 제 106 항 중 어느 한 항에 있어서, 단계 e)의 개량된 표현형 성능이 관심 생성물의 부피 생산성, 관심 생성물의 비 생산성, 관심 생성물의 수율, 관심 생성물의 역가 및 이의 조합으로 이루어진 그룹으로부터 선택되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1455] 108. 제 100 항 내지 제 107 항 중 어느 한 항에 있어서, 단계 e)의 개량된 표현형 성능이 관심 생성물의 증가된 또는 보다 효율적인 생산이며, 상기 관심 생성물은 소분자, 효소, 펩타이드, 아미노산, 유기산, 합성 화합물, 연료, 알코올, 1차 세포외 대사산물, 2차 세포외 대사산물, 세포내 성분 분자 및 이의 조합으로 이루어진 그룹으로부터 선택되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1456] 109. 제 108 항에 있어서, 관심 생성물이 스피노신, 스피노사드, 스피네토람, 제니스테인, 콜린 옥시다제, 쿠마미딘 화합물, 에리스로마이신, 이버멕틴 아글리콘, HMG-CoA 환원효소 억제제, 카복실산 이성질체, 알파-메틸 메티오닌, 티알리신, 알파-케토부티레이트, 아스파테이트 하이드록사메이트, 아자세린, 5-플루오로인돌, 베타-하이드록시노르발린, 세룰레닌, 퓨린, 피리미딘 및 이의 유사체로 이루어진 그룹으로부터 선택되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1457] 110. 제 109 항에 있어서, 스피노신이 스피노신 A, 스피노신 D, 스피노신 J, 스피노신 L 또는 이의 조합인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1458] 111. 제 100 항 내지 제 110 항 중 어느 한 항에 있어서, 트랜스포존이 기능상실(LoF) 트랜스포존 또는 기능획

득(GoF) 트랜스포존을 포함하는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1459] 112. 제 111 항에 있어서, GoF 트랜스포존이 용해도 태그, 프로모터 및/또는 카운터 선택 마커를 포함하는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1460] **항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리의 생성 및 생산 사카로폴리스포라 균주의 표현형 성능 개량을 위한 이의 이용**

[1461] 113. 다음 단계를 포함하는 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리를 생성하는 방법:

[1462] a) 소정의 대사산물 및/또는 발효 생성물에 저항성인 사카로폴리스포라 균주를 선택하여, 상기 복수의 개별 사카로폴리스포라 균주의 각각의 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 사카로폴리스포라 균주를 포함하는 초기 사카로폴리스포라 균주 라이브러리를 생성하는 단계, 여기서 상기 독특한 유전자 변이의 적어도 하나는 소정의 대사산물 및/또는 발효 생성물에 대한 저항성을 초래한다; 및

[1463] b) 소정의 대사산물 및/또는 발효 생성물에 저항성인 사카로폴리스포라 균주를 수집하여 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리를 생성하는 단계.

[1464] 114. 제 113 항에 있어서, 소정의 대사산물 및/또는 발효 생성물이 스피노신 합성 경로에 관련된 분자, SAM/메티오닌 경로에 관련된 분자, 라이신 생산 경로에 관련된 분자, 트립토판 경로에 관련된 분자, 트레오닌 경로에 관련된 분자, 아세틸-CoA 생산 경로에 관련된 분자, 및 드-노보(de-novo) 또는 회수(salvage) 퓨린 및 피리미딘 경로에 관련된 분자로 이루어진 그룹으로부터 선택되는 것인 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리를 생성하는 방법.

[1465] 115. 제 113 항 내지 제 114 항 중 어느 한 항에 있어서, 1) 스피노신 합성 경로에 관련된 분자는 스피노신이고, 및 선택적으로 각각의 균주는 약 50ug/ml 내지 약 2mg/ml의 스피노신 J/L에 저항성이고;

[1466] 2) SAM/메티오닌 경로에 관련된 분자는 알파-메틸 메티오닌(aMM) 또는 노르류신이고, 및 선택적으로 각각의 균주는 약 1mM 내지 약 5mM 알파-메틸 메티오닌(aMM)에 저항성이고;

[1467] 3) 라이신 생산 경로에 관련된 분자는 티알리신 또는 알파-케토부티레이트 및 아스파테이트 하이드록사메이트의 혼합물이고;

[1468] 4) 트립토판 경로에 관련된 분자는 아자세린 또는 5-플루오로인돌이고;

[1469] 5) 트레오닌 경로에 관련된 분자는 베타-하이드록시노르발린이고;

[1470] 6) 아세틸-CoA 생산 경로에 관련된 분자는 세룰레닌이고; 및

[1471] 7) 드-노보 또는 회수 퓨린 및 피리미딘 경로에 관련된 분자는 퓨린 또는 피리미딘 유사체인 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리를 생성하는 방법.

[1472] 116. 제 113 항 내지 제 115 항 중 어느 한 항에 있어서, b) 기본 사카로폴리스포라 균주의 표현형 성능과 비교하여 측정된 표현형 변수에서 적어도 1 증가를 나타내는 서브서열 사카로폴리스포라 균주 라이브러리를 선택하는 단계를 추가로 포함하는 것인 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리를 생성하는 방법.

[1473] 117. 제 116 항에 있어서, 서브서열 사카로폴리스포라 균주 라이브러리의 각각의 균주가 스피노신의 증가된 합성을 나타내는 것인 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리를 생성하는 방법.

[1474] 118. 다음 단계를 포함하는 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법:

[1475] a) 복수의 개별 사카로폴리스포라 균주의 각각의 균주 내에서 발견된 독특한 유전자 변이를 갖는 복수의 개별 사카로폴리스포라 균주를 포함하는 초기 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리를 제공하는 단계, 여기서 유전자 변이 각각은 하나 이상의 유전자 변이, 여기서 유전자 변이는 소정의 대사산물 또는 발효 생성물에 대한 저항성을 부여한다;

[1476] b) 기준 사카로폴리스포라 균주에 대한 표현형 성능 개량을 위한 초기 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리의 개별 사카로폴리스포라 균주를 선별 및 선택하여, 표현형 성능 개량을 부여하는 독특한 유전자 변이를 확인하는 단계;

[1477] c) 선행 단계에서 선별된 적어도 2개의 개별 사카로폴리스포라 균주에 존재하는 유전자 변이로부터 독특한 유전자 변이의 조합을 각각 포함하는 후속 복수의 사카로폴리스포라 균주를 제공하여, 후속 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리를 생성하는 단계;

[1478] d) 기준 사카로폴리스포라 균주에 대한 표현형 성능 개량을 위한 후속 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리의 개별 사카로폴리스포라 균주를 선별 및 선택하여, 추가 표현형 성능 개량을 부여하는 유전자 변이의 독특한 조합을 확인하는 단계; 및

[1479] e) 사카로폴리스포라 균주가 생산 사카로폴리스포라 균주의 표현형 성능에 비해 원하는 수준의 개량된 표현형 성능을 나타낼 때까지 선형 또는 비선형 방식으로 단계 c)-d)를 1회 이상 반복하는 단계, 여기서 각각의 후속 반복은 미생물 균주의 새로운 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리를 생성하며, 여기서 새로운 라이브러리의 각각의 균주는 선행 라이브러리의 적어도 2개의 개별 사카로폴리스포라 균주 중에서 선택된 유전자 변이의 조합인 유전자 변이를 포함한다.

[1480] 119. 제 118 항에 있어서, 후속 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리는 초기 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1481] 120. 제 118 항 내지 제 119 항 중 어느 한 항에 있어서, 후속 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리는 초기 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리의 서브세트인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1482] 121. 제 118 항 내지 제 120 항 중 어느 한 항에 있어서, 후속 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리는 선행 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1483] 122. 제 118 항 내지 제 122 항 중 어느 한 항에 있어서, 후속 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리는 선행 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리의 완전한 조합 라이브러리의 서브세트인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1484] 123. 제 118 항 내지 제 122 항 중 어느 한 항에 있어서, 후속 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리의 사카로폴리스포라 균주의 표현형 성능이 생산 사카로폴리스포라 균주의 표현형 성능과 비교하여 측정된 표현형 변수에서 적어도 10% 증가를 나타낼 때까지 단계 c)-d)가 반복되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1485] 124. 제 118 항 내지 제 123 항 중 어느 한 항에 있어서, 후속 항-대사산물/발효 생성물 저항성 사카로폴리스포라 균주 라이브러리의 사카로폴리스포라 균주의 표현형 성능이 생산 사카로폴리스포라 균주의 표현형 성능과 비교하여 측정된 표현형 변수에서 적어도 1배 증가를 나타낼 때까지 단계 c)-d)가 반복되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1486] 125. 제 118 항 내지 제 124 항 중 어느 한 항에 있어서, 단계 e)의 개량된 표현형 성능이 관심 생성물의 부피 생산성, 관심 생성물의 비 생산성, 관심 생성물의 수율, 관심 생성물의 역가 및 이의 조합으로 이루어진 그룹으로부터 선택되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1487] 126. 제 125 항에 있어서, 단계 e)의 개량된 표현형 성능이 관심 생성물의 증가된 또는 보다 효율적인 생산이며, 상기 관심 생성물은 소분자, 효소, 펩타이드, 아미노산, 유기산, 합성 화합물, 연료, 알코올, 1차 세포외 대사산물, 2차 세포외 대사산물, 세포내 성분 분자 및 이의 조합으로 이루어진 그룹으로부터 선택되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1488] 127. 제 126 항에 있어서, 관심 생성물이 스피노신, 스피노사드, 스피네토람, 제니스테인, 콜린 옥시다제, 쿠마미딘 화합물, 에리스로마이신, 이버멕틴 아글리콘, HMG-CoA 환원효소 억제제, 카복실산 이성질체, 알파-메틸 메티오닌, 티알리신, 알파-케토부티레이트, 아스파테이트 하이드록사메이트, 아자세린, 5-플루오로인돌, 베타-하이드록시노르발린, 세룰레닌, 퓨린, 피리미딘 및 이의 유사체로 이루어진 그룹으로부터 선택되는 것인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1489] 128. 제 127 항에 있어서, 스피노신이 스피노신 A, 스피노신 D, 스피노신 J, 스피노신 L 또는 이의 조합인 생산 사카로폴리스포라 균주의 표현형 성능을 개량시키는 방법.

[1490] **사카로폴리스포라 숙주 세포 및 균주 라이브러리**

[1491] 129. 숙주 세포의 내인성 유전자에 작동 가능하게 연결된 프로모터를 포함하는 사카로폴리스포라 숙주 세포로, 여기서 프로모터는 내인성 유전자와 이종이며, 프로모터는 서열번호 1 내지 69로 이루어진 그룹으로부터 선택된 서열을 가지는 것인 사카로폴리스포라 숙주 세포.

[1492] 130. 제 129 항에 있어서, 내인성 유전자가 사카로폴리스포라 숙주 세포에서 스피노신의 합성에 관련된 것인 사카로폴리스포라 숙주 세포.

[1493] 131. 제 129 항 내지 제 130 항 중 어느 한 항에 있어서, 사카로폴리스포라 숙주 세포가 내인성 유전자에 작동 가능하게 연결된 프로모터가 없는 기준 사카로폴리스포라 균주의 표현형 성능과 비교하여 원하는 수준의 개량된 표현형 성능을 갖는 것인 사카로폴리스포라 숙주 세포.

[1494] 132. 사카로폴리스포라 균주 라이브러리로, 여기서 라이브러리의 각각의 사카로폴리스포라 균주는 숙주 세포의 내인성 유전자에 작동 가능하게 연결된 프로모터를 포함하고, 프로모터는 내인성 유전자와 이종이며, 프로모터는 서열번호 1 내지 69로 이루어진 그룹으로부터 선택된 서열을 갖는 것인 사카로폴리스포라 균주 라이브러리.

[1495] 133. 숙주 세포의 내인성 유전자에 연결된 터미네이터를 포함하는 사카로폴리스포라 숙주 세포로, 여기서 터미네이터는 내인성 유전자와 이종이며, 프로모터는 서열번호 70 내지 80으로 이루어진 그룹으로부터 선택된 서열을 갖는 것인 사카로폴리스포라 숙주 세포.

[1496] 134. 제 133 항에 있어서, 내인성 유전자가 사카로폴리스포라 숙주 세포에서 스피노신의 합성에 관련된 것인 사카로폴리스포라 숙주 세포.

[1497] 135. 제 133 항 내지 제 134 항 중 어느 한 항에 있어서, 사카로폴리스포라 숙주 세포가 내인성 유전자에 작동 가능하게 연결된 프로모터가 없는 기준 사카로폴리스포라 균주의 표현형 성능과 비교하여 원하는 수준의 개량된 표현형 성능을 갖는 것인 사카로폴리스포라 숙주 세포.

[1498] 136. 사카로폴리스포라 균주 라이브러리로, 여기서 라이브러리의 각각의 사카로폴리스포라 균주는 숙주 세포의 내인성 유전자에 연결된 터미네이터를 포함하고, 터미네이터는 내인성 유전자와 이종이며, 터미네이터는 서열번호 70 내지 80으로 이루어진 그룹으로부터 선택된 서열을 갖는 것인 사카로폴리스포라 균주 라이브러리.

[1499] 137. 숙주 세포의 내인성 유전자에 작동 가능하게 연결된 리보솜 결합 부위를 포함하는 사카로폴리스포라 숙주 세포로, 여기서 리보솜 결합 부위는 내인성 유전자와 이종이며, 리보솜 결합 부위는 서열번호 97 내지 127로 이루어진 그룹으로부터 선택된 서열을 갖는 것인 사카로폴리스포라 숙주 세포.

[1500] 138. 제 137 항에 있어서, 내인성 유전자가 사카로폴리스포라 숙주 세포에서 스피노신의 합성에 관련된 것인 사카로폴리스포라 숙주 세포.

[1501] 139. 제 133 항 내지 제 138 항 중 어느 한 항에 있어서, 사카로폴리스포라 숙주 세포가 내인성 유전자에 작동 가능하게 연결된 RBS 부위가 없는 기준 사카로폴리스포라 균주의 표현형 성능과 비교하여 원하는 수준의 개량된 표현형 성능을 갖는 것인 사카로폴리스포라 숙주 세포.

[1502] 140. 사카로폴리스포라 균주 라이브러리로, 여기서 라이브러리의 각각의 사카로폴리스포라 균주는 숙주 세포의 내인성 유전자에 작동 가능하게 연결된 리보솜 결합 부위를 포함하고, 리보솜 결합 부위는 내인성 유전자와 이종이며, 리보솜 결합 부위는 서열번호 97 내지 127로 이루어진 그룹으로부터 선택된 서열을 갖는 것인 사카로폴리스포라 균주 라이브러리.

[1503] 141. 트랜스포존을 포함하는 사카로폴리스포라 숙주 세포로, 여기서 사카로폴리스포라 숙주 세포는 트랜스포존이 없는 기준 사카로폴리스포라 균주의 표현형 성능과 비교하여 원하는 수준의 개량된 표현형 성능을 갖는 것인 사카로폴리스포라 숙주 세포.

[1504] 142. 제 141 항에 있어서, 트랜스포존이 기능상실(LoF) 트랜스포존 또는 기능획득(GoF) 트랜스포존인 사카로폴리스포라 숙주 세포.

[1505] 143. 제 142 항에 있어서, 기능획득(GoF) 트랜스포존이 프로모터, 카운터 선택 마커 및/또는 용해도 태그를 포함하는 것인 사카로폴리스포라 숙주 세포.

[1506] 144. 제 141 항 내지 제 143 항 중 어느 한 항에 있어서, 트랜스포존이 서열번호 128 내지 131로 이루어진 그룹으로부터 선택된 서열을 포함하는 것인 사카로폴리스포라 숙주 세포.

[1507] 145. 사카로폴리스포라 균주 라이브러리로, 여기서 라이브러리의 각각의 사카로폴리스포라 균주는 서열번호 128 내지 131로 이루어진 그룹으로부터 선택된 서열을 갖는 트랜스포존을 포함하고, 각각의 균주의 트랜스포존은 상이한 게놈 유전자좌에 있는 것인 사카로폴리스포라 균주 라이브러리.

[1508] 146. 사카로폴리스포라 균주 라이브러리로, 여기서 라이브러리의 각각의 사카로폴리스포라 균주는 다음에 대한 균주의 저항성을 초래하는 유전자 변이를 포함하는 것인 사카로폴리스포라 균주 라이브러리:

[1509] 1) 스피노신 합성 경로에 관련된 분자,

[1510] 2) SAM/메티오닌 경로에 관련된 분자,

[1511] 3) 라이신 생산 경로에 관련된 분자,

[1512] 4) 트립토판 경로에 관련된 분자,

[1513] 5) 트레오닌 경로에 관련된 분자,

[1514] 6) 아세틸-CoA 생산 경로에 관련된 분자, 및/또는

[1515] 7) 드-노보 또는 회수 퓨린 및 피리미딘 경로에 관련된 분자.

[1516] 147. 제 146 항에 있어서,

[1517] 1) 스피노신 합성 경로에 관련된 분자는 스피노신이고;

[1518] 2) SAM/메티오닌 경로에 관련된 분자는 알파-메틸 메티오닌(aMM) 또는 노르류신이고;

[1519] 3) 라이신 생산 경로에 관련된 분자는 티알리신 또는 알파-케토부티레이트 및 아스파테이트 하이드록사메이트의 혼합물이고;

[1520] 4) 트립토판 경로에 관련된 분자는 아자세린 또는 5-플루오로인돌이고;

[1521] 5) 트레오닌 경로에 관련된 분자는 베타-하이드록시노르발린이고;

[1522] 6) 아세틸-CoA 생산 경로에 관련된 분자는 세룰레닌이고; 및

[1523] 7) 드-노보 또는 회수 퓨린 및 피리미딘 경로에 관련된 분자는 퓨린 또는 피리미딘 유사체인 사카로폴리스포라 균주 라이브러리.

[1524] 148. 제 147 항에 있어서, 분자가 스피노신 J/L이고, 각각의 균주가 약 50ug/ml 내지 약 2mg/ml 스피노신 J/L에 저항성인 사카로폴리스포라 균주 라이브러리.

[1525] 149. 제 147 항에 있어서, 분자가 알파-메틸 메티오닌(aMM)이고, 각각의 균주가 약 1mM 내지 약 5mM aMM에 저항성인 사카로폴리스포라 균주 라이브러리.

[1526] 150. 리포터 유전자를 포함하는 사카로폴리스포라 균주로, 리포터 유전자는 다음으로 이루어진 그룹으로부터 선택되는 것인 사카로폴리스포라 균주:

[1527] a) 녹색 형광 리포터 단백질을 코딩하는 유전자, 선택적으로 유전자는 사카로폴리스포라에서의 발현에 최적화된 코돈이다;

[1528] b) 녹색 형광 리포터 단백질을 코딩하는 유전자, 선택적으로 유전자는 사카로폴리스포라에서의 발현에 최적화된 코돈이다; 및

[1529] c) 베타-글루쿠로니다제(gusA) 단백질을 코딩하는 유전자, 선택적으로 유전자는 사카로폴리스포라에서의 발현에 최적화된 코돈이다.

[1530] 151. 제 150 항에 있어서,

[1531] a) 녹색 형광 리포터 단백질은 서열번호 143의 아미노산 서열을 가지고;

[1532] b) 적색 형광 리포터 단백질은 서열번호 144의 아미노산 서열을 가지고; 및

[1533] c) gusA 단백질은 서열번호 145의 아미노산 서열을 가지는 것인 사카로폴리스포라 균주.

[1534] 152. 제 150 항에 있어서,

[1535] a) 녹색 형광 리포터 단백질을 코딩하는 유전자는 서열번호 81의 서열을 가지고;

[1536] b) 적색 형광 리포터 단백질을 코딩하는 유전자는 서열번호 82의 서열을 가지고; 및

[1537] c) gusA 단백질을 코딩하는 유전자는 서열번호 83의 서열을 가지는 것인 사카로폴리스포라 균주.

[1538] 153. 제 150 항 내지 제 153 항 중 어느 한 항에 있어서, 균주가 녹색 형광 리포터 단백질을 코딩하는 유전자 및 적색 형광 리포터 단백질을 코딩하는 유전자 둘 다를 포함하고, 녹색 형광 리포터 단백질 및 적색 형광 리포터 단백질의 형광 여기 및 방출 스펙트럼은 서로 구별되는 것인 사카로폴리스포라 균주.

[1539] 154. 제 150 항 내지 제 153 항 중 어느 한 항에 있어서, 균주가 녹색 형광 리포터 단백질을 코딩하는 유전자 및 적색 형광 리포터 단백질을 코딩하는 유전자 둘 다를 포함하고, 녹색 형광 리포터 단백질 및 적색 형광 리포터 단백질의 형광 여기 및 방출 스펙트럼은 사카로폴리스포라 균주의 내인성 형광과 구별되는 것인 사카로폴리스포라 균주.

[1540] 155. 사카로폴리스포라 균주의 게놈에서 하나 이상의 중성 통합 부위(neutral integration site)에 통합된 DNA 단편을 포함하는 사카로폴리스포라 균주로, 여기서 중성 통합 부위는 서열번호 132 내지 142로부터 선택된 서열을 갖는 게놈 단편, 또는 서열번호 132 내지 142 중 어느 하나와 상동성인 게놈 단편 내의 위치의 그룹으로부터 선택되는 것인 사카로폴리스포라 균주.

[1541] 156. 제 155 항에 있어서, 사카로폴리스포라 균주는 통합된 DNA 단편이 없는 기준 사카로폴리스포라 균주의 표현형 성능과 비교하여 원하는 수준의 개량된 표현형 성능을 갖는 것인 사카로폴리스포라 균주.

[1542] 157. 제 156 항에 있어서, 사카로폴리스포라 균주가 통합된 DNA 단편이 없는 기준 사카로폴리스포라 균주의 표현형 성능과 비교하여 원하는 수준의 개량된 스피노신 생산을 갖는 것인 사카로폴리스포라 균주.

[1543] 158. 제 155 항 내지 제 157 항 중 어느 한 항에 있어서, 통합된 DNA 단편이 리포터 단백질을 코딩하는 서열을 포함하는 것인 사카로폴리스포라 균주.

[1544] 159. 제 155 항 내지 제 158 항 중 어느 한 항에 있어서, 통합된 DNA 단편이 트랜스포존을 포함하는 것인 사카로폴리스포라 균주.

[1545] 160. 제 155 항 내지 제 159 항 중 어느 한 항에 있어서, 통합된 DNA 단편이 상응하는 인테그라제(integrase)에 의해 인식될 수 있는 부착 부위(attB)를 포함하는 것인 사카로폴리스포라 균주.

[1546] **사카로폴리스포라 균주에서 DNA 단편을 통합시키기 위한 중성 통합 부위(Neutral integration sites; NISs)**

[1547] 161. 사카로폴리스포라 균주의 게놈 내로 DNA 단편을 통합시키는 방법으로, 여기서 DNA 단편은 사카로폴리스포라 균주의 게놈의 중성 통합 부위에 통합되며, 중성 통합 부위는 서열번호 132 내지 142로부터 선택된 서열을 갖는 게놈 단편, 또는 서열번호 132 내지 142 중 어느 하나와 상동성인 게놈 단편 내의 위치의 그룹으로부터 선택되는 것인 사카로폴리스포라 균주의 게놈 내로 DNA 단편을 통합시키는 방법.

[1548] 162. 제 161 항에 있어서, DNA 단편이 상응하는 인테그라제에 의해 인식될 수 있는 부착 부위(attB)를 포함하는 것인 사카로폴리스포라 균주의 게놈 내로 DNA 단편을 통합시키는 방법.

[1549] 163. 다음 단계를 포함하는 적어도 2개의 부모 사카로폴리스포라 균주로부터 유래된 유전자 돌연변이를 신속하게 통합시키는 방법:

[1550] (1) 적어도 2개의 부모 사카로폴리스포라 균주를 제공하는 단계, 여기서 각각의 균주는 다른 균주에는 존재하지 않는 독특한 게놈 돌연변이를 포함한다;

[1551] (2) 각각의 부모 균주로부터 원형질체를 준비하는 단계;

[1552] (3) 부모 균주로부터의 원형질체를 융합시켜 2개의 부모 사카로폴리스포라 균주의 게놈을 포함하는 융합된 원형질체를 생성하는 단계, 여기서 각각의 부모 균주의 게놈 사이에 상동성 재조합이 일어난다;

[1553] (4) 단계 (3)에서 생성된 융합된 원형질체로부터 사카로폴리스포라 세포를 회수하는 단계; 및

[1554] (5) 제 1 부모 사카로폴리스포라 균주의 독특한 게놈 돌연변이를 포함하는 사카로폴리스포라 세포를 선택하는 단계; 및

[1555] (6) 단계 (5)에서 얻어진 사카로폴리스포라 세포를 제 2 부모 균주의 독특한 게놈 돌연변이의 존재에 대해 유전

자형 분석(genotyping)하는 단계,

[1556] 이에 의해 2개의 부모 사카로폴리스포라 균주로부터 유래된 독특한 게놈 돌연변이를 포함하는 새로운 사카로폴리스포라 균주를 수득한다.

[1557] 164. 제 163 항에 있어서, 독특한 게놈 돌연변이 중 하나는 선택 가능한 마커에 연결되고, 다른 독특한 게놈 돌연변이는 임의의 선택 가능한 마커에 연결되지 않는 것인 방법.

[1558] 165. 제 164 항에 있어서, 단계 (3)에서, 선택 가능한 마커에 연결된 독특한 게놈 돌연변이를 원래 함유하는 균주의 원형질체:선택 가능한 마커에 연결되지 않은 독특한 게놈 돌연변이를 원래 함유하는 균주의 원형질체의 비는 1:1 미만인 방법.

[1559] 166. 제 165 항에 있어서, 비가 약 1:10 내지 약 1:100 이하인 방법.

[1560] 167. 제 163 항 내지 제 166 항 중 어느 한 항에 있어서, 단계 (4)에서, 원형질체 세포가 한천 오버레이를 사용하지 않고 삼투적으로 안정화된 매질 상에 플레이팅되는 것인 방법.

[1561] 168. 제 163 항 내지 제 167 항 중 어느 한 항에 있어서, 단계 (5)는 독특한 게놈 돌연변이 중 하나가 선택 약물에 저항성을 초래하는 선택 가능한 마커에 연결되는 경우, 적절한 선택 약물 항생제를 성장 세포 상에 오버레이함으로써 달성되는 것인 방법.

[1562] 169. 제 163 항 내지 제 168 항 중 어느 한 항에 있어서, 단계 (5)는 독특한 게놈 돌연변이 중 어느 것도 선택 가능한 마커에 연결되지 않은 경우, 유전자형 분석에 의해 달성되는 것인 방법.

[1563] 170. 제 163 항 내지 제 170 항 중 어느 한 항에 있어서, 2개 초과의 균주로부터 유래된 유전자 돌연변이가 단일 통합 공정 동안 랜덤하게 통합되는 것인 방법.

[1564] 171. 제 163 항 내지 제 171 항 중 어느 한 항에 있어서, 단계 (2)에서 원형질체를 약 5000xg의 속도로 약 5분 동안 원심분리하여 초기에 수집하는 것인 방법.

[1565] 172. 제 163 항 내지 제 172 항 중 어느 한 항에 있어서, 방법이 면모(cotton wool)를 통해 원형질체를 여과하는 것을 포함하지 않는 것인 방법.

[1566] 173. 제 163 항 내지 제 173 항 중 어느 한 항에 있어서, 융합된 원형질체가 탑-한천(top-agar)이 아닌 R2YE 매질에서 회수되는 것인 방법.

[1567] 174. 제 173 항에 있어서, R2YE 매질이 0.5M 소르비톨 및 0.5M 만노스를 포함하는 것인 방법.

[1568] **사카로폴리스포라 균주에서의 표적화된 게놈 편집**

[1569] 175. 다음을 포함하는 사카로폴리스포라 균주에서 표적화된 게놈 편집 방법:

[1570] a) 선택 마커, 카운터 선택 마커, 편집될 사카로폴리스포라 균주의 게놈 유전자좌와 상동성을 갖는 DNA 단편, 및 플라스미드 백본 서열을 포함하는 플라스미드를 기본 사카로폴리스포라 균주에 도입하는 단계;

[1571] b) 게놈에서 선택 마커의 존재에 기초하여 통합 이벤트를 갖는 사카로폴리스포라 균주를 선택하는 단계;

[1572] c) 카운터 선택 마커 유전자의 부재에 기초하여 루프 아웃된 플라스미드 백본을 가진 사카로폴리스포라 균주를 선택하는 단계, 여기서 카운터 선택 마커는 sacB 유전자 또는 pheS 유전자이다.

[1573] 176. 제 175 항에 있어서, 편집된 게놈을 갖는 생성된 사카로폴리스포라 균주가 편집이 없는 부모 균주에 비해 더 우수한 성능을 갖는 것인 방법.

[1574] 177. 제 176 항에 있어서, 생성된 사카로폴리스포라 균주는 편집이 없는 부모 균주에 비해 증가된 스피노신 생산을 가지는 것인 방법.

[1575] 178. 제 175 항 내지 제 177 항 중 어느 한 항에 있어서, sacB 유전자가 사카로폴리스포라 스피노사(*Saccharopolyspora spinosa*)에 대해 코돈-최적화되는 것인 방법.

[1576] 179. 제 178 항에 있어서, sacB 유전자가 서열번호 146에 의해 코딩된 아미노산 서열에 대해 90% 서열 동일성을 갖는 아미노산 서열을 코딩하는 것인 방법.

[1577] 180. 제 175 항 내지 제 177 항 중 어느 한 항에 있어서, pheS 유전자가 사카로폴리스포라 스피노사에 대해 코

돈-최적화되는 것인 방법.

[1578] 181. 제 180 항에 있어서, pheS 유전자가 서열번호 147 또는 서열번호 148에 의해 코딩된 아미노산 서열에 대해 90% 서열 동일성을 갖는 아미노산 서열을 코딩하는 것인 방법.

[1579] **컨쥬게이션을 이용한 공여자(donor) 미생물 세포로부터 사카로폴리스포라 미생물의 수용자(recipient) 세포로의 유전자 물질의 전달**

[1580] 182. 다음을 포함하는 공여자(donor) 미생물 세포로부터 사카로폴리스포라 미생물의 수용자(recipient) 세포로 유전자 물질을 전달하는 방법:

[1581] 1) 선택적으로, 수용자 세포를 후기-대수(late-exponential) 또는 정지(stationary) 상으로 계대배양하는 단계;

[1582] 2) 선택적으로, 공여자 세포를 중기-대수(mid-exponential) 상으로 계대배양하는 단계;

[1583] 3) 공여자 및 수용자 세포를 결합하는 단계;

[1584] 4) 컨쥬게이션 매질 상에 공여자 및 수용자 세포 혼합물을 플레이팅하는 단계;

[1585] 5) 세포를 컨쥬게이션시키기 위해 플레이트를 배양하는 단계;

[1586] 6) 공여자 세포에 대한 항생제 선택을 적용하는 단계;

[1587] 7) 비-통합 수용자 세포에 대한 항생제 선택을 적용하는 단계; 및

[1588] 8) 통합된 수용자 세포의 성장을 허용하기 위해 플레이트를 추가로 배양하는 단계.

[1589] 183. 제 182 항에 있어서, 공여자 미생물 세포가 대장균 세포인 방법.

[1590] 184. 제 182 항 내지 제 183 항 중 어느 한 항에 있어서, 다음 조건 중 적어도 2개, 3개, 4개, 5개, 6개, 7개 또는 그 이상이 이용되는 것인 방법:

[1591] 1) 수용자 세포는 컨쥬게이션 전에 세척된다;

[1592] 2) 공여자 세포 및 수용자 세포는 약 30℃의 온도에서 컨쥬게이션된다;

[1593] 3) 수용자 세포를 컨쥬게이션하기 전에 적어도 약 48시간 동안 계대배양한다;

[1594] 4) 컨쥬게이션을 위한 공여자 세포:수용자 세포의 비는 약 1:0.6 내지 1:1.0이다;

[1595] 5) 공여자 세포에 대한 선택을 위한 항생제 약물은 공여자 세포와 수용자 세포가 혼합된 후 약 15 내지 24시간 후에 혼합물로 전달된다;

[1596] 6) 수용자 세포에 대한 선택을 위한 항생제 약물은 공여자 세포 및 수용자 세포가 혼합된 후 약 40 내지 48시간 후에 혼합물에 전달된다;

[1597] 7) 공여자 및 수용자 세포 혼합물로 플레이팅된 컨쥬게이션 매질을 적어도 약 3시간 내지 10시간 동안 건조시킨다;

[1598] 8) 컨쥬게이션 매질은 적어도 약 3g/L 글루코스를 포함한다;

[1599] 9) 공여자 세포의 농도는 약 OD600=0.1 내지 0.6이다;

[1600] 10) 수용자 세포의 농도는 약 OD540=5.0 내지 15.0이다.

[1601] 185. 제 184 항에 있어서, 공여자 세포에 대한 선택을 위한 항생제 약물이 날리딕산(nalidixic)이고, 농도가 약 50 내지 약 150㎍/ml인 방법.

[1602] 186. 제 185 항에 있어서, 공여자 세포에 대한 선택을 위한 항생제 약물이 날리딕산이고, 농도가 약 100㎍/ml인 방법.

[1603] 187. 제 184 항에 있어서, 수용자 세포에 대한 선택을 위한 항생제가 아프라마이신(apramycin)이고, 농도가 약 50 내지 약 250㎍/ml인 방법.

[1604] 188. 제 187 항에 있어서, 수용자 세포에 대한 선택을 위한 항생제가 아프라마이신이고, 농도가 약 100㎍/ml인 방법.

[1605] 189. 제 182 항 내지 제 188 항 중 어느 한 항에 있어서, 방법은 고 처리량 방법에서 수행되는 것인 방법.

[1606] 190. 제 189 항에 있어서, 방법은 48-웰 Q-트레이에서 수행되는 것인 방법.

[1607] 191. 제 189 항에 있어서, 고 처리량 방법은 자동화되는 것인 방법.

[1608] 192. 제 191 항에 있어서, 공여자 세포 및 수용자 세포의 혼합물이 액체 혼합물이고, 액체 혼합물의 충분한 부피가 요동 운동(rocking motion)을 하는 배지 상에 플레이팅되고, 액체 혼합물이 배지의 전체 영역에 걸쳐 분산되는 것인 방법.

[1609] 193. 제 191 항에 있어서, 방법은 공여자 세포에 의해 제공된 통합된 DNA를 갖는 수용자 세포의 후속 접종을 위해 효모 핀을 사용하는 콜로니 피킹에 의해 컨쥬게이션 완료체(exconjugant)를 전달하는 자동화된 방법을 포함하는 것인 방법.

[1610] 194. 제 193 항에 있어서, 콜로니 피킹은 침지 운동(dipping motion) 또는 교반 운동(stirring motion)에서 수행되는 것인 방법.

[1611] 195. 제 184 항 내지 제 194 항 중 어느 한 항에 있어서, 컨쥬게이션 매질이 약 3-10g/L 글루코스를 포함하는 변형된 ISP4 매질인 방법.

[1612] 196. 제 184 항 내지 제 194 항 중 어느 한 항에 있어서, 혼합물에서 공여자 세포 또는 수용자 세포의 총 개수가 약 $5x10^6$ 내지 약 $9X10^6$인 방법.

[1613] 197. 제 182 항 내지 제 196 항 중 어느 한 항에 있어서, 방법은 다음 조건 중 적어도 4개를 사용하여 수행되는 것인 방법:

[1614] 1) 수용자 세포는 컨쥬게이션 전에 세척된다;

[1615] 2) 공여자 세포 및 수용자 세포는 약 30℃의 온도에서 컨쥬게이션된다;

[1616] 3) 수용자 세포를 컨쥬게이션하기 전에 적어도 약 48시간 동안 계대배양한다;

[1617] 4) 컨쥬게이션을 위한 공여자 세포:수용자 세포의 비는 약 1:0.8이다;

[1618] 5) 공여자 세포에 대한 선택을 위한 항생제 약물은 공여자 세포와 수용자 세포가 혼합된 후 약 20시간 후에 혼합물에 전달된다;

[1619] 6) 혼합물에서 공여자 세포의 양 또는 수용자 세포의 양은 약 $7x10^6$이다; 및

[1620] 7) 컨쥬게이션 매질은 약 6g/L 글루코스를 포함한다.

[1621] **사카로폴리스포라 균주에서 표적화된 게놈 편집의 상처없는(scarless) 방법**

[1622] 198. 다음을 포함하는 표적화된 게놈 유전자좌에서 유전자 변이를 함유하는 상처없는(scarless) 사카로폴리스포라 균주를 초래하는 사카로폴리스포라 균주에서 표적화된 게놈 편집 방법:

[1623] a) 플라스미드를 사카로폴리스포라 균주 내로 도입하는 단계로, 상기 플라스미드는 다음을 포함한다:

[1624] i. 선택 마커,

[1625] ii. 카운터 선택 마커,

[1626] iii. 표적 유전자좌에서 사카로폴리스포라 게놈으로 통합될 유전자 변이를 함유하는 DNA 단편, 상기 DNA 단편은 원하는 유전자 변이의 측면에 있는(flanking) 표적 게놈 유전자좌에 상동성 암(homology arm)을 가진다, 및

[1627] iv. 플라스미드 백본 서열;

[1628] b) 초기 상동성 재조합을 겪고 게놈에서 선택 마커의 존재에 기초하여 표적 유전자좌에 통합된 유전자 변이를 갖는 사카로폴리스포라 균주를 선택하는 단계; 및

[1629] c) 표적 유전자좌에 통합된 유전자 변이를 갖지만 카운터 선택 마커의 부재에 기초하여 플라스미드 백본을 루프아웃시키는 추가적인 상동성 재조합을 겪은 사카로폴리스포라 균주를 선택하는 단계,

[1630] 여기서 상기 표적화된 게놈 유전자좌는 코딩 DNA 모듈의 반복 세그먼트를 포함하지 않는 게놈 영역을 포함하는 사카로폴리스포라 게놈의 임의의 영역을 포함할 수 있다.

[1631] 199. 제 198 항에 있어서, 플라스미드가 온도 민감성 레플리콘을 포함하지 않는 것인 방법.

[1632] 200. 제 198 항 내지 제 199 항 중 어느 한 항에 있어서, 플라스미드가 복제 기점(origin of replication)을 포함하지 않는 것인 방법.

[1633] 201. 제 198 항 내지 제 200 항 중 어느 한 항에 있어서, 선택 단계 c)가 통합된 플라스미드의 복제 없이 수행되는 것인 방법.

[1634] 202. 제 198 항 내지 제 201 항 중 어느 한 항에 있어서, 플라스미드가 단일 상동성 재조합 벡터인 방법.

[1635] 203. 제 198 항 내지 제 202 항 중 어느 한 항에 있어서, 플라스미드가 이중 상동성 재조합 벡터인 방법.

[1636] 204. 제 198 항 내지 제 203 항 중 어느 한 항에 있어서, 카운터 선택 마커가 sacB 유전자 또는 pheS 유전자인 방법.

[1637] 205. 제 204 항에 있어서, sacB 유전자 또는 pheS 유전자가 사카로폴리스포라 스피노사에 대해 코돈-최적화되는 방법.

[1638] 206. 제 205 항에 있어서, sacB 유전자가 서열번호 146에 의해 코딩된 아미노산 서열과 90% 동일성을 갖는 아미노산 서열을 코딩하는 것인 방법.

[1639] 207. 제 205 항에 있어서, pheS 유전자가 서열번호 147 또는 서열번호 148에 의해 코딩된 아미노산 서열에 대해 90% 서열 동일성을 갖는 아미노산 서열을 코딩하는 것인 방법.

[1640] 208. 제 198 항 내지 제 207 항 중 어느 한 항에 있어서, 플라스미드가 형질전환에 의해 사카로폴리스포라 균주로 도입되는 것인 방법.

[1641] 209. 제 198 항 내지 제 208 항 중 어느 한 항에 있어서, 형질전환이 원형질체 형질전환인 방법.

[1642] 210. 제 198 항 내지 제 209 항 중 어느 한 항에 있어서, 플라스미드가 컨쥬게이션에 의해 사카로폴리스포라 균주로 도입되고, 사카로폴리스포라 균주는 수용자 세포이고, 플라스미드를 포함하는 공여자 세포는 플라스미드를 사카로폴리스포라 균주로 전달하는 것인 방법.

[1643] 211. 제 198 항 내지 제 210 항 중 어느 한 항에 있어서, 컨쥬게이션이 플라스미드를 포함하는 대장균 공여자 세포를 기초로 하는 것인 방법.

[1644] 212. 제 198 항 내지 제 211 항 중 어느 한 항에 있어서, 표적 유전자좌가 사카로폴리스포라 균주에서 관심 화합물의 생산과 관련된 유전자좌인 방법.

[1645] 213. 제 198 항 내지 제 212 항 중 어느 한 항에 있어서, 생성된 사카로폴리스포라 균주가 게놈 편집이 없는 대조군 균주에 비해 관심 화합물의 증가된 생산을 가지는 것인 방법.

[1646] 214. 제 212 항 또는 제 213 항에 있어서, 관심 화합물이 스피노신인 방법.

[1647] 215. 제 198 항 내지 제 214 항 중 어느 한 항에 있어서, 방법은 고 처리량 공정으로서 수행되는 것인 방법.

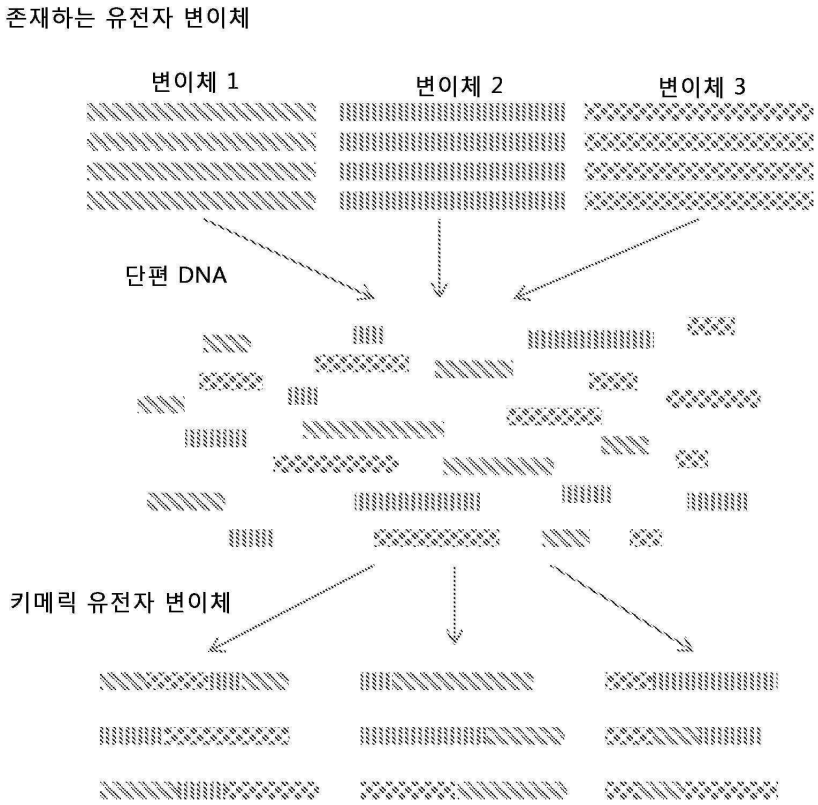
[1648] **참조에 의한 통합**

[1649] 본 발명에 인용된 모든 참고문헌, 기사, 간행물, 특허, 특허 간행물 및 특허 출원은 모든 목적을 위해 그 전문이 참조로 포함된다. 그러나 본 발명에 언급된 모든 참고문헌, 기사, 출판물, 특허, 특허 간행물 및 특허 출원은 세계 어느 나라에서든 유효한 선행 기술을 구성하거나 보통의 일반 지식의 일부를 구성한다는 인정 또는 어떠한 형태의 제안으로 간주되어서는 안 된다.

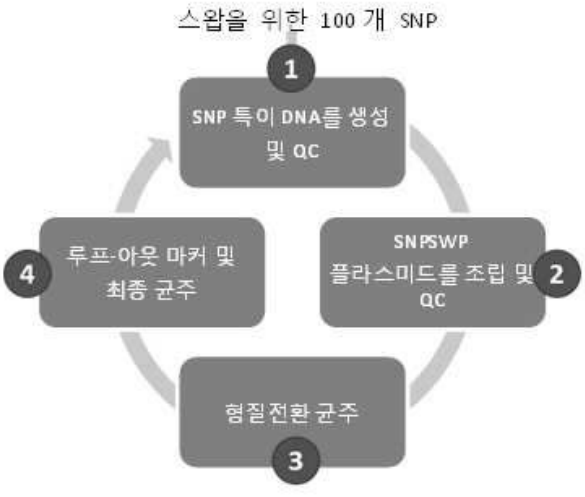
[1650] 또한, 2015년 12월 7일에 출원된 U.S. Provisional Application No. 62/264,232에 대한 우선권을 주장하는 2016년 12월 7일에 출원된 International Application No. PCT/US2016/065464, 2016년 4월 27일에 출원된 U.S. Nonprovisional Application No. 15/140,296, 2017년 4월 27일에 출원된 International Application No. PCT/US2017/29725, 2016년 12월 30일에 출원된 U.S. Nonprovisional Application No. 15/396,230 및 2016년 7월 29일에 출원된 U.S. Provisional Application No. 62/368,786은 모든 목적을 위해 모든 명세서, 참조, 도면 및 청구항을 포함하는 전체가 참조로서 본 명세서에 포함된다.

도면

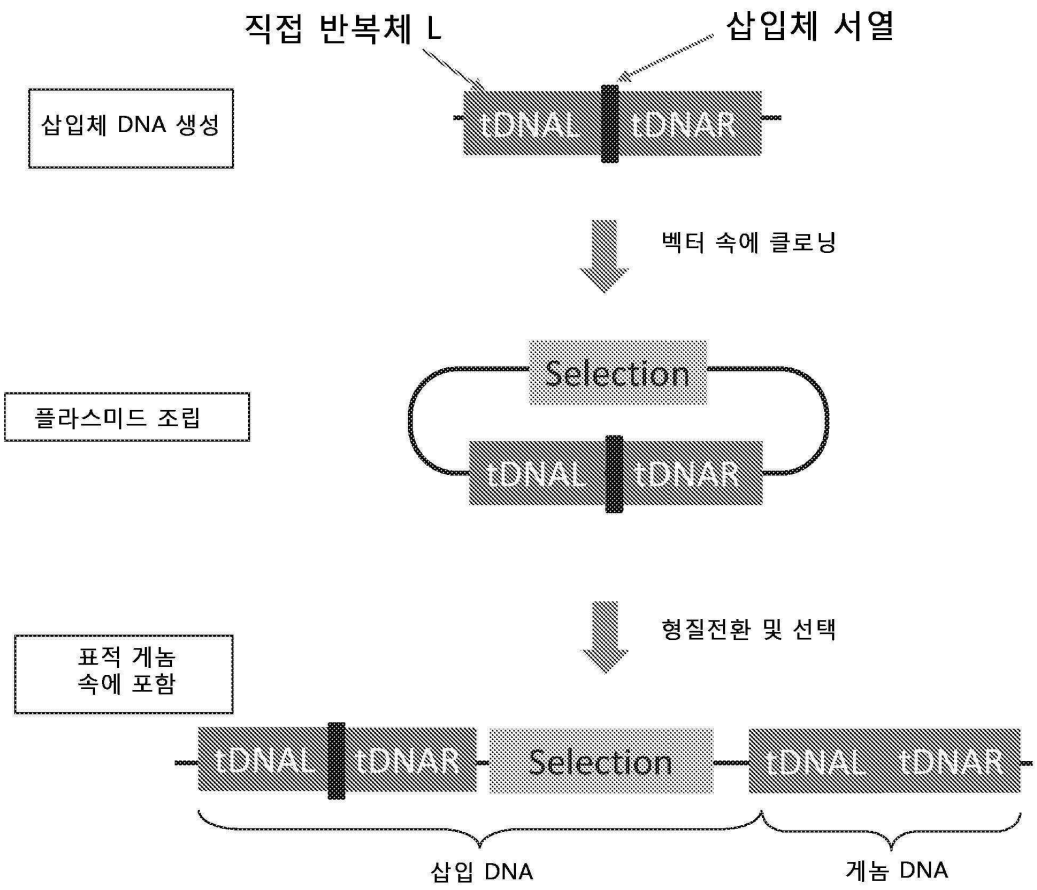
도면1


존재하는 유전자 변이체
변이체 1
변이체 2
변이체 3
단편 DNA
키메릭 유전자 변이체

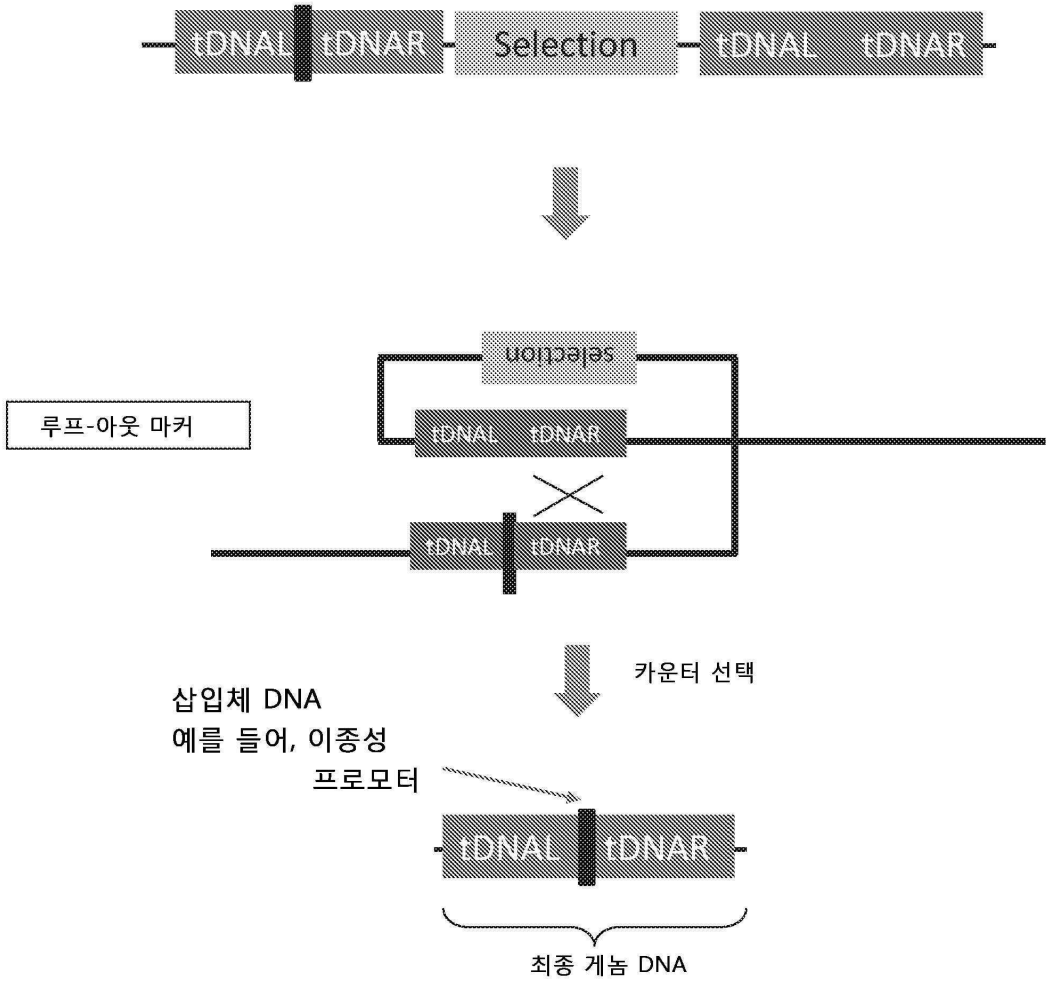
도면2


스왑을 위한 100 개 SNP
1
SNP 특이 DNA를 생성 및 QC
2
SNPSWP 플라스미드를 조립 및 QC
3
형질전환 균주
4
루프-아웃 마커 및 최종 균주

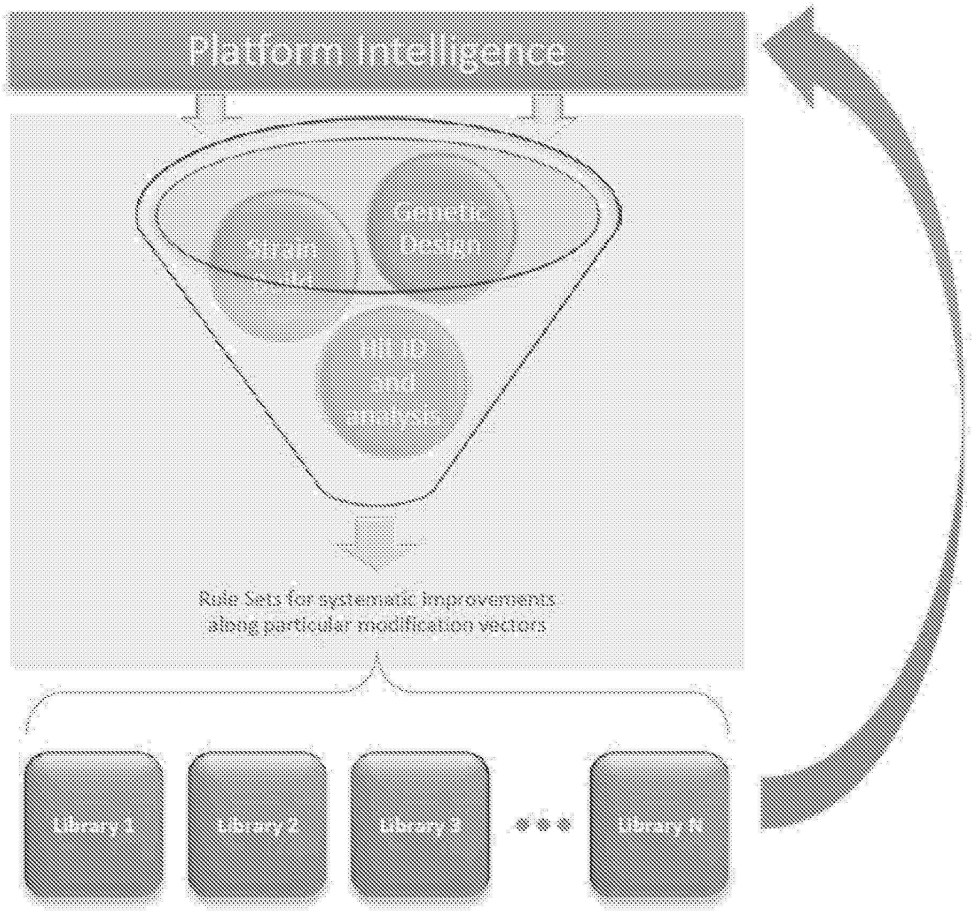
도면3


직접 반복체 L
삽입체 서열
삽입체 DNA 생성
tDNAL
tDNAR
벡터 속에 클로닝
Selection
플라스미드 조립
tDNAL
tDNAR
형질전환 및 선택
표적 게놈
속에 포함
tDNAL
tDNAR
Selection
tDNAL
tDNAR
삽입 DNA
게놈 DNA

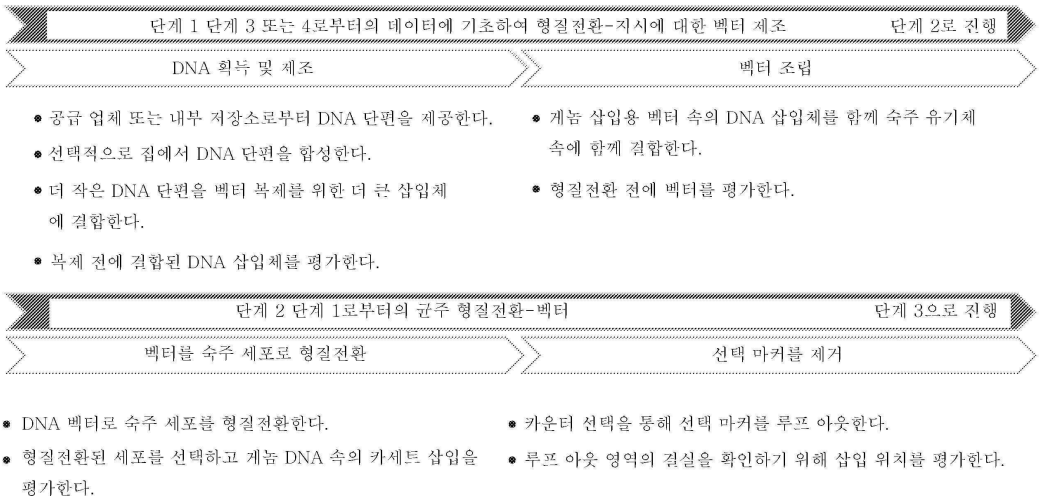
도면4


tDNAL
tDNAR
Selection
tDNAL
tDNAR
selection
루프-아웃 마커
tDNAL
tDNAR
tDNAL
tDNAR
카운터 선택
삽입체 DNA
예를 들어, 이종성
프로모터
tDNAL
tDNAR
최종 게놈 DNA

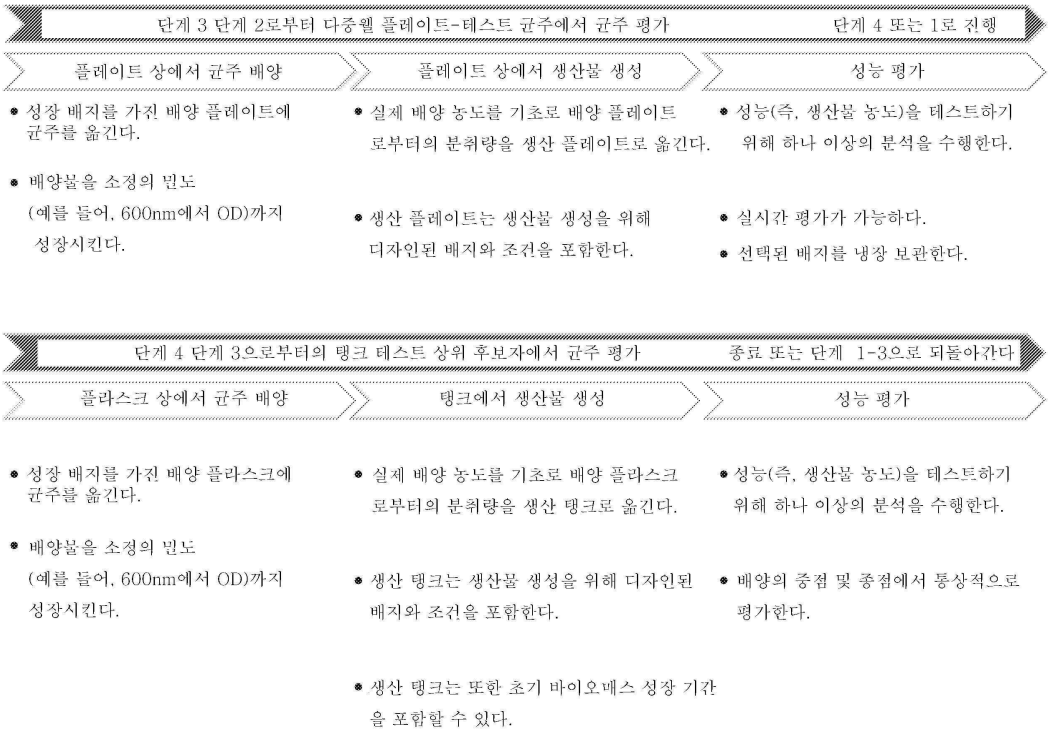
도면5


Platform Intelligence
Genetic Design
Rule Sets for systematic improvements along particular modification vectors
Library 1
Library 2
Library 3
Library N

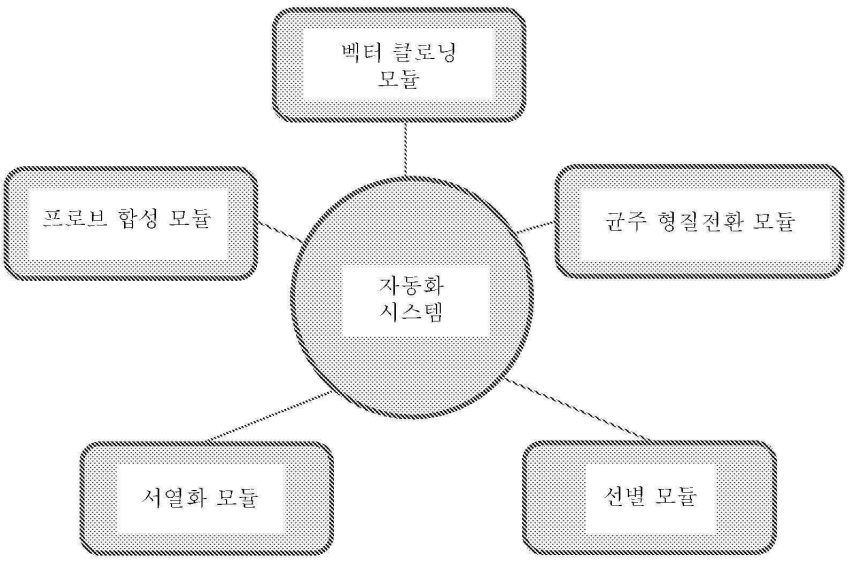
도면6a


단계 1 단계 3 또는 4로부터의 데이터에 기초하여 형질전환-지시에 대한 벡터 제조
단계 2로 진행
DNA 획득 및 제조
벡터 조립
공급 업체 또는 내부 저장소로부터 DNA 단편을 제공한다.
선택적으로 집에서 DNA 단편을 합성한다.
더 작은 DNA 단편을 벡터 복제를 위한 더 큰 삽입체에 결합한다.
복제 전에 결합된 DNA 삽입체를 평가한다.
게놈 삽입용 벡터 속의 DNA 삽입체를 함께 숙주 유기체 속에 함께 결합한다.
형질전환 전에 벡터를 평가한다.
단계 2 단계 1로부터의 균주 형질전환-벡터
단계 3으로 진행
벡터를 숙주 세포로 형질전환
선택 마커를 제거
DNA 벡터로 숙주 세포를 형질전환한다.
형질전환된 세포를 선택하고 게놈 DNA 속의 카세트 삽입을 평가한다.
카운터 선택을 통해 선택 마커를 루프 아웃한다.
루프 아웃 영역의 결실을 확인하기 위해 삽입 위치를 평가한다.

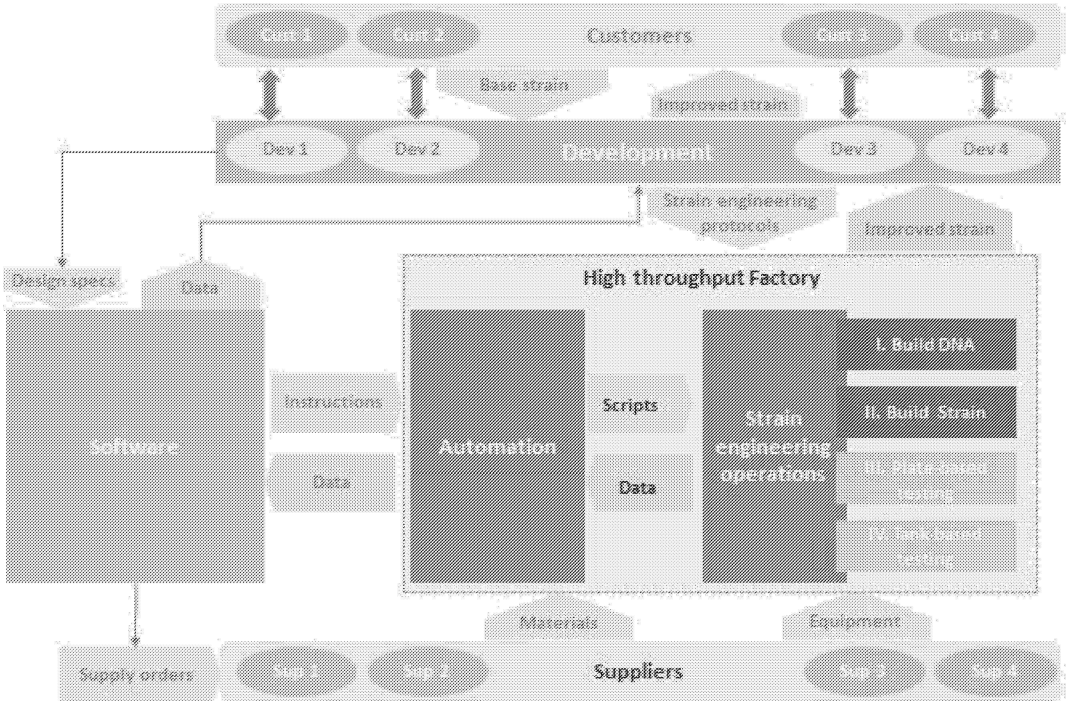
도면6b


단계 3 단계 2로부터 다중웰 플레이트-테스트 균주에서 균주 평가
단계 4 또는 1로 진행
플레이트 상에서 균주 배양
플레이트 상에서 생산물 생성
성능 평가
• 성장 배지를 가진 배양 플레이트에 균주를 옮긴다.
• 배양물을 소정의 밀도 (예를 들어, 600nm에서 OD)까지 성장시킨다.
• 실제 배양 농도를 기초로 배양 플레이트로부터의 분취량을 생산 플레이트로 옮긴다.
• 생산 플레이트는 생산물 생성을 위해 디자인된 배지와 조건을 포함한다.
• 성능(즉, 생산물 농도)을 테스트하기 위해 하나 이상의 분석을 수행한다.
• 실시간 평가가 가능하다.
• 선택된 배지를 냉장 보관한다.
단계 4 단계 3으로부터의 탱크 테스트 상위 후보자에서 균주 평가
종료 또는 단계 1-3으로 되돌아간다
플라스크 상에서 균주 배양
탱크에서 생산물 생성
성능 평가
• 성장 배지를 가진 배양 플라스크에 균주를 옮긴다.
• 배양물을 소정의 밀도 (예를 들어, 600nm에서 OD)까지 성장시킨다.
• 실제 배양 농도를 기초로 배양 플라스크로부터의 분취량을 생산 탱크로 옮긴다.
• 생산 탱크는 생산물 생성을 위해 디자인된 배지와 조건을 포함한다.
• 생산 탱크는 또한 초기 바이오매스 성장 기간을 포함할 수 있다.
• 성능(즉, 생산물 농도)을 테스트하기 위해 하나 이상의 분석을 수행한다.
• 배양의 중점 및 종점에서 통상적으로 평가한다.

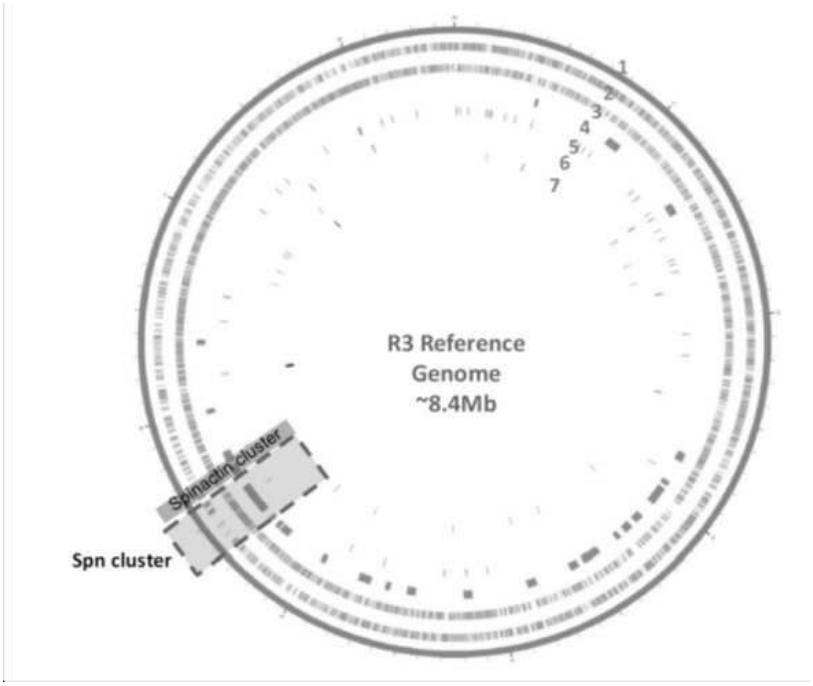
도면7


벡터 클로닝 모듈
프로브 합성 모듈
균주 형질전환 모듈
자동화 시스템
서열화 모듈
선별 모듈

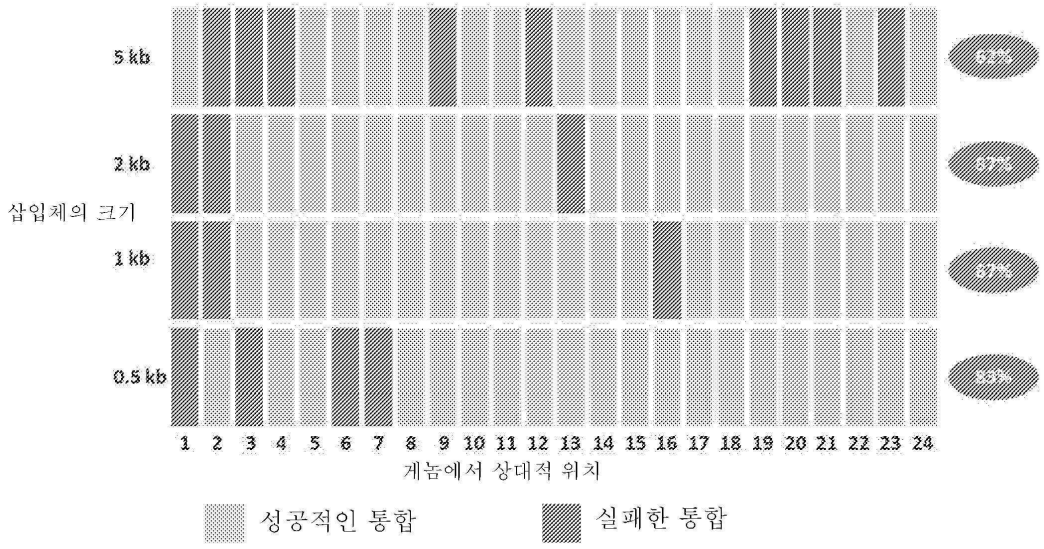
도면8


Cust 1
Cust 2
Customers
Cust 3
Cust 4
Base strain
Improved strain
Dev 1
Dev 2
Development
Dev 3
Dev 4
Strain engineering protocols
Improved strain
Design specs
Data
High throughput Factory
Software
Instructions
Data
Automation
Scripts
Data
Strain engineering operations
I. Build DNA
II. Build Strain
III. Plate-based testing
IV. Bank-based testing
Materials
Equipment
Supply orders
Sup 1
Sup 2
Suppliers
Sup 3
Sup 4

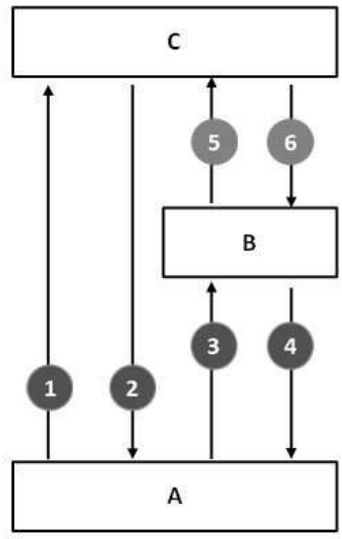
도면9


1
2
3
4
5
6
7
R3 Reference Genome ~8.4Mb
Spn cluster

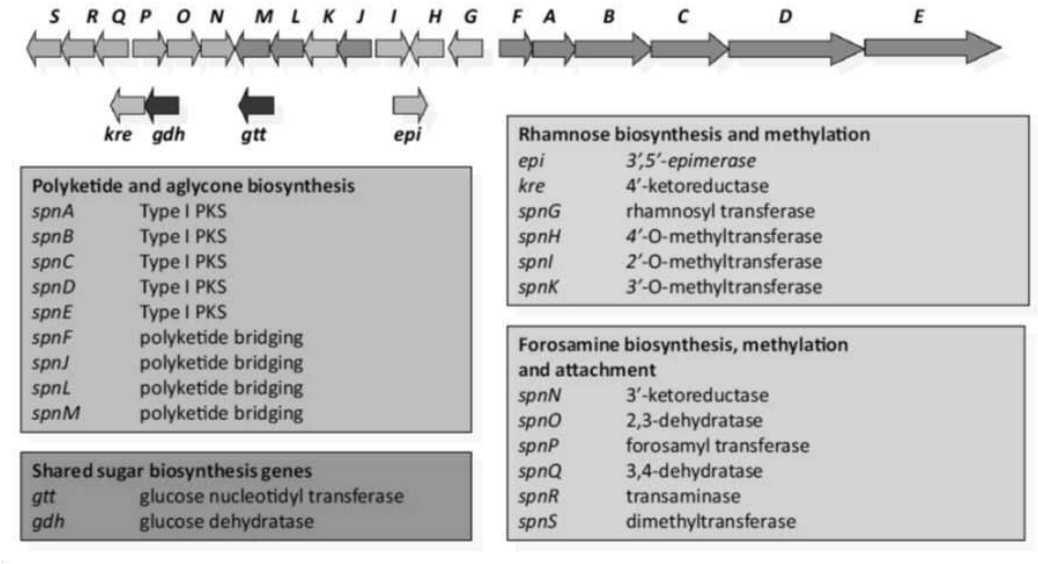
도면10


5 kb
2 kb
1 kb
0.5 kb
삽입체의 크기
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
게놈에서 상대적 위치
성공적인 통합
실패한 통합

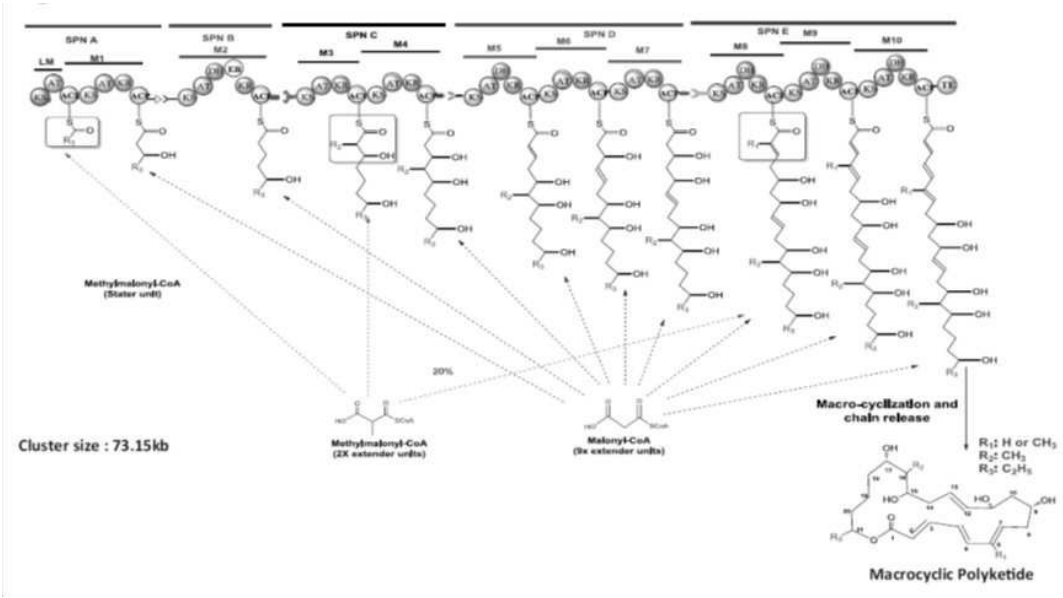
도면11


C
5
6
B
3
4
1
2
A

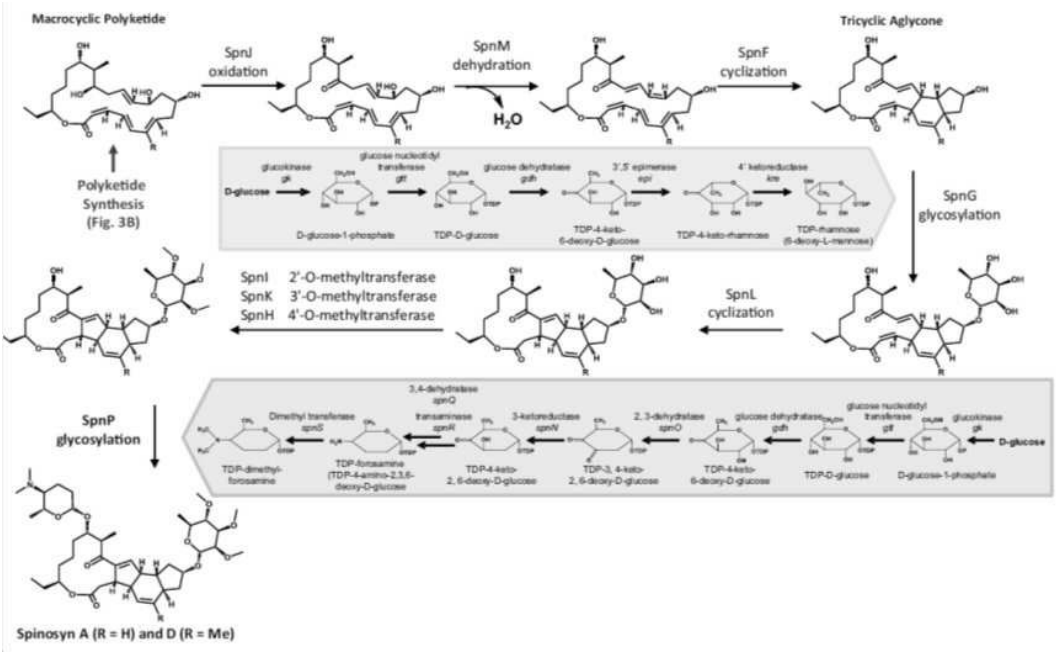
도면12a


S R Q P O N M L K J I H G F A B C D E
kre gdh gtt epi
Polyketide and aglycone biosynthesis
spnA Type I PKS
spnB Type I PKS
spnC Type I PKS
spnD Type I PKS
spnE Type I PKS
spnF polyketide bridging
spnJ polyketide bridging
spnL polyketide bridging
spnM polyketide bridging
Shared sugar biosynthesis genes
gtt glucose nucleotidyl transferase
gdh glucose dehydratase
Rhamnose biosynthesis and methylation
epi 3',5'-epimerase
kre 4'-ketoreductase
spnG rhamnosyl transferase
spnH 4'-O-methyltransferase
spnI 2'-O-methyltransferase
spnK 3'-O-methyltransferase
Forosamine biosynthesis, methylation and attachment
spnN 3'-ketoreductase
spnO 2,3-dehydratase
spnP forosamyl transferase
spnQ 3,4-dehydratase
spnR transaminase
spnS dimethyltransferase

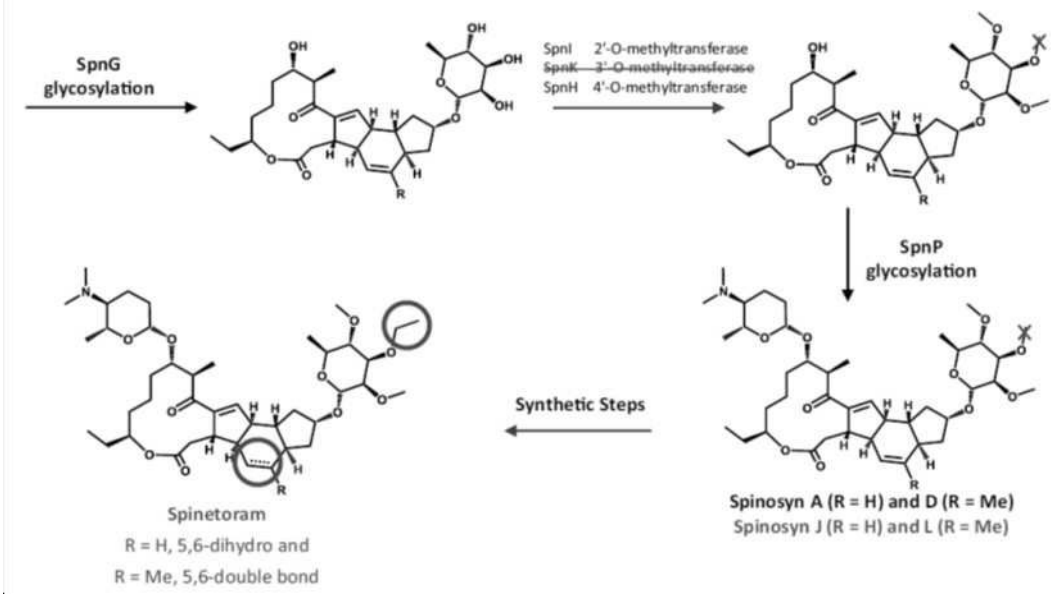
도면12b


SPN A
SPN B
SPN C
SPN D
SPN E
Methylmalonyl-CoA (Stater unit)
Cluster size : 73.15kb
Macro-cyclization and chain release
Macrocyclic Polyketide

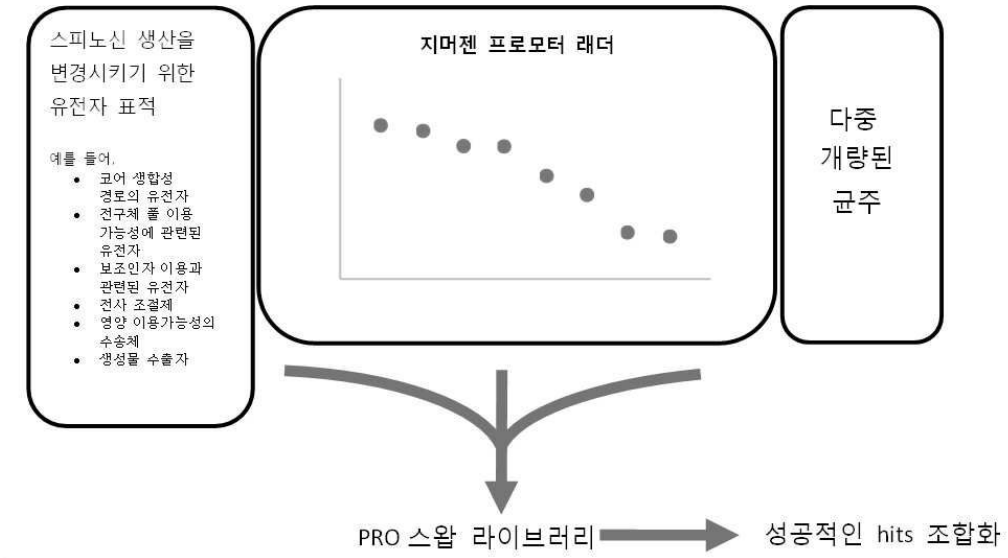
도면12c


Macrocyclic Polyketide
SpnJ oxidation
SpnM dehydration
H_2O
SpnF cyclization
Tricyclic Aglycone
Polyketide Synthesis (Fig. 3B)
D-glucose
D-glucose-1-phosphate
TDP-D-glucose
TDP-4-keto-6-deoxy-D-glucose
TDP-4-keto-rhamnose
TDP-rhamnose (6-deoxy-L-mannose)
SpnG glycosylation
SpnI 2'-O-methyltransferase
SpnK 3'-O-methyltransferase
SpnH 4'-O-methyltransferase
SpnL cyclization
SpnP glycosylation
TDP-dimethyl-forosamine
TDP-forosamine (TDP-4-amino-2,3,6-deoxy-D-glucose)
TDP-4-keto-2, 6-deoxy-D-glucose
TDP-3, 4-keto-2, 6-deoxy-D-glucose
TDP-4-keto-6-deoxy-D-glucose
TDP-D-glucose
D-glucose-1-phosphate
Spinosyn A (R = H) and D (R = Me)

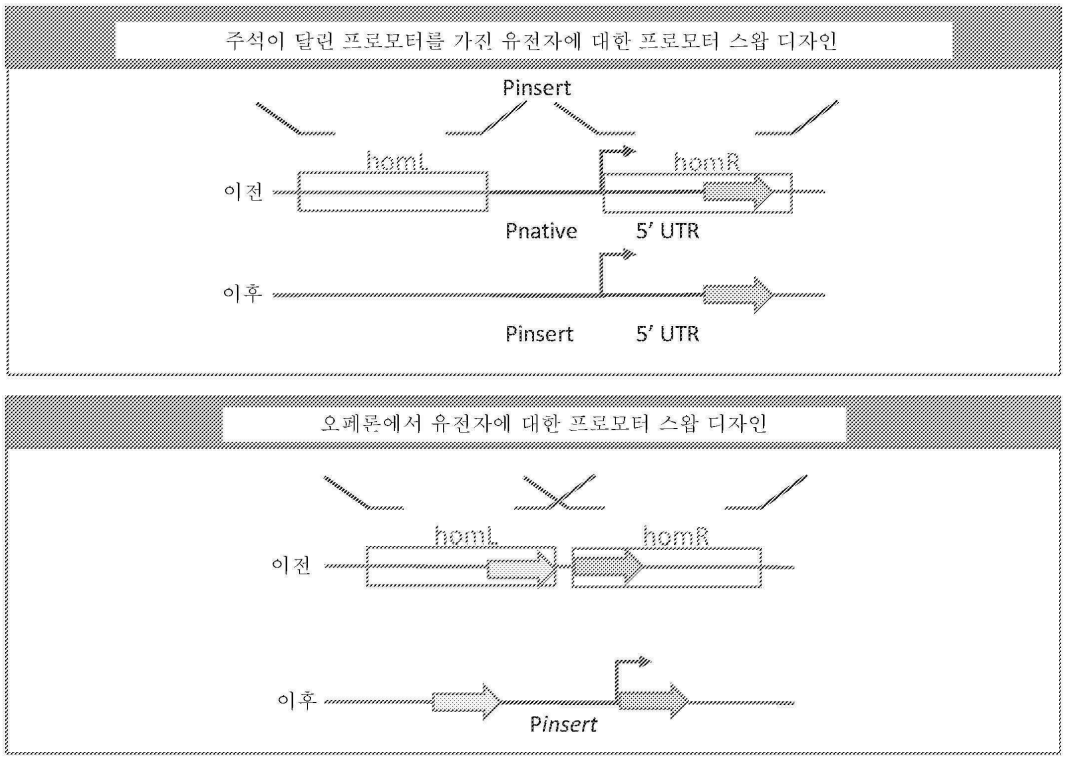
도면12d


SpnG glycosylation
SpnI 2'-O-methyltransferase
SpnK 3'-O-methyltransferase
SpnH 4'-O-methyltransferase
SpnP glycosylation
Synthetic Steps
Spinetoram
R = H, 5,6-dihydro and
R = Me, 5,6-double bond
Spinosyn A (R = H) and D (R = Me)
Spinosyn J (R = H) and L (R = Me)

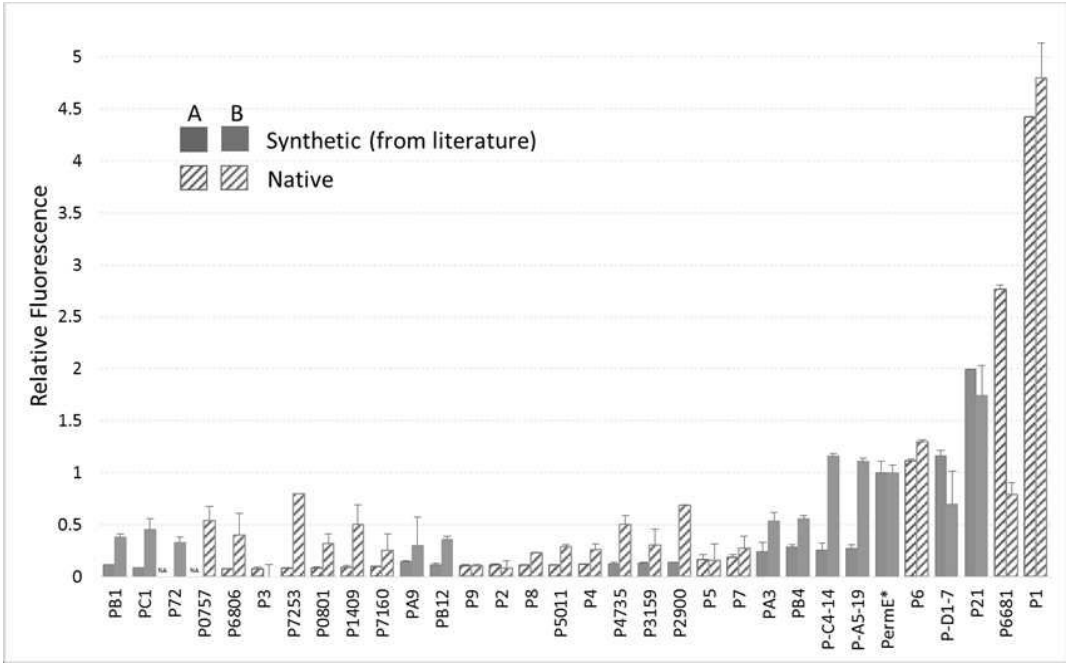
도면13


스피노신 생산을
변경시키기 위한
유전자 표적
예를 들어,
코어 생합성 경로의 유전자
전구체 풀 이용 가능성에 관련된 유전자
보조인자 이용과 관련된 유전자
전사 조절제
영양 이용가능성의 수송체
생성물 수출자
지머젠 프로모터 래더
다중
개량된
균주
PRO 스왑 라이브러리
성공적인 hits 조합화

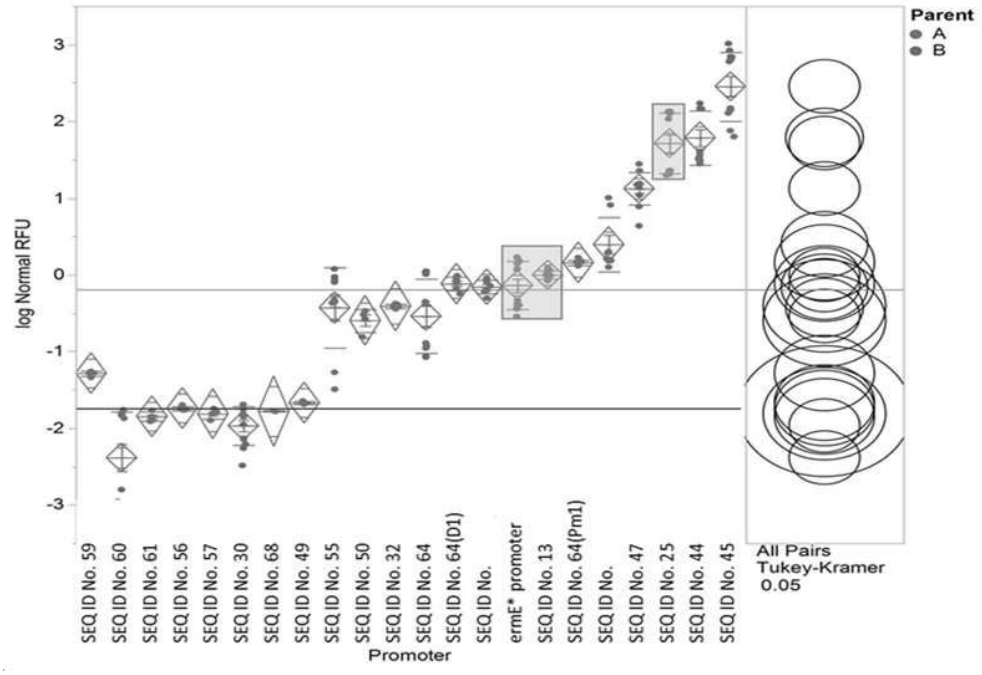
도면14


주석이 달린 프로모터를 가진 유전자에 대한 프로모터 스왑 디자인
Pinsert
이전
homL
homR
Pnative
5′ UTR
이후
Pinsert
5′ UTR
오페론에서 유전자에 대한 프로모터 스왑 디자인
이전
homL
homR
이후
Pinsert

도면15

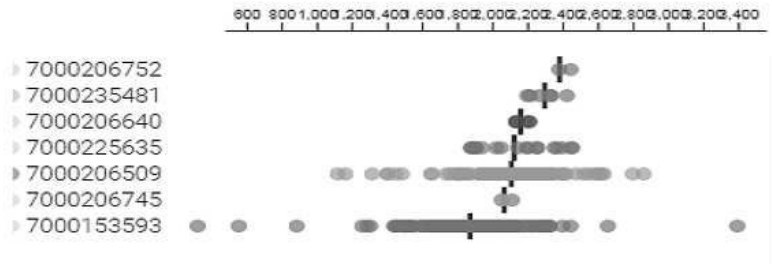

A B
Synthetic (from literature)
Native
Relative Fluorescence
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
PB1
PC1
P72
P0757
P6806
P3
P7253
P0801
P1409
P7160
PA9
PB12
P9
P2
P8
P5011
P4
P4735
P3159
P2900
P5
P7
PA3
PB4
P-C4-14
P-A5-19
PermE*
P6
P-D1-7
P21
P6681
P1

도면16


Parent
A
B
log Normal RFU
3
2
1
0
-1
-2
-3
SEQ ID No. 59
SEQ ID No. 60
SEQ ID No. 61
SEQ ID No. 56
SEQ ID No. 57
SEQ ID No. 30
SEQ ID No. 68
SEQ ID No. 49
SEQ ID No. 55
SEQ ID No. 50
SEQ ID No. 32
SEQ ID No. 64
SEQ ID No. 64(D1)
SEQ ID No.
ermE* promoter
SEQ ID No. 13
SEQ ID No. 64(Pm1)
SEQ ID No.
SEQ ID No. 47
SEQ ID No. 25
SEQ ID No. 44
SEQ ID No. 45
Promoter
All Pairs
Tukey-Kramer
0.05

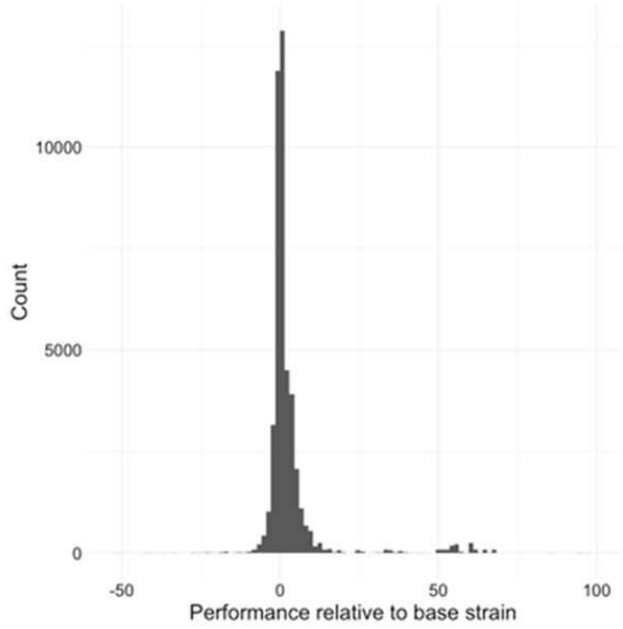
도면17

균주 성능 (스피노신 J+L 생산)

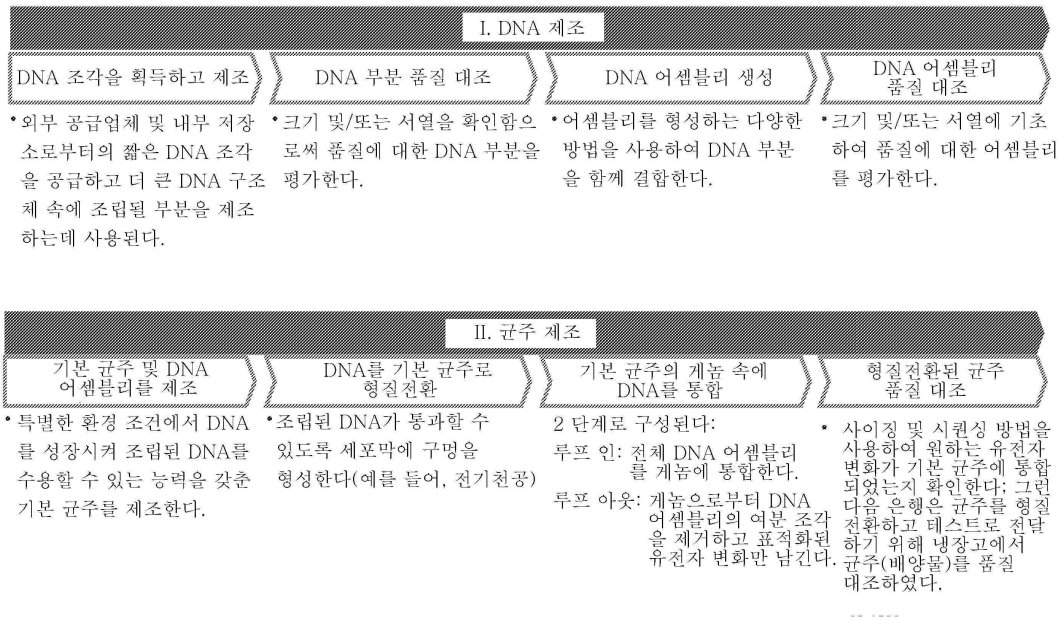


Strain ID	Gene function of PROSWP target
7000225635	1- core biosynthetic pathway
7000206640	2- precursor pools
7000206509	2- precursor pools
7000206745	3- cofactor utilization
7000206752	4- transcriptional regulator
7000235481	5- gene of unknown function
7000153593	Control (parent strain)

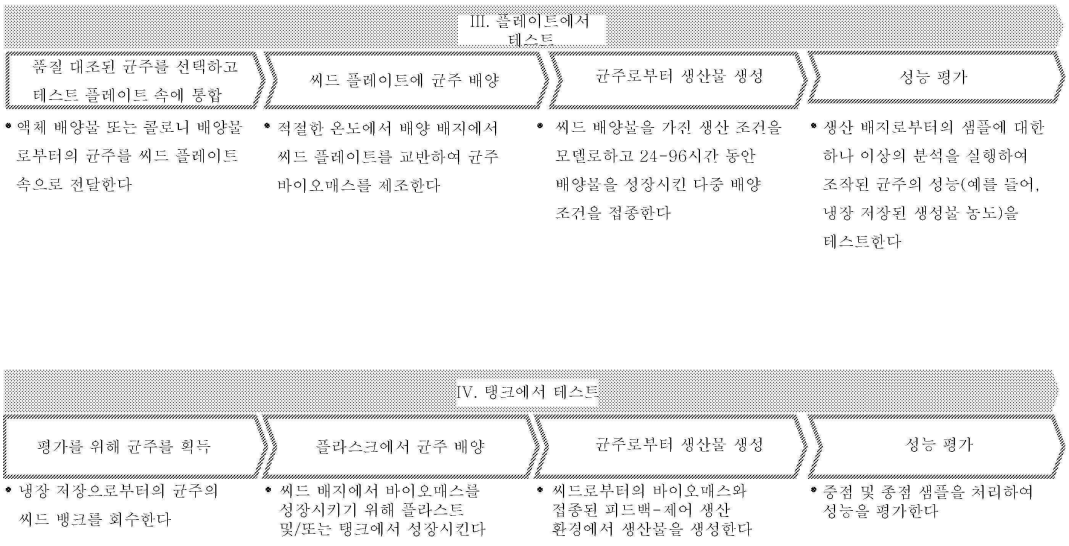
도면18



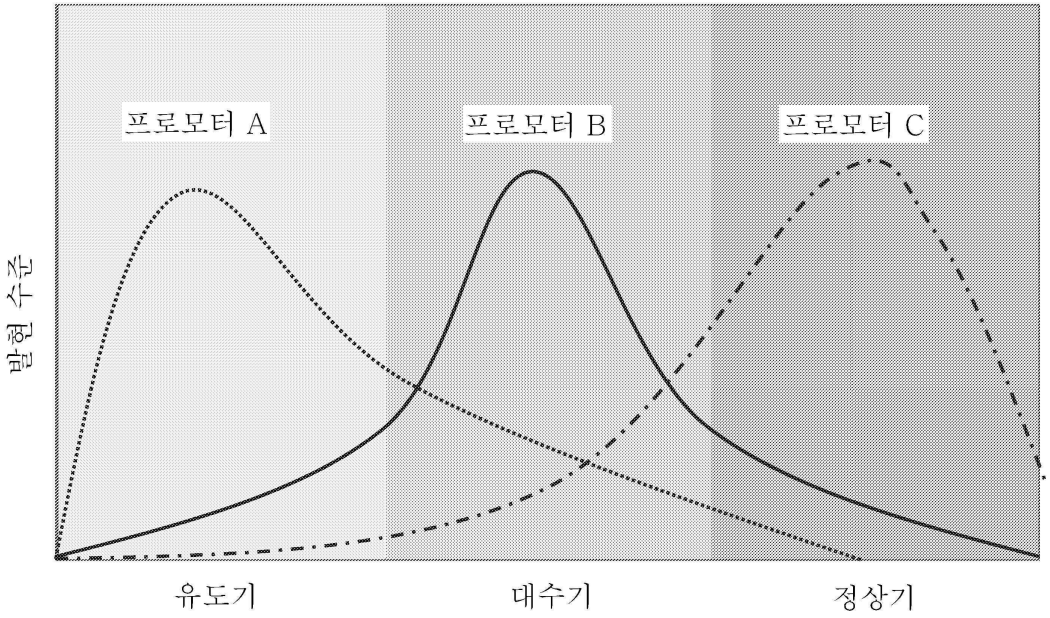
도면19


I. DNA 제조
DNA 조각을 획득하고 제조
DNA 부분 품질 대조
DNA 어셈블리 생성
DNA 어셈블리 품질 대조
• 외부 공급업체 및 내부 저장소로부터의 짧은 DNA 조각을 공급하고 더 큰 DNA 구조체 속에 조립될 부분을 제조하는데 사용된다.
• 크기 및/또는 서열을 확인함으로써 품질에 대한 DNA 부분을 평가한다.
• 어셈블리를 형성하는 다양한 방법을 사용하여 DNA 부분을 함께 결합한다.
• 크기 및/또는 서열에 기초하여 품질에 대한 어셈블리를 평가한다.
II. 균주 제조
기본 균주 및 DNA 어셈블리를 제조
DNA를 기본 균주로 형질전환
기본 균주의 게놈 속에 DNA를 통합
형질전환된 균주 품질 대조
• 특별한 환경 조건에서 DNA를 성장시켜 조립된 DNA를 수용할 수 있는 능력을 갖춘 기본 균주를 제조한다.
• 조립된 DNA가 통과할 수 있도록 세포막에 구멍을 형성한다(예를 들어, 전기천공)
2 단계로 구성된다:
루프 인: 전체 DNA 어셈블리를 게놈에 통합한다.
루프 아웃: 게놈으로부터 DNA 어셈블리의 여분 조각을 제거하고 표적화된 유전자 변화만 남긴다.
• 사이징 및 시퀀싱 방법을 사용하여 원하는 유전자 변화가 기본 균주에 통합되었는지 확인한다; 그런 다음 은행은 균주를 형질전환하고 테스트로 전달하기 위해 냉장고에서 균주(배양물)를 품질 대조하였다.

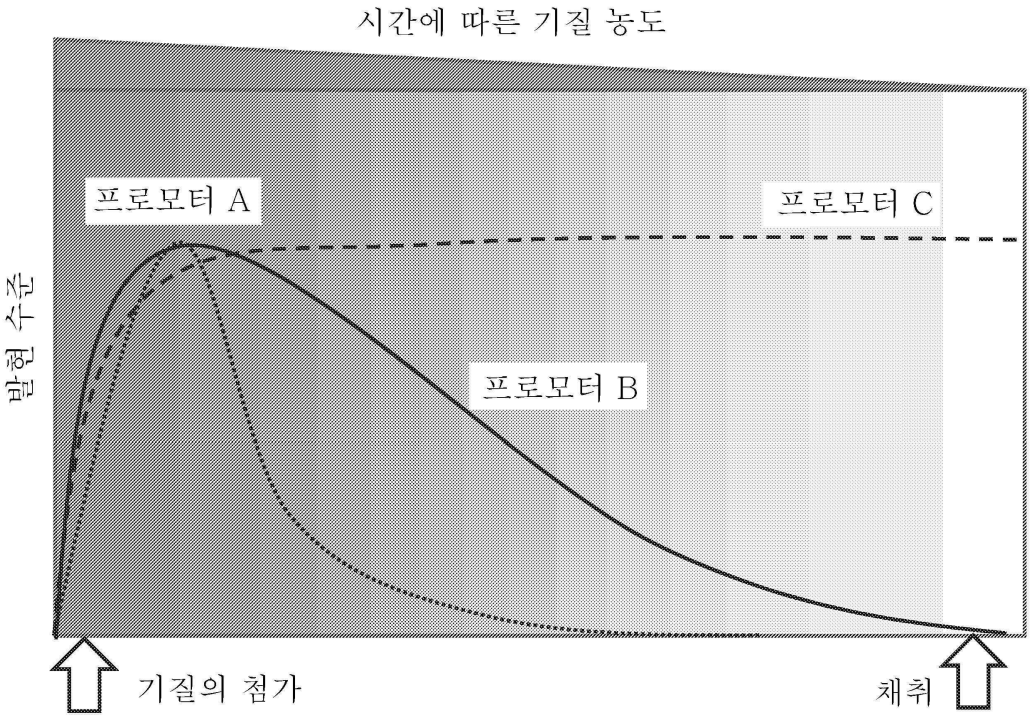
도면20


III. 플레이트에서 테스트
품질 대조된 균주를 선택하고 테스트 플레이트 속에 통합
씨드 플레이트에 균주 배양
균주로부터 생산물 생성
성능 평가
• 액체 배양물 또는 콜로니 배양물로부터의 균주를 씨드 플레이트 속으로 전달한다
• 적절한 온도에서 배양 배지에서 씨드 플레이트를 교반하여 균주 바이오매스를 제조한다
• 씨드 배양물을 가진 생산 조건을 모델로하고 24-96시간 동안 배양물을 성장시킨 다중 배양 조건을 접종한다
• 생산 배지로부터의 샘플에 대한 하나 이상의 분석을 실행하여 조작된 균주의 성능(예를 들어, 냉장 저장된 생성물 농도)을 테스트한다
IV. 탱크에서 테스트
평가를 위해 균주를 획득
플라스크에서 균주 배양
균주로부터 생산물 생성
성능 평가
• 냉장 저장으로부터의 균주의 씨드 뱅크를 회수한다
• 씨드 배지에서 바이오매스를 성장시키기 위해 플라스크 및/또는 탱크에서 성장시킨다
• 씨드로부터의 바이오매스와 접종된 피드백-제어 생산 환경에서 생산물을 생성한다
• 중점 및 종점 샘플을 처리하여 성능을 평가한다

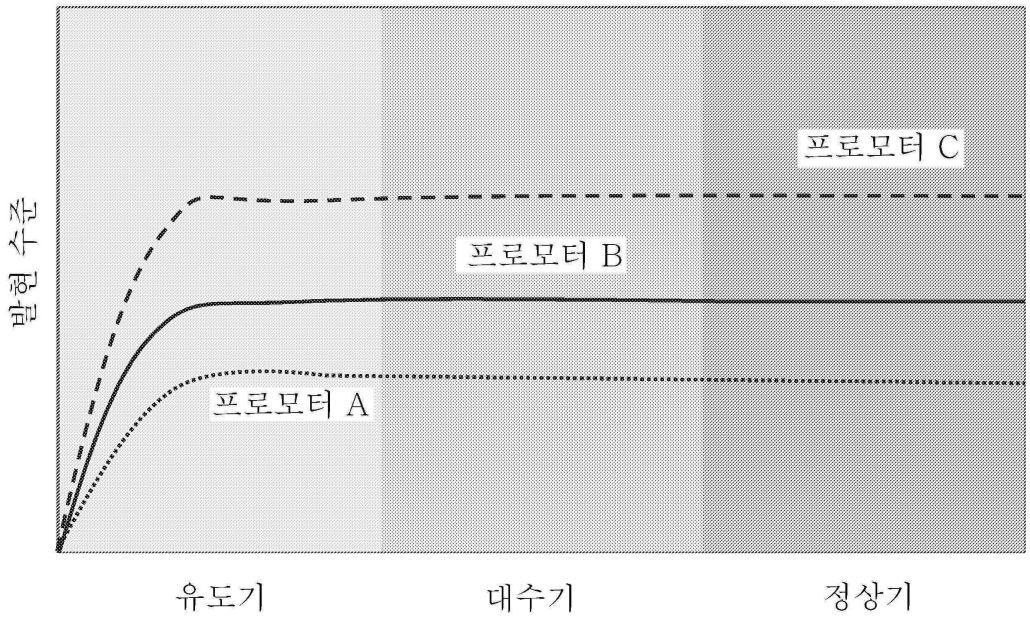
도면21


프로모터 A
프로모터 B
프로모터 C
발현 수준
유도기
대수기
정상기

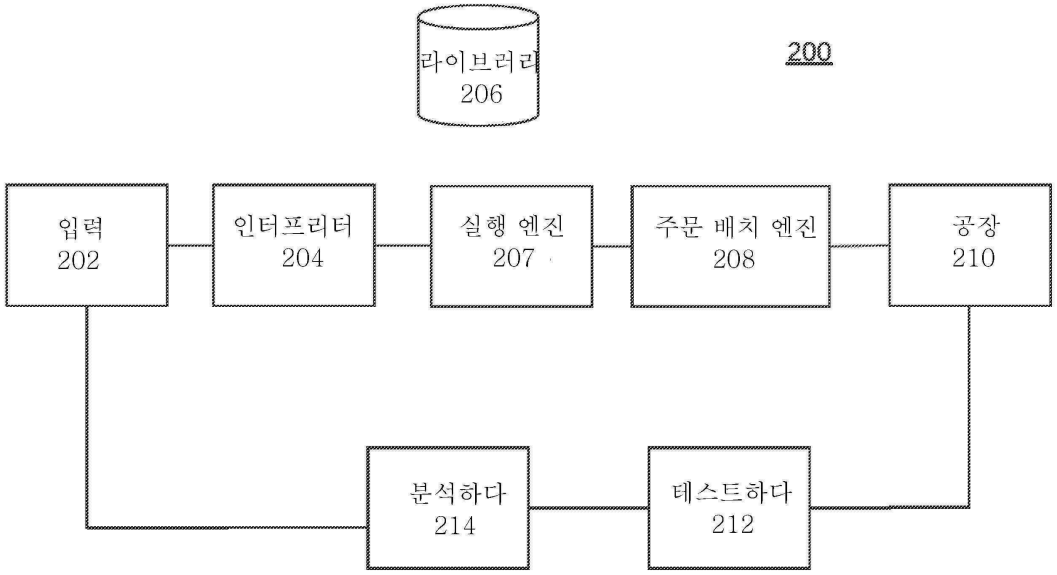
도면22


시간에 따른 기질 농도
프로모터 A
프로모터 C
프로모터 B
발현 수준
기질의 첨가
채취

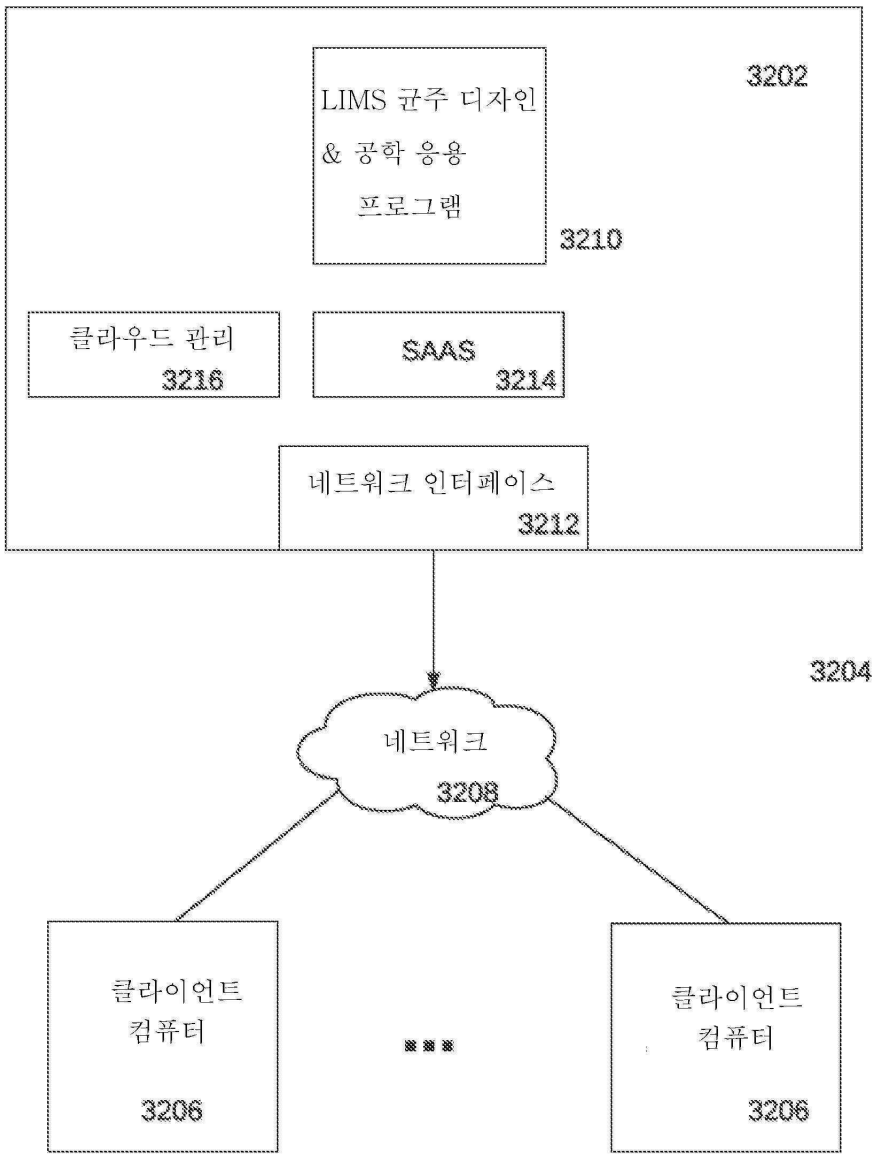
도면23


프로모터 C
프로모터 B
프로모터 A
발현 수준
유도기
대수기
정상기

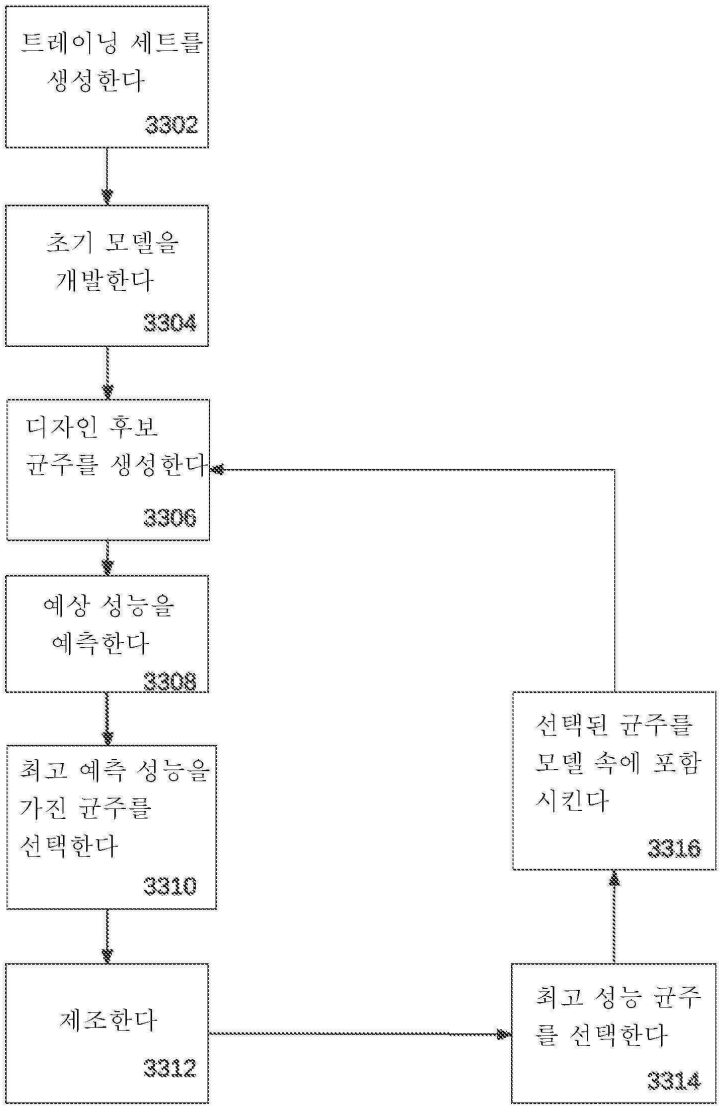
도면24


라이브러리
206
200
입력
202
인터프리터
204
실행 엔진
207
주문 배치 엔진
208
공장
210
분석하다
214
테스트하다
212

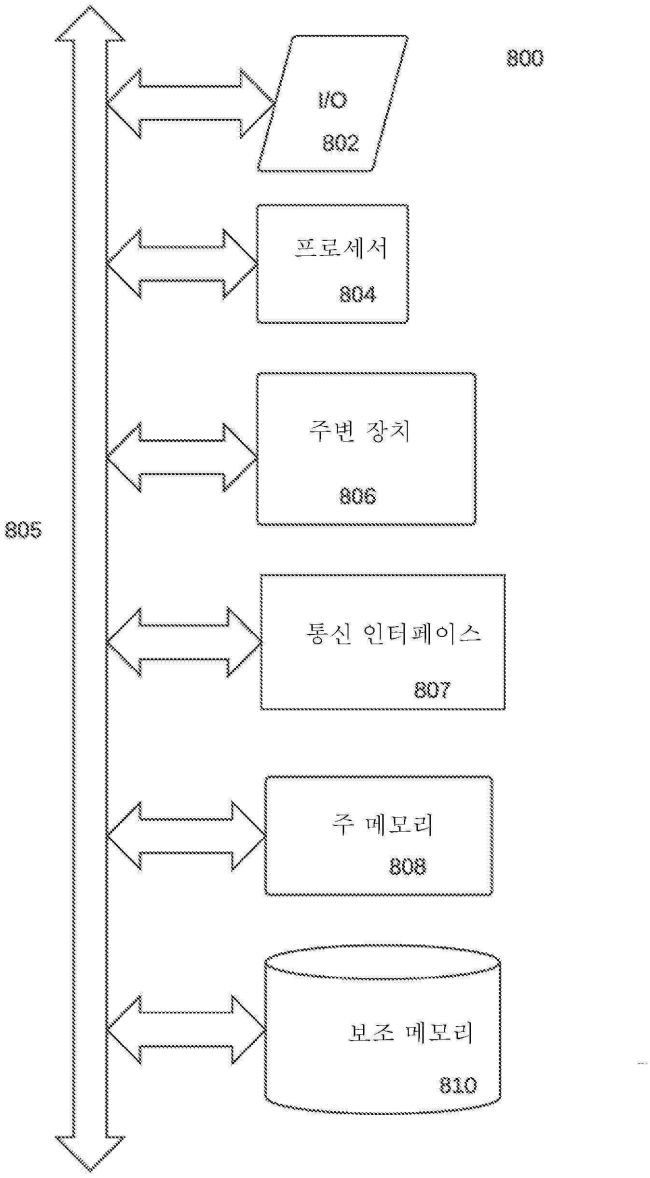
도면25


3202
LIMS 균주 디자인
& 공학 응용
프로그램
3210
클라우드 관리
3216
SAAS
3214
네트워크 인터페이스
3212
3204
네트워크
3208
클라이언트
컴퓨터
3206
클라이언트
컴퓨터
3206

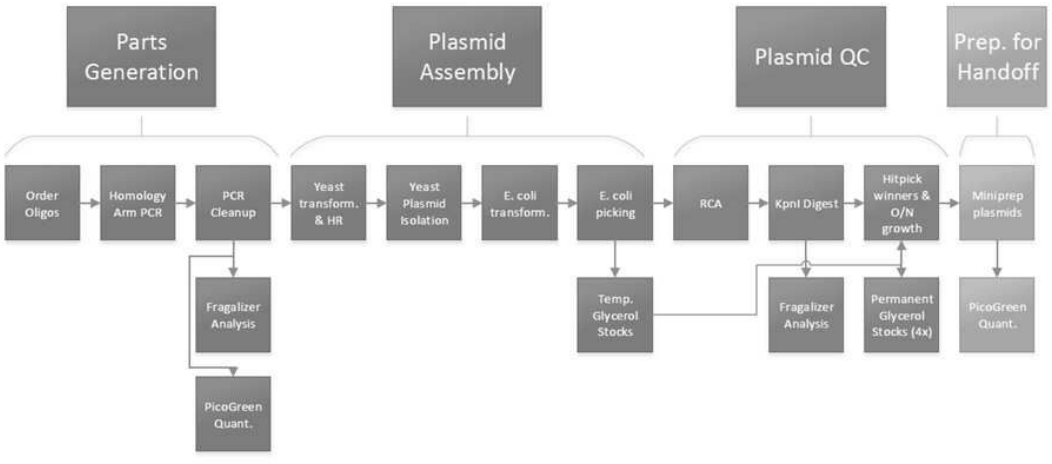
도면26


트레이닝 세트를 생성한다
3302
초기 모델을 개발한다
3304
디자인 후보 균주를 생성한다
3306
예상 성능을 예측한다
3308
최고 예측 성능을 가진 균주를 선택한다
3310
제조한다
3312
최고 성능 균주를 선택한다
3314
선택된 균주를 모델 속에 포함시킨다
3316

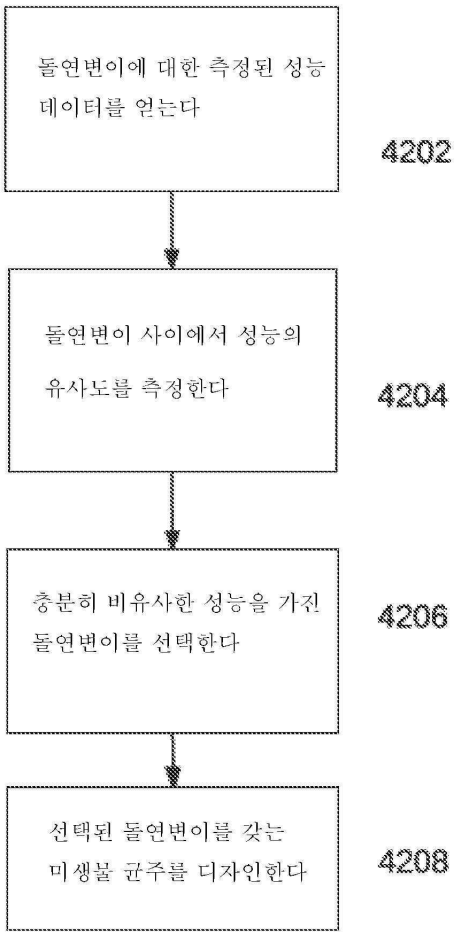
도면27


800
I/O
802
프로세서
804
주변 장치
806
805
통신 인터페이스
807
주 메모리
808
보조 메모리
810

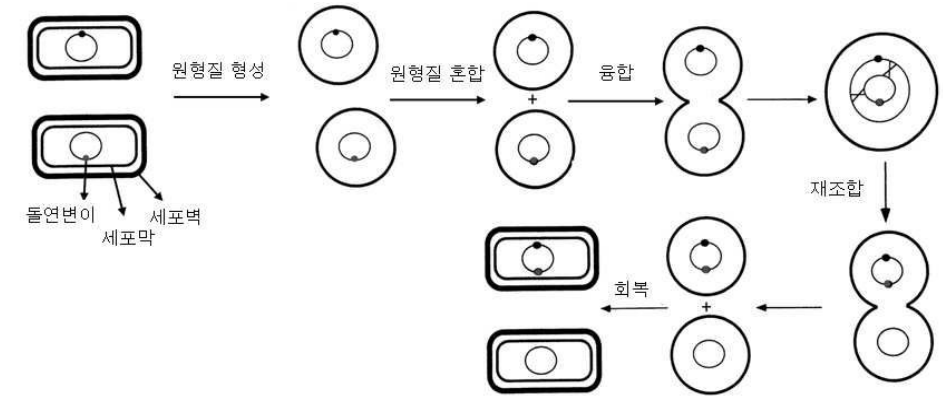
도면28


Parts Generation
Plasmid Assembly
Plasmid QC
Prep. for Handoff
Order Oligos
Homology Arm PCR
PCR Cleanup
Yeast transform. & HR
Yeast Plasmid Isolation
E. coli transform.
E. coli picking
RCA
KpnI Digest
Hitpick winners & O/N growth
Miniprep plasmids
Fragalizer Analysis
PicoGreen Quant.
Temp. Glycerol Stocks
Fragalizer Analysis
Permanent Glycerol Stocks (4x)
PicoGreen Quant.

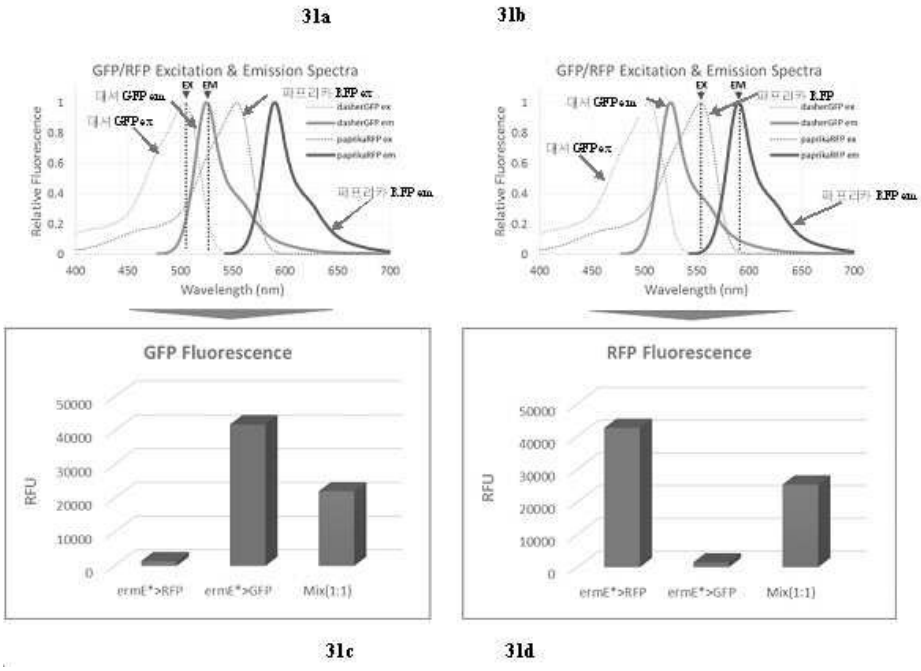
도면29


돌연변이에 대한 측정된 성능 데이터를 얻는다
4202
돌연변이 사이에서 성능의 유사도를 측정한다
4204
충분히 비유사한 성능을 가진 돌연변이를 선택한다
4206
선택된 돌연변이를 갖는 미생물 균주를 디자인한다
4208

도면30


원형질 형성
원형질 혼합
융합
재조합
회복
돌연변이
세포벽
세포막

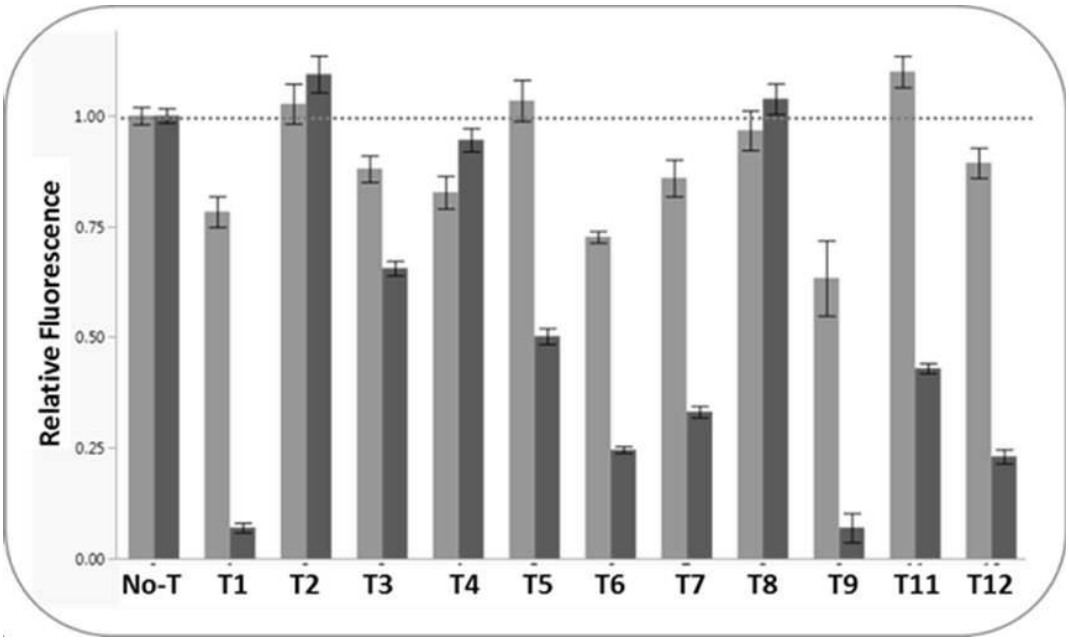
도면31



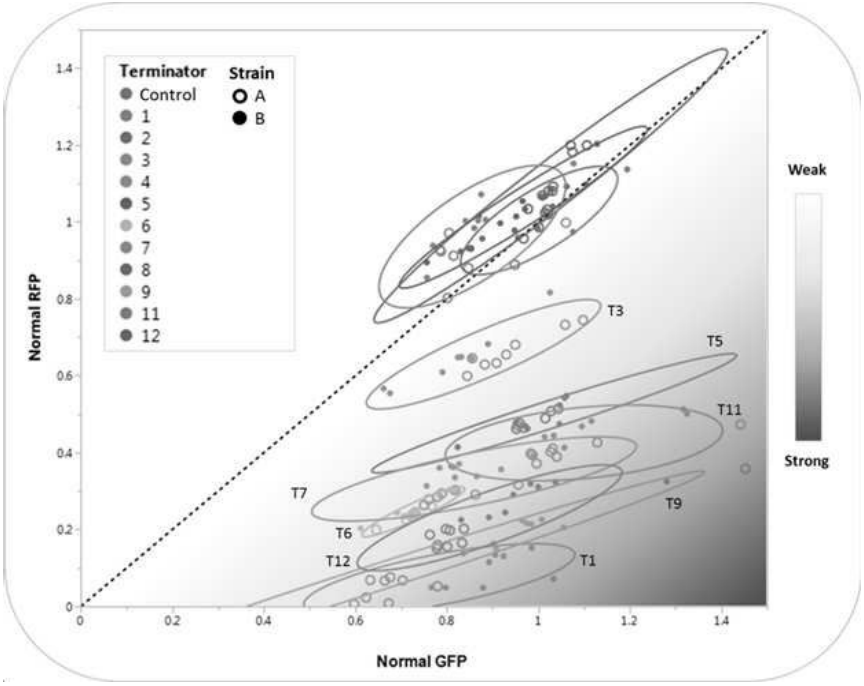
도면32



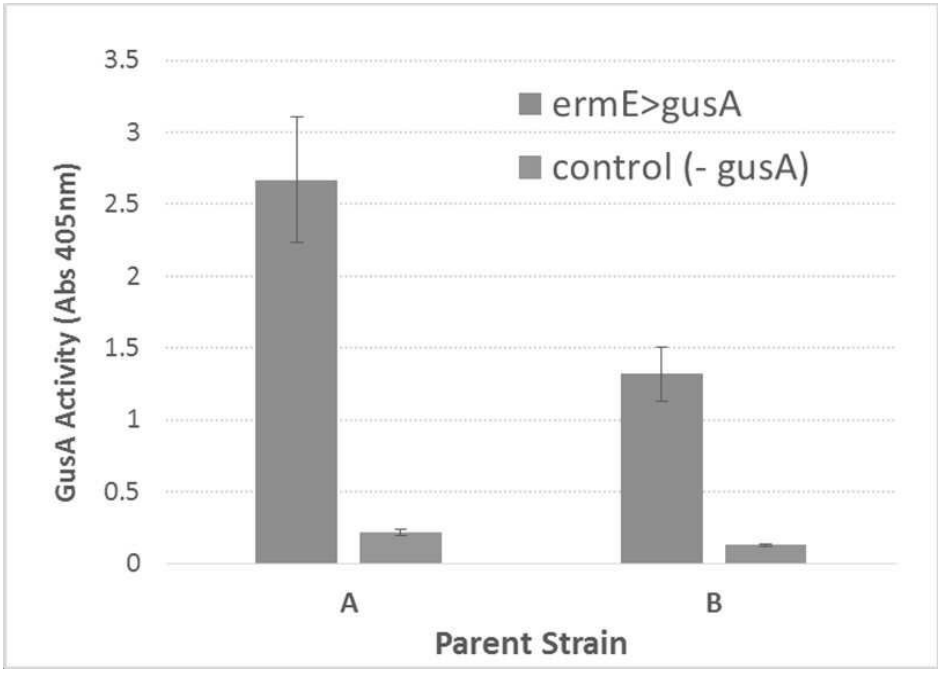
도면33



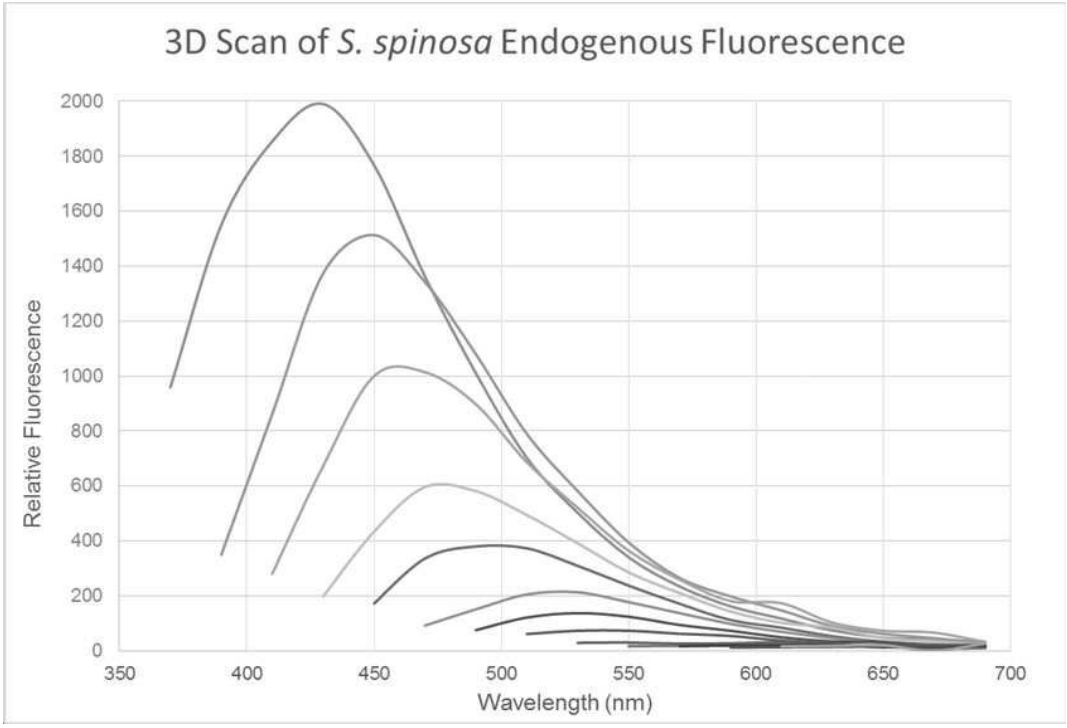
도면34


Terminator
Control
1
2
3
4
5
6
7
8
9
11
12
Strain
A
B
Weak
Strong
Normal RFP
Normal GFP
T3
T5
T11
T7
T9
T6
T12
T1
0
0.2
0.4
0.6
0.8
1
1.2
1.4

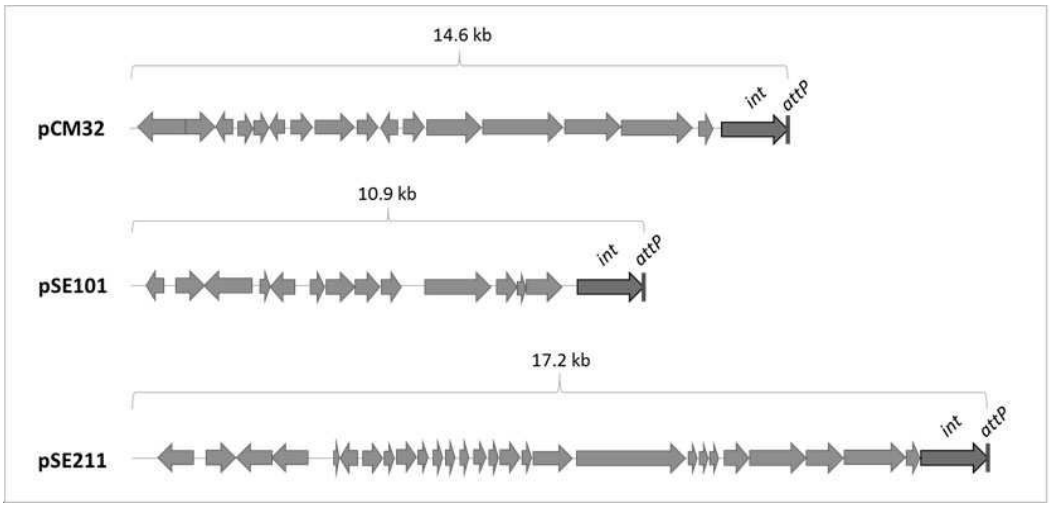
도면35


ermE>gusA
control (- gusA)
GusA Activity (Abs 405nm)
3.5
3
2.5
2
1.5
1
0.5
0
A
B
Parent Strain

도면36


3D Scan of S. spinosa Endogenous Fluorescence
Relative Fluorescence
2000
1800
1600
1400
1200
1000
800
600
400
200
0
350
400
450
500
550
600
650
700
Wavelength (nm)

도면37


14.6 kb
pCM32
int
attP
10.9 kb
pSE101
int
attP
17.2 kb
pSE211
int
attP

도면38

```
Sequence ID: Query_234253Length: 8373170Number of Matches: 13
Related Information
Range 1: 2710919 to 2711068GraphicsNext MatchPrevious Match
Alignment statistics for match #1
Score     Expect    Identities          Gaps      Strand
266 bits(294)       7e-72     149/150(99%)        0/150(0%) Plus/Minus
Query  1        GAAAGAGCCCCGGTCGCCATGTCGCGTCCGGGGCTCTGACCAGCGGAGGATACGGGATTC  60
                ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct  2711068  GAAAGAGCCCCGGTCGCGATGTCGCGTCCGGGGCTCTGACCAGCGGAGGATACGGGATTC  2711009

Query  61       GAACCCGTGAGGGCTATTAACCCAACACGATTTCCAATCGTGCGCCTTAGGCCGCTCGGC  120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2711008  GAACCCGTGAGGGCTATTAACCCAACACGATTTCCAATCGTGCGCCTTAGGCCGCTCGGC  2710949

Query  121      CAATCCTCCGTGCGGAAGCCTAGCAGGGCG  150
                ||||||||||||||||||||||||||||||
Sbjct  2710948  CAATCCTCCGTGCGGAAGCCTAGCAGGGCG  2710919
```

도면39

```
Sequence ID: Query_181603Length: 8373170Number of Matches: 47
Related Information
Range 1: 2471978 to 2472088GraphicsNext MatchPrevious Match
Alignment statistics for match #1
Score   Expect  Identities     Gaps    Strand
168 bits(186)  1e-42  104/111(94%)   0/111(0%)       Plus/Plus
Query  1        GCGCGACGTAGCGCACGCCGCTGTAGCTCAGTTGGTAGAGCGCCCGCCTTGTAAGCGGAC  60
                || ||||| || ||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2471978  GCACGACGAAGTGCATGCCGCTGTAGCTCAGTTGGTAGAGCGCCCGCCTTGTAAGCGGAC  2472037

Query  61       GGTCAGGGGTTCGAGTCCCCTCAGCGGCTCCGAAACTACCAGCGGAAACGC  111
                ||||||||||||||||||||||||||||||||||   ||||||||||||||
Sbjct  2472038  GGTCAGGGGTTCGAGTCCCCTCAGCGGCTCCGACCGTACCAGCGGAAACGC  2472088
```

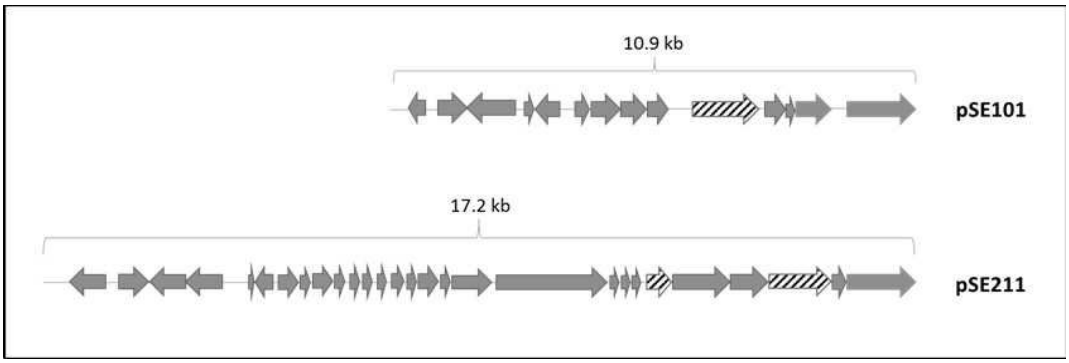
도면40

```
Sequence ID: Query_167805Length: 8373170Number of Matches: 1
Related Information
Range 1: 3350688 to 3350824GraphicsNext MatchPrevious Match
Alignment statistics for match #1
Score   Expect  Identities     Gaps    Strand
170 bits(188)  4e-43  122/138(88%)   2/138(1%)       Plus/Plus
Query  3        TTGAGCCGGGCGCGAACGGCCAGGTAGCTCAGTTGGTACGAGCGTCCGCCTGAAAAGCGG  62
                ||||||||  ||| | ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3350688  TTGAGCCGA-CGCTATCGGCCAGGTAGCTCAGTTGGTACGAGCGTCCGCCTGAAAAGCGG  3350746

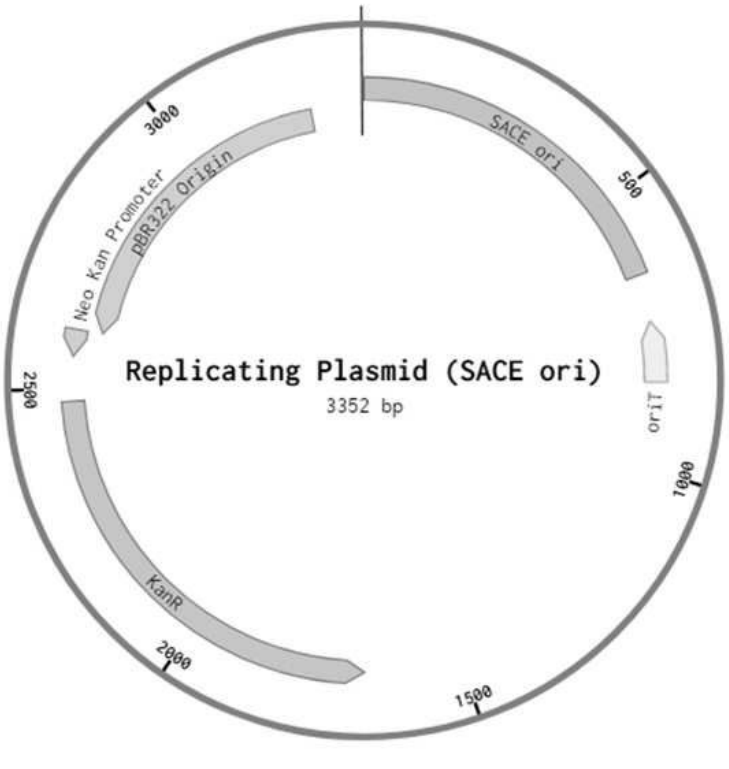
Query  63       AAGGTCGGCGGTTCGACCCCGCCCCTGGCCACCACCAAGAGCTGCGAACA-GCCCCTGCC  121
                |||||||||||||||||||||||||||||||| |  ||| | || || | |||||||||
Sbjct  3350747  AAGGTCGGCGGTTCGACCCCGCCCCTGGCCACCGCTCAGACCAGCCAAAACGCCCCTGCC  3350806

Query  122      GATCTCGGTGGGGGCTTT  139
                ||  |||| |||||| ||
Sbjct  3350807  GAGATCGGCGGGGGCGTT  3350824
```

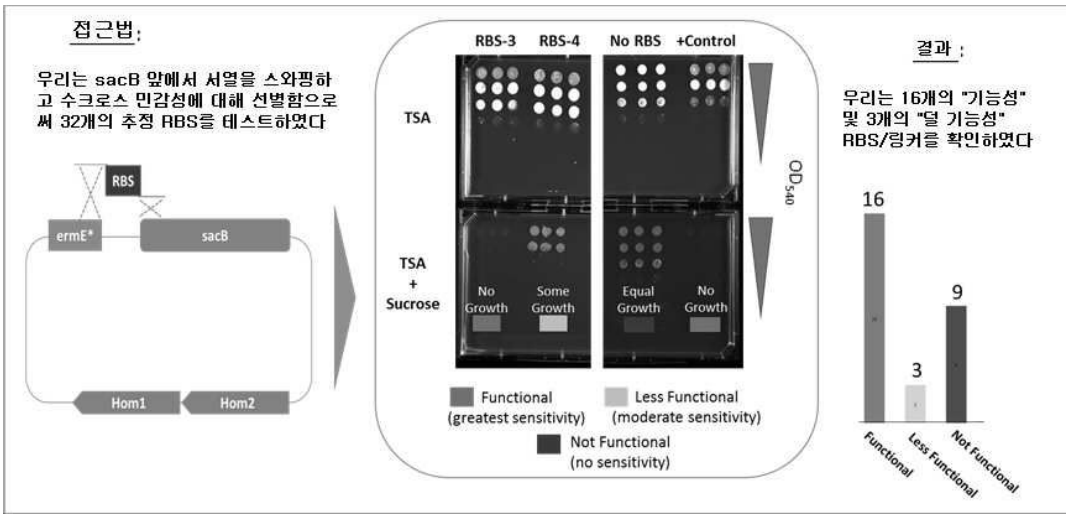
도면41a


10.9 kb
pSE101
17.2 kb
pSE211

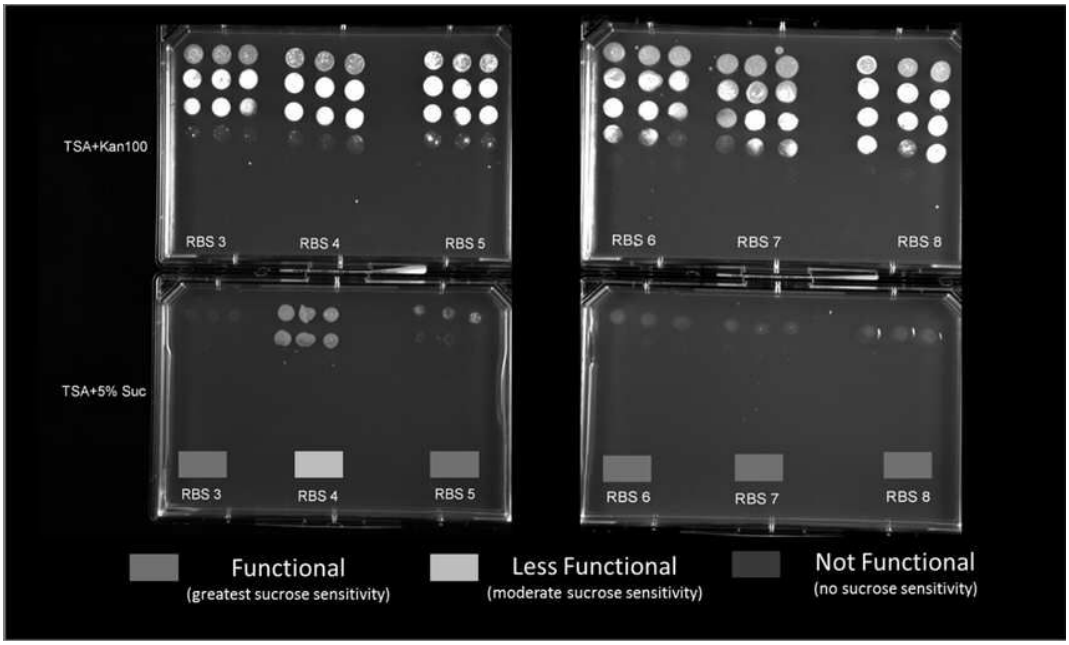
도면41b


Replicating Plasmid (SACE ori)
3352 bp
SACE ori
Neo Kan Promoter
pBR322 Origin
oriT
KanR
500
1000
1500
2000
2500
3000

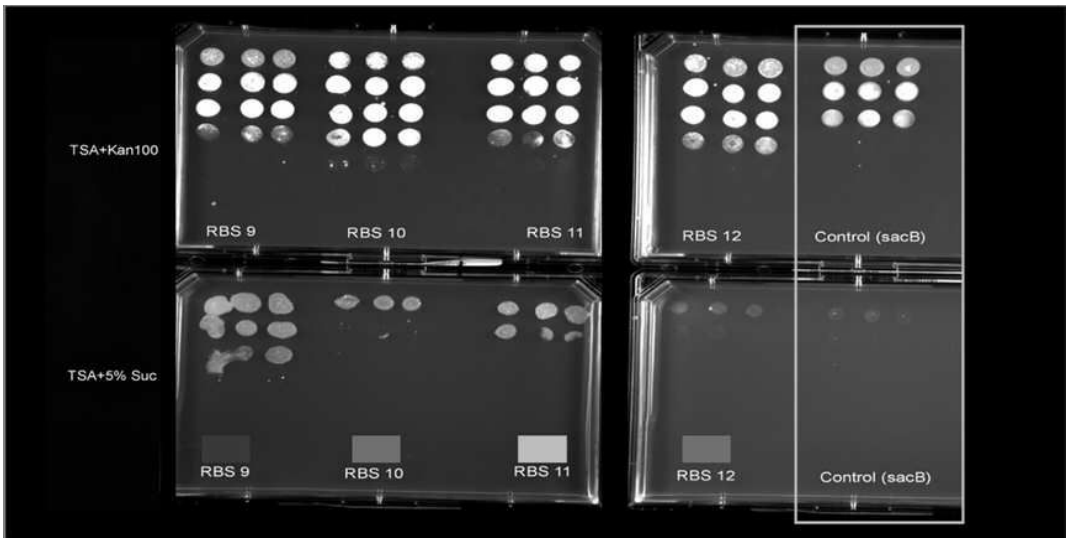
도면42


접근법:
우리는 sacB 앞에서 서열을 스와핑하고 수크로스 민감성에 대해 선별함으로써 32개의 추정 RBS를 테스트하였다
RBS
ermE*
sacB
Hom1
Hom2
RBS-3
RBS-4
No RBS
+Control
TSA
TSA + Sucrose
No Growth
Some Growth
Equal Growth
No Growth
OD600
Functional (greatest sensitivity)
Less Functional (moderate sensitivity)
Not Functional (no sensitivity)
결과 :
우리는 16개의 "기능성" 및 3개의 "덜 기능성" RBS/링커를 확인하였다
16
3
9
Functional
Less Functional
Not Functional

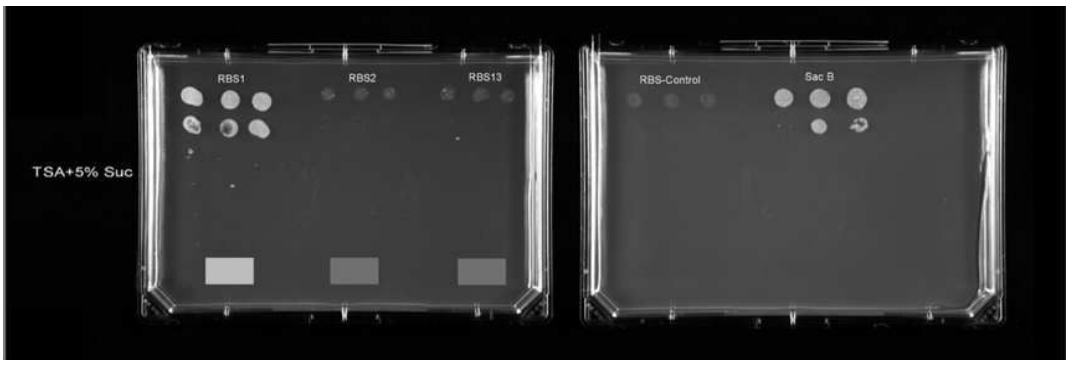
도면43a


TSA+Kan100
TSA+5% Suc
RBS 3
RBS 4
RBS 5
RBS 6
RBS 7
RBS 8
Functional
(greatest sucrose sensitivity)
Less Functional
(moderate sucrose sensitivity)
Not Functional
(no sucrose sensitivity)

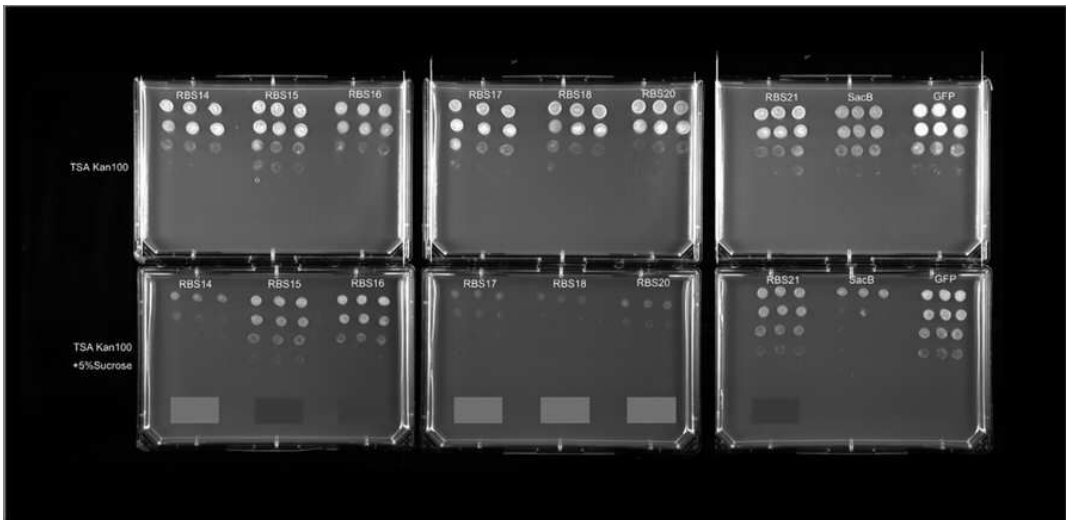
도면43b



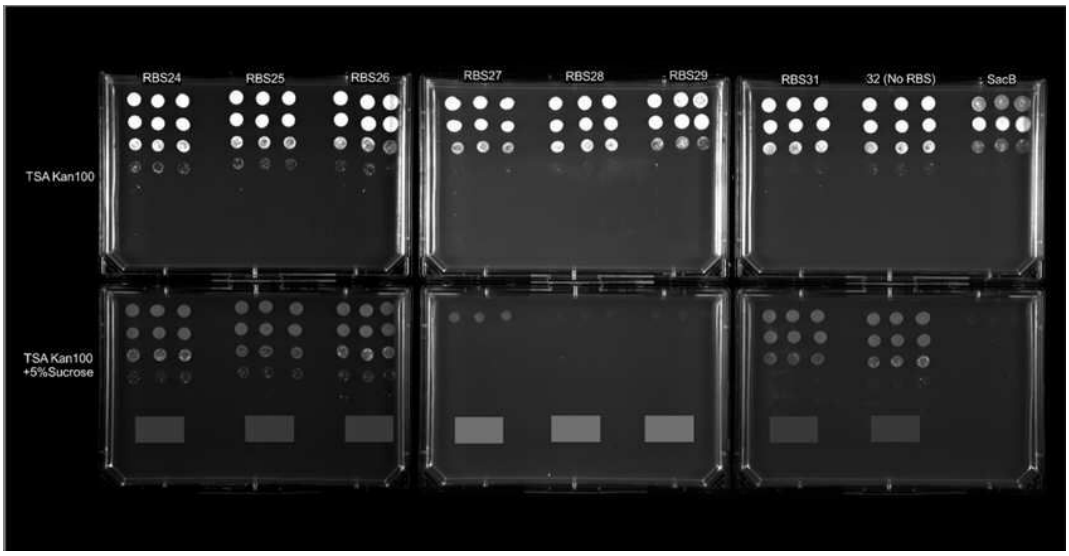
도면43c



도면43d



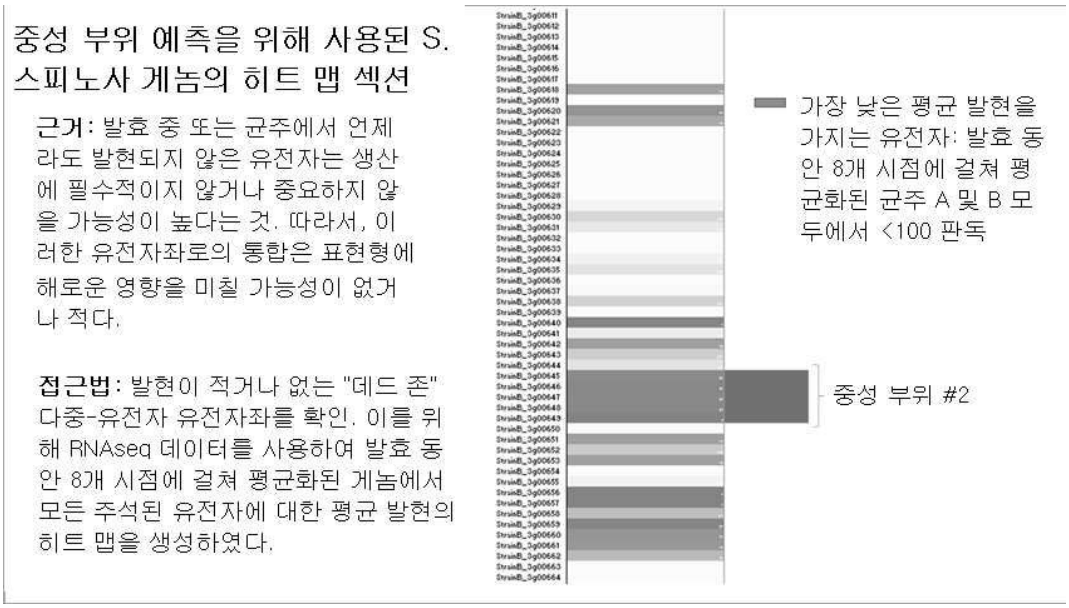
도면43e


RBS24
RBS25
RBS26
RBS27
RBS28
RBS29
RBS31
32 (No RBS)
SacB
TSA Kan100
TSA Kan100
+5%Sucrose

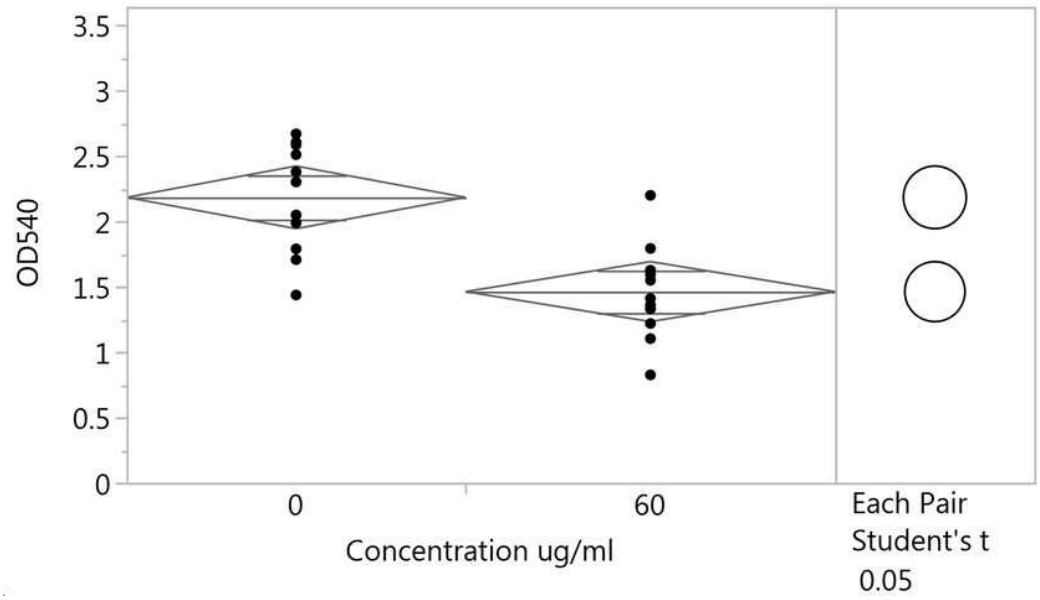
도면44


기능상실 트랜스포존 - 예시적 다이어그램
ME
Selectable Marker ‡
ME
기능획득 트랜스포존 - 예시적 다이어그램
ME
Selectable Marker ‡
Gain of Function Element*
ME
기능획득 재활용 가능 트랜스포존 - 예시적 다이어그램
ME
Gain of Function Element*
Selectable Marker ‡
Counter-Selectable Marker
Gain of Function Element*
ME
‡ 선택 가능한 마커는 모듈형이며 숙주 세포에 따라 다양할 수 있다
* GoF 요소는 모듈형이며 용해도 태그, 프로모터, 카운터 선택 등을 포함할 수 있다
฿ 이들 영역은 직접 반복되어 재조합 및 루프-아웃을 가능하게 할 수 있다

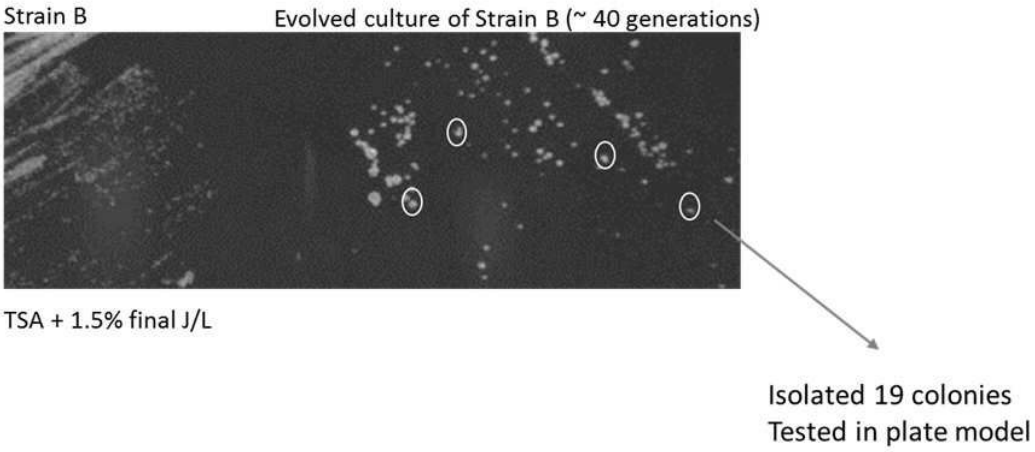
도면45


중성 부위 예측을 위해 사용된 S. 스피노사 게놈의 히트 맵 섹션
근거: 발효 중 또는 균주에서 언제라도 발현되지 않은 유전자는 생산에 필수적이지 않거나 중요하지 않을 가능성이 높다는 것. 따라서, 이러한 유전자좌로의 통합은 표현형에 해로운 영향을 미칠 가능성이 없거나 적다.
접근법: 발현이 적거나 없는 "데드 존" 다중-유전자 유전자좌를 확인. 이를 위해 RNAseq 데이터를 사용하여 발효 동안 8개 시점에 걸쳐 평균화된 게놈에서 모든 주석된 유전자에 대한 평균 발현의 히트 맵을 생성하였다.
가장 낮은 평균 발현을 가지는 유전자: 발효 동안 8개 시점에 걸쳐 평균화된 균주 A 및 B 모두에서 <100 판독
중성 부위 #2

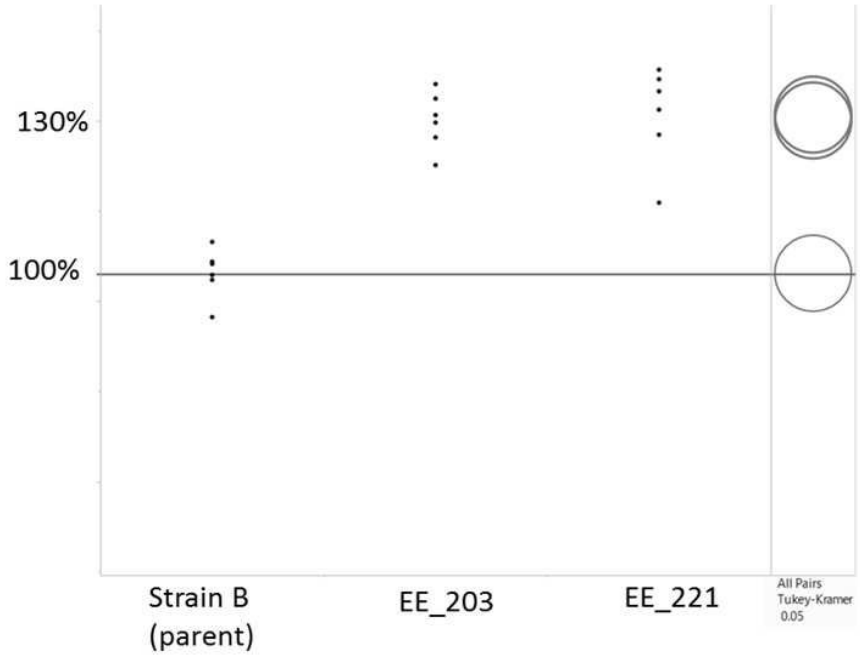
도면46


OD540
3.5
3
2.5
2
1.5
1
0.5
0
0
60
Concentration ug/ml
Each Pair
Student's t
0.05

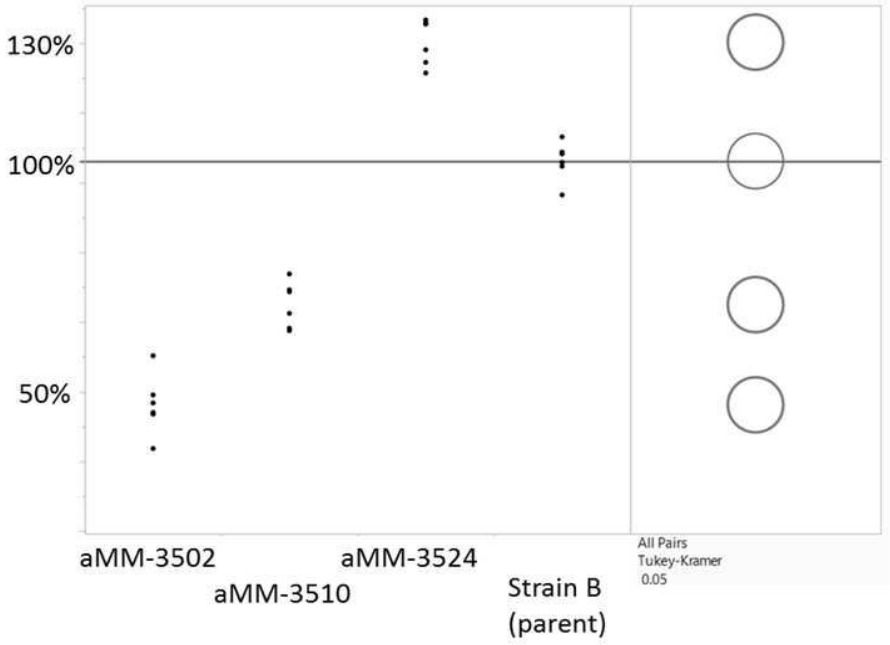
도면47


Strain B
Evolved culture of Strain B (~ 40 generations)
TSA + 1.5% final J/L
Isolated 19 colonies
Tested in plate model

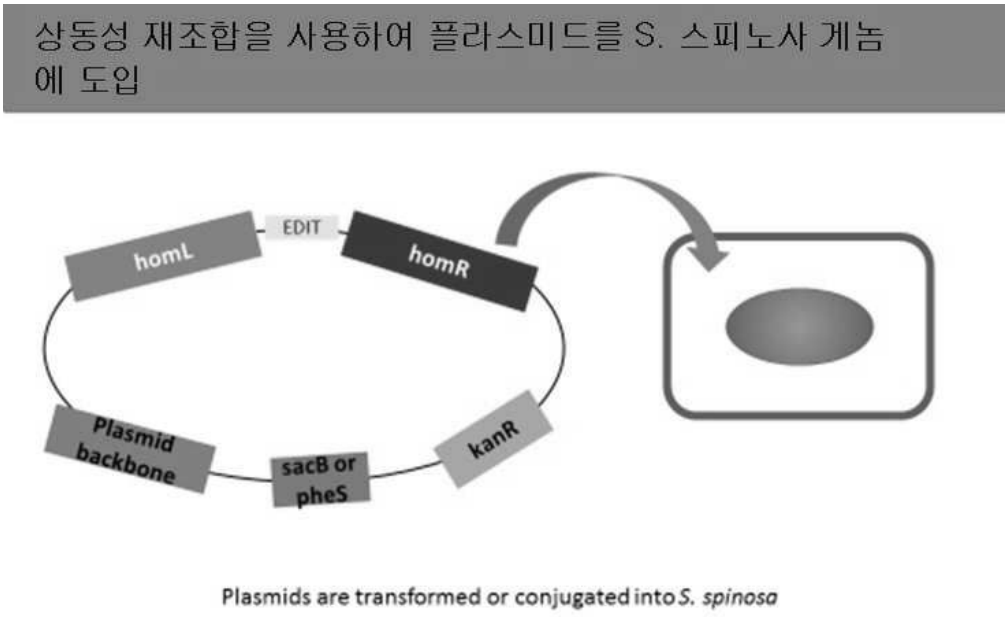
도면48a


130%
100%
Strain B
(parent)
EE_203
EE_221
All Pairs
Tukey-Kramer
0.05

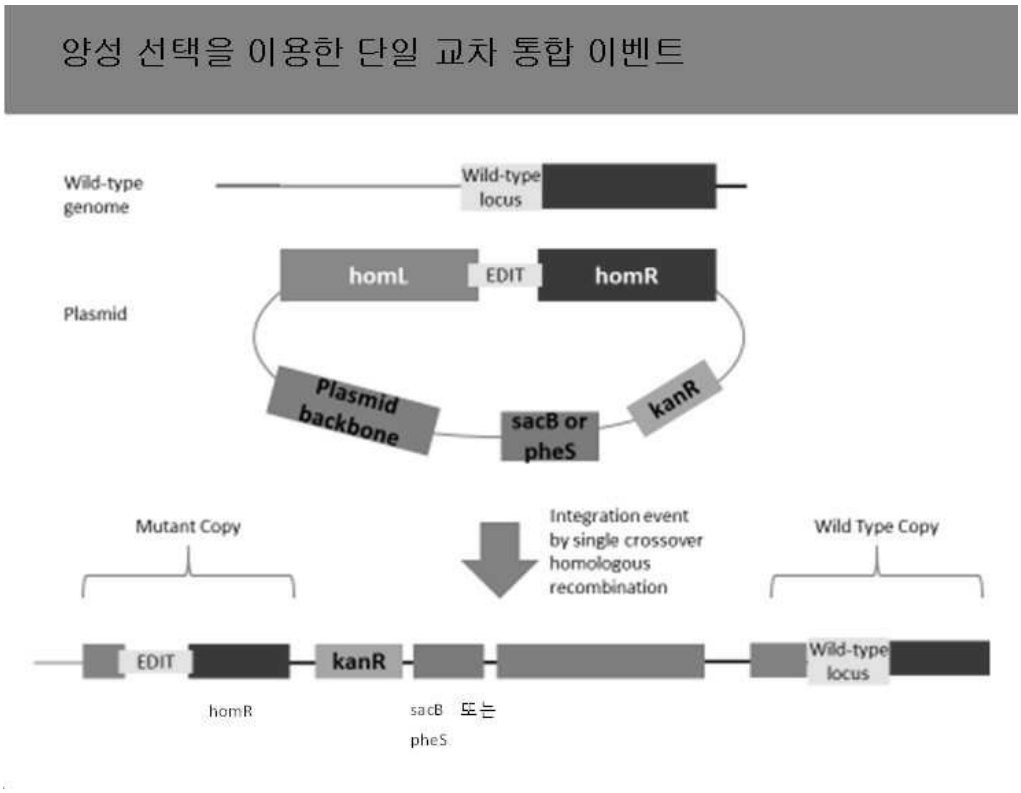
도면48b


130%
100%
50%
aMM-3502
aMM-3510
aMM-3524
Strain B
(parent)
All Pairs
Tukey-Kramer
0.05

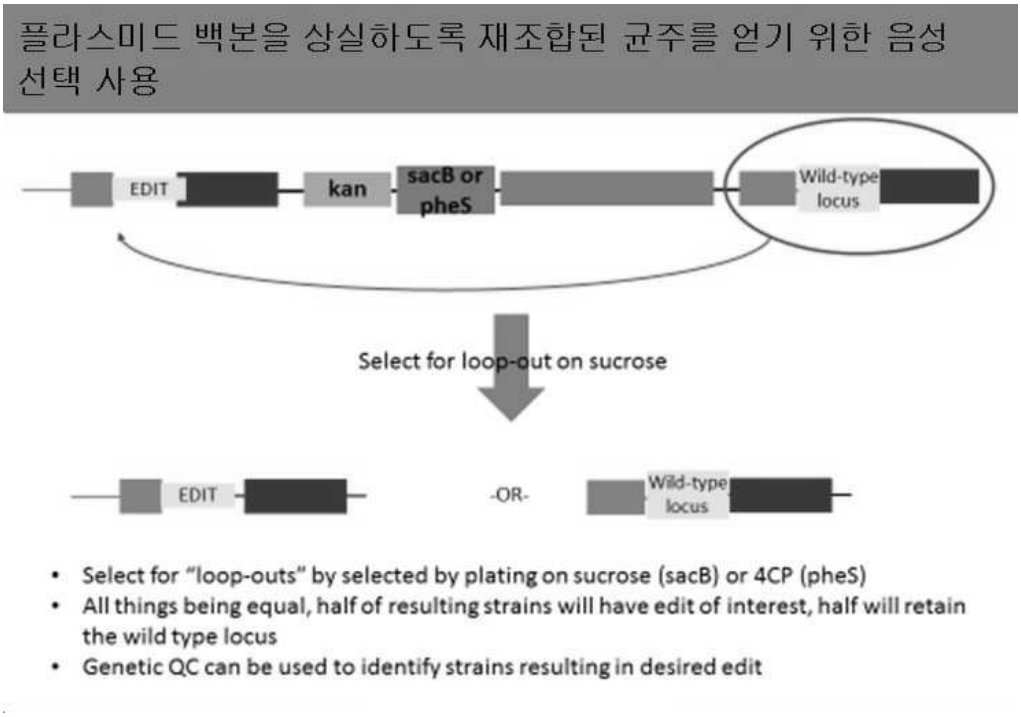
도면49a


상동성 재조합을 사용하여 플라스미드를 S. 스피노사 게놈에 도입
homL
EDIT
homR
Plasmid backbone
sacB or pheS
kanR
Plasmids are transformed or conjugated into S. spinosa

도면49b

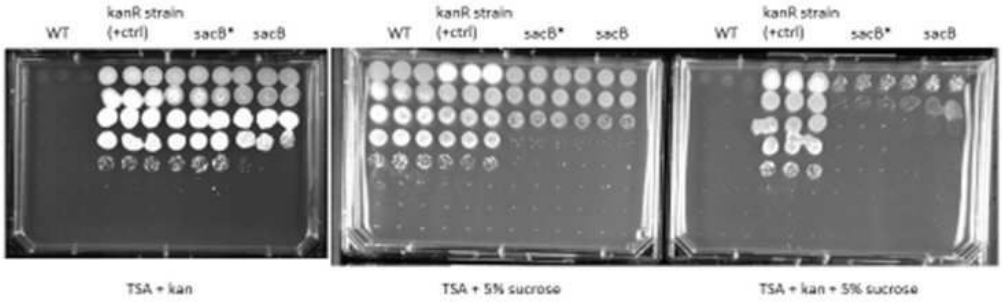

양성 선택을 이용한 단일 교차 통합 이벤트
Wild-type genome
Wild-type locus
Plasmid
homL
EDIT
homR
Plasmid backbone
sacB or pheS
kanR
Mutant Copy
Integration event by single crossover homologous recombination
Wild Type Copy
EDIT
kanR
Wild-type locus
homR
sacB 또는 pheS

도면49c


플라스미드 백본을 상실하도록 재조합된 균주를 얻기 위한 음성 선택 사용
EDIT
kan
sacB or pheS
Wild-type locus
Select for loop-out on sucrose
EDIT
-OR-
Wild-type locus
• Select for "loop-outs" by selected by plating on sucrose (sacB) or 4CP (pheS)
• All things being equal, half of resulting strains will have edit of interest, half will retain the wild type locus
• Genetic QC can be used to identify strains resulting in desired edit

도면50

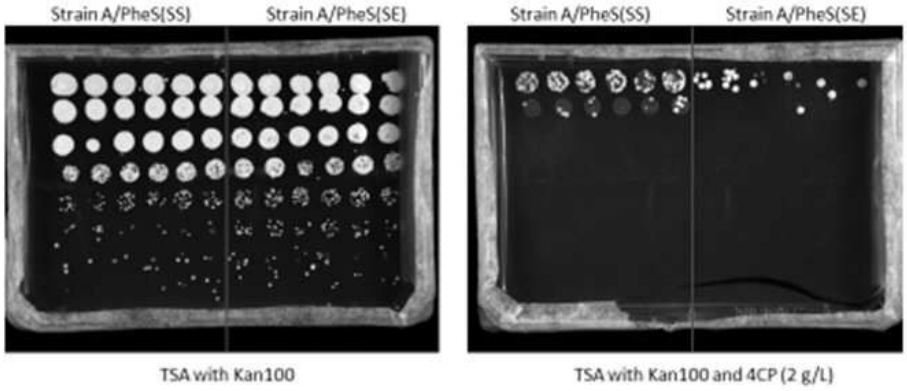
SacB는 S. 스피노사의 추정 카운터 선택 마커이다 - 이는 선택 매질상에서의 제한 성장을 유발한다



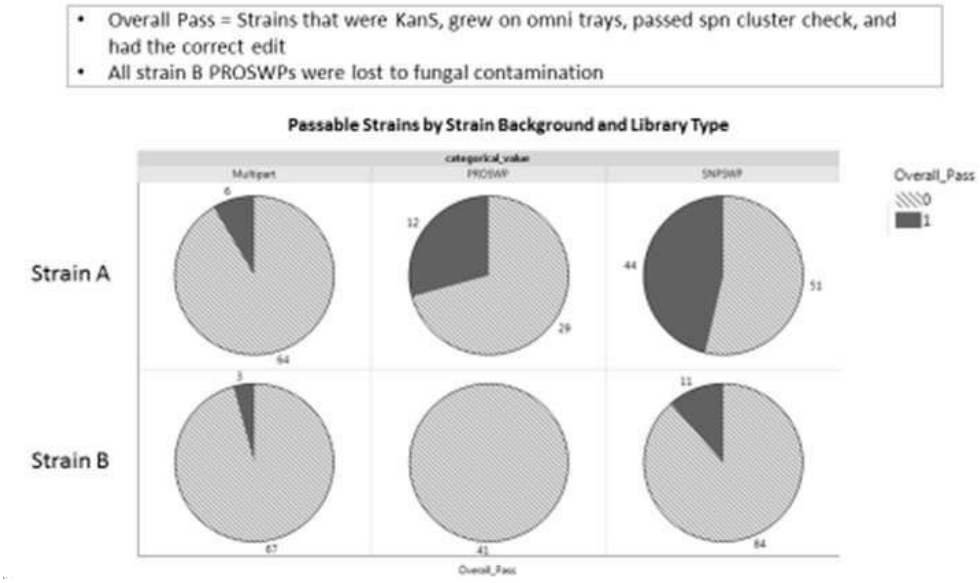
도면51

균주 A에서 CSM으로서 PheS(SS) 및 PheS(SE) 평가

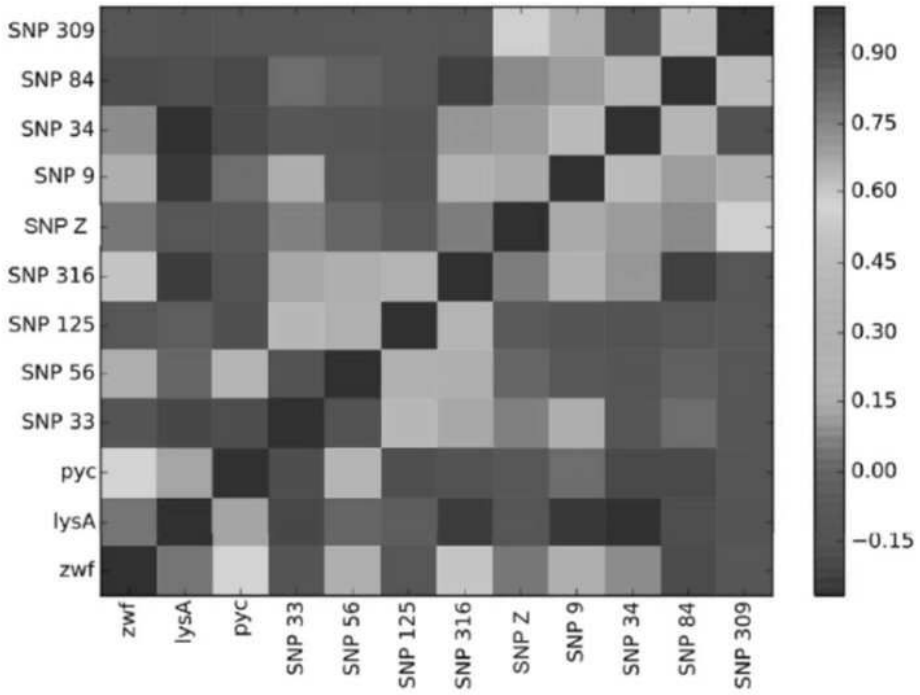
Strain A/PheS(SS) and strain A/PheS(SE) are currently tested for 4CP sensitivity at 2 g/L. A culture dilution series was spotted in six replicates onto TSA/Kan100 and TSA/Kan100 containing 4CP.



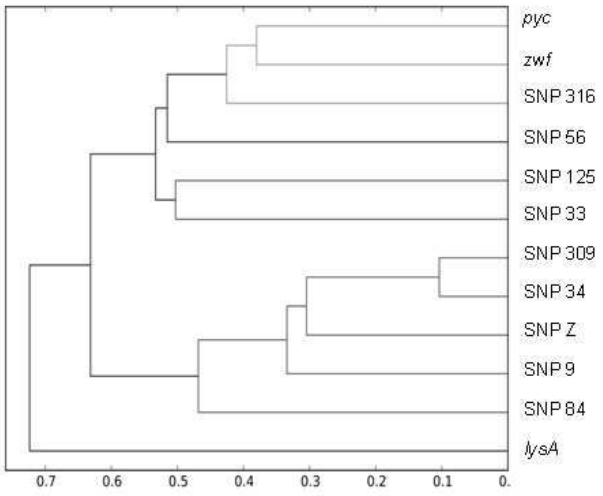
도면52


• Overall Pass = Strains that were KanS, grew on omni trays, passed spn cluster check, and had the correct edit
• All strain B PROSWPs were lost to fungal contamination
Passable Strains by Strain Background and Library Type
Multipart
PROSWP
SNPSWP
Overall_Pass
0
1
Strain A
Strain B
6
64
12
29
44
51
3
67
11
84
41

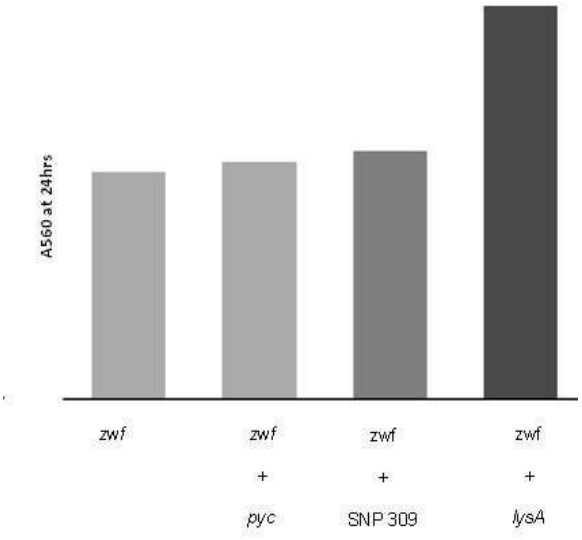
도면53


SNP 309
SNP 84
SNP 34
SNP 9
SNP Z
SNP 316
SNP 125
SNP 56
SNP 33
pyc
lysA
zwf
0.90
0.75
0.60
0.45
0.30
0.15
0.00
−0.15

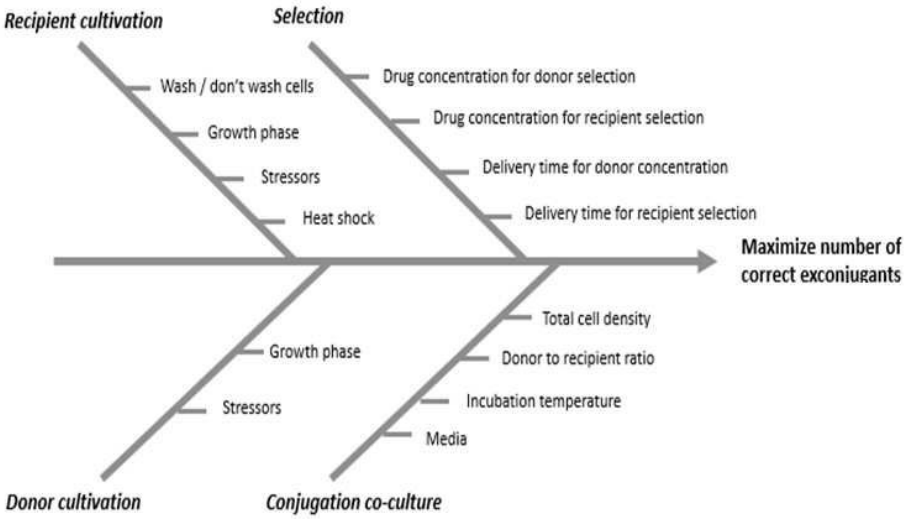
도면54a


pyc
zwf
SNP 316
SNP 56
SNP 125
SNP 33
SNP 309
SNP 34
SNP Z
SNP 9
SNP 84
lysA
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.

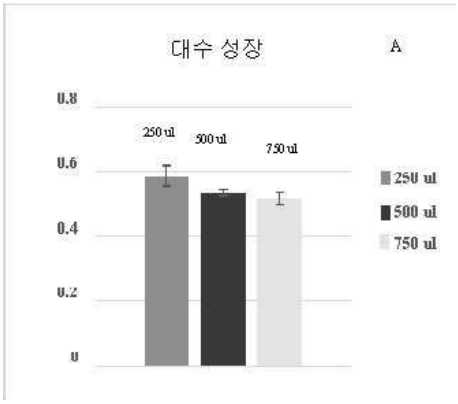
도면54b


A560 at 24hrs
zwf
zwf + pyc
zwf + SNP 309
zwf + lysA

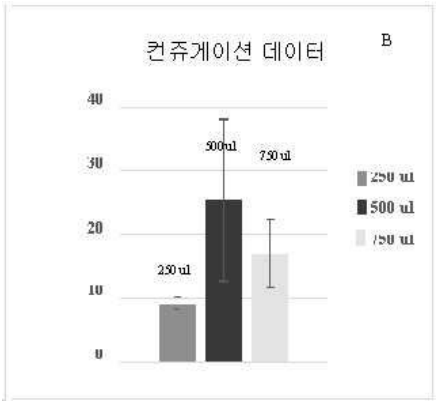
도면55



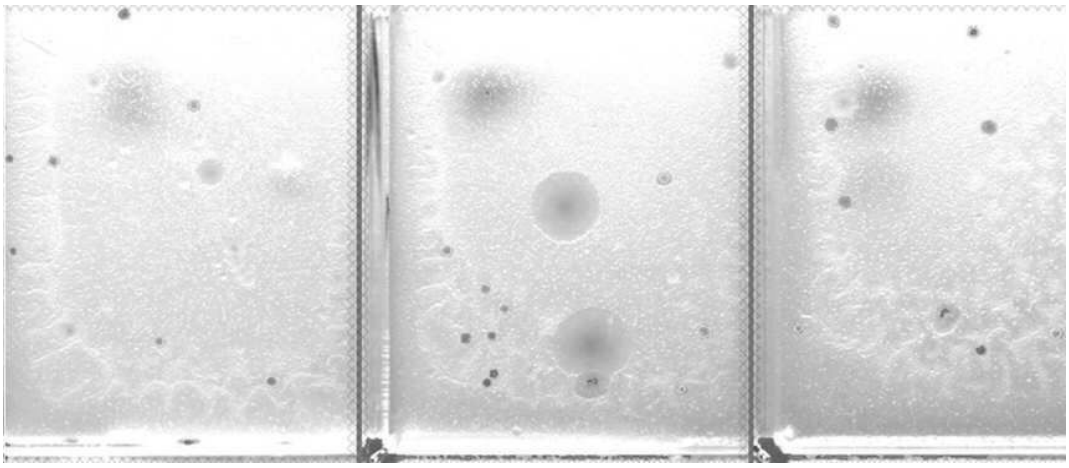
도면56a



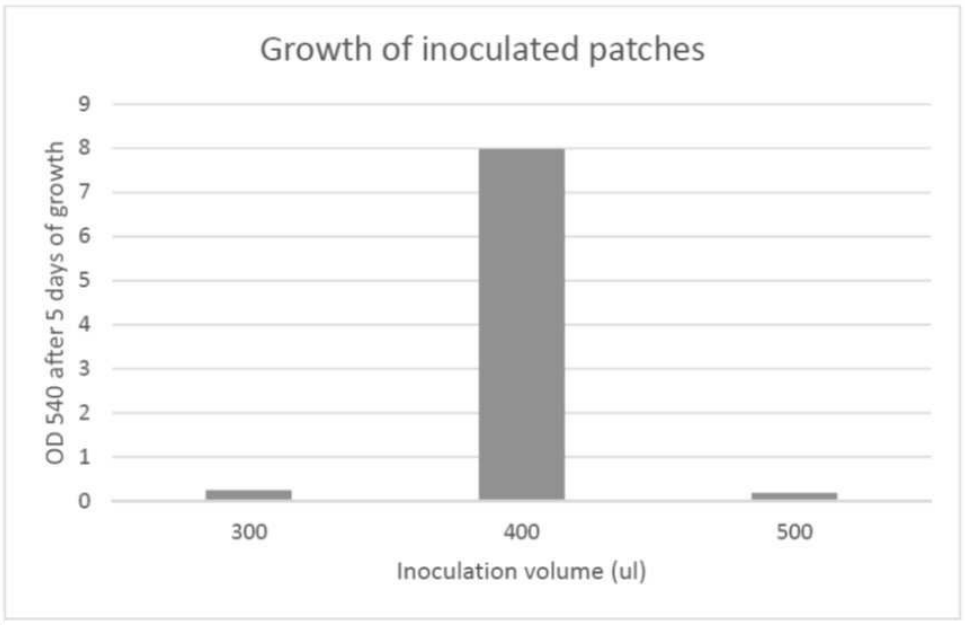
도면56b



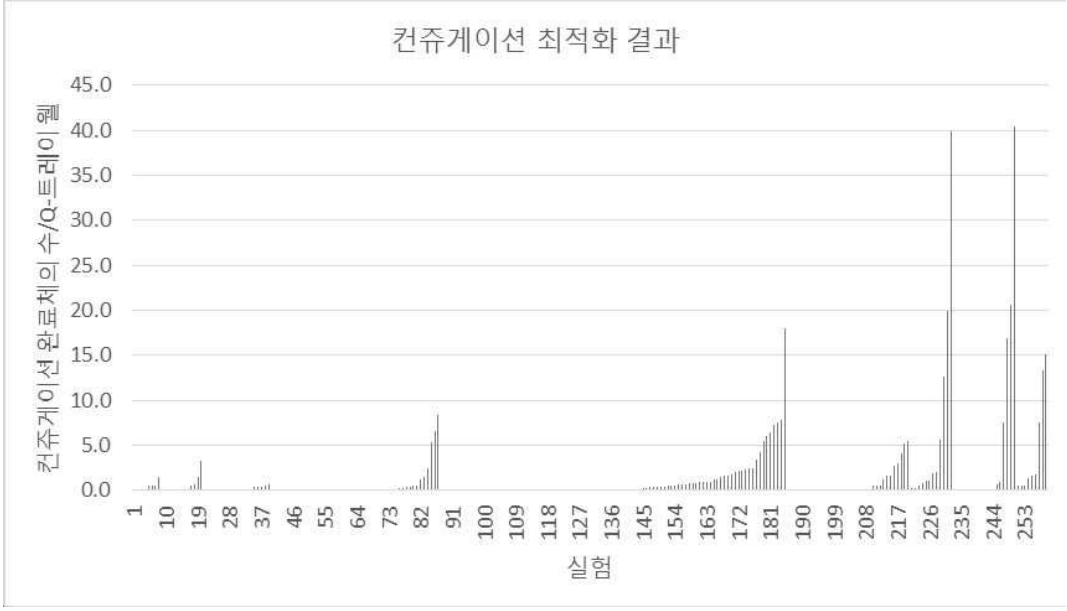
도면57



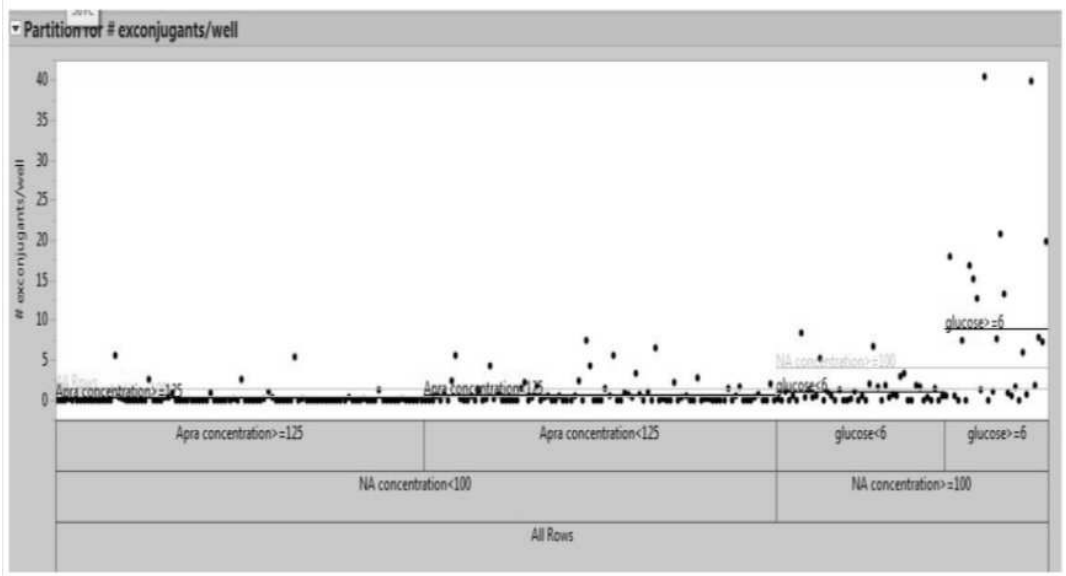
도면58


Growth of inoculated patches
OD 540 after 5 days of growth
0
1
2
3
4
5
6
7
8
9
300
400
500
Inoculation volume (ul)

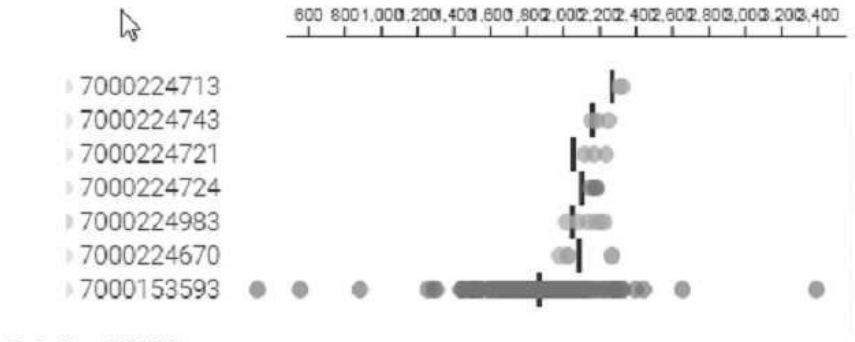
도면59


컨쥬게이션 최적화 결과
컨쥬게이션 완료체의 수/Q-트레이 웰
45.0
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
1
10
19
28
37
46
55
64
73
82
91
100
109
118
127
136
145
154
163
172
181
190
199
208
217
226
235
244
253
실험

도면60

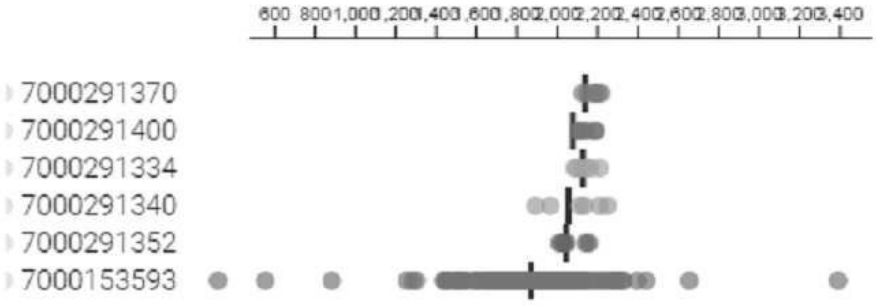

Partition of # exconjugants/well
exconjugants/well
40
35
30
25
20
15
10
5
0
glucose>=6
Apra concentration>=125
Apra concentration<125
glucose<6
glucose>=6
NA concentration<100
NA concentration>=100
All Rows

도면61



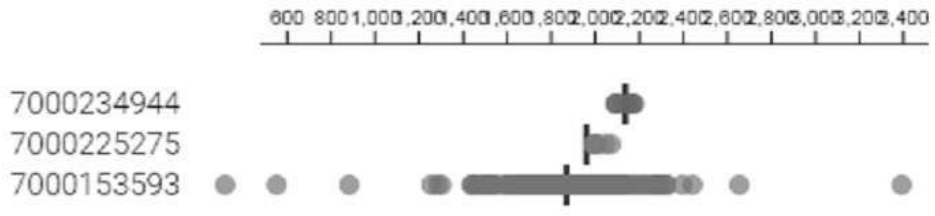
Strain ID	SNPSWP type
7000224713	SNPSWP-NON SYNONYMOUS Coding
7000224721	SNPSWP-NON SYNONYMOUS Coding
7000224743	SNPSWP-NON SYNONYMOUS Coding
7000224724	SNPSWP-STOP LOST
7000224983	SNPSWP-NON SYNONYMOUS Coding
7000224670	SNPSWP-STOP LOST
7000153593	Control (parent strain)

도면62



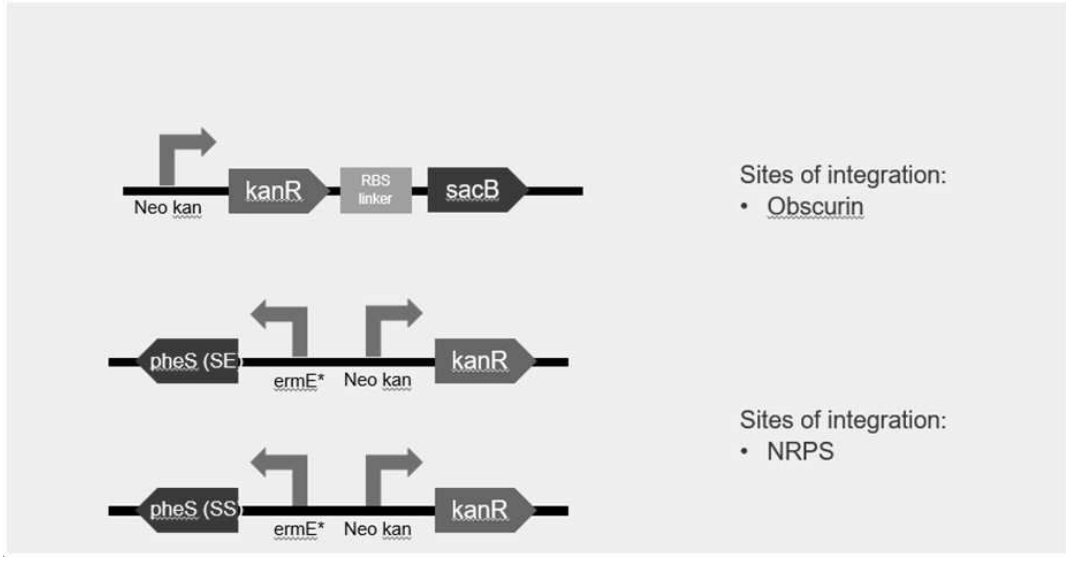
Strain ID	Details of terminator insertion
7000291370	T1 insertion upstream of potassium transporter
7000291334	T1 insertion upstream of DNA binding protein
7000291400	T1 insertion upstream of specific transcriptional regulator
7000291340	T1 insertion upstream of specific transcriptional regulator
7000291352	T1 insertion upstream of specific transcriptional regulator
7000153593	Control (parent strain)

도면63

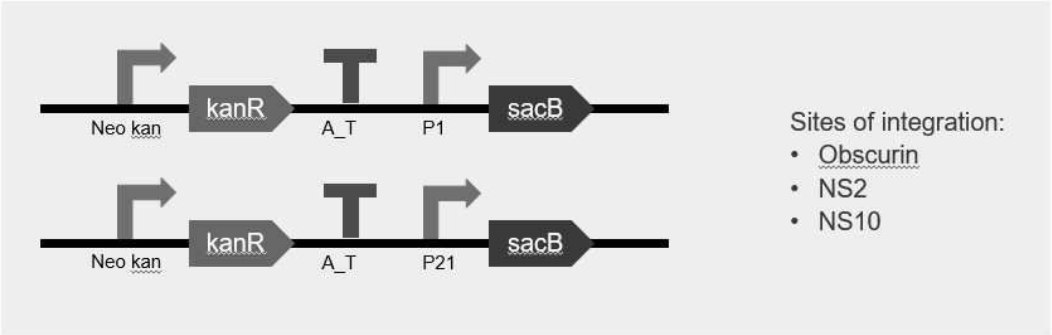


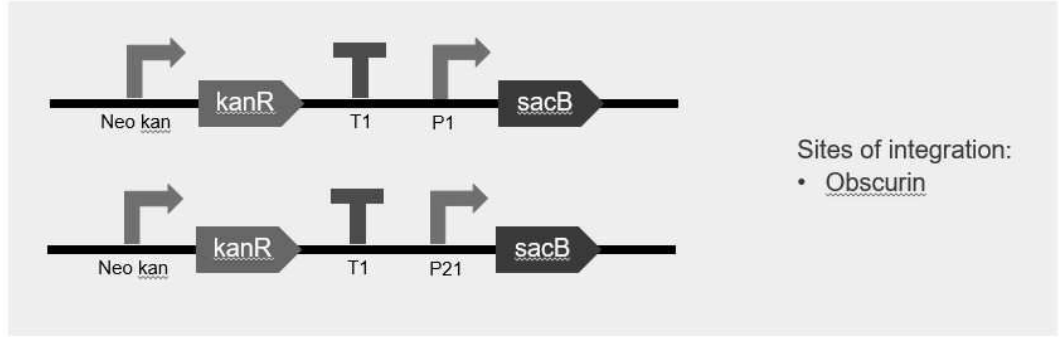
Strain ID	Details of RBSSWP
7000234944	RBS13 - spnM
7000225275	RBS14 - spnN
7000153593	Control (parent strain)

도면64a

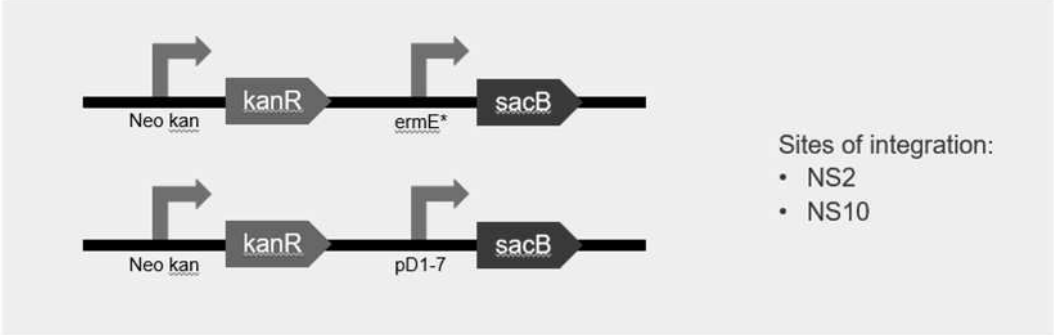

Neo kan
kanR
RBS linker
sacB
Sites of integration:
• Obscurin
pheS (SE)
ermE*
Neo kan
kanR
Sites of integration:
• NRPS
pheS (SS)
ermE*
Neo kan
kanR

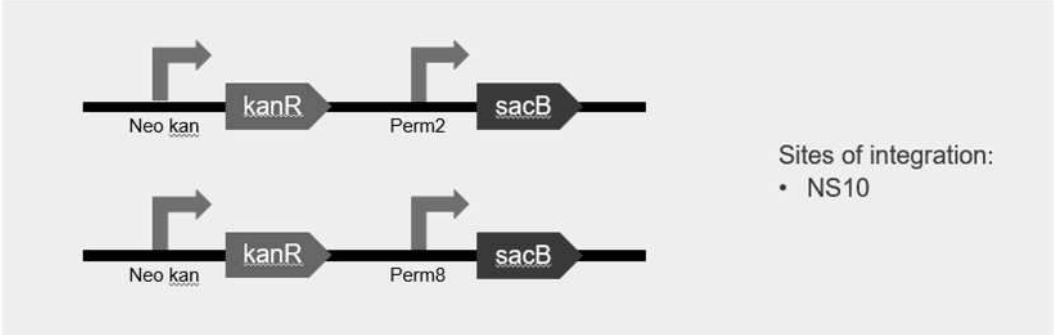
도면64b


Neo kan
kanR
A_T
P1
sacB
Sites of integration:
• Obscurin
• NS2
• NS10
Neo kan
kanR
A_T
P21
sacB

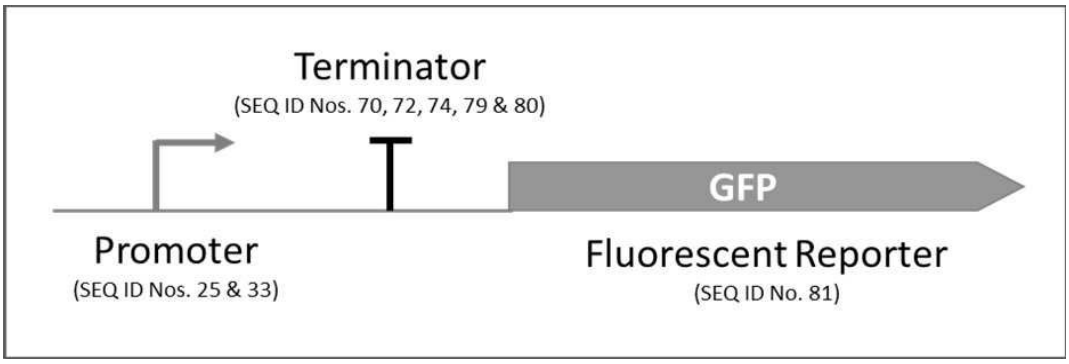

Neo kan
kanR
T1
P1
sacB
Sites of integration:
• Obscurin
Neo kan
kanR
T1
P21
sacB

도면64c

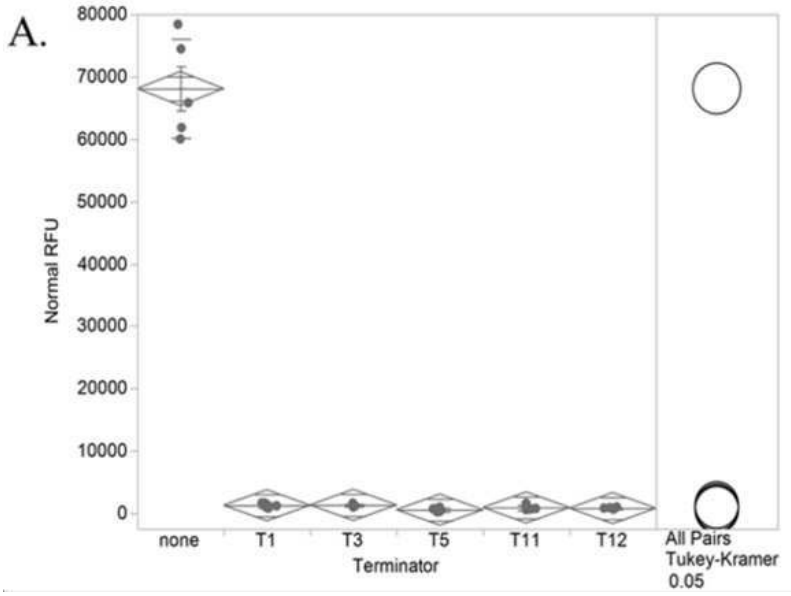

Neo kan
kanR
ermE*
sacB
Neo kan
kanR
pD1-7
sacB
Sites of integration:
• NS2
• NS10


Neo kan
kanR
Perm2
sacB
Neo kan
kanR
Perm8
sacB
Sites of integration:
• NS10

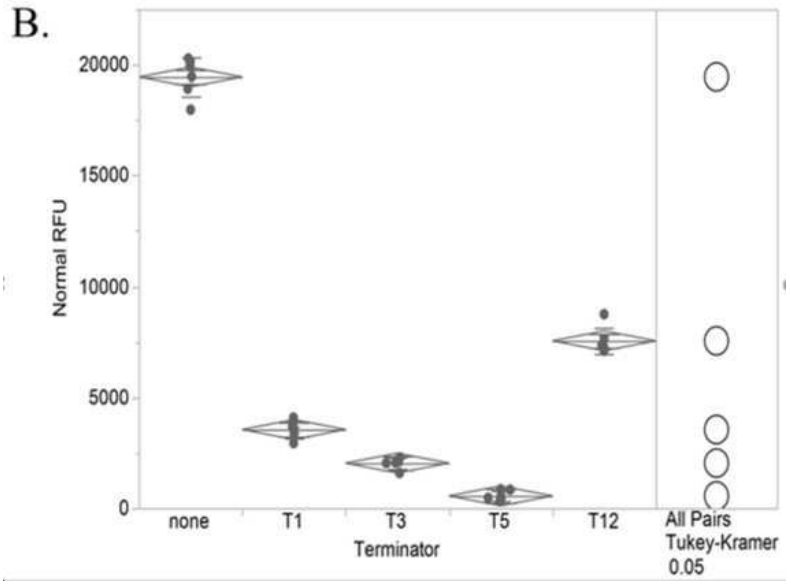
도면65


Terminator
(SEQ ID Nos. 70, 72, 74, 79 & 80)
GFP
Promoter
(SEQ ID Nos. 25 & 33)
Fluorescent Reporter
(SEQ ID No. 81)

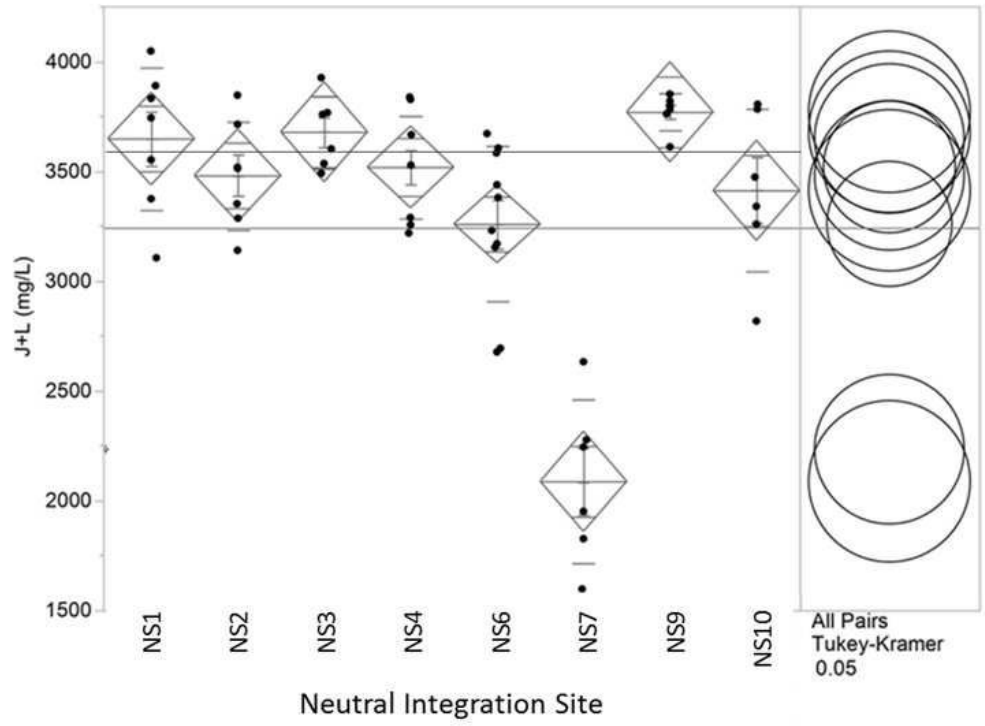
도면66a


A.
80000
70000
60000
50000
40000
30000
20000
10000
0
Normal RFU
none
T1
T3
T5
T11
T12
Terminator
All Pairs
Tukey-Kramer
0.05

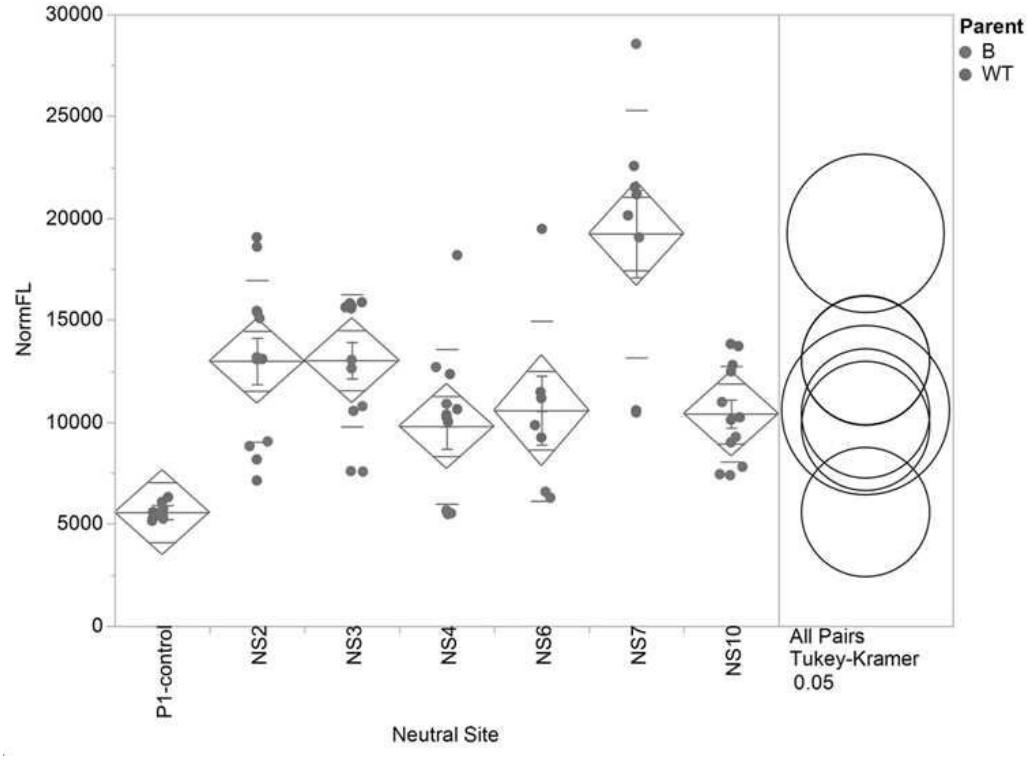
도면66b


B.
20000
15000
10000
5000
0
Normal RFU
none
T1
T3
T5
T12
Terminator
All Pairs
Tukey-Kramer
0.05

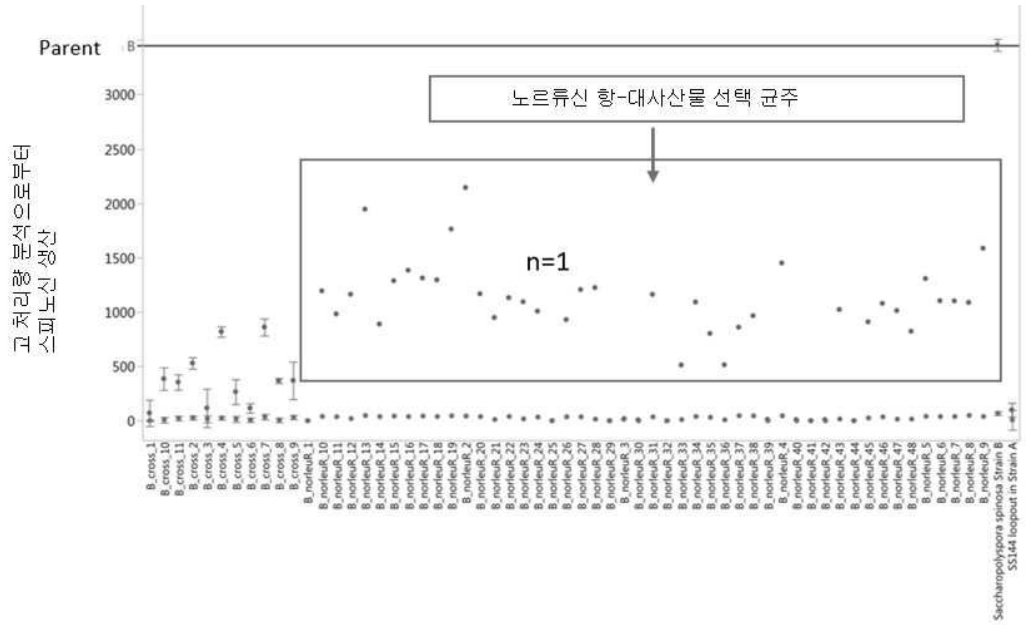
도면67


J+L (mg/L)
4000
3500
3000
2500
2000
1500
NS1
NS2
NS3
NS4
NS6
NS7
NS9
NS10
All Pairs
Tukey-Kramer
0.05
Neutral Integration Site

도면68


NormFL
30000
25000
20000
15000
10000
5000
0
Parent
B
WT
P1-control
NS2
NS3
NS4
NS6
NS7
NS10
All Pairs
Tukey-Kramer
0.05
Neutral Site

도면69



서 열 목 록

SEQUENCE LISTING

<110> Zymergen Inc.

Mason, Benjamin

Goranov, Alexi

Kelly, Peter

Kim, Youngnyun

Modi, Sheetal

Pasumarthi, Nihal

Mijts, Benjamin

Enyeart, Peter

<120> A HIGH-THROUGHPUT (HTP) GENOMIC ENGINEERING PLATFORM FOR IMPROVING SACCHAROPOLYSPORA SPINOSA

<130> ZYMR-013/01WO 327574-2059

<150> US 62/515,934

<151> 2017-06-06

<160> 175

<170> PatentIn version 3.5

<210> 1

<211> 141

<212> DNA

<

213> Saccharopolyspora spinosa

<400> 1

gccgcaccaa gcgagcaatg ccgccccggc ggtcccgacc gcgggacccc ggggcggtcg 60

cacgtccggg gcagcgggac ttgtcgatgg aacaggtacg gcctcaatag atcaggtacc 120

gatgaagggc tgttggaatc a 141

<210> 2

<211> 198

<212> DNA

<213> Saccharopolyspora spinosa

<400> 2

ggaccgagcg ggaggcaacg cctcgcgaag gcgaccgggg agcaatcccc tccagttcgg 60

cggcggacgg gccgccaccc cgcaaggaca gtgttcttcc gggatcggcg gcccgctcgt 120

cacctacccg acaggactcc gcctggcaca acaagtcgta cggcggaaag ttaacaagtc 180

caggaggaca atccagtg 198

<210> 3

<211> 232

<212> DNA

<213> Saccharopolyspora spinosa

<400> 3

gggactgttt gaaagtggct agcgtagcgg tgcgggtagc ggaacctcag aggccttctc 60

gctctgggat ccccgacatc atgaatgcca attcaccagg tcggggctgt cctcgcgaga 120

agacccctga gaacccgcgg cggtgcgagt tgagtcccac accgcaagcg gctaccgccg 180

cttataagac aggctctaac cgagtgaaag gcgctgagag ttgagcaccc tc 232

<210> 4

<211> 361

<212> DNA

<213> Saccharopolyspora spinosa

<400> 4

tagaaactgt tcatcgactg gctccgcgtg gcggtgcgga tagcggaacc taaaattctc 60

gctgtgggat ccccgacatc atgaatgcca ttcacctagt cgggggtgtc ctcgcgagat 120

ggctcggcgt gtggggtgtc ctcgcaggac actgttgaga acccgcggcg gcggtgcgag 180

ttgcgcgcgt gggctaagcg gcttcgccgc ttgaaagaca aagacgtagc gggagtgagt 240

gccagggcgg gccgttgtcc gctttggcac tcgcgtccgt ttcgggggcc ggtggtcggc 300

ggactgcggg gttctggttc gatcttgggt cgtagctccg ggtaattccc ggggatctac 360

c 361

<210> 5

<211> 303

<212> DNA

<213> Saccharopolyspora spinosa

<400> 5

tcgcccacag gacaggaaca cagcgtgtcg atgaaacgtc atactggtgt tggacgaaaa 60

cccagatgga gcagtaccga gcaaagtcga cttcgagtgg ggcatttcga gtgcggtcga 120

tgatcattga cgcgagtgga ccggggctca ttccccgcaa gctggtcttt cctgatcgat 180

tttgtgaccg acctcgtcga acgaacggcc gtactgtgtg tcaacctcgc gaatcgggcc 240

gctagcctgg tacctgagtg tgtctgtaaa tcacgagcat atggatcgag gcgaacgcca 300

gcc 303

<210> 6

<211> 253

<212> DNA

<213> Saccharopolyspora spinosa

<400> 6

ctggcgaggt gccgaccata actcgatcta cacgagcgtg gacttgcaac gttgaccgtt 60

tacatccgtg tagatatcct cgtcggagtc cggccaggga gatggcgtct gcacgtcgac 120

gatgctgccc ggcggaccgc tttgccagcc gaaggatgag cgccgtgacg gtgtcgtcaa 180

ccaaaccggc ccgctgtgcg acgaggcggc ttcccgcccg cggcacgggg tcacgcacga 240

aaggagtgcg gcg 253

<210> 7

<211> 168

<212> DNA

<213> Saccharopolyspora spinosa

<400> 7

tatccacgcg ctgtggttct cgtgggcacc ctctgcccac agcaaaagaa gatctttcgt 60

cgctgcgcag ttctagcgga aggatattgt gtagatgcgg atttctgatg taatttttct 120

tgacaagtga tgcggggcac cgcaccacac ggcgccggga ggccgaac 168

<210> 8

<211> 363

<212> DNA

<213> Saccharopolyspora spinosa

<400> 8

ctgctcgcca tactcaacga cattcttccg tcactggaaa cgacgaaggc gatgtcgctg 60

cccgaaagtg tggctacaat acgggacctc ggcatcgtaa tgggctccgt caagaggcac 120

ggcgttgaac cagtggtcgc gattcccgaa ctggaagcgc cgttgcgtgc cataggcgaa 180

cgcaccggca tgatcccacg ggacaccatc catcactaca tacggtggaa tccgacgggg 240

cggcgtgagc gcatgtacac gggcgagccg atggagaagt tgctcatggc ctctgtgcgc 300

atcagtttgc ccaggctgag cgctgctgtc gatgtgtgca ccgctctgca caccgccgag 360

gcc 363

<210> 9

<211> 300

<212> DNA

<213> Saccharopolyspora spinosa

<400> 9

gccttttgc cgaaatcgcg cggccagttc gtccaatgcg caagtcgcga tctccggcgg 60

cgcggtgcga agcggccgcg cggtgcaggt gaccccacag tctttccggg gcggttcttg 120

gcggggttcg gcgcgaaggg attcaagatc gttcttctga aacgtccaag acatcgatct 180

ttgtgccgct tttaacgtgt tcgactgcgt tgccgccctc cgcctcttgg cggaactggc 240

agtcttaagg tggaacctgt tggcacaatg aggtgccgtc aagcgtggag ctgcctgaag 300

<210> 10

<211> 322

<212> DNA

<213> Saccharopolyspora spinosa

<400> 10

gcgaacgcaa gccatttcgg cgcgggccag gccgtgtcgg gtcgggctgg gtgcgggcgg 60

gcttacctgc gcaactctgg ctgtgcaagg gatcgcttat ccgccatgcc atcaggttga 120

ccagctctcc gagtgcagtc gtcggctgcc ggcagggcgt tgacatcggg cgctttgacg 180

ccagcggggt gaggtgatgt gtacaagcgc cgttgccgac gggtggatct tgctcgcggc 240

ttcggcgcag acaggttcgg cgaaaactac tcttgcgtgc tcggataatc cgtgctcgga 300

ttacatgcgg aggtggtcaa cg 322

<210> 11

<211> 149

<212> DNA

<213> Saccharopolyspora spinosa

<400> 11

ctcaggatct ggacactaaa ttccatcttt tgggtgaaag ttgactggaa cgatttagaa 60

ggtgacggct ttgtgacggg gcattgctgt gaaatggttc tcacttatgt ttacgctcgt 120

ctgacgcggc ggtgaatgat ccgccgcgc 149

<210> 12

<211> 199

<212> DNA

<213> Saccharopolyspora spinosa

<400> 12

acgacggagc gaccctaaca tcgacacacc ggtcgcctcc cgtgacagca cgaccgaaga 60

atctaaagct gccctttta actagagaat tctgaacaaa aaggcaagat gtcaccctgg 120

tcacaatccg gccttccgcg cgcggcattg acgcggtaaa gtcccgggtc gccatcgaca 180

cgaggcaggg tgccctggc 199

<210> 13

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic promoter P21

<400> 13

tgtgcgggct ctaacacgtc ctagtatggt aggatgagca a 41

<210> 14

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic promoter PA9

<400> 14

ccgggcggct tcctcatgct tgacttgact aggataaagg g 41

<210> 15

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic promoter PA3

<400> 15

tagcagggct ccaaaactaa cgcctgatgt aggatcagat g 41

<210

> 16

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic promoter PB4

<400> 16

gctgtaggct gttaatatat ttcggtgtgt aggatacggg c 41

<210> 17

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic promoter PB12

<400> 17

cgggatggct tatgaaggat tgtctcactt aggatagagc a 41

<210> 18

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic promoter PB1

<400> 18

cgtcagggct actctggcaa ccaagcgatt aggattgaag g 41

<210> 19

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic promoter PC1

<400> 19

actttcggct aaaaagcaat tcattcaatt aggatggaag a 41

<210> 20

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic promoter P72

<400> 20

ctaattggct acgtcataga gagattcttt aggatgagaa a 41

<210> 21

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic promoter P-C4-1

<400> 21

ggcaactagg ttgacgtatt tttccgttag gcctagggtg agtg 44

<210> 22

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic promoter P-A5-19

<400> 22

tatgcgttgc ttgaccaaac ctatgtatag ggatagggtt ggtc 44

<210> 23

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic promoter P-C4-14

<400> 23

ccctcgctgg ttgacacagt tagtcagatt gcctacgatt tcgt 44

<210> 24

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic promoter P-D1-7

<400> 24

aattgcccac ttgacgttga gagtgaagca atataggtta acct 44

<210> 25

<211> 242

<212> DNA

<213> Saccharopolyspora spinosa

<400> 25

aaccactgcg acgagcggta tttgggggaa gtaaagaggt gacccacgac tcactgtcgg 60

tgatcgatat cgacccggaa acgacttgat aacgacgctc tgatcagcac aaataccccg 120

gatcgaagca cccaccccca ctgttactgt gatcagcgtc acatgatctc aggtttccga 180

tctccgtgtt ggttacgtag tgtcgctcct cggtcggccc cgaaccgatc agcaaggagc 240

gg 242

<210> 26

<211> 65

<212> DNA

<213> Saccharopolyspora spinosa

<400> 26

cagcaaacct gtggaccatc accaacaccg aaacgtctaa tgggcaagtc aaccttcgcg 60

gcgaa 65

<210> 27

<211> 201

<212> DNA

<213> Saccharopolyspora spinosa

<400> 27

tccggtggcc tggtcaaacg ccgccgggca cggcgtctgg tatctctgaa tgtgtgacga 60

tgacggcccc ctccccccac tcgctgttca acgatcttcc cttgcccggt acggtgagct 120

cggcgtggcg tgcgaggcac gcgatttccg ccggacaaat ccgaatcgct tgaacgcgta 180

acaccagggc tgctgtctgc g 201

<210> 28

<211> 300

<212> DNA

<213> Saccharopolyspora spinosa

<400> 28

gcggcaaacc gtcgcagaca ccccgaaacg tcgtgatcgt cctgccctac tgcccttgtg 60

aagatcgtct cggatcttcg ctcgtggtcc cacctccatt ccggtggcct ggtcaaacgc 120

cgccgggcac ggcgtctggt atctctgaat gtgtgacgat gacggccccc tccccccact 180

cgctgttcaa cgatcttccc ttgcccggta cggtgagctc ggcgtggcgt gcgaggcacg 240

cgatttccgc cggacaaatc cgaatcgctt gaacgcgtaa caccagggct gctgtctgcg 300

<210> 29

<211> 220

<212> DNA

<213>

Saccharopolyspora spinosa

<400> 29

ctgaatgcag ccgtaagtta ttggatcacc taggaatcgg gtcacttttc ccctgccgga 60

atgtgtgcct gcttacttag cgtgccttgt tcacctctcg ttcacttcga tggcggcgat 120

cgtccactcc gactccttag cgtccgtgtc gagcggccaa agcacgagcc tgcgcgaggc 180

tcggccgcgc aaccgcaggg tttccaactg gaggaacgaa 220

<210> 30

<211> 177

<212> DNA

<213> Saccharopolyspora spinosa

<400> 30

actttcccc tgccggaatg tgtgcctgct tacttagcgt gccttgttca cctctcgttc 60

acttcgatgg cggcgatcgt ccactccgac tccttagcgt ccgtgtcgag cggccaaagc 120

acgagcctgc gcgaggctcg gccgcgcaac cgcagggttt ccaactggag gaacgaa 177

<210> 31

<211> 200

<212> DNA

<213> Saccharopolyspora spinosa

<400> 31

gtgattccgg ctagactgct actttgcgct gccctctttc cgtctgtcct gcaccggacc 60

gtaggatggt gggcgccatt gcacccttga cagctgtgtt agcggagtgt gacagcggat 120

acggaccccg tcggtcgcat tcgccgggca cctttcgccg acgcggctgt agccagttca 180

gagtcccgga aggacgcatc 200

<210> 32

<211> 299

<212> DNA

<213>

Saccharopolyspora spinosa

<400> 32

tgatcgaagc gtgatctctt gactggcggc gcgcgcgggt tcactctagt cctcaacgcg 60

gggctgggct gccgtcggtg tgcccctcg acagctggcg tgattccggc tagactgcta 120

ctttgcgctg ccctctttcc gtctgtcctg caccggaccg taggatggtg ggcgccattg 180

caccttgac agctgtgtta gcggagtgtg acagcggata cggaccccgt cggtcgcatt 240

cgccgggcac ctttcgccga cgcggctgta gccagttcag agtcccggaa ggacgcatc 299

<210> 33

<211> 240

<212> DNA

<213> Saccharopolyspora spinosa

<400> 33

cgctccgccg catcggttac ggcgcttgca ctcgactggc gagagtgcta aacacggtat 60

tggcactcag caaggttgag tgccaggtcg ggacggtgag gccgtctccg gcggtgccac 120

cagacggcgc cgccgcacgg tcgtccgtcg cgggcaccga gcctggccga gcacgagtcc 180

tgccgtgggg tgcgcaaacc caccaccgcg gcgtccagac aggtggagga ccacaccgca 240

<210> 34

<211> 242

<212> DNA

<213> Saccharopolyspora spinosa

<400> 34

accggatcag cagttccacg ccgatctaat aaggaccaac tcggctcggc ggaagtccgg 60

taggagcgaa gttactgtcc tcagaggtct gaggtccagt ggaagaggcg acgaaacaag 120

gagattcgtc tctcaccgta aagagtgaaa aaatctagcg aggcggctga cggctttcgg 180

ttcgacttgc gagtcggcta ggttcgtgat cacgaactcc gattgaaggt cctaacagga 240

gt 242

<210> 35

<211> 300

<212> DNA

<213> Saccharopolyspora spinosa

<400> 35

ctccagccac agccggagca tctccctgat ctcgatcacc gaaacctccc ggtagctcat 60

tccgccaata cgagcggccg agcgaccgga aaccccgcag ccacgccggg cggtcccata 120

actggcaaac caccagccca gacggtccca tcagtggcaa acaggtggtc ccatgctcct 180

ggcaaaaccg gctctgaggt ggtcccttac tcctggcagc cgacaggcgt cgttgcgcag 240

tggcggacag ggcacggatg actcctatga ggtagtcgat tacgtgctac cgtctacgcc 300

<210> 36

<211> 359

<212> DNA

<213> Saccharopolyspora spinosa

<400> 36

ggcggactaa gacttaggtc tacctaagcg cggtgcaact taacccagcg ttccgcgcgg 60

cacaagcacc ttgttcgaaa ccgggttcca ggtcacctgg aaacccattt gagaccggat 120

cgacgcactt tcccacgcct atacgggtat ctggcgaatg gcggaatctg actttgggtt 180

cggcagtgac ctggacttat atgtcgatgt gcgcatcagt cgacgtgata ctcgcgacta 240

accgcaggtg atttgccgaa cgggtctgcg tatttccctg cggcgagtta gggtgccctt 300

gcttgccttg aacattgctc tacctcatca ggactccttc gaagggaagt gagctgctc 359

<210> 37

<211> 433

<212> DNA

<213> Saccharopolyspora spinosa

<400> 37

aaccactgcg acgagcggta tttgggggaa gtaaagaggt gacccacgac tcactgtcgg 60

tgatcgatat cgacccggaa acgacttgat aacgacgctc tgatcagcac aaataccccg 120

gatcgaagca cccacccca ctgttactgt gatcagcgtc acatgatctc aggtttccga 180

tctccgtgtt ggttacgtag tgtcgctcct cggtcggccc cgaaccgatc agcaaggagc 240

ggaagcccgc agcgccgaac cctgtccagc aggcttccag accccgaaac gaagaacacc 300

ggacagggac gggggaccca acacccgggc tccccgaagc cctaggggtg aagccggctc 360

ccccgagccg gccgggctgc ctctcagccc gaacccgaca gctcacctcg caggcgcggc 420

aggagagagg aac 433

<210> 38

<211> 115

<212> DNA

<213> Saccharopolyspora spinosa

<400> 38

gggcgcgccg caattcgatg acgttcatgc gccgtgtcgg ggaatcgccg gtggcggcgc 60

cagcagagac tgaacttact ggtggtgtgt ccaggaatcg gaggggcagt accga 115

<210> 39

<211> 496

<212> DNA

<213>

Saccharopolyspora spinosa

<400> 39

acttcggacc ccagtctctt tcccccgatt agcgcagcag cccctactcc cattggccag 60

gatttggaaa atgcgctgcg tatgtcgatc gccgttgacg tccaacggac ttccggcggc 120

aacaatagtg tgtcacggca ggaatgtcac gcgaccatcg aagatctttg ggtcgccgca 180

cctggtttca cgcgaacgag tgaaatgcgc gagctccgct cgatcggggt gggccggacc 240

tgtacggtga tcaccgttgg ttctgcgggg attcatgggg aagatttgcg ctggctgttt 300

gcctcctggc cggatagtta tagtcggtac cgccgcatgc ggcggtaacc gcgaattaac 360

tgacggctag tttgccgtct tttctctctg tgtgtttcct gctcggttcc agaaaattac 420

gagaaggtga acgttgcaga gatcaggcat accggtgttg ccaggtggcg caccaacatc 480

gcagcaggtt gggcag 496

<210> 40

<211> 202

<212> DNA

<213> Saccharopolyspora spinosa

<400> 40

cactcgttcg cgtgaaacca ggtgcggcga cccaaagatc ttcgatggtc gcgtgacatt 60

cctgccgtga cacactattg ttgccgccgg aagtccgttg gacgtcaacg gcgatcgaca 120

tacgcagcgc attttccaaa tcctggccaa tgggagtagg ggctgctgcg ctaatcgggg 180

gaaagagact ggggtccgaa gt 202

<210> 41

<211> 379

<212> DNA

<213> Saccharopolyspora spinosa

<400> 41

gaactctccg atcgcaattg aacacccggg aagcatgcca agaatcacag aaatctctga 60

tatcccccgg gaaacgccgc tttcgcaagc caaatcttag gccttccagg tgatggtagc 120

gatcttgaca agcgcgagca ggtcgttccc gctagcctgg gctctaccga gtcgggtgtg 180

ccgggtagat cgaggatttc tgagtcaatg agcgcttctc cttgctccgc tgtcctgatg 240

tcccgcaccg catcgaacca gggcaggaag gtgtaaggcg ccgagacagc acactgtccc 300

gctgggacgt cataacgcga ttcgccacgg gcatcgctca tctcctgaag gcaaggcgcg 360

aagactgatc gtcgcctgc 379

<210> 42

<211> 261

<212> DNA

<213> Saccharopolyspora spinosa

<400> 42

gcgatcttga caagcgcgag caggtcgttc ccgctagcct gggctctacc gagtcgggtg 60

tgccgggtag atcgaggatt tctgagtcaa tgagcgcttc tccttgctcc gctgtcctga 120

tgtcccgcac cgcatcgaac cagggcagga aggtgtaagg cgccgagaca gcacactgtc 180

ccgctgggac gtcataacgc gattcgccac gggcatcgct catctcctga aggcaaggcg 240

cgaagactga tcgtcgcctg c 261

<210> 43

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> P21_mutant synthetic promoter

<220><221> misc_feature

<222> (7)..(10)

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

<220><221> misc_feature

<222> (30)..(35)

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

<400> 43

tgtgcgnnnn ctaacacgtc ctagtatggn nnnnngagca a 41

<210> 44

<211> 40

<212> DNA

<213> Saccharopolyspora spinosa

<400> 44

cccggatcga agcacccacc cccactgtta ctgtgatcag 40

<210

> 45

<211> 209

<212> DNA

<213> Saccharopolyspora spinosa

<400> 45

aaccactgcg acgagcggta tttgggggaa gtaaagaggt gacccacgac tcactgtcgg 60

tgatcgatat cgacccggaa acgacttgat aacgacgctc tgatcagcac aaataccccg 120

gatcgaagca cccacccca ctgttactgt gatcagcgtc acatgatctc aggtttccga 180

tctccgtgtt ggttacgtag tgtcgctcc 209

<210> 46

<211> 433

<212> DNA

<213> Saccharopolyspora spinosa

<400> 46

aaccactgcg acgagcggta tttgggggaa gtaaagaggt gacccacgac tcactgtcgg 60

tgatcgatat cgacccggaa acgacttgat aacgacgctc tgatcagcac aaataccccg 120

gatcgaagca cccacccca ctgttactgt gatcagcgtc acatgatctc aggtttccga 180

tctccgtgtt ggttacgtag tgtcgctcct cggtcggccc cgaaccgatc agcaaggagc 240

ggaagcccgc agcgccgaac cctgtccagc aggcttccag accccgaaac gaagaacacc 300

ggacagggac gggggaccca acacccgggc tccccgaagc cctaggggtg aagccggctc 360

ccccgagccg gccgggctgc ctctcagccc gaacccgaca gctcacctcg caggcgcggc 420

aggagagagg aac 433

<210> 47

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> P21-P1 synthetic promoter

<400> 47

tgtgcgggct ctaacacgtc gaagtatggt aggatgagtg ttactgtgat cag 53

<210> 48

<211> 52

<212> DNA

<213> Artificial Sequence

<220><223> P1-P21 synthetic promoter

<400> 48

cccggatcga agcacccggg ctctactgtt actgtgatgg taggatgagc aa 52

<210> 49

<211> 332

<212> DNA

<213> Saccharopolyspora spinosa

<400> 49

gggcaggccc agcttctcgc cgcgccagtc gggccgcgcc tcggcggcct gctcgcgcgc 60

gtggttgaat gcgctcttcg ggccgtccag ccacgaaccg atgactcttc gcacgtcatc 120

cagggtacgt gcccctgcca tacggccagc ggagcatcac gctcggccgg tgcgcgcaac 180

cccgaccacc caaacgggcg gcagttaaca cccacgaaac attcaggtga cgacagggca 240

acacccctaa cataacgtgg actacgagcc ggcggtggaa cctttggcgt tgcgtcggtg 300

agctgtacga gtgcgtgaag gagccaccga gg 332

<210> 50

<211> 300

<212> DNA

<213> Saccharopolyspora spinosa

<400> 50

cacgagcgcg ctcgacggac cagctcaccg agaagtaggt cggagcaccg ttagcgggaa 60

aagtggggtt atcggcgttg cactaagcac gatggaccat ttgaggtaat gcgatgtagc 120

ccaaccggct ggttggcgtg ttgatgttgc ggttgaatgc cgcgttacgc gtcccgggca 180

aattcgactt aaatgtcgcc tgtatcacaa ttctgttact tctgacggac tgtcgcttag 240

agtacctctc cgggttcagc cagcgataaa tagtcgctgg ctccgtctgg ggggatggga 300

<210> 51

<211> 247

<212> DNA

<213> Saccharopolyspora spinosa

<400> 51

cccggcaatt gttgggccca cagcaaaaat catgctcaag ggctcgccct gtaaccgggg 60

gacgattgtt tgtgggtggt gtgttgtggt gcgggtgccg ggtggttgtc ctcgcccggt 120

cgctgggttg tggtgtgtcg gcgcgctggt gggtcgcggt gcgggcatct agtggggcgg 180

aaggcctgat ttcggttgct gttggtggtg ttctgggttc tggcggcgtt ggggtcgggt 240

gggtgtg 247

<210> 52

<211> 230

<212> DNA

<213> Saccharopolyspora spinosa

<400> 52

gctggcggag gggcacgagc gggcgatcgc agcgggtgcg acgccgctgc cgcagccgcc 60

cgaccagcag ggcgcgagct tccgcgtcta cgccgacccc gatggtcacc cgttctgcat 120

gtgcgcctgc gaggagtgag cgcgctctcc ggtcaggggc gcaagcagtc tgcttgcagg 180

agctagcact tgtggttatc gtcgttggtg accggagagg tgctagcccg 230

<210> 53

<211> 300

<212> DNA

<213> Saccharopolyspora spinosa

<400> 53

gcggcaaacc gtcgcagaca ccccgaaacg tcgtgatcgt cctgccctac tgcccttgtg 60

aagatcgtct cggatcttcg ctcgtggtcc cacctccatt ccggtggcct ggtcaaacgc 120

cgccgggcac ggcgtctggt atctctgaat gtgtgacgat gacggccccc tccccccact 180

cgctgttcaa cgatcttccc ttgcccggta cggtgagctc ggcgtggcgt gcgaggcacg 240

cgatttccgc cggacaaatc cgaatcgctt gaacgcgtaa caccagggct gctgtctgcg 300

<210> 54

<211> 300

<212> DNA

<213> Saccharopolyspora spinosa

<400> 54

tgtcgccgtg ggtagagtgt ctggacgagg aagacgtgag ctcgcaccca ggagatgggc 60

gccaccgcag cgccgggtga tcgagcaggt caaggggacc ccgttttgac cctctgagaa 120

cggggcaggt atctttgttt atcggacccg acacccatcg ggtcgatccg catgcccgtg 180

catgacgtgg tcgccagccg ctggttgaga ccgcgccagc gctgaggacg tgtggaactc 240

ctctcaacaa ccctctgggg tcgttcccat tggggcgcat cggcgccgaa aggccgagga 300

<210> 55

<211> 298

<212> DNA

<213>

Saccharopolyspora spinosa

<400> 55

acccgcgagg cgcccgaacc attgattgcg caaatttttc acaatccgcg ttcgtattac 60

gtcgcttggc cacgctccgc cgttacggag atagctcata gtcacccaaa agagcgatac 120

gatcatgttc aggtaacaac tcgatcggga tagatacccg attgatcgtc ccgctgaccc 180

gcttgggcgg ttacgctgcc cccgacgaca ccactttcgg tcgatcagtg gcgcggctga 240

ttggtcgaga gtcccggtgc cggtcggggg gcacgaccga ctccagggag gtagtgac 298

<210> 56

<211> 300

<212> DNA

<213> Saccharopolyspora spinosa

<400> 56

ctccagccac agccggagca tctccctgat ctcgatcacc gaaacctccc ggtagctcat 60

tccgccaata cgagcggccg agcgaccgga aaccccgcag ccacgccggg cggtcccata 120

actggcaaac caccagccca gacggtccca tcagtggcaa acaggtggtc ccatgctcct 180

ggcaaaaccg gctctgaggt ggtcccttac tcctggcagc cgacaggcgt cgttgcgcag 240

tggcggacag ggcacggatg actcctatga ggtagtcgat tacgtgctac cgtctacgcc 300

<210> 57

<211> 300

<212> DNA

<213> Saccharopolyspora spinosa

<400> 57

ggtggcactt accgatccgt cggggacgcg tccttcgagc cagtgcttcg gtacgccgcc 60

gagccagcgc accggatccg gttcgccggt gactacttcg cgccgctggg gcagatggag 120

gtagccgtga ccgcgggtaa agacgcggcc gaggcggtca tccgtgatcg cgcgggagcg 180

catcagcgtc gcttcggctc ggtcaaaccc aggtgaaccc cttgacctac ctcagggaag 240

cagcgatatt catacgtaga cggtagtcga ttaccgatca gataccaccc tggaggaaga 300

<210> 58

<211> 248

<212> DNA

<213> Saccharopolyspora spinosa

<400> 58

ggccggattc ctcggcggac aggcgcagcc ccggctgatc gagaggatcg gggtgggggc 60

cgaggccgcg atcatcccgg cgttcatggg aggacatcgg tgtgctcacc gatcccgagc 120

gccgggtgtc gctgatcatg cgcgacggcg tgctggtgaa ggaccgcccg acggtgtgat 180

atttgcctgt taaccccgct ttcatcccag gtcaagcgcc tgcaatacag cgcgggaagc 240

atggtggc 248

<210> 59

<211> 250

<212> DNA

<213> Saccharopolyspora spinosa

<400> 59

cggcgcgagc actccagcac cccgagccgg ggctcgtcgg agagcaacag cagcaaccgg 60

gcgtccaacg cgtcgagccc ctcagcattg ggagccatgt catatccctt gttcaggctg 120

accaataaag ccagcgattt ggatcaaata cttatcattt tgtgcagcga aatcaaatac 180

tgttgctcag gatgacctgc ccggcgcacg ctgaccccgt cactgcttcg gcgaaacagg 240

ggaggacatc 250

<210> 60

<211> 250

<212> DNA

<213> Saccharopolyspora spinosa

<400> 60

tgagaccgga tcgacgcact ttcccacgcc tatacgggta tctggcgaat ggcggaatct 60

gactttgggt tcggcagtga cctggactta tatgtcgatg tgcgcatcag tcgacgtgat 120

actcgcgact aaccgcaggt gatttgccga acgggtctgc gtatttccct gcggcgagtt 180

agggtgccct tgcttgcctt gaacattgct ctacctcatc aggactcctt cgaagggaag 240

tgagctgctc 250

<210> 61

<211> 298

<212> DNA

<213> Saccharopolyspora spinosa

<400> 61

tcgctttgac gcggacaccc cggccttttt gtcgctcgct agaggtaaaa acgacacacg 60

atcgagcgat tacctttttgt gtaacactcc aaactcaacg gtttccccga cccgtttgtc 120

ctgatcaagc ggcagatgcg ggtgatcggg atcaggtcgg tccgcgtcgg tggccgcaga 180

gtgcgcgtta gcgtcccggc ggcacttgat ctcgcagatc aggctgcccg ctgcaggtgc 240

ttcccggcct tcgccggagc tgccgacaac aggtacgcca acgggccagg agctgatc 298

<210> 62

<211> 296

<212> DNA

<213>

Saccharopolyspora spinosa

<400> 62

caagaagccg aaaggcggcg aactctcggc ggcagataag aaaaacaaca aaacgatctc 60

atcgctacga tctgccgtcg agcgatgcat cgcacattta aagaattgga agatacttgc 120

caccgggtac cgaggacggc tcgctgaact ctccaacatc atccgcatcg tcacggcgct 180

cgaattctat cgactcggct ggtaactcac gtgaataacg ctcttcgtgt tcagcaaacc 240

tgtggaccat caccaacacc gaaacgtcta atgggcaagt caaccttcgc ggcgaa 296

<210> 63

<211> 300

<212> DNA

<213> Saccharopolyspora spinosa

<400> 63

gcggcaaacc gtcgcagaca ccccgaaacg tcgtgatcgt cctgccctac tgcccttgtg 60

aagatcgtct cggatcttcg ctcgtggtcc cacctccatt ccggtggcct ggtcaaacgc 120

cgccgggcac ggcgtctggt atctctgaat gtgtgacgat gacggccccc tccccccact 180

cgctgttcaa cgatcttccc ttgcccggta cggtgagctc ggcgtggcgt gcgaggcacg 240

cgatttccgc cggacaaatc cgaatcgctt gaacgcgtaa caccagggct gctgtctgcg 300

<210> 64

<211> 226

<212> DNA

<213> Saccharopolyspora spinosa

<400> 64

gggtgatcga gcaggtcaag gggaccccgt tttgaccctc tgagaacggg gcaggtatct 60

ttgtttatcg gacccgacac ccatcgggtc gatccgcatg cccgtgcatg acgtggtcgc 120

cagccgctgg ttgagaccgc gccagcgctg aggacgtgtg gaactcctct caacaaccct 180

ctggggtcgt tcccattggg gcgcatcggc gccgaaaggc cgagga 226

<210> 65

<211> 191

<212> DNA

<213> Saccharopolyspora spinosa

<400> 65

acggccagcg gagcatcacg ctcggccggt gcgcgcaacc ccgaccaccc aaacgggcgg 60

cagttaacac ccacgaaaca ttcaggtgac gacagggcaa cacccctaac ataacgtgga 120

ctacgagccg gcggtggaac ctttggcgtt gcgtcggtga gctgtacgag tgcgtgaagg 180

agccaccgag g 191

<210> 66

<211> 177

<212> DNA

<213> Saccharopolyspora spinosa

<400> 66

acttttcccc tgccggaatg tgtgcctgct tacttagcgt gccttgttca cctctcgttc 60

acttcgatgg cggcgatcgt ccactccgac tccttagcgt ccgtgtcgag cggccaaagc 120

acgagcctgc gcgaggctcg gccgcgcaac cgcagggttt ccaactggag gaacgaa 177

<210> 67

<211> 299

<212> DNA

<213> Saccharopolyspora spinosa

<400> 67

tgatcgaagc gtgatctctt gactggcggc gcgcgcgggt tcactctagt cctcaacgcg 60

gggctgggct gccgtcggtg tgccccctcg acagctggcg tgattccggc tagactgcta 120

ctttgcgctg ccctctttcc gtctgtcctg caccggaccg taggatggtg ggcgccattg 180

caccttgac agctgtgtta gcggagtgtg acagcggata cggaccccgt cggtcgcatt 240

cgccgggcac ctttcgccga cgcggctgta gccagttcag agtcccggaa ggacgcatc 299

<210> 68

<211> 266

<212> DNA

<213> Saccharopolyspora spinosa

<400> 68

atcgggacac tacgccgcgc actcgcgtgc acggaccact gggctcgttt tcccggacaa 60

cccgaaacca cgcgccaatt tgccgacgca acgagaactg tgcgacacga cacatccgga 120

tcacacgatg gtgttaacga ggtgtttttt cttcggtttg cctggatatc tttcacgccg 180

agtacgcccc ctcccggctg aacttatggg gtgcgcagcc ggggaaggag cgctctgctt 240

catcccatcc aacaaggagc aacaaa 266

<210> 69

<211> 163

<212> DNA

<213> Saccharopolyspora spinosa

<400> 69

ttatatgtcg atgtgcgcat cagtcgacgt gatactcgcg actaaccgca ggtgatttgc 60

cgaacgggtc tgcgtatttc cctgcggcga gttagggtgc ccttgcttgc cttgaacatt 120

gctctacctc atcaggactc cttcgaaggg aagtgagctg ctc 163

<210> 70

<211> 37

<212> DNA

<213> Saccharopolyspora spinosa

<400> 70

cccgaacctt cgggggcggg ccctcttgct tttcaat 37

<210> 71

<211> 49

<212> DNA

<213> Saccharopolyspora spinosa

<400> 71

cgggcaataa tacgtgcccg gacggtagtg cgagcacgag gtgggtacg 49

<210> 72

<211> 41

<212> DNA

<213> Saccharopolyspora spinosa

<400> 72

agtttgtcga accggcggcg ttcgccggct ttaccttgcg c 41

<210> 73

<211> 42

<212> DNA

<213> Saccharopolyspora spinosa

<400> 73

ggtttctcga accagtgctt tgcgtactgg ttgtcgttgc ag 42

<210> 74

<211> 37

<212> DNA

<213> Saccharopolyspora erythraea

<400> 74

cggagccaga gggcgcctga gtgcctgttt ttgatcc 37

<210> 75

<211> 39

<212> DNA

<213> Saccharopolyspora erythraea

<400> 75

aaacgccccc ggctccggcc gggggcgttt ttggttgtg 39

<210> 76

<211> 37

<212> DNA

<213> Saccharopolyspora erythraea

<400> 76

agacgcagga ggtctcgtga ggggcttttc cgcgagc 37

<210> 77

<211> 35

<212> DNA

<213> Saccharopolyspora erythraea

<400> 77

cgtgtgactt gtcccactcg gggtttttgt cgcga 35

<210> 78

<211> 39

<212> DNA

<213> Saccharopolyspora erythraea

<400> 78

ggattcgtcc ggccgaggcc aatcggcttt tcggggccc 39

<210> 79

<211> 38

<212> DNA

<213> Saccharopolyspora erythraea

<400> 79

gctttcgtcg gccgggaacg ccctggtgtt tcttaccg 38

<210> 80

<211> 38

<212> DNA

<213> Saccharopolyspora erythraea

<400> 80

ttgggtggat tcacccctac cgggtgtttt tctcggct 38

<210> 81

<211> 711

<212> DNA

<213> Artificial Sequence

<220><223> codon optimized reporter gene DasherGFP

<400> 81

atgacggcat tgacggaagg tgcaaaactg tttgagaaag agatcccgta tatcaccgaa 60

ctggaaggcg acgtcgaagg tatgaaattt atcattaaag gcgagggtac cggtgacgcg 120

accacgggta ccattaaagc gaaatacatc tgcactacgg gcgacctgcc ggtcccgtgg 180

gcaaccctgg tgagcaccct gagctacggt gttcagtgtt tcgccaagta cccgagccac 240

atcaaggatt tctttaagag cgccatgccg gaaggttata cccaagagcg taccatcagc 300

ttcgaaggcg acggcgtgta caagacgcgt gctatggtta cctacgaacg cggttctatc 360

tacaatcgtg tcacgctgac tggtgagaac tttaagaaag acggtcacat tctgcgtaag 420

aacgttgcat tccaatgccc gccaagcatt ctgtatattc tgcctgacac cgttaacaat 480

ggcatccgcg ttgagttcaa ccaggcgtac gatattgaag gtgtgaccga aaaactggtt 540

accaaatgca gccaaatgaa tcgtccgttg gcgggctccg cggcagtgca tatcccgcgt 600

tatcatcaca ttacctacca caccaaactg agcaaagacc gcgacgagcg ccgtgatcac 660

atgtgtctgg tagaggtcgt gaaagcggtt gatctggaca cgtatcagta a 711

<210> 82

<211> 792

<212> DNA

<213> Artificial Sequence

<220><223> codon optimized reporter gene PaprikaRFP

<400> 82

atggtgagca agggtgagga actgattaaa gagaatatgc gcatgaagct gtacatggaa 60

ggcacggtga ataaccacca cttcaaatgc accagcgagg gtgagggtaa accgtatgaa 120

ggcacccaaa cgatgcgtat caaagttgtt gagggtggcc cgttgccgtt tgcgttcgac 180

attttagcga cgagctttat gtatggctct cgtacgttta tcaagtaccc gaagggtatt 240

ccggactttt tcaaacaatc ttttccagag ggtttcacct gggagcgcgt gactcgctac 300

gaagatggcg gcgtcgtgac cgttatgcag gatacctccc tggaagatgg ctgcctggtc 360

taccacgttc aggtccgtgg tgtcaatttc ccgagcaatg gtccggttat gcagaagaaa 420

accaagggtt gggaaccgaa caccgagatg ttgtatcctg cagatggtgg cctggaaggt 480

cgcagcgaca tggcattgaa actggtcggt ggcggccatc tgagctgtag cttcgtgacc 540

acgtatcgtt cgaagaagcc ggcgaagaac ctgaaaatgc cgggtattca cgcggttgac 600

caccgtctgg agcgcctgga agaatccgac aacgagatgt tcgtggtgca aagagaacat 660

gccgttgcgc gttattgtga tctgccgagc aagctgggcc ataagctgaa cagcggtctg 720

cgtagccgcg ctcaggccag caattccgcg gtcgatggta ccgctggtcc gggtagcacg 780

ggtagccgtt aa 792

<210> 83

<211> 1812

<212> DNA

<213> Artificial Sequence

<220><223> codon optimized reporter gene gusA

<400> 83

atgctgcgcc ccgtggaaac cccgacgcgc gaaatcaaga agctggacgg cctctgggcc 60

ttctccctgg accgggagaa ctgcgggatc gaccagcgct ggtgggagtc cgccctgcag 120

gagtcgcgcg ccatcgccgt gccggggagc ttcaacgacc agttcgcgga cgccgacatc 180

cgcaactacg cgggcaacgt gtggtaccag cgcgaggtct tcatcccgaa gggctgggcg 240

ggccagcgga tcgtcctgcg cttcgacgcc gtgacccact acggcaaggt ctgggtgaac 300

aaccaggagg tcatggaaca ccagggcggg tacaccccgt tcgaggccga cgtcaccccg 360

tacgtcatcg ccggcaagag cgtccgcatc accgtctgcg tcaacaacga gctgaactgg 420

cagacgatcc cccccggcat ggtcatcacc gacgagaacg ggaagaagaa gcagagctac 480

ttccacgact tcttcaacta cgccggcatc caccgctcgg tgatgctgta cacgaccccc 540

aacacctggg tcgacgacat cacggtggtg acccacgtcg cccaggactg caaccacgcc 600

agcgtggact ggcaggtggt ggccaacggc gacgtctccg tggagctccg cgacgcggac 660

cagcaggtcg tcgccaccgg ccaggggacc tcgggcaccc tgcaggtggt caacccgcac 720

ctctggcagc ccggcgaggg ctacctctac gagctgtgcg tcacggcgaa gtcgcagacc 780

gagtgcgaca tctacccccct gcgcgtcggc atccggtccg tggccgtcaa gggcgagcag 840

ttcctgatca accacaagcc cttctacttc accggcttcg gccgccacga ggacgccgac 900

ctccggggca agggcttcga caacgtcctg atggtccacg accacgcgct gatggactgg 960

atcggcgcca actcctaccg cacctcccac tacccctacg ccgaggagat gctcgactgg 1020

gccgacgagc acgggatcgt cgtgatcgac gagaccgccg ccgtcggctt caacctctcg 1080

ctcgggatcg gcttcgaagc ggggaacaag cccaaggagc tctactccga ggaagccgtc 1140

aacggcgaga cccagcaggc ccacctgcag gcgatcaagg agctgatcgc gcgcgacaag 1200

aaccacccga gcgtcgtcat gtggagcatc gccaacgaac cggacacgcg cccgcagggt 1260

gcgcgggaat acttcgcccc gctcgccgaa gccacccgca agctcgaccc cacgcgcccc 1320

atcacctgcg tcaacgtgat gttctgcgac gcgcacaccg acaccatctc cgacctgttc 1380

gacgtcctgt gcctgaaccg ctactacggc tggtacgtcc agtccgggga cctggagacg 1440

gcggaaaagg tgctggagaa ggagctcctg gcgtggcagg agaagctgca ccagcccatc 1500

atcatcacgg agtacggggt cgacaccctg gccggcctcc actccatgta cacggacatg 1560

tggagcgagg agtaccagtg cgcctggctg gacatgtacc accgcgtctt cgaccgcgtg 1620

agcgcggtcg tgggcgaaca ggtctggaac ttcgccgact tcgcgacgtc gcagggcatc 1680

ctgcgcgtgg ggggcaacaa gaagggcatc ttcacccgcg accgcaagcc caagtccgcc 1740

gccttcctcc tgcagaagcg gtggaccggg atgaacttcg gcgagaagcc ccagcagggg 1800

ggcaagcagt ga 1812

<210> 84

<211> 1740

<212> DNA

<213> Saccharopolyspora endophytica

<400> 84

atgccgcgta agaaccgcga tgaaggcacc cgggcgccca acggcgcgag cagcatctac 60

aagggcaaag acggctactg gcacggccgc gtctggatgg gcaccaagga cgacggcagt 120

gaggaccgtc gccacaggtc agcgaagagc gaaacagagc tcctcaataa ggttcgcaag 180

ctcgaacggg agcgggacag cggcaaggtg cagaagcctg gccgcgcctg gaccgtcgag 240

aaatggctta cgcactgggt ggagaacatc gccgctccca ccgtgcggcc gaccacgatg 300

gtcggctacc gcgcctcggt gtataagcat ctgatccccg gcgtgggcaa gcaccggatc 360

gacaggttgc agccggaaca cctcgaaaag ctctacgcca agatgcagcg cgatggactc 420

aaggccgcga cagcgcacct cgcgcaccgg acggtgcggg tcgcgctgaa cgaggccaag 480

aagcgacgtc acatcaccga gaacccggcc aatatcgcga agccgcccag ggtggacgag 540

gaggagattg tcccgttcac ggtggatgaa gcccgccgga tcctcgcagc agctgcggag 600

acgcggaacg gcgctcgctt tgtcatcgcg ctgacccttg gcctgcgcag gggtgaagca 660

ctcgggttga agtggtcgga tctctcgatc acctggaagc acggatgccg gaaggggagc 720

gcgtgccggg tgggtcgccg agccgagcag tgcggcgagc gtcgcggcag cggcacgctc 780

gtcatccggc gcgcgattca gcagcaggtt tggcagcacg gttgctcaga ggacaagccg 840

tgcgaccacc gctacggcgc tcactgcccg cgccggcata gcggcggtgt ggtcgtgacc 900

gatgtgaagt ccagggcggg tcggcgaacc gtgggccttc cgcacccggt ggtggaagcg 960

ctcgaagagc accgcgcccg ccagcggaca gagcgggaga aggcgcgcaa cgagtgggac 1020

gacgccgatt gggtcttcac gaacaggtgg ggtcgcccgg ttcatccgac cgttgactac 1080

gacgcctgga aggcactgct cagggcagcg aacgtgcgca acgcgcggtt gcacgacgca 1140

cgccacaccg cggcgacgat gttgctggtg ttgaaggttc cgctgcctgc ggtcatggaa 1200

atcatgggct ggtcggaagc ctctatggcc aagcgctaca tgcacgtgcc gcacgagctc 1260

gtgaccgcga tcgcggacca ggtgggtgac ctggtgtggc ccgtcccaga gaccgaggag 1320

gaggcgccac cgcctgagga ggagtgggcg ctggacgcca accaggtggc ggcgatccgg 1380

aagctggccg gagctctccc gccgcagttg cgggagcagt tcgaggcgct gctgcccggc 1440

gacgacgagg acgacggccc gacttcggga gtggtcatcc ctgcgtaacc agtgcggcca 1500

gaacccggcc taacggggcc tactgagacg aaaactgaga ctggacatgc gagaggcccg 1560

gaagcgagat cgcttccggg cctctgacct gcggaggata cgggattcga acccgtgagg 1620

gctattaacc caacacgatt tccaattccg atggcgcgag tgccaggggg tagctgaacg 1680

tgccttttgc ctggtcagtg gcactacggc aacatcaggt gtggcttgat ccgtgcgcgt 1740

<210> 85

<211> 495

<212> PRT

<213> Saccharopolyspora endophytica

<400> 85

Met Pro Arg Lys Asn Arg Asp Glu Gly Thr Arg Ala Pro Asn Gly Ala
1 5 10 15

Ser Ser Ile Tyr Lys Gly Lys Asp Gly Tyr Trp His Gly Arg Val Trp
20 25 30

Met Gly Thr Lys Asp Asp Gly Ser Glu Asp Arg Arg His Arg Ser Ala
35 40 45

Lys Ser Glu Thr Glu Leu Leu Asn Lys Val Arg Lys Leu Glu Arg Glu
50 55 60

Arg Asp Ser Gly Lys Val Gln Lys Pro Gly Arg Ala Trp Thr Val Glu
65 70 75 80

Lys Trp Leu Thr His Trp Val Glu Asn Ile Ala Ala Pro Thr Val Arg
85 90 95

Pro Thr Thr Met Val Gly Tyr Arg Ala Ser Val Tyr Lys His Leu Ile
100 105 110

Pro Gly Val Gly Lys His Arg Ile Asp Arg Leu Gln Pro Glu His Leu
115 120 125

Glu Lys Leu Tyr Ala Lys Met Gln Arg Asp Gly Leu Lys Ala Ala Thr
130 135 140

Ala His Leu Ala His Arg Thr Val Arg Val Ala Leu Asn Glu Ala Lys
145 150 155 160

Lys Arg Arg His Ile Thr Glu Asn Pro Ala Asn Ile Ala Lys Pro Pro
165 170 175

Arg Val Asp Glu Glu Glu Ile Val Pro Phe Thr Val Asp Glu Ala Arg
180 185 190

Arg Ile Leu Ala Ala Ala Ala Glu Thr Arg Asn Gly Ala Arg Phe Val
195 200 205

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

210 215 220

Trp Ser Asp Leu Ser Ile Thr Trp Lys His Gly Cys Arg Lys Gly Ser

225 230 235 240

Ala Cys Arg Val Gly Arg Arg Ala Glu Gln Cys Gly Glu Arg Arg Gly

245 250 255

Ser Gly Thr Leu Val Ile Arg Arg Ala Ile Gln Gln Gln Val Trp Gln

260 265 270

His Gly Cys Ser Glu Asp Lys Pro Cys Asp His Arg Tyr Gly Ala His

275 280 285

Cys Pro Arg Arg His Ser Gly Gly Val Val Val Thr Asp Val Lys Ser

290 295 300

Arg Ala Gly Arg Arg Thr Val Gly Leu Pro His Pro Val Val Glu Ala

305 310 315 320

Leu Glu Glu His Arg Ala Arg Gln Arg Thr Glu Arg Glu Lys Ala Arg

325 330 335

Asn Glu Trp Asp Asp Ala Asp Trp Val Phe Thr Asn Arg Trp Gly Arg

340 345 350

Pro Val His Pro Thr Val Asp Tyr Asp Ala Trp Lys Ala Leu Leu Arg

355 360 365

Ala Ala Asn Val Arg Asn Ala Arg Leu His Asp Ala Arg His Thr Ala

370 375 380

Ala Thr Met Leu Leu Val Leu Lys Val Pro Leu Pro Ala Val Met Glu

385 390 395 400

Ile Met Gly Trp Ser Glu Ala Ser Met Ala Lys Arg Tyr Met His Val

405 410 415

Pro His Glu Leu Val Thr Ala Ile Ala Asp Gln Val Gly Asp Leu Val

420 425 430

Trp Pro Val Pro Glu Thr Glu Glu Glu Ala Pro Pro Pro Glu Glu Glu

435 440 445

Trp Ala Leu Asp Ala Asn Gln Val Ala Ala Ile Arg Lys Leu Ala Gly

450 455 460

Ala Leu Pro Pro Gln Leu Arg Glu Gln Phe Glu Ala Leu Leu Pro Gly

465 470 475 480

Asp Asp Glu Asp Asp Gly Pro Thr Ser Gly Val Val Ile Pro Ala

485 490 495

<210> 86

<211> 1525

<212> DNA

<213> Saccharopolyspora erythraea

<400> 86

atgccccgca aacgccgccc agaaggcacc cgagccccca acggcgccag cagcatctac 60

tacagcgaga cggacggcta ctggcacggg cgcgtcacga tgggcgtccg cgacgacggc 120

aagcccgacc gtcgccacgt ccaagccaag accgagaccg aggtcatcga taaggtccgc 180

aagctcgaac gtgaccggga tagcggcaac gcgcggaagc ctggtcgcgc gtggacagtc 240

gagaagtggc tgactcactg ggtcgagaac atcgcggtgc actccgttcg gtacaagacg 300

cttcagggct accgaacggc ggtctacaag cacctgatcc ccggtatcgg cgcgcaccgg 360

atggaccgca tcgagccgga gcacttcgag cggttctacg ccaggatgca ggccgccggc 420

gccagtgcag ggaccgcaca tcaggtgcac cggactgcca aaacggcatt caacgaatac 480

ttccggcggc agcgcatcac cgggaacccc atcgccttcg tgaaagcgcc gcgcgtcgag 540

gaaaaggaag tggagccgtt cacgccgcag gaagccaaga gcatcatcac ggccgcgctc 600

aagcggcgca acggcgtgcg atacgtcgtc gccttggctc tcggttgtcg ccaaggcgaa 660

gccctggggt tcaagtggga ccgcctcgac cgcgggaacc ggctttaccg cgtacggcag 720

gcattgcagc ggcaggcttg gcaacacgga tgcgacgacc cgcacgcctg cggagcacga 780

cttcatcggg tggcgtgccc ggacaactgc acccagcatc gcaaccgcaa gagctgcatt 840

cgcgacgaga agggccacca ccgtccgtgc ccgccgaact gcaccaggca cgcgagcagt 900

tgcccgcagc ggcacggtgg tgggctcgtc gaggtcgacg tgaagtcgaa ggctggtcgc 960

cggagcttcg ttctgccaga tgaggtcttc gatctgctga tgcgccacga gcaggcgcag 1020

cagcgggagc gcaagcacgc cggtagcgag tggcaggagg ggggctgggt cttcacccag 1080

cccaacggcc ggccgatcga tccgcggcgc gactggggtg agtggaagga catcttgggg 1140

gaggcaggtg ttcgggatgc tcggctgcac gacgcgcgcc acactgcggc gacggtcctc 1200

atgctgctcc gcgttccaga ccgggccgtc caggatcaca tgggctggtc ctcgatccgg 1260

atgaaggagc gctacatgca cgtcaccgag gaactgcgac gagagatcgc cgatcagctc 1320

aacgggtact tctgggacgt caactgagac ggaaagtgag acgaaaagcg cctggtcagg 1380

gacctgtcga cggcgtttcc gctggtagtt tcggagccgc tgaggggact cgaacccctg 1440

accgtccgct tacaaggcgg gcgctctacc aactgagcta cagcggcgtg cgctacgtcg 1500

cgcgcgaaca tcgtaagcgt ccacc 1525

<210> 87

<211> 448

<212> PRT

<213> Saccharopolyspora erythraea

<400> 87

Met Pro Arg Lys Arg Arg Pro Glu Gly Thr Arg Ala Pro Asn Gly Ala
1 5 10 15

Ser Ser Ile Tyr Tyr Ser Glu Thr Asp Gly Tyr Trp His Gly Arg Val
20 25 30

Thr Met Gly Val Arg Asp Asp Gly Lys Pro Asp Arg Arg His Val Gln
35 40 45

Ala Lys Thr Glu Thr Glu Val Ile Asp Lys Val Arg Lys Leu Glu Arg
50 55 60

Asp Arg Asp Ser Gly Asn Ala Arg Lys Pro Gly Arg Ala Trp Thr Val
65 70 75 80

Glu Lys Trp Leu Thr His Trp Val Glu Asn Ile Ala Val His Ser Val
85 90 95

Arg Tyr Lys Thr Leu Gln Gly Tyr Arg Thr Ala Val Tyr Lys His Leu
100 105 110

Ile Pro Gly Ile Gly Ala His Arg Met Asp Arg Ile Glu Pro Glu His
115 120 125

Phe Glu Arg Phe Tyr Ala Arg Met Gln Ala Ala Gly Ala Ser Ala Gly
130 135 140

Thr Ala His Gln Val His Arg Thr Ala Lys Thr Ala Phe Asn Glu Tyr
145 150 155 160

Phe Arg Arg Gln Arg Ile Thr Gly Asn Pro Ile Ala Phe Val Lys Ala

165 170 175

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

Ala Leu Gln Arg Gln Ala Trp Gln His Gly Cys Asp Asp Pro His Ala
245 250 255
Cys Gly Ala Arg Leu His Arg Val Ala Cys Pro Asp Asn Cys Thr Gln
260 265 270
His Arg Asn Arg Lys Ser Cys Ile Arg Asp Glu Lys Gly His His Arg
275 280 285
Pro Cys Pro Pro Asn Cys Thr Arg His Ala Ser Ser Cys Pro Gln Arg
290 295 300

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

Arg Arg Asp Trp Gly Glu Trp Lys Asp Ile Leu Gly Glu Ala Gly Val
370 375 380
Arg Asp Ala Arg Leu His Asp Ala Arg His Thr Ala Ala Thr Val Leu
385 390 395 400
Met Leu Leu Arg Val Pro Asp Arg Ala Val Gln Asp His Met Gly Trp
405 410 415

Ser Ser Ile Arg Met Lys Glu Arg Tyr Met His Val Thr Glu Glu Leu
420 425 430

Arg Arg Glu Ile Ala Asp Gln Leu Asn Gly Tyr Phe Trp Asp Val Asn
435 440 445

<210> 88

<211> 2172

<212> DNA

<213> Saccharopolyspora erythraea

<400> 88

acgtcaccca actcgccgcc acgctcgcct cgctcgcggc cctgctcgcc gaacagcagc 60

ccgccccgga acccgagccc gaaccggccg cccgcaggct gcccaaccgc gtgctgctca 120

cggtcgagga agcggccaag caactggggc tcggcaggac caagacctac gcgctggtgg 180

cgtctggcga gatcgaatct gtccggatcg gtcggctcag gcgcatcccg cgcaccgcca 240

tcgacgacta cgccgcccga ctcatcgccc agcagagcgc cgcctgaagg gaaccactat 300

ggaacaaaag cgcacccgaa accccaacgg tcgatcgacg atctacctcg ggaacgacgg 360

ctactggcac ggccgcgtca ccatgggcat cggcgacgac ggcaagcctg accggcgcca 420

cgtcaagcgc aaggacaagg acgaagttgt cgaggaggtc ggcaagctcg aacgggagcg 480

ggactccggc aacgtccgca agaagggcca gccgtggaca gtcgagcggt ggctgacgca 540

ctgggtggag agcatcgcgc cgctgacctg ccggtacaag accatgcggg gctaccagac 600

ggccgtgtac aagcacctca tccccggttt gggcgcgcac aggctcgatc ggatccagaa 660

ccatccggag tacttcgaga agttctacct gcgaatgatc gagtcgggac tgaagccggc 720

gacggctcac caggtacacc gcacggcgcg aacggctttc ggcgaggcgt acaagcgggg 780

acgcatccag aggaacccgg tttcgatcgc aaaggcacct cgggtggaag aggaggaggt 840

cgaaccgctt gaggtcgagg acatgcagct ggtcatcaag gccgccctgg aacgccgaaa 900

cggcgtccgc tacgtcatcg cactggctct cggaactcgg cagggcgaat cgctcgcgct 960

gaagtggccg cggctgaacc ggcagaagcg cacgctgcgg atcaccaagg cactccaacg 1020

tcagacgtgg aagcacgggt gctctgaccc gcatcggtgc ggcgcgacct accacaagac 1080

cgagccgtgc aaggcggcct gcaagcggca cacgcgagct tgtccgccgc catgcccgcc 1140

agcttgcacc gaacacgccc ggtggtgccc gcagcgaacc ggtggcgggc tggtcgaggt 1200

cgacgtcaag tcgagggctg gacgacggac cgtgacgctg cccgaccaac tgttcgactt 1260

gatcctcaag cacgaaaagc ttcaggggcc cgaacgggag ctcgcgggca cggagtggca 1320

cgacggcgag tggatgttca cccagcccaa cggcaagccg atcgatccac gtcaggacct 1380

cgacgagtgg aaagcaatcc ttgttgaagc cggagtccgc gaggcgcggc tacatgacgc 1440

acggcacacc gccgcgactg tgctgttggt cctcggagtg cccgaccggg tcgtgatgga 1500

gctgatgggc tggtcgtccg tcaccatgaa gcagcggtac atgcacgtca tcgactccgt 1560

ccggaacgac gtagcggacc gcctgaacac ctacttctgg ggcaccaact gagacccaga 1620

ctgagaccca aaacgccccc gtcgagatcg acgggggcgt tttggcagct cttggtggtg 1680

gccaggggcg gggtcgaacc gccgaccttc cgcttttcag gcggacgctc gtaccaactg 1740

agctacctgg ccgttcgcgc ccggctcaaa gccgaaccgc tgtggcgacc cagacgggac 1800

tcgaacccgc gacctccgcc gtgacagggc ggcgcgctaa ccaactgcgc cactgggcca 1860

tgttctgttg ttgcgtaccc ccaacgggat tcgaacccgc gctaccgcct tgaaagggcg 1920

gcgtcctagg ccgctagacg atgggggctt ggccgattcg gaaccgaccc ggcctcgcct 1980

ccaaccggct ttccctttcg gggcgccccg ttgggagcag tgaaagctta cgacacaccc 2040

cccagcgccc cacaacgggg gggtccccaa acctcacgag cccccgcgcg gcccacgccc 2100

gccggtcacg tcggtcgcca ccatatgcca tctgaccagc cttttccatc gcctatcctc 2160

agtcggccca ct 2172

<210> 89

<211> 437

<212> PRT

<213> Saccharopolyspora erythraea

<400> 89

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

1 5 10 15

Leu Gly Asn Asp Gly Tyr Trp His Gly Arg Val Thr Met Gly Ile Gly

20 25 30

Asp Asp Gly Lys Pro Asp Arg Arg His Val Lys Arg Lys Asp Lys Asp

35 40 45

Glu Val Val Glu Glu Val Gly Lys Leu Glu Arg Glu Arg Asp Ser Gly

50 55 60

Asn Val Arg Lys Lys Gly Gln Pro Trp Thr Val Glu Arg Trp Leu Thr

65 70 75 80

His Trp Val Glu Ser Ile Ala Pro Leu Thr Cys Arg Tyr Lys Thr Met

85 90 95

Arg Gly Tyr Gln Thr Ala Val Tyr Lys His Leu Ile Pro Gly Leu Gly

100 105 110

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

115 120 125

Phe Tyr Leu Arg Met Ile Glu Ser Gly Leu Lys Pro Ala Thr Ala His

130 135 140

Gln Val His Arg Thr Ala Arg Thr Ala Phe Gly Glu Ala Tyr Lys Arg

145 150 155 160

Gly Arg Ile Gln Arg Asn Pro Val Ser Ile Ala Lys Ala Pro Arg Val

165 170 175

Glu Glu Glu Glu Val Glu Pro Leu Glu Val Glu Asp Met Gln Leu Val

180 185 190

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

195 200 205

Leu Ala Leu Gly Thr Arg Gln Gly Glu Ser Leu Ala Leu Lys Trp Pro

210 215 220

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

225 230 235 240

Arg Gln Thr Trp Lys His Gly Cys Ser Asp Pro His Arg Cys Gly Ala

245 250 255

Thr Tyr His Lys Thr Glu Pro Cys Lys Ala Ala Cys Lys Arg His Thr

260 265 270

Arg Ala Cys Pro Pro Pro Cys Pro Pro Ala Cys Thr Glu His Ala Arg

275 280 285

Trp Cys Pro Gln Arg Thr Gly Gly Gly Leu Val Glu Val Asp Val Lys

290 295 300

Ser Arg Ala Gly Arg Arg Thr Val Thr Leu Pro Asp Gln Leu Phe Asp

305 310 315 320

Leu Ile Leu Lys His Glu Lys Leu Gln Gly Ala Glu Arg Glu Leu Ala

325 330 335

Gly Thr Glu Trp His Asp Gly Glu Trp Met Phe Thr Gln Pro Asn Gly
340 345 350
Lys Pro Ile Asp Pro Arg Gln Asp Leu Asp Glu Trp Lys Ala Ile Leu
355 360 365
Val Glu Ala Gly Val Arg Glu Ala Arg Leu His Asp Ala Arg His Thr
370 375 380
Ala Ala Thr Val Leu Leu Val Leu Gly Val Pro Asp Arg Val Val Met

385 390 395 400
Glu Leu Met Gly Trp Ser Ser Val Thr Met Lys Gln Arg Tyr Met His
405 410 415
Val Ile Asp Ser Val Arg Asn Asp Val Ala Asp Arg Leu Asn Thr Tyr
420 425 430
Phe Trp Gly Thr Asn
435

<210> 90
<211> 1521
<212> DNA
<213> Saccharopolyspora spinosa
<400> 90
atgccacgca aacgccgccc ggaaggcacc cgggcaccca acggagccag cagcatctac 60

ctcggcaagg acggctactg gcacggccgc gtcaccgtcg gagttcgcga cgacggtaag 120
cccgaccgcc ctcacgtcca ggccaagacc gaggccgaag tcatcgacaa ggtgcgcaag 180
ctcgaacgcg atcgcgatgc ggggaaggtg cgaaagcctg gccgggcctg gaccgtcgag 240
aagtggctta cgcactgggt cgagaacatc gccgcgccat ccgtccgtta caagaccctt 300
cagggctacc gcacggcggt gtacaagcac ttgatccccg gcatcggcgc gcaccggatc 360
gaccgaattg aaccggagca cttcgagaag ctctacgcga agatgcagga atccggcgcg 420
aaagcgggaa ccgcgcacca ggtgcaccgc accgctcggg ccgcctttaa cgaagccttc 480

cggcgtcggc acctcaccga aagcccggtg cggttcgtga aagcgccgaa ggtcgaagaa 540
gaggaagtcg agcccttcac gccgaaggaa gcccagcaga tcattacggc cgcgctcaat 600
cgtcgaaacg gcgtgcgatt cgtgatcgct ctcgcactgg gctgccgcca gggtgaagcg 660
ctgggcttca agtgggaacg gctcgaccgg gaaaacaggc tctaccacgt tcggagggcg 720
cttcagcgtc aagcctggca acacggctgt gaagatccgc acaactgcgg tgcgaggttc 780

caccgggttg cttgcgccga gaactgcaag cggcaccgca atcggaagaa ctgcattcgc 840

aacgagaagg gacacgctcg accgtgcccg ccgaactgcg accgacacgc cagcagctgc 900

ccgaaacggc acggcggagg cctgcgcgag gtggatgtga agtcgaaggc tggccgccgg 960

cggttcgttc ttcctgacga gatcttcgac ctgctcatgc ggcatgagga agtccagcgg 1020

cacgaacggg ttcacgccgg taccgagtgg caggagggcg gctggatctt cacgcagccc 1080

aacggcaggc cgatcgatcc gcgccgcgat tggggcgagt ggaaggagat cctcgcggag 1140

gccggtgttc gggatgcccg gctgcacgac gcgcggcaca ccgcagcgac ggtgctcatg 1200

ctgctccgtg ttccggaccg ggccgttcag gaccacatgg gatggtcgtc gatccggatg 1260

aaagagcggt acatgcacgt caccgaggaa ctgcgccgcg agatcgccga tcagctgaat 1320

gggtatttct ggaaccccaa ctgagaccga aagtgagacg gatcgcgcct ggtcaccggg 1380

tgggcaggcg cgtttccgct ggtacggtcg gagccgctga ggggactcga acccctgacc 1440

gtccgcttac aaggcgggcg ctctaccaac tgagctacag cggcatgcac ttcgtcgtgc 1500

ggggacatcg taagcggcga t 1521

<210> 91

<211> 447

<212> PRT

<213> Saccharopolyspora spinosa

<400> 91

Met Pro Arg Lys Arg Arg Pro Glu Gly Thr Arg Ala Pro Asn Gly Ala
1 5 10 15

Ser Ser Ile Tyr Leu Gly Lys Asp Gly Tyr Trp His Gly Arg Val Thr
20 25 30

Val Gly Val Arg Asp Asp Gly Lys Pro Asp Arg Pro His Val Gln Ala
35 40 45

Lys Thr Glu Ala Glu Val Ile Asp Lys Val Arg Lys Leu Glu Arg Asp
50 55 60

Arg Asp Ala Gly Lys Val Arg Lys Pro Gly Arg Ala Trp Thr Val Glu
65 70 75 80

Lys Trp Leu Thr His Trp Val Glu Asn Ile Ala Ala Pro Ser Val Arg
85 90 95

Tyr Lys Thr Leu Gln Gly Tyr Arg Thr Ala Val Tyr Lys His Leu Ile

100 105 110

Pro Gly Ile Gly Ala His Arg Ile Asp Arg Ile Glu Pro Glu His Phe

115 120 125

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

130 135 140

Ala His Gln Val His Arg Thr Ala Arg Ala Ala Phe Asn Glu Ala Phe

145 150 155 160

Arg Arg Arg His Leu Thr Glu Ser Pro Val Arg Phe Val Lys Ala Pro

165 170 175

Lys Val Glu Glu Glu Glu Val Glu Pro Phe Thr Pro Lys Glu Ala Gln

180 185 190

Gln Ile Ile Thr Ala Ala Leu Asn Arg Arg Asn Gly Val Arg Phe Val

195 200 205

Ile Ala Leu Ala Leu Gly Cys Arg Gln Gly Glu Ala Leu Gly Phe Lys

210 215 220

Trp Glu Arg Leu Asp Arg Glu Asn Arg Leu Tyr His Val Arg Arg Ala

225 230 235 240

Leu Gln Arg Gln Ala Trp Gln His Gly Cys Glu Asp Pro His Asn Cys

245 250 255

Gly Ala Arg Phe His Arg Val Ala Cys Ala Glu Asn Cys Lys Arg His

260 265 270

Arg Asn Arg Lys Asn Cys Ile Arg Asn Glu Lys Gly His Ala Arg Pro

275 280 285

Cys Pro Pro Asn Cys Asp Arg His Ala Ser Ser Cys Pro Lys Arg His

290 295 300

Gly Gly Gly Leu Arg Glu Val Asp Val Lys Ser Lys Ala Gly Arg Arg

305 310 315 320

Arg Phe Val Leu Pro Asp Glu Ile Phe Asp Leu Leu Met Arg His Glu

325 330 335

Glu Val Gln Arg His Glu Arg Val His Ala Gly Thr Glu Trp Gln Glu

340 345 350

Gly Gly Trp Ile Phe Thr Gln Pro Asn Gly Arg Pro Ile Asp Pro Arg
355 360 365

Arg Asp Trp Gly Glu Trp Lys Glu Ile Leu Ala Glu Ala Gly Val Arg
370 375 380

Asp Ala Arg Leu His Asp Ala Arg His Thr Ala Ala Thr Val Leu Met
385 390 395 400

Leu Leu Arg Val Pro Asp Arg Ala Val Gln Asp His Met Gly Trp Ser
405 410 415

Ser Ile Arg Met Lys Glu Arg Tyr Met His Val Thr Glu Glu Leu Arg
420 425 430

Arg Glu Ile Ala Asp Gln Leu Asn Gly Tyr Phe Trp Asn Pro Asn
435 440 445

<210> 92

<211> 1669

<212> DNA

<213> Saccharopolyspora spinosa

<400> 92

atgccacgca agcgccgccc ggaaggcacc cgggcaccca acggagccag cagcatctac 60

ctcggaaacg acggctactg gcacggccgc gtcacgatgg gaacccgtga cgacggccgc 120

cccgaccgac ggcatgtcca gggcaagacc gaggccgaag tcatagacaa agtgcgcaag 180

ctcgaacgcg accgcgacgc cggacggatg cgcaagcctg gccgggcctg gaccgtcgag 240

aagtggctga tgcactggct ggagcacatt gcgaagccat cggtccggcc gaaaaccgtc 300

gcccggtatc ggacttccgt cgagcaatac ctgattcctg gtctcggtgc gcaccgcatc 360

gaccgcttgc agccggagaa cattgagaag ctgtacgcaa aattgctcgc tcgcgggttg 420

gcgccgtcca ctgtgcacca tgttcaccgg actctgcgcg tcgctttcaa cgaggcgttc 480

aagcgggaac acatcacgaa aaacccggtc ctcgttgcga aagcgccgaa gctggtcgaa 540

ccggagatcg agccgttcac cgtggccgaa gcacaacgaa ttctcgatgt tgcacggaca 600

cggcggaatg gtgctcggtt cgcactcgcg ctcgcgctgg gaatgcgcca gggcgaagct 660

ctcggactca agtggtccga cctgcgaatc acctggcacc acgggtgcgc atccggactc 720

accgaagaac agcaggcggc catcgaaatg ctcgcgaagg tcgatccgca gcgatggaag 780

cggcctgacg attccgggtg cggattcaag gacgtggagg actgcccgca ggctcacccg 840

gccgcgacac tgaacattcg gcgcgcattg cagcgccaca cctggcaaca cgggtgcggt 900

gacaaaccga cgtgcggcaa gaaacggggc gcggactgcc cgcagcgtca tggcggcggc 960
ttggccatcg tcccggtgaa gtcgagggcg gggacgcgct cgatcagcgt gcctgagccg 1020
ctgattcatg cgttgctcga tcacgacgag gcgcaggatg aggaacggca cttggcccgg 1080
aacctgtggc acgacgatgg atggatgttc gctcagccca acgggaaggc gacggacccg 1140
agggccgact atggcgaatg gcgcgagctg ctggacgccg cgaaggttcg gccggcgcgg 1200
ctgcacgacg cgcggcacac cgccgcgacg atgttgctgg ttctcaaggt cgcaccacgg 1260
gcaatcatgg acgtgatggg ctggtcggag gcgtcgatgc tgacccgcta cgtccacgtg 1320

ccggacgaga tcaagcaggg catcgcgggc caggtcggcg gactgctgtg gaaggactgg 1380
cagcagcccg acgacggccc agacgacgag gacggcggca ccgccgggca ccctgtcccg 1440
gcctgacgtg cccactgcca gaggaggcgt ttgagccgga aactgagccg gaacgacacc 1500
aggcgctttc cgtgtccacg gaaagcgcct ggtgagagcg gagccgccta agggaatcga 1560
acccttgacc tacgcattac gagtgcgtcg ctctagccga ctgagctaag gcggcgttgc 1620
acggccaagt gtagcgggcc ggacctcgcc gtcgttcatg gccccgact 1669

<210> 93
<211> 481
<212> PRT
<213> Saccharopolyspora spinosa

<400> 93

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

Arg Asp Ala Gly Arg Met Arg Lys Pro Gly Arg Ala Trp Thr Val Glu
65 70 75 80
Lys Trp Leu Met His Trp Leu Glu His Ile Ala Lys Pro Ser Val Arg
85 90 95

Pro Lys Thr Val Ala Arg Tyr Arg Thr Ser Val Glu Gln Tyr Leu Ile
100 105 110
Pro Gly Leu Gly Ala His Arg Ile Asp Arg Leu Gln Pro Glu Asn Ile
115 120 125

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

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

Gln Arg Trp Lys Arg Pro Asp Asp Ser Gly Cys Gly Phe Lys Asp Val
260 265 270
Glu Asp Cys Pro Gln Ala His Pro Ala Ala Thr Leu Asn Ile Arg Arg
275 280 285
Ala Leu Gln Arg His Thr Trp Gln His Gly Cys Gly Asp Lys Pro Thr
290 295 300
Cys Gly Lys Lys Arg Gly Ala Asp Cys Pro Gln Arg His Gly Gly Gly
305 310 315 320

Leu Ala Ile Val Pro Val Lys Ser Arg Ala Gly Thr Arg Ser Ile Ser
325 330 335
Val Pro Glu Pro Leu Ile His Ala Leu Leu Asp His Asp Glu Ala Gln

340 345 350

Asp Glu Glu Arg His Leu Ala Arg Asn Leu Trp His Asp Asp Gly Trp

355 360 365

Met Phe Ala Gln Pro Asn Gly Lys Ala Thr Asp Pro Arg Ala Asp Tyr

370 375 380

Gly Glu Trp Arg Glu Leu Leu Asp Ala Ala Lys Val Arg Pro Ala Arg

385 390 395 400

Leu His Asp Ala Arg His Thr Ala Ala Thr Met Leu Leu Val Leu Lys

405 410 415

Val Ala Pro Arg Ala Ile Met Asp Val Met Gly Trp Ser Glu Ala Ser

420 425 430

Met Leu Thr Arg Tyr Val His Val Pro Asp Glu Ile Lys Gln Gly Ile

435 440 445

Ala Gly Gln Val Gly Gly Leu Leu Trp Lys Asp Trp Gln Gln Pro Asp

450 455 460

Asp Gly Pro Asp Asp Glu Asp Gly Gly Thr Ala Gly His Pro Val Pro

465 470 475 480

Ala

<210> 94

<211> 643

<212> DNA

<213> Saccharopolyspora erythraea

<400> 94

ggtcggatct cccggttact acaggcacaa tggccacggc tggccatgct tggtctggtc 60

aggtgtccca ctgttctcaa gccgcagccg ctcagccgac accgagcggt gtgcatggca 120

cctgaacgcg tcaggcagtg acatttgccc aagcgacatg cccgctcgtg aagcggggga 180

ctgtacgaga gtacgcagag gccttcaccc ggtcgaatcc ggggtgcctt cggccgcttc 240

gacgcgggct cggccagtag cctgccacac cgccgatctt ggcgcaccgc gatggggccg 300

cttgtcgccc gacgagcgcc ttgctagcgt gcgcccctgc gatccggccc ggcgaggtgg 360

gagtcaccgg gccccgcgag gtgtgcctga cctcggtcgc agccgtgctg ccaggcaggt 420

ggggcgttcc gggggaccca cagcggaacg atcgagtact ctgtcgtacc ttcgtacaca 480

gttgtggaca actctgtgga cacctgtgtg ggcaggtgca gtcagggact gacccgcaag 540

gcgcagcgct cgccgatccc gctcgtgact cgtcgagtgc ggtgtccggc ctgggcagca 600

gcgacgcagg tcattgcgca ggcgagggga ggggagcaca ccg 643

<210> 95

<211> 11001

<212> DNA

<213> Saccharopolyspora erythraea

<400> 95

tctccggtcg cgtcttccgc agggtcggcg cgccgggcgg cgggccgggt ccgggcgcct 60

tcctcgtgcc ggcggtggac gcttacgatg ttcgcgcgcg acgtagcgca cgccgctgta 120

gctcagttgg tagagcgccc gccttgtaag cggacggtca ggggttcgag tcccctcagc 180

ggctccgaaa ctaccagcgg aaacgccgtc gacaggtccc tgaccaggcg cttttcgtct 240

cactttccgt ctcagttgac gtcccagaag tacccgttga gctgatcggc gatctctcgt 300

cgcagttcct cggtgacgtg catgtagcgc tccttcatcc ggatcgagga ccagcccatg 360

tgatcctgga cggcccggtc tggaacgcgg agcagcatga ggaccgtcgc cgcagtgtgg 420

cgcgcgtcgt gcagccgagc atcccgaaca cctgcctccc ccaagatgtc cttccactca 480

ccccagtcgc gccgcggatc gatcggccgg ccgttgggct gggtgaagac ccagcccccc 540

tcctgccact cgctaccggc gtgcttgcgc tcccgctgct gcgcctgctc gtggcgcatc 600

agcagatcga agacctcatc tggcagaacg aagctccggc gaccagcctt cgacttcacg 660

tcgacctcga cgagcccacc accgtgccgc tgcgggcaac tgctcgcgtg cctggtgcag 720

ttcggcgggc acggacggtg gtggcccttc tcgtcgcgaa tgcagctctt gcggttgcga 780

tgctgggtgc agttgtccgg gcacgccacc cgatgaagtc gtgctccgca ggcgtgcggg 840

tcgtcgcatc cgtgttgcca agcctgccgc tgcaatgcct gccgtacgcg gtaaagccgg 900

ttcccgcggt cgaggcggtc ccacttgaac cccagggctt cgccttggcg acaaccgaga 960

gccaaggcga cgacgtatcg cacgccgttg cgccgcttga gcgcggccgt gatgatgctc 1020

ttggcttcct gcggcgtgaa cggctccact tccttttcct cgacgcgcgg cgctttcacg 1080

aaggcgatgg ggttcccggt gatgcgctgc cgccggaagt attcgttgaa tgccgttttg 1140

gcagtccggt gcacctgatg tgcggtccct gcactggcgc cggcggcctg catcctggcg 1200

tagaaccgct cgaagtgctc cggctcgatg cggtccatcc ggtgcgcgcc gataccgggg 1260

atcaggtgct tgtagaccgc cgttcggtag ccctgaagcg tcttgtaccg aacggagtgc 1320

accgcgatgt tctcgaccca gtgagtcagc cacttctcga ctgtccacgc gcgaccaggc 1380

ttccgcgcgt tgccgctatc ccggtcacgt tcgagcttgc ggaccttatc gatgacctcg 1440

gtctcggtct tggcttggac gtggcgacgg tcgggcttgc cgtcgtcgcg gacgcccatc 1500

gtgacgcgcc cgtgccagta gccgtccgtc tcgctgtagt agatgctgct ggcgccgttg 1560

ggggctcggg tgccttctgg gcggcgtttg cggggcatgg atggttcctc tcggtggtcg 1620

gtttcgtggc tggttgtggg cgcgcggggt cgccgcgcga gagcgcggtg gggggttaga 1680

tccggagtta gatcggcgga ggcgagttag atatctaact cgggggttag atcgttctgc 1740

tgctgggctg acctgcggaa agttaggcaa gttagccgag ttacggggtt agcgcccagg 1800

ttgggagcag cctgggaggc tcccaactcg gggtgagcga gggtgccagg tggctaactc 1860

gtctaactcg tctagcgctt cgcccgggtc gccgagcggg ccgcgttcgg tggcgtggtg 1920

ggacgctggc ccctgtcccg ctaggcgcta gctcgctggt cggtgccggt ccgggccgct 1980

agcgctcccg ctaggtctag cgggtgtggg cgctaggtct agcggtgcgg agccggggtg 2040

tgggtacggc ggtgccggcg cggtcgaggg cgagcatgag gccgcggagt gcggccagtg 2100

cggcgaagcc accgagggcg agggtgcgca ccatcccgag ggtcaccgcg agcgcagcga 2160

tgaggatcac gggcagcacc gcccgccaca gcaggacggc taccgcggcg gcggtgagca 2220

gggcgagggc ggcgagcacg gccagcacgc ggccggtgat gtccagcggg gtcatcgtgg 2280

gtctcccggg tgtggttgtg ggcggtcggt tcgaacgggg tgggtggacg gtgcggactg 2340

cccggcaccg ccgcgcagag gtgtttccgg gcggggtgtt cacccggtcc acctgtccac 2400

ctgggcgggg ttgggcctgg tggcgggtgg tggagtgagt gggtgggcgc gtgggtggga 2460

atccacccgg aatcgcggcc tccacccctg atccaccggg catccacttg cccacttgtg 2520

gggcgttctc gcaggccgtg cccagtgggc cgggccagtg gatggccagt ggatatccac 2580

tggcccctgc cgggctcact ggcccgttca ctggcctcgc gcggcccggc ggtacggggt 2640

gaggggggtag ccgtcccagc cgtcccaagc gtcccagtgc tggtcaggct gggacggctt 2700

gcaggcgtgg gacggctcgt gccgtcccag cagtgcgcgc cgctggggtc ggtgggacgg 2760

ctcgtgccgt cccagggcgt gttgccgtcc cgcgctgacc tgcggtggga cggctgggac 2820

ggctgggacg gcacccccctc ccaccccgat attgggccga ctgggcgacg tgacggcggt 2880

ggtcatgccg cgccccgggc caggcggtcg atgtaggcgg tgagcgcttc gacggggacg 2940

cggcgggcgt ggccgatgcg taccgagtcg acgtcgccgg acttgatgag cttgaacatc 3000

gtggtgcggc cgaccgacaa ccggcgggcg gcctgctcga cggtcagcag cagccgcgtt 3060

gcgtccccgg actcccgggg ccgggggagg gaagcggggg ccggtgtcga caaccccgga 3120

ttccgggggt gtcgggcggt gtgcgcgggg cgttccctcg ttccctgttg tggggtttgc 3180

gcaggtcgag ccgagggagt ggcggaggga acggggtagg gagaattccc tgcctccccg 3240

gcggttccct gccgggttcc ctgtgccgtg acggcggtcg ggcggtgggc ggtgtgcatg 3300

gcggctcctg tggggacgaa gttgacgaaa ttacgaactt tggcgctgac ctggggtttc 3360

gtggggagca aagttcgtag ctacgaactt tgctttcagg gtttgccctg gtcaggggca 3420

ttgttcgtaa gttcgtggtg ttcgtaggca atgaagatct cggtggcgcg gcctcccttc 3480

cggccctccg ggcggcgctg ggtcttggcg atggtgtggt cggcgatgag ctggtccagc 3540

agcgcttcga tgcgggtctt cctggcgttg ccctggaaga agcctgcggt gatctcggtg 3600

cgggtgcgtc cctgtggccc ggcctcgcgg atgtaggcgg cgagctgggc cagcgccggg 3660

tcggcttggc ggaacaccgc gcgggcggag tcgacggcgt agcggatgaa cgccgccgcg 3720

gcgaccaggt gcgcgggctg gatggtgggg gtgccgtcga gggcggcgtg gatgccggcc 3780

acccggaggc agttgggtgc ggcgcgggag aggaactgct cgatcggtcc atcttcggtg 3840

ctgaagccgc cgaactcgac gtagagccgc cgccacaggt aggcggcctc gtcgctgaag 3900

ccgagctgtc cgagcgtggc gccctggtcc aggcgtgccc gcaggtcgac cgccaggcgc 3960

tcgatcagcg ccggatcggc ccccgtgcct gccgggagga actgggtctg ggcgacgaag 4020

accgggagaa accggttgta ggtgcctccg gccatgtcgg agtggctgac cttggcgtgg 4080

aactcgccgg gggtgatgtg ggtcaggatg ccgacgtggg cgtcgcgcac gatgcgggcg 4140

gtgacgccga gggtggacag gttgccgccc tcccacgccg cgcgcagcgt cgccgagagg 4200

gtgttgccct cgcggcgcat ccgcgccatc accgaggccc actccggttc gaaggccagc 4260

agtcgccgat ccccggaagg cagcagtgcg cgggggcgcg gcttgccttt gccgggctcg 4320

gtcccgtccc cctcgtcttc gtcggcgaac gcctgggtca ggccctcgcc ggaggtcagg 4380

ccgctgtgga tgttggaggc cacgaacccg gaatcggcgg ccgtcaggag gcgtttggcg 4440

gccgaccagc ccgcgccctt gcggccgatg ccggtccggc cgatgaccat cggccacacc 4500

agcagcgggt gccggtcgtc gccgacccga atgtgtgggc gtccaccgag gtggaccccg 4560

gttccggcca gcagggaggc caggatgttg gtggggtcgg cctcgctggt cggctgcacc 4620

tggcggacca ggtcgccgag gaaggtctcg aacatcgcct catcgcgggc ggggagttcg 4680

gcgcgggcgc tggcgatctg agcgagcgct tggcgttcgg tttccgggtc cggcggcaac 4740

atcggtcccg tccctgtgcc gggagcctcc gggctcgctg tgggctggtc ggagtcggag 4800

accaggcgca gccgccgagg ctgttccgct ggggtggctg cggggtgttc gggcttgctc 4860

atgggctacc gctccgatgg gtcgtgagtg ctgttcagag gggttgggaa cccgggtccg 4920

cacccgggtt cacggcttct gcgccgcctt ctccggcggg ttctcgctgc cgttctcggt 4980

gatcttggtg aggttcttcc catgcaggtg agaagcggtg agaacccgga ggaatcgacc 5040

tcgattcgtg caggccaacc aggtggatcg aggccgatcg agtcccggtc gatcgacccc 5100

cgatccccgc tcggctgggt cgttgccggg tctccgggcg actactggcg gtgacgctgc 5160

ggcggtgctc gggtcgggtt cactgaggct ggaacagtgg ggggcgggaa tccgggtctt 5220

aaacccgcaa acccgcaaaa gggggtgtga cctggtgttt tgtggtgcgg gtttggtgcg 5280

ggtttgcggg tttggtgggc tgggtttggc gggtttgcgg gtttggtgac ggctgtggcg 5340

ggtttggctc cccctcctgt cccggcttct ttttcctggt cattggggtt ggttgcgggg 5400

tttgcgggtt tgcgggtttg ctacggcgct tggcgatgtg ccgggcggcg tgcccgtggt 5460

gacgggcagg gcgccggtca cgccgcgtag tgcgaacggt ctgcctcggg tccgcagccc 5520

gcgttcgctg gtggtttcgg gctcccaacc cgcgcggagt tgggagcacc cccgtcgaac 5580

gcggcgggtc ccggggcggc tggtctggag gggtgggaga cccggcccgg tggcggcggt 5640

tcaaggccca cacggggtgg gtggcgggat ggcgccaccg ggccggggtg gtcgggatcg 5700

gcggggagga ccggcgccgg tagtggcttc tggccccaca cggggtgggg tggcggttgc 5760

gctaccgggc cggggtttga gctgttcgcg aacctgtcgg gacgcttccg acagggcaaa 5820

acacgccctg tggctgcgac cagcaggaac agccgtttcc gctgttcccg gcgggtgtgc 5880

ctgcgaccag cggaaacagt gtgtttgcgc tgttccctgg gggtgtggcc gcgaccagcg 5940

ggaacgcccg ggctagttcg gccagatcgc gtccggctcg agctccaccg ggtgcaccgc 6000

cggggtgtcc cgtcccggtg ccggcgggcg ggtgaccgcc cggaccgcgt cggcccgggc 6060

ctgagccagg gcgtagcggc gccggtcgaa cgtcggccca ccggaggcgg ccagcaccgc 6120

cagcccgttc tccagccgct cgatcgcctt gcggggcgtc ttgcggtgga ccagggccac 6180

gaaccgagcc aggtcgtaga cctcggcgtc gcgcagctgc gcgtcggtca gcgggcaggc 6240

gtagagcacg gtcaccacgc catgcacggt caccttgggc acggtcatca cgcggccacc 6300

ccttccagct cgatgcgctc ggcgccggac aggccgccga tgagcgggtc cttcgggctg 6360

gagaccacct gtggcgtcgg ggcgccgcgc tcgatgtcgg cgcgggtgac caccgggtgg 6420

ttgcgggacc aagcgcccgc gatgcccagc atgcccgggg tggtgccgta gacctcggcc 6480

cagagctgct ggtaggtcga cgagaccacc cgcacgccgc gttcccactc gccgagcagc 6540

ccgagcaggg tgcccggcga gtgcgcccgg atgccccggg accgggccgc ttcggccagg 6600

tgctggcagg cggtcttctg cgaccagccc cgggctgtgc ggtactgccg cagtcgcacg 6660

ctcatgtgcc tcacgccctc gtcagtgcgc ggcgcggtgc gggcaccccg gggcggccgt 6720

agagctcggc gacccggcgg cggatcatca tcagcagcgc gtagtccccg ctgcccggct 6780

tgatgtcctc gatctcttcg cgggcgaggc agatgcggtt cgagccctcg tgcaggaagg 6840

tctccagcac catctcgtgt gcgtactggt cgttgccgag ttcgcgctcg tccagcaaca 6900

tcccgccgtt gacggcgatc cagaacacgg cctcgatgtc ttccagcgac accggcacgc 6960

tcatgctcac ctgcatcacg atcggggtgt cgttgtcctc gcggtgggtg gggatgaacg 7020

tcgcgtggtg gtcgccgacc atgccctgtg cccgctcggt gtgcagtcgc atcgtcacgc 7080

cacctccccc gcggtctcgt cacggtcgtc ggcggcggcc gactcggtgc cggtcacgaa 7140

cgcctcgtag gcggcctggc cgcactgctc caagtagatg ttcagggcgg agccctcgat 7200

tcggagggtg cccttgccgg taccgagctt gagggcgtcg agcttcccgg actcgatcgc 7260

ccggtagatg gtggccttgg agatgttcag caggtcggcg actgccttga ccttgaacac 7320

ctggttatcc tcgaagcgca tcgtttctcc atcggtttcg atgtgcgccc tggtcggtgt 7380

tgcagcactg accagggctt tttcttgtcc ggctgaggcg gttgaggcat ctgcaacaac 7440

cggaacgaga gcaacagtag cgacgagagt taagtgagtc aagtctctgt cgtgacttcc 7500

ttgagttgac gatcgcgtac ccttgctgac gtggcaaaac gcgaggtaga gaactctcag 7560

gcgccctacc gtcaggtcgc cgacgccatc cgcgcgagca tcacgaacgg gacgtacggg 7620

ccgggggaga agctccctac cggtaaggac ttggctgacg agtacggcgt ggctatcaac 7680

accgcccgca gcgcgctgga aatcctgcgg caggaggggc tgatcgctgt ccgccatggc 7740

caggggtcgt acgtcctcag ccagccggag tccggcggag ccgatgccga gaacgctgac 7800

gtgcccagcg tcgcggcggt gcatcaacag cttgcggaga tcaaccggcg actcgcggcg 7860

atcgaggaac gcctgaacga acttgcccgc tgacctgcca gcagctaccc ggcaactgca 7920

ccgaagaacg tcttgagcgc gccgaacgtg gtcttcacga actccgcgga ttcggtcggc 7980

tgctgcacga ccataaccac gaccacgatc agcacagccc aaccgaacag cttcgggagc 8040

accccgccgc cggtcttcat catcggaaac gactgcttgg ccatgacctt ctccctcgtt 8100

cagagtgttg actgctttcg ctgtcatgcc tcaagaatcg atcgactttg cgatgatggg 8160

gagatgagtc gcgggtgagt cggaggtgag tcgttcgact cacctggctc tgaccagggg 8220

gaacgccagg ggagggaccg tgcccgtaga ggaacccgaa gcaacgcctt tgaagcgcgc 8280

acgtctcgcg agtggcatgt cgcagatgga gacccgcacc aagctccgcg aagtccgacg 8340

tcgtcgcggc aagatgcccc cgaaagacgt gagcctgaag cggatgtaca cggagtggga 8400

gcagggtcgc gtccttccga ccgactggcg cgacgaacta tgcgaggtct tcgcgctccc 8460

gccagcggct ctaggcttcg tggacacggc gccgccgccg tctgcattgg acattccatc 8520

ggcgctggag atcaccagga tcgacgccga gatcgtcgaa atgttggagc agcagaccga 8580

ccactaccgt ctcatggacc ggaaggtcgg cgcggcgatc atcccgcaga ccgtcgcaca 8640

tgttgagcac atggagaagc tgttacgcac tgccctgcct ggaaagcact cccacctggc 8700
cgcggtcgca cttgcggagg gcgcagcctt ggcgggctgg caggcgctcg atgcaggtga 8760
cgtcacgaag gcgtggaacc tgcacgacgt cgcgaaagct gctgcccggc aaggtgaaga 8820
ccctgctgtg ttcgctcacg tgacggctca gcaggcgtac gcgcttctcg acgccggccg 8880
ggccaccgaa gcagtcgagc tagtccagta cgcacacacc cccgaaatcg cgcggcgtgt 8940
cccggctcgg cttcgtgcct ggcttgcagc tgcggaggcg gagtttctcg ctgccgctgg 9000
ccagcaccgc caggcgctca cgatgctcga tcaggcggcc gacgccctac ccgaaggcga 9060

caccgatccg gaactgcctt tcctgatgct gaacagcacg cacctcgccc gttggcgtgg 9120
tcactgcctc gccaggctag gcgctgacga ggcagtcgat gaccttacga gagccctgga 9180
gggtagtcag gtcttgtcct cgaagcgggc ggagtcgggc cttcgggtcg atctcgcctt 9240
agcgctgcgg aagcgtggcg acatcgaagg ttcgaaaacg catgcgcaga gagccgcgga 9300
actggcgggc agtacgggat cagcacgaca gcgcgcacgg atcgcgaagc tgctcgtgga 9360
ctagctcacc tgccgagagc caagtagtgg agaaggccga cgagggtgcc ggagttcttc 9420
acctttccct cgcggatcag gttgggcacg ttggccacgg ggacccactc gaacttgcct 9480

tcgttgacct ccgtggcgtc tccgaccagc tcagcgcccc ggccgagaaa gacgtggtga 9540
ggattgcgca gcatgcctac tgctggttcg aatgtgacta gtggttcgat ctttcgaggc 9600
ttatagccgg tttcctcgat cagctcccgc aacgctgtct cctgcgggtc ttccccctcg 9660
tcgatgatcc ctccggggag ctcccagttc cagatatccg gcgcgaaccg gtgccgccat 9720
gccatgagca catgctcttc tgcgtcgtcg agaagcacgg tcatcgcagg aggtggaaac 9780
cacacagtgt ggtgctcgaa gcgctctcct gaaggttgcg agatatcggc caagccgacg 9840
cgtacccacg cgctctcgta cacaggacgc tcaccatgca cggtccacgt gccctcttgg 9900

tcagccactg atcagccctt ctctcggact atcgaaagtt cgttctcttg gtgttccccg 9960
gtgcagctag cgagaacggt aggcgtctcg cctatgagcg acggtcacca cggtgacaag 10020
gatctgctcg tcgttgatcg agtacagggc tcggtattcg ccgcgacggg tgctgtgcag 10080
ttcctcgaat ggcgcatcga ggcgcttacc cagacggtgc gggttgtcag cgaccggccc 10140
cgtgaggtgg tcgtgcatcg ccatggccgc agccaagggg aggcggttca tctggcgcct 10200
ggcggcagcc gtgtagctga tcgtgtagct gcctcgtggg ctcattcgtt ggccgcgcgt 10260
gctcggaggt cctctgccat ctgatcagcg gaaaccgttc ggccggctgc catgtcggcg 10320

agtccttcgc ggaggtcccg tacgagctcg gagtcggaga gcacttcgat cgtttctcgc 10380

agccgatcga tctcgcccga cggcacgacg tcggcaacgt gctcgccatg cgcgatgacg 10440

ggcacgctct cgccacgttc ggcacgctgg acggcttcgt tccatgacgc caggccatcg 10500

gcggacacgc tcactggctg ctcgctcatg gtgttcattg tgcggcatcc tctcttcggt 10560

cgaccagcgc cgaaggctgg aacagggttt gtcagggagt gccctctagg agctagaccg 10620

ctggtgaggc cgtgaacgcc gctctagcgg cccctctaga tctagaggcc gtcacctcta 10680

gatctagagg gctccgtagc ggccctcagg ccgcctctga cgaaccgcgt gcgtcctgcc 10740

gaggcctggt ggggcgagta gcgccacagc cgggagatcg gcgctcaaga ggccctgaga 10800

tcgccacaga cgccgcgaag gcgtccaccc ggtaccagga ggccgggcaa ccgggttgtg 10860

ggccgcgccg cgctccgctt cgtcggctca gttcccggct cagttcgccc ccgtacaacc 10920

acccaaccac tgttccagca attgcgctga ccagcagttt tgtacgactt tgtgctgatg 10980

tgaaccacag accggagatc t 11001

<210> 96

<211> 17501

<212> DNA

<213> Saccharopolyspora erythraea

<400> 96

ccgcattcgg acgccgcctt gttggaaggt gcgcgcactg gcaatcagtc gaacgtgcgt 60

tctatgcttt gcgtgctcgg tggtggggaa gaccgccggg ccccgcgggc cggtcactgc 120

accccattgg tggcgcggcc cgccctttac atcggggcgg ggttggtgac gagcagagat 180

gaccaatggt ggggtacccg ccggcctgtg ggcgagccca ggcctgtgtg ggtttacctg 240

ccgttggtct tcccgcctgc gcgggtggtc agcagggagg ttccgattgg tgtccgggcc 300

tacgggatcg acaacaccgc caccgtccct ggtgagctgc tggcgtggca gttgaccgcc 360

actggcgatt ggtgggctca gatccggctc accttgcaca accgcaacca gcgcggcgcg 420

ctggagaccg agctgtgggc gcccggagcc gccgtacgcc cgcggtagca ggtcaatggg 480

tatcgcggtc gccgctgcgc ccggcgcggg cgtgttcgac gctggcccgc aacgcctgca 540

tgaggtcgac ggcttcggtc tcgcgtggcg gtgggatgcc gggttcgaga gtgccgccgg 600

cgcgtttggc gtcgatgagt gcctggacct gttgctggta ggtgtcgtgg tagtcggccg 660

ggtcccagtc gatggccatg gcctcgatca actgcacagc catccgcaac tcctgctcgg 720

atgcaccgct gtggtcgggc aggctggcac cgagttcgcg gtgggggtcg cggacttcgt 780

cggcccagtg cagcaggtag accgcgagca cctcgtgttc ggccttgagc gcggcgaggt 840

actggcggtt gcgcatgacg aaggtggcga tccctgcccg gttggcggtg gccagggctt 900

cccggagcag gccgtagacc ttggtgtact gcggtccaga tggggcgagg tagtaggtgt 960
cgcggaagta gaccggggcg atctcgtcga ggtcgacgaa cccgacgatc tccagcgtcc 1020
gggagcggcc cggggcgatg tcctccagtt ccttcgggtc gacgatgacc cagccgtctt 1080
cggtggggta gcccttgacg acctcctcga aaggcacctc ctggccggtg cgctcgttga 1140
cccgcttctg ccgcacccgg tcaccggtgc cgcgctggac ctgatggaag cgaatcgtgt 1200
gccgctcgac cgcggtgaac aactgcaccg gcaccgacac cagaccgaaa cgacagcact 1260
ccaccccaca ccgcagacgc catcagcacc acctccctcc agccgacgtg accgacccca 1320

gcaccaactg accacgggaa ccagaggcag aggtcacccc ttggcaaaac ccccagcccg 1380
tcgagaagtc cgctctacgg cgatctcaga ggctcttgag aaagcatgtg cggtcgctcg 1440
tccagggcca ccggataagg gtcaccccgg gtcttgggcc ctgtgtcgca gcctcagcgg 1500
cgacgaggcc acccttctgc gtcagctagc catgccgcgt gccgggctcc taagcctccg 1560
cgcgaggtag cttcgatgca tggagctggt tccccgcgcc gccacgatcg cagagcactt 1620
cgtccgccca cttggtcggc gatggctgca cgtccaagct gtcgccgaac gcgcgcatga 1680
actgagccac gcagtgccgg ccgccgaccg ggacatgctc gtggcggctg cctggctgca 1740

cgacattgga tactcacccg agattggaca caccgggttc catccgctcg atggggcgcg 1800
gtacctgcaa gccgaggact ggcccgaggt gctggttaac ctcgtcgcgc accattccgg 1860
tgcgcggttc gaagcagcag agcgagggat ggcaggtgag ctagcagagt tcccgttcga 1920
cgactccccg ctgctcgaca ctcttgcgac tgccgacctc acgacgggac catccggcga 1980
acggctgacc tacgacgaac gcatggacga gatcctcagc cggtactcac ccgacgaccc 2040
ggtgtatcgc acgtggacca aggcgaggcc gatcatcgcg gaagccatcg cgcgtaccga 2100
ccgccgcctt gccggcagtc atccgatgta gggctcggtc cgcgcgtcgt cgagagcgtg 2160

gtcgatacgg agccgcatag acgggtggat attgaggccg ttgaggtccg ctgggggaac 2220
ccagcggacc tctttggttt cactgccgtc ctctcgcggc gtgccaccaa tccatcggcc 2280
ctcgaagcac agagagaact gctgacggac ctcaccgtcg tcataagcca tgacatggtg 2340
cgggttggtg tacgtgccaa cgagcctgac aacctcaatc gtcagcccgg tttcctcctg 2400
gacctcgcgg accacagtgt cggcgatgct ctctccagcg tcatgtccgc cgccgggtag 2460
agcccagagg tcgttatcca ctttgtggat cagcaagatc tcgccgcgct cgttcctcac 2520
cgcagccgtc acagacggca ccaccgagtt ggcctccggc gcgttcggat cgttgaagta 2580

gtccaccctt gccattcgtg cctcctacgt cggttcgggc gtggcctgct cccacacccg 2640

gtcaaaggac tccatgtagt gcttccacat ccgaccacca ggcaactcgc gaaggtgcat 2700

gacggggttc tgtccggcga gcgcgccgaa ggcatgtccg ttcacgagca actgtccgtc 2760

gaaccggtag agcgagttgt agaggatcgt tccgtgtgtc cgcacatcga ccttgggcag 2820

gcccgacacc tcgctcaggt aacggcgcat catctgcacg cggccttcga gcccacccgt 2880

ggtgccttcc tccatggctc gttggacgac tgcgggcgat ttctcatccc cgaccacgaa 2940

gcggaactga acgccttgtt cggccttctg tcggacgatc gggatgatgt tgtgctgctc 3000

gacaaggaac tgccccgaga agacgaggac ctccatcctc tccttcacgc cagcgatcag 3060

ctcgttccac agcgagtacg gcacgtccga acgtgtcggg tagaggtgga ccagctcggt 3120

ctgcttggtc ggctgcgtat ggaggtcgcc cgccaggtca ggccataggt gtacctcgtc 3180

gatgcttagc aaccccgcaa gcttcttccg cgtcgatcgg tgcggcttcc ggccgtcgtg 3240

ggtgatccac cgctcgacgg tcttgcggtc gacctgtaga cgttctgaaa cgtcctcgat 3300

cgtcagcccc gcagcgttga aagctgcacg taggcggtca ttcggcactc tgcccgctcc 3360

cctcgcggga cgttcccagg acggattggg acgttacccg gtcgatcaga acgtcttcaa 3420

gatccctgca aaggccgtgg gagcgtcccg cgaattgccg tgaactagag ccatcccagg 3480

agaaaccaac tccaacaacc ggcgccctcc agggcaacgg cggagagagc ttaaggagat 3540

ggcgatggcc aagagtccga ttccgatgat gaagaccggc ggcggtcttc taccgaagtt 3600

cgttggcgcg ctgctcacgc tggcgttcct cgccatggtg atcaagcagc ccgccgccgc 3660

ggccgagatg ctgaccggcg cgggcgcggc actcggcgcg gcggtggaag ggctcatgag 3720

cttcctgctg cagctcggaa agtagagagc gctcatggcg aagtcctgct acgtcgtgtc 3780

gctccccgac cgttcggcga gcattgccag gcgggcttcg aggtcggcca gctcagcgct 3840

gacgcggtcg acctcgaccc ggaggccgac gacggcttgg cgaacggcgt tgatctcccc 3900

tgcgacgtcg acgtcttccg tgcgatcacg gacctgcgcg ccaacgtgcc gcttgatcgt 3960

cacgaggccc gcgtctcgca ccagcttcac cgcggcgatc gccgtgttca ccgccacgtc 4020

gtagtgatcg gcgagctggc gatacggcgg cagtgaagcg ccaggagcga actcgtcagc 4080

ctcgatccga gcgatcaggt catcggcgac acgtcggcta gctgggcgtt tgtcttcgga 4140

ctcagcgctc gcactcacgt gatcagccta cccaaaacgt cttgagacat ccggttgaca 4200

aggccacccc caacccttaa tgtctcaaca catcaaggat gtctcaagac atcagagctt 4260

aggagaacga catggagtac ctggacttcg agccgtcggc agatgagctg gcgacctgcg 4320

aggcagagct ggaggactgg gagcggctgg ccgtacggga ctactgggcc gcggtcatcg 4380

ccgaagagtt gggagccccc gagccgcagg ctcggccgtc ggtgtggcgg tcgatgggcc 4440

cgggcgagcg ggagcgcttg cgccgcaccg cgcggcgggc ggccaacacc acgctgcgac 4500

tggtggccga caacgtcgcg gtcgagcagg gctcggcagc tcccaacttc ggggaggcgg 4560

cgtgatggcc ggcctcgagc ctgtcgacat cacggcgctg acacggacgt cgattccgct 4620

ggaggcggtg aaccgcttca acggcggaac ggtgtggagc acgtggggac agttccacct 4680

cgctccgtgc gactgccaca gcctctactg gtctgaccgc gacgtcttct acgtccgcga 4740

cggggagttc tacgcgtgcc cgctgtgggg gaccgagtga tggcgcggca gtggccgatc 4800

gtcaagacgg tgctgtggga cggcacgtgg gagttctgcc acggccacct cgacgggctg 4860

ccggtgttcg cctggcgcgg tcgcccttgc tcggtcccgt cccggctcga gctggccacc 4920

cgccgccagc tccggcagat gggcctgtgc cccggtgggg cggatccggt ggcgctgctg 4980

cggttccgcc accgccagcc ctaccggcgc gaagagcttg ccgagctgtt ccgcatcgac 5040

ctggccaagc ccaagcgcac cgccaccccc gcgcagcggg aagcgatcga gcgggcgctg 5100

accgcccgtc gtacgtgccc gacctgcccg cccgggcagc aggtcaagcc ctactacctg 5160

ccgacctcgg ccgggcagtg ctgggactgc tacctgcccg acaccgccgc ggcctgacga 5220

accacacagc accaacacga cccatcgcat caggagaacg agcgtcatgt ccgagccccg 5280

catccagatc ggcggcaaga cctacaccag cgagaactgg cacgagatcg ccgaggaccc 5340

gaacgtcacc ccggccaacg ggcaggtcgt caacgtcgcc cacggcgaca accacggcat 5400

ccaggccggc gccatccacg gcggcatcac cttcaaccgc aactgaccct cagggagaga 5460

tttcgtgtcc agcatcgacg acatccgcag cgtgctgtcg caggccaacc tccagtccgg 5520

cgaggtgatg gccaccctcg tcgagtccca gcgccacatc gaggccctgg ccaacatggt 5580

cgccaccgtc acccagggct cggacaacga gctggtgcag cgggccctgg ccaccttcgg 5640

cctcgcccgg acccagctct ccgaggtcgt cggcgtcatc cacgacggaa ccgaccacct 5700

cagcaactac cagcagaccc tctgaccccg ttcccccagc acaccaaccc cttcttcagg 5760

aggaagccgt catggccggt atcgaagacg tccgcgcgaa cctgtccgcc gccaccaccc 5820

aggccagcga agcgctctac gcgctcaagc aggccgcgct gaccatcaac gaggtccagc 5880

gcgtgctcga cgacaccgtg gccagcagcg cccgggagtc ggcccagcac gccatcagcg 5940

cgttccacca ggcgttcagc caggccgaac aggcccagga gctggtgatc tccggcaggg 6000

actcgatcga cacctacgcc gcccagctct aaccccaccc gcacatccac ccgctctgtc 6060

gagaacgcga ggagaacgaa ccgcaatgtc gagcatcgaa caggtccgcg ccgccatgca 6120

gtccgccacc taccgcaccg aacaggtcgt ggccgcgctg cagtcctccg cgctggagct 6180

ggaccagatc gacgccctgc tgcagaacgt cggccagggc agcagcaacg agcactacaa 6240

ccgcggcgcg ctggcgctga tccaggccac cgaaccgctg cggcaggccc tggagctggt 6300

gcgcaccggc gagaccgacc tcaacaccta cgccgcccag atctagccgc ccgcctgcga 6360

gtccgaaagg gaccaccacc gtgatccagc tcgccattac cagcttcggc tacctccacg 6420
gccccgcgcc ggaggccacc gccgtcatcg acctgcgcaa ccacctgcgc gacccccacg 6480
tcgaccccgc cttccgccag ctcaccggct tcgacctcgc cgtgcacgac aaggtcctgg 6540
ccgcgcccgg cgccgccaat atgcgcgtag ccctggcgga gttggccgcc gccctgctcc 6600
acaccgggag cgaaaagctc gtgaccatcg ccctgggctg cgccggcgga cgccaccgct 6660
ccgtcgtgct ggccaacgac ctggccaacg tcatgcgtgt ctgcggctgg cagggcgaac 6720
tcgaacaccg cgacatcgac aagcccgtca tccaccgcac caccaagtga gaggagcaca 6780

tccgatgagc ctgaggagtc ggaactggga ccgctcgccc gaaaccgagg gtgaccgccg 6840
cttccacgac ctgcgcgaca gcggctacac cggcccgatc gaccaggacg gaaaccccgt 6900
caccagcggc cgggacgccg acatcctccg ccgcatggcc gaggaacgcg gcgaaaccgt 6960
cgactggtga ccccgccccg gcccggcagc cgcaaccacc acgaagcacc tgccgggccg 7020
ggctccccac tgatccttcg aatctgcaag gagaagcccg atgctagacc ccgagaccgg 7080
cgacgtgacg gtcacgatcc ccgctgacct cgcggccaag ctgagcggcg cctacgaggc 7140
gtggcgcgcc gccctcgacg acgtcgacgc cagcggcgtc gacacggacc caacccatct 7200

ggctgccctg tggcgggccc gctcgcgtac cgagctggct ctcgccgatg tctacggcga 7260
gttggcctac tccgtcggcc cgttcgtgcc cctgatcaat gcggtgtttc gcggcgtcga 7320
cgcctgccga ggcgccgccg accgctacgt cgggatcgcc gaacgcctcg aagggagggc 7380
gtcgtgaatc gcgtcttctc gatgctgcga gtcgcggtcc gcttcacggg ctgggtggcg 7440
gtcccggtgc tgctgctggt ggcgctgtcg ctgctgtccg gcgtggtgcg cacgaccgcc 7500
ctgggcgcct tctgtgtggt cgtggcgctc aaggccgcga tgaccgtgct cgacctggtc 7560
ggcgagtacg ccccgaaccc ccgttacggc ttccgattgg aggtgaccgg tcgatgaccg 7620

cggcgaagag tgcctaccca gtgccggtgg acgagctgct gccgcaggcc cgccagctcg 7680
ccgacgacct gggcacgatc ccgccccgca accggctgat gtccgagctg aagatcgggg 7740
caccgaaggc caacgcactg ctggacaagc tcaaggccga cccggacccg gccacggccc 7800
gtccggcggg tctgcacctg gtggctgagg gccgggctct gcccgagtcg acgaccgagc 7860
cggacccggt cagcgacgcc cctgccgagc ctgccgcacc gaccccgtcg gttccggggg 7920
cggacccggc cacggtcgag gcgcacaccc cggcggcgga ggggtcgtcg gagcaggtca 7980
ctgcgcccgc agacccgggc acgccccccgg gcgaggtccg ggagcgcaag gcggtgtcga 8040

cgtggccggt gctgctgctg gcgctgccgg cgttcgtggc gatctggtcg ggctgggtgg 8100

gcctgggcgg gctcaccggg ttcgggatcg tgcacccgct tccggggatc tgggacgagt 8160

tcgaactcaa caccgcgatc acgctgccga tcggggtgga gacctacgcc gcctacgccc 8220

tgcgggtctg gctctccggg caggtcccgg cgcgagcccg ccacttcgcg aaggtctccg 8280

cgctgggctc gctggccctg ggcgcgctcg gacagatcgc ctaccacctg atgaccgcgg 8340

caggcatgac cgccgccccg tggtggatca ccaccatcgt cgcctgcctt cccgtggccg 8400

tgctcggcat gggcgccgcc ctgacccacc tgctgcacac ccccgacctg gaggtgaccc 8460

gatgaccacc cacgaccacg agccccagca cacccccgag gtcaccgaac ccgacaccga 8520

ggcgcaggtg ttccacctgc ccgtcgaacg cgccgagccc gccgacaccg caggtgccgg 8580

tgagggtgag ggtgaggtga tcgagggcga gatcgtcgag ccgccccagg ttgaccagcc 8640

cgagccccgc ggcaccgggt ccgcgctggc gcggacggag aagcgcgagc cgatccttcc 8700

cagttgggcc aaggactccc aggagttcgc cgacaccgca cggtgggcgc tgggctatgc 8760

cggacatacc gccgggtttc acgcggtgcg ctgcccggtc tacatcgccc gcatcatcgc 8820

ccgggtgccg cagggcacgc tgcggctgct gcgcgggctg gctcgatggt tgaccgacgc 8880

cgaaggccgc ccggtgcgca acgccgcggc ccgtcgcgag gacgccacgg agtacctgaa 8940

gctgtccacg cagcgcgacg ggcggatccg ctcccgcgcg gtcctcacgg ccgggctggg 9000

ggcggccggc atcgcggcgt tcctgctcgg ccgggcgatg ctgcccgagc tggtgcagtg 9060

gtcgatcgtg gccgccgcga tcggtgggct gggctggctc ggtgctccgg cggacaagcc 9120

gatcgccagc cgggcgatcg aggccacccg ggtgcccaag ctgaccagcg acgcggtgac 9180

ccgcgcgctg gctgcgctgg ggatttccca gatcaaccag gccatgggca agggtgggga 9240

gggcatcggc tacccgcggc cgatcagccg cgatggcaag ggctggcgcg ccgacatcga 9300

cctgccccac ggcgtgaccc cgggcgacat catggaccgc cgcgaaaagc tcgcgtccgg 9360

tctgcggcgc ccgaccggct gcgtgtggcc cgagtccgac aacgccgagc acgccggtcg 9420

gctggtgctg tgggtcggtg accaggacat gcgcaaggcc aaacagtccg cgtgggcgct 9480

gcgcaagggc gtgcaggtcg acctgttcca gccccagccg ttcggcaccg accagcgcgg 9540

ccggtgggtc gacctgcggc tgatgttcac cagcgtcgcg atcggcgcga ttccccgcat 9600

gggcaagacc ttcgccctgc gcgagctgct gctcatcgcc gccctggacc cgcgcgccga 9660

gctgcacacc tacgacctca agggcaccgg cgacctcgac ccgctggaga aggtcaccca 9720

cgcccacggt gtcggcgatg acgacctgga cctgcacctg gccgacatgc gggccgtgcg 9780

caccgagctg cggcgacggg ccaagctcat ccggcagttg gccaagcagg gcctcgcccc 9840

ggagaacaag gtcaccccgg agctggcctc cacgaagtcg ctgggcctgc acccgatcgt 9900

catcggcgtg gatgaatgcc aggtgtggtt cgagcacgcc aagtacggcg aggagttcga 9960

agagatctgc actgacctgg tgaaacgggg tccggcgctg ggaatcatca tcatgctggc 10020

cacccagcgc ccggacgcca agagcctgcc caccggcatc tcggcgaacg tctccacccg 10080

gttctgcctg aaggtgcagg gccagaccga aaacgacatg gtgctgggca cctcgaagta 10140

caagcagggc gtgcgggcca ccacgttcgc ctgggccgac aagggcatcg gttacctggt 10200

cggcgagggc tccgacgccc agatcgtgcg caccgtcgcc ggactcgacg gccccgccgc 10260

cgaaaaggtc gccgcctacg cccgccacct gcgcgagcag gccggaaccc tgtccggaca 10320

cgccatcggc gagaccgtca cgtccgatga ggaccaccgc cgggacacgc tgctggacga 10380

catcctcgcc gtgaccccgg aaaccgaggc gaaggtgtgg aacgagacca ccgtggcccg 10440

gctcgccgag ctgcgccccg aggtctacgg gagctgggaa gccgaccagc tctccgccgc 10500

gctcaagccc cacggcatcc gagccaaccg tcaggtgtgg ggcaccgacg agagcggcga 10560

gggccgcaac cgccgcggct tccaccgcga cgacatcacc aaaaccgtca ctgaacgtga 10620

ccgaggacgg gaagcgagct agcccccgcc gggtcgctag gtctagcggc cggccccgct 10680

agacctagcg gctccgctag caacccatct ccaccgccac cagcgcccta gtccctagcg 10740

gcgcctggcc aaacacgccc aaaaacgcct gctggaggca cctttgaccc ccgccgacct 10800

cgctaacacc agcctgaccc tgctcctcgg aatggtcacg ctctgttact cgctggtgtg 10860

tgtgatctgg ccgttcaagg actgccgcac ctgccgcggc accggccgcc tccgctcccc 10920

gttcctgcgc agctaccgcc tctgccccgc ctgcgaagcc accggcctgc ggctgcgcac 10980

cggacgcaaa gccgtcaacg ccctgcgccg cgtccaccgc cgcaaccgcg gccactgaac 11040

cggcaaccgg aaggacacca ccgtgatgca catcgtcggc agcgcccgct ccgccctgca 11100

ccagctcgga ctcaccagcg ccaagaccct ctgcggcaaa aagctccgcc agccagagga 11160

cggcgccgcg aagaacgccc cgctgtgccc ggagtgctac cgccgctccg gttggaccca 11220

cgacaagcgc cggcagtgaa caccctccag aaaggacaca tcgccatgag caggcatcgc 11280

ctcgaacccc gcaaccccaa caacgttcgc gaggtggtcg tcggctggga ctcgcccatg 11340

cggaccttct acgccgtcgt cgaagaccac agcggcgcca tcccggtcga cctcggcgac 11400

tccatcgagc ccgtgctccg tcccggggcc gtgctcgatg cggtccgccc ctacgccgcc 11460

atcccgcccg gcctcgacga cgagctgctg cgcgacgcgc tggccgaccc cggcatccgc 11520

gccgcctgac ccgaaccccg tccacgcggc cccgcccgcc attccgccac agaacccggg 11580

cggggccacc acccctcaaa gataggtaac cgtcatgttc gcagagatcg ccatccccgt 11640

ccttaccggc ggatgggccg ccgtcgccac cgccaccacc atcaactacc gacgccgcac 11700

gctgaccgac ccggtcaccg gcatcggcaa ccgcgccgcc ctctaccgca ccgcccgccg 11760

caccaccgcc cgcagcgggc tcgtcggact gctcatggtc gacctcgacc ggttcaagca 11820

gatcaacgac acccacggcc acccttcgg caaccgcgtc ctgaccgcca tcgccacccg 11880
cctgattgag aacacgctgc gcggggagcg cgcggtgcgg ctgcacggcg acgagttcgc 11940
gatctggctc ggacgcatca cctccaccgc ccgcgccgaa ggccgcgccc tccagatcgc 12000
cgacgccctc gccgaacccc tccagatcgg cggccgtcgg ctcgtggtgc ccggcagcgt 12060
cggcgtcgcc gtcgcccccg cccgcacccc actcggcgaa ctgctgaaca ccgccgacca 12120
gcacatgtac caggtcaagg ccacccacca cctgcccgca ctgcccgccg gcgagccgcg 12180
gcgcgcccgc gaccgggcca ccccgcccga ccacgccgcc tgaaacccga ccacacccac 12240

gacaggaggt gactccgcga tgaccgaccc gatccggctc gcgcactggc tgatcgagca 12300
cggcatgtac gtcttcccgc tgcgccccta ctccaagcgc ccgttcggca actgccgccg 12360
ctgcaaggac aaccgctgca cccagccgtg cccgtgcctg accgccgacc gtccctgcca 12420
cggctacctc gccgccacca accagcaccg ccgggcccgc cgctggttca cccgtatgcc 12480
ggcagccaac gtcggcatca gcaccgacct ctccgacacg gtcgtgctcg acctcgaccg 12540
caagcccaag gctcccgcag ccgccgcgca cgacgtgccc atcctcgtcg ccgacggatt 12600
gggagccctc gacgcaatca ccacgcacga gggcgccgac tggcccgaca ccctgaccat 12660

tgccaccccg tccgaagggc gacacctgta cttccgccgc cctgcggggt tggaggtcgc 12720
cagcgacgcc aacggccggg tcgggcacca gatcgacatc cgcgcccagg gtggctacgt 12780
cgtcgccccc ggctgccaga tcaccgcacc acccgaagac gtcttcggca cctacacccg 12840
cgtatcgacc acagtggaca tcgcaccgct gccggactgg ctgcgccctc gggtcacccc 12900
gccacccgcg acaccgaccg gaccggggaa ggctcccaac ctcgggcgaa tccgtcacgg 12960
ggacggtcat gagcccggtt actggaaaac ggtgtggaag agcgtgctcg acaaggtcga 13020
gtacgaggac ggcgagcgct ggaagttggt ctacaacgcc gcccgccgcc tggccaacct 13080

cgccgtgcac gacggcgccc cctggaccga gcacgaggtg ctcgacgagc tggaggccgc 13140
cgcgatccgc cgccgcgagc acaccggcaa acccaccgag cccgccaccg cacgccgcaa 13200
cgcccagcgc ggctgggacc gcggcaccca cgacggcccc gactccctga tcggcctggg 13260
cggcgcggca tgagcccgac atcccggaag gaatccctcg tgaccagccc acggcttctg 13320
gacctgttct gcggcgccgg cggcgccggc aagggctacg ccgacgccgg attcgacgtc 13380
gtcggcgtcg acatcgcccc ccagcccgac tacccgttcg agttccacca ggccgacgcg 13440
ctgaccttcc tcgccgccca cggcaccgag ttcgacgtcg tgcacgcctc gcctccgtgc 13500

caggcgtcca gcgcgttgac caagggcacc aaccgcggcc ggtcttaccc ccaattgatc 13560

ccccagaccc gcaccgccct ggtgcagctc ggggtgccgt gggtgatcga gaacgtcgcc 13620

ggcgccccga tccgcaagga cctcatgctc tgcggcgaga tgttcggcct cgccgtcctg 13680

cggcaccggt tcttcgagct gggcggctgg accacgccga gacccccgca ccccgcccac 13740

cgtggccggg tctccggaat gcgccacggc cagtggttca ccggccccta cttcgccgtc 13800

tacggcgacg gcggcggcaa aggcaccgtc gcccagtggc agcaggccat gggcatcacc 13860

tggaccgacg tccgcaagtc cctcgccgaa gccatcccgc ccgcctacac ccaccacctc 13920

ggcaccgcac tactggccgc ccgcgccgcc agccctgcgg cgacggccgc atgaccaccc 13980

gatccgcacc cgatggaggt gagcacgtga ccacacccgc tcagcagccc agcgacggcg 14040

ccgtcctgct cgacgaactg cacgccaccc tgaccaaatt cgtgatcctg cccagcccgc 14100

aggccatcga cgccgtggtg ctctggatcg ccgccaccca cgcccaaccc gcctgggccc 14160

acgccccccg cctggtcatc cgcgcccccg agaagcgctg cggcaagtcc cgactgctcg 14220

acatcgtcga aggcacctgc cacgagccgt tcctgaccgt caacgcctcc ccctccgccg 14280

tgtaccggtc gatcagcgac gacccgccca ccatgctcgt cgacgaggcc gacaccatct 14340

tcggccccga cgccggcacc aacgaagaag tccgcgggct gctcaacgcc ggacaccagc 14400

gcaaccgacc cgccaagcgc tacgacgccg catccggccg cgtcgagtcc atccccacct 14460

tcgccatggc cgcactagcc ggaatcggcg ccatgcccga caccatcgag gaccgcgccg 14520

tcatcgtgcg catgcggcgc cgcgcccccg gcgaaaccgt cgcgccctac cggcaccgcc 14580

gcgaccgccc acacctgacc gcgctcgcga agcggctcgc cgcctggctc cgcgcctcga 14640

tgcccgacct cgaacgcgcc gaacccgaca tgccgctgga ggaccgggcc gccgacacct 14700

gggaaccgct catcatcgtc gccgaccacg ccggcggcga ctggcctacc cgagcccgca 14760

acgccgcggt cgacctgctg gccgaagccg ccgacaacga ccaaggctcc ctgcggaccc 14820

ggctgctcgt cgactgccgc accgcattcg gtgaccaccc cacgctgtcc accaccgaac 14880

tgctgcgcca gctcaactcc gaccccgaag caccctggcc cacctacggc aagaccggac 14940

tcaacgccgc caagctctcc aagctgctcg ccgaattcga catccgctcc gccaacgtcc 15000

gcttccccga cggcacccag gccaagggct accagcgagc ccacttcttc gacgcctgga 15060

cccgctactg ccccgacgcc ccgcacgacc ggccagaggg ggtgccgtcc cagccgtccc 15120

aagcgtccca ccgcaggtca gagcgggacg gcttgaccct ctgggacggc atcagccgtc 15180

ccaacgacga acccgaccca gacctctggg acggcacaag ccgtcccacc gcaccgagcc 15240

gtcccagcct gacctgcatt gggacggctg ggacggctgg gacggacacc cctcccagca 15300

ctaacaccaa gggggccgca tgaccacgaa cgtcacccaa ctcgccgcca cgctcgcctc 15360

gctcgcggcc ctgctcgccg aacagcagcc cgccccggaa cccgagcccg aaccggccgc 15420

ccgcaggctg cccaaccgcg tgctgctcac ggtcgaggaa gcggccaagc aactggggct 15480

cggcaggacc aagacctacg cgctggtggc gtctggcgag atcgaatctg tccggatcgg 15540

tcggctcagg cgcatcccgc gcaccgccat cgacgactac gccgcccgac tcatcgccca 15600

gcagagcgcc gcctgaaggg aaccactatg gaacaaaagc gcacccgaaa ccccaacggt 15660

cgatcgacga tctacctcgg gaacgacggc tactggcacg gccgcgtcac catgggcatc 15720

ggcgacgacg gcaagcctga ccggcgccac gtcaagcgca aggacaagga cgaagttgtc 15780

gaggaggtcg gcaagctcga acgggagcgg gactccggca acgtccgcaa gaagggccag 15840

ccgtggacag tcgagcggtg gctgacgcac tgggtggaga gcatcgcgcc gctgacctgc 15900

cggtacaaga ccatgcgggg ctaccagacg gccgtgtaca agcacctcat ccccggtttg 15960

ggcgcgcaca ggctcgatcg gatccagaac catccggagt acttcgagaa gttctacctg 16020

cgaatgatcg agtcgggact gaagccggcg acggctcacc aggtacaccg cacggcgcga 16080

acggctttcg gcgaggcgta caagcgggga cgcatccaga ggaacccggt ttcgatcgca 16140

aaggcacctc gggtggaaga ggaggaggtc gaaccgcttg aggtcgagga catgcagctg 16200

gtcatcaagg ccgccctgga acgccgaaac ggcgtccgct acgtcatcgc actggctctc 16260

ggaactcggc agggcgaatc gctcgcgctg aagtggccgc ggctgaaccg gcagaagcgc 16320

acgctgcgga tcaccaaggc actccaacgt cagacgtgga agcacgggtg ctctgacccg 16380

catcggtgcg gcgcgaccta ccacaagacc gagccgtgca aggcggcctg caagcggcac 16440

acgcgagctt gtccgccgcc atgcccgcca gcttgcaccg aacacgcccg gtggtgcccg 16500

cagcgaaccg gtggcgggct ggtcgaggtc gacgtcaagt cgagggctgg acgacggacc 16560

gtgacgctgc ccgaccaact gttcgacttg atcctcaagc acgaaaagct tcaggggccc 16620

gaacgggagc tcgcgggcac ggagtggcac gacggcgagt ggatgttcac ccagcccaac 16680

ggcaagccga tcgatccacg tcaggacctc gacgagtgga aagcaatcct tgttgaagcc 16740

ggagtccgcg aggcgcggct acatgacgca cggcacaccg ccgcgactgt gctgttggtc 16800

ctcggagtgc ccgaccgggt cgtgatggag ctgatgggct ggtcgtccgt caccatgaag 16860

cagcggtaca tgcacgtcat cgactccgtc cggaacgacg tagcggaccg cctgaacacc 16920

tacttctggg gcaccaactg agacccagac tgagacccaa aacgcccccg tcgagatcga 16980

cgggggcgtt ttggcagctc ttggtggtgg ccaggggcgg ggtcgaaccg ccgaccttcc 17040

gcttttcagg cggacgctcg taccaactga gctacctggc cgttcgcgcc cggctcaaag 17100

ccgaaccgct gtggcgaccc agacgggact cgaacccgcg acctccgccg tgacagggcg 17160

gcgcgctaac caactgcgcc actgggccat gttctgttgt tgcgtacccc caacgggatt 17220

cgaacccgcg ctaccgcctt gaaagggcgg cgtcctaggc cgctagacga tgggggcttg 17280

gccgattcgg aaccgacccg gcctcgcctc caaccggctt tcccttcgg ggcgccccgt 17340
tgggagcagt gaaagcttac gacacacccc ccagcgcccc acaacggggg ggtccccaaa 17400
cctcacgagc ccccgcgcgg cccacgcccg ccggtcacgt cggtcgccac catatgccat 17460
ctgaccagcc ttttccatcg cctatcctca gtcggcccac t 17501

<210> 97
<211> 12
<212> DNA
<213> Saccharopolyspora spinosa
<400> 97
aggaggtccc at 12

<210> 98
<211> 25
<212> DNA
<213> Saccharopolyspora spinosa

<400> 98
ccaggaatcg gaggggcagt accga 25

<210> 99
<211> 25
<212> DNA
<213> Saccharopolyspora spinosa
<400> 99
gcaacttcct ggagggaaac gccac 25

<210> 100
<211> 25
<212> DNA
<213> Saccharopolyspora spinosa
<400> 100
tcgtcacggc agtgagggat tgggc 25

<210> 101
<211> 25
<212> DNA
<213> Saccharopolyspora spinosa

<400> 101

cgaaatcccg gcgaggaagg gcgcg 25

<210> 102

<211> 28

<212> DNA

<213> Saccharopolyspora spinosa

<400> 102

cgcctcggcc cccttcagga ggagacag 28

<210> 103

<211> 26

<212> DNA

<213> Saccharopolyspora spinosa

<400> 103

ctccagacgc ccacgcaagg agaccc 26

<210> 104

<211> 19

<212> DNA

<213> Saccharopolyspora spinosa

<400> 104

actagtaagg aggtccaac 19

<210> 105

<211> 15

<212> DNA

<213> Saccharopolyspora spinosa

<400> 105

aagaggtata tatta 15

<210> 106

<211> 20

<212> DNA

<213> Saccharopolyspora spinosa

<400> 106

ccaccgctgg aggtatccgg 20

<210> 107

<211> 13

<212> DNA

<213> Saccharopolyspora spinosa

<400> 107

aggagagatc ggc 13

<210> 108

<211> 12

<212> DNA

<213> Saccharopolyspora spinosa

<400> 108

aaagaggaga aa 12

<210> 109

<211> 15

<212> DNA

<213> Saccharopolyspora spinosa

<400> 109

attaaagagg agaaa 15

<210> 110

<211> 19

<212> DNA

<213> Saccharopolyspora spinosa

<400> 110

agaaggtgga ggtcacacc 19

<210> 111

<211> 16

<212> DNA

<213> Saccharopolyspora spinosa

<400> 111

aagggctgtt ggaatc 16

<210> 112

<211> 18

<212> DNA

<213> Saccharopolyspora spinosa

<400> 112

attgaggtcg agggtcgg 18

<210> 113

<211> 22

<212> DNA

<213> Saccharopolyspora spinosa

<400> 113

ggcggtgaat gatccgccgc gc 22

<210> 114

<211> 20

<212> DNA

<213> Saccharopolyspora spinosa

<400> 114

gacgaggaag aggcgccaca 20

<210> 115

<211> 16

<212> DNA

<213> Saccharopolyspora spinosa

<400> 115

acgttacgct cgtcgc 16

<210> 116

<211> 19

<212> DNA

<213> Saccharopolyspora spinosa

<400> 116

gggacgttac gctcgtcgc 19

<210> 117

<211> 21

<212> DNA

<213> Saccharopolyspora spinosa

<400> 117

tcgtgacctc ggtgctgaac a 21

<210> 118

<211> 15

<212> DNA

<213> Saccharopolyspora spinosa

<400> 118

aggaggaaca atcca 15

<210> 119

<211> 18

<212> DNA

<213> Saccharopolyspora spinosa

<400> 119

ccgcaggaag tgagtgac 18

<210> 120

<211> 20

<212> DNA

<213> Saccharopolyspora spinosa

<400> 120

cgtgacctcg gtgctgaaca 20

<210> 121

<211> 18

<212> DNA

<213> Saccharopolyspora spinosa

<400> 121

aattcccggg gatctacc 18

<210> 122

<211> 17

<212> DNA

<213> Saccharopolyspora spinosa

<400> 122

cgaggcgaac gccagcc 17

<210> 123

<211> 16

<212> DNA

<213> Saccharopolyspora spinosa

<400> 123

gcgaaggaga gccccc 16

<210> 124

<211> 17

<212> DNA

<213> Saccharopolyspora spinosa

<400> 124

ccgaaaggaa cgccgac 17

<210> 125

<211> 17

<212> DNA

<213> Saccharopolyspora spinosa

<400> 125

gaggaaagga aaacgaa 17

<210> 126

<211> 18

<212> DNA

<213> Saccharopolyspora spinosa

<400> 126

gaacggaagg gacgcctg 18

<210> 127

<211> 21

<212> DNA

<213> Saccharopolyspora spinosa

<400> 127

cggcgggtcg gagaggagtg c 21

<210> 128

<211> 1048

<212> DNA

<213> Artificial Sequence

<220><223> transposon mutagenesis payload sequence LoF

<400> 128

ctgtctctta tacacatctc cggaattgcc agctggggcg ccctctggta aggttgggaa 60

gccctgcaaa gtaaactgga tggctttctt gccgccaagg atctgatggc gcaggggatc 120

aagatctgat caagagacag gatgaggatc gtttcgcatg attgaacaag atggattgca 180

cgcaggttct ccggccgctt gggtggagag gctattcggc tatgactggg cacaacagac 240

aatcggctgc tctgatgccg ccgtgttccg gctgtcagcg caggggcgcc cggttctttt 300

tgtcaagacc gacctgtccg gtgccctgaa tgaactgcag gacgaggcag cgcggctatc 360

gtggctggcc acgacgggcg ttccttgcgc agctgtgctc gacgttgtca ctgaagcggg 420

aagggactgg ctgctattgg gcgaagtgcc ggggcaggat ctcctgtcat ctcaccttgc 480

tcctgccgag aaagtatcca tcatggctga tgcaatgcgg cggctgcata cgcttgatcc 540

ggctacctgc ccattcgacc accaagcgaa acatcgcatc gagcgagcac gtactcggat 600

ggaagccggt cttgtcgatc aggatgatct ggacgaagag catcaggggc tcgcgccagc 660

cgaactgttc gccaggctca aggcgcgcat gcccgacggc gaggatctcg tcgtgaccca 720

tggcgatgcc tgcttgccga atatcatggt ggaaaatggc cgcttttctg gattcatcga 780

ctgtggccgg ctgggtgtgg cggaccgcta tcaggacata gcgttggcta cccgtgatat 840

tgctgaagag cttggcggcg aatgggctga ccgcttcctc gtgctttacg gtatcgccgc 900

tcccgattcg cagcgcatcg ccttctatcg ccttcttgac gagttcttct gaatcgatag 960

ccgccccgca gggcgctccg caggccgctt ccggaccact ccggaagcgg ccgtgcggtc 1020

ggaggtacca gatgtgtata agagacag 1048

<210> 129

<211> 1352

<212> DNA

<213> Artificial Sequence

<220><223> transposon mutagenesis payload sequence Gain of Function -

promoter

<400> 129

ctgtctctta tacacatctc cggaattgcc agctggggcg ccctctggta aggttgggaa 60

gccctgcaaa gtaaactgga tggctttctt gccgccaagg atctgatggc gcaggggatc 120

aagatctgat caagagacag gatgaggatc gtttcgcatg attgaacaag atggattgca 180

cgcaggttct ccggccgctt gggtggagag gctattcggc tatgactggg cacaacagac 240

aatcggctgc tctgatgccg ccgtgttccg gctgtcagcg caggggcgcc cggttctttt 300

tgtcaagacc gacctgtccg gtgccctgaa tgaactgcag gacgaggcag cgcggctatc 360

gtggctggcc acgacgggcg ttccttgcgc agctgtgctc gacgttgtca ctgaagcggg 420

aagggactgg ctgctattgg gcgaagtgcc ggggcaggat ctcctgtcat ctcaccttgc 480

tcctgccgag aaagtatcca tcatggctga tgcaatgcgg cggctgcata cgcttgatcc 540

ggctacctgc ccattcgacc accaagcgaa acatcgcatc gagcgagcac gtactcggat 600

ggaagccggt cttgtcgatc aggatgatct ggacgaagag catcaggggc tcgcgccagc 660

cgaactgttc gccaggctca aggcgcgcat gcccgacggc gaggatctcg tcgtgaccca 720

tggcgatgcc tgcttgccga atatcatggt ggaaaatggc cgcttttctg gattcatcga 780

ctgtggccgg ctgggtgtgg cggaccgcta tcaggacata gcgttggcta cccgtgatat 840

tgctgaagag cttggcggcg aatgggctga ccgcttcctc gtgctttacg gtatcgccgc 900

tcccgattcg cagcgcatcg ccttctatcg ccttcttgac gagttcttct gaatcgatag 960

ccgccccgca gggcgctccg caggccgctt ccggaccact ccggaagcgg ccgtgcggtc 1020

ggaggtaccg gtaccagccc gacccgagca cgcgccggca cgcctggtcg atgtcggacc 1080

ggagttcgag gtacgcggct tgcaggtcca ggaaggggac gtccatgcga gtgtccgttc 1140

gagtggcggc ttgcgcccga tgctagtcgc ggttgatcgg cgatcgcagg tgcacgcggt 1200

cgatcttgac ggctggcgag aggtgcgggg aggatctgac cgacgcggtc cacacgtggc 1260

accgcgatgc tgttgtgggc acaatcgtgc cggttggtag gatccccacc caacgcaccc 1320

caggaggtcc catagatgtg tataagagac ag 1352

<210> 130

<211> 3068

<212> DNA

<213> Artificial Sequence

<220><223> transposon mutagenesis payload sequence Gain of Function - counterselection

<400> 130

ctgtctctta tacacatctg gtaccagccc gacccgagca cgcgccggca cgcctggtcg 60

atgtcggacc ggagttcgag gtacgcggct tgcaggtcca ggaaggggac gtccatgcga 120

gtgtccgttc gagtggcggc ttgcgcccga tgctagtcgc ggttgatcgg cgatcgcagg 180

tgcacgcggt cgatcttgac ggctggcgag aggtgcgggg aggatctgac cgacgcggtc 240

cacacgtggc accgcgatgc tgttgtgggc acaatcgtgc cggttggtag gatccccacc 300

caacgcaccc caggaggtcc cataagaggt atatattacc ggaattgcca gctggggcgc 360

cctctggtaa ggttgggaag ccctgcaaag taaactggat ggctttcttg ccgccaagga 420

tctgatggcg caggggatca agatctgatc aagagacagg atgaggatcg tttcgcatga 480

ttgaacaaga tggattgcac gcaggttctc cggccgcttg ggtggagagg ctattcggct 540

atgactgggc acaacagaca atcggctgct ctgatgccgc cgtgttccgg ctgtcagcgc 600

aggggcgccc ggttcttttt gtcaagaccg acctgtccgg tgccctgaat gaactgcagg 660
acgaggcagc gcggctatcg tggctggcca cgacgggcgt tccttgcgca gctgtgctcg 720
acgttgtcac tgaagcggga agggactggc tgctattggg cgaagtgccg gggcaggatc 780
tcctgtcatc tcaccttgct cctgccgaga aagtatccat catggctgat gcaatgcggc 840
ggctgcatac gcttgatccg gctacctgcc cattcgacca ccaagcgaaa catcgcatcg 900
agcgagcacg tactcggatg gaagccggtc ttgtcgatca ggatgatctg gacgaagagc 960
atcaggggct cgcgccagcc gaactgttcg ccaggctcaa ggcgcgcatg cccgacggcg 1020

aggatctcgt cgtgacccat ggcgatgcct gcttgccgaa tatcatggtg gaaaatggcc 1080
gcttttctgg attcatcgac tgtggccggc tgggtgtggc ggaccgctat caggacatag 1140
cgttggctac ccgtgatatt gctgaagagc ttggcggcga atgggctgac cgcttcctcg 1200
tgctttacgg tatcgccgct cccgattcgc agcgcatcgc cttctatcgc cttcttgacg 1260
agttcttctg atgcgcggcc ggacccgcac acacccgctc cagacgccca cgcaaggaga 1320
cccatgaaca tcaagaagtt cgccaagcgg gcgaccgtcc tgaccttcac caccgccctg 1380
ctcgcgggcg gggccaccca ggccttcgcc aaggagaaca cccagaagcc ctacaaggag 1440

acgtacgggg tgtcgcacat cacccgccac gacatgctcc agatccccaa gcagcagcag 1500
agcgagaagt accaggtccc gcagttcgac cagtccacca tcaagaacat cgaatcggcc 1560
aagggcctcg acgtgtggga ctcctggccc ctgcagaacg ccgacggcac cgtggccgag 1620
tacaacgggt accacgtggt gttcgccctg gcgggctccc ccaaggacgc cgacgacacc 1680
tcgatctaca tgttctacca gaaggtcggc gacaacagca tcgactcctg gaagaacgcg 1740
ggccgcgtct tcaaggacag cgacaagttc gacgcgaacg acgagatcct gaaggagcag 1800
acccaggagt ggtccggctc cgccaccttc acgtccgacg gcaagatccg gctcttctac 1860

acggacttct ccggcacgca ctacgggaag cagagcctca ccacggcgca ggtcaacgtg 1920
tcgaagtccg acgacaccct caagatcaac ggcgtggagg accacaagac gatcttcgac 1980
ggcgacggca agacctacca gaacgtgcag cagttcatcg acgagggcaa ctacacgtcg 2040
ggcgacaacc acacgctgcg cgacccccac tacgtggagg acaaggggca caagtacctg 2100
gtcttcgagg ccaacaccgg caccgacaac ggctaccagg gcgaggaatc cctgttcaac 2160
aaggcgtact acggcggcag cacgaacttc ttccgcaagg agagccagaa gctccagcag 2220
tcggccaaga agcgggacgc cgagctcgcc aacggcgcgc tgggcatggt ggagctgaac 2280

gacgactaca cgctgaagaa ggtcatgaag ccgctcatca cctccaacac cgtgacggac 2340

gagatcgagc gggcgaacgt cttcaagatg aacggcaagt ggtacctgtt caccgactcc 2400

cgcggctcca agatgaccat cgacggcatc aactcgaacg acatctacat gctgggttac 2460

gtctccaaca gcctgaccgg gccgtacaag ccgctcaaca agaccggcct ggtgctccag 2520

atgggcctgg acccgaacga cgtcaccttc acctactccc acttcgcggt gccccaggcg 2580

aagggcaaca acgtggtcat cacctcgtac atgacgaacc ggggcttctt cgaggacaag 2640

aaggccacct tcgccccctc cttcctgatg aacatcaagg gcaagaagac ctccgtggtg 2700

aagaacagca tcctggagca gggccagctc accgtcaaca actgaggtac cagcccgacc 2760

cgagcacgcg ccggcacgcc tggtcgatgt cggaccggag ttcgaggtac gcggcttgca 2820

ggtccaggaa ggggacgtcc atgcgagtgt ccgttcgagt ggcggcttgc gcccgatgct 2880

agtcgcggtt gatcggcgat cgcaggtgca cgcggtcgat cttgacggct ggcgagaggt 2940

gcggggagga tctgaccgac gcggtccaca cgtggcaccg cgatgctgtt gtgggcacaa 3000

tcgtgccggt tggtaggatc cccacccaac gcaccccagg aggtcccata gatgtgtata 3060

agagacag 3068

<210> 131

<211> 1716

<212> DNA

<213> Artificial Sequence

<220><223> transposon mutagenesis payload sequence Gain of Function - solubility tag

<400> 131

ctgtctctta tacacatctc cggaattgcc agctggggcg ccctctggta aggttgggaa 60

gccctgcaaa gtaaactgga tggctttctt gccgccaagg atctgatggc gcaggggatc 120

aagatctgat caagagacag gatgaggatc gtttcgcatg attgaacaag atggattgca 180

cgcaggttct ccggccgctt gggtggagag gctattcggc tatgactggg cacaacagac 240

aatcggctgc tctgatgccg ccgtgttccg gctgtcagcg caggggcgcc cggttctttt 300

tgtcaagacc gacctgtccg gtgccctgaa tgaactgcag gacgaggcag cgcggctatc 360

gtggctggcc acgacgggcg ttccttgcgc agctgtgctc gacgttgtca ctgaagcggg 420

aagggactgg ctgctattgg gcgaagtgcc ggggcaggat ctcctgtcat ctcaccttgc 480

tcctgccgag aaagtatcca tcatggctga tgcaatgcgg cggctgcata cgcttgatcc 540

ggctacctgc ccattcgacc accaagcgaa acatcgcatc gagcgagcac gtactcggat 600

ggaagccggt cttgtcgatc aggatgatct ggacgaagag catcaggggc tcgcgccagc 660

cgaactgttc gccaggctca aggcgcgcat gcccgacggc gaggatctcg tcgtgaccca 720

tggcgatgcc tgcttgccga atatcatggt ggaaaatggc cgcttttctg gattcatcga 780

ctgtggccgg ctgggtgtgg cggaccgcta tcaggacata gcgttggcta cccgtgatat 840

tgctgaagag cttggcggcg aatgggctga ccgcttcctc gtgctttacg gtatcgccgc 900

tcccgattcg cagcgcatcg ccttctatcg ccttcttgac gagttcttct gaatcgatag 960

ccgccccgca gggcgctccg caggccgctt ccggaccact ccggaagcgg ccgtgcggtc 1020

ggaggtacca tgtcccctat actaggttat tggaaaatta agggccttgt gcaacccact 1080

cgacttcttt tggaatatct tgaagaaaaa tatgaagagc atttgtatga gcgcgatgaa 1140

ggtgataaat ggcgaaacaa aaagtttgaa ttgggtttgg agtttcccaa tcttccttat 1200

tatattgatg gtgatgttaa attaacacag tctatggcca tcatacgtta tatagctgac 1260

aagcacaaca tgttgggtgg ttgtccaaaa gagcgtgcag agatttcaat gcttgaagga 1320

gcggttttgg atattagata cggtgtttcg agaattgcat atagtaaaga ctttgaaact 1380

ctcaaagttg attttcttag caagctacct gaaatgctga aaatgttcga agatcgttta 1440

tgtcataaaa catatttaaa tggtgatcat gtaacccatc ctgacttcat gttgtatgac 1500

gctcttgatg ttgttttata catggaccca atgtgcctgg atgcgttccc aaaattagtt 1560

tgttttaaaa aacgtattga agctatccca caaattgata agtacttgaa atccagcaag 1620

tatatagcat ggcctttgca gggctggcaa gccacgtttg gtggtggcga ccatcctcca 1680

aaatcggatc tggttccaga tgtgtataag agacag 1716

<210> 132

<211> 723

<212> DNA

<213> Saccharopolyspora spinosa

<400> 132

atgaccacgt tgagcctgca cggggcgaca acgctgctgt acgccgcgcc ggtctcgacc 60

gagctgctgt cccagctgcc gttggacaac ctggccgcct acgtcgccac gatggccgcc 120

gacctggcgg ccagggaccg ggaacggctg gagcagggat tggcggcggc ggtcgagcgc 180

ggcgggccgt ggttcgagcg tgaccgctac gagctggccc ggtccctcgc gagggccgtc 240

caggtcgagc ccgaggcgtc cgggtcgagc tgagcccgga ccccggattc gaaggtttcc 300

cggccggtgg tgccagtccc gccgccccgt gtcgcacggg gtttgaacac cgccaccggc 360

cgggttccca accgatcagc ggaccgtttc ggaaccgccg ggaacagcac gaaccagctc 420

ccaaccccgg ggttgggagc aatcccggaa ccaggttccc ggccgccggg aagaagggtt 480

cttcacccct tctcacctgc atgggaagaa ccccgagaag cggcaccgga gaagcaccga 540

gaacccgccg agaacccctc tcccgaccga gccgatcgac cgggcgaccg agccgaccag 600

ccgggccgcc gcagccgacc gggcgaccga gccgaccagc cgggccgccg cagccgaccg 660

ggcgaccgag ccgaccagcc gggccgccgc agccgaccgg gcgaccgagc cggtgacctg 720

cat 723

<210> 133

<211> 1259

<212> DNA

<213> Saccharopolyspora spinosa

<400> 133

tcacagcttc ctgaagtgct cgatgagccg ggagtagttg gcgttggcgc ggcccgcctc 60

gatcgccctg gagaccacgg cgacggtgta gcccgggagt tcgtcgtcga tgccccgcgc 120

ccggctctcg ctgactaggt cctcagccga cgacaggtgg tgttgcagcg ccccgaattc 180

caccgggtac tcgccgcgat cgacctcgtc gggccgggac accggcatcg gccgagatca 240

gtccgccccg gccgccgatg agttcgagcg tcgcacggta gcggtcgtgg gccgactcgg 300

cgccgccgaa gaacaccacg tcgtccgggt gcccgattac ctgagcgatc gtcatgatct 360

gcccgtgcag gtactccgcg cccgctgagg tgggccattc cgcgaccgct cgggcctcgg 420

ccgaggtgcc atccgtcagg ttgaccacgg tgtggccggc gagcgcgtcg cccgccgatt 480

ccaggagttc gcgcgcggtg ctgctgccct acaccgtgac gacgacgagc gggctggccg 540

cgacggcgtc atgaacggtc gcggcccgct gcgcccctg gcgaccaggt cgtccgcctt 600

cccgctcgtg cggttccaca ccgcggtcag tgtccggcgg ccaggaacgc gctcgcgagc 660

gcgaagccca tctcgcccaa accgatcacg gtcaccgccg gtcggctatc gatgttgctg 720

gtcatggctc atgccctcgc tcgccccaag atgatcgaga gcacgctagg gacttcgagc 780

gaacgcgagg tcaagacccc tgttcagccc ggtggcgtca ggcccaggga cttgatcagg 840

tcgtggatct cggcccggta gagcttgacc ggttccctgg tctgggggct caggcggccg 900

taggtttcca gcgaaaccgg cccgtgcatg cgacgccaga cgcgccggcc gccgtcgcgg 960

gctcactgcg gcagttcggg gaacgactcg cggacctcgt aggccaggcg cgcttcgaag 1020

tccgatcagt cgtggtcgcc gtcggcctgg cgggtgcccg cctgcggcca ggcagcggcc 1080

accaggccgg tgcggccagc gcaagcccgg tgcccggcct cggcggcgag cccgccgccg 1140

ggcggtgcct ggtagccggg gactggggcg ccatagatcc gcaaccccac aatccggcgc 1200

accacgtgca cgggcgtgcg cgtcgggctg tgatgcggtc gtgagctgcg atcggccat 1259

<210> 134

<211> 1687

<212> DNA

<213> Saccharopolyspora spinosa

<400> 134

ctaccgcgcg aatgcgggat cgccgtccag cacgcgaatc gccgcacttg cttgcagtgc 60

aacgacattc agcacatggc cgaccgagtc gtgtagttcg gcggcgatcc ggtttcgctc 120

ggacaacgcc tcgacctgcc accgtaacca ctcgtgatcg gacggtccca gcagcgtccg 180

ggcgagccgc gccaacaccg cgccgacacc ctcgaccagg tggaaacccg cgacgatcaa 240

gcccaggccg accagcggcg cccatcaatg cccgatgccg aaagcccgga acggcgagat 300

gccgacgagc tcgggcggca gcgtcagcgc cggatcggcg agcagcagca ggatgaacgg 360

cggaacgatc agcgtcagca ggccgatgcc ggagccgagc gcgaggtgga cgacgagcca 420

gcagacggtc cgcaaccgat ccggccgggt ctcggcggcg ccggattccg gcagcccgag 480

cagggattcc gccgtggcat ccgacagcga ccgcgtcgcc ggcagcaacc cggtggccac 540

cccggcgagc acgacgaacg cgccttgcgc gagaccgacc gcgacaggcc gcaggccgag 600

cccgatgtcc gggcccagca gcgccccgag caccagcccg gacaccgcca ggtacgccag 660

cggccctgag gatcaggtgc acccatcgcc ggtacgtgga gcgcggtgaa gggtctgatc 720

aggcgggacg gcatggccac atcgtctctg cggaaccggc gcgcgatcgc aggttccgca 780

cggcttgcgg cgcgtcgccc gctggtacag gtagaccgac gacccgagca ccagctccca 840

gcacatccag ctcacgtaga acagcgcgcc gggccagtag cgcatcccgg cgatcgtgtt 900

cgacctgccg gagatcaggt tggggaggac cgtggcgagc tggccgggcc cccatgcatg 960

gatgaaaagc agcgcggcgc cgaaccagcc gccggcgagc agcagcccgc gcggcgtcgg 1020

acggccagcc aggccgggaa cccagcgcgg ccagatccgc ccgcaccggt ggacgacggc 1080

cagttacagc acgatgccga tcagcgcggc gggcacgacc agccaaccgc tggcgaagcc 1140

gtcccgcgcc gggtcggggt agtccagtcc gatggtgccg cccagcgccc agttggtctt 1200

gatcatcaga tagggcagcg cgaacacgag gccgagatat ccgagccagc gcgccggcgg 1260

cccggccggc ccgccgcagc gttcgcaggc ccggtgatct cccggacccg caggactgtc 1320

ccggacccgc aggactgtcg cctggaacag caggatcgcg ctcagcaacg cgaaggcgcg 1380

ggtggcgaac atcggccagc caggggaaag cccatcaggc cggacacccg cgacagtcga 1440

acaggatgcc gccggtggcc cagagcagca gcaggacggc cgcgccgttg atcatgatga 1500

gcatgcggcg aggtcgcaac ggcgacatca gcacggcggt gccgaccacg ccgaccgctg 1560

cggccagtgg tgtcgcccag cccgggagct cgaagaggac gaatccgcgc ggcaggacgt 1620

gccgcgattc gacggagagg aaggtcaggg tcggcgcgat cacggccgcc cgactcggct 1680

tgagcat 1687

<210> 135

<211> 1286

<212> DNA

<213> Saccharopolyspora spinosa

<400> 135

gtgatcaacc gggacaccgc gtacttctgg gagggcgcgg cggccggtga gctgcgcatc 60

cagcgctgcg gccggtgcgg gctgctgcgg cacccgccgg ggccgatgtg cccggaatgc 120

ggcgccgcga cgcggacgca cctggtctcc gaagggctcg gcgaggtcta cagctccatc 180

gtccaccacc acccgctgat gccgggcaag gacctaccgc tggtcgtggc gctggtcgag 240

ctggacgaag gcgtgcgggt actgggcgag ctgctcggag tggccccgga ggacgtccgg 300

atcggccacc gggtggcggt cgacttccag cggatcgacg acgaactggt gctgccggga 360

tggaggctcc atgagtgacg gaataccggt ggcgctggct cggacgcggt cgctcgccga 420

gctgaccatc ggcgatgcgc tgccggagga gcgcatcgag gtatcgccga ccttcgtagc 480

acgtcagcgg cacggttcgg gcggcgttgc cggaggtcgc gcggtgagca cgatttcggg 540

cagcgcggcg atcgtcggga tcggtgcgac ggagttcttc aagggatccg ggcgcagcga 600

accgcaactg gccgccgagg ccgttcgggc cgccctcgcg gacgccgggc tggaaccgtc 660

cgatgtggat ggtctggtga ccttcaacat ggacaacaac accgaaaccg ccgtcgtcag 720

ggaactgggc atcccgagct gaccttcttc agccgcatcc actacggcga cggcgcggcc 780

tgcgcgactg tgcagcagac cgccacatgg ctgtcgccac cggcacggcc gacgtcgtgg 840

tctgctagcg ggcgttcaac gagcggtccg ggcgccggtt cgggcaggtg caggcggcgg 900

cagcaggcac gccgacctcg gcggggctgg acaacagctg gtcctacccg gtcggtggcc 960

acgcccggca cgcaggtcgc gatgttcgcc cgccgctacc tgcacaccta cggcgcgacc 1020

agcgaagact tcggccgcgt cgcggtggcc gaccgaaagt gcgccgcgac cagtcccaac 1080

gcctggttct accagcggcc gaccaccctg gccgatcacc aggcgtcgcg ctgggtcacc 1140

gagccgctgc ggttgctgga ctgctgccag gaaagcgacg gcggggtggc gatcgtggtg 1200

acgagccggg accgggccgc ggacctgcgg cactcgccgg ccgtggtggc ggcggcccag 1260

ggcagcggcg ccgaccagtt caccat 1286

<210> 136

<211> 1864

<212> DNA

<213> Saccharopolyspora spinosa

<400> 136

tcaatgcgac gttaccgggc gggatggcgg attcacacct gtttcaaggg ttttccccgc 60

ggccccggct acctcggtcg gccgggtacg ccggggcggg tttgccacgg gtgcgggccc 120

cgagcgggcg gtactcgacg ccgaaggcat ccagccgggg caggtggtgg tcgaccaggc 180

ggcgccggac gtccgcgaag ccgcggtcgc ggctgctcca gggccaagat cccgcatccg 240

gtcgtgctcg gctccggtcg ccgcccggtc gtcgtagccg acgcacaaca accgcagggt 300

gccctggggc acgacgtagg cggcgttggc ggcgatgccg gtgtcgagcc ggtcgaggta 360

ttccgcgacc gctgcgccag ttccagtcga accccggcgg atcatcgttc cagcccgcga 420

tctgggtgcg cacggtgacc agcgtggcgt cggttaccgg gcgtacgaca acccgtcctg 480

tccgatgacc tccagcgtca cgccctgcga caccttcgcc gtgtggtcgg gtgcggcaag 540

gatctgcagg tcggagtgtg cgtgcatgtc gatgaaaccg ggcgcgagca ctaggccgtc 600

ggcgtcgatc gcgcgttttc cgctcaactg tgcgggttcg gcgatctcgg cgatccgccc 660

gcgctcgatc ccgacgtcgg cgcggtagcc aggcccgccg tctccgtcga cgacgctcgc 720

cccgtggacc acgatctcca tgccggaacc tccccgtgct aggagttctt gcgcagcggt 780

cgtcccaatt agaagaacgt gcgctccagg tcgatcacgg tgtcgtcgtc gatgaccggg 840

atcaggggtc acttgtcgaa aaccgtgcag ggatgggaga ggccaaagcc gatccagtcg 900

ccgatctgcg gatcggtgta ccgcatggtc aggagcgtgt gctggtcggc cagcgacgtc 960

acctcgcacg aaccggcggc cagctcccgg atctcgccct ccgcgccgcg gatcacctgc 1020

ggttccggca ggtcttggtc gaacgacacg tcccggcggc ccacggtcag cagcgccagg 1080

ccccgctcag gccgggaggt gacctgcgcc cacacccgca tcgcggggcg gaacggcgat 1140

tccgcgcccg ccagccggtg ctcgcggccg aacggcgagc ggacccggta gaaaccgtcg 1200

tcgtgcagca ggtatccgcc gctgcgcagc accggcacga ccggcagccc ggccggccac 1260

ggctcggtga acgcctcggc gacctggtcg aagtacgcgc tgccgccgca ggtgacgatc 1320

acctcgtcca ggcccgagaa gtggcccgcc tgggccagtt ccgccacgag gtcgcgcaac 1380

cgccccaggt atctgtccac agcggacagc gactgctcgc tgacgtcgtg ccccagcgcg 1440

ccctcgtagc cgccgacccg accagtcgca gcatcgggct ttccagcacc aggtcggcga 1500

tctcgcgcgc ctcggcgatc gtgcgcgcac cgggccgggc gccgtcgctg ccgagctcga 1560

ccagcacgtc gacctgccgg gacgcgccgt ggtgggccag tgcctcggcc atcagctgcg 1620

ctccgcgcac cgagtccacg aagcagctga actcgaattc ggggttggcg tccagttcgt 1680

cggccagcca ggccagcccg ctggggtcca gcagttggtt ggcgagcatg atccgctcga 1740

cgccgaaggc gcggtaaacc cgcagctggc tgacgttcgc cgcggtgatg ccccacgaac 1800

cggtagccgg tattgcgtaa aatgcaacat cggttgcgta gaatgccgga ttggtgtgta 1860

ccat 1864

<210> 137

<211> 1212

<212> DNA

<213> Saccharopolyspora spinosa

<400> 137

ctagcccgct tccacgatgt cgatcgtgcg ggcgagcacg tggtgcgggt gcggctggtg 60

ggcgttgccc cattccacgg ccgccgaatg cccggcggtc gcgacgatca gcccgacggc 120

ccggtcctgc cgtggtccgc cgaggccgcc gtcttcggcg tggatgcccg ccgcgtactc 180

gccgctgcgg tgcggggcga tgctctgggc gaactgcgcg ccgcctcggg cacggtccat 240

gacgagctcg tgggtctcga ggccgcgcag cagctcgccg agcccctggt ggttgaggac 300

gacttccaca gtggacatca gccggtcacc ctttccaggt tgatcacgca cacgcccgag 360

cccggcacga acgggctcag gtagcgctgg ggtccgccgg tcaccgccca gtgggtgcca 420

tccggcagca cgatccggtc ggtggccagt acgtcggcct ccggtccggt gtagacggtc 480

atgcgggcgg tgacctgatc ggcgctgtcg gtgtcctcgg tcgaaccgcc cggggccagc 540

acgcagccgg ggatctcgtg ctcggtgggc ggcagcggat cgccgtagcg attggtgccc 600

gccggtctgg agaccgtgag tacctgcccc gccaggccct cggggatcac gggacgtcca 660

ggatcgcgct gccggtgccg acgctgaagg ccgcgccgag cccggctgcc ttccgtagcc 720

ggatgcgctg ttccctgctg agatccatgc cctgggggcc cccgctgtcg gtataggtcg 780

ccgcgtacgg gccgatcgtg gtcgaggccg caccgcccgg cgtgggcacc gccgctttgg 840

ccacctccag ggccaccgcg gtgacgtcgg gcggcacagg gtcgggcagc tcgccgatct 900

cgccgaggat cgccgcgagc gtccaggtgt ggatcgctgc gccggtctcg gccgagaccg 960

ggcggcccac cagcgcggcc agggtctcgg ctgtgaacag ctcggccatc acgtgcgccg 1020

cgttctccgg gccttggccg gggcgtccgt gccgggatcc tcatccccgc cggcctgacg 1080

tactccgagc tcggccggga gggtgcgcag ccgggccgcc gtcgcggcgt cgacctcggc 1140

cacgccatcg gtgaaccgca cacccaaatc gcgcacgtac aggccgggat gccgctggca 1200

ggtgaacttc ac 1212

<210> 138

<211> 1070

<212> DNA

<213> Saccharopolyspora spinosa

<400> 138

ctaaaccggt ctgggcgaca cctcgcacca agcccagtca cgaatgaagg cctggcaggc 60

taccgcgaag cctgccaggc cgcttcggtt gtcggtcaga agtcgaaggc caccgaggtg 120

ggtgttcttg atgcagtcgt tgtcgaaggt ccgcgagtag ctgaccggcc tgccctcccg 180

cgtaccggtg gcggtcacct ccaccgggcg tcgcctccac cgggcgcgag atcatcgggc 240

agggggcgtc cggggattgc atgggaatgc tgtcggaatt cccccgaacc cccggcagca 300

gcacgcaggc gccagccgcg tccgggtacg agccaccagc agcgtcgcag gtgagcgtca 360

cgctcgaccg cgcactctgc ccccttcgtg gtggtcaccc cggagtaaaa ccgcagcttc 420

cgggtgattt gcactccccc caccccggct atccaccccg ccggatgtac tcgtccacac 480

tggacttctc tacggaaaac cccgcccgtg atggcaagtt cggcggttga ccccccgggc 540

gatcgacctg ctgaagatgg acatcgaggg cgccgaggcc gacgcgctcg ccggaatcgg 600

cgatcacgac tgggccagga cccggcaggt cgtactcgaa gtacatgact tcgccggtgc 660

cccggcggcc gtgcgcacca cgctggcgga acgcggcttc accatcaccg cggaccggcc 720

ggagggcatg ctaggcgggc tgggtaccaa gaccgtttac gcacggcggt gaacgcggca 780

agatgcggat catcctgacc agcctgccct actactccca cctggtgccg gttgtggtcc 840

cggtggccca cgccctgcgg cgcgccgggc acaccgtcgc cgtcgcgacc gcaccgtaca 900

tggcctcgga actctcgcgg cactgcgtcg aacacctgcc gctgccgaac gtccaaaccc 960

tcgaacagct gctcgccgcc ccgcgttcgt caccagcccc ggcatgcccg gagcggaagg 1020

ggaagccgcc gaggacaccg cgcgcacccg cgagaatccc ggtccgctga 1070

<210> 139

<211> 999

<212> DNA

<213> Saccharopolyspora spinosa

<400> 139

atggaacatc tgccgttcat cgcaacgctg cagccgagct cagctcggct gcggcgggca 60

accaagtccg ggcactgctg ctccagctgg aaagcccagg gcaccgcagg gttggagtgc 120

attcacggat ttccactgct gcagcgcgac gtgcgcgccg ctgagcggcg agccaaacgg 180

atcccggtgc cagaggcggc cgagcgggaa ccgtccgagt gcgtgtgact atgagccagt 240

cctttcgtcc tccggtgtcg tctgaatctc cgcgccgcta ttcagtgcca ccggtaaccc 300

cgcgcggcat atggatcgcg cggcacggcg gtcaccgagt ggtccacgtg cacctcgtgt 360

gcaccgaggg gtcctccacc tcaccactgg gcgttcaccg gacgttgtgc gatctgcggc 420

tgggcgacga ttgggttcac ttccgccgtg tggctccatt gccgagctgg ttgacggtgt 480

gcccgctgtg cttggccagt cccaggattt cgctggacct cgccttgcgc atcgagcggc 540

atctcatcag ccgcgaaccg ctgaccgccg cacggctgat ggatcgccaa gacgccaccg 600

gattgattcc ccgagttccg ttcaccgaca gtctggtatt cgaggtcgat cgacatgctc 660

gcgatggaac acccgtcatt cgtgctggaa caagcgcgag cgcagcagca gcaccgtacg 720

ctgctgcctc cggtgtacct cggcgacttc cgcacggcgc tatacagagc gatgccgaac 780

gtccggtacc acagttgtcc cgacgacggc gggatcgcgc acctattcat catcgacgcc 840

ggagctgccg actcgtgggt ggtctggtgg cagcgcgaag acggctccca caccattgcg 900

gaagccggcc cgtacgcgct ccgcgaccgc atcaaccaga tcgctggcga ttggttgatc 960

cacatcgcca ccaccgtgga gatcgcgcac atcgcgtga 999

<210> 140

<211> 1103

<212> DNA

<213> Saccharopolyspora spinosa

<400> 140

ttagcgctcc atgaaaccgg tcaagccgcc cttgggctcg gaccactgct ccgcgacgtg 60

gtcgacgtgg cccggggttt cgccggaacg cagcgcagtc aacagcttct cgcacgcttc 120

ccgggggcct tcggcgacca cctcgacgcg gttcgcgggc cggttcgtgg cgctgcccac 180

caggccgagt tccagcgccc gggaacgcgt ccaccagcgg aagccgacgc cctgcacccg 240

cccgtgcacc catgcggtca gtcgtgcctg ttgcgaatcc gtcacgccgg gcatgatgcc 300

gcacggccgc tgcgcgcgcg gccctggctt gcgccgatca tcaccactgc ccgggcaccg 360

gtatgatccg ccacgattcg tggtagggcc gggtgttctg gggcatccgc gcagaccgcc 420

cgacgagccg acgagtgggg caccgtggac ccgatgttgc cggaagaccc ggacgcccgg 480

ccgaagcgcc tggtgatgct ggccgcatcg agcatcgccg tggccaccgt gttcgcgacg 540

accgccgcga tggtcgccgg cgcgcagcag cacgacgcgg cgacgccgga cttcacgctg 600

aagccgaccg ccacgtcccg gccggcctcg acgttctcca cgccgtcttc gttgctgccc 660

ccgccgccga tgagcagcga aacctcggat atctcgacga cctccagcag gtcgacttcg 720

tccgcttcga agagcagcga aacgctcacg gagatcccgc ctccgatgcc gccagagccg 780

ccccgcccc acactccgcc gcggccgacg aagaccacca cgacacccac gaccacgacg 840

accaccacca cgacgaccac gaccaccacg accacggagc cgaccgactc agagacctcc 900

ggcggtgacg actagtgacg ttggccgtgc acgcctgcca gtcgctgtgc tcgtggcacc 960

ggactcccaa gcagctcaac gggtttccgc tgctggcctg tcgcggctgc gggtcgcagt 1020

ggatccgcag cgagccgtgg acgcccatcg accacaccgg ccggatcccc gacgacgttc 1080

gggcggagct cgccgagcgc tga 1103

<210> 141

<211> 1191

<212> DNA

<213> Saccharopolyspora spinosa

<400> 141

ttaggccgtc gcgggttcgg cttcgtccgt gcgggcccag cggttcggtt cacccgtgat 60

ggcgaggcag gtgaccgtga tcagcgcgcc gatcagcagg taggccgcga tcggccagga 120

cgccccgccg gaggtcgtga gcagcgcggt agccaccagc ggcgacaacc cgccgcccag 180

caccgaaccg ctctggtagc cgagggaaac acccgtgtag cggaccttcg tggcgaacag 240

tccggcgcag aacgccgcca tcggcccgaa aaccgccgcc gaaccggcat atcccagcgt 300

catggccagc acgatcagcg ccgggctcgc ggtgtccagc agccagaaca tcggaaatgc 360

gtacaacgcc cggaaaatcc cgccgccgag catcacctgg gtgcggcaga tgaggtccga 420

caggtgcgag aacagcagcc ctcagcagca gccaattcgc ccgcaaagct cgccctgact 480

ggctccgcct ggtggccgac gtcgaggcgg gccgcgtaaa cgttttgatc atctgggagc 540

cgtcccgagc gtcccggctg ttatctgcgt ggtcgacgct gctggaaacc tgccaacggc 600

tcggcgtact gatccacgtc accagccacc agcagaccta cgatctcgac aatccgcgcc 660

actggcgcac actcgccgaa gacggcgttg acgcggtgta cgaagccgag aagacgtagc 720

atcgcgtccg acggggcaca aagtcctccg cggcggccgg ccgaccgcac gggaaagtcc 780

tttacggcta ccgccgggtg tacgacccag agacgcgcgc attgggcgga aacgacttcg 840

acgccgacaa gctcgtgtct cgcgtcgccg agacggtgcg ccgcaccatc gctgacgcca 900

cgatcaacgt cgatgtgaac gtgcagcggg gggatggcgc caatgtctga ccggctccgc 960

gcctggctgc gcaccaccat ccctgcggca tggtccgcgc tcgtcgcgtg gctgatcgcg 1020

gccggcgtcc cggactggct caccggcccg ctcggcgcgg ccggtgacgt cctggtcgta 1080

ctcggcgccc tctacgcgct gttgcgctgg accgagccgc acatgccgcc gtggctgacc 1140

cggatactgc tcggctccaa cacaccaccc acctacccgc cgaccgagta g 1191

<210> 142

<211> 951

<212> DNA

<213> Saccharopolyspora spinosa

<400> 142

atgaccggct tgacggtcga tccgctcgac ccggcggtgg tcccgctccg cgagggccgg 60

accgtcctgg gcgcggggtt cctggtcgca ccgggtgtgg tcgccacctg cgcccacgtg 120

gtcggcagag caacgccggt cgccgatttc ccgttgctgc gcggccacga ccacgccgtc 180

gaagtgctgt cgcaggacga cgacctggac gtcgcgatcc tgcggctggc ggacacacca 240

ccgggagcgc tgccggttcc ggcacgcaga ccgcgccgac gttggagctc gacgtgcggc 300

ccgcctggga gtcgccgcca ggtggtccga agatcgcgct ccgcccggga gatccggacc 360

agtcgctggc gccagaccag cgacggcctc tcggtgtggg acctgcggac cagaacgcag 420

gcgcacgcgt tccggctcgg ggtcaccgac ctggtcgtgt ccgccgacgg ctccgccgcc 480

gcgatgacgg accaggcgaa caactcgatc gggctggtcg acctgatcaa gatggattag 540

atcgcgccgc tgatcgcccc gggcctgaga ctcatcggga tgtccgtgcc gtacttgttc 600

gccaaggacg tcaccgcggt gcaggtccgt ttcggactac aagcacgcaa atcgctcaag 660

acgttcccgc tggacgggtt gtccgtggcc gatcgaccgg tcgtcacccg gacgggtccg 720

tggccacgac gatcaccggc gattcagttg ctttgatcga cctggagaac atggccccgt 780

tgccgccgct gatcggcaag gtcgacgaga tcgaggcctc tccggcggct acctgatctc 840

cgaggagacc ctggacttcc aggtctggaa cttgcaggaa cgccgcctgg tcaccgcgat 900

caccctcgac gacacggatt cgcggcgcgc cgtcgagaac ggcgatctca t 951

<210> 143

<211> 236

<212> PRT

<213> Artificial Sequence

<220><223> codon optimized reporter gene DasherGFP

<400> 143

Met Thr Ala Leu Thr Glu Gly Ala Lys Leu Phe Glu Lys Glu Ile Pro

1 5 10 15

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

20 25 30

Lys Gly Glu Gly Thr Gly Asp Ala Thr Thr Gly Thr Ile Lys Ala Lys

35 40 45

Tyr Ile Cys Thr Thr Gly Asp Leu Pro Val Pro Trp Ala Thr Leu Val

50 55 60

Ser Thr Leu Ser Tyr Gly Val Gln Cys Phe Ala Lys Tyr Pro Ser His

65 70 75 80

Ile Lys Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Thr Gln Glu

85 90 95

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

100 105 110

Val Thr Tyr Glu Arg Gly Ser Ile Tyr Asn Arg Val Thr Leu Thr Gly

115 120 125

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

130 135 140

Gln Cys Pro Pro Ser Ile Leu Tyr Ile Leu Pro Asp Thr Val Asn Asn

145 150 155 160

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

165 170 175

Glu Lys Leu Val Thr Lys Cys Ser Gln Met Asn Arg Pro Leu Ala Gly

180 185 190

Ser Ala Ala Val His Ile Pro Arg Tyr His His Ile Thr Tyr His Thr

195 200 205

Lys Leu Ser Lys Asp Arg Asp Glu Arg Arg Asp His Met Cys Leu Val

210 215 220

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

225 230 235

<210> 144

<211> 263

<212> PRT

<213> Artificial Sequence

<220><223> codon optimized reporter gene PaprikaRFP

<400> 144

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

1 5 10 15

Leu Tyr Met Glu Gly Thr Val Asn Asn His His Phe Lys Cys Thr Ser

20 25 30

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

35 40 45

Val Val Glu Gly Gly Pro Leu Pro Phe Ala Phe Asp Ile Leu Ala Thr

50 55 60

Ser Phe Met Tyr Gly Ser Arg Thr Phe Ile Lys Tyr Pro Lys Gly Ile

65 70 75 80

Pro Asp Phe Phe Lys Gln Ser Phe Pro Glu Gly Phe Thr Trp Glu Arg

85 90 95

Val Thr Arg Tyr Glu Asp Gly Gly Val Val Thr Val Met Gln Asp Thr

100 105 110

Ser Leu Glu Asp Gly Cys Leu Val Tyr His Val Gln Val Arg Gly Val

115 120 125

Asn Phe Pro Ser Asn Gly Pro Val Met Gln Lys Lys Thr Lys Gly Trp

130 135 140

Glu Pro Asn Thr Glu Met Leu Tyr Pro Ala Asp Gly Gly Leu Glu Gly

145 150 155 160

Arg Ser Asp Met Ala Leu Lys Leu Val Gly Gly Gly His Leu Ser Cys

165 170 175

Ser Phe Val Thr Thr Tyr Arg Ser Lys Lys Pro Ala Lys Asn Leu Lys

180 185 190

Met Pro Gly Ile His Ala Val Asp His Arg Leu Glu Arg Leu Glu Glu

195 200 205

Ser Asp Asn Glu Met Phe Val Val Gln Arg Glu His Ala Val Ala Arg

210 215 220

Tyr Cys Asp Leu Pro Ser Lys Leu Gly His Lys Leu Asn Ser Gly Leu

225 230 235 240

Arg Ser Arg Ala Gln Ala Ser Asn Ser Ala Val Asp Gly Thr Ala Gly

245 250 255

Pro Gly Ser Thr Gly Ser Arg
260

<210> 145

<211> 603

<212> PRT

<213> Artificial Sequence

<220><223> codon optimized reporter gene gusA

<400

> 145

Met Leu Arg Pro Val Glu Thr Pro Thr Arg Glu Ile Lys Lys Leu Asp
1 5 10 15

Gly Leu Trp Ala Phe Ser Leu Asp Arg Glu Asn Cys Gly Ile Asp Gln
20 25 30

Arg Trp Trp Glu Ser Ala Leu Gln Glu Ser Arg Ala Ile Ala Val Pro
35 40 45

Gly Ser Phe Asn Asp Gln Phe Ala Asp Ala Asp Ile Arg Asn Tyr Ala
50 55 60

Gly Asn Val Trp Tyr Gln Arg Glu Val Phe Ile Pro Lys Gly Trp Ala
65 70 75 80

Gly Gln Arg Ile Val Leu Arg Phe Asp Ala Val Thr His Tyr Gly Lys
85 90 95

Val Trp Val Asn Asn Gln Glu Val Met Glu His Gln Gly Gly Tyr Thr
100 105 110

Pro Phe Glu Ala Asp Val Thr Pro Tyr Val Ile Ala Gly Lys Ser Val
115 120 125

Arg Ile Thr Val Cys Val Asn Asn Glu Leu Asn Trp Gln Thr Ile Pro
130 135 140

Pro Gly Met Val Ile Thr Asp Glu Asn Gly Lys Lys Lys Gln Ser Tyr
145 150 155 160

Phe His Asp Phe Phe Asn Tyr Ala Gly Ile His Arg Ser Val Met Leu
165 170 175

Tyr Thr Thr Pro Asn Thr Trp Val Asp Asp Ile Thr Val Val Thr His

180 185 190

Val Ala Gln Asp Cys Asn His Ala Ser Val Asp Trp Gln Val Val Ala
195 200 205

Asn Gly Asp Val Ser Val Glu Leu Arg Asp Ala Asp Gln Gln Val Val
210 215 220

Ala Thr Gly Gln Gly Thr Ser Gly Thr Leu Gln Val Val Asn Pro His
225 230 235 240

Leu Trp Gln Pro Gly Glu Gly Tyr Leu Tyr Glu Leu Cys Val Thr Ala
245 250 255

Lys Ser Gln Thr Glu Cys Asp Ile Tyr Pro Leu Arg Val Gly Ile Arg
260 265 270

Ser Val Ala Val Lys Gly Glu Gln Phe Leu Ile Asn His Lys Pro Phe
275 280 285

Tyr Phe Thr Gly Phe Gly Arg His Glu Asp Ala Asp Leu Arg Gly Lys
290 295 300

Gly Phe Asp Asn Val Leu Met Val His Asp His Ala Leu Met Asp Trp
305 310 315 320

Ile Gly Ala Asn Ser Tyr Arg Thr Ser His Tyr Pro Tyr Ala Glu Glu
325 330 335

Met Leu Asp Trp Ala Asp Glu His Gly Ile Val Val Ile Asp Glu Thr
340 345 350

Ala Ala Val Gly Phe Asn Leu Ser Leu Gly Ile Gly Phe Glu Ala Gly
355 360 365

Asn Lys Pro Lys Glu Leu Tyr Ser Glu Glu Ala Val Asn Gly Glu Thr
370 375 380

Gln Gln Ala His Leu Gln Ala Ile Lys Glu Leu Ile Ala Arg Asp Lys
385 390 395 400

Asn His Pro Ser Val Val Met Trp Ser Ile Ala Asn Glu Pro Asp Thr
405 410 415

Arg Pro Gln Gly Ala Arg Glu Tyr Phe Ala Pro Leu Ala Glu Ala Thr
420 425 430

Arg Lys Leu Asp Pro Thr Arg Pro Ile Thr Cys Val Asn Val Met Phe
435 440 445

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

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

Pro Lys Ser Ala Ala Phe Leu Leu Gln Lys Arg Trp Thr Gly Met Asn
580 585 590
Phe Gly Glu Lys Pro Gln Gln Gly Gly Lys Gln
595 600

<210> 146
<211> 1422
<212> DNA
<213> Artificial Sequence
<220><223> codon optimized sacB gene
<400> 146
atgaacatca agaagttcgc caagcgggcg accgtcctga ccttcaccac cgccctgctc 60
gcgggcgggg ccacccaggc cttcgccaag gagaacaccc agaagcccta caaggagacg 120
tacggggtgt cgcacatcac ccgccacgac atgctccaga tccccaagca gcagcagagc 180

gagaagtacc aggtcccgca gttcgaccag tccaccatca agaacatcga atcggccaag 240

ggcctcgacg tgtgggactc ctggcccctg cagaacgccg acggcaccgt ggccgagtac 300

aacgggtacc acgtggtgtt cgccctggcg ggctccccca aggacgccga cgacacctcg 360

atctacatgt tctaccagaa ggtcggcgac aacagcatcg actcctggaa gaacgcgggc 420

cgcgtcttca aggacagcga caagttcgac gcgaacgacg agatcctgaa ggagcagacc 480

caggagtggt ccggctccgc caccttcacg tccgacggca agatccggct cttctacacg 540

gacttctccg gcacgcacta cgggaagcag agcctcacca cggcgcaggt caacgtgtcg 600

aagtccgacg acaccctcaa gatcaacggc gtggaggacc acaagacgat cttcgacggc 660

gacggcaaga cctaccagaa cgtgcagcag ttcatcgacg agggcaacta cacgtcgggc 720

gacaaccaca cgctgcgcga cccccactac gtggaggaca aggggcacaa gtacctggtc 780

ttcgaggcca acaccggcac cgacaacggc taccagggcg aggaatccct gttcaacaag 840

gcgtactacg gcggcagcac gaacttcttc cgcaaggaga gccagaagct ccagcagtcg 900

gccaagaagc gggacgccga gctcgccaac ggcgcgctgg gcatggtgga gctgaacgac 960

gactacacgc tgaagaaggt catgaagccg ctcatcacct ccaacaccgt gacggacgag 1020

atcgagcggg cgaacgtctt caagatgaac ggcaagtggt acctgttcac cgactcccgc 1080

ggctccaaga tgaccatcga cggcatcaac tcgaacgaca tctacatgct gggttacgtc 1140

tccaacagcc tgaccgggcc gtacaagccg ctcaacaaga ccggcctggt gctccagatg 1200

ggcctggacc cgaacgacgt caccttcacc tactcccact tcgcggtgcc ccaggcgaag 1260

ggcaacaacg tggtcatcac ctcgtacatg acgaaccggg gcttcttcga ggacaagaag 1320

gccaccttcg ccccctcctt cctgatgaac atcaagggca agaagacctc cgtggtgaag 1380

aacagcatcc tggagcaggg ccagctcacc gtcaacaact ga 1422

<210> 147

<211> 1068

<212> DNA

<213> Artificial Sequence

<220><223> codon optimized S. erythraea pheS gene

<400> 147

atgtccggtg cgaacgaccc ctacgacccc aagcaggtgg ccgcgctgtc cgccgaaacc 60

ctggaacggg cggtggccga cgcgcgggaa gccttcgaca aggccggtga cctcgacgaa 120

ctggccgccg ccaagccggc ccacctgggt gaacggagcc cgctgctgac ggcccggcgg 180

gagatcggtg ccctgccccc gaaggcccgc tccgacgcgg gtaagcgcgt gaacgaggcg 240

cgggaggcga tccagggcgc cttcgacgag cggcgggcgg ccctccaggc ggaacgcgac 300

gaacgggtgc tgcgcgaaga agccgtcgac gtcaccctcc cctgggaccg cgtgcccgtg 360

ggtgcgcgcc acccgatcac ccagctgatc gagcacgtgg ccgacacgtt cgtggccatg 420

ggttgggaag tcgccgaagg ccccgagctc gagaccgaat ggttcaactt cgacgccctg 480

aacttcggca aggaccaccc ggcgcgcacc atgcaggaca ccttctacgt cggtccgaag 540

gaatccggtc tcgtcctccg gacgcacacc agcccggccc aggtccgcgc cctgctggac 600

cgggaactgc cggtgtacgt ggtgtgcccc ggccgcacct tccggaccga cgagctggac 660

tccacgcaca cgccggtctt ccaccaggtg gaagggctcg ccgtggacaa gggtctcacg 720

atggcccacc tcaagggcac gctcgacgcg ttcgcgcgcg tcatgttcgg tcccgaatcc 780

aagacgcgcc tgcgcccgag cttcttcccg ttcgccgaac cctcgggtga agtcgacgtc 840

tggttcccgc agaagaaggg cggtcccggc tgggtcgaat ggggcggctg cggtatggtc 900

aacccgaacg tcctgcgcgc ctgcggtgtc gaccccgaaa cccacaccgg tttcggcttc 960

gggatgggtc tcgaacggac gctccagttc cgcaacggta tcccggacat gcgggacatg 1020

gtggaaggtg acgtgcagtt cacgcagccc ttcggtatcg actcctga 1068

<210> 148

<211> 1062

<212> DNA

<213> Artificial Sequence

<220><223> codon optimized S. spinosa pheS gene

<400> 148

gtgtccggcg ccaacgaccc gtacgacccg aaggaagtgg cggcgctctc gccggagacg 60

ctggatcgcg cggtggtcga ggcgagcaag gcgttcgcca cggcgacgga cctggacgcg 120

ctcgccgtgg tgaagccggc gcatctcggc gatcgtagcc cgctgctcac cgcgcgtcgc 180

gaaatcggtg cgctgccgcc caaggcgcgc agcgaagcgg gcaagcgcgt gaatgaagcg 240

cgcgaggcca tccagtcggc gttcgacgag cgccgcgccg ccttgcaggc tgagcgcgat 300

gaacgggtcc tccgcgagga gaccgttgac gtgaccctgc cgtgggaccg ggtctccgcg 360

ggggcccgcc acccgatcac ccagctggct gaggatattg aagacacgtt cgtggcgatg 420

ggttgggagg tcgcggaggg gccggagttg gaagccgaat ggttcaattt cgacgccctg 480

aacttcggta aggatcatcc ggcgcgcacg atgcaggaca ccttctatgt cgcccccgaa 540

aactcggggc tggtcttgcg gacccacacg tccccgtcgc aggtccgggc cctcctggat 600

cgcgagctgc cggtttacgt ggtttgtccc ggccgtacct tccggacgga cgaattggat 660

gcgacccaca cgccggtctt tagccaagtt gaagggctgg cggttgacaa gggtctgagc 720

atggcccact tgaaggggac gctggatgcg tttgcgcggt cgatgttcgg tccggaatcg 780

aagacccggc tgcggccgtc gtacttcccg ttttcggagc cgagcgcgga aatggacgtg 840

tggttcccgg agaagaaggg gggcgcgggc tgggtggagt ggggagggtg tggtatggtc 900

aaccccaacg tgctccgcgc gtgcggcgtg gacccggagg tctacaccgg tttcggtttc 960

ggtatgggcc tggagcggac cctgatgttc cgcaacggca tcccggacat gcgggatatg 1020

gtcgaggggg atgtgcgttt cacgcagccg tttgggatct ga 1062

<210> 149

<211> 309

<212> DNA

<213> Saccharopolyspora erythraea

<400> 149

agcttggtac cagcccgacc cgagcacgcg ccggcacgcc tggtcgatgt cggaccggag 60

ttcgaggtac gcggcttgca ggtccaggaa ggggacgtcc atgcgagtgt ccgttcgagt 120

ggcggcttgc gcccgatgct agtcgcggtt gatcggcgat cgcaggtgca cgcggtcgat 180

cttgacggct ggcgagaggt gcggggagga tctgaccgac gcggtccaca cgtggcaccg 240

cgatgctgtt gtgggcacaa tcgtgccggt tggtaggatc cccacccaac gcaccccagg 300

aggtcccat 309

<210> 150

<211> 77

<212> DNA

<213> Artificial Sequence

<220><223> terminator sequence for GFP, RFP

<400> 150

atcgatagcc gccccgcagg gcgctccgca ggccgcttcc ggaccactcc ggaagcggcc 60

gtgcggtcgg aggtacc 77

<210> 151

<211> 261

<212> PRT

<213> Artificial Sequence

<220><223> selection marker gene aac(3)IV conferring resistance to Apramycin

<400> 151

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

35 40 45
Arg Ala Ala Leu Gly Pro Gly Gly Thr Leu Val Met Pro Ser Trp Ser
50 55 60
Gly Leu Asp Asp Glu Pro Phe Asp Pro Ala Thr Ser Pro Val Thr Pro
65 70 75 80
Asp Leu Gly Val Val Ser Asp Thr Phe Trp Arg Leu Pro Asn Val Lys
85 90 95
Arg Ser Ala His Pro Phe Ala Phe Ala Ala Ala Gly Pro Gln Ala Glu

100 105 110
Gln Ile Ile Ser Asp Pro Leu Pro Leu Pro Pro His Ser Pro Ala Ser
115 120 125
Pro Val Ala Arg Val His Glu Leu Asp Gly Gln Val Leu Leu Leu Gly
130 135 140
Val Gly His Asp Ala Asn Thr Thr Leu His Leu Ala Glu Leu Met Ala
145 150 155 160
Lys Val Pro Tyr Gly Val Pro Arg His Cys Thr Ile Leu Gln Asp Gly

165 170 175
Lys Leu Val Arg Val Asp Tyr Leu Glu Asn Asp His Cys Cys Glu Arg
180 185 190
Phe Ala Leu Ala Asp Arg Trp Leu Lys Glu Lys Ser Leu Gln Lys Glu
195 200 205
Gly Pro Val Gly His Ala Phe Ala Arg Leu Ile Arg Ser Arg Asp Ile
210 215 220
Val Ala Thr Ala Leu Gly Gln Leu Gly Arg Asp Pro Leu Ile Phe Leu

225 230 235 240
His Pro Pro Glu Gly Gly Met Arg Arg Met Arg Cys Arg Ser Pro Val

245 250 255

Asp Trp Leu Ser Ser

260

<210> 152

<211> 177

<212> PRT

<213> Artificial Sequence

<220><223> selection marker gene aacC1 conferring resistance to Gentamycin

<400> 152

Met Leu Arg Ser Ser Asn Asp Val Thr Gln Gln Gly Ser Arg Pro Lys

1 5 10 15

Thr Lys Leu Gly Gly Ser Ser Met Gly Ile Ile Arg Thr Cys Arg Leu

20 25 30

Gly Pro Asp Gln Val Lys Ser Met Arg Ala Ala Leu Asp Leu Phe Gly

35 40 45

Arg Glu Phe Gly Asp Val Ala Thr Tyr Ser Gln His Gln Pro Asp Ser

50 55 60

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

65 70 75 80

Ala Phe Asp Gln Glu Ala Val Val Gly Ala Leu Ala Ala Tyr Val Leu

85 90 95

Pro Arg Phe Glu Gln Pro Arg Ser Glu Ile Tyr Ile Tyr Asp Leu Ala

100 105 110

Val Ser Gly Glu His Arg Arg Gln Gly Ile Ala Thr Ala Leu Ile Asn

115 120 125

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

130 135 140

Gln Ala Asp Tyr Gly Asp Asp Pro Ala Val Ala Leu Tyr Thr Lys Leu

145 150 155 160

Gly Ile Arg Glu Glu Val Met His Phe Asp Ile Asp Pro Ser Thr Ala

165 170 175

Thr

<210> 153

<211> 286

<212> PRT

<213> Artificial Sequence

<220><223> selection marker gene aacC8 conferring resistance to Neomycin B

<400> 153

Met Asp Glu Lys Glu Leu Ile Glu Arg Ala Gly Gly Pro Val Thr Arg
1 5 10 15

Gly Arg Leu Val Arg Asp Leu Glu Ala Leu Gly Val Gly Ala Gly Asp
20 25 30

Thr Val Met Val His Thr Arg Met Ser Ala Ile Gly Tyr Val Val Gly
35 40 45

Gly Pro Gln Thr Val Ile Asp Ala Val Arg Asp Ala Val Gly Ala Asp
50 55 60

Gly Thr Leu Met Ala Tyr Cys Gly Trp Asn Asp Ala Pro Pro Tyr Asp
65 70 75 80

Leu Ala Glu Trp Pro Pro Ala Trp Arg Glu Ala Ala Arg Ala Glu Trp
85 90 95

Pro Ala Tyr Asp Pro Leu Leu Ser Glu Ala Asp Arg Gly Asn Gly Arg
100 105 110

Val Pro Glu Ala Leu Arg His Gln Pro Gly Ala Val Arg Ser Arg His
115 120 125

Pro Asp Ala Ser Phe Val Ala Val Gly Pro Ala Ala His Pro Leu Met
130 135 140

Asp Asp His Pro Trp Asp Asp Pro His Gly Pro Asp Ser Pro Leu Ala
145 150 155 160

Arg Leu Ala Gly Ala Gly Gly Arg Val Leu Leu Leu Gly Ala Pro Leu
165 170 175

Asp Thr Leu Thr Leu Leu His His Ala Glu Ala Arg Ala Glu Ala Pro
180 185 190

Gly Lys Arg Phe Val Ala Tyr Glu Gln Pro Val Thr Val Gly Gly Arg

195 200 205

Arg Val Trp Arg Arg Phe Arg Asp Val Asp Thr Ser Arg Gly Val Pro
210 215 220
Tyr Gly Arg Val Val Pro Glu Gly Val Val Pro Phe Thr Val Ile Ala
225 230 235 240
Gln Asp Met Leu Ala Ala Gly Ile Gly Arg Thr Gly Arg Val Ala Ala
245 250 255
Ala Pro Val His Leu Phe Glu Ala Ala Asp Val Val Arg Phe Gly Val
260 265 270

Glu Trp Ile Glu Ser Arg Met Gly Gly Ala Ala Gly Gly Ala
275 280 285

<210> 154
<211> 262
<212> PRT
<213> Artificial Sequence
<220><223> selection marker gene aadA conferring resistance to Spectinomycin, Streptomycin
<400> 154

Met Arg Glu Ala Val Ile Ala Glu Val Ser Thr Gln Leu Ser Glu Val
1 5 10 15
Val Gly Val Ile Glu Arg His Leu Glu Pro Thr Leu Leu Ala Val His
20 25 30

Leu Tyr Gly Ser Ala Val Asp Gly Gly Leu Lys Pro His Ser Asp Ile
35 40 45
Asp Leu Leu Val Thr Val Thr Val Arg Leu Asp Glu Thr Thr Arg Arg
50 55 60
Ala Leu Ile Asn Asp Leu Leu Glu Thr Ser Ala Ser Pro Gly Glu Ser
65 70 75 80
Glu Ile Leu Arg Ala Val Glu Val Thr Ile Val Val His Asp Asp Ile
85 90 95

Ile Pro Trp Arg Tyr Pro Ala Lys Arg Glu Leu Gln Phe Gly Glu Trp

100 105 110

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

115 120 125

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

130 135 140

Leu Val Gly Pro Ala Ala Glu Glu Leu Phe Asp Pro Val Pro Glu Gln

145 150 155 160

Asp Leu Phe Glu Ala Leu Asn Glu Thr Leu Thr Leu Trp Asn Ser Pro

165 170 175

Pro Asp Trp Ala Gly Asp Glu Arg Asn Val Val Leu Thr Leu Ser Arg

180 185 190

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

195 200 205

Ala Asp Trp Ala Met Glu Arg Leu Pro Ala Gln Tyr Gln Pro Val Ile

210 215 220

Leu Glu Ala Arg Gln Ala Tyr Leu Gly Gln Glu Glu Asp Arg Leu Ala

225 230 235 240

Ser Arg Ala Asp Gln Leu Glu Glu Phe Val His Tyr Val Lys Gly Glu

245 250 255

Ile Thr Lys Val Val Gly

260

<210> 155

<211> 126

<212> PRT

<213> Artificial Sequence

<220><223> selection marker gene ble conferring resistance to Bleomycin

<400> 155

Met Thr Asp Gln Ala Thr Pro Asn Leu Pro Ser Arg Asp Phe Asp Ser

1 5 10 15

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

20 25 30

Ala Gly Trp Met Ile Leu Gln Arg Gly Asp Leu Met Leu Glu Phe Phe

35 40 45

Ala His Pro Gly Leu Asp Pro Leu Ala Ser Trp Phe Ser Cys Cys Leu

50 55 60

Arg Leu Asp Asp Leu Ala Glu Phe Tyr Arg Gln Cys Lys Ser Val Gly

65 70 75 80

Ile Gln Glu Thr Ser Ser Gly Tyr Pro Arg Ile His Ala Pro Glu Leu

85 90 95

Gln Glu Trp Gly Gly Thr Met Ala Ala Leu Val Asp Pro Asp Gly Thr

100 105 110

Leu Leu Arg Leu Ile Gln Asn Glu Leu Leu Ala Gly Ile Ser

115 120 125

<210> 156

<211> 219

<212> PRT

<213> Artificial Sequence

<220><223> selection marker gene cat conferring resistance to Chloramphenicol

<400> 156

Met Glu Lys Lys Ile Thr Gly Tyr Thr Thr Val Asp Ile Ser Gln Trp

1 5 10 15

His Arg Lys Glu His Phe Glu Ala Phe Gln Ser Val Ala Gln Cys Thr

20 25 30

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

35 40 45

Lys Lys Asn Lys His Lys Phe Tyr Pro Ala Phe Ile His Ile Leu Ala

50 55 60

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

65 70 75 80

Glu Leu Val Ile Trp Asp Ser Val His Pro Cys Tyr Thr Val Phe His

85 90 95

Glu Gln Thr Glu Thr Phe Ser Ser Leu Trp Ser Glu Tyr His Asp Asp

100 105 110

Phe Arg Gln Phe Leu His Ile Tyr Ser Gln Asp Val Ala Cys Tyr Gly

115 120 125

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

130 135 140

Val Ser Ala Asn Pro Trp Val Ser Phe Thr Ser Phe Asp Leu Asn Val

145 150 155 160

Ala Asn Met Asp Asn Phe Phe Ala Pro Val Phe Thr Met Gly Lys Tyr

165 170 175

Tyr Thr Gln Gly Asp Lys Val Leu Met Pro Leu Ala Ile Gln Val His

180 185 190

His Ala Val Cys Asp Gly Phe His Val Gly Arg Met Leu Asn Glu Leu

195 200 205

Gln Gln Tyr Cys Asp Glu Trp Gln Gly Gly Ala

210 215

<210> 157

<211> 381

<212> PRT

<213> Artificial Sequence

<220><223> selection marker gene ermE conferring resistance to Erythromycin

<400> 157

Met Ser Ser Ser Asp Glu Gln Pro Arg Pro Arg Arg Arg Asn Gln Asp

1 5 10 15

Arg Gln His Pro Asn Gln Asn Arg Pro Val Leu Gly Arg Thr Glu Arg

20 25 30

Asp Arg Asn Arg Arg Gln Phe Gly Gln Asn Phe Leu Arg Asp Arg Lys

35 40 45

Thr Ile Ala Arg Ile Ala Glu Thr Ala Glu Leu Arg Pro Asp Leu Pro

50 55 60

Val Leu Glu Ala Gly Pro Gly Glu Gly Leu Leu Thr Arg Glu Leu Ala

65 70 75 80

Asp Arg Ala Arg Gln Val Thr Ser Tyr Glu Ile Asp Pro Arg Leu Ala

85 90 95

Lys Ser Leu Arg Glu Lys Leu Ser Gly His Pro Asn Ile Glu Val Val

100 105 110

Asn Ala Asp Phe Leu Thr Ala Glu Pro Pro Pro Glu Pro Phe Ala Phe

115 120 125

Val Gly Ala Ile Pro Tyr Gly Ile Thr Ser Ala Ile Val Asp Trp Cys

130 135 140

Leu Glu Ala Pro Thr Ile Glu Thr Ala Thr Met Val Thr Gln Leu Glu

145 150 155 160

Phe Ala Arg Lys Arg Thr Gly Asp Tyr Gly Arg Trp Ser Arg Leu Thr

165 170 175

Val Met Thr Trp Pro Leu Phe Glu Trp Glu Phe Val Glu Lys Val Asp

180 185 190

Arg Arg Leu Phe Lys Pro Val Pro Lys Val Asp Ser Ala Ile Met Arg

195 200 205

Leu Arg Arg Arg Ala Glu Pro Leu Leu Glu Gly Ala Ala Leu Glu Arg

210 215 220

Tyr Glu Ser Met Val Glu Leu Cys Phe Thr Gly Val Gly Gly Asn Ile

225 230 235 240

Gln Ala Ser Leu Leu Arg Lys Tyr Pro Arg Arg Arg Val Glu Ala Ala

245 250 255

Leu Asp His Ala Gly Val Gly Gly Gly Ala Val Val Ala Tyr Val Arg

260 265 270

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

275 280 285

Pro Arg Gly Gly Gln Pro Gln Arg Gly Arg Arg Thr Gly Gly Arg Asp

290 295 300

His Gly Asp Arg Arg Thr Gly Gly Gln Asp Arg Gly Asp Arg Arg Thr

305 310 315 320

Gly Gly Arg Asp His Arg Asp Arg Gln Ala Ser Gly His Gly Asp Arg

325 330 335

Arg Ser Ser Gly Arg Asn Arg Asp Asp Gly Arg Thr Gly Glu Arg Glu

340 345 350

Gln Gly Asp Gln Gly Gly Arg Arg Gly Pro Ser Gly Gly Gly Arg Thr

355 360 365

Gly Gly Arg Pro Gly Arg Arg Gly Gly Pro Gly Gln Arg

370 375 380

<210> 158

<211> 332

<212> PRT

<213> Artificial Sequence

<220

><223> selection marker gene hyg onferring resistance to Hygromycin

<400> 158

Met Thr Gln Glu Ser Leu Leu Leu Asp Arg Ile Asp Ser Asp Asp

1 5 10 15

Ser Tyr Ala Ser Leu Arg Asn Asp Gln Glu Phe Trp Glu Pro Leu Ala

20 25 30

Arg Arg Ala Leu Glu Glu Leu Gly Leu Pro Val Pro Pro Val Leu Arg

35 40 45

Val Pro Gly Glu Ser Thr Asn Pro Val Leu Val Gly Glu Pro Asp Pro

50 55 60

Val Ile Lys Leu Phe Gly Glu His Trp Cys Gly Pro Glu Ser Leu Ala

65 70 75 80

Ser Glu Ser Glu Ala Tyr Ala Val Leu Ala Asp Ala Pro Val Pro Val

85 90 95

Pro Arg Leu Leu Gly Arg Gly Glu Leu Arg Pro Gly Thr Gly Ala Trp

100 105 110

Pro Trp Pro Tyr Leu Val Met Ser Arg Met Thr Gly Thr Thr Trp Arg

115 120 125

Ser Ala Met Asp Gly Thr Thr Asp Arg Asn Ala Leu Leu Ala Leu Ala

130 135 140

Arg Glu Leu Gly Arg Val Leu Gly Arg Leu His Arg Val Pro Leu Thr

145 150 155 160

Gly Asn Thr Val Leu Thr Pro His Ser Glu Val Phe Pro Glu Leu Leu

165 170 175

Arg Glu Arg Arg Ala Ala Thr Val Glu Asp His Arg Gly Trp Gly Tyr

180 185 190

Leu Ser Pro Arg Leu Leu Asp Arg Leu Glu Asp Trp Leu Pro Asp Val

195 200 205

Asp Thr Leu Leu Ala Gly Arg Glu Pro Arg Phe Val His Gly Asp Leu

210 215 220

His Gly Thr Asn Ile Phe Val Asp Leu Ala Ala Thr Glu Val Thr Gly

225 230 235 240

Ile Val Asp Phe Thr Asp Val Tyr Ala Gly Asp Ser Arg Tyr Ser Leu

245 250 255

Val Gln Leu His Leu Asn Ala Phe Arg Gly Asp Arg Glu Ile Leu Ala

260 265 270

Ala Leu Leu Asp Gly Ala Gln Trp Lys Arg Thr Glu Asp Phe Ala Arg

275 280 285

Glu Leu Leu Ala Phe Thr Phe Leu His Asp Phe Glu Val Phe Glu Glu

290 295 300

Thr Pro Leu Asp Leu Ser Gly Phe Thr Asp Pro Glu Glu Leu Ala Gln

305 310 315 320

Phe Leu Trp Gly Pro Pro Asp Thr Ala Pro Gly Ala

325 330

<210> 159

<211> 264

<212> PRT

<213> Artificial Sequence

<220><223> selection marker gene neo conferring resistance to Kanamycin

<400> 159

Met Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala Ala Trp Val

1 5 10 15

Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Gln Thr Ile Gly Cys Ser

20 25 30

Asp Ala Ala Val Phe Arg Leu Ser Ala Gln Gly Arg Pro Val Leu Phe

35 40 45

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

50 55 60

Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys Ala Ala Val

65 70 75 80

Leu Asp Val Val Thr Glu Ala Gly Arg Asp Trp Leu Leu Leu Gly Glu

85 90 95

Val Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro Ala Glu Lys

100 105 110

Val Ser Ile Met Ala Asp Ala Met Arg Arg Leu His Thr Leu Asp Pro

115 120 125

Ala Thr Cys Pro Phe Asp His Gln Ala Lys His Arg Ile Glu Arg Ala

130 135 140

Arg Thr Arg Met Glu Ala Gly Leu Val Asp Gln Asp Asp Leu Asp Glu

145 150 155 160

Glu His Gln Gly Leu Ala Pro Ala Glu Leu Phe Ala Arg Leu Lys Ala

165 170 175

Arg Met Pro Asp Gly Glu Asp Leu Val Val Thr His Gly Asp Ala Cys

180 185 190

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

195 200 205

Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile Ala Leu Ala

210 215 220

Thr Arg Asp Ile Ala Glu Glu Leu Gly Gly Glu Trp Ala Asp Arg Phe

225 230 235 240

Leu Val Leu Tyr Gly Ile Ala Ala Pro Asp Ser Gln Arg Ile Ala Phe

245 250 255

Tyr Arg Leu Leu Asp Glu Phe Phe

260

<210> 160

<211> 1644

<212> DNA

<213> Artificial Sequence

<220><223> selection marker gene amdSYM

<400> 160

atgccccagt cctgggagga gctggcggcc gacaagcgcg cgcgcctcgc gaagacgatc 60

ccggacgagt ggaaggtcca gacgctgccc gcggaggact ccgtgatcga cttccccaag 120

aagtcgggga tcctctccga ggcggagctg aagatcaccg aagcctccgc ggccgacctg 180

gtcagcaagc tggcggccgg cgagctgacc agcgtcgaag tcaccctggc cttctgcaag 240

cgggcggcca tcgcgcagca gctcacgaac tgcgcccacg agttcttccc cgacgccgcc 300

ctcgcccagg cgcgcgagct ggacgagtac tacgccaagc acaagcgccc ggtggggccc 360

ctccacgggc tgccgatctc gctgaaggac cagctccggg tcaagggcta cgagacctcg 420

atggggtaca tctcgtggct gaacaagtac gacgagggcg actccgtcct gaccaccatg 480

ctgcggaagg ccggcgccgt cttctacgtc aagacctcgg tcccgcagac cctcatggtg 540

tgcgagacgg tgaacaacat catcggccgg accgtgaacc cccggaacaa gaactggtcc 600

tgcggcggct cctccggcgg ggaggggggcc atcgtcggca tccgcggcgg cgtcatcggc 660

gtgggcaccg acatcggcgg ctccatccgg gtgcccgccg ccttcaactt cctctacggc 720

ctgcgcccgt cccacgggcg cctcccgtac gcgaagatgg ccaactccat ggagggccag 780

gagaccgtgc actcggtggt gggccccatc acccactcgg tcgaagacct gcgcctgttc 840

acgaagagcg tcctgggcca ggaaccgtgg aagtacgaca gcaaggtgat cccgatgccg 900

tggcgccagt ccgagtcgga catcatcgcc tccaagatca agaacggggg cctgaacatc 960

gggtactaca acttcgacgg caacgtgctc ccgcaccccc cgatcctgcg cggggtcgag 1020

accacggtgg ccgccctggc caaggccggc cacaccgtca cgccctggac cccgtacaag 1080

cacgacttcg gccacgacct catctcccac atctacgcgg cggacggcag cgccgacgtg 1140

atgcgcgaca tctcggcctc cggggaaccg gcgatcccca acatcaagga cctgctgaac 1200

cccaacatca aggccgtcaa catgaacgag ctgtgggaca cccacctgca gaagtggaac 1260

taccagatgg aatacctcga gaagtggcgc gaggccgagg agaaggcggg caaggagctg 1320

gacgcgatca tcgccccgat caccccccacc gcggcggtgc ggcacgacca gttccggtac 1380

tacggctacg cctcggtcat caacctcctg gacttcacct ccgtcgtcgt cccggtgacg 1440

ttcgcggaca agaacatcga caagaagaac gaatcgttca aggcggtctc ggagctggac 1500

gccctcgtgc aggaggagta cgacccggaa gcctaccacg gcgccccggt cgccgtccag 1560

gtgatcgggc gccgcctgtc ggaggagcgc accctcgcga tcgccgagga ggtgggcaag 1620

ctgctgggca acgtcgtgac gccc 1644

<210> 161

<211> 1215

<212> DNA

<213> Artificial Sequence

<220><223> Counter selection marker tetA gene

<400> 161

atgaaccgca ccgtgatgat ggcgctcgtc atcatcttcc tcgacgccat gggcatcggc 60

atcatcatgc cggtcctgcc ggccctgctg cgggagttcg tgggcaaggc gaacgtggcg 120

gagaactacg gcgtcctcct cgcgctgtac gccatgatgc aggtgatctt cgcgccgctg 180

ctcggccggt ggtcggaccg catcggccgc cggccggtcc tgctgctctc gctcctcggg 240

gcgaccctgg actacgccct catggcgacg gcgtccgtcg tgtgggtcct ctacctgggc 300

cggctgatcg ccggcatcac gggcgccacc ggcgccgtcg cggcgtcgac gatcgccgac 360

gtcaccccgg aggagtcgcg cacccactgg ttcggcatga tgggcgcctg cttcggcggc 420

gggatgatcg ccggccccgt gatcggcggc ttcgccggcc agctctcggt gcaggccccc 480

ttcatgttcg ccgccgccat caacggcctg gccttcctgg tgtcgctgtt catcctgcac 540

gagacccaca acgccaacca ggtgtccgac gaactgaaga acgaaaccat caacgagacg 600

acctcgtcga tccgggagat gatctccccc ctgtccggcc tgctcgtggt cttcttcatc 660

atccagctga tcggccagat ccccgcgacc ctctgggtgc tgttcggcga ggaacgcttc 720

gcctgggacg gcgtgatggt gggggtgtcc ctcgcggtgt tcgggctcac ccacgcgctg 780

ttccagggcc tggcggcggg cttcatcgcc aagcacctgg gcgagcgcaa ggccatcgcc 840

gtcgggatcc tggccgacgg ctgcggcctg ttcctgctgg cggtgatcac ccagtcctgg 900

atggtctggc cggtcctgct gctcctggcc tgcggcggga tcaccctgcc ggcgctccag 960

ggcatcatct ccgtccgcgt cggccaggtc gcgcaggggc agctgcaggg cgtgctgacg 1020

tcgctcaccc acctcacggc cgtgatcggg ccgctcgtgt tcgccttcct gtactccgcg 1080

acccgcgaga cctggaacgg ctgggtgtgg atcatcggct gcggcctgta cgtcgtggcc 1140

ctcatcatcc tgcgcttctt ccaccccggc cgggtgatcc accccatcaa caagtccgac 1200

gtccagcagc ggatc 1215

<210> 162

<211> 1251

<212> DNA

<213> Artificial Sequence

<220><223> Counter selection marker lacY gene

<400> 162

atgtactacc tgaagaacac caacttctgg atgttcggcc tgttcttctt cttctacttc 60

ttcatcatgg gcgcctactt cccgttcttc cccatctggc tgcacgacat caaccacatc 120

tcgaagtcgg acaccgggat catcttcgcg gccatctcgc tcttctcgct cctgttccag 180

ccgctcttcg ggctgctctc ggacaagctg ggcctccgca agtacctgct ctggatcatc 240

acgggcatgc tggtcatgtt cgcgccgttc ttcatcttca tcttcggccc cctcctgcag 300

tacaacatcc tggtggggtc gatcgtcggc gggatctacc tgggcttctg cttcaacgcc 360

ggggcgccgg ccgtcgaggc cttcatcgag aaggtctcgc gccggtcgaa cttcgagttc 420

gggcgcgccc ggatgttcgg ctgcgtgggc tgggccctct gcgcctccat cgtgggcatc 480

atgttcacga tcaacaacca gttcgtcttc tggctggggt ccggctgcgc cctcatcctc 540

gcggtgctgc tgttcttcgc caagaccgac gccccgagca gcgcgacggt cgcgaacgcc 600

gtgggggcga accactccgc cttctcgctc aagctcgcgc tggagctgtt ccggcagccc 660

aagctgtggt tcctgtcgct gtacgtcatc ggcgtcagct gcacgtacga cgtgttcgac 720

cagcagttcg ccaacttctt cacctcgttc ttcgccaccg gcgagcaggg cacccgggtc 780

ttcggctacg tgaccacgat gggggagctg ctcaacgcct cgatcatgtt cttcgccccc 840

ctgatcatca accgcatcgg cggcaagaac gccctcctcc tggccggcac catcatgtcc 900

gtccgcatca tcggctccag cttcgcgacc tccgccctgg aggtcgtgat cctgaagacc 960

ctgcacatgt tcgaggtccc gttcctcctg gtcggctgct tcaagtacat cacctcccag 1020

ttcgaggtcc gcttctcggc cacgatctac ctggtctgct tctgcttctt caagcagctg 1080

gcgatgatct tcatgagcgt cctcgcgggc aacatgtacg aaagcatcgg cttccagggc 1140

gcctacctgg tgctgggcct ggtggccctc ggcttcaccc tcatcagcgt cttcaccctc 1200

tccggcccgg gcccgctgtc cctgctccgc cgccaggtca acgaggtggc g 1251

<210> 163

<211> 1419

<212> DNA

<213> Artificial Sequence

<220><223> Counter selection marker sacB gene

<400> 163

atgaacatca agaagttcgc caagcgggcg accgtcctga ccttcaccac cgccctgctc 60

gcgggcgggg ccacccaggc cttcgccaag gagaacaccc agaagcccta caaggagacg 120
tacggggtgt cgcacatcac ccgccacgac atgctccaga tccccaagca gcagcagagc 180
gagaagtacc aggtcccgca gttcgaccag tccaccatca agaacatcga atcggccaag 240
ggcctcgacg tgtgggactc ctggcccctg cagaacgccg acggcaccgt ggccgagtac 300
aacgggtacc acgtggtgtt cgccctggcg ggctccccca aggacgccga cgacacctcg 360
atctacatgt tctaccagaa ggtcggcgac aacagcatcg actcctggaa gaacgcgggc 420
cgcgtcttca aggacagcga caagttcgac gcgaacgacg agatcctgaa ggagcagacc 480

caggagtggt ccggctccgc caccttcacg tccgacggca agatccggct cttctacacg 540
gacttctccg gcacgcacta cgggaagcag agcctcacca cggcgcaggt caacgtgtcg 600
aagtccgacg acaccctcaa gatcaacggc gtggaggacc acaagacgat cttcgacggc 660
gacggcaaga cctaccagaa cgtgcagcag ttcatcgacg agggcaacta cacgtcgggc 720
gacaaccaca cgctgcgcga cccccactac gtggaggaca aggggcacaa gtacctggtc 780
ttcgaggcca acaccggcac cgacaacggc taccagggcg aggaatccct gttcaacaag 840
gcgtactacg gcggcagcac gaacttcttc cgcaaggaga gccagaagct ccagcagtcg 900

gccaagaagc gggacgccga gctcgccaac ggcgcgctgg gcatggtgga gctgaacgac 960
gactacacgc tgaagaaggt catgaagccg ctcatcacct ccaacaccgt gacggacgag 1020
atcgagcggg cgaacgtctt caagatgaac ggcaagtggt acctgttcac cgactcccgc 1080
ggctccaaga tgaccatcga cggcatcaac tcgaacgaca tctacatgct gggctacgtc 1140
tccaacagcc tgaccgggcc gtacaagccg ctcaacaaga ccggcctggt gctccagatg 1200
ggcctggacc cgaacgacgt caccttcacc tactcccact tcgcggtgcc ccaggcgaag 1260
ggcaacaacg tggtcatcac ctcgtacatg acgaaccggg gcttcttcga ggacaagaag 1320

gccaccttcg ccccctcctt cctgatgaac atcaagggca agaagacctc cgtggtgaag 1380
aacagcatcc tggagcaggg ccagctcacc gtcaacaac 1419

<210> 164
<211> 1068
<212> DNA
<213> Artificial Sequence
<220><223> Counter selection marker pheS gene derived from S. erythraea
<400> 164

atgtccggtg cgaacgaccc ctacgacccc aagcaggtgg ccgcgctgtc cgccgaaacc 60

ctggaacggg cggtggccga cgcgcgggaa gccttcgaca aggccggtga cctcgacgaa 120

ctggccgccg ccaagccggc ccacctgggt gaacggagcc cgctgctgac ggcccggcgg 180

gagatcggtg ccctgccccc gaaggcccgc tccgacgcgg gtaagcgcgt gaacgaggcg 240

cgggaggcga tccagggcgc cttcgacgag cggcgggcgg ccctccaggc ggaacgcgac 300

gaacgggtgc tgcgcgaaga agccgtcgac gtcaccctcc cctgggaccg cgtgcccgtg 360

ggtgcgcgcc acccgatcac ccagctgatc gagcacgtgg ccgacacgtt cgtggccatg 420

ggttgggaag tcgccgaagg ccccgagctc gagaccgaat ggttcaactt cgacgccctg 480

aacttcggca aggaccaccc ggcgcgcacc atgcaggaca ccttctacgt cggtccgaag 540

gaatccggtc tcgtcctccg gacgcacacc agcccggccc aggtccgcgc cctgctggac 600

cgggaactgc cggtgtacgt ggtgtgcccc ggccgcacct tccggaccga cgagctggac 660

tccacgcaca cgccggtctt ccaccaggtg gaagggctcg ccgtggacaa gggtctcacg 720

atggcccacc tcaagggcac gctcgacgcg ttcgcgcgcg tcatgttcgg tcccgaatcc 780

aagacgcgcc tgcgcccgag cttcttcccg ttcgccgaac cctcgggtga agtcgacgtc 840

tggttcccgc agaagaaggg cggtcccggc tgggtcgaat ggggcggctg cggtatggtc 900

aacccgaacg tcctgcgcgc ctgcggtgtc gaccccgaaa cccacaccgg tttcggcttc 960

gggatgggtc tcgaacggac gctccagttc cgcaacggta tcccggacat gcgggacatg 1020

gtggaaggtg acgtgcagtt cacgcagccc ttcggtatcg actcctga 1068

<210> 165

<211> 1038

<212> DNA

<213> Artificial Sequence

<220><223> Counter selection marker pheS gene derived from Corynebacterium

<400> 165

atgagcgaga tccagctgac cgaggcctcg ctcaacgagg ccgccgacgc cgccatcaag 60

gccttcgacg gcgcgcagaa cctcgacgaa ctcgcggcgc tgcgccggga ccacctgggc 120

gacgccgccc ccatcccgca ggcccgccgc tccctcggga ccatcccgaa ggaccagcgc 180

aaggacgcgg gtcgcttcgt gaacatggcc ctcggtcgcg cggaaaagca cttcgcccag 240

gtcaaggtgg tgctcgaaga aaagcgcaac gcggaggtcc tcgagctcga acgggtggac 300

gtgaccgtcc cgaccacccg ggaacaggtg ggtgcgctcc acccgatcac catcctgaac 360

gaacagatcg cggacatctt cgtcggtatg ggctgggaaa tcgccgaagg tccggaggtg 420

gaggcggagt acttcaactt cgacgcgctc aacttcctcc ccgaccaccc cgcgcgcacg 480

ctccaggaca ccttccacat cgcgcccgag ggttcgcgcc aggtgctgcg gacgcacacc 540

tccccggtgc aggtccgcac catgctgaac cgcgaagtgc ccatctacat cgcgtgcccg 600

ggtcgggtgt tccgcacgga cgaactcgac gcgacccaca ccccggtctt ccaccagatc 660

gaggggctgg cggtcgacaa gggtctgacg atggcccacc tgcgcgggac gctggaccac 720

ctggccaagg agctgttcgg gccggaaacc aagacgcgca tgcgctccaa ctacttcccg 780

ttctcggagc cctccgcgga ggtcgacgtc tggttcccga acaagaaggg tggggccggc 840

tggatcgaat ggggcgggtg cggcatggtg aaccccaacg tcctccgcgc cgtgggtgtc 900

gacccggaag agtacaccgg gttcggcttc ggcatgggca tcgaacggac cctgcagttc 960

cgcaacggtc tgagcgacat gcgggacatg gtcgagggtg acatccggtt cacgctcccg 1020

ttcggcatcc aggcctga 1038

<210> 166

<211> 668

<212> DNA

<213> Artificial Sequence

<220><223> GST Solubility Tag

<400> 166

atgtccccta tactaggtta ttggaaaatt aagggccttg tgcaacccac tcgacttctt 60

ttggaatatc ttgaagaaaa atatgaagag catttgtatg agcgcgatga aggtgataaa 120

tggcgaaaca aaaagtttga attgggtttg gagtttccca atcttcctta ttatattgat 180

ggtgatgtta aattaacaca gtctatggcc atcatacgtt atatagctga caagcacaac 240

atgttgggtg gttgtccaaa agagcgtgca gagatttcaa tgcttgaagg agcggttttg 300

gatattagat acggtgtttc gagaattgca tatagtaaag actttgaaac tctcaaagtt 360

gattttctta gcaagctacc tgaaatgctg aaaatgttcg aagatcgttt atgtcataaa 420

acatatttaa atggtgatca tgtaacccat cctgacttca tgttgtatga cgctcttgat 480

gttgttttat acatggaccc aatgtgcctg gatgcgttcc caaaattagt ttgttttaaa 540

aaacgtattg aagctatccc acaaattgat aagtacttga aatccagcaa gtatatagca 600

tggcctttgc agggctggca agccacgttt ggtggtggcg accatcctcc aaaatcggat 660

ctggttcc 668

<210> 167

<211> 150

<212> DNA

<213> Saccharopolyspora endophytica

<400> 167

gcgagaggcc cggaagcgag atcgcttccg ggcctctgac ctgcggagga tacgggattc 60

gaacccgtga gggctattaa cccaacacga tttccaattc cgatggcgcg agtgccaggg 120

ggtagctgaa cgtgcctttt gcctggtcag 150

<210> 168

<211> 45

<212> DNA

<213> Saccharopolyspora erythraea

<400> 168

tcggagccgc tgaggggact cgaacccctg accgtccgct tacaa 45

<210> 169

<211> 52

<212> DNA

<213> Saccharopolyspora spinosa

<400> 169

ggcagctctt ggtggtggcc aggggcgggg tcgaaccgcc gaccttccgc tt 52

<210> 170

<211> 48

<212> DNA

<213> Saccharopolyspora spinosa

<400> 170

tcggagccgc tgaggggact cgaacccctg accgtccgct tacaaggc 48

<210> 171

<211> 76

<212> DNA

<213> Saccharopolyspora spinosa

<400> 171

ggagccgcct aagggaatcg aacccttgac ctacgcatta cgagtgcgtc gctctagccg 60

actgagctaa ggcggc 76

<210> 172

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> synthetic promoter Pmut-1

<400

> 172

tgtgcggtgg ctaacacgtc ctagtatggt atcatgagca a 41

<210> 173

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> synthetic promoter B2

<400> 173

tgtgcgctgg ctaacacgtc ctagtatggt atagtgagca a 41

<210> 174

<211> 41

<212> DNA

<213> Artificial Sequence

<220><223> synthetic promoter D1

<400> 174

tgtgcggttc ctaacacgtc ctagtatggt actatgagca a 41

<210> 175

<211> 41

<212> DNA

<213

> Artificial Sequence

<220><223> synthetic promoter D2

<400> 175

tgtgcggtgg ctaacacgtc ctagtatggt atcatgagca a 41