



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2016년09월30일
(11) 등록번호 10-1661519
(24) 등록일자 2016년09월26일

(51) 국제특허분류(Int. Cl.)
C12N 15/52 (2006.01) C12N 1/21 (2006.01)
C12N 9/00 (2006.01) C12P 13/04 (2006.01)
(21) 출원번호 10-2010-7019996
(22) 출원일자(국제) 2009년02월19일
심사청구일자 2013년12월16일
(85) 번역문제출일자 2010년09월07일
(65) 공개번호 10-2010-0121642
(43) 공개일자 2010년11월18일
(86) 국제출원번호 PCT/US2009/034534
(87) 국제공개번호 WO 2009/105551
국제공개일자 2009년08월27일
(30) 우선권주장
61/030,835 2008년02월22일 미국(US)
(56) 선행기술조사문헌
KR1020010084864 A

(73) 특허권자
스캐랩 게노믹스, 엘엘씨
미국, 위스콘신주 53713, 메디슨, 앤 스트리트
1202
(72) 발명자
블래트너, 프레데릭, 알.
미국, 위스콘신 53711, 제퍼슨 스트리트 메디슨
1547
김선창
대전광역시 유성구 궁동로 5, 다솔 103동 702호
(궁동)
이준형
대전광역시 유성구 장대로80번길 41 (장대동)
(74) 대리인
신관호

전체 청구항 수 : 총 14 항

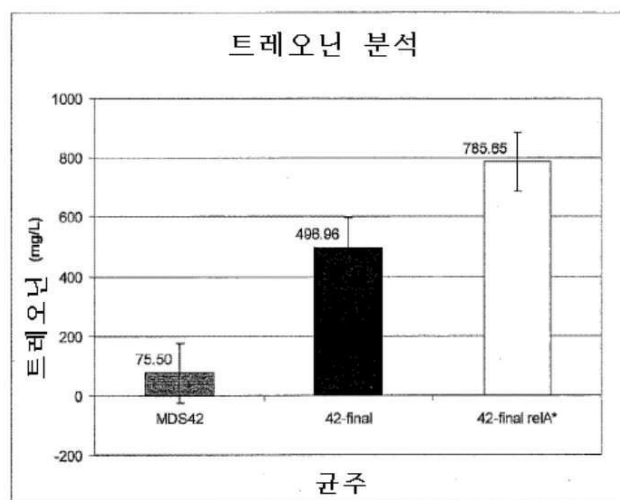
심사관 : 이현석

(54) 발명의 명칭 아미노산 생합성을 위한 조성물들과 방법들

(57) 요약

일반적으로, 본 발명은 특히 아미노산류와 같은 상업적인 생물학적 제품을 생성하는데 유용한 방법들 및 조성물들과 관련된다. 더욱 상세하게는, 본 발명은, 상업적인 제품의 생산을 위하여 유전적으로 변형된 미생물의 균주들과 이들의 이용에 관한 것이다. 본 발명은 또한, 특히, 신규의 분리된 새로운 DNA, 핵산, 벡터들 및 감소된 계놈 박테리아를 제공한다.

대표도 - 도1



명세서

청구범위

청구항 1

- (a) SEQ ID NO: 1,
- (b) SEQ ID NO: 2,
- (c) 상기 뉴클레오타이드 서열들 (a)-(b)중에서 어느 한 개의 보체 및
- (d) 상기 뉴클레오타이드 서열들 (a)-(b)중에서 어느 한 개의 축퇴된 서열로 구성되는 그룹으로부터 선택되는 한 개의 뉴클레오타이드 서열을 포함하는 분리된 핵산.

청구항 2

제 1항에 있어서,
상기 분리된 핵산은 분리된 DNA가 되는 분리된 핵산.

청구항 3

- 제 1항에 있어서,
상기 핵산은,
- (a) SEQ ID NO: 1,
 - (b) SEQ ID NO: 2,
 - (c) 상기 뉴클레오타이드 서열들 (a)-(b)중에서 어느 한 개의 보체 및
 - (d) 상기 뉴클레오타이드 서열들 (a)-(b)중에서 어느 한 개의 축퇴된 서열로 구성되는 그룹으로부터 선택되는 어느 한 개의 뉴클레오타이드 서열을 포함하는 분리된 핵산.

청구항 4

발현 조절 서열에 작동 가능하게 링크된 청구항 1항의 상기 핵산을 포함하는 발현 벡터.

청구항 5

청구항 4항의 벡터를 포함하는 배양된 대장균.

청구항 6

청구항 4항의 벡터로 형질감염된 배양된 대장균.

청구항 7

relA 유전자를 함유하는 게놈을 포함하며, 상기 *relA* 유전자는, 상기 *relA* 유전자의 암호 서열의 위치 547에서 적어도 하나의 점 돌연변이를 함유하며,

상기 점 돌연변이는 SEQ ID NO: 1에 명시된 위치 547에서의 G→A 돌연변이, SEQ ID NO: 2에 명시된 위치 547에서의 G→T 돌연변이 중의 하나 이상으로부터 선택되는 재조합형 대장균.

청구항 8

제 7항에 있어서,

상기 대장균의 상기 게놈은, 고유 모균주의 게놈보다 적어도 5% 더 작게 되도록 유전자 조작되어 있는 재조합형 대장균.

청구항 9

아미노산의 생합성하는 방법에 있어서,

상기 방법은, 청구항 7항의 대장균을 제공하며, 상기 대장균의 아미노산을 발현하는 아미노산의 생합성 방법.

청구항 10

제 9항에 있어서,

상기 아미노산은 Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Val, Leu, Lys, Met, Phe, Pro, Ser, Thr, Tyr 및 Trp로 이루어지는 그룹에서 선택되는 아미노산의 생합성 방법.

청구항 11

아미노산의 생합성 방법에 있어서,

(1) 청구항 8항의 대장균을 제공하기;

(2) 프로모터에 의해 작동 가능하도록 조절되는 아미노산을 위하여 하나 이상의 생합성 유전자들을 함유하는 플라스미드를 상기 대장균에 도입하기;

(3) 상기 대장균으로부터 아미노산을 발현하기를 포함하는 아미노산의 생합성 방법.

청구항 12

삭제

청구항 13

삭제

청구항 14

제 11항에 있어서,

상기 아미노산을 위한 상기 하나 이상의 생합성 유전자들은 재조합 타크(tac) 프로모터에 의해 작동 가능하도록 조절되는 *thrABC*를 포함하는 아미노산의 생합성 방법.

청구항 15

삭제

청구항 16

삭제

청구항 17

제 7항에 있어서,

relA 유전자들을 함유하는 게놈을 포함하며, 상기 *relA* 유전자는, 상기 *relA* 유전자의 암호 서열의 위치 547에서 하나의 점 돌연변이를 함유하며,

상기 점 돌연변이는 SEQ ID NO: 1에 명시된 위치 547에서의 G->A 돌연변이가 되는 재조합형 대장균.

청구항 18

제 8항에 있어서,

relA 유전자들을 함유하는 게놈을 포함하며, 상기 *relA* 유전자는, 상기 *relA* 유전자의 암호 서열의 위치 547에서 하나의 점 돌연변이를 함유하며,

상기 점 돌연변이는 SEQ ID NO: 1에 명시된 위치 547에서의 G->A 돌연변이가 되는 재조합형 대장균.

발명의 설명

기술 분야

- [0001] 본 출원은 2008년 2월 22일에 출원된 미국 가출원 번호 제61/030,835호의 이익을 주장하며, 여기에 참조로서 상기 내용들이 포함되어 있다.
- [0002] 일반적으로, 본 발명은 특히 아미노산류와 같은 상업적인 생물학적 제품들을 생산하는데 유용한 방법들 및 조성물들과 관련된다. 더욱 상세하게는, 본 발명은, 유전적으로 변형된 미생물의 균주들과 상업적인 제품들의 생산을 위한 이들의 이용에 관한 것이다. 본 발명은 또한 특히 신규의 분리된(isolated) DNA, 핵산, 벡터들 및 감소된(reduced) 계능 박테리아를 제공한다.

발명의 내용

해결하려는 과제

- [0003] 일 실시예에 있어서, 본 발명은 분리된 DNA를 제공하며, 이것의 뉴클레오티드 서열은 SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 또는 SEQ ID NO: 4 또는 이들의 보체(complement) 중 어느 하나를 포함하거나 또는 그들 중 어느 하나로 구성된다.
- [0004] 또 하나의 실시예에 있어서, 본 발명은 SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 또는 SEQ ID NO: 4 또는 이들의 보체 중 어느 하나를 포함하거나 또는 그들 중 어느 하나로 구성되는 분리된 핵산을 제공한다.
- [0005] 또 하나의 실시예에 있어서, 본 발명은 SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 또는 SEQ ID NO: 4(또는 이들의 보체)의 뉴클레오티드 서열을 포함하거나 이들로 구성되는, 또는 이들의 축퇴된 서열(degenerate variant)을 포함하거나 이들로 구성되는 분리된 핵산을 제공한다.
- [0006] 또 하나의 실시예에 있어서, 본 발명은 발현 벡터, 예를 들어, 발현 조절 서열에 작동 가능하게 링크된 SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 또는 SEQ ID NO: 4의 핵산을 포함하는 플라스미드를 제공한다. 또 하나의 실시예에 있어서, 본 발명은 상기 벡터들 중 하나를 포함하는 배양 세포를 제공한다. 또 다른 실시예에 있어서, 본 발명은 상기 벡터들 중 하나 또는 상기 세포의 자세포(progeny)로 형질감염된(transfected) 배양 세포를 제공한다.
- [0007] 또 다른 실시예에 있어서, 본 발명은 감소된 계능 박테리움, 예를 들면, *relA* 유전자를 함유하는 계능을 가지는 대장균(*E. coli*)을 제공하며, 여기서, *relA* 유전자는 최소한 한 개의 점 돌연변이를 함유한다. 관련된 실시예에 있어서, 상기 돌연변이는 *relA* 유전자의 위치 547 또는 548에 있다.
- [0008] 또 다른 실시예에 있어서, 본 발명은 유전자 조작되어 원래의 모균주(parent strain)의 계능보다 적어도 5%, 적어도 8% 또는 적어도 10%가 더 작게 되는 계능을 가지는 박테리움을 제공하며, 여기서 상기 박테리움은 적어도 하나의 *relA* 돌연변이를 포함한다. 일 실시예에 있어서, 상기 적어도 하나의 *relA* 돌연변이는 상기 *relA* 유전자의 위치 547 또는 548에서 발생한다. 또 하나의 실시예에 있어서, 상기 돌연변이는 상기 *relA* 유전자의 위치 547에서의 G→A 점 돌연변이, 상기 *relA* 유전자의 위치 547에서의 G→T 점 돌연변이, 상기 *relA* 유전자의 위치 548에서의 C→G 점 돌연변이, 또는 상기 *relA* 유전자의 위치 548에서의 C→T 점 돌연변이 중 하나 이상으로부터 선택된다. 또 하나의 실시예에 있어서, 상기 박테리움은 적은 규모의 수, 예를 들어, 상기 *relA* 유전자의 위치 540과 550 사이에서 발생하는 1 내지 약 10, 1 내지 약 8 또는 1 내지 약 5 *relA* 점 돌연변이들을 포함한다.
- [0009] 또 하나의 실시예에 있어서, 본 발명은 본 명세서에서 기술하는 바와 같이, 하나 이상의 감소된 계능 박테리움을 사용하여 아미노산을 생합성하는 방법을 제공한다. 상기 방법은: (1) 고유(native) *relA* 유전자를 포함하는 DNA를 가지는 박테리움을 제공하기; (2) *relA* 돌연변이 박테리움을 형성하기 위하여 고유 *relA* 유전자를 돌연변이 *relA* 유전자로 대체하기; 및 (3) 상기 돌연변이 *relA* 박테리움으로부터 아미노산을 표현하기를 포함한다. 일 실시예에 있어서, 상기 아미노산은 Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Val, Leu, Lys, Met, Phe, Pro, Ser, Thr, Tyr 및 Trp로부터 선택된다.

도면의 간단한 설명

- [0010] 도 1은, MDS42 세포 라인, 트레오닌을 생성하기 위해 조작된 MDS42 세포 라인, 그리고 *relA* 점 돌연변이를 가지는 트레오닌을 생성하기 위해 조작된 MDS42 세포 라인에 의한 트레오닌 생성을 보여준다.

도 2는, MDS42의 *relA*를, 유전자내의 위치 547에서, G에서 A로의 전환(transversion)을 포함하는 단일 점 돌연변이를 함유하는 돌연변이 *relA* 유전자로 대체하는 스킴(scheme)을 도시한 것이다.

발명을 실시하기 위한 구체적인 내용

- [0011] 본 발명은 여러 가지 형태로 구현될 수 있는 한편, 몇 가지 실시예들의 하기 서술은 본 개시가 상기 발명의 대표적인 예시로 간주된다는 이해로 만들어진 것이며, 본 발명을 기재된 특정 실시예들에 한정하려고 하는 것은 아니다. 표제들은 단지 편의를 위해 제공되며 어떤 방식으로든 본 발명을 제한하는 것으로 해석되지 않는다. 어느 표제 하에 기재된 실시예들은 다른 표제 하에 기재된 실시예들과 결합될 수 있다.
- [0012] 본 발명에 기재된 어떤 수 또는 데이터에 의해 형성될 수 있는 어느 범위들, 비율들 및 비율들의 범위들은 본 발명의 추가적인 실시예들을 나타낸다는 것이 이해될 것이다. 이것은 상한 및/또는 하한 경계를 포함하거나 또는 포함하지 않고 형성될 수 있는 범위들을 포함한다. 그러므로, 당업자들은 본 명세서에서 제시되는 데이터와 수들로부터 다수의 그러한 비율들, 범위들 및 비율들의 범위들이 명확하게 도출될 수 있으며, 모두 본 발명의 실시예들을 나타낼 수 있다는 것을 알게 될 것이다.
- [0013] 본 화합물들, 생성물들과 조성물들 및 방법들이 개시되고 설명되기 전에, 여기에서 사용되는 용어는 단지 특정한 실시예들을 설명하기 위한 것이며, 제한하려고 의도하는 것은 아니다. 명세서와 첨부된 청구범위에서 사용되는 바와 같이, 단일한 형태들, "a," "an"과 "the"는 문맥상에서 명확하게 달리 지시하지 않으면, 복수 개의 지시 대상물을 포함한다는 것이 주목되어야 한다.
- [0014] 아래의 표 1 내지 4는 돌연변이 *relA* DNA 서열들을 설명한다.
- [0015] 표 1. 위치 547 (SEQ ID NO: 1)에서의 G→A 돌연변이를 포함하는 대장균 *relA* 코딩 서열. 돌연변이된 아미노산에는 밑줄을 긋고 볼드체로 표시하였다.
- [0016] ATGGTTGCGGTAAGAAGTGCACATATCAATAAGGCTGGTGAATTTGATCC
- [0017] GGAAAAATGGATCGCAAGTCTGGGTATTACCAGCCAGAAGTCGTGTGAGT
- [0018] GCTTAGCCGAAACCTGGGCGTATTGTCTGCAACAGACGCAGGGGCATCCG
- [0019] GATGCCAGTCTGTTATTGTGGCGTGGTGTGAGATGGTGGAGATCCTCTCG
- [0020] ACATTAAGTATGGACATTGACACGCTGCGGGCGGCGCTGCTTTTCCCTCTG
- [0021] GCGGATGCCAACGTAGTCAGCGAAGATGTGCTGCGTGAGAGCGTCGGTAA
- [0022] GTCGGTCGTTAACCTTATTCACGGCGTGCCTGATATGGCGGCGATCCGCCA
- [0023] GCTGAAAGCGACGCACACTGATTCTGTTTCCCTCCGAACAGGTCGATAACGT
- [0024] TCGCCGGATGTTATTGGCGATGGTCGATGATTTTCGCTGCGTAGTCATCAA
- [0025] ACTGGCGGAGCGTATTGCTCATCTGCGCGAAGTAAAAGATGCGCCGGAAG
- [0026] ATGAACGTGTACTGGCGGCAAAAGAGGTACCAACATCTACACACCGCTG
- [0027] GCTAACCGTCTCGGAATCGGACAACCTGAAATGGGAACCTGGAAGATTACTG
- [0028] CTTCCGTTACCTCCATCCAACCGAATACAAACGAATTGCCAAACTGCTGCA
- [0029] TGAACGGCGTCTCGACCGCAACACTACATCGAAGAGTTCGTTGGTCATCT
- [0030] GCGCGCTGAGATGAAAGCTGAAGGCGTTAAAGCGGAAGTGATGGTCGTC
- [0031] CGAAACACATCTACAGCATCTGGCGTAAATGCAGAAAAAGAACCTCGCC
- [0032] TTTGATGAGCTGTTTGATGTGCGTGCGGTACGTATTGTGCGCGAGCGTTTA
- [0033] CAGGATTGCTATGCCGCACTGGGGATAGTGACACTCACTATCGCCACCTG
- [0034] CCGGATGAGTTTGACGATTACGTCGCTAACCCGAAACCAACGGTTATCA
- [0035] GTCTATTTCATACCGTGGTTCTGGGGCCGGGTGGAAAAACCGTTGAGATCCA

[0036]	AATCCGCACCAAACAGATGCATGAAGATGCAGAGTTGGGTGTTGCTGCGC
[0037]	ACTGGAATATAAGAGGGCGCGGCTGCTGGCGGCGCACGTTTCGGGACAT
[0038]	GAAGACCGGATTGCCTGGCTGCGTAACTGATTGCGTGGCAGGAAGAGAT
[0039]	GGCTGATTCCGGCGAAATGCTCGACGAAGTACGTAGTCAGGTCTTTGACG
[0040]	ACCGGTGTACGTCTTTACGCCGAAAGGTGATGTCGTTGATTGCTGCGG
[0041]	GATCAACGCCGTGGACTTCGCTTACCACATCCACAGTGATGTCGGACACC
[0042]	GCTGCATCGGGGCAAAATTTGGCGGGCGCATTTGTGCCGTTACCTACCAG
[0043]	CTGCAGATGGGCGACCAGATTGAAATTATCACCCAGAAACAGCCGAACCC
[0044]	CAGCCGTGACTGGTTAAACCCAAACCTCGGTTACGTCACAACCAGCCGTG
[0045]	GGCGTTCGAAAATTCACGCCTGGTTCGGTAAACAGGACCGTGACAAAAAC
[0046]	ATTCTGGCTGGGCGGCAATCCTTGACGACGAGCTGGAACATCTGGGGAT
[0047]	CAGCCTGAAAGAAGCAGAAAAACATCTGCTGCCGCGTTACAACCTCAATG
[0048]	ATGTCGACGAGTTGCTGGCGGCGATTGGTGGCGGGGATATCCGTCTCAATC
[0049]	AGATGGTGAACCTTCTGCAATCGCAATTTAATAAGCCGAGTGCCGAAGAG
[0050]	CAGGACGCCGCGCGCTGAAGCAACTTCAGCAAAAAAGCTACACGCCGCA
[0051]	AAACCGCAGTAAAGATAACGGTCGCGTGGTAGTCGAAGGTGTTGGCAACC
[0052]	TGATGCACCACATCGCGCGCTGCTGCCAGCCGATTCTGGAGATGAGATTG
[0053]	TCGGCTTCATTACCCAGGGGCGCGGTATTTAGTACACCGCGCCGATTGCG
[0054]	AACAACCTGGCGGAACTGCGCTCCCATGCGCCAGAACGCATTGTTGACGCG
[0055]	GTATGGGGTGAGAGCTACTCCGCGGATATTCGCTGGTGGTCCGCTGGTA
[0056]	GCTAATGATCGTAGTGGGTGTGTACGTGATATCACGACCATTTCTGCCAAC
[0057]	GAGAAGGTGAACGTGCTTGGCGTTGCCAGCCGTAGCGACACCAACAGCA
[0058]	ACTGGCGACCATCGACATGACCATGAGATTTACAACCTGCAAGTGCTGG
[0059]	GGCGCGTGCTGGGTAACTCAACCAGGTGCCGGATGTTATCGACGCGCGT
[0060]	CGGTTGCACGGGAGTTAG
[0061]	표 2. 위치 547 (SEQ ID NO: 2)에서의 G->T 돌연변이를 포함하는 대장균 <i>relA</i> 코딩 서열. 돌연변이된 아미노산에는 밑줄을 긋고 볼드체로 표시하였다.
[0062]	ATGGTTGCGGTAAGAAGTGCACATATCAATAAGGCTGGTGAATTTGATCC
[0063]	GGAAAAATGGATCGCAAGTCTGGGTATTACCAGCCAGAAGTCGTGTGAGT
[0064]	GCTTAGCCGAAACCTGGGCGTATTGTCTGCAACAGACGCAGGGGCATCCG
[0065]	GATGCCAGTCTGTTATTGTGGCGTGGTGTGAGATGGTGGAGATCCTCTCG
[0066]	ACATTAAGTATGGACATTGACACGCTGCGGGCGGCGCTGCTTTTCCCTCTG
[0067]	GCGGATGCCAACGTAGTCAGCGAAGATGTGCTGCGTGAGAGCGTCGGTAA
[0068]	GTCGGTCGTTAACCTTATTCACGGCGTGCGTGATATGGCGGCGATCCGCCA
[0069]	GCTGAAAGCGACGCACACTGATTCTGTTTCTCCGAACAGGTCGATAACGT
[0070]	TCGCCGGATGTTATTGGCGATGGTCGATGATTTTCGCTGCGTAGTCATCAA

[0071] ACTGGCGGAGCGTATTGCTCATCTGCGCGAAGTAAAAGATGCGCCGGAAG

[0072] ATGAACGTGTACTGGCGGCAAAAGAGGTACCAACATCTACTCACCCTG

[0073] GCTAACCGTCTCGGAATCGGACAACCTGAAATGGAACTGGAAGATTACTG

[0074] CTTCCGTTACCTCCATCCAACCGAATACAAACGAATTGCCAACTGCTGCA

[0075] TGAACGGCGTCTCGACCGGAACACTACATCGAAGAGTTCGTTGGTCATCT

[0076] GCGCGCTGAGATGAAAGCTGAAGGCGTTAAAGCGGAAGTGATGGTCGTC

[0077] CGAAACACATCTACAGCATCTGGCGTAAAATGCAGAAAAAGAACCTCGCC

[0078] TTTGATGAGCTGTTTGTATGTGCGTGCGGTACGTATTGTGCGCGAGCGTTTA

[0079] CAGGATTGCTATGCCGCACTGGGGATAGTGCACACTCACTATCGCCACCTG

[0080] CCGGATGAGTTTGACGATTACGTGCTAACCCGAAACCAACCGTTATCA

[0081] GTCTATTTCATACCGTGGTTCTGGGGCCGGTGAAAAACCGTTGAGATCCA

[0082] AATCCGCACCAACAGATGCATGAAGATGCAGAGTTGGGTGTTGCTGCGC

[0083] ACTGGAAATATAAGAGGGCGCGGCTGCTGGCGGCGCACGTTCCGGACAT

[0084] GAAGACCGGATTGCCTGGCTGCGTAACTGATTGCGTGCGAGGAAGAGAT

[0085] GGCTGATTCCGGCGAAATGCTCGACGAAGTACGTAGTCAGGTCTTTGACG

[0086] ACCGGGTGTACGTCTTTACGCCGAAAGTGATGTCGTTGATTTGCCTGCGG

[0087] GATCAACGCCGTGGACTTCGCTTACCACATCCACAGTGATGTCGGACACC

[0088] GCTGCATCGGGCAAAAATTGGCGGGCGCATTGTGCCGTTACCTACCAG

[0089] CTGCAGATGGGCGACCAGATTGAAATTATCACCAGAAACAGCCGAACCC

[0090] CAGCCGTGACTGGTTAAACCCAAACCTCGGTTACGTACAAACCAGCCGTG

[0091] GGCGTTCGAAAATTCACGCCTGGTTCCGTAAACAGGACCGTGACAAAAAC

[0092] ATTCTGGCTGGGCGGCAAATCCTTGACGACGAGCTGGAACATCTGGGGAT

[0093] CAGCCTGAAAGAAGCAGAAAAACATCTGCTGCCGCGTTACAACCTCAATG

[0094] ATGTCGACGAGTTGCTGGCGGCGATTGGTGGCGGGGATATCCGTCTCAATC

[0095] AGATGGTGAACCTTCCTGCAATCGCAATTTAATAAGCCGAGTCCGAAGAG

[0096] CAGGACGCCGCCGCTGAAGCAACTTCAGCAAAAAAGCTACACGCCGCA

[0097] AAACCGCAGTAAAGATAACGGTCGCGTGGTAGTCGAAGGTGTTGGCAACC

[0098] TGATGCACCACATCGCGCGCTGCTGCCAGCCGATTCTGGAGATGAGATTG

[0099] TCGGCTTCATTACCCAGGGGCGCGGTATTTAGTACACCGCGCCGATTGCG

[0100] AACAACTGGCGGAACTGCGCTCCCATGCGCCAGAACGCATTGTTGACGCG

[0101] GTATGGGGTGAGAGCTACTCCGCCGGATATTCGCTGGTGGTCCGCGTGGTA

[0102] GCTAATGATCGTAGTGGGTGTTACGTGATATCACGACCATTCGCCAAC

[0103] GAGAAGGTGAACGTGCTTGGCGTTGCCAGCCGTAGCGACACCAACAGCA

[0104] ACTGGCGACCATCGACATGACCATTGAGATTTACAACCTGCAAGTGCTGG

[0105] GGCGCGTGCTGGGTAACTCAACCAGGTGCCGGATGTTATCGACGCGCT

[0106] CGGTTGCACGGGAGTTAG

[0107]	표3. 위치 548 (SEQ ID NO: 3)에서의 C->G 돌연변이를 포함하는 대장균 <i>relA</i> 코딩 서열. 돌연변이된 아미노산에는 밑줄을 긋고 볼드체로 표시하였다.
[0108]	ATGGTTGCGGTAAGAAGTGCACATATCAATAAGGCTGGTGAATTTGATCC
[0109]	GGAAAAATGGATCGCAAGTCTGGGTATTACCAGCCAGAAGTCGTGTGAGT
[0110]	GCTTAGCCGAAACCTGGGCGTATTGTCTGCAACAGACGCAGGGGCATCCG
[0111]	GATGCCAGTCTGTTATTGTGGCGTGGTGTGAGATGGTGGAGATCCTCTCG
[0112]	ACATTAAGTATGGACATTGACACGCTGCGGGCGGCGCTGCTTTTCCCTCTG
[0113]	GCGGATGCCAACGTAGTCAGCGAAGATGTGCTGCGTGAGAGCGTCGGTAA
[0114]	GTCGGTCGTTAACCTTATTCACGGCGTGCGTGATATGGCGGCGATCCGCCA
[0115]	GCTGAAAGCGACGCACACTGATTCTGTTTCCCTCCGAACAGGTCGATAACGT
[0116]	TCGCCGGATGTTATTGGCGATGGTCGATGATTTTCGCTGCGTAGTCATCAA
[0117]	ACTGGCGGAGCGTATTGCTCATCTGCGCGAAGTAAAAGATGCGCCGGAAG
[0118]	ATGAACGTGTACTGGCGGCAAAAGAGGTACCAACATCTACG G ACCGCTG
[0119]	GCTAACCGTCTCGGAATCGGACAACCTGAAATGGGAAGTGAAGATTACTG
[0120]	CTTCCGTTACCTCCATCCAACCGAATACAAACGAATTGCCAAACTGCTGCA
[0121]	TGAACGGCGTCTCGACCGGAACACTACATCGAAGAGTTCGTTGGTCATCT
[0122]	GCGCGCTGAGATGAAAGCTGAAGGCGTTAAAGCGGAAGTGATGGTCGTC
[0123]	CGAAACACATCTACAGCATCTGGCGTAAATGCAGAAAAAGAACCTCGCC
[0124]	TTTGATGAGCTGTTTGATGTGCGTGCGGTACGTATTGTGCGCGAGCGTTTA
[0125]	CAGGATTGCTATGCCGCACTGGGGATAGTGCACTCACTATCGCCACCTG
[0126]	CCGGATGAGTTTGACGATTACGTCGTAACCCGAAACCAACCGTTATCA
[0127]	GTCTATTATACCGTGGTTCTGGGGCCGGTGAAAAACCGTTGAGATCCA
[0128]	AATCCGCACCAACAGATGCATGAAGATGCAGAGTTGGGTGTTGCTGCGC
[0129]	ACTGGAAATATAAGAGGGCGCGGCTGCTGGCGGCGCACGTTCCGGACAT
[0130]	GAAGACCGGATTGCCTGGCTGCGTAACTGATTGCTGGCAGGAAGAGAT
[0131]	GGCTGATTCCGGCGAAATGCTCGACGAAGTACGTAGTCAGGTCTTTGACG
[0132]	ACCGGGTGTACGTCTTTACGCCGAAAGGTGATGTCGTTGATTTCGCTGCGG
[0133]	GATCAACGCCGCTGGACTTCGCTTACCACATCCACAGTGATGTCGGACACC
[0134]	GCTGCATCGGGGCAAAAATTGGCGGGCGCATTGTGCCGTTACCTACCAG
[0135]	CTGCAGATGGGCGACCAAGATTGAAATTATCACCCAGAAACAGCCGAACCC
[0136]	CAGCCGTGACTGGTTAAACCCAAACCTCGGTTACGTCACAACCAAGCCGTG
[0137]	GGCGTTCGAAAATTCACGCCTGGTTCCGTAAACAGGACCGTGACAAAAAC
[0138]	ATTCTGGCTGGGCGGCAATCCTTGACGACGAGCTGGAACATCTGGGGAT
[0139]	CAGCCTGAAAGAAGCAGAAAAACATCTGCTGCCGCGTTACAACCTCAATG
[0140]	ATGTCGACGAGTTGCTGGCGGCGATTGGTGGCGGGGATATCCGTCTCAATC
[0141]	AGATGGTGAACCTTCCTGCAATCGCAATTTAATAAGCCGAGTGCCGAAGAG

[0142]	CAGGACGCCGCCGCTGAAGCAACTTCAGCAAAAAAGCTACACGCCGCA
[0143]	AAACCCGAGTAAAGATAACGGTCGCGTGGTAGTCGAAGGTGTTGGCAACC
[0144]	TGATGCACCACATCGCGCGCTGCTGCCAGCCGATTCTGGAGATGAGATTG
[0145]	TCGGCTTCATTACCCAGGGGCGCGGTATTTAGTACACCGCGCCGATTGCG
[0146]	AACAACCTGGCGGAACTGCGCTCCCATGCGCCAGAACGCATTGTTGACGCG
[0147]	GTATGGGGTGAGAGCTACTCCGCCGGATATTCGCTGGTGGTCCGCGTGGTA
[0148]	GCTAATGATCGTAGTGGGTGTTACGTGATATCACGACCATTCTCGCCAAC
[0149]	GAGAAGGTGAACGTGCTTGGCGTTGCCAGCCGTAGCGACACCAAAACAGCA
[0150]	ACTGGCGACCATCGACATGACCATTGAGATTTACAACCTGCAAGTGCTGG
[0151]	GGCGCGTGCTGGGTAAACTCAACCAGGTGCCGGATGTTATCGACGCGCGT
[0152]	CGGTTGCACGGGAGTTAG
[0153]	표4. 위치 548 (SEQ ID NO: 4)에서의 C->T 돌연변이를 포함하는 대장균 <i>relA</i> 코딩 서열. 돌연변이된 아미노산에 는 밑줄을 긋고 볼드체로 표시하였다.
[0154]	ATGGTTGCGGTAAGAAGTGCACATATCAATAAGGCTGGTGAATTTGATCC
[0155]	GGAAAAATGGATCGCAAGTCTGGGTATTACCAGCCAGAAGTCGTGTGAGT
[0156]	GCTTAGCCGAAACCTGGGCGTATTGTCTGCAACAGACGCAGGGGCATCCG
[0157]	GATGCCAGTCTGTTATTGTGGCGTGGTGTGAGATGGTGGAGATCCTCTCG
[0158]	ACATTAAGTATGGACATTGACACGCTGCGGGCGGCGCTGCTTTTCCCTCTG
[0159]	GCGGATGCCAACGTAGTCAGCGAAGATGTGCTGCGTGAGAGCGTCGGTAA
[0160]	GTCGGTCGTTAACCTTATTCACGGCGTGCGTGATATGGCGGCGATCCGCCA
[0161]	GCTGAAAGCGACGCACACTGATTCTGTTTCCCTCCGAACAGGTCGATAACGT
[0162]	TCGCCGGATGTTATTGGCGATGGTCGATGATTTTCGCTGCGTAGTCATCAA
[0163]	ACTGGCGGAGCGTATTGCTCATCTGCGCGAAGTAAAAGATGCGCCGGAAG
[0164]	ATGAACGTGTACTGGCGGCAAAAGAGGTACCAACATCTACG T ACCGCTG
[0165]	GCTAACCGTCTCGGAATCGGACAACCTGAAATGGGAACTGGAAGATTACTG
[0166]	CTTCCGTTACCTCCATCCAACCGAATACAAACGAATTGCCAAACTGCTGCA
[0167]	TGAACGGCGTCTCGACCGCAACACTACATCGAAGAGTTCGTTGGTCATCT
[0168]	GCGCGCTGAGATGAAAGCTGAAGGCGTTAAAGCGGAAGTGATGGTCGTC
[0169]	CGAAACACATCTACAGCATCTGGCGTAAAATGCAGAAAAAGAACCTCGCC
[0170]	TTTGATGAGCTGTTGATGTGCGTGCGGTACGTATTGTCGCCGAGCGTTTA
[0171]	CAGGATTGCTATGCCGCACTGGGGATAGTGCACACTCACTATCGCCACCTG
[0172]	CCGGATGAGTTTGACGATTACGTCGCTAACCCGAAACCAAACGGTTATCA
[0173]	GTCTATTCATACCGTGGTTCTGGGGCCGGGTGAAAAACCGTTGAGATCCA AATCCGCACCAACAGATGCATGAAGATGCAGAGTTGGGTGTTGCTGCGC ACTGGAATATAAAGAGGGCGCGGTGCTGGCGGCGCACGTTCCGGACAT GAAGACCGGATTGCCTGGCTGCGTAAACTGATTGCGTGCGAGGAAGAGAT
[0174]	GGCTGATTCCGGCGAAATGCTCGACGAAGTACGTAGTCAGGTCTTTGACG
[0175]	ACCGGTGTACGTCCTTACGCCGAAAGGTGATGTCGTTGATTGCTGCGG

- [0176] GATCAACGCCGCTGGACTTCGCTTACCACATCCACAGTGATGTCGGACACC
- [0177] GCTGCATCGGGGCAAAAATTGGCGGGCGCATTGTGCCGTTACCTACCAG
- [0178] CTGCAGATGGGCGACCAGATTGAAATTATCACCCAGAAACAGCCGAACCC
- [0179] CAGCCGTGACTGGTTAAACCCAAACCTCGGTTACGTACACAACCAGCCGTG
- [0180] GCGGTTGCAAAATTCACGCCTGGTTCCGTAAACAGGACCGTGACAAAAAC
- [0181] ATTCTGGCTGGGCGGCAAATCCTTGACGACGAGCTGGAACATCTGGGGAT
- [0182] CAGCCTGAAAGAAGCAGAAAAACATCTGCTGCCGCGTTACAACCTCAATG
- [0183] ATGTCGACGAGTTGCTGGCGGCGATTGGTGGCGGGGATATCCGTCTCAATC
- [0184] AGATGGTGAACTTCCTGCAATCGCAATTTAATAAGCCGAGTGCCGAAGAG
- [0185] CAGGACGCCCGCGCTGAAGCAACTTCAGCAAAAAAGCTACACGCCGCA
- [0186] AAACCGCAGTAAAGATAACGGTCGCGTGGTAGTCGAAGGTGTTGGCAACC
- [0187] TGATGCACCACATCGCGCGCTGCTGCCAGCCGATTCTGGAGATGAGATTG
- [0188] TCGGCTTCATTACCCAGGGGCGCGGTATTTAGTACACCGCGCGGATTGCG
- [0189] AACAACCTGGCGGAACCTGCGCTCCCATGCGCCAGAACGCATTGTTGACGCG
- [0190] GTATGGGGTGAGAGCTACTCCGCCGATATTCGCTGGTGGTCCGCGTGGTA
- [0191] GCTAATGATCGTAGTGGGTGTTACGTGATATCACGACATTCTCGCCAAC
- [0192] GAGAAGTGAAACGTGCTTGGCGTTGCCAGCCGTAGCGACACCAACAGCA
- [0193] ACTGGCGACCATCGACATGACCATTGAGATTTACAACCTGCAAGTGCTGG
- [0194] GGCGCGTGCTGGGTAAACTCAACCAGGTGCCGGATGTTATCGACGCGCGT
- [0195] CGGTTGCACGGGAGTTAG
- [0196] 일 실시예에 있어서, 본 발명은 분리된 DNA를 제공한다. 이것의 뉴클레오티드 서열이 SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 또는 SEQ ID NO: 4 또는 이들의 보체(complement)를 포함하거나 또는 이들 중 어느 하나로 구성된다.
- [0197] 또 하나의 실시예에 있어서, 본 발명은 SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 또는 SEQ ID NO: 4 또는 이들의 보체를 포함하거나 또는 이들 중 어느 하나로 구성되는 분리된 핵산을 제공한다.
- [0198] 또 하나의 실시예에 있어서, 본 발명은 SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 또는 SEQ ID NO: 4(또는 이들의 보체)의 뉴클레오티드 서열 또는 이들의 축퇴된 서열을 포함하거나 또는 이들 중 어느 하나로 구성되는 분리된 핵산을 제공한다.
- [0199] 또 하나의 실시예에 있어서, 본 발명은 발현 벡터, 예를 들어, 발현 조절 서열에 작동 가능하게 링크된 SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 또는 SEQ ID NO: 4의 핵산을 포함하는 플라스미드를 제공한다. 또 하나의 실시예에 있어서, 본 발명은 상기 벡터들 중 하나를 포함하는 배양 세포를 제공한다. 또 다른 실시예에 있어서, 본 발명은 상기 벡터들 중 하나, 또는 상기 세포의 자세포(progeny)로 형질감염된(transfected) 배양 세포를 제공한다.
- [0200] 또 다른 실시예에 있어서, 본 발명은 감소된 계능 박테리움, 예를 들면, *relA* 유전자를 함유하는 대장균을 제공하며, 여기서, *relA* 유전자는 적어도 하나의 점 돌연변이를 함유한다. 관련된 실시예에 있어서, 상기 돌연변이는 상기 *relA* 유전자의 위치 547 또는 548에서 발생한다.
- [0201] 또 다른 실시예에 있어서, 본 발명은 고유 모균주(parent strain)의 계능보다 적어도 5%, 적어도 8% 또는 적어도 10%가 더 작게 되도록 유전자 조작된 계능을 가지는 박테리움을 제공하며, 여기서 상기 박테리움은 적어도 하나의 *relA* 돌연변이를 포함한다. 일 실시예에 있어서, 상기 적어도 하나의 *relA* 돌연변이는 상기 *relA* 유전자의 위치 547 또는 548에서 발생한다. 또 하나의 실시예에 있어서, 상기 돌연변이는 상기 *relA* 유전자의 위치 547에

서의 G→A 점 돌연변이, 상기 *relA* 유전자의 위치 547에서의 G→T 점 돌연변이, 상기 *relA* 유전자의 위치 548에서의 C→G 점 돌연변이, 또는 상기 *relA* 유전자의 위치 548에서의 C→T 점 돌연변이 중 하나 이상으로부터 선택된다. 또 하나의 실시예에 있어서, 박테리움은 적은 규모의 수, 예를 들어, 상기 *relA* 유전자의 위치 540과 550 사이에서 발생하는 1 내지 약 10, 1 내지 약 8 또는 1 내지 약 5 *relA* 점 돌연변이들을 포함한다.

[0202] 또 하나의 실시예에 있어서, 본 발명은 본 명세서에서 서술되듯이 하나 이상의 감소된 계능 박테리움을 사용하여 아미노산을 생합성하는 방법을 제공하며, 상기 방법은: (1) 고유 *relA* 유전자를 포함하는 DNA를 가지는 박테리움을 제공하기; (2) *relA* 돌연변이 박테리움을 형성하기 위하여 상기 고유 *relA* 유전자를 돌연변이 *relA* 유전자로 대체하기; (3) 상기 돌연변이 *relA* 박테리움으로부터 아미노산을 발현하기를 포함한다. 일 실시예에 있어서, 아미노산은 Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Val, Leu, Lys, Met, Phe, Pro, Ser, Thr, Tyr 및 Trp로부터 선택된다.

[0203] 정의들

[0204] 본 명세서에서 사용되는 "염기쌍(Base pair)"은 예를 들어, 이중 나선 DNA 분자에 있어서, 아데닌(A)과 티민(T)의, 또는 시토신(C)과 구아닌(G)의 수소 결합된 뉴클레오티드들을 지칭한다. RNA에서는, 우라실(uracil : U)이 티민 대신에 사용된다. 염기쌍은 또한 DNA 길이 측정의 단위로서 사용될 수 있다. 삽입 서열(insert sequence)과 벡터와 관련하여 사용되는 "클론(Clone)"은 경우에 따라, 상동적(homologous), 위치특이적 또는 비정통적(illegitimate) 재조합에 의해 상기 삽입 서열을 상기 벡터 안으로 결합(ligation) 또는 그것의 도입을 의미할 수 있다. 삽입 서열, 벡터 및 숙주세포(host cell)와 관련하여 사용되는 경우, 상기 용어는 주어진 삽입 서열의 복제들을 만드는 것을 의미할 수 있다. 상기 용어는 또한 복제된 삽입 서열을 운반하는 숙주세포 또는 복제된 삽입 서열 자체를 지칭할 수도 있다.

[0205] 본 명세서에서 사용되는 "보체(Complement)," "상보적인(complementary)" 또는 "상보성(complementarity)"은 핵산 분자들의 뉴클레오티드 유사체들(analogs) 또는 뉴클레오티드들 사이에서의 왓슨-크릭(Watson-Crick) 또는 휴그스틴(Hoogsteen) 염기쌍 형성(base pairing)을 의미할 수 있다. 예를 들어, 서열 5'-A-G-T-3'은 서열 3'-T-C-A-5'에 대하여 상보적이다. 상보성(Complementarity)은 "부분적(partial)"일 수 있으며, 뉴클레오티드들 중 단지 일부만이 염기쌍 형성 규칙에 따라 매칭된다(matched). 또는, 상기 핵산들 사이에는 "완전한(complete)" 또는 "전체적인(total)" 상보성이 존재할 수 있다. 핵산 나선(strand)들 사이의 상보성의 정도는 핵산 나선들 사이의 혼성화(hybridization)의 강도 및 효율에 영향을 미칠 수 있다.

[0206] 핵산을 나타낼 때 본 명세서에서 사용되는 "인코딩(encoding) 또는 코딩(coding)"은 뉴클레오티드들의 서열을 의미할 수 있으며, 이것은, RNA로의 전사(transcription) 및 단백질로의 후속적인 번역(translation)과 동시에, 주어진 단백질, 펩타이드 또는 아미노산 서열의 합성으로 이끌 것이다. 그러한 전사와 번역은 인 비트로(*in vitro*) 또는 인 비보(*in vivo*)에서 실제로 일어날 수 있거나, 표준 유전 코드에 근거하여 완전히 이론적일 수 있다.

[0207] 본 명세서에서 사용되는 "효소(Enzyme)"는 다른 화합물들 내에서 화학적인 변화를 유도하기 위해 촉매제로서 작용함으로써, 하나 이상의 기질(substrate)들로부터 하나 이상의 생성물들을 만들어내는 단백질을 의미할 수 있다. 효소들은 2004년 3월 11일자로 국제 생화학 및 분자 생물학 연맹의 명명법 위원회(the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology)에 의해 추천된 바와 같이, 그들의 EC 번호에 의해서, 또는 표준 명명법을 사용하여 본 명세서에서 나타내어진다.

[0208] 본 명세서에서 사용되는 "발현조절 서열(Expression control sequence)"은 작동 가능하게 링크된 핵산의 전사를 지시하는 전사인자 결합 위치의 배열 또는 프로모터(promoter)를 의미할 수 있다.

[0209] 이중 나선을 가진 핵산에 관하여 사용되는 "자유단(Free end)"은 블런트 자유단들(blunt free ends) 또는 스틱키 자유단들(sticky free ends), 또는 이들의 조합을 가지는 선형 핵산을 의미할 수 있다.

[0210] 본 명세서에서 사용되는 "유전자(gene)"는 폴리펩타이드 또는 그것의 전구체(precursor)를 인코딩하는(encoding) 핵산 서열을 포함하는 핵산(예를 들어, DNA 또는 RNA)을 지칭한다. 상기 폴리펩타이드는 전-길이(full-length) 또는 단편(fragment)의 기능적인 성질들(예를 들어, 효소적인 활동, 리간드 결합(ligand binding), 신호 전달(transduction), 항원성(antigenicity) 등) 또는 원하는 활동이 유지되는 한, 전 길이의 코딩 서열에 의해서 또는 상기 코딩 서열의 어느 부분에 의해서 인코딩될(encoded) 수 있다. 상기 용어는 또한 전-길이 mRNA로 전사되는 상기 유전자에 기여하는 상기 5' 및 3' 단부 모두에서 코딩 영역에 인접한 상기 서열들을 포함한다. 용어 "유전자(gene)"는 cDNA와 유전자의 계능 형태들(genomic forms) 모두를 포함한다. 유전자

의 클론 또는 게놈 형태는 (예를 들면, 인트론(introns))이라 칭해지는 비코딩 서열들로 간접되는 코딩 영역을 함유할 수 있다.

[0211] 핵산을 균주(strain)에 추가하는 것과 관련하여 사용되는 경우, "도입하다(introduce)" 또는 "도입된(introduced)"은, 상기 핵산이 상기 균주의 염색체로 통합되어질 수 있거나 혹은 상기 균주 내의 플라스미드(plasmid)와 같은 벡터상에 포함될 수 있다는 것을 의미한다.

[0212] 본 명세서에서 사용되는 "라이브러리(Library)"는 삽입 서열을 각각 포함하는 복수의 벡터들 또는 복수의 뉴클레오티드 단편들을 지칭한다.

[0213] 본 명세서에서 사용되는 "돌연변이(mutant)" 또는 "돌연변이유발(mutagenesis)"은, 모균주(parent strain)의 핵산에 대한 어떤 변형을 의미할 수 있다. 핵산의 돌연변이유발은 삭제, 삽입, 치환, 재배치, 및 억제 유전자와 점 돌연변이들을 포함하나 이에 제한되지 않는 임의의 형태가 될 수 있다.

[0214] 본 명세서에서 사용되는 "핵산(Nucleic acid)"은 DNA 또는 RNA를 포함하나 이에 제한되지 않는 분자를 함유하는 핵산을 의미할 수 있다. 상기 용어는 4-아세틸시토신, 8-하이드록시-N6-메틸아데노신, 아지리딘시토신(aziridinylcytosine), 슈도이소시토신(pseudoisocytosine), 5- 카르복시하이드록실메틸)우라실, 5-플루오로우라실, 5-브로모우라실, 5-카르복시메틸아미노메틸-2-티오우라실, 5 카르복시메틸아미노메틸우라실, 디하이드로우라실, 이노신, N6-이소펜테닐아데닌, 1-메틸아데닌, 1-메틸슈도우라실, 1-메틸구아닌, 1-메틸이노신, 2,2-디메틸구아닌, 2-메틸아데닌, 2-메틸구아닌, 3-메틸시토신, 5-메틸시토신, N6-메틸아데닌, 7-메틸구아닌, 5-메틸아미노메틸우라실류, 5-메톡시아미노메틸-2-티오우라실, y-D-마니노실큐에오신, 5'- 메톡시카르보닐메틸우라실, 5-메톡시우라실, 2-메틸티오-N6-이소펜테닐아데닌, 우라실-5-옥시 아세트산 메틸에스테르, 우라실-5-옥시 아세트산, 옥시부톡소신, 슈도우라실, 큐에오신, 2-티오시토신, 5-메틸-2-티오우라실, 2-티오우라실, 4-티오우라실, 5-메틸우라실, N-우라실-5- 옥시 아세트산 메틸 에스테르, 우라실-5-옥시 아세트산, 슈도우라실, 큐에오신, 2-티오시토신, 및 2, 6-디아미노푸린을 포함하나, 이에 제한되지 않는 DNA와 RNA의 임의의 염기 유사체들을 함유하는 서열들을 포함한다.

[0215] 본 명세서에서 사용되는 "뉴클레오티드(Nucleotide)"는 오탄당 부분(pentose sugar moiety), 인산염 그룹(phosphate group)과 질소함유 헤테로 환식의 염기(nitrogenous heterocyclic base)로 구성되는 핵산(DNA 또는 RNA)의 단량체 단위(monomeric unit)를 지칭할 수 있다. 상기 염기는 글리코시딕(glycosidic) 탄소(상기 오탄(pentose)의 1' 탄소를 통해 상기 당(sugar) 부분에 연결될 수 있다. 상기 염기와 당의 조합은 뉴클레오사이드(nucleoside)로 칭해진다. 상기 뉴클레오사이드가 상기 오탄(pentose)의 3' 또는 5' 위치에 결합된 인산염 그룹을 포함하는 경우에, 그것은 뉴클레오티드라고 지칭될 수 있다. 작동 가능하게 링크된 뉴클레오티드들의 서열은 "염기 서열" 또는 "뉴클레오티드 서열" 또는 "핵산 서열"로 지칭될 수 있으며, 좌측에서 우측 방향이 5'-종점(terminus)에서 3'-종점의 전통적인 방향으로 있는 공식에 의하여 본 명세서에서 표현될 수 있다.

[0216] 본 명세서에서 사용되는 "작동 가능하게 링크된(Operably linked)"이라는 용어는 발현조절 서열과 다운스트림 폴리뉴클레오티드(downstream polynucleotide)를 지칭할 수 있으며, 상기 발현조절 서열에서 폴리뉴클레오티드의 생산적인 전사가 개시되도록 한다.

[0217] 본 명세서에서 사용되는 "초과 발현(Overexpressing)"은 유전자에 의해 인코딩되는 단백질의 전체적인 세포 활동이 증가되는 것을 의미할 수 있다. 단백질의 전체적인 세포 활동은 단백질의 증가된 세포의 양, 또는 상기 단백질의 증가된 반감기(half-life)에 기인한다. 단백질의 전체적인 세포 양은 상기 단백질을 코딩하는 상기 유전자의 증폭, 또는 강한 프로모터를 상기 단백질을 코딩하는 상기 유전자에 작동 가능하게 링크시키는 것을 포함하나, 이에 제한되지 않는 방법들에 의해 증가될 수 있다.

[0218] 본 명세서에서 사용되는 "단백질(protein)"은, 단편(fragment), 유도체들(derivatives), 동족체들(homologs), 변이체들(variants)과 그들의 융합체들(fusions) 뿐만 아니라, 고유 형태이든 재조합 형태이든, 펩타이드, 폴리펩타이드 및 단백질을 의미할 수 있다.

[0219] 게놈을 언급하는 경우에 본 명세서에서 사용되는 "비교영역(Region of comparison)"은, 1×10^7 , 1.5×10^7 , 2×10^7 , 2.5×10^7 , 3×10^7 , 3.5×10^6 , 4×10^7 또는 그 이상의 뉴클레오티드들 또는 염기쌍들이 될 수 있으며, 핵산 서열을 언급하는 경우에는, 50, 100, 250, 500, 10^3 , 5×10^3 , 10^4 , 5×10^4 , 10^5 , 5×10^5 , 10^6 또는 그 이상의 뉴클레오티드들 또는 그 이상의 염기쌍들이 될 수 있다.

[0220] 본 명세서에서 사용되는 "엄격한 혼성화 조건들(Stringent hybridization conditions)"은 제 1의 핵산 서열이,

핵산들의 복합 혼합물에서와 같이, 제 2의 핵산 서열에 특이적으로 또는 선택적으로 혼성화될 것이나, 다른 서열들에 대해서는 혼성화되지 않는 조건을 의미할 수 있다. 엄격한 조건들은 서열 의존적이며 다른 환경에서는 달라질 것이다. 일반적으로, 엄격한 조건들은 정의된 이온 강도 pH에서 특정 서열을 위한 열적 용점(T_m)보다 더 낮은 약 5-10°C가 되도록 선택된다. 상기 T_m 은 평형상태(상기 제 2의 서열이 T_m 에서 과도하게 존재함에 따라, 평형 상태에서 상기 제 1의 서열의 50 %가 사용되고 있는)에서 상기 제 1의 서열의 50 %가 상기 제 2의 서열에 혼성화하는(정의된 이온 강도, pH, 및 핵 농도하에서의) 온도일 수 있다. 엄격한 조건들은, 염의 농도가 약 1.0 M 나트륨 이온 미만이며, pH 7.0-8.3에서 전형적으로 약 0.01-1.0 M 나트륨 이온 농도(또는 다른 염들)이며, 상기 온도는 짧은 핵산 서열들(예를 들어, 약 10-50 뉴클레오티드들)에 대해서는 적어도 약 30°C 그리고, 긴 핵산 서열들(예를 들어, 약 50 뉴클레오티드들 보다 큰)에 대해서는 적어도 약 60°C가 되는 조건들일 수 있다. 엄격한 조건들은 또한 포름아미드(formamide)와 같은 불안정화제(destabilizing agents)의 첨가로 얻어질 수 있다. 선택적인 또는 특이적인 혼성화를 위해, 포지티브 신호가 백그라운드 혼성화(background hybridization)의 적어도 2 내지 10배일 수 있다. 예시적인 엄격한 혼성화 조건들은 다음을 포함한다: 65°C에서 0.2x SSC 와 0.1 % SDS에서 세척과 함께, 50 % 포름아미드, 5x SSC, 및 1 % SDS, 42°C에서의 배양, 또는 5x SSC, 1 % SDS, 65°C에서의 배양.

[0221] 본 명세서에서 사용되는 "실질적으로 상보적인(Substantially complementary)"은 제 1의 서열이 비교영역에 걸쳐서 제 2의 서열의 보체(complement)와 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98% 또는 99%가 동일하거나 또는 상기 두 개의 서열들이 엄격한 혼성화 조건들하에서 혼성화되는 것을 의미할 수 있다.

[0222] 본 명세서에서 사용되는 "실질적으로 동일한(Substantially identical)"은, 제 1의 서열과 제 2의 서열이 비교영역에 걸쳐서 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98% 또는 99%가 동일하거나 또는 실질적으로 상보적인 것을 의미할 수 있다. 참조 서열(reference sequence)과 시험 서열(test sequence)은 수작업으로 또는 컴퓨터 알고리즘(예를 들어, GAP, BESTFIT, FASTA 및 TFAST)에 의하여 정렬될 수 있으며, 동일성의 퍼센티지는 "동일한 잔기들(residues)의 전체 수"를 "참조 서열 내의 잔기들의 전체 수"로 나눈 다음 100만큼 곱함으로써 계산된다.

[0223] 감소된 게놈 균주

[0224] 자연환경에서의 박테리아는 표준 산업 또는 실험실 성장에서 보통 경험할 수 없는 많은 조건들에 노출되어 있고, 따라서 많은 수의 조건-의존적, 스트레스-유발된 유전자, 또는 그렇지 않으면 유기체들의 산업 또는 실험실 용도에서 필요하지 않을 수도 있는 불필요한 유전자를 운반한다. 박테리아 균주의 게놈 안에 함유된 많은 유전 정보는 산업적 또는 실험실적 중요 과정에서 박테리아 배양(cultures)의 사용에 해로운 영향 없이 삭제될 수 있다는 것이 인식되었다. 감소된 게놈을 가진 박테리움은 많은 산업 및 실험실 응용(application)에 있어서 고유 균주들보다 유리할 것이라는 것이 또한 인식되었다. 예를 들어, 감소된 게놈을 가진 박테리움은 적어도 다소 적은 대사적인 요구를 필요로 하며, 따라서 보다 효율적으로 원하는 생성물을 생성할 수 있다. 또한, 감소된 게놈은 더 적은 고유 생성물들과 더 낮은 레벨(level)의 특정 고유 단백질로 이어질 수 있어서 상기 남아있는 박테리아성 단백질로부터 원하는 단백질의 정제를 쉽게 한다. 게다가, 일부 박테리아성 유전자 서열들은 표준 산업적 또는 실험실 관행에 저촉될 수 있는 불안정성과 연관되고, 고비용의 부담스러운 품질 조절 절차를 수반할 것이다.

[0225] 상기 감소된 게놈 균주는 고유 모균주의 게놈보다 적어도 2%, 5%, 7% 에서 8%까지, 14%까지, 18%까지, 20%까지, 그리고 40%까지 60%까지 더 작은 게놈을 가질 수 있다. 일련의 삭제 이후 더 작아진 게놈의 퍼센티지는 "모든 삭제 이후 삭제된 염기쌍의 전체 수"를 "모든 삭제 이전의 모균주의 게놈 내의 염기쌍의 전체 수"로 나눈 다음 100을 곱해 계산된다.

[0226] 상기 감소된 게놈 균주는 삭제된 단백질 인코딩 유전자의 약 5% 내지 약 10%, 약 10% 내지 약 20%, 약 30% 내지 약 40%, 또는 약 60%를 가지는 게놈을 또한 가질 수 있다. 삭제된 단백질 인코딩 유전자를 가졌던 게놈에의 퍼센티지는 "모든 삭제 이후 삭제된 단백질 인코딩 유전자의 전체 수"를 "모든 삭제 이전의 모균주의 게놈 내의 단백질 인코딩 유전자의 전체 수"로 나눈 다음, 100을 곱해 계산된다.

[0227] 상기 게놈으로부터 삭제된 다른 핵산 서열들과 상기 유전자들은 특정 성장 조건들하에서 상기 감소된 게놈 균주의 증식 및 생존의 비율에 부정적인 영향을 주지 않을 것이다. 부정적 영향의 정도가 용인될 수 있을지는 특정 응용(application)에 의존한다. 예를 들어, 증식 비율에 있어서 30% 감소는 한 응용에 대해서는 용인되나 다른 것에는 안될 수도 있다. 또한, 상기 게놈으로부터 핵산 서열을 삭제한 부정적 영향은 배양 조건 변경과 같은 조치에 의하여 감소될 수 있다. 그러한 조치는 용인할 수 없는 부정적 영향을 용인할 수 있는 것으로 바꿀 수 있

다. 상기 증식 비율은 증가되든지 또는 대략 모균주와 같아질 수 있다. 상기 증식 비율은 또한 모균주의 그것보다 약 5%, 10%, 15%, 20%, 30%, 40%에서 약 50%까지 더 낮아질 수 있다. 상기 감소된 계놈 균주의 배가 시간은 약 30분에서 약 3시간 사이까지의 범위가 될 수 있다.

[0228] 모본(Parent)

[0229] 상기 모균주는 상기 감소된 계놈 균주가 유래된 중간체 균주 뿐만 아니라, 임의의 박테리아 균주나 다른 유기체가 될 수도 있다. 모균주들의 대표적 예는 K-12 또는 B 같은 대장균 균주이거나, 그와 실질적으로 동일한 계놈 서열을 가진 균주를 포함하나, 이에 제한되지 않는다. 상기 K-12 균주는 MG 1655가 될 수도 있다. 상기 모균주의 계놈의 뉴클레오티드 서열은 부분적으로 또는 완전히 알려질 수 있다. 대장균의 몇몇 균주들 및 다른 통상적으로 사용되는 실험실 미생물들의 완전한 계놈 서열은 알려져 있다.(예를 들어, Blattner *et al.*, Science, 277: 1453-74,1997; GenBank Accession No. U00096; Fema *et al.*, Nature, 409,529-533,2001; Hayashi *et al.*, DNA Res., 8,11-22,2001; Welch *et al.*, Proc. Natl. Acad. Sci, USA 99: 17020-17024,2002 및 GenBank Accession No. AE014075을 보라, 이들 각각은 참조로서 본 명세서에 포함되었다.)

[0230] 대장균 MG1655(주석(annotated) 버전 m56) (NCBI 접수 번호 U00096.1)의 핵산 서열은 4,639,675 뉴클레오티드들 또는 염기쌍들의 전체 크기를 가지고 있다. 대장균 MG1655의 계놈 서열의 최초 공개는 주석 버전 m54(4,639,221 뉴클레오티드들 또는 염기쌍들) 이었다. 대장균 MG 1655 계놈의 최초 공개에서 발생할 것으로 예측되었던 각 유전자는 소문자 "b"로 시작하는 고유한 숫자 식별기호가 배정되었다. 첫 번째 예측된 유전자는 b0001을 가진 *thrL*이었다; 마지막 유전자는 b4403를 가진 *lasT*이었다. 최초 공개 이후, 몇몇 신규의 유전자들이 발견되었는데, 이들은 예를 들어, yaa V와 같은 다양한 그룹들에 의해 "a" 또는 ".1" 접미사를 갖는 인증되지 않는 b-숫자가 부여되었다. 이들 유전자들의 일부는 가장 최근의 주석(버전 m56)에 받아들여졌고 b4404부터 시작하는 새로운 b-숫자들이 배정되었다. m54 버전의 몇몇 유전자들은 이후로 시퀀싱(sequencing) 오류들로 인해 따로 분리되거나 함께 병합되는 것으로 나타났다. 이들은 모두 b4460부터 b4500까지의 새로운 b-숫자들이 재발행되었고 오래된 b-숫자는 삭제되었다.

[0231] 계놈 삭제

[0232] 상기 감소된 계놈 균주는, 내용이 본 명세서에 참조로서 포함된, 국제 특허 출원 공개 번호 WO 2007/024756의 표 1에 제시된 하나 이상의 핵산 영역이 결핍되거나, 서열들이 거기에 실질적으로 동일할 수 있다. WO 2007/024756의 표 1은 대장균 MG 1655(주석 버전 m56)에 대한 계놈의 다음의 주석이 달린 특성들을 제공한다: 컬럼 1(타입)은 주석이 달린 특성(코딩 서열 (CDS), 프로파지, 복제 개시점(rep_origin), 반복 영역, rRNA, tRNA, 기타_RNA(misc_RNA) 또는 기타_특성(misc_feature))의 타입을 열거한다; 컬럼 2(S)는 참조 가닥(+; 전방으로, -: 상보적인)을 열거한다; 컬럼 3(좌측)은 상기 특성의 좌측 단부의 위치를 열거한다; 컬럼 4(우측)는 상기 특성의 우측 단부의 위치를 열거한다; 컬럼 5(이름)는 상기 특성의 이름을 열거한다; 그리고 컬럼 6(B)은 상기 특성의 블래트너 숫자 또는 특성의 b-숫자를 열거한다. 상기 감소된 계놈 균주는 MDS41, MDS42, MDS43, MDS44, MDS45, MDS46, MDS47, MDS48, MDS49, MDS50, MDS51, MDS52, MDS53, MDS54, MDS55, MDS56, MDS57, MDS58, MDS59, MDS60 또는 실질적으로 이들과 동일한 계놈을 가진 균주일 수 있다. 또한 상기 감소된 계놈 균주는 MDS42recA- 또는 실질적으로 이것과 동일한 계놈을 가진 균주일 수도 있다.

[0233] 삭제될 수 있는 핵산 서열의 한 형태는 상기 유기체 또는 그 유기체의 유전자 생성물의 안정성에 악영향을 줄 수 있는 것들을 포함한다. 불안정을 야기하는 그러한 인자들은 전이 인자들, 삽입 서열들, 및 계놈 불안정에 역할을 할 수 있는 다른 "이기적인 DNA(selfish DNA)" 인자들을 포함한다. 이들 인자들은 손상되지 않았거나 결함이 있더라도, 벡터와 같은 세포내 다른 핵산 서열들 또는 상기 계놈내의 한 포인트에서 다른 포인트로 이동할 수 있다. 이들 인자들은 박테리아 생존 및 배양 환경에서의 생장에 대하여 중요하지 않을 수 있다. 대장균 MG1655(주석 버전 m54)의 계놈 지도상의 IS 인자들의 위치는, 내용이 본 명세서에 참조로서 포함된 국제 특허 공개 번호 WO 2003/070880 및 미국 특허 공개 번호 제20030138937호의 표 1 및 도 1에 나타내어진다. 계놈 불안정성에 연관된 다른 핵산 서열들도 또한 삭제될 수 있다.

[0234] 삭제될 수 있는 핵산 서열의 또 하나의 형태는 제한 수정 시스템 유전자들 및 생성물들이 외래 DNA를 파괴할 수 있는 다른 내생적 뉴클레아제이다. 이들 유전자들은 박테리아 생존 및 배양 환경에서의 생장에 대하여 중요하지 않을 수 있다. 이들 유전자들은 박테리움 안으로 도입된 벡터들을 파괴함으로써 유전자 조작을 또한 방해할 수도 있다. 대장균 MG 1655(주석 버전 m54)의 계놈 지도상의 제한 수정 시스템 유전자들의 위치는 내용이 본 명세서에 참조로서 포함된 국제 특허 공개 번호 WO 2003/070880 및 미국 특허 공개 번호 제20030138937호의 표 1 및 도 1에 나타내어진다. 이중유래 유전자들을 포함하는, DNA 메틸라아제 유전자들은, 예를 들어, 진핵생물 메틸라

아제 유전자처럼 특정 용도를 위하여 상기 균주를 최적화하기 위해, 상기 삭제된 균주에 다시 추가될 수 있다.

- [0235] 삭제될 수 있는 핵산 서열의 또 하나의 형태는 플라젤라(flagella) 유전자 패밀리어며, 이것은 박테리아를 위해 운동성을 제공할 수 있다. 배양 환경에서, 박테리아 운동성은 세포 생존 및 성장을 위해 중요하지 않을 수 있으며, 유영동작(swimming action)은 세포 에너지의 1% 이상을 소비하면서 아무 도움이 되지 않으므로, 대사적으로 아주 고비용이 될 수 있다. 대장균 MG 1655(주석 버전 m54)의 게놈 지도상의 플라젤라 유전자들의 위치는 내용이 본 명세서에 참조로서 포함된 국제 특허 공개 번호 WO 2003/070880 및 미국 특허 공개 번호 제20030138937호의 표 1 및 도 1에 나타내어진다.
- [0236] 삭제될 수 있는 핵산 서열의 또 하나의 형태는 Rhs 인자들이다. Rhs 인자들은 3.7 Kb Rhs 핵(core)을 공유할 수 있는데, 상기 핵은 상동 재조합을 거쳐 게놈 재배열의 수단을 제공할 수 있는 큰 상동 반복 영역(대장균 K-12내에 다섯 개의 복제 부분이 있을 수 있다.)이 될 수 있다. Rhs 인자들은 일부 다른 배경에서 주로 진화되는 부속 인자들일 수 있고, 종으로서 대장균의 분기(divergence) 이후 수평 교환에 의해 대장균으로 퍼질 수 있다. 대장균 MG 1655(주석 버전 m54)의 게놈 지도상의 Rhs 인자들의 위치는 내용이 본 명세서에 참조로서 포함된 국제 특허 공개 번호 WO 2003/070880 및 미국 특허 공개 번호 제20030138937호의 표 1 및 도 1에 나타내어진다.
- [0237] 삭제될 수 있는 핵산 서열의 또 하나의 형태는 세포의 생존 및 증식에 덜 중요할 수 있는 비-전사 영역들이다. 삭제될 수 있는 핵산 서열의 또 하나의 형태는 주 제한 수정 유전자 패밀리를 인코딩할 수 있는 hsd 영역들이다. 대장균 MG 1655(주석 버전 m54)의 게놈 지도상의 hsd 영역들 및 비-전사 영역들의 위치는 내용이 본 명세서에 참조로서 포함된 국제 특허 공개 번호 WO 2003/070880 및 미국 특허 공개 번호 제20030138937호의 표 1 및 도 1에 나타내어진다.
- [0238] 삭제될 수 있는 핵산 서열의 다른 형태들은 프로파지류, 슈도유전자류(pseudogenes), 독소 유전자류, 병원성 유전자류(pathogenicity genes), 주변세포질(periplasmic) 단백질 유전자류, 막 단백질 유전자 및 [tonA](FhuA) 및/또는 용해성 파지(lytic phage) T1을 위하여 수용체(receptor)를 인코딩하는 그것의 완전한 오페론[fhu]ABC와 같은, 박테리오파지 수용체들을 포함한다.
- [0239] 삭제될 수 있는 핵산 서열의 다른 형태들은 하나의 박테리아성 균주의 게놈과 하나 이상의 다른 균주들을 비교함으로써 식별될 수 있다. 상기 균주들의 둘이나 셋에서 존재하지 않는 핵산 서열들은 기능적으로 본질적일 가능성이 적고, 따라서 삭제 후보가 될 수 있다. 대장균 K-12(앞에, Blattner, *et al.*을 보라)의 서열이 그와 가까운 관계인 0157:H7(앞에, Perna *et al.*을 보라)의 서열과 비교된 후, 단백질 인코딩 유전자들의 22%(K-12) 및 46%(0157:H7)가 상대적으로 일정한 백분율로 랜덤하게 삽입되는 1 내지 약 85 kb로부터의 균주-특이적인 아일랜드(islands)상에 위치될 수 있다는 것이 발견되었다. 내용이 본 명세서에 참조로서 포함된 국제 특허 공개 번호 WO 2003/070880 및 미국 특허 공개 번호 제20030138937호는 대장균 균주들 0157:H7 EDL933 및 K-12 MG1655의 게놈 서열들의 비교가 12개의 표적(target)의 삭제와 식별로 이어지고, 약 8% 더 작은 게놈을 가진 박테리아 균주를 결과적으로 가져온다는 것을 기술하고 있다. 상기 감소된 게놈을 가진 박테리아는 고유 모균주 MG 1655와 실질적으로 같은 비율로 성장하였다.
- [0240] 요로감염(uropathogenic) 대장균 균주 CFT073 H7(앞에, Welch *et al.*을 보라)의 DNA 서열은 최근에 결정되었고, 그 서열은 상기 K-12(MG 1655) 및 0157:H7와 비교되었다. 상기 게놈들의 어느 하나에서 발견된 모든 코딩 유전자들의 단지 약 40%만이 모든 게놈들에 존재하고, CFT073, K-12 및 0157:H7는 67%, 43% 및 68% 균주 특이적인 아일랜드 유전자들로 구성된다는 것을 결과는 보여준다. 이 정보에 근거하여, 단백질 코딩 서열들의 약 60% 정도가 대장균으로부터 삭제될 수 있다. 다른 균주들의 성장에는 필요하지 않으나 한 균주의 성장에는 필수적인 유전자들이 있을 수 있다는 것이 주목되어야 한다. 그러한 경우, 그 균주의 성장에 필수적인 유전자는 상기 균주로부터 삭제되지 않을 수도 있고, 또는 만일 삭제된다면, 상기 균주의 성장을 허용하도록 하기 위하여 보완 기능을 가진 다른 유전자로 교체될 수 있다.
- [0241] 삭제될 수 있는 핵산 서열들의 다른 형태들은 주변세포질 단백질들을 코딩하는 유전자들이다. 대장균과 같은 그람 음성 박테리아는, 두 개의 세포막, 내부 세포막 및 외부 세포막을 가지며, 이들은 주변세포질 공간(periplasmic space (PS))에 의해 분리되어 있다. 내용이 본 명세서에 참조로서 포함된 국제 특허 공개 번호 WO 2003/070880 에서 보여지는 것처럼, 9 개의 알려진 및 3 개의 추정상의 주변세포질 단백질 유전자들은, 최소한의 매체상에서 성장하는 상기 유기체의 능력에 상당히 영향을 미치지 않고, MDS40을 구축하면서 성공적으로 삭제되었다. 이들 돌연변이들은, 아미노산 흡수(uptake), 무기 대사, 세포막 유지, 당 대사 및 부착(adhesion)을 포함하는 기능들의 범위에 영향을 미친다. 주변세포질 단백질들의 제거는 주변세포질(periplasm)에 발현된 재조합 단백질들 내의 오염물질을 감소시킬 수 있다.

- [0242] 신호-펩타이드 서열들에 의해 식별된 알려진 또는 추정상의 막 단백질들을 위해 코딩하는 대략 85 유전자들이 삭제되었다. 이들 중 33은 플라젤라 구조 또는 생합성에 포함될 수 있다; 9는 항원(fimbrial) 구조 또는 생합성에 포함될 수 있다; 그리고 13은 일반적인 분비 경로들에 포함될 수 있다. 나머지는 상기 세포막들 내에서 다양한 알려진 또는 추정상의 기능들을 가질 수 있다. 이들 단백질의 다수는 주변세포질 공간에서 처리될 수 있다. 내용이 본 명세서에 참조로서 포함된 국제 특허 공개 번호 WO 2003/070880 에서 보여지는 것처럼, 그들은 최소한의 매체상에서 성장하는 상기 유기체의 능력에 상당히 영향을 미치지 않고 MDS40를 구축하면서 또한 삭제되었다.
- [0243] 주석이 달린 MG 1655 데이터베이스 내의 신호 펩타이드-유사 서열들을 검색함으로써 그리고 문헌과 이들을 상호 관련시킴으로써, 내재하는 주변세포질 단백질이 될 수 있는 181 단백질들이 식별되었다. 다수의 이들 단백질들이, 다음을 제외한, 기능에 따라 몇 가지 그룹으로 분류된다: 부착 및 이동성; 영양 및 염분 흡수(uptake), 미량 원소 흡수(uptake); 환경 감지; 방어 및 보호; 및 주변세포질 단백질 분비와 처리. 삭제될 수 있는 전체 오페론들 또는 상기 유전자들 중에는, 말하자면 생물 의약품 제조를 위하여, 한정된 최소한의 매체에 있어서 필요하지 않을 수도 있는, 당 및 아미노산 운반 단백질을 위하여 코딩하는 것이 있다.
- [0244] 실험자는, 상기 계놈으로부터 하나 또는 수 개의 유전자들 또는 다른 핵산 서열들을 제거한 결과를 시험할 수 있다. 예를 들어, 상기 계놈의 하나 또는 수 개의 유전자들 또는 다른 핵산 서열들이 삭제된 후에, 실험자는 결과로서 생기는 박테리아의 생존 및 증식 비율을 측정할 수 있다. 상기 식별된 유전자들 또는 다른 핵산 서열들의 대부분이 원하는 생성물을 생산하기 위하여 해로운 영향 없이 삭제될 수 있음에도 불구하고, 특정 유전자나 다른 핵산 서열의 삭제는, 세포의 사멸이나 증식 비율에 있어서 용인될 수 없는 레벨의 감소와 같은 용인할 수 없는 결과를 가져올 수 있다. 유전자 기능에 있어서 쓸모없는 중복 및 생물학적 경로들 사이의 상호 작용들 때문에, 이러한 가능성이 존재한다. 추가적인 삭제가 없이 균주 내에서 실행 가능한 몇몇 삭제들은 단지 다른 삭제들과 결합하여서만 해로울 것이다. 상기 가능성은 삭제 후보들을 식별하는데 사용되는 소정의 방법들 때문에 또한 존재한다. 예를 들어, 삭제 후보들을 식별하는데 사용되는 한 가지 방법은 두 대장균 균주들을 비교하고, 두 균주들에 나타나지 않는 다른 DNA 서열들이나 유전자들을 선택하는 것이다. 이들 유전자들 및 다른 DNA 서열들의 대부분이 기능적으로 본질적이지 않을 것 같은 반면, 그들의 일부는 특유한 균주를 위해 중요할 수 있다. 삭제 후보들을 식별하는데 사용하는 또 하나의 방법은 비전사 영역들 및 계놈 안정성에 중요할 수 있는 소정의 비전사 영역들이 존재하는 가능성을 식별하는 것이다.
- [0245] 하나 또는 몇 개의 유전자들 또는 시험될 다른 DNA 서열들을 삭제한 결과는 응용의 목적에 달려있다. 예를 들어, 많은 응용들에 대하여 참인, 높은 생성 효율이 주요 관심사인 경우에, 증식 비율 및 매체 소비 비율에서의 삭제 효과는 시험된 결과가 될 수 있다. 이 경우, 시험된 결과는 또한 생성 속도 양 및 특정 생성물의 세포당 수율로서 더 구체적일 수 있다. 고유 단백질을 제거할 때 오염이 주요 관심사이며, 더 적은 고유 단백질들 및 더 낮은 고유 단백질 레벨, 또는 특정 고유 단백질의 결핍은 시험된 결과가 될 수 있다.
- [0246] 유전자 또는 다른 DNA 서열 삭제의 결과를 테스트하는 것은, 상기 유전자 또는 상기 DNA 서열에 대해 별로 알려지지 않았을 때 중요할 수 있다. 이것은 감소된 계놈으로 박테리움을 만들에 있어서 삭제 후보들을 식별하기 위한 또 하나의 실행 가능한 방법이다. 이 방법은 다른 방법들에 의해 식별된 후보들이 삭제되고 추가적인 후보들이 탐색되고 있을 때에 특별히 유용하다.
- [0247] 유전자 또는 다른 DNA 서열 삭제의 결과가 일련의 조건들 하의 박테리아의 생존력에 영향을 미치는 경우, 특정 유전자 또는 다른 DNA 서열을 삭제하지 않는 대안의 하나는 상기 해로운 영향들을 완화할 수 있는 수단이 있는지 결정하는 것이다. 예를 들어, 지질 다당류(LPS) 유전자들을 삭제하는 것이, LPS 단백질들의 막 안팎 영역의 세포막에서의 결핍으로 야기되는 더 많은 다공성의 세포막들로 인한 열악한 생존 결과에 이른다면, 배양 조건들은 다공성을 더욱 많이 가지는(more porous) 세포막들을 수용하도록 변화하여 LPS 유전자들을 운반하는 박테리아뿐만 아니라 LPS 유전자들이 부족한 박테리아들도 생존할 수 있도록 할 수 있다.
- [0248] 삭제하는 방식
- [0249] 당업자들에게 알려진 임의의 몇 가지 계놈 핵산 삭제방법들을 사용하여 상기 핵산 서열들을 삭제함으로써 상기 균주들이 만들어질 수 있다. 삭제 위치에서 어떤 다른 돌연변이들을 생성하지 않으면서 뒤에 어떤 삽입된 핵산을 남기지 않고(스카리스 삭제) 상기 계놈으로부터 상기 핵산 서열들이 삭제될 수 있다. 박테리아성 계놈으로부터 몇 개의 순차적인 삭제가 이루어지고 있다면, 뒤에 어떤 삽입된 핵산 서열들을 남기지 않는 것이 중요할 수 있다. 그러한 삽입된 서열들이 뒤에 남겨진다면, 그 서열들은, 박테리아로부터 남아있는, 계놈의 특징이 없고 아마도 중요한 부분들을 삭제하거나 또는 불리한 효과들로 다른 예상하지 못한 계놈 재배열을

야기하는, 바람직하지 못한 재조합 사건들을 위한 후보 위치들이 될 수 있다.

- [0250] 박테리아의 게놈에 있어서 삭제하기 위한 대표적인 방법들이 미국 특허 공개 번호 제20030138937호와 국제 특허 공개 번호 WO 2003/070880, Posfai, G. et al., *J Bacteriol.* 179: 4426-4428 (1997), Muyrers, J.P.P. et al., *Nucl. Acids Res.* 27: 1555-1557 (1999), Datsenko, K.A. et al., *Proc. Natl Acad. Sci.* 97:6640-6649 (2000) 및 Posfai, G. et al., *Nucl. Acids Res.* 27: 4409-4415 (1999)에 기술되어 있으며, 이들 각각은 본 명세서에 참조로서 포함되어 있다. 상기 삭제 방법들은 선형 DNAs에 기초한 것들과 자살 플라스미드(suicide plasmids)들에 기초한 것들로 분류될 수 있다. J.P.P. et al., *Nucl. Acids Res.* 27: 1555-1557 (1999) 및 Datsenko, K.A. et al., *Proc. Natl. Acad. Sci.* 97:6640-6649 (2000)에서 개시된 방법들은 선형 DNA에 기초한 방법들이며, Posfai, G. et al., *J Bacteriol.* 179: 4426-4428 (1997) 및 Posfai, G. et al., *Nucl. Acids Res.* 27: 4409-4415 (1999)에서 개시된 방법들은 자살 플라스미드에 기초한 방법들이다.
- [0251] 방법들
- [0252] 상기 감소된 게놈 균주는, 재조합 단백질을, 핵산류, 치료제류, 대사 중간생성물 및 최종생성물류와 같은 원하는 생성물들의 생산을 위해 사용될 수 있다. 재조합 단백질의 대표적인 예들은 인슐린, 인터루킨류, 사이토카인류, 성장 호르몬류, 성장 인자들, 적혈구 생성촉진 인자, 콜로니 촉진 인자들, 인터페론, 항체들 및 항체 단편들을 포함하나 이에 제한되는 것은 아니다. 치료제류의 대표적인 예들은 백신 성분, 진단 제품, 또는 연구 시약을 포함하나 이에 제한되는 것은 아니다. 대사 중간생성물류 및 최종생성물류의 대표적인 예들은 아미노산류, 지방산류, 비타민류 등과, 박테리아 내에서 자연적으로 생성되지는 않지만, 대사 경로 설계 또는 다른 유전자 조작의 결과로서 생성되는 화학 화합물들을 포함하나 이에 제한되는 것은 아니다(예를 들어, 미국 특허 번호 제6,472,16호 및 제6,372,476호를 보라, 이들 모두는 본 명세서에 참조로서 포함되어 있다.).
- [0253] 재조합 단백질을 주변세포질 또는 세포질로 발현될 수 있다. 주변세포질에서 단백질들의 발현은 산업용으로 통상적으로 사용되며 Hanahan, [*J. Mol. Biol.*], 166:557-80, 1983; Hockney, [*Trends Biotechnol.*], 12:456-632, 1994; 및 Hamig et al., [*Trends Biotechnol.*], 16:54-60, 1998에서 검토되었으며, 이들 각각은 참조로서 본 명세서에 포함되어 있다. 주변세포질 공간으로 분비를 일으키는 신호 펩타이드에 재조합 단백질이 부착되는 융합 단백질을 발현함으로써, 재조합 단백질들이 주변세포질에서 생성될 수 있다. 특정한 신호 펩타이드에 의해 상기 신호 펩타이드가 분해될 수 있다.
- [0254] 주변 세포질 발현에 유용한 구조는, 원하는 단백질을 인코딩하는 제 2핵산 서열에 작동 가능하게 링크된 주변세포질 공간으로 단백질의 운송을 중재할 수 있는 신호 펩타이드를 인코딩하는 제 1핵산 서열을 포함할 수 있다. 상기 신호 서열은 발현되고 있는 단백질 그대로일 수 있다. 상기 주변세포질 공간으로 운송된 단백질은 생물학적으로 활동적일 수 있다. 재조합 구조의 발현은 유도성 프로모터 또는 숙주 균주내에서 구조적으로 발현되는 프로모터의 제어에 따라 가능하다. 유도성 프로모터의 사용은 포화시킬 수 있는 Sec 시스템을 이용할 때 유리할 수 있다. 예를 들어, 비대사성 갈락토스 유도체, IPTG에 의해 유도 가능한 [*lac*]-기반 프로모터/억제물질이 이용될 수 있다. 그러한 프로모터들은 주변세포질 발현을 최적화하는 Sec 시스템을 통하여 발현과 분비의 미세 조절을 가능하게 한다.
- [0255] 상기 재조합 단백질은 또한 상기 재조합 단백질의 적당한 폴딩(folding)을 제공할 수 있는 샤페론/이황화 결합(chaperones/disulfide-bond) 형성 효소들로 공동 발현될 수 있다. 재조합 단백질의 주변세포질 발현(periplasmic expression)에 유용한 핵산 서열들은 미국 특허 번호 제5,747,662호; 제5,578,464호; 제6,335,178호; 및 제6,022,952호; Thomas et al., *Mol-Micro [Mol Micro]*, (2001) 39 (1) 47-53; Weiner et al., [*Cell*], (1998) 93, 93-101; 및 *Current Protocols in Molecular Biology [Current Protocols in Molecular Biology]* (1994) 16.6.1-16.6.14(Copyright 2000 by John Wiley [et al.] and Sons)에 기재된 것들을 포함하나 이에 제한되지는 않으며, 그들 각각은 참조로서 본 명세서에 포함되어 있다.
- [0256] 상기 감소된 게놈 균주는 또한 핵산을 복제하거나 증폭하는데 사용될 수 있다. 상기 감소된 게놈 균주는 고품질 핵산들의 생성을 허용할 수 있는 깨끗하고 최소한의 유전적 배경을 제공할 수 있다. IS 인자들과 같은, 이기적 DNA 인자들이 부족한 감소된 게놈 균주는 또한 다른 독성 유전자들의 복제를 허용할 수 있다. 상기 감소된 게놈 균주는 게놈 라이브러리와 같은 라이브러리를 생성하는데 또한 이용될 수 있다.
- [0257] 상기 감소된 게놈 균주는, 백신으로 접종된 숙주 내에 면역 반응을 유도할 수 있는 항원들을 인코딩하는 이종유래의 유전자들(heterologous genes)을 도입함으로써 백신으로서 또한 사용될 수 있다. 감소된 게놈 백신들은, 숙주 내에서 바람직한 생리학적 반응(즉, 면역 반응)을 유도하는 것으로 알려진 DNA를 함유하는 DNA에 기초한

백신들일 수 있다.

[0258] 실시예들

[0259] 다량의 트레오닌을 생성할 수 있는 MDS42 유도체는 상기 감소된 계능 박테리아 안으로 많은 변형을 조작함으로써 생성되었다. 이들 변형은, 분해관련 트레오닌 탈수소효소(dehydrogenase)를 인코딩하는 유전자 *tdh*를 삭제하는 것; 및 재조합 타크 프로모터(tac promoter)에 의해 작동 가능하게 조절되는 트레오닌 생합성 유전자들 *thrABC*를 함유하는 플라스미드를 도입하는 것을 포함한다. 이 플라스미드상의 *thrA* 유전자는 호모세린 탈수소효소의 피드백 저항 형태를 인코딩한다. 이 균주에 의한 트레오닌의 생성은 *lacI* 유전자를 삭제하여 플라스미드가 매개하는(plasmid-borne) *thrABC* 유전자의 발현을 활성화하여 더욱 개선된다. 돌연변이 트레오닌과 호모세린 운반체(transporter)를 인코딩하는 최종 돌연변이 *rhtA23*은 매체에서 트레오닌의 분비를 증가시킨다. 결과로 나온 균주는 도 1에서 42-final로 칭해진다.

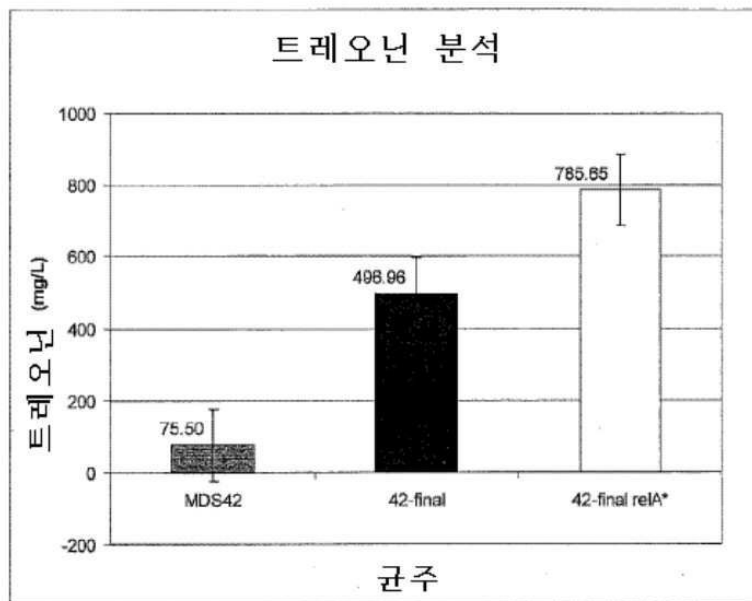
[0260] 상업적인 트레오닌 과잉 생산 균주들의 전사적 조절(transcriptional regulation)의 마이크로어레이 분석은 상기 세포들이 기아 상태로 영구적으로 고착되었음을 나타내었고, 이것은 긴축감응(stringent response)의 정상적인 조절이 이들 균주들에서는 상실되었다는 것을 암시한다. 이러한 데이터는, 상기 긴축감응을 조절하는 것에 연루된 단백질을 인코딩하는 유전자들 내에서의 돌연변이들을 본 발명자들이 찾도록 자극하였다. 2개의 유전자들이 초기에 조사를 위하여 표적으로 삼아졌다. 상기 *relA* 유전자는 전하를 띤 아미노 아실(amino acyl) tRNA's의 유용성을 감지함에 있어서 중심 역할을 하며, ppGpp를 생산하고, ppGpp는 차례로 RNA 중합효소를 방해하고, 그것에 의해 긴축 조절을 개시한다. 두 번째 유전자 *spoT*는 ppGpp를 퇴화(degrading)시키는데 연루되며, 그것에 의해 ppGpp의 세포내 레벨을 조절함으로써 상기 긴축감응을 또한 간접적으로 조절한다. *relA*의 특정한 돌연변이는 (G->A)547로 밝혀졌으며, 이것은Ala 에서 Thr의 돌연변이로 귀착된다. 비교적 미묘한 변화로 나타남에도 불구하고, 잔기는 틸 안으로 깊숙히 위치를 차지하여 상기 단백질의 구조적 모델의 주요한 구조적 특징을 나타낸다. 이 틸은 효소의 활성 영역을 나타내는 것 같다.

[0261] MDS42-final에 의해 트레오닌 생성을 향상시키는 이 돌연변이의 능력은 상기 돌연변이를 MDS42-final로 도입함으로써 조사되었다. 상기 유전자는 도 2에 개략적으로 도시된 방법에 의해 조작되었다. 간단하게, 돌연변이된 서열을 함유하는 돌연변이 *relA* 유전자의 영역을 함유하는 PCR 단편이 조립되어 상기 *relA* 돌연변이가 클로람페니콜 아세틸트랜스퍼라제(chloramphenicol acetyltransferase) 유전자에 연결되도록 하였다. 람다 레드-감(Lambda red-gam) 재조합 효소를 위하여 유전자들을 인코딩하는 플라스미드를 함유하는 MDS42-final의 내부로 상기 선형 PCR 단편이 일렉트로포레이트(electroporate) 되었다. 상기 선형 단편의 재조합은 염색체상의 *relA* 유전자를 목표로 하고 있으며, 클로람페니콜 저항성 *relA* 부분이매체(merodiploid)를 결과적으로 발생시킨다. 상기 람다 레드-감 플라스미드는 클로람페니콜 저항성 세포로부터 치료되었고 상기 세포들은 유도성 I-SceI 엔도뉴클레아제(endonuclease)를 함유하는 플라스미드와 함께 후속적으로 변형되었다. 엔도뉴클레아제의 발현을 유도하는 것은 클로람페니콜 저항성 세포에서 상기 *relA* 유전자 단편에 인접한 이중 나선 염색체 브레이크(break)로 귀착된다. 상기 세포가 생존할 수 있는 유일한 방법은 야생형(wild-type) 서열에서 (G->A)547 돌연변이 서열로 염색체상에 존재하는 *relA* 대립유전자(allele)를 변환하도록 이바지하는 상기 *relA* 유전자 서열들의 *recA* 중재된 재조합을 거치는 것이다. 재조합체들은 클로람페니콜에 대한 민감도에 의해 식별된다. 상기 I-SceI 함유 플라스미드는 그 다음으로 클로람페니콜 민감성 세포들로부터 치료되며, 상기 세포들은 상기 *thrABC* 발현 플라스미드와 함께 변형된다.

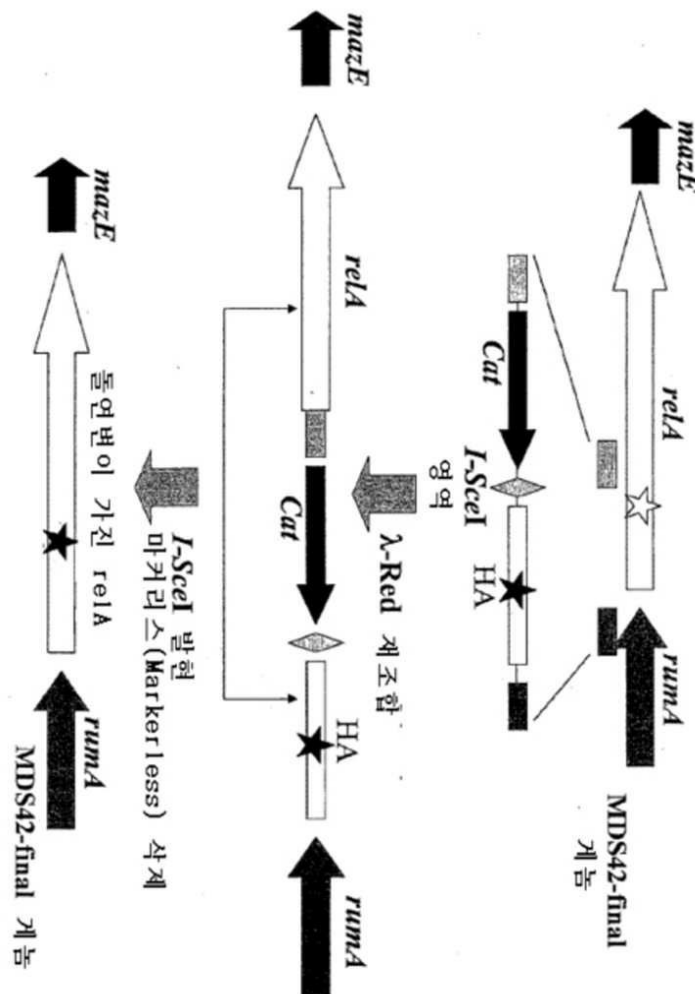
[0262] MDS42, MDS42-final 및 MDS42-final *relA**에 의한 트레오닌 생성은 최소한의 글루코스 매체에서 개개의 균주들을 정지상(stationary phase)으로 되게 하고, 원심분리법 및 여과법에 의해 세포들을 제거하고, 그리고 나서 필터가 소독된 사용된 매체를 새로운 최소한의 매체로 회석하고, 트레오닌 돌연변이 균주들의 생장을 지지하는 조합된 매체의 능력을 시험함으로써 분석되었다. 샘플들 내에서 트레오닌의 절대적 레벨은 상기 실험 샘플들의 지지된 생장과 화학적으로 합성된 트레오닌의 알려진 농도들로 생성된 표준 곡선을 비교함으로써 결정된다. 도 1에서 나타난 결과는, 상기 *relA** 돌연변이가 상기 MDS42-final 균주의 능력을 현저하게 증가시킨다는 것을 나타낸다. 이것은, *relA**와 유사한 *relA* 돌연변이가 높은 수준의 아미노산을 생성하도록 하기 위하여 대장균 균주들의 능력을 개선한다는 가설을 뒷받침한다.

도면

도면1



도면2



서열 목록

SEQUENCE LISTING

- <110> Scarab Genomics, LLC
- Blattner, Frederick R.
- Lee, Jun Hyong R.
- Kim, Sun Chang
- <120> Compositions and Methods for Amino Acid Biosynthesis
- <130> 02730.0025.00PC00
- <150> US 61/030,835
- <151> 2008-02-22
- <160> 4
- <170> PatentIn version 3.3
- <210> 1

<211> 2235

<212> DNA

<213> Escherichia coli

<400> 1

atggttgcgg taagaagtgc acatatcaat aaggctggtg aatttgatcc ggaaaaatgg	60
atcgcaagtc tgggtattac cagccagaag tcgtgtgagt gcttagccga aacctgggcg	120
tattgtctgc aacagacgca ggggcatccg gatgccagtc tgttattgtg gcgtggtgtt	180
gagatggtgg agatcctctc gacattaagt atggacattg acacgctgcg ggccggcgctg	240
cttttcctc tggcggatgc caacgtagtc agcgaagatg tgctgcgtga gagcgtcggg	300
aagtcggtcg ttaaccttat tcacggcgtg cgtgatatgg cggcgatccg ccagctgaaa	360
gcgacgcaca ctgattctgt ttctccgaa caggctcgata acgttcgccg gatgttattg	420
gcgatggtcg atgattttcg ctgcgtagtc atcaaaactgg cggagcgtat tgctcatctg	480
cgcgaagtaa aagatgcgcc ggaagatgaa cgtgtactgg cggcaaaaga gtgtaccaac	540
atctacacac cgctggctaa ccgtctcgga atcggacaac tgaaatggga actggaagat	600
tactgcttcc gttacctcca tccaaccgaa tacaacgaa ttgccaaact gctgcatgaa	660
cggcgtctcg acccggaaca ctacatcgaa gatttcgttg gtcactcgcg cgctgagatg	720
aaagctgaag gcgttaaagc ggaagtgtat ggtcgtccga aacacatcta cagcatctgg	780
cgtaaaatgc agaaaaagaa cctcgccctt gatgagctgt ttgatgtgcg tgcggtacgt	840
attgtcgccg agcgtttaca ggattgctat gccgcactgg ggatagtga cactcactat	900
cgccacctgc cggatgagtt tgacgattac gtcgctaacc cgaaacaaa cggttatcag	960
tctattcata ccgtggttct ggggccgggt ggaaaaaccg ttgagatcca aatccgcacc	1020
aaacagatgc atgaagatgc agagtgggt gttgctgcgc actggaaata taaagagggc	1080
gcggctgctg gcggcgacg ttccgggacat gaagaccgga ttgcctggct gcgtaaactg	1140
attgcgtggc aggaagagat ggctgattcc ggcgaaatgc tcgacgaagt acgtagtcag	1200
gtctttgacg accgggtgta cgtctttacg ccgaaagggtg atgtcgttga tttgcctgcg	1260
ggatcaacgc cgtggactt cgcttaccac atccacagtg atgtcggaca ccgtgcac	1320
ggggcaaaaa ttggcgggcg cattgtgccg ttcacctacc agctgcagat gggcgaccag	1380
attgaaatta tcaccagaa acagccgaac ccagccgtg actggttaaa cccaaacctc	1440
ggttacgtca caaccagccg tgggcgttcg aaaattcacg cctggttcg taaacaggac	1500
cgtgacaaaa acattctggc tgggcggcaa atccttgacg acgagctgga acatctgggg	1560
atcagcctga aagaagcaga aaaacatctg ctgccgcgtt acaacttcaa tgatgtcgac	1620

gagttgctgg cggcgattgg tggcggggat atccgtctca atcagatggt gaatttcctg 1680
 caatcgcaat ttaataagcc gagtgccgaa gagcaggacg cggccgcgct gaagcaactt 1740
 cagcaaaaaa gctacacgcc gcaaaaccgc agtaaagata acggtcgcgt ggtagtcgaa 1800

ggtgttggca acctgatgca ccacatcgcg cgtgctgcc agccgattcc tggagatgag 1860
 attgtcggct tcattaccca ggggcgcggt atttcagtac accgcgccga ttgcgaacaa 1920
 ctggcggaaac tgcgctccca tgcgccagaa cgcattgttg acgcggtatg gggtagagagc 1980
 tactccgccg gatattcgct ggtggtccgc gtggtagcta atgatcgtag tgggttgta 2040
 cgtgatatca cgaccattct cgccaacgag aaggtgaacg tgcttggcgt tgccagccgt 2100
 agcgacacca aacagcaact ggcgaccatc gacatgacca ttgagattta caacctgcaa 2160
 gtgctggggc gcgtgctggg taaactcaac cagggtgccg atgttatcga cgcgcgtcgg 2220

ttgcacggga gttag 2235

<210> 2
 <211> 2235
 <212> DNA
 <213> Escherichia coli
 <400> 2

atggttgccg taagaagtgc acatatcaat aaggctggtg aatttgatcc ggaataatgg 60
 atcgcaagtc tgggtattac cagccagaag tcgtgtgagt gcttagccga aacctgggcg 120
 tattgtctgc aacagacgca ggggcacccg gatgccagtc tgttattgtg gcgtggtgtt 180
 gagatggtgg agatcctctc gacattaagt atggacattg acacgctgcg ggccggcgctg 240
 cttttccctc tggcggatgc caacgtatgc agcgaagatg tgctgcgtga gagcgtcgg 300

aagtcggtcg ttaaccttat tcacggcgtg cgtgatatgg cggcgatccg ccagctgaaa 360
 gcgacgcaca ctgattctgt ttcctccgaa caggctcgata acgttcgccg gatgttattg 420
 gcgatggtcg atgattttcg ctgcgtatgc atcaaactgg cggagcgtat tgctcatctg 480
 cgcaagtaa aagatgcgcc ggaagatgaa cgtgtactgg cggcaaaaga gtgtaccaac 540
 atctactcac cgctggctaa ccgtctcgga atcggaacac tgaaatggga actggaagat 600
 tactgttcc gttacctcca tccaaccgaa tacaacgaa ttgccaaact gctgcatgaa 660
 cggcgtctcg accgcaaca ctacatcgaa gagttcgttg gtcactctgc cgctgagatg 720

aaagctgaag gcgttaaagc ggaagtgtat ggtcgtccga aacacatcta cagcatctgg 780
 cgtaaaatgc agaaaaagaa cctgccttt gatgagctgt ttgatgtgcg tgcggtacgt 840
 attgtcgccg agcgtttaca ggattgctat gccgcactgg ggatagtgca cactcactat 900

cgccacctgc cggatgagtt tgacgattac gtcgctaacc cgaacacaaa cggttatcag 960
tctattcata ccgtgggttct ggggccgggt ggaaaaaccg ttgagatcca aatccgcacc 1020
aaacagatgc atgaagatgc agagtgggt gttgctgcgc actggaaata taaagagggc 1080
gcggctgctg gcggcgacg ttcgggacat gaagaccgga ttgcctggct gcgtaaacgt 1140

attgctggc aggaagagat ggctgattcc ggcgaaatgc tcgacgaagt acgtagtcag 1200
gtctttgacg accgggtgta cgtctttacg ccgaaagggt atgtcgttga ttgctgctg 1260
ggatcaacgc cgctggactt cgcttaccac atccacagtg atgtcggaca ccgctgcatc 1320
ggggcaaaaa ttggcggcg cattgtgccg ttcacctacc agctgcagat gggcgaccag 1380
attgaaatta tcaccagaa acagccgaac ccagccgtg actggttaaa cccaaacctc 1440
ggttacgtca caaccagccg tggcggttcg aaaattcacg cctggttccg taaacaggac 1500
cgtgacaaaa acattctggc tggcgggcaa atccttgacg acgagctgga acatctgggg 1560

atcagcctga aagaagcaga aaaacatctg ctgccgcgtt acaacttcaa tgatgtcgac 1620
gagttgctgg cggcgattgg tggcggggat atccgtctca atcagatggt gaacttcctg 1680
caatcgcaat ttaataagcc gagtgccgaa gagcaggacg ccgccgcgt gaagcaactt 1740
cagcaaaaaa gtiacacgcc gcaaaaccgc agtaaagata acggtcgcgt ggtagtcgaa 1800
ggtgttggca acctgatgca ccacatcgcg cgtgctgcc agccgattcc tggagatgag 1860
attgtcggct tcattacca gggcgccggt atttcagtac accgcgccga ttgcgaacaa 1920
ctggcggaac tgcgtccca tgcgccagaa cgcattgttg acgcggtatg gggtgagagc 1980

tactccgccg gatattcgct ggtggtccgc gtggtagcta atgatcgtag tgggttgta 2040
cgtgatatca cgaccattct cgccaacgag aagggtgaac tgcttggcgt tgccagccgt 2100
agcgacacca aacagcaact ggcgaccatc gacatgacca ttgagattta caacctgcaa 2160
gtgctggggc gcgtgctggg taaactcaac cagggtgccg atgttatcga cgcgcgtcgg 2220
ttgcacggga gttag 2235

<210> 3
<211> 2235
<212> DNA
<213> Escherichia coli
<400> 3

atggttgcgg taagaagtgc acatatcaat aaggctggtg aatttgatcc ggaaaaatgg 60

atcgcaagtc tgggtattac cagccagaag tcgtgtgagt gcttagccga aacctgggcg 120
tattgtctgc aacagacgca ggggcatccg gatgccagtc tgttattgtg gcgtggtgtt 180

gagatggtgg agatcctctc gacattaagt atggacattg acacgctgcg ggcggcgctg 240
 cttttccctc tggcggatgc caacgtagtc agcgaagatg tgctgctga gagcgtcgg 300
 aagtcggtcg ttaaccttat tcacggcgtg cgtgatatgg cggcgatccg ccagctgaaa 360
 gcgacgcaca ctgattctgt ttccctccgaa caggctgata acgttcgccg gatgttattg 420
 gcgatggtcg atgattttcg ctgcgtagtc atcaaaactgg cggagcgtat tgctcatctg 480

 cgcaagtaa aagatgcgcc ggaagatgaa cgtgtactgg cggcaaaaga gtgtaccaac 540
 atctacggac cgctggctaa ccgtctcgga atcggacaac tgaatggga actggaagat 600
 tactgtcttc gttacctcca tccaaccgaa tacaacgaa ttgccaaact gctgcatgaa 660
 cggcgctctg accgcgaaca ctacatcgaa gagttcgttg gtcactgctg cgctgagatg 720
 aaagctgaag gcgttaaagc ggaagtgtat ggtcgtccga aacacatcta cagcatctgg 780
 cgtaaatgc agaaaaagaa cctcgccctt gatgagctgt ttgatgtgcg tgcggtacgt 840
 attgtcgccg agcgtttaca ggattgctat gccgcactgg ggatagtgca cactcactat 900

 gccacactgc cggatgagtt tgacgattac gtcgctaacc cgaaaccaa cggttatcag 960
 tctattcata ccgtggttct ggggcccggg gaaaaaccg ttgagatcca aatccgcacc 1020
 aaacagatgc atgaagatgc agagtgggt gttgctgctg actggaaata taaagagggc 1080
 gcggctgctg gcggcgacg ttccgggacat gaagaccgga ttgcctggct gcgtaaactg 1140
 attgctggc aggaagagat ggctgattcc ggcgaaatgc tcgacgaagt acgtagtcag 1200
 gtctttgacg accgggtgta cgtctttacg ccgaaagggt atgtcgttga tttgcctgcg 1260
 ggatcaacgc cgtggactt cgttaccac atccacagt atgtcgga ccgctgcac 1320

 ggggcaaaaa ttggcggcg cattgtgccg ttacactacc agctgcagat gggcgaccag 1380
 attgaaatta tcaccagaa acagccgaac ccagccgtg actggttaaa cccaaacctc 1440
 ggttacgtca caaccagccg tggcggttcg aaaattcacg cctggttcg taaacaggac 1500
 cgtgacaaaa acattctggc tggcgggcaa atccttgacg acgagctgga acatctgggg 1560
 atcagcctga aagaagcaga aaaacatctg ctgcccggtt acaacttcaa tgatgtcgac 1620
 gagttgctgg cggcgattgg tggcggggat atccgtctca atcagatggt gaacttcctg 1680
 caatcgcaat ttaataagcc gagtgccgaa gagcaggacg ccgccgcgt gaagcaactt 1740

 cagcaaaaaa gctacacgcc gcaaaaccgc agtaaagata acggtcgcgt ggtagtcgaa 1800
 ggtgttgga acctgatga ccacatcgcg cgctgctgcc agccgattcc tggagatgag 1860
 attgtcggct tcattacca gggcgcggtt atttcagtac accgcgccga ttgcgaacaa 1920
 ctggcggaac tgcgctccca tgcgccagaa cgcattgttg acgcggtatg gggtagagac 1980
 tactccgccg gatattcgct ggtggtccgc gtggtagcta atgatcgtag tgggttgta 2040

cgtgatatca cgaccattct cgccaacgag aaggtgaacg tgcttggcgt tgccagccgt 2100
 agcgacacca aacagcaact ggcgaccatc gacatgacca ttgagattta caacctgcaa 2160

 gtgctggggc gcgtgctggg taaactcaac cagggtgccg atgttatcga cgcgcgtcgg 2220
 ttgcacggga gttag 2235
 <210> 4
 <211> 2235
 <212> DNA
 <213> Escherichia coli
 <400> 4
 atggttgcgg taagaagtc acatatcaat aaggctggtg aatttgatcc ggaaaaatgg 60
 atcgcaagtc tgggtattac cagccagaag tcgtgtgagt gcttagccga aacctgggcg 120
 tattgtctgc aacagacgca ggggcatccg gatgccagtc tgttattgtg gcgtggtgtt 180
 gagatggtgg agatcctctc gacattaagt atggacattg acacgtgcg gccggcgctg 240

 cttttccctc tggcggatgc caacgtagtc agcgaagatg tgctgcgtga gagcgtcgg 300
 aagtcggtcg ttaaccttat tcacggcgtg cgtgatatgg cggcgatccg ccagctgaaa 360
 gcgacgcaca ctgattctgt ttctccgaa caggctgata acgttcgccg gatgttattg 420
 gcgatggtcg atgattttcg ctgcgtagtc atcaaactgg cggagcgtat tgctcatctg 480
 cgcgaagtaa aagatgcgcc ggaagatgaa cgtgtactgg cggcaaaaga gtgtaccaac 540
 atctacgtac cgtggctaa ccgtctcgga atcggacaac tgaaatggga actggaagat 600
 tactgtctcc gttacctcca tccaaccgaa taaaacgaa ttgccaaact gctgcatgaa 660

 cggcgtctcg accgcgaaca ctacatcgaa gagttcgttg gtcactctcg cgtgagatg 720
 aaagctgaag gcgttaaagc ggaagtgtat ggtcgtccga aacacatcta cagcatctgg 780
 cgtaaaatgc agaaaaagaa cctcgccctt gatgagctgt ttgatgtgcg tgcggtacgt 840
 attgtcgccg agcgtttaca ggattgctat gccgcactgg ggatagtga cactcactat 900
 cgccacctgc cggatgagtt tgacgattac gtcgctaacc cgaaacaaa cggttatcag 960
 tctattcata ccgtggttct ggggccgggt ggaaaaaccg ttgagatcca aatccgcacc 1020
 aaacagatgc atgaagatgc agagtgggt gttgctgcgc actggaaata taaagagggc 1080

 gcggctgctg gcggcgcacg ttcgggacat gaagaccgga ttgcctggct gcgtaaactg 1140
 attgctggc aggaagagat ggctgattcc ggcgaaatgc tcgacgaagt acgtagtcag 1200
 gtctttgacg accgggtgta cgtctttacg ccgaaagggt atgtcgttga tttgcctgcg 1260
 ggatcaacgc cgttgactt cgcttaccac atccacagt atgtcggaca ccgtgcatc 1320

ggggcaaaaa ttggcgggcg cattgtgccg ttcacctacc agctgcagat gggcgaccag	1380
attgaaatta tcaccagaa acagccgaac ccagccgtg actggttaaa cccaaacctc	1440
ggttacgtca caaccagccg tgggcgttcg aaaattcacg cctggttccg taaacaggac	1500
cgtgacaaaa acattctggc tgggcggcaa atccttgacg acgagctgga acatctgggg	1560
atcagcctga aagaagcaga aaaacatctg ctgccgcgtt acaacttcaa tgatgtcgac	1620
gagttgctgg cggcgattgg tggcggggat atccgtctca atcagatggt gaacttcctg	1680
caatcgcaat ttaataagcc gagtgccgaa gaggaggacg ccgccgcgct gaagcaactt	1740
cagcaaaaaa gctacacgcc gcaaaaccgc agtaaagata acggtcgcgt ggtagtcgaa	1800
ggtgttggca acctgatgca ccacatcgcg cgtgctgcc agccgattcc tggagatgag	1860
attgtcggct tcattacca ggggcgcggt atttcagtac accgcgccga ttgcgaacaa	1920
ctggcggaac tgcgtccca tgcgccagaa cgcattgttg acgcggtatg gggtagagac	1980
tactcgcgcg gatattcgct ggtggtccgc gtggtagcta atgatcgtag tgggttgta	2040
cgtgatatca cgaccattct cgccaacgag aaggtgaacg tgcttggcgt tgccagccgt	2100
agcgacacca aacagcaact ggcgaccatc gacatgacca ttgagattta caacctgcaa	2160
gtgctggggc gcgtgctggg taaactcaac caggtgccgg atgttatcga cgcgcgtcgg	2220
ttgcacggga gttag	2235