



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2020년09월24일  
(11) 등록번호 10-2159691  
(24) 등록일자 2020년09월18일

- (51) 국제특허분류(Int. Cl.)  
C12N 15/81 (2006.01) C12N 9/10 (2006.01)  
C12N 9/16 (2006.01) C12N 9/88 (2006.01)  
C12P 5/00 (2006.01) C12P 7/54 (2006.01)
- (52) CPC특허분류  
C12N 15/81 (2013.01)  
C12N 9/1029 (2013.01)
- (21) 출원번호 10-2015-7028637
- (22) 출원일자(국제) 2014년03월14일  
심사청구일자 2019년02월19일
- (85) 번역문제출일자 2015년10월12일
- (65) 공개번호 10-2015-0132310
- (43) 공개일자 2015년11월25일
- (86) 국제출원번호 PCT/US2014/028421
- (87) 국제공개번호 WO 2014/144135  
국제공개일자 2014년09월18일
- (30) 우선권주장  
61/800,356 2013년03월15일 미국(US)
- (56) 선행기술조사문헌  
US20120276587 A1  
EP02546336 A1

- (73) 특허권자  
아미리스 인코퍼레이티드  
미국, 캘리포니아주 94608, 에머리빌, 스위트  
100, 5885 홀리스 스트리트  
토탈 마케팅 서비스  
프랑스 에프-92800 뷔도 푸르 미셀레 24
- (72) 발명자  
호킨스 크리스티 미셸  
미국 94608 캘리포니아주 에머리빌 스위트 100 홀  
리스 스트리트 5885  
마하트데쿨-미도우즈 티나 티파완  
미국 94608 캘리포니아주 에머리빌 스위트 100 홀  
리스 스트리트 5885  
(뒷면에 계속)
- (74) 대리인  
유미특허법인

전체 청구항 수 : 총 35 항

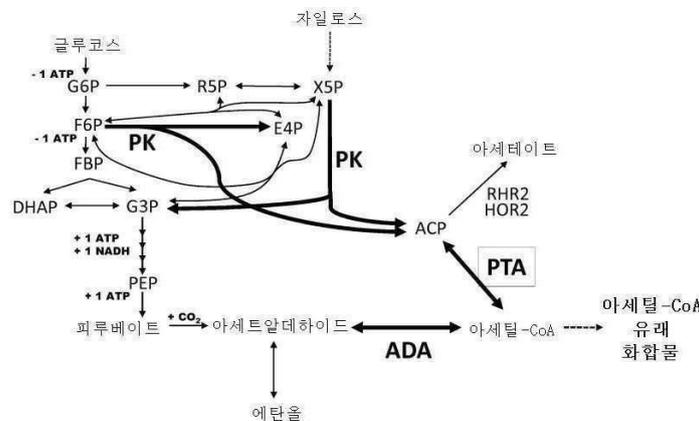
심사관 : 최준호

(54) 발명의 명칭 아세틸-코엔자임 A 유래 화합물을 생산하기 위한 포스포케톨라제 및 포스포트랜스아세틸라제의 용도

(57) 요약

본 발명은 숙주 세포에서 아세틸-CoA 또는 아세틸-CoA 유래 화합물의 생산을 개선시키는 조성물 및 방법을 제공한다. 일부 구현예들에서, 숙주 세포는 포스포케톨라제 (PK)를 코딩하는 이종의 뉴클레오티드 서열과 아세틸 포스페이트를 아세테이트로 변환하는 내인성 효소의 기능적 파괴를 포함하도록 유전자 변형된다. 일부 구현예들에 (뒷면에 계속)

대표도



서, 숙주 세포는 포스포트랜스아세틸라제 (PTA)를 코딩하는 이중의 뉴클레오티드 서열을 더 포함한다. 일부 구현예들에서, 아세틸 포스페이트를 아세테이트로 변환하는 효소는 글리세롤-1-포스파타제이다. 일부 구현예들에서, 글리세롤-1-포스파타제는 GPP1/RHR2이다. 일부 구현예들에서, 글리세롤-1-포스파타제는 GPP2/HOR2이다. 본원에 기술된 조성물과 방법은 비제한적으로 이소프레노이드, 폴리케티드 및 지방산 등의 아세틸-CoA-유래 화합물을 이중 생산하기 위한 유효한 경로를 제공해준다.

(52) CPC특허분류

- C12N 9/16* (2013.01)
- C12N 9/88* (2013.01)
- C12P 5/007* (2013.01)
- C12P 7/54* (2013.01)
- C12Y 203/01008* (2013.01)
- C12Y 301/03021* (2013.01)
- C12Y 401/02009* (2013.01)
- Y02P 20/52* (2015.11)

(72) 발명자

**미도우즈 애덤 레온**

미국 94608 캘리포니아주 에머리빌 스위트 100 홀리스 스트리트 5885

**피켄스 로렌 바바라**

미국 94608 캘리포니아주 에머리빌 스위트 100 홀리스 스트리트 5885

**타이 애나**

미국 94608 캘리포니아주 에머리빌 스위트 100 홀리스 스트리트 5885

**썬 애니 에닝**

미국 94608 캘리포니아주 에머리빌 스위트 100 홀리스 스트리트 5885

**명세서**

**청구범위**

**청구항 1**

- (a) 포스포케탈라제 (PK; EC 4.1.2.9)를 코딩하는 이중의 핵산;
- (b) 포스포트랜스아세틸라제 (PTA; EC 2.3.1.8)를 코딩하는 이중의 핵산; 및
- (c) 아세틸 포스페이트를 아세테이트로 변환하는 내인성 글리세롤-1-포스파타제 (EC 3.1.3.21)의 기능적 파괴를 포함하는,  
유전자 변형된 효모 숙주 세포.

**청구항 2**

제1항에 있어서,  
상기 글리세롤-1-포스파타제가 GPP1/RHR2, GPP2/HOR2, 및 이의 상동체로 이루어진 군으로부터 선택되는, 유전자 변형된 효모 숙주 세포.

**청구항 3**

제2항에 있어서,  
상기 GPP1/RHR2 또는 이의 상동체가 기능적으로 파괴된, 유전자 변형된 효모 숙주 세포.

**청구항 4**

제2항에 있어서,  
상기 GPP2/HOR2 또는 이의 상동체가 기능적으로 파괴된, 유전자 변형된 효모 숙주 세포.

**청구항 5**

제2항에 있어서,  
상기 GPP1/RHR2 및 GPP2/HOR2 둘다, 또는 상기 GPP1/RHR2의 상동체 및 상기 GPP2/HOR2의 상동체 모두가 기능적으로 파괴된, 유전자 변형된 효모 숙주 세포.

**청구항 6**

제1항에 있어서,  
상기 유전자 변형된 숙주 세포가 이소프레노이드를 생산할 수 있는, 유전자 변형된 효모 숙주 세포.

**청구항 7**

제6항에 있어서,  
상기 유전자 변형된 숙주 세포가 이소펜테닐 피로포스페이트를 제조하기 위한 메발로네이트 (MEV) 경로의 하나 이상의 효소를 코딩하는 하나 이상의 이중의 핵산을 포함하는, 유전자 변형된 효모 숙주 세포.

**청구항 8**

제7항에 있어서,  
상기 MEV 경로의 하나 이상의 효소가 NADH-이용성 HMG-CoA 리덕타제를 포함하는, 유전자 변형된 효모 숙주 세포.

**청구항 9**

제7항에 있어서,

상기 MEV 경로의 하나 이상의 효소가 아세틸-CoA 2 분자를 축합하여 아세토아세틸-CoA를 형성하는 효소를 포함하는, 유전자 변형된 효모 숙주 세포.

**청구항 10**

제7항에 있어서,

상기 MEV 경로의 하나 이상의 효소가 아세토아세틸-CoA를 아세틸-CoA와 축합하여 HMG-CoA 형성하는 효소를 포함하는, 유전자 변형된 효모 숙주 세포.

**청구항 11**

제7항에 있어서,

상기 MEV 경로의 하나 이상의 효소가 HMG-CoA를 메발로네이트로 변환하는 효소를 포함하는, 유전자 변형된 효모 숙주 세포.

**청구항 12**

제7항에 있어서,

상기 MEV 경로의 하나 이상의 효소가 메발로네이트를 메발로네이트 5-포스페이트로 인산화하는 효소를 포함하는, 유전자 변형된 효모 숙주 세포.

**청구항 13**

제7항에 있어서,

상기 MEV 경로의 하나 이상의 효소가 메발로네이트 5-포스페이트를 메발로네이트 5-피로포스페이트로 변환하는 효소를 포함하는, 유전자 변형된 효모 숙주 세포.

**청구항 14**

제7항에 있어서,

상기 MEV 경로의 하나 이상의 효소가 메발로네이트 5-피로포스페이트를 이소펜테닐 피로포스페이트로 변환하는 효소를 포함하는, 유전자 변형된 효모 숙주 세포.

**청구항 15**

제7항에 있어서,

상기 MEV 경로의 하나 이상의 효소가 HMG-CoA 신타제, 메발로네이트 키나제, 포스포메발로네이트 키나제 및 메발로네이트 피로포스페이트 데카르복실라제로부터 선택되는, 유전자 변형된 효모 숙주 세포.

**청구항 16**

제7항에 있어서,

상기 숙주 세포가 MEV 경로의 전체 효소를 코딩하는 복수의 이중의 핵산을 포함하는, 유전자 변형된 효모 숙주 세포.

**청구항 17**

제7항에 있어서,

상기 MEV 경로의 하나 이상의 효소를 코딩하는 하나 이상의 이중의 핵산이 하나의 전사 조절인자의 통제 하에 있는, 유전자 변형된 효모 숙주 세포.

**청구항 18**

제7항에 있어서,

상기 MEV 경로의 하나 이상의 효소를 코딩하는 하나 이상의 이중의 핵산이 복수의 이중의 전사 조절인자들의 통제 하에 있는, 유전자 변형된 효모 숙주 세포.

**청구항 19**

제7항에 있어서,

이소펜테닐 피로포스페이트 (IPP)를 다이메틸알릴 피로포스페이트 (DMAPP)로 변환할 수 있는 효소를 코딩하는 이중의 핵산을 더 포함하는, 유전자 변형된 효모 숙주 세포.

**청구항 20**

제7항에 있어서,

IPP 및/또는 DMAPP 분자를 축합하여 폴리프레닐 화합물을 형성할 수 있는 효소를 코딩하는 이중의 핵산을 더 포함하는, 유전자 변형된 효모 숙주 세포.

**청구항 21**

제7항에 있어서,

IPP 또는 폴리프레닐을 변형시켜 이소프레노이드 화합물을 형성할 수 있는 효소를 코딩하는 이중의 핵산을 더 포함하는, 유전자 변형된 효모 숙주 세포.

**청구항 22**

제21항에 있어서,

상기 IPP 또는 폴리프레닐을 변형하여 이소프레노이드 화합물을 형성할 수 있는 효소가 카렌 신타제, 게라니올 신타제, 리날롤 신타제, 리모넨 신타제, 미르센 신타제, 옥시멘 신타제, α-피넨 신타제, β-피넨 신타제, γ-테르피넨 신타제, 테르피놀렌 신타제, 아모르파다이엔 신타제, α-파르네센 신타제, β-파르네센 신타제, 파르네솔 신타제, 네롤리돌 신타제, 패출리올 신타제, 누트카톤 신타제 및 아비에타다이엔 신타제로 이루어진 군으로부터 선택되는, 유전자 변형된 효모 숙주 세포.

**청구항 23**

제6항에 있어서,

상기 이소프레노이드가 헤미테르펜, 모노테르펜, 다이테르펜, 트리테르펜, 테트라테르펜, 세스퀴테르펜 및 폴리테르펜으로 이루어진 군으로부터 선택되는, 유전자 변형된 효모 숙주 세포.

**청구항 24**

제6항에 있어서,

상기 이소프레노이드가 C<sub>5</sub>-C<sub>20</sub> 이소프레노이드인, 유전자 변형된 효모 숙주 세포.

**청구항 25**

제6항에 있어서,

상기 이소프레노이드가 아비에타다이엔, 아모르파다이엔, 카렌, α-파르네센, β-파르네센, 파르네솔, 게라니올, 게라닐게라니올, 이소프렌, 리날롤, 리모넨, 미르센, 네롤리돌, 옥시멘, 패출롤, β-피넨, 사비넨, γ-테르피넨, 테르피놀렌 및 발렌센으로 이루어진 군으로부터 선택되는, 유전자 변형된 효모 숙주 세포.

**청구항 26**

제1항에 있어서,

상기 효모가 사카로마이세스 세레비지애 (*Saccharomyces cerevisiae*)인, 유전자 변형된 효모 숙주 세포.

**청구항 27**

제1항 내지 제26항 중 어느 한 항에 있어서,

상기 유전자 변형된 숙주 세포가, 아세틸 포스페이트를 아세테이트로 변환하는 내인성 효소의 기능적 파괴를 포함하지 않는 효모 세포와 비교해, 아세틸-CoA 유래 화합물을 증가된 양으로 생산하는, 유전자 변형된 효모 숙주 세포.

**청구항 28**

이소프레노이드의 제조 방법으로서,

(a) 제6항 내지 제26항 중 어느 한항에 따른 유전자 변형된 숙주 세포의 개체군을, 이소프레노이드 화합물 생산에 적합한 조건 하에, 탄소원이 첨가된 배지에서 배양하는 단계; 및

(b) 상기 배지로부터 이소프레노이드 화합물을 회수하는 단계를 포함하는, 방법.

**청구항 29**

효모 숙주 세포에서 아세틸-CoA 또는 아세틸-CoA 유래 화합물의 생산을 증가시키는 방법으로서,

(a) 숙주 세포에서 포스포케톨라제 (PK; EC 4.1.2.9)를 코딩하는 이종의 핵산 및 포스포트랜스아세틸라제 (PTA; EC 2.3.1.8)를 코딩하는 이종의 핵산을 발현시키는 단계; 및

(b) 아세틸 포스페이트를 아세테이트로 변환하는 내인성 글리세롤-1-포스파타제 (EC 3.1.3.21) 효소를 기능적으로 파괴하는 단계를 포함하는, 방법.

**청구항 30**

제29항에 있어서,

상기 글리세롤-1-포스파타제가 GPP1/RHR2, GPP2/HOR2, 및 이의 상동체로부터 선택되는, 방법.

**청구항 31**

제30항에 있어서,

상기 GPP1/RHR2 또는 이의 상동체가 기능적으로 파괴되는, 방법.

**청구항 32**

제30항에 있어서,

상기 GPP2/HOR2 또는 이의 상동체가 기능적으로 파괴되는, 방법.

**청구항 33**

제30항에 있어서,

상기 GPP1/RHR2 및 GPP2/HOR2 둘다, 또는 상기 GPP1/RHR2의 상동체 및 상기 GPP2/HOR2의 상동체 모두가 기능적으로 파괴되는, 방법.

**청구항 34**

제29항에 있어서,

상기 효모가 사카로마이세스 세레비지에 (*Saccharomyces cerevisiae*)인, 방법.

**청구항 35**

제29항 내지 제34항 중 어느 한 항에 있어서,

상기 숙주 세포가, 아세틸 포스페이트를 아세테이트로 변환하는 내인성 효소의 기능적 파괴를 포함하지 않는 효모 세포와 비교해, 아세틸-CoA 또는 아세틸-CoA 유래 화합물을 증가된 양으로 생산하는, 방법.

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

### 발명의 설명

#### 기술 분야

[0001] 본 출원은 2013년 3월 15일자 미국 가출원 61/800,356에 대한 우선권을 주장하며, 이 문헌의 전체 내용은 원용에 의해 본 명세서에 포함된다.

[0002] 본 발명은 조작된 숙주 세포에서 아세틸-CoA 유래 화합물을 생산하기 위한 조성물 및 방법에 관한 것이다.

#### 배경 기술

[0003] 아세틸 코엔자임 A (아세틸-CoA)는 폴리케티드, 지방산, 이소프레노이드, 페놀류, 알칼로이드, 비타민 및 아미노산 등의 필수적인 생물학적 화합물의 합성에 있어, 중요한 중간체이다. 특히 아세틸-CoA로부터 유래되는 대사산물들은 공업적으로 유용한 화합물을 포함하는 일차 및 이차 대사산물이다. 효모에서, 아세틸-CoA는 피루베이트 대사로부터 생합성된다 (도 1). 그러나, 이 생합성 경로에서, 피루베이트 카르복실라제 및/또는 피루베이

트 데하이드로게나제에 의해 촉매되는 반응들을 통해 CO<sub>2</sub>가 소실된다. 공업용 발효 조건에서, 피루베이트 대사 및 저급 당분해에 대한 대안을 제공하는 한가지 이점은, 피루베이트의 탈카르복시화 반응에서 CO<sub>2</sub>가 적게 생산되며, 따라서 탄소가 더 많이 최종 생성물에 포집됨으로써, 이론적인 최대 수율이 높아진다는 것이다. 두번째 이점은, NADH가 더 적게 만들어짐에 따라서 이를 재산화하기 위한 산소의 필요량이 훨씬 적어진다는 것이다. 이는 포스포아세틸트랜스퍼라제 (PTA; EC 2.3.1.8)와 더불어 포스포케톨라제 (PK; EC 4.1.2.9)를 발현시킴으로써, 달성할 수 있다.

[0004] PK와 PTA는 반응을 촉매하여, 프럭토스-6-포스페이트 (F6P) 또는 자일로스-5-포스페이트 (X5P)를 아세틸-CoA로 변환시킨다. 도 1에 도시된 바와 같이, PK는 펜토스 포스페이트 중간산물인 자일로스 5-포스페이트로부터 또는 상류 당분해 중간산물인 D-프럭토스 6-포스페이트 (F6P)로부터 나온다. PK는 X5P를 글리세르알데하이드 3-포스페이트 (G3P)와 아세틸 포스페이트로, 또는 F6P를 에리트로스 4-포스페이트 (E4P)와 아세틸 포스페이트로 분해한다. 그런 후, PTA가 아세틸 포스페이트를 아세틸-CoA로 변환한다. G3P는 저급 당분해 경로에 재유입될 수 있으며, E4P는 트랜스알돌라제 및 트랜스케톨라제의 비-산화적 펜토스 포스페이트 경로 네트워크를 통한 순환에 의해 펜토스 포스페이트 경로 또는 당분해로 재유입될 수 있다.

[0005] 본 출원인은 PK 및 PTA 효소의 도입으로 달성할 수 있는, 이종의 (heterologous) 이소프레노이드 생산의 효능 개선에 대해 언급한 바 있다. 2012년 11월 9일자 미국 출원번호 13/673,819 (미국 특허번호 제 8,415,136)를 참조하며, 이의 전체 내용이 원용에 의해 본 명세서에 포함된다. 특히, 세포질 아세틸-CoA가 천연 효모 대사 네트워크에서 이루어지는 화학 반응만을 이용해 글루코스로부터 합성되는 경우, 메발로네이트 경로를 통한 글루코스의 이소프레노이드 파르네센으로의 변환에 대해 가능한 최대 화학량론적 수율은 23.6 wt%이다. 아세트알데하이드 데하이드로게나제, 아세틸화 (ADA; EC 1.2.1.10) 및 NADH-이용성 HMG-CoA 리덕타제에 의해 촉매되는 반응이 메발로네이트 생산용 대사 네트워크에 통합되면, 최대 화학량론적 이론적 수율은 25.2 wt%까지 개선된다. PK 및 PTA가 추가로 도입되면, 반응 네트워크는 바람직하게는 중량 수율이 29.8 wt% 이상으로, 최대 이론 수율의 현저한 증가가 이루어질 수 있다.

[0006] Sondregger 등도 역시 자일로스-이용 효소 균주에서 에탄올을 생산하는데 있어 PK와 PTA의 이점을 언급한 바 있다. Sondregger *et al.*, *Applied and Environmental Microbiology* 70(5):2892-2897 (2004)을 참조하며, 이 문헌은 전체 내용이 원용에 의해 본 명세서에 포함된다. 피키아 스티피티스 (*Pichia stipitis*)로부터 유래된 NAD(P)H-의존형 자일로스 리덕타제 및 NAD<sup>+</sup>-의존형 자일리톨 데하이드로게나제의 과발현으로 인한 에탄올의 낮은 수율을 해결하기 위해, 이종의 포스포케톨라제 경로 (PK, PTA 및 ADA)를 사카로마이세스 세레비지에에 도입하였다. 2가지 옥시도리덕타제 반응에서 조인자 (cofactor) 선호성이 상이하어, 환원된 반응 중간산물인 자일리톨의 축적량이 많아지고 에탄올 수율이 낮아지게 되는, 혐기성 레독스 밸런스 문제 (anaerobic redox balancing problem)가 유발되었다. 레독스 대사는 포스포케톨라제 경로를 도입함으로써 밸런스 문제가 해소되었는데, 포스포케톨라제 경로는 에탄올로 변환되는 자일로스 1개 당 NADH 총 1개를 재산화하며, 에탄올 수율이 25%까지 향상시킨다. 그러나, PK의 과발현 역시 아세테이트 축적 증가 및 발효율 감소를 유발한다. 어느 정도의 아세테이트 축적은 아세트알데하이드를 아세테이트로 변환하는 ALD6의 돌연변이와 포스포케톨라제 경로를 조합함으로써 낮출 수 있었지만, 제조할 포스포케톨라제 경로를 통한 흐름 (flux)은 자일리톨과 글리세롤 축적을 완전히 없애는데 필요한 최적 흐름의 약 30%이었다. 저자들은, 포스포케톨라제 경로에 의한 아세테이트 형성을 방지하기 위해서는 활성이 더 높은 포스포트랜스아세틸라제 및/또는 아세트알데하이드 데하이드로게나제가 필수적일 수 있다는 것을, 제안하였다.

[0007] 따라서, 이종의 PK 경로를 도입하여 아세틸-CoA 유래 화합물의 수율에 실질적인 향상을 유도할 수 있기는 하지만, PK 및 PTA를 통한 최적의 탄소 흐름을 달성하기 위해 상기한 경로에 대한 추가적인 개선이 필요할 것으로 보인다. 본원의 조성물 및 방법은 이러한 필요성을 해결하며, 관련 이점을 물론 제공한다.

**발명의 내용**

[0008] 본 발명은 상업적으로 유용한 화합물을 제조하기 위해 포스포케톨라제 (PK) 및 포스포트랜스아세틸라제 (PTA)의 활용성을 향상시키는 조성물 및 방법을 제공한다. 이들 조성물과 방법은, 포스포케톨라제 경로-기반의 아세테이트 축적이 PK 촉매 반응의 생성물인 아세틸 포스페이트의 효소-촉매화된 가수분해로부터 유발된다는 놀라운 발견 사실을 토대로 한다. 아세틸 포스페이트의 가수분해는, 탄소 고갈을 통해, 이소프레노이드, 폴리케티드 및 지방산 등의 아세틸-CoA로부터 유래되는 임의 타입의 생성물의 생산에 부정적으로 작용할 수 있는, 부적절한 부반응이다. 숙주 세포에서 아세틸 포스페이트 가수분해를 촉매하는 천연 효소의 기능적인 파괴를 통해, 아세

테이트 축적을 감소시키고, PK/PTA 경로를 통한 아세틸-CoA 생산으로의 탄소 흐름을 증가시킨다.

- [0009] 본원에 제공되는 조성물 및 방법은, 효소에서 아세틸 포스페이트를 아세테이트로 가수분해하는 천연 효소, 즉 GPP1/RHR2 및 이의 관련성 높은 상동체 GPP2/HOR2의 예상치 못한 발견을, 또한 토대로 한다. 이들 효소는 모두 기존에 글리세롤-1-포스파타제 (EC 3.1.3.21; "글리세롤-3-포스파타제"라고도 함) 활성을 가진 것으로 특정되었을 뿐, 이들 효소의 다른 아세틸-포스파타제 활성은 예상하지 못하였다. PK와 PTA를 이중적으로 발현하는 세포에서, RHR2 및 HOR2를 코딩하는 유전자들 중 하나 또는 둘다를 제거하면 아세테이트 축적이 감소되며, RHR2 코딩 유전자만 제거하면 아세테이트 수준이 실질적으로 감소하게 된다. 아울러, PK, PTA 및 메발로네이트 경로를 포함하도록 조작된 세포에서 *RHR2* 유전자를 제거하면, 아세틸-CoA 유래 이소프레노이드인 파르네센의 생산이 실질적으로 증가하였다.
- [0010] 따라서, 본 발명은 상업적으로 유용한 화합물을 제조하기 위한 유전자 변형된 숙주 세포 및 이의 이용 방법을 제공한다. 일 측면에서, 본 발명은 포스포케톨라제 (PK; EC 4.1.2.9)를 코딩하는 이중의 핵산; 및 아세틸 포스페이트를 아세테이트로 변환하는 내인성 효소의 기능적 파괴를 포함하는 유전자 변형된 숙주 세포를 제공한다. 일부 구현예들에서, 유전자 변형된 숙주 세포는 포스포트랜스아세틸라제 (PTA; EC 2.3.1.8)를 코딩하는 이중의 핵산을 더 포함한다.
- [0011] 다른 측면에서, 본 발명은 포스포트랜스아세틸라제 (PTA; EC 2.3.1.8)를 코딩하는 이중의 핵산; 및 아세틸 포스페이트를 아세테이트로 변환하는 내인성 효소의 기능적 파괴를 포함하는 유전자 변형된 숙주 세포를 제공한다. 일부 구현예들에서, 유전자 변형된 숙주 세포는 포스포케톨라제 (PK; EC 4.1.2.9)를 코딩하는 이중의 핵산을 더 포함한다.
- [0012] 일부 구현예들에서, 아세틸 포스페이트를 아세테이트로 변환하는 효소는 글리세롤-1-포스파타제 (EC 3.1.3.21)이다. 일부 구현예들에서, 글리세롤-1-포스파타제는 GPP1/RHR2, GPP2/HOR2 및 이들의 상동체 및 변이체로 이루어진 군으로부터 선택된다. 일부 구현예들에서, 유전자 변형된 숙주 세포는 GPP1/RHR2의 기능성 파괴를 포함한다. 일부 구현예들에서, 유전자 변형된 숙주 세포는 GPP2/HOR2의 기능성 파괴를 포함한다. 일부 구현예들에서, 유전자 변형된 숙주 세포는 GPP1/RHR2 및 GPP2/HOR2 모두의 기능성 파괴를 포함한다.
- [0013] 일부 구현예들에서, 유전자 변형된 숙주 세포는 아실화 아세틸알데하이드 데하이드로게나제 (ADA; EC 1.2.1.10)를 코딩하는 이중의 핵산을 더 포함한다. 일부 구현예들에서, 유전자 변형된 숙주 세포는 천연적인 피루베이트 데하이드로게나제 (PDH)-바이패스의 하나 이상의 효소의 기능성 파괴를 더 포함한다. 일부 구현예들에서, PDH-바이패스의 하나 이상의 효소는 아세틸-CoA 신태타제 1 (ACS1), 아세틸-CoA 신태타제 2 (ACS2) 및 알데하이드 데하이드로게나제 6 (ALD6)로부터 선택된다.
- [0014] 일부 구현예들에서, 유전자 변형된 숙주 세포는 이중의 아세틸-CoA 유래 화합물을 생산할 수 있다. 일부 구현예들에서, 이중의 아세틸-CoA 유래 화합물은 이소프레노이드, 폴리케티드 및 지방산으로 이루어진 군으로부터 선택된다. 특정 구현예에서, 유전자 변형된 숙주 세포는 이소프레노이드를 생산할 수 있다.
- [0015] 일부 구현예들에서, 유전자 변형된 숙주 세포는 이소펜테닐 피로포스페이트를 제조하기 위한 메발로네이트 (MEV) 경로의 하나 이상의 효소를 코딩하는 하나 이상의 이중의 핵산을 포함한다. 일부 구현예들에서, MEV 경로의 하나 이상의 효소는 NADH-이용성 HMG-CoA 리덕타제를 포함한다. 일부 구현예들에서, MEV 경로의 하나 이상의 효소는 아세틸-CoA 2 분자를 축합하여 아세토아세틸-CoA를 형성하는 효소를 포함한다. 일부 구현예들에서, MEV 경로의 하나 이상의 효소는 아세토아세틸-CoA를 아세틸-CoA와 축합하여 HMG-CoA를 합성하는 효소를 포함한다. 일부 구현예들에서, MEV 경로의 하나 이상의 효소는 HMG-CoA를 메발로네이트로 변환하는 효소를 포함한다. 일부 구현예들에서, MEV 경로의 하나 이상의 효소는 메발로네이트를 메발로네이트 5-포스페이트로 인산화하는 효소를 포함한다. 일부 구현예들에서, MEV 경로의 하나 이상의 효소는 메발로네이트 5-피로포스페이트로 변환하는 효소를 포함한다. 일부 구현예들에서, MEV 경로의 하나 이상의 효소는 메발로네이트 5-피로포스페이트를 이소펜테닐 피로포스페이트로 변환하는 효소를 포함한다. 일부 구현예들에서, 숙주 세포는 MEV 경로의 전체 효소들을 코딩하는 이중의 복수의 핵산을 포함한다. 일부 구현예들에서, MEV 경로의 하나 이상의 효소를 코딩하는 하나 이상의 이중의 핵산은 하나의 전사 조절인자의 통제 하에 배치된다. 일부 구현예들에서, MEV 경로의 하나 이상의 효소를 코딩하는 하나 이상의 이중의 핵산은 복수의 이중의 전사 조절인자의 통제 하에 배치된다. 일부 구현예들에서, 유전자 변형된 숙주 세포는 이소펜테닐 피로포스페이트 (IPP)를 다이메틸알릴 피로포스페이트 (DMAPP)로 변환할 수 있는 효소를 코딩하는 이중의 핵산을 더 포함한다. 일부 구현예들에서,

유전자 변형된 숙주 세포는 IPP 및/또는 DMAPP 분자를 축합하여 폴리프레닐 화합물을 형성할 수 있는 효소를 코딩하는 이중의 핵산을 더 포함한다. 일부 구현예들에서, 유전자 변형된 숙주 세포는 IPP 또는 폴리프레닐을 변형시켜 이소프레노이드 화합물을 형성할 수 있는 효소를 코딩하는 이중의 핵산을 더 포함한다.

[0016] 일부 구현예들에서, IPP 또는 폴리프레닐을 변형시켜 이소프레노이드 화합물을 형성할 수 있는 효소는, 카렌 신타제, 게라니올 신타제, 리날롤 신타제, 리모넨 신타제, 미르센 신타제, 옥시멘 신타제, α-피넨 신타제, β-피넨 신타제, 감마-테르피넨 신타제, 테르피놀렌 신타제, 아모르파다이엔 신타제, α-파르네센 신타제, β-파르네센 신타제, 파르네솔 신타제, 네롤리돌 신타제, 패출리올 (patchouliol) 신타제, 누트카톤 (nootkatone) 신타제 및 아비에타다이엔 (abietadien) 신타제로 이루어진 군으로부터 선택된다.

[0017] 일부 구현예들에서, 이소프레노이드는 헤미테르펜, 모노테르펜, 다이테르펜, 트리테르펜, 테트라테르펜, 세스퀴테르펜 및 폴리테르펜으로 이루어진 군으로부터 선택된다. 일부 구현예들에서, 이소프레노이드는 세스퀴테르펜이다. 일부 구현예들에서, 이소프레노이드는 C<sub>5</sub>-C<sub>20</sub> 이소프레노이드이다. 일부 구현예들에서, 이소프레노이드는 아비에타다이엔, 아모르파다이엔, 카렌, α-파르네센, β-파르네센, 파르네솔, 게라니올, 게라닐게라니올, 이소프렌, 리날롤, 리모넨, 미르센, 네롤리돌, 옥시멘, 패출롤, β-피넨, 사비넨, γ-테르피넨, 테르피놀렌 및 발렌센으로 이루어진 군으로부터 선택된다.

[0018] 다른 측면에서, 본 발명은, 이소펜테닐 피로포스페이트를 제조하기 위한 메발로네이트 (MEV) 경로의 하나 이상의 효소를 코딩하는 하나 이상의 이중의 핵산; 포스포케톨라제 (PK)를 코딩하는 이중의 핵산; 포스포트랜스아세틸라제 (PTA)를 코딩하는 이중의 핵산; 및 글리세롤-1-포스파타제 (EC 3.1.3.21)의 기능성 파괴를 포함하는, 이소프레노이드를 생산할 수 있는 유전자 변형된 숙주 세포를 제공한다. 일부 구현예들에서, 글리세롤-1-포스파타제는 GPP1/RHR2 또는 이의 상동체 또는 변이체이다. 일부 구현예들에서, 글리세롤-1-포스파타제는 GPP2/HOR2 또는 이의 상동체 또는 변이체이다.

[0019] 다른 측면에서, 본 발명은, 이소펜테닐 피로포스페이트를 제조하기 위한 메발로네이트 (MEV) 경로의 하나 이상의 효소를 코딩하는 하나 이상의 이중의 핵산; 아세틸알데하이드 데하이드로게나제, 아세틸화 (ADA)를 코딩하는 이중의 핵산; 포스포케톨라제 (PK)를 코딩하는 이중의 핵산; 포스포트랜스아세틸라제 (PTA)를 코딩하는 이중의 핵산; 및 글리세롤-1-포스파타제 (EC 3.1.3.21)의 기능성 파괴를 포함하는, 이소프레노이드를 생산할 수 있는 유전자 변형된 숙주 세포를 제공한다. 일부 구현예들에서, 글리세롤-1-포스파타제는 GPP1/RHR2 또는 이의 상동체 또는 변이체이다. 일부 구현예들에서, 글리세롤-1-포스파타제는 GPP2/HOR2 또는 이의 상동체 또는 변이체이다.

[0020] 다른 측면에서, 본 발명은, 이소펜테닐 피로포스페이트를 제조하기 위한 메발로네이트 (MEV) 경로의 하나 이상의 효소를 코딩하는 하나 이상의 이중의 핵산; 아세틸알데하이드 데하이드로게나제, 아세틸화 (ADA)를 코딩하는 이중의 핵산; 아세틸-CoA 신테타제 1 (ACS1), 아세틸-CoA 신테타제 2 (ACS2) 및 알데하이드 데하이드로게나제 6 (ALD6)로 이루어진 군으로부터 선택되는 천연적인 PDH-바이패스의 하나 이상의 효소의 기능성 파괴; 포스포케톨라제 (PK)를 코딩하는 이중의 핵산; 포스포트랜스아세틸라제 (PTA)를 코딩하는 이중의 핵산; 및 글리세롤-1-포스파타제 (EC 3.1.3.21)의 기능성 파괴를 포함하는, 이소프레노이드를 생산할 수 있는 유전자 변형된 숙주 세포를 제공한다. 일부 구현예들에서, 글리세롤-1-포스파타제는 GPP1/RHR2 또는 이의 상동체 또는 변이체이다. 일부 구현예들에서, 글리세롤-1-포스파타제는 GPP2/HOR2 또는 이의 상동체 또는 변이체이다.

[0021] 다른 측면에서, 본 발명은, 이소펜테닐 피로포스페이트를 제조하기 위한 메발로네이트 (MEV) 경로의 하나 이상의 효소를 코딩하는 하나 이상의 이중의 핵산으로서, 상기 하나 이상의 효소는 NADH-이용성 HMG-CoA 리덕타제를 포함하는 것인, 핵산; 아세틸알데하이드 데하이드로게나제, 아세틸화 (ADA)를 코딩하는 이중의 핵산; 아세틸-CoA 신테타제 1 (ACS1), 아세틸-CoA 신테타제 2 (ACS2) 및 알데하이드 데하이드로게나제 6 (ALD6)로 이루어진 군으로부터 선택되는 천연적인 PDH-바이패스의 하나 이상의 효소의 기능성 파괴; 포스포케톨라제 (PK)를 코딩하는 이중의 핵산; 포스포트랜스아세틸라제 (PTA)를 코딩하는 이중의 핵산; 및 글리세롤-1-포스파타제 (EC 3.1.3.21)의 기능성 파괴를 포함하는, 이소프레노이드를 생산할 수 있는 유전자 변형된 숙주 세포를 제공한다. 일부 구현예들에서, 글리세롤-1-포스파타제는 GPP1/RHR2 또는 이의 상동체 또는 변이체이다. 일부 구현예들에서, 글리세롤-1-포스파타제는 GPP2/HOR2 또는 이의 상동체 또는 변이체이다.

[0022] 다른 측면에서, 본 발명은, 이소펜테닐 피로포스페이트를 제조하기 위한 메발로네이트 (mev) 경로의 복수의 효소를 코딩하는 하나 이상의 이중의 핵산으로서, 상기 복수의 효소가 아세틸-CoA:말로닐-CoA 아세틸트랜스퍼라제를 포함하는 것인, 핵산; 아세틸알데하이드 데하이드로게나제, 아세틸화 (ADA)를 코딩하는 이중의 핵산; 아세틸-CoA 신테타제 1 (ACS1), 아세틸-CoA 신테타제 2 (ACS2) 및 알데하이드 데하이드로게나제 6 (ALD6)로 이루어진

균으로부터 선택되는 천연적인 PDH-바이패스의 하나 이상의 효소의 기능성 파괴; 포스포케톨라제 (PK)를 코딩하는 이중의 핵산; 포스포트랜스아세틸라제 (PTA)를 코딩하는 이중의 핵산; 및 글리세롤-1-포스파타제 (EC 3.1.3.21)의 기능성 파괴를 포함하는, 이소프레노이드를 생산할 수 있는 유전자 변형된 숙주 세포를 제공한다. 일부 구현예들에서, 글리세롤-1-포스파타제는 GPP1/RHR2 또는 이의 상동체 또는 변이체이다. 일부 구현예들에서, 글리세롤-1-포스파타제는 GPP2/HOR2 또는 이의 상동체 또는 변이체이다.

[0023] 다른 측면에서, 본 발명은, 폴리케티드 생합성 경로의 하나 이상의 효소를 코딩하는 하나 이상의 이중의 핵산; 포스포케톨라제 (PK)를 코딩하는 이중의 핵산; 포스포트랜스아세틸라제 (PTA)를 코딩하는 이중의 핵산; 및 글리세롤-1-포스파타제 (EC 3.1.3.21)의 기능성 파괴를 포함하는, 폴리케티드를 생산할 수 있는 유전자 변형된 숙주 세포를 제공한다. 일부 구현예들에서, 글리세롤-1-포스파타제는 GPP1/RHR2 또는 이의 상동체 또는 변이체이다. 일부 구현예들에서, 글리세롤-1-포스파타제는 GPP2/HOR2 또는 이의 상동체 또는 변이체이다.

[0024] 다른 측면에서, 본 발명은, 지방산 생합성 경로의 하나 이상의 효소를 코딩하는 하나 이상의 이중의 핵산; 포스포케톨라제 (PK)를 코딩하는 이중의 핵산; 포스포트랜스아세틸라제 (PTA)를 코딩하는 이중의 핵산; 및 글리세롤-1-포스파타제 (EC 3.1.3.21)의 기능성 파괴를 포함하는, 지방산을 생산할 수 있는 유전자 변형된 숙주 세포를 제공한다. 일부 구현예들에서, 글리세롤-1-포스파타제는 GPP1/RHR2 또는 이의 상동체 또는 변이체이다. 일부 구현예들에서, 글리세롤-1-포스파타제는 GPP2/HOR2 또는 이의 상동체 또는 변이체이다.

[0025] 일부 구현예들에서, 본원에 제공되는 유전자 변형된 숙주 세포는 박테리아 세포, 진균류 세포, 조류 세포, 곤충 세포 및 식물 세포로 이루어진 균으로부터 선택된다. 일부 구현예들에서, 세포는 효모 세포이다. 일부 구현예들에서, 효모는 사카로마이세스 세레비지애 (*Saccharomyces cerevisiae*)이다.

[0026] 일부 구현예들에서, 유전자 변형된 숙주 세포는, 아세틸 포스페이트를 아세테이트로 변환하는 내인성 효소의 기능적 파괴를 포함하지 않는 효모 세포와 비교하여, 아세틸-CoA 유래 화합물 (예, 이소프레노이드, 폴리케티드 또는 지방산)을 증가된 양으로 생산한다.

[0027] 다른 측면에서, 본 발명은, 본원에 기술된 이중의 아세틸-CoA 유래 화합물을 생산할 수 있는 유전자 변형된 숙주 세포 개체군을 상기 이중의 아세틸-CoA 유래 화합물을 생산하는데 적합한 조건 하에 탄소원이 첨가된 배지에서 배양하는 단계; 및 상기 배지로부터 이중의 아세틸-CoA 유래 화합물을 회수하는 단계를 포함하는, 이중의 아세틸-CoA 유래 화합물의 생산 방법을 제공한다. 일부 구현예들에서, 이중의 아세틸-CoA 유래 화합물은 이소프레노이드, 폴리케티드 및 지방산으로 이루어진 균으로부터 선택된다.

[0028] 다른 측면에서, 본 발명은, 숙주 세포에서 포스포케톨라제 (PK; EC 4.1.2.9)를 코딩하는 이중의 핵산을 발현하는 단계; 및 아세틸 포스페이트를 아세테이트로 변환하는 내인성 효소를 기능적으로 파괴하는 단계를 포함하는, 숙주 세포에서 아세틸-CoA 또는 아세틸-CoA 유래 화합물의 생산을 증가시키는 방법을 제공한다. 일부 구현예들에서, 본 방법은 숙주 세포에서 포스포트랜스아세틸라제 (PTA; EC 2.3.1.8)를 코딩하는 이중의 핵산을 발현하는 단계를 더 포함한다.

[0029] 다른 측면에서, 본 발명은, 숙주 세포에서 포스포트랜스아세틸라제 (PTA; EC 2.3.1.8)를 코딩하는 이중의 핵산을 발현하는 단계; 및 아세틸 포스페이트를 아세테이트로 변환하는 내인성 효소를 기능적으로 파괴하는 단계를 포함하는, 숙주 세포에서 아세틸-CoA의 생산을 증가시키는 방법을 제공한다. 일부 구현예들에서, 본 방법은 숙주 세포에서 포스포케톨라제 (PK; EC 4.1.2.9)를 코딩하는 이중의 핵산을 발현하는 단계를 더 포함한다.

[0030] 일부 구현예들에서, 아세틸 포스페이트를 아세테이트로 변환하는 효소는 글리세롤-1-포스파타제 (EC 3.1.3.21)이다. 일부 구현예들에서, 글리세롤-1-포스파타제는 GPP1/RHR2, GPP2/HOR2 및 이들의 상동체 및 변이체로부터 선택된다. 일부 구현예들에서, GPP1/RHR2 또는 이의 상동체 또는 변이체는 기능적으로 파괴된다. 일부 구현예들에서, GPP2/HOR2 또는 이의 상동체 또는 변이체는 기능적으로 파괴된다. 일부 구현예들에서, GPP1/RHR2 및 GPP2/HOR2 둘다, 또는 GPP1/RHR2의 상동체 또는 변이체 및 GPP2/HOR2의 상동체 또는 변이체 둘다, 기능적으로 파괴된다. 일부 구현예들에서, 숙주 세포는 박테리아 세포, 진균류 세포, 조류 세포, 곤충 세포 및 식물 세포로 이루어진 균으로부터 선택된다. 일부 구현예들에서, 숙주 세포는 효모 세포이다. 일부 구현예들에서, 효모는 사카로마이세스 세레비지애이다. 일부 구현예들에서, 숙주 세포는, 아세틸 포스페이트를 아세테이트로 변환하는 내인성 효소의 기능적 파괴를 포함하지 않는 효모 세포와 비교하여, 아세틸-CoA 또는 아세틸-CoA 유래 화합물을 증가된 양으로 생산한다.

**도면의 간단한 설명**

[0031]

**도 1**은 효모 숙주 세포에서 당 (글루코스 및 자일로스)을 아세틸-CoA 및 아세틸-CoA 유래 화합물로 변환하는데 관여하는 경로들을 개략적으로 도시한 것이다. 진한 화살 표시는 재조합 포스포케톨라제 경로를 나타낸다. 아세틸 포스페이트는 아세틸-CoA로 향하는 포스포케톨라제 (PK) / 포스포트랜스아세틸라제 (PTA) 경로의 중간 산물이며, RHR2 및 HOR2에 의해 아세테이트로 가수분해된다. 약어: G6P, 글루코스-6-포스페이트; R5P, 리불로스-5-포스페이트; X5P, 자일로스-5-포스페이트; F6P, 프럭토스-6-포스페이트; E4P, 에리트로스-4-포스페이트; FBP, 프럭토스-1,6-바이포스페이트; DHAP, 다이하이드록시아세톤 포스페이트; G3P, 글리세르알데하이드-3-포스페이트; PEP, 포스포에놀피루베이트; ADA, 아세트알데하이드 데하이드로게나제, 아세틸화; ACP, 아세틸 포스페이트.

**도 2**는 이소프레노이드를 생산하기 위한 메발로네이트 경로의 효소들을 도시한 것이다. 약어: AcCoA, 아세틸-CoA; AcAcCoA, 아세토아세틸-CoA; HMGCoA, 3-하이드록시-3-메틸글루타릴-CoA; Mev5P, 메발로네이트-5-포스페이트; Mev5DP, 메발로네이트-5-다이포스페이트; IPP, 이소펜테닐 다이포스페이트; DMAPP, 다이메틸알릴 피로포스페이트; Erg10, 아세틸-CoA 티올라제; ACC1, 아세틸-CoA 카르복실라제; AACS, 아세토아세틸-CoA 신타제; Erg13, 3-하이드록시-3-메틸글루타릴-CoA 신타제; HMGr, 3-하이드록시-3-메틸글루타릴-CoA 리덕타제; Erg12, 메발로네이트 키나제; Erg8, 포스포메발로네이트 키나제; Erg19, 메발로네이트 피로포스페이트 데카르복실라제.

**도 3A-3B**는 야생형 (균주 Y967, 좌), 및 이중의 아세트알데하이드 데하이드로게나제 아실화 (Dz.eutE)를 포함하며 천연적인 PDH-바이패스 (*acs1Δ acs2Δ ald6Δ*)가 결손되며 (균주 Y12869; 중간); 이중의 포스포케톨라제 (Lm.PK) 및 포스포트랜스아세틸라제 (Ck.PTA) (균주 Y12746; 우)를 더 포함하는, 재조합 효모 세포 (중간, 우)의, 당 소비 (A) 및 아세테이트 생산 (B)을 도시한 것이다.

**도 3C-3D**는 이중의 아세트알데하이드 데하이드로게나제 아실화 (Dz.eutE)를 포함하며 천연적인 PDH-바이패스 (*acs1Δ acs2Δ ald6Δ*)가 결손되며 (균주 Y12869; 좌); 이중의 포스포케톨라제 (Lm.PK) (균주 Y19390; 중간) 또는 포스포트랜스아세틸라제 (Ck.PTA) (균주 Y19391; 우)를 더 포함하는, 재조합 효모 세포의, 당 소비 (C) 및 아세테이트 생산 (D)을 도시한 것이다.

**도 4**는 야생형 사카로마이세스 세레비지에 균주 Y967의 세포-무함유 추출물 (CFE)에서 120분의 시간 경과 동안 아세틸 포스페이트 가수분해를 나타낸 것이다. CFE 단독 (좌); CFE + 30 mM 소듐 플루오라이드, 광역 포스파타제 저해제 (중양); 및 열에 의해 불활화된 CFE (우)가 도시된다.

**도 5**는 Y967 세포-무함유 추출물의 음이온 교환 크로마토그래피의 결과를 도시한 것이다. 단백질은 유속 0.5 mL/분으로 컬럼 부피의 30배로 완충제 B (20 mM Tris-Cl pH 7, 1M NaCl, 10% 글리세롤)를 0-100% 농도 구배로 사용하여 용출시키고, 분획 1 mL을 수집하여, 단백질 겔 전기영동으로 분석 (**도 5B**)하고, 아세틸 포스파타제 활성 (**도 5A**)을 분석하였다. ACP, 아세틸 포스페이트.

**도 6A**는 Y967의 10번 분획의 음이온 교환 크로마토그래피 결과를 도시한 것이다. 이 정제 공정에서 가장 고 활성인 분획은 14번으로, 이를 질량 분광측정으로 분석하여, 분획내 단백질의 정체를 확인하였다 (**도 6B**). RHR2는 활성 분획에서 포스파타제로서 동정되었다.

**도 7**은 야생형 효모 균주 (Y968) 또는 RHR2, HOR2, 또는 RHR2와 HOR2가 결손된 재조합 효모 균주의 CFE에 대한 아세틸 포스파타제 활성 분석 결과를 도시한 것이다.

**도 8A-8C**는 재조합 효모 균주 개체군의 아세테이트 수준 (A), 글리세롤 수준 (B) 및 광학 밀도 (C)를 나타낸 것이다. 균주 Y12746.ms63909.ms64472은 PDH-바이패스 (*acs1Δ acs2Δ ald6Δ*)의 결손을 포함하며, 아세트알데하이드 데하이드로게나제 아세틸화 (Dz.eutE), 포스포케톨라제 (Lm.PK), 포스포트랜스아세틸라제 (Ck.PTA) 및 파르네센 생산 경로의 유전자들을 이중적으로 발현한다. 균주 Y12746.ms63909.ms64472 rhr2<sup>+</sup>는 균주 Y12746.ms63909.ms64472와 동계이며, RHR2 (rhr2<sup>+</sup>) 결손을 추가로 포함한다.

**도 8D-8E**는 재조합 효모 균주 개체군의 아세테이트 수준 (D)과 광학 밀도 (E)를 나타낸 것이다. 균주 Y12745는 PDH-바이패스 (*acs1Δ acs2Δ ald6Δ*)의 결손을 포함하며, 아세트알데하이드 데하이드로게나제 아세틸화 (Dz.eutE), 포스포케톨라제 (Lm.PK) 및 포스포트랜스아세틸라제 (Ck.PTA)를 이중적으로 발현한다. 균주 Y12746 rhr2<sup>+</sup>는 Y12746과 동계이지만, RHR2 (rhr2<sup>+</sup>) 결손을 추가로 포함한다.

**도 9**는 RHR2 유전자가 온전 (RHR2<sup>+</sup>)하거나 결손 (rhr2<sup>-</sup>)된 재조합 효모 균주 개체군의 상대적인 파르네센 수준 (상단) 및 상대적인 광학 밀도 (하단)를 나타낸 것이다. Y968 (우측 패널)은 야생형 효모 균주이다. Y12869.ms63907.ms64472 ("Y12869"; 우측 패널에서 2번째)는 PDH-바이패스의 결손 (*acs1Δ acs2Δ ald6Δ*)을

포함하며, 아세트알데하이드 데하이드로게나제 아세틸화 (Dz.eutE) 및 파르네센 생산 경로의 유전자를 이중적으로 발현하지만, 포스포케틀라제 또는 포스포트랜스아세틸라제는 발현하지 않는다. Y12746.ms63907.ms64472 ("Y12746"; 좌측 패널에서 2번째)는 PDH-바이패스의 결손 (*acs1Δ acs2Δ ald6Δ*)을 포함하며, 아세트알데하이드 데하이드로게나제 아세틸화 (Dz.eutE) 및 파르네센 생산 경로의 유전자를 이중적으로 발현하며, 파르네센의 합성에 사용되는 세포질 아세틸-CoA 생산하기 위한 경로로서 포스포케틀라제와 포스포트랜스아세틸라제를 이용한다. Y12745.ms63907.ms64472 ("Y12745"; 좌측 패널)는 PDH-바이패스의 결손 (*acs1Δ acs2Δ ald6Δ*)과 파르네센 생산 경로의 유전자를 포함하며, 파르네센의 합성에 사용되는 세포질 아세틸-CoA 생산하기 위한 경로로서 포스포케틀라제와 포스포트랜스아세틸라제를 이용한다.

**발명을 실시하기 위한 구체적인 내용**

[0032]

**5. 구현예들에 대한 상세한 설명**

[0033]

**5.1 용어**

[0034]

본원에서, 용어 "이종의"는, 통상 천연적으로 발견되지 않는 것을 지칭한다. 용어 "이종의 뉴클레오티드 서열"은 자연에서 소정의 세포에서 정상적으로 발견되지 않는 뉴클레오티드 서열을 지칭한다. 이와 같이, 이종의 뉴클레오티드 서열은, (a) 이의 숙주 세포에 외래적 (즉, 세포에 "외인성")이거나; (b) 숙주 세포에서 천연적으로 발견되지만 (즉, "내인성"), 세포에서 비천연적인 양으로 존재 (즉, 숙주 세포에서 천연적으로 확인되는 양보다 많거나 적음)하거나; 또는 (c) 숙주 세포에서 본래 확인되지만 이의 본래의 위치에서 벗어난 위치에 있을 수 있다. 용어 "이종의 효소"는 자연에서 소정의 세포에서 정상적으로는 확인되지 않는 효소를 지칭한다. 이 용어는, (a) 소정의 세포에 외인성이고 (즉, 숙주 세포에 본래 존재하지 않거나 또는 해당 문맥에서 숙주 세포에서 본래 존재하지 않는 뉴클레오티드 서열에 의해 코딩됨); 및 (b) 숙주 세포에서 본래 발견되지만 (예, 효소가 세포에 내인성인 뉴클레오티드 서열에 의해 코딩됨), 숙주 세포에서 비정상적인 양 (예, 본래 확인되는 양보다 많거나 적게)으로 제조되는 효소를 포괄한다.

[0035]

한편, 본원에서, 분자, 특히 효소 및 핵산에 대해 사용되는 "천연적인" 또는 "내인성"이라는 용어는, 천연 미생물에서 분자의 발현 수준 보다 낮거나, 동일하거나 또는 높을 수 있는 발현 수준과는 무관하게, 자연에서 기원하거나 발견되는 유기체에서 발현되는 분자를 지칭한다. 천연 효소 또는 폴리뉴클레오티드의 발현은 재조합 미생물에서 변형될 수 있는 것으로 이해된다.

[0036]

본원에서, 타겟 유전자, 예컨대 PDH-바이패스의 하나 이상의 유전자의, 예컨대 "기능적으로 파괴한다" 또는 "기능적 파괴"는, 타겟 유전자가, 타겟 유전자에 의해 코딩된 단백질의 활성을 숙주 세포에서 낮추는 방식으로 변형되는 것을 의미한다. 일부 구현예들에서, 타겟 유전자의 기능적 파괴로, 기능적으로 파괴되지 않았을 경우의 발현과 비교해, 타겟 유전자의 발현 수준이 적어도 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% 또는 100%까지 감소된다. 마찬가지로, 예를 들어, 타겟 단백질, 예로 아세틸 포스포타제 활성을 가진 단백질의 "기능적으로 파괴한다" 또는 "기능적 파괴"는, 타겟 단백질이 숙주 세포에서 단백질의 활성을 감소시키는 방식으로 변형되는 것을 의미한다. 일부 구현예들에서, 타겟 단백질의 기능적 파괴는, 기능적으로 파괴되지 않았을 경우의 활성 또는 발현과 비교해, 타겟 단백질의 활성 또는 발현 수준을 적어도 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% 또는 100%까지 감소시킨다. 일부 구현예들에서, 타겟 유전자에 의해 코딩되는 타겟 단백질의 활성이 숙주 세포에서 소거된다. 다른 구현예들에서, 타겟 유전자에 의해 코딩되는 타겟 단백질의 활성이 숙주 세포에서 감소된다. 타겟 유전자의 기능적 파괴는, 유전자 발현이 없어지거나 감소되도록, 또는 유전자 산물의 활성이 사라지거나 감소되도록, 유전자 전체 또는 일부를 결손시킴으로써 달성될 수 있다. 또한, 타겟 유전자의 기능적 파괴는, 발현이 소거 또는 감소되도록 유전자의 조절인자, 예컨대, 유전자의 프로모터를 돌연변이시키거나, 또는 유전자 산물의 활성이 없어지거나 또는 감소되도록 유전자의 코딩 서열을 돌연변이시킴으로써, 달성될 수 있다. 일부 구현예들에서, 타겟 유전자의 기능적 파괴로 타겟 유전자의 오픈 리딩 프레임 전체가 제거된다.

[0037]

본원에서, 용어 "모 세포"는, 변형된 숙주 세포에서 조작되는 한가지 이상의 특정 유전자에 대한 변형, 예컨대, ADA의 이중 발현, NADH-이용성 HMG-CoA 리덕타제의 이중 발현, AACs의 이중 발현, 포스포케틀라제의 이중 발현, 포스포트랜스아세틸라제의 이중 발현 및 메발로네이트 경로의 하나 이상의 효소의 이중 발현으로 이루어진 군으로부터 선택되는 한가지 이상의 변형을 포함하지 않는다는 것을 제외하고는, 본원에 언급된 유전자 변형된 숙주 세포와 동일한 유전자 백그라운드를 가지는 세포를 지칭한다.

[0038]

본원에서, 용어 "생산"은 일반적으로 일정량의 이소프레노이드가 본원에 제시된 유전자 변형된 숙주 세포에 의

해 제조되는 것을 지칭한다. 일부 구현예들에서, 생산은 숙주 세포에서의 이소프레노이드의 수율로서 표시된다. 다른 구현예들에서, 생산은 이소프레노이드 생산에 대한 숙주 세포의 생산성으로 표시된다.

[0039] 본원에서, 용어 "생산성"은, 숙주 세포를 경시적으로 (시간 당) (부피) 배양한 발효 배지 양 당 제조되는 이소프레노이드 (중량) 양으로 표시되는, 숙주 세포에 의한 이소프레노이드의 생산을 지칭한다.

[0040] 본원에서, 용어 "수율"은 숙주 세포에 의해 소비되는 탄소원의 양 당 생산되는 이소프레노이드의 양을 중량으로 표시한, 숙주 세포에 의한 이소프레노이드의 생산을 지칭한다.

[0041] 본원에서, "아세틸-CoA 유래 화합물"이라는 표현은 이의 합성에 기질로서 아세틸-CoA가 이용된 화합물을 지칭한다. 예시적인 아세틸-CoA 유래 화합물로는 이소프레노이드, 폴리케티드, 지방산 및 알코올을 포함하지만, 이들로 한정되는 것은 아니다. 일부 구현예들에서, 아세틸-CoA 유래 화합물은 에탄올이고, 예를 들어 미국 특허 7,253,001에 기술된 바와 같이 펜토스 기질로부터 제조되는 바이오에탄올이며, 상기 문헌은 그 전체가 원용에 의해 본 명세서에 포함된다.

[0042] 본원에서, "변이체"라는 용어는 아미노산 삽입, 결손, 돌연변이 및 치환에 의해 구체적으로 언급된 "기준" 폴리펩타이드 (예, 야생형 서열)와 차이가 있지만, 기준 폴리펩타이드와 실질적으로 유사한 활성을 유지하는, 폴리펩타이드를 지칭한다. 일부 구현예들에서, 변이체는 돌연변이 유발 등의 재조합 DNA 기법에 의해 구축된다. 일부 구현예들에서, 변이체 폴리펩타이드는 다른 하나의 염기성 잔기 (즉, Lys의 경우 Arg), 다른 소수성 잔기 (즉, Ile의 경우 Leu) 또는 다른 하나의 방향족 잔기 (즉, Tyr의 경우 Phe)로의 치환 등에 의해 기준 폴리펩타이드와 차이가 있다. 일부 구현예들에서, 변이체는 보존적인 치환으로 기준 서열의 실질적인 구조적 유사성을 획득하는, 유사체를 포함한다. 이러한 보존적인 치환의 예로는 아스파르트산에 대한 글루탐산 및 그 역; 아스파라긴에 대한 글루타민 및 그 역; 트레오닌에 대한 세린 및 그 역; 아르기닌에 대한 라이신 및 그 역; 또는 이소루신, 발린 또는 루신 중 어느 하나에 대한 이들 중 임의의 것을 포함하나, 이들로 한정되지 않는다.

[0043] **5.2 숙주 세포**

[0044] 본원에 제공되는 조성물과 방법에 사용가능한 숙주 세포는 고세균 (archae), 원핵생물 또는 진핵생물 세포를 포함한다.

[0045] 적합한 원핵생물 숙주로는, 비제한적인 예로, 다양한 그람-양성, 그람-음성 또는 그람-가변성 박테리아 (gram-variable bacteria) 모두를 포함한다. 예로는, 비제한적으로, 다음과 같은 속에 속하는 세포를 포함한다: 아그로박테리움 (*Agrobacterium*), 알리사이클로바실러스 (*Alicyclobacillus*), 아나바에나 (*Anabaena*), 아나시스티스 (*Anacystis*), 아르트로박터 (*Arthrobacter*), 아조박터 (*Azobacter*), 바실러스 (*Bacillus*), 브레비박테리움 (*Brevibacterium*), 크로마티움 (*Chromatium*), 클로스트리듐 (*Clostridium*), 코리네박테리움 (*Corynebacterium*), 엔테로박터 (*Enterobacter*), 에르위니아 (*Erwinia*), 에스케리키아 (*Escherichia*), 락토바실러스 (*Lactobacillus*), 락토코커스 (*Lactococcus*), 메소리조븀 (*Mesorhizobium*), 메틸로박테리움 (*Methylobacterium*), 마이크로박테리움 (*Microbacterium*), 포르미듐 (*Phormidium*), 슈도모나스 (*Pseudomonas*), 로도박터 (*Rhodobacter*), 로도슈도모나스 (*Rhodopseudomonas*), 로도스피릴룸 (*Rhodospirillum*), 로도코커스 (*Rhodococcus*), 살모넬라 (*Salmonella*), 세네데스문 (*Scenedesmun*), 세라티아 (*Serratia*), 시겔라 (*Shigella*), 스타필로코커스 (*Staphylococcus*), 스트렙토마이세스 (*Streptomyces*), 신네코쿠스 (*Synnecoccus*) 및 자이모모나스 (*Zymomonas*). 원핵생물 균주의 예로는, 비제한적으로, 바실러스 썩틸리스 (*Bacillus subtilis*), 바실러스 아밀로리퀘팩시네스 (*Bacillus amyloliquefacines*), 브레비박테리움 암모니아케네스 (*Brevibacterium ammoniagenes*), 브레비박테리움 임마리오필럼 (*Brevibacterium immariophilum*), 클로스트리듐 베이예린키이 (*Clostridium beigerinckii*), 엔테로박터 사카자키 (*Enterobacter sakazakii*), 에스케리키아 콜라이 (*Escherichia coli*), 락토코커스 락티스 (*Lactococcus lactis*), 메소리조븀 로티 (*Mesorhizobium loti*), 슈도모나스 에어루지노사 (*Pseudomonas aeruginosa*), 슈도모나스 메발로니 (*Pseudomonas mevalonii*), 슈도모나스 푸디카 (*Pseudomonas pudica*), 로도박터 캡슐라투스 (*Rhodobacter capsulatus*), 로도박터 스페어로이데스 (*Rhodobacter sphaeroides*), 로도스피릴룸 루브럼 (*Rhodospirillum rubrum*), 살모넬라 엔테리카 (*Salmonella enterica*), 살모넬라 티피 (*Salmonella typhi*), 살모넬라 티피무리움 (*Salmonella typhimurium*), 시겔라 다이센테리아 (*Shigella dysenteriae*), 시겔라 플렉스네리 (*Shigella flexneri*), 시겔라 소네이 (*Shigella sonnei*) 및 스타필로코커스 아우레우스 (*Staphylococcus aureus*)를 포함한다. 특정 구현예에서, 숙주 세포는 에스케리키아 콜라이 (*Escherichia coli*) 세포이다.

[0046] 적합한 고세균 숙주로는, 비제한적인 예로, 다음과 같은 속에 속하는 세포를 포함한다: 에어로피룸

(*Aeropyrum*), 아르케글로부스 (*Archaeoglobus*), 할로박테리움 (*Halobacterium*), 메타노코커스 (*Methanococcus*), 메타노박테리움 (*Methanobacterium*), 피로코커스 (*Pyrococcus*), 설폴로부스 (*Sulfolobus*), 및 썬모플라스마 (*Thermoplasma*). 고세균 균주의 예로는, 비제한적으로, 아르케오글로부스 풀기두스 (*Archaeoglobus fulgidus*), 할로박테리움 (*Halobacterium*) sp., 메타노코커스 야나스키 (*Methanococcus jannaschii*), 메타노박테리움 썬모토토트로피쿰 (*thermoautotrophicum*), 썬모플라스마 액시도필룸 (*Thermoplasma acidophilum*), 썬모플라스마 볼카눔 (*Thermoplasma volcanium*), 피로코커스 호리코시 (*Pyrococcus horikoshii*), 피로코커스 아비시 (*Pyrococcus abyssi*) 및 에어로피룸 퍼닉스 (*Aeropyrum pernix*)를 포함한다.

[0047] 적합한 진핵생물의 숙주로는, 비제한적인 예로, 진균류 세포, 조류 세포, 곤충 세포 및 식물 세포를 포함한다. 일부 구현예에서, 본 발명의 방법에 사용가능한 효모는 미생물 기탁 기관 (예, IFO, ATCC 등)에 기탁된 효모를 포함하며, 특히 액시쿨로코니둠 (*Aciculoconidium*), 암브로시오지마 (*Ambrosiozyma*), 아르트로아스쿠스 (*Arthroascus*), 아릭시오지마 (*Arxiozyma*), 아쉬비아 (*Ashbya*), 바브제비아 (*Babjevia*), 벤싱턴니아 (*Bensingtonia*), 보트리오아스쿠스 (*Botryoascus*), 보트리오지마 (*Botryozyma*), 브레타노마이세스 (*Brettanomyces*), 불레라 (*Bullera*), 불레로마이세스 (*Bulleromyces*), 칸디다 (*Candida*), 시테로마이세스 (*Citeromyces*), 클라비스포라 (*Clavispora*), 크립토코커스 (*Cryptococcus*), 시스토필로바시둠 (*Cystofilobasidium*), 데바리오마이세스 (*Debaryomyces*), 데카라 (*Dekkara*), 디포다스콕시스 (*Dipodascopsis*), 디포다스쿠스 (*Dipodascus*), 에니엘라 (*Eeniella*), 엔도마이콕셀라 (*Endomycopsella*), 에레마스쿠스 (*Eremascus*), 에레모테시움 (*Eremothecium*), 에리트로바시둠 (*Erythrobasidium*), 펠로마이세스 (*Fellomyces*), 필로바시둠 (*Filobasidium*), 갈락토마이세스 (*Galactomyces*), 제오트리쿰 (*Geotrichum*), 킬리리에르몽델라 (*Guilliermondella*), 한세니아스포라 (*Hanseniaspora*), 한세놀라 (*Hansenula*), 하세가와이 (*Hasegawaea*), 폴터마니아 (*Holtermannia*), 호르모아스쿠스 (*Hormoascus*), 하이포피키아 (*Hyphopichia*), 이사첸티아 (*Issatchenkia*), 클로엑케라 (*Kloeckera*), 클로엑케라스포라 (*Kloeckeraspora*), 클루이베로마이세스 (*Kluyveromyces*), 콘도아 (*Konoda*), 쿠라이시아 (*Kuraishia*), 쿠르츠마모마이세스 (*Kurtzmanomyces*), 루코스포리둠 (*Leucosporidium*), 리포마이세스 (*Lipomyces*), 로테로마이세스 (*Lodderomyces*), 말라세지아 (*Malassezia*), 메츠니코위아 (*Metschnikowia*), 므라키아 (*Mrakia*), 므소지마 (*Myxozyma*), 나드소니아 (*Nadsonia*), 나카자와이 (*Nakazawaea*), 네마토스포라 (*Nematospora*), 오가타에 (*Ogataea*), 오스포리디움 (*Oosporidium*), 파키솔렌 (*Pachysolen*), 파키티코스포라 (*Phachytichospora*), 파피아 (*Phaffia*), 피키아 (*Pichia*), 로도스포리둠 (*Rhodosporeidium*), 로도토룰라 (*Rhodotorula*), 사카로마이세스 (*Saccharomyces*), 사카로마이코데스 (*Saccharomycodes*), 사카로마이콕시스 (*Saccharomycopsis*), 사이토엘라 (*Saitoella*), 사카구키아 (*Sakaguchia*), 사투라노스포라 (*Saturnospora*), 스킴조블라스토스포리온 (*Schizoblastosporion*), 스킴조사카로마이세스 (*Schizosaccharomyces*), 슈반니오마이세스 (*Schwanniomyces*), 스포리디오볼루스 (*Sporidiobolus*), 스포로볼로마이세스 (*Sporobolomyces*), 스포로파키더미아 (*Sporopachydermia*), 스테파노아스쿠스 (*Stephanoascus*), 스테리그마토마이세스 (*Sterigmatomyces*), 스테리그마토스포리둠 (*Sterigmatosporidium*), 심비오타프리나 (*Symbiotaphrina*), 심포디오마이세스 (*Sympodiomyces*), 심포디오마이콕시스 (*Sympodiomycopsis*), 톨라스포라 (*Torulaspora*), 트리코스포리엘라 (*Trichosporiella*), 트리코스포론 (*Trichosporon*), 트리코넵시스 (*Trigonopsis*), 츠키야에아 (*Tsuchiyaea*), 우데니오마이세스 (*Udeniomyces*), 발토마이세스 (*Waltomyces*), 윌커하미아 (*Wickerhamia*), 윌커하미엘라 (*Wickerhamiella*), 윌리옵시스 (*Williopsis*), 야마다지마 (*Yamadazyma*), 야로위니아 (*Yarrowia*), 자이코아스쿠스 (*Zygoascus*), 자이코사카로마이세스 (*Zygosaccharomyces*), 자이코빌리옵시스 (*Zygowilliopsis*), 및 자이코지마 (*Zygozyma*)의 속에 속한다.

[0048] 일부 구현예에서, 숙주 미생물은 사카로마이세스 세레비지애 (*Saccharomyces cerevisiae*), 피키아 패스토리스 (*Pichia pastoris*), 스킴조사카로마이세스 폼베 (*Schizosaccharomyces pombe*), 데케라 브룩셀렌시스 (*Dekkera bruxellensis*), 클루이베로마이세스 락티스 (*Kluyveromyces lactis*) (예전에는 사카로마이세스 락티스 (*Saccharomyces lactis*)로 지칭됨), 클루이베로마이세스 마르시아누스 (*Kluyveromyces marxianus*), 아르술라 아데니니보란스 (*Arxula adenivorans*), 또는 한세놀라 폴리모르파 (*Hansenula polymorpha*) (지금은 피키아 앙구스타 (*Pichia angusta*)라고도 함)이다. 일부 구현예에서, 숙주 미생물은 칸디다 (*Candida*) 속의 균주, 예컨대 칸디다 리폴리티카 (*Candida lipolytica*), 칸디다 킬리에르몽디 (*Candida guilliermondii*), 칸디다 크루세이 (*Candida krusei*), 칸디다 슈도트로피칼리스 (*Candida pseudotropicalis*) 또는 칸디다 유틸리스 (*Candida utilis*)이다.

[0049] 특정 구현예에서, 숙주 미생물은 사카로마이세스 세레비지애 (*Saccharomyces cerevisiae*)이다. 일부 구현예들에서, 숙주는 빵 효모, CBS 7959, CBS 7960, CBS 7961, CBS 7962, CBS 7963, CBS 7964, IZ-1904, TA, BG-1, CR-1, SA-1, M-26, Y-904, PE-2, PE-5, VR-1, BR-1, BR-2, ME-2, VR-2, MA-3, MA-4, CAT-1, CB-1, NR-1, BT-1

및 AL-1으로 이루어진 균으로부터 선택되는 사카로마이세스 세레비지에 (*Saccharomyces cerevisiae*) 균주이다. 일부 구현예들에서, 숙주 미생물은 PE-2, CAT-1, VR-1, BG-1, CR-1 및 SA-1으로 이루어진 균으로부터 선택되는 사카로마이세스 세레비지에 (*Saccharomyces cerevisiae*) 균주이다. 특정 구현예에서, 사카로마이세스 세레비지에 (*Saccharomyces cerevisiae*) 균주는 PE-2이다. 다른 특정 구현예에서, 사카로마이세스 세레비지에 (*Saccharomyces cerevisiae*) 균주는 CAT-1이다. 다른 특정 구현예에서, 사카로마이세스 세레비지에 (*Saccharomyces cerevisiae*) 균주는 BG-1이다.

[0050] 일부 구현예들에서, 숙주 미생물은 공업 발효에 적합한 미생물이다. 특정 구현예에서, 미생물은, 공업적인 발효 환경에서 공인된 스트레스 조건인, 용매 고농도, 고온, 확장된 기질 이용성, 영양 제한, 당 및 염으로 인한 삼투 스트레스, 산도, 설파이트 및 박테리아 오염 또는 이들의 조합 하에 존속하도록 조건화된다.

[0051] **5.3 아세틸-CoA까지의 포스포케톨라제 (PK) / 포스포트랜스아세틸라제 (PTA) 경로**

[0052] 일부 구현예들에서, 포스포케톨라제 경로는, 포스포케톨라제, 선택적으로 포스포트랜스아세틸라제를 코딩하는 폴리뉴클레오티드 및/또는 폴리펩타이드를 발현하도록 세포를 조작함으로써, 본원에 제공된 유전자 변형된 숙주 세포에서 활성화된다. 따라서, 일부 구현예들에서, 본원에 제공된 유전자 변형된 숙주 세포는 포스포케톨라제 활성을 가진 폴리펩타이드를 코딩하는 이종의 폴리뉴클레오티드를 포함한다. 다른 구현예에서, 특히 아세틸 포스페이트가 포스포케톨라제 활성과는 무관하게 대사 중간산물로서 공급되는 경우, 본원에 제공되는 유전자 변형된 숙주 세포는 포스포트랜스아세틸라제 활성을 가진 폴리펩타이드를 코딩하는 이종의 폴리뉴클레오티드를 포함한다. 다른 구현예에서, 본원에 제공되는 유전자 변형된 숙주 세포는 포스포케톨라제 활성을 가진 폴리펩타이드를 코딩하는 이종의 폴리뉴클레오티드와 포스포트랜스아세틸라제 활성을 가진 폴리펩타이드를 코딩하는 이종의 폴리뉴클레오티드를 포함한다.

[0053] **5.3.1 포스포케톨라제 (PK)**

[0054] 포스포케톨라제 (EC 4.1.2.9)는 자일롤로스 5-포스페이트를 글리세르알데하이드 3-포스페이트와 아세틸 포스페이트로 변환하는 반응; 및/또는 프럭토스-6-포스페이트를 에리트로스-4-포스페이트와 아세틸 포스페이트를 변환하는 반응을 촉매한다. 포스포케톨라제 활성은 단독 탄소원으로 자일로스를 이용하여 증식한 수종의 효모 균주에서 동정되지만, 글루코스로 증식한 효모 균주에서는 동정되지 않는다 (Evans and Ratledge, *Arch. Microbiol.* 139: 48-52; 1984). 포스포케톨라제의 저해제로는, 비제한적으로 에리트로스 4-포스페이트 및 글리세르알데하이드 3-포스페이트를 포함한다.

[0055] 포스포케톨라제 활성을 코딩하는 폴리뉴클레오티드, 유전자 및 폴리펩타이드에 대한 다수 예들이 당해 기술 분야에 공지되어 있으며, 본원에 제시된 유전자 변형된 숙주 세포에서 이용할 수 있다. 일부 구현예에서, 상기한 폴리뉴클레오티드, 유전자 및/또는 폴리펩타이드는 락토바실러스 펜토서스 (*Lactobacillus pentosus*) MD363의 자일롤로스 5-포스페이트케톨라제 (XpkA)이다 (Posthuma *et al.*, *Appl. Environ. Microbiol.* 68: 831-7; 2002). XpkA는 유산균에서 포스포케톨라제 경로 (PKP)의 핵심 효소이며, 비활성 4.455  $\mu\text{mol}/\text{min}/\text{mg}$ 을 가진다 (Posthuma *et al.*, *Appl. Environ. Microbiol.* 68: 831-7; 2002). 다른 구현예에서, 이러한 폴리뉴클레오티드, 유전자 및/또는 폴리펩타이드는 루코노스톡 메센테로이데스 (*Leuconostoc mesenteroides*)의 포스포케톨라제이며 (Lee *et al.*, *Biotechnol Lett.* 27(12):853-858 (2005)), 비활성이 9.9  $\mu\text{mol}/\text{min}/\text{mg}$ 이고, pH 4.5 이상에서 안정적이다 (Goldberg *et al.*, *Methods Enzymol.* 9: 515-520; 1966). 이 포스포케톨라제는 D-자일롤로스 5-포스페이트에 대한 Km은 4.7 mM이고, 프럭토스 6-포스페이트에 대한 Km은 29 mM이다 (Goldberg *et al.*, *Methods Enzymol.* 9: 515-520; 1966). 루코노스톡 메센테로이데스 (*Leuconostoc mesenteroides*)의 예시적인 포스포케톨라제 뉴클레오티드 서열은 등재 번호 AY804190 및 본원에 제시된 서열번호 1을 포함한다. 루코노스톡 메센테로이데스 (*Leuconostoc mesenteroides*)의 예시적인 포스포케톨라제 단백질 서열로는 등재 번호 YP\_819405, AAV66077.1과 본원에 제시된 서열번호 2를 포함한다. 다른 구현예에서, 이러한 폴리뉴클레오티드, 유전자 및/또는 폴리펩타이드는, 예를 들어 Sonderegger *et al.* (*Appl. Environ. Microbiol.* 70: 2892-7; 2004)에서 펜토스-대사성 사카로마이세스 세레비지에 균주에서 언급된 바와 같은, 락토바실러스 락티스 유래의 D-자일롤로스 5-포스페이트/D-프럭토스 6-포스페이트 포스포케톨라제 유전자 xfp이다.

[0056] 다른 사용가능한 포스포케톨라제로는 비제한적으로, 비피도박테리움 덴티움 (*Bifidobacterium dentium*) ATCC 27678 (ABIX02000002.1:2350400..2352877; EDT46356.1); 비피도박테리움 애니말리스 (*Bifidobacterium animalis*) (NC\_017834.1:1127580..1130057; YP\_006280131.1); 및 비피도박테리움 슈돌롱검 (*Bifidobacterium pseudolongum*) (AY518216.1:988..3465; AAR98788.1); 아스퍼질러스 니둘란스 (*Aspergillus nidulans*) FGSC A4 (CBF76492.1); 비피도박테리움 롱검 (*Bifidobacterium longum*) (AAR98787.1); 비피도박테리움 비피덤

(*Bifidobacterium bifidum*) NCIMB 41171 (ZP 03646196.1); 비피도박테리움 애니말리스 아종. 락티스 (*Bifidobacterium animalis subsp. lactis*) HN019 (ZP 02962870.1); 락토바실러스 플라타룸 (*Lactobacillus plantarum*) WCFS1 (NP\_786060.1); 락토바실러스 브레비스 아종 그라베센시스 (*Lactobacillus brevis subsp. gravesensis*) ATCC 27305 (ZP\_03940142.1); 락토바실러스 루테리 (*Lactobacillus reuteri*) 100-23 (ZP\_03073172.1); 및 루코노스톡 메센테로이데스 아종 메센테로이데스 (*Leuconostoc mesenteroides subsp. mesenteroides*) ATCC 8293 (YP\_818922.1)로부터 유래된 것을 포함한다.

[0057]

다른 이용가능한 포스포케탈라제로는, 국제 공개공보 WO 2011/15985에 언급된 것을 포함하며, 이 문헌은 그 전체가 원용에 의해 본 명세서에 포함된다. 이들 포스포케탈라제로는, (YP\_001601863.1; 글루콘아세트도박터 디아조트로피쿠스 (*Gluconacetobacter diazotrophicus*) Pal 5), (YP\_001093221.1; 슈와넬라 로이히카 (*Shewanella loihica*) PV-4), (YP\_926792.1; 슈와넬라 아마존엔시스 (*Shewanella amazonensis*) SB2B), (YP\_735093.1; 슈와넬라 sp. MR-4), (YP\_001049439.1; 슈와넬라 발티카 (*Shewanella baltica*) OS155), (ZP\_02157884.1; 슈와넬라 벤티카 (*Shewanella benthica*) KT99), (YP\_001472925.1; 슈와넬라 세디미니스 (*Shewanella sediminis*) HAW-EB3), (YP\_001759669.1; 슈와넬라 우디 (*Shewanella woodyi*) ATCC 51908), (YP\_001673352.1; 슈와넬라 할리팩센시스 (*Shewanella halifaxensis*) HA W-EB4), (YP\_563733.1; 슈와넬라 데니트리피칸스 (*Shewanella denitrificans*) OS217), (ZP\_05111697.1; 레지오넬라 드란코우르티 (*Legionella drancourtii*) LLAP 12), (EEQ84307.1; 아젤로마이세스 더마티티디스 (*Ajellomyces dermatitidis*) ER-3), (XP\_002626734.1; 아젤로마이세스 더마티티디스 SLH14081), (XP\_001539009.1; 아젤로마이세스 캡슐라투스 (*Ajellomyces capsulatus*) Nam1), (EEH04133.1; 아젤로마이세스 캡슐라투스 G1864R), (EEH20258.1; 파라코크시디오이데스 브라실리엔시스 (*Paracoccidioides brasiliensis*) Pb03), (EEH44652.1; 파라코크시디오이데스 브라실리엔시스 (*Paracoccidioides brasiliensis*) Pb 18), (XP\_002582752.1; 운시노카르푸스 리시 (*Uncinocarpus reesii*) 1704), (EER26377.1; 코크시디오이데스 포사다시 (*Coccidioides posadasii*) C735 delta SOWgp), (EEQ28085.1; 미이크로스포럼 카니스 (*Microsporium canis*) CBS 113480), (XP\_001819785.1; 아스퍼질러스 오리제 (*Aspergillus oryzae*) RIB40), (XP\_001399780.1; 아스퍼질러스 나이저 (*Aspergillus niger*)), (XP\_001263382.1; 네오사르토르야 피셔리 (*Neosartorya fischeri*) NRRL 181), (XP\_001271080.1; 아스퍼질러스 클라카투스 (*Aspergillus clavatus*) NRRL 1), (XP\_001213784.1; 아스퍼질러스 테레우스 (*Aspergillus terreus*) NIH2624), (CBF76492.1; 아스퍼질러스 니들란스 (*Aspergillus nidulans*) FGSC44), (XP\_002561913.1; 페니실리움 크리소게눔 위스콘신 (*Penicillium chrysogenum Wisconsin*) 54-1255), (XP\_002480391.1; 탈라로마이세스 스티피타투스 (*Talaromyces stipitatus*) ATCC 10500), (XP\_002144014.1; 페니실리움 스티피타투스 (*Penicillium stipitatus*) ATCC 10500), (XP\_002144014.1; 페니실리움 마메페이 (*Penicillium mameffei*) ATCC 18224), (XP\_754543.1; 아스퍼질러스 푸미가투스 (*Aspergillus fumigatus*) Af293), (XP\_001556635.1; 보트리오티니아 펙켈리아나 (*Botryotinia fuckeliana*) B05.1 0), (XP\_001592549.1; 스크레로티니아 스크레로티오럼 (*Sclerotinia sclerotiorum*) 1980), (XP\_386729.1; 지베렐라 지에 (*Gibberella zeae*) PH-1), (EEU47171.1; 넥트리아 해마토코카 (*Nectria haematococca*) mp VI 77-13-4), (EEY16637.1; 버티실리움 알보아트룸 (*Verticillium alboatrum*) ValMs.1 02), (XP\_956649.1; 뉴로스포라 크라사 (*Neurospora crassa*) OR74A), (XP\_364271.2; 마그나포르테 그리세아 (*Magnaporthe grisea*) 70-15), (XP\_001904585.1; 포도스포라 안세린 (*Podospora anserine*)), (XP\_001836159.1; 모프리놉시스 시네리아 (*Coprinopsis cinerea*) okayama7#130), (NP\_595963.1; 시조사카로마이세스 폼베 (*Schizosaccharomyces pombe*)), (XP\_002173441.1; 시조사카로마이세스 자포니쿠스 (*Schizosaccharomyces japonicus*) yFS275), (XP\_570860.1; 크립토크커스 네오포르만스 버 네오포르만스 (*Cryptococcus neoformans var. neoformans*) JEC21), (XP\_759561.1; 우스틸라고 메이디스 (*Ustilago maydis*) 521), (ZP\_05027078.1; 마이크로콜레우스 크토노플라스테스 (*Microcoleus chthonoplastes*) PCC 7420), (YP\_003101114.1; 액티노시네마 미럼 (*Actinosynnema mirum*) DSM 43827), (ZP\_03568244.1; 아토포비움 리마에 (*Atopobium rimae*) ATCC 49626), (YP\_003180237.1; 아토포비움 파블럼 (*Atopobium parvulum*) DSM 20469), (ZP\_03946928.1; 아토포비움 바기나에 (*Atopobium vaginae*) DSM 15829), (ZP\_03296299.1; 콜린셀라 스테코리스 (*Collinsella stercoris*) DSM 13279), (AAR98787.1; 비피도박테리움 롱검 (*Bifidobacterium longum*)), (ZP\_03618909.1; 비피도박테리움 브레이브 (*Bifidobacterium breve*) DSM 20213), (ZP\_03646196.1; 비피도박테리움 비피덤 (*Bifidobacterium bifidum*) NCIMB 41171), (ZP\_04448101.1; 비피도박테리움 앙굴라툼 (*Bifidobacterium angulatum*) DSM 20098), (ZP\_03324204.1; 비피도박테리움 카테놀라툼 (*Bifidobacterium catenulatum*) DSM 16992), (AAR98790.1; 비피도박테리움 sp. CFAR 172), (AAR98789.1; 비피도박테리움 풀로럼 (*Bifidobacterium pullorum*)), (ZP\_03937610.1; 가드네렐라 바기날리스 (*Gardnerella vaginalis*) ATCC 14019), (ZP\_05965201.1; 비피도박테리움 갈리쿰 (*Bifidobacterium gallicum*) DSM 20093), (ZP\_02962870.1; 비

피도박테리움 애니말리스 아종 락티스 (*Bifidobacterium animalis subsp. lactis*) HNO19), (AAR98788.1; 비피도박테리움 슈돌롱검 아종 글로보숨 (*Bifidobacterium pseudolongum subsp. Globosum*)), (ZP\_03946518.1; 아토포비움 바기나에 (*Atopobium vaginae*) DSM 15829), (YP\_001511171.1; 프랭키아 (*Frankia*) sp. EAN1pec), (YP\_713678.1; 프랭키아 알니 (*Frankia alni*) ACN14a), (YP\_002778395.1; 로도코커스 오파쿠스 (*Rhodococcus opacus*) BA), (YP\_701466.1; 로도코커스 조스티이 (*Rhodococcus jostii*) RHA1), (ZP\_04383880.1; 로도코커스 에리트르폴리스 (*Rhodococcus erythropolis*) SK121), (YP\_947598.1; 아르트로박터 아우레센스 (*Arthrobacter aurescens*) TC 1), (CAD48946.1; 프로피오니박테리움 프레우텐레이키이 아종 셔마니이 (*Propionibacterium freudenreichii subsp. Shermanii*)), (NP\_791495.1; 슈도모나스 시링에 pv. 토마토 str. (*Pseudomonas syringae* pv. *Tomato* str.) DC3000), (YP\_003125992.1; 키티노파가 피넨시스 (*Chitinophaga pinensis*) DSM 2588), (ABX56639.1; 베루코마이ক্র로비에 박테리움 (*Verrucomicrobiae bacterium*) V4), (YP\_002371883.1; 시아노테시 sp. (*Cyanothecce* sp.) PCC 8801), (YP\_001806596.1; 시아노테시 sp. ATCC 51142), (ZP\_01730652.1; 시아노테시 sp. CCY0110), (CAQ48286.1; 플란кто트릭스 루베센스 (*Planktothrix rubescens*) NIVA-CYA 98), (ZP\_03276298.1; 아르트로스피라 맥시마 (*Arthrospira maxima*) CS-328), (ZP\_03157277.1; 시아노테시 sp. PCC 7822), (YP\_002379031.1; 시아노테시 sp. PCC 7424), (YP\_001658501.1; 마이크로시스티스 에어루지노사 (*Microcystis aeruginosa*) NIES-843), (ZP\_01621774.1; 링비야 (*Lyngbya*) sp. PCC 8106), (NP\_485524.1; 노스톡 (*Nostoc*) sp. PCC 7120), (ZP\_05036350.1; 시네코코커스 (*Synechococcus*) sp. PCC 7335), (YP\_001514813.1; 아카리오콜로리스 마리나 (*Acaryochloris marina*) MBIC 11 017), (ZP\_05039537.1; 시네코코커스 sp. PCC 7335), (ZP\_02886235.1; 버크홀데리아 그라미니스 (*Burkholderia graminis*) C4 DIM), (ZP\_03264503.1; 버크홀데리아 sp. H160), (ZP\_01085819.1; 시네코코커스 sp. WH 5701), (ZP\_05045603.1; 시아노비움 sp. PCC 7001), (ZP\_01123645.1; 시네코코커스 sp. WH 7805), (YP\_001223932.1; 시네코코커스 sp. WH 7803), (ZP\_01079038.1; 시네코코커스 sp. RS9917), (YP\_001889002.1; 버크홀데리아 피토퍼만스 (*Burkholderia phytofirmans*) PsJN), (YP\_553967.1; 버크홀데리아 제노보란스 (*Burkholderia xenovorans*) LB400), (ZP\_02881709.1; 버크홀데리아 그라미니스 (*Burkholderia graminis*) CADIM), (ZP\_03270532.1; 버크홀데리아 sp. H160), (YP\_001861620.1; 버크홀데리아 피마툼 (*Burkholderia phymatum*) STM815), (YP\_002755285.1; 액시도박테리움 캡슐라툼 (*Acidobacterium capsulatum*) ATCC 51196), (EDZ38884.1; 렙토스피릴룸 (*Leptospirillum*) sp. Group II '5-way CO'), (EES53204.1; 렙토스피릴룸 페로디아조트로피 (*Leptospirillum ferrodiazotrophum*)), (YP\_172723.1; 시네코코커스 일롱가투스 (*Synechococcus elongatus*) PCC 6301), (NP\_681976.1; 썬모시네코코커스 일롱가투스 BP-1), (YP\_114037.1; 메틸로코커스 캡슐라투스 (*Methylococcus capsulatus*) str. Bath), (YP\_002482577.1; 시아노테시 sp. PCC 7425), (NP\_442996.1; 시네코시스티스 (*Synechocystis*) sp. PCC 6803), (YP\_002482735.1; 시아노테시 sp. PCC 7425), (ZP\_04774866.1; 알로크로마티움 비노숨 (*Allochromatium vinosum*) DSM 180), (ZP\_01453148.1; 마리프로퓌두스 페로옥시단스 (*Mariprofundus ferrooxydans*) PV-1), (ZP\_04830548.1; 갈리오넬라 페루기니아 (*Gallionella ferruginea*) ES-2), (XP\_001273863.1; 아스퍼질러스 클라바투스 (*Aspergillus clavatus*) NRRL 1), (XP\_001258643.1; 네오사르토르야 피셔리 (*Neosartorya fischeri*) NRRL 181), (XP\_001727680.1; 아스퍼질러스 오리제 (*Aspergillus oryzae*) RIB40), (XP\_001396306.1; 아스퍼질러스 나이저 (*Aspergillus niger*)), (XP\_001216075.1; 아스퍼질러스 테레우스 (*Aspergillus terreus*) NIH2624), (XP\_002567130.1; 페니실리움 크리소게눔 위스콘신 (*Penicillium chrysogenum Wisconsin*) 54-1255), (XP\_002143851.1; 페니실리움 마르네페이 (*Penicillium marneffeii*) ATCC 18224), (XP\_002480216.1; 탈라로마이세스 스티피타투스 (*Talaromyces stipitatus*) ATCC 10500), (XP\_001559949.1; 보트리오티니아 펙켈리아나 (*Botryotinia fuckeliana*) B05.10), (XP\_001593100.1; 스크레로티니아 스크레로티어럼 (*Sclerotinia sclerotiorum*) 1980), (XP\_001932192.1; 피레노포라 트리티시레펜티스 (*Pyrenophora tritici-repentis*) Pt-1C-BFP), (XP\_001793729.1; 패오스페리아 노도룸 (*Phaeosphaeria nodorum*) SN 15), (XP\_567776.1; 크립토코커스 네오포르난스 var. 네오포르난스 (*Cryptococcus neoformans* var. *neoformans*) JEC21), (XP\_386504.1; 오이베렐라 지이 (*Ooibberella zeae*) PH-1), (EEU46265.1; 넥트리아 헤마토코카 (*Nectria haematococca*) mp VI 77-13-4), (AC024516.1; 메타리지움 아니솅리에 (*Metarhizium anisopliae*)), (XP\_959985.1; 뉴로스포라 크라사 (*Neurospora crassa*) OR74A), (XP\_001904686.1; 포도스포라 안세린 (*Podospora anserine*)), (YP\_002220141.1; 액시디티오바실러스 페로옥시단스 (*Acidithiobacillus ferrooxidans*) ATCC 53993), (YP\_001220128.1; 액시디필리움 크립툼 (*Acidiphilium cryptum*) JF -5), (YP\_001471202.1; 서모토가 레팅에 (*Thermotoga lettingae*) TMO), (YP\_002352287.1; 디티오글로무스 투르기둠 (*Dictyoglomus turgidum*) DSM 6724), (YP\_571790.1; 니트로박터 햄버겐시스 (*Nitrobacter hamburgensis*) X14), (ZP\_01092401.1; 발라스토피렐룰라 마리나 (*Blastopirellula marina*) DSM 3645), (YP\_001340809.1; 마리노모나스 (*Marinomonas*) sp. MWYL1), (NP\_866384.1; 로도피렐룰라 발티카 (*Rhodopirellula baltica*) SH 1),

(ZP\_05108502.1; 레기오넬라 드란코르티이 (*Legionella drancourtii*) LLAP 12), (ZP\_04995817.1; 스트렙토마이세스 *sp. Mg1*), (ZP\_04023055.1; 락토바실러스 루테리 (*Lactobacillus reuteri*) SD2112), (ZP\_03960060.1; 락토바실러스 바기날리스 (*Lactobacillus vaginalis*) ATCC 49540), (ZP\_03073172.1; 락토바실러스 루테리 (*Lactobacillus reuteri*) 100-23), (ZP\_05553031.1; 락토바실러스 콜레오호미니스 (*Lactobacillus coleohominis*) 101-4-CHN), (ZP\_05863347.1; 락토바실러스 퍼멘툼 (*Lactobacillus fermentum*) 28-3-CHN), (ZP\_04021289.1; 락토바실러스 액시도필러스 (*Lactobacillus acidophilus*) ATCC 4796), (ZP\_03995194.1; 락토바실러스 크리스파투스 (*Lactobacillus crispatus*) IV-VOI), (ZP\_04010922.1; 락토바실러스 울투넨시스 (*Lactobacillus ultunensis*) DSM 16047), (ZP\_05549961.1; 락토바실러스 크리스파투스 (*Lactobacillus crispatus*) 125-2-CRN), (ZP\_03951361.1; 락토바실러스 가세리 (*Lactobacillus gasserii*) IV-V03), (ZP\_05744515.1; 락토바실러스 이너스 (*Lactobacillus iners*) DSM 13335), (YP\_618635.1; 락토바실러스 델브루에키이 아종 불가리쿠스 (*Lactobacillus delbrueckii* subsp. *bulgaricus*) ATCC 11842), (ZP\_03955917.1; 락토바실러스 옌세나이 (*Lactobacillus jensenii*) IV-VI6), (ZP\_03942415.1; 락토바실러스 부크네리 (*Lactobacillus buchneri*) ATCC 11577), (ZP\_01544800.1; 오에노코쿠스 오에니 (*Oenococcus oeni*) ATCC BAA-1163), (NP\_786060.1; 락토바실러스 플라타룸 (*Lactobacillus plantarum*) WCFST), (Q937F6; XPKA\_LACPE), (YP\_394903.1; 락토바실러스 사케이 아종 사케이 (*Lactobacillus sakei* subsp. *sakei*) 23K), (YP\_803891.1; 페디오코쿠스 펜토사세우스 (*Pediococcus pentosaceus*) ATCC 25745), (BAI40727.1; 락토바실러스 람노수스 (*Lactobacillus rhamnosus*) GG), (ZP\_03940142.1; 락토바실러스 바레비스 아종 그라베센시스 (*Lactobacillus brevis* subsp. *Gravesensis*) ATCC 27305), (ZP\_04009273.1; 락토바실러스 살리바리우스 (*Lactobacillus salivarius*) ATCC 11741), (ZP\_03958643.1; 락토바실러스 루미니스 (*Lactobacillus ruminis*) ATCC 25644), (ZP\_04431433.1; 바실러스 코아굴란스 (*Bacillus coagulans*) 36DI), (ZP\_04601906.1; 킹엘라 오랄리스 (*Kingella oralis*) ATCC 51147), (ZP\_05736927.1; 그라눌리카텔라 아디아센스 (*Granulicatella adiacens*) ATCC 49175), (YP\_001449631.1; 스트렙토코쿠스 고르도니이 str. 칼리스 아종 (*Streptococcus gordonii* str. *Challis* substr.) CHI), (NP\_736274.1; 스트렙토코쿠스 아갈락티에 (*Streptococcus agalactiae*) NEM316), (ZP\_04442854.1; 리스테리아 그레이 (*Listeria grayi*) DSM 20601), (ZP\_05646360.1; 엔테로코쿠스 카셀리플라부스 (*Enterococcus casseliflavus*) EC30), (ZP\_05650322.1; 엔테로코쿠스 갈리나룸 (*Enterococcus gallinarum*) EG2), (ZP\_05675307.1; 엔테로코쿠스 패슘 (*Enterococcus faecium*) Com12), (BAH69929.1; 마이코플라스마 퍼멘탄스 (*Mycoplasma fermentans*) PG 18), (YP\_002000006.1; 마이코플라스마 아르트리티디스 (*Mycoplasma arthritidis*) 15 8L3-1), (YP\_001256266.1; 마이코플라스마 아갈락티에 (*Mycoplasma agalactiae*) PG2), (YP\_001988835.1; 락토바실러스 카세이 (*Lactobacillus casei*) BL23), (NP\_786753.1; 락토바실러스 플라타룸 (*Lactobacillus plantarum*) WCFS 1), (ZP\_04009976.1; 락토바실러스 살리바리우스 (*Lactobacillus salivarius*) ATCC 11741), (YP\_818922.1; 루코노스톡 메센테로이데스 아종 메센테로이데스 (*Leuconostoc mesenteroides* subsp. *Mesenteroides*) ATCC 8293), (YP\_794669.1; 락토바실러스 브레비스 (*Lactobacillus brevis*) ATCC 367), (ZP\_04782553.1; 바이셀라 파라메센테로이데스 (*Weissella paramesenteroides*) ATCC 33313), (YP\_001727454.1; 루코노스톡 시트레움 (*Leuconostoc citreum*) KM20), (YP\_819405.1; 루코노스톡 메센테로이데스 아종 메센테로이데스 (*Leuconostoc mesenteroides* subsp. *mesenteroides*) ATCC 8293), (ABX75772.1; 락토코쿠스 락티스 아종 락티스 (*Lactococcus lactis* subsp. *Lactis*)), (YP\_811314.1; 오에노코쿠스 오에니 (*Oenococcus oeni*) PSU-1), (ZP\_02951191.1; 클로스트리듐 부티리킴 (*Clostridium butyricum*) 5521), (ZP\_05390294.1; 클로스트리듐 카복시디보란스 (*Clostridium carboxidivorans*) P7), (NP\_347971.1; 클로스트리듐 아세트부틸리킴 (*Clostridium acetobutylicum*) ATCC 824), (ZP\_03800296.1; 코프로코쿠스 코메스 (*Coprococcus comes*) ATCC 27758), (ZP\_04857624.1; 루미노코쿠스 (*Ruminococcus*) *sp.* 5\_1\_39B FAA), (ZP\_04743029.2; 로세부리아 인테스티날리스 (*Roseburia intestinalis*) L 1-82), (ZP\_02038271.1; 박테로이데스 카필로수스 (*Bacteroides capillosus*) ATCC 29799), (XP\_002180542.1; 패오닥틸럼 트리코무툼 (*Phaeodactylum tricomutum*) CCAP 1055/1), (YP\_568630.1; 로도슈도모나스 팔루스트리스 (*Rhodopseudomonas palustris*) BisB5), (YP\_487462.1; 로도슈도모나스 팔루스트리스 (*Rhodopseudomonas palustris*) HaA2), (NP\_947019.1; 로도슈도모나스 팔루스트리스 CGA009), (YP\_533660.1; 로도슈도모나스 팔루스트리드 BisB18), (YP\_973512.1; 폴라로모나스 나프탈레니보란스 (*Polaromonas naphthalenivorans*) CJ2), (ZP\_01464191.1; 스티그마텔라 아우란티아카 (*Stigmatella aurantiaca*) DW4/3-1), (YP\_001267778.1; 슈도모나스 푸티타 (*Pseudomonas putida*) F1), (YP\_829644.1; 아르트로박터 (*Arthrobacter*) *sp.* FB24), (YP\_002486392.1; 아르트로박터 클로로페놀리쿠스 (*Arthrobacter chlorophenolicus*) A6), (ZP\_05816651.1; 상우이박터 케디에이 (*Sanguibacter keddieii*) DSM 10542), (YP\_002883053.1; 베우텐버지아 카베매 (*Beutenbergia cavemae*) DSM 12333), (YP\_003161540.1; 조네시아 데니트리피칸스 (*Jonesia denitrificans*) DSM 20603), (ZP\_03911482.1; 크

실라니모나스 셀룰로실리티카 (*Xylanimonas cellulositytica*) DSM 15894), (CAJ57850.1; 셀룰로모나스 플라비게나 (*Cellulomonas flavigena*)), (YP\_001134605.1; 미코박테리움 길붐 (*Mycobacterium gilvum*) PYR-GCK), (YP\_953877.1; 미코박테리움 반바알레나이 (*Mycobacterium vanbaalenii*) PYR-1), (YP\_003155611.1; 브래키박테리움 패시움 (*Brachybacterium faecium*) DSM 4810), (YP\_003148127.1; 키토코커스 세덴타리우스 (*Kytococcus sedentarius*) DSM 20547), (YP\_001221168.1; 클라비박터 미키가넨시스 아종 미키가넨시스 (*Clavibacter michiganensis* subsp. *michiganensis*) NCPPB 382), (YP\_001158426.1; 살리니스포라 트로피카 (*Salinispora tropica*) CNB-440), (YP\_001536420.1; 살리니스포라 아레니콜라 (*Salinispora arenicola*) CNS-205), (ZP\_04608302.1; 마이크로모노스포라 (*Micromonospora*) sp. ATCC 39149), (YP\_887914.1; 미코박테리움 스메그마티스 (*Mycobacterium smegmatis*) str. MC2 155), (YP\_639956.1; 미코박테리움 (*Mycobacterium*) sp. MCS), (ZP\_04749157.1; 미코박테리움 칸사시 (*Mycobacterium kansasii*) ATCC 12478), (YP\_001851039.1; 마이코박테리움 마리눔 (*Mycobacterium marinum*) M), (NP\_960507.1; 미코박테리움 아비움 아종 파라투베르쿨로시스 (*Mycobacterium avium* subsp. *paratuberculosis*) K-10), (ZP\_05224330.1; 미코박테리움 인트라셀룰라레 (*Mycobacterium intracellulare*) ATCC 13950), (YP\_001703240.1; 미코박테리움 앵세스 (*Mycobacterium abscessus*)), (ZP\_00995133.1; 자니박터 (*Janibacter*) sp. HTCC2649), (YP\_291026.1; 썬모비피다 푸스카 (*Thermobifida fusca*) YX), (ZP\_04031845.1; 썬모모노스포라 커바타 (*Thermomonospora curvata*) DSM 43183), (ZP\_04475514.1; 스트렙토스포랑기움 로세움 (*Streptosporangium roseum*) DSM 43021), (ZP\_04335641.1; 노카르디움시스 다손빌레이 아종 다손빌레이 (*Nocardiosis dassonvillei* subsp. *dassonvillei*) DSM 43111), (ZP\_04482201.1; 스택케브랜디아 나사우엔시스 (*Stackebrandtia nassauensis*) DSM 44728), (YP\_003099712.1; 액티노시네마 미룸 (*Actinosynnema mirum*) DSM 43827), (NP\_733508.1; 스트렙토마이세스 코엘리콜러 (*Streptomyces coelicolor*) A3(2)), (CAJ88379.1; 스트렙토마이세스 암보팩시엔스 (*Streptomyces ambofaciens*) ATCC 23877), (ZP\_05536883.1; 스트렙토마이세스 그리세오플라부스 (*Streptomyces griseoflavus*) Tu4000), (ZP\_05020421.1; 스트렙토마이세스 스비세우스 (*Streptomyces sviveus*) ATCC 29083), (CBG67625.1; 스트렙토마이세스 스카비에이 (*Streptomyces scabiei*) 87.22), (NP\_822448.1; 스트렙토마이세스 아베르미틸리스 (*Streptomyces avermitilis*) MA-4680), (ZP\_04689547.1; 스트렙토마이세스 가나엔시스 (*Streptomyces ghanaensis*) ATCC 14672), (ZP\_05530021.1; 스크렙토마이세스 비리도크로모게네스 (*Streptomyces viridochromogenes*) DSM 40736), (ZP\_05512501.1; 스트렙토마이세스 히그로스코피커스 (*Streptomyces hygroscopicus*) ATCC 53653), (ZP\_05800927.1; 스트렙토마이세스 플라이오그리세우스 (*Streptomyces flayogriseus*) ATCC 33331), (YP\_001828275.1; 스트렙토마이세스 그리세우스 아종 그리세우스 (*Streptomyces griseus* subsp. *griseus*) NBRC 13350), (ZP\_04705493.1; 스트렙토마이세스 알부스 (*Streptomyces albus*) J1074), (ZP\_04996963.1; 스트렙토마이세스 sp. Mg1), (ZP\_05485309.1; 스트렙토마이세스 sp. SPB78), (ZP\_03860882.1; 크리벨라 플라이다 (*Kribbella flayida*) DSM 17836), (YP\_117539.1; 노카르디아 파르시니카 (*Nocardia farcinica*) IFM 10152), (YP\_001505556.1; 프렌키아 (*Frankia*) sp. EAN1pec), (YP\_482627.1; 프렌키아 sp. Cc13), (YP\_003116893.1; 카테눌리스포라 액시디필라 (*Catenulispora acidiphila*) DSM 44928), (YP\_872280.1; 액시도썬무스 롤리티쿠스 (*Acidothermus lolyticus*) I1B), (YP\_924807.1; 노카르디오이데스 sp. JS614), (YP\_001104157.1; 사카로폴리스포라 에리트라에 (*Saccharopolyspora erythraea*) NRRL 2338), (YP\_002282673.1; 리조븀 레구미노사룸 by. 트리폴리 (*Rhizobium leguminosarum* by. *trifolii*) WSM2304), (YP\_002977256.1; 리조븀 레구미노사룸 by. 트리폴리 (*Rhizobium leguminosarum* by. *trifolii*) WSM1325), (YP\_001979796.1; 리조븀 에틀리 (*Rhizobium etli*) CIAT 652), (YP\_470926.1; 리조븀 에틀리 (*Rhizobium etli*) CFN 42), (YP\_002540633.1; 아그로박테리움 라디오박터 (*Agrobacterium radiobacter*) K84), (ZP\_05182366.1; 브루셀라 (*Bruceella*) sp. 83/13), (ZP\_04683384.1; 오크로박트럼 인터메디움 (*Ochrobactrum intermedium*) LMG 3301), (YP\_001373254.1; 오크로박트럼 안트로피 (*Ochrobactrum anthropi*) ATCC 49188), (YP\_001204109.1; 브라디리조븀 (*radyrhizobium*) sp. ORS278), (YP\_001238418.1; 브라디리조븀 sp. BTAl1), (NP\_769158.1; 브라디리조븀 자포니쿰 (*Bradyrhizobium japonicum*) USDA 110), (YP\_577164.1; 니트로박터 햄버겐시스 (*Nitrobacter hamburgensis*) X14), (YP\_002961612.1; 메틸로박테리움 엑스토르퀼스 (*Methylobacterium extorquens*) AM 1), (YP\_674792.1; 메소리조븀 (*Mesorhizobium*) sp. BNCI), (ZP\_05813617.1; 메소리조븀 오퍼투니스툼 (*Mesorhizobium opportunistum*) WSM2075), (YP\_318559.1; 니트로박터 위노그라드스키 (*Nitrobacter winogradskyi*) Nb-255), (YP\_001755280.1; 메틸로박테리움 라디오톨레란스 (*Methylobacterium radiotolerans*) JCM 2831), (YP\_001753119.1; 메틸로박테리움 라디오톨레란스 JCM 2831), (YP\_003066011.1; 메틸로박테리움 엑스토르퀼스 (*Methylobacterium extorquens*) DM4), (YP\_002964777.1; 메틸로박테리움 엑스토르퀼스 (*Methylobacterium extorquens*) AM 1), (YP\_002501292.1; 메틸로박테리움 노둘란스 (*Methylobacterium nodulans*) ORS 2060), (YP\_002495265.1; 메틸로박테리움 노둘란스 ORS 2060), (YP\_001770387.1; 메틸로박테리움

(*Methylobacterium*) sp.4-46), (YP\_002944712.1; 바리오보락스 파라독수스 (*Variovorax paradoxus*) S110), (ZP\_01156757.1; 오서니콜라 그라눌로수스 (*Oceanicola granulosus*) HTCC2516), (ZP\_01628787.1; 노둘라리아 스포게나 (*Nodularia spumigena*) CCY9414), (YP\_001865546.1; 노스톡 푼크티폼 (*Nostoc punctiforme*) PCC 73102), (YP\_321015.1; 아나바에나 바리아빌리스 (*Anabaena variabilis*) ATCC 29413), (ZP\_03769140.1; 노스톡 아졸레이 (*Nostoc azollae*)의 0708), (NP\_923943.1; 글로에오박터 바이올라세우스 (*Gloeobacter violaceus*) PCC 7421), (YP\_477385.1; 시네코코커스 sp. JA-2-3B'a(2-13)), (YP\_001328659.1; 시노리조비움 메디카 (*Sinorhizobium medicae*) WSM419), (YP\_765670.1; 리조븀 레구미노사룸 bv. 비시아 (*Rhizobium leguminosarum* bv. *viciae*) 3841), (NP\_384212.2; 시노리조비움 멜릴로티 (*Sinorhizobium meliloti*) 1021), (ZP\_02928455.1; 베루코미크로비움 스피노섬 (*Verrucomicrobium spinosum*) DSM 4136), (YP\_001637539.1; 메틸로박테리움 엑스토르퀼스 (*Methylobacterium extorquens*) Pa1), (ZP\_01045825.1; 니트로박터 sp. Nb-311A), (ZP\_02736602.1; 게마타 옵스쿠리글로부스 (*Gemmata obscuriglobus*) UQM 2246), (YP\_003157871.1; 데설포마이크로비움 바쿨라툼 (*Desulfomicrobium baculatum*) DSM 4028), (ZP\_03631304.1; 박테리움 엘린514), (ZP\_04577558.1; 옥살로박터 포르미게네스 (*Oxalobacter formigenes*) HOxBS), (ZP\_04579712.1; 옥살로박터 포르미게네스 *Oxalobacter formigenes*) OXCC13), (YP\_826169.1; 솔리박터 우시타투스 (*Solibacter usitatus*) Ellin6076), (YP\_002018753.1; 펠로딕티온 페오클라트라티포름 (*Pelodictyon phaeoclathratiforme*) BU-1), (YP\_002016285.1; 프로스테코클로리스 에어스투아리이 (*Prosthecochloris aestuarii*) DSM 271), (YP\_001943369.1; 클로로비움 리미콜라 (*Chlorobium limicola*) DSM 245), (NP\_662409.1; 클로로비움 테피둠 (*Chlorobium tepidum*) TLS), (ZP\_01386179.1; 클로로비움 페로옥시단스 (*Chlorobium ferrooxidans*) DSM 13031), (YP\_375422.1; 클로로비움 루테올룸 (*Chlorobium luteolum*) DSM 273), (YP\_285277.1; 데클로로모나스 아로마티카 (*Dechloromonas aromatica*) RCB), (YP\_314589.1; 티오바실러스 데니트리피칸스 (*Thiobacillus denitrificans*) ATCC 25259), (YP\_545002.1; 메틸로바실러스 플라젤라투스 (*Methylobacillus flagellatus*) KT), (NP\_842139.1; 니트로소모나스 유로파에 (*Nitrosomonas europaea*) ATCC 19718), (YP\_748274.1; 니트로소모나스 유포파 (*Nitrosomonas eutropha*) C91), (YP\_411688.1; 니트로소스피라 멀티포르미스 (*Nitrosospora multiformis*) ATCC 25196), (YP\_344700.1; 니트로소코커스 오서니 (*Nitrosococcus oceani*) ATCC 19707), (YP\_007004.1; 칸디다투스 프로토클라미디아 아모에보필라 (*Candidatus Protochlamydia amoebophila*) UWE25), (NP\_435833.1; 시노리조비움 멜릴로티 (*Sinorhizobium meliloti*) 1021), (ZP\_04421874.1; 설푸로스피릴룸 델레이아눔 (*Sulfurospirillum deleyianum*) DSM 6946), (NP\_107054.1; 메소리조븀 로티 (*Mesorhizobium loti*) MAFF303099), (YP\_002289797.1; 올리고트로파 카르복시도보란스 (*Oligotropha carboxidovorans*) OM5), (YP\_001833312.1; 베이에린키아 인디카 아종 인디카 (*Beijerinckia indica* subsp. *indica*) ATCC 9039)를 포함한다.

[0058] 또한, 본 발명의 조성물 및 방법에서 사용가능한 포스포케톨라제로는 본원에 언급된 임의의 포스포케톨라제의 "유도체"로 지칭되는 분자를 포함한다. 이러한 "유도체"는 다음과 같은 특징을 가진다: (1) 본원에 언급된 임의의 포스포케톨라제와 실질적인 상동성을 공유하며, (2) X5P를 글리세르알데하이드 3-포스페이트 (G3P)와 아세틸 포스페이트로; 또는 F6P를 에리트로스 4-포스페이트 (E4P)와 아세틸 포스페이트로 변환하는 반응을 촉매할 수 있음. 유도체의 아미노산 서열이 포스포케톨라제와 80% 이상, 더 바람직하게는 90% 이상, 가장 바람직하게는 95% 이상 동일하다면, 포스포케톨라제의 유도체는 포스포케톨라제와 "실질적인 상동성"을 공유한다고 한다.

[0059] **5.3.2 포스포트랜스아세틸라제 (PTA)**

[0060] 일부 구현예들에서, 본원에 제공되는 유전자 변형된 숙주 세포는 포스포트랜스아세틸라제를 코딩하는 이중의 뉴클레오티드 서열을 포함한다. 포스포트랜스아세틸라제 (EC 2.3.1.8)는 아세틸 포스페이트를 아세틸-CoA로 변환한다.

[0061] 포스포트랜스아세틸라제 활성을 코딩하는 폴리뉴클레오티드, 유전자 및 폴리펩타이드에 대한 다수 예들이 당해 기술 분야에 공지되어 있으며, 본원에 제시된 유전자 변형된 숙주 세포에 사용될 수 있다. 일부 구현예에서, 이러한 폴리뉴클레오티드, 유전자 및/또는 폴리펩타이드는 클로스트리듐 클루이베리 (*Clostridium kluverii*)로부터 유래되는 포스포트랜스아세틸라제이다. 클로스트리듐 클루이베리 (*Clostridium kluverii*)의 대표적인 포스포트랜스아세틸라제 뉴클레오티드 서열은 등재 번호 NC\_009706.1:1428554..1429555와 본원에 제시된 서열번호 3을 포함한다. 클로스트리듐 클루이베리 (*Clostridium kluverii*)의 대표적인 포스포트랜스아세틸라제 단백질 서열은 등재 번호 YP\_001394780과 본원에 제시된 서열번호 4를 포함한다. 다른 이용가능한 포스포트랜스아세틸라제로는, 비제한적으로, 다음과 같은 균주로부터 유래된 것을 포함한다: 락토바실러스 루테리 (*Lactobacillus reuteri*) (NC\_010609.1:460303..461277; YP\_001841389.10); 바실러스 섭틸리스 (*Bacillus subtilis*)

(NC\_014479.1:3671865..3672836; YP\_003868063.1); 메타노사르시나 서모필라 (*Methanosarcina thermophila*) (L23147.1:207..1208; AAA72041.1); 락토바실러스 산프란시센시스 (*Lactobacillus sanfranciscensis*) (BAB19267.1); 락토바실러스 플라타룸 (*Lactobacillus plantarum*) WCFS1 (NP\_784550.1); 락토바실러스 퍼멘텀 (*Lactobacillus fermentum*) ATCC 14931 (ZP\_03944466.1); 바실러스 썩틸리스 아종 썩틸리스 (*Bacillus subtilis subsp. subtilis*) str. 168 (NP\_391646.1); 메타노사르시나 서모필리아 (*Methanosarcina thermophila*) (AAA72041.1); 클로스트리듐 서모셀룸 (*Clostridium thermoCELLUM*) DSM 4150 (ZP\_03152606.1); 클로스트리듐 아세토부틸리쿰 (*Clostridium acetobutylicum*) ATCC 824 (NP\_348368.1); 클로스트리듐 클루이베리 (*Clostridium kluyveri*) DSM 555 (YP\_001394780.1); 베일로넬라 파블라 (*Veillonella parvula*) DSM 2008 (ZP\_03855267.1); 및 살모넬라 엔테리카 아종 엔테리카 혈청형 파라티피 A (*Salmonella enterica subsp. enterica serovar Paratyphi A*) str. ATCC 9150 (YP\_149725.1).

[0062]

다른 이용가능한 포스포트랜스아세틸라제는 국제 출원 WO 2011/15985에 기술된 것을 포함하며, 상기 문헌은 그 전체가 원용에 의해 본 명세서에 포함된다. 이들 포스포트랜스아세틸라제로는 다음과 같은 것을 포함한다: (ZP\_05427766.1; 유박테리움 사페눔 (*Eubacterium saphenum*) ATCC 49989), (ZP\_03627696.1; 박테리움 Ellin514), (ZP\_03131770.1; 크토니오박터 플라부스 (*Chthonio bacter flavus*) Ellin428), (YP\_001878031.1; 아케만시아 무시니필라 (*Akkermansia muciniphila*) TCCBAA- 835), (ZP\_04562924.1; 시트로박터 sp.30\_2), (YP\_001451936.1; 시트로박터 코세리 (*Citrobacter koseri*) ATCC BAA-895), (YP\_149725.1; 살모넬라 엔테리카 아종 엔테리카 혈청형 파라티피 A (*Salmonella enterica subsp. enterica serovar Paratyphi A*) str. ATCC 9150), (YP\_001569496.1; 살모넬라 엔테리카 아종 안조네 혈청형 62 (*Salmonella enterica subsp. anzonae serovar 62*):z4,z23:--), (NP\_416953.1; 에스케리키아 콜라이 (*Escherichia coli*) str. K-12 substr. MG1655), (YP\_002920654.1; 클렙시엘라 뉴모니 (*Klebsiella pneumoniae*) NTUH-K2044), (ZP\_04637797.1; 예르시니아 인터메디아 (*Yersinia intermedia*) ATCC 29909), (ZP\_01222604.1; 포토박테리움 프로퐁둠 (*Photobacterium profundum*) 3TCK), (ZP\_02156855.1; 슈와넬라 벤티카 (*Shewanella benthica*) KT99), (YP\_958508.1; 마리노박터 아쿠에어올레이 (*Marinobacter aquaeolei*) VT8), (YP\_066771.1; 데설포탈레아 사이 크로필라 (*Desulfotalea psychrophila*) Lsv54), (YP\_002780531.1; 로도코커스 오파쿠스 (*Rhodococcus opacus*) B4), (YP\_703506.1; 로도코커스 요스티이 (*Rhodococcus jostii*) RHA1), (ZP\_05479963.1; 스트렙토마이세스 sp. A44), (YP\_002761398.1; 게마키모나스 아우란티아카 (*Gemmatimonas aurantiaca*) T-27), (ZP\_04670189.1; 클로스트리디알레스 박테리움 (*Clostridiales bacterium*) L\_7\_47FAA), (ZP\_05493958.1; 클로스트리듐 파피소솔벤스 (*Clostridium papyrosolvans*) DSM 2782), (YP\_003143506.1; 슬랙키아 헬리오트리니레두센스 (*Slackia heliotrinireducens*) DSM 20476), (ZP\_05090822.1; 루에게리아 (*Ruegeria*) sp. R11), (ZP\_01748021.1; 사기틀라 스텔라타 (*Sagittula stellata*) E-37), (NP\_604069.1; 푸조박테리움 뉴클레아툼 아종 뉴클레아툼 (*Fusobacterium nucleatum subsp. nucleatum*) ATCC 25586), (ZP\_05814734.1; 푸조박테리움 (*Fusobacterium*) sp. 3\_1\_33), (ZP\_06026613.1; 푸조박테리움 페리오돈티쿰 (*Fusobacterium periodonticum*) ATCC 33693), (ZP\_05617632.1; 푸조박테리움 sp. 3\_1\_5R), (ZP\_05628030.1; 푸조박테리움 sp. D12), (ZP\_04860946.1; 푸조박테리움 바눔 (*Fusobacterium vanum*) ATCC 27725), (ZP\_04567444.1; 푸조박테리움 모르티페룸 (*Fusobacterium mortiferum*) ATCC 9817), (YP\_001489437.1; 아르코박터 부트즐레리 (*Arcobacter butzleri*) RM4018), (YP\_003163236.1; 렙토티리키아 부칼리스 (*Leptotrichia buccalis*) C-1013-b), (ZP\_05902420.1; 렙토티리키아 호프스타디이 (*Leptotrichia hofstadii*) F0254), (ZP\_06011308.1; 렙토티리키아 굿펠로우이 (*Leptotrichia goodfellowii*) F0264), (ZP\_04479548.1; 스트렙토바실러스 모닐리포르미스 (*Streptobacillus moniliformis*) DSM 12112), (ZP\_03855267.1; 베일로넬라 파르블라 (*Veillonella parvula*) DSM 2008), (ZP\_03928523.1; 액시다미노코커스 (*Acidaminococcus*) sp. D21), (NP\_970659.1; 트레포네마 덴티콜라 (*Treponema denticola*) ATCC 35405), (ZP\_05621510.1; 트레포네마 빈센티이 (*Treponema vincentii*) ATCC 35580), (NP\_218534.1; 트레포네마 팔리둠 아종 팔리둠 (*Treponema pallidum subsp. pallidum*) str. Nichols), (ZP\_04047318.1; 브라키스피라 무르도키이 (*Brachyspira murdochii*) DSM 12563), (YP\_002720478.1; 브라키스피라 히오디센테리아 (*Brachyspira hyodysenteriae*) WA1), (YP\_001740706.1; 칸디다투스 클로아카모나스 액시다미노보란스 (*Candidatus Cloacamonas acidaminovorans*)), (EER05013.1; 퍼킨수스 마누스 (*Perkinsus mannus*) ATCC 50983), (YP\_945582.1; 보렐리아 터리카티 (*Borrelia turicatae*) 91E135), (YP\_001884013.1; 보렐리아 험시이 (*Borrelia hermsii*) DAH), (YP\_002222233.1; 보렐리아 두토니이 (*Borrelia duttonii*) Ly), (ZP\_03675306.1; 보렐리아 스피엘만니 (*Borrelia spielmanii*) A14S), (ZP\_03435394.1; 보렐리아 아프젤리 (*Borrelia afzelii*) ACA-1), (ZP\_03540018.1; 보렐리아 가리니이 (*Borrelia garinii*) Far04), (ZP\_03672928.1; 보렐리아 발라이시아나 (*Borrelia valaisiana*) VS116), (NP\_212723.1; 보렐리아 버그도르페리 (*Borrelia burgdorferi*))

B31), (YP\_001956287.1; 배양되지 않은 터마이트 그룹 1 박테리움 필로타입 *Rs-D17*), (NP\_975268.1; 미코플라스마 미코이데스 아종 미코이데스 (*Mycoplasma mycoides subsp. mycoidesuncultured SC str. PGI*), (YP\_424216.1; 미코플라스마 카프리컬럼 아종 카프리컬럼 (*Mycoplasma capricolum subsp. capricolum*) ATCC 27343), (YP\_053283.1; 메조플라스마 플로럼 (*Mesoplasma florum*) LI), (CAK99540.1; 스피로플라스마 시트리 (*Spiroplasma citri*)), (NP\_072966.1; 미코플라스마 게니탈리움 (*Mycoplasma genitalium*) G37), (NP\_110116.1; 미코플라스마 뉴모니 (*Mycoplasma pneumomae*) M129), (NP\_853403.1; 미코플라스마 갈리셉티컴 (*Mycoplasma gallisepticum*) R), (NP\_757889.1; 미코플라스마 페네트란스 (*Mycoplasma penetrans*) HF-2), (YP\_116016.1; 미코플라스마 히오뉴모니아 (*Mycoplasma hyopneumoniae*) 232), (YP\_002960607.1; 미코플라스마 컨정티비 (*Mycoplasma conjunctivae*)), (YP\_001256282.1; 미코플라스마 아갈락티이 (*Mycoplasma agalactiae*) PG2), (BAH69503.1; 미코플라스마 퍼멘탄스 (*Mycoplasma fermentans*) PG18), (YP\_278771.1; 미코플라스마 시노비아 (*Mycoplasma synoviae*) 53), (NP\_326068.1; 미코플라스마 풀모니스 (*Mycoplasma pulmonis*) UAB CTIP), (YP\_015865.1; 미코플라스마 모빌 (*Mycoplasma mobile*) 163K), (YP\_001256630.1; 미코플라스마 아갈락티이 (*Mycoplasma agalactiae*) PG2), (YP\_802685.1; 부채라 아피디콜라 (*Buchnera aphidicola*) str. Cc (*Cinara cedri*)), (YP\_001885432.1; 클로스트리디움 보툴리눔 (*Clostridium botulinum*) B str. Eklund 17B), (YP\_001308302.1; 클로스트리디움 베이예린키이 (*Clostridium beijerinckii*) NCIMB 8052), (ZP\_05131280.1; 클로스트리디움 (*Clostridium*) sp. 7\_2\_43FAA), (ZP\_02948604.1; 클로스트리디움 부티리컴 (*Clostridium butyricum*) 5521), (NP\_562641.1; 클로스트리디움 퍼프린겐스 (*Clostridium perfringens*) str. 13), (ZP\_05391232.1; 클로스트리디움 카복시디보란스 (*Clostridium carboxidivorans*) P7), (YP\_001394780.1; 클로스트리디움 클루이베리 (*Clostridium kluyveri*) DSM 555), (ZP\_02995419.1; 클로스트리디움 스포로게네스 (*Clostridium sporogenes*) ATCC 15579), (NP\_781870.1; 클로스트리디움 테타니 (*Clostridium tetani*) E88), (ZP\_04862192.1; 클로스트리디움 보툴리눔 (*Clostridium botulinum*) D str. 1873), (YP\_878298.1; 클로스트리디움 노비이 (*Clostridium novyi*) NT), (ZP\_04804960.1; 클로스트리디움 셀룰로보란스 (*Clostridium cellulovorans*) 743B), (NP\_348368.1; 클로스트리디움 아세토부틸리컴 (*Clostridium acetobutylicum*) ATCC 824), (ACA51668.1; 서모안에어로 박테리움 사카롤라이티컴 (*Thermoanaero bacterium saccharolyticum*)), (ZP\_05336886.1; 서모안에어로박테리움 서모사카롤라이티컴 (*Thermoanaero bacterium thermosaccharolyticum*) SM 571), (NP\_623097.1; 서모안에어로박터 텡콘젠시스 (*Thermoanaero bacter tengcongensis*) MB4), (YP\_001663354.1; 서모안에어로박터 sp. X514), (YP\_002508771.1; 할로서모트릭스 오레니이 (*Halothermoth rix orenii*) H 168), (YP\_003190679.1; 테셀포토마쿨럼 아세톡시단스 (*Desulfotomaculum acetoxidans*) DSM 771), (YP\_001917776.1; 나트란에어로비우스 서모필스 (*Natranaerobius thermophiles*) JWIM-WN-LF), (YP\_360288.1; 카르복시도서무스 하이드로게노포르만스 (*Carboxydotherrmus hydrogenoformans*) Z-2901), (EY83551.1; 박테로이데스 sp.2\_1\_33B), (ZP\_02033408.1; 파라박테로이데스 머디 (*Parabacteroides merdae*) ATCC 43184), (NP\_905297.1; 포르피로모나스 강기발리스 (*Porphyromonas gingivalis*) W83), (ZP\_04056000.1; 포르피로모나스 우에노니스 (*Porphyromonas uenonis*) 60-3), (ZP\_04389884.1; 포르피로모나스 엔도돈탈리스 (*Porphyromonas endodontalis*) ATCC 35406), (ZP\_02068815.1; 박테로이데스 유니포르미스 (*Bacteroides uniformis*) ATCC 8492), (ZP\_03460749.1; 박테로이데스 에거티이 (*Bacteroides eggerthii*) DSM 20697), (ZP\_03676944.1; 박테로이데스 셀룰로실리티쿠스 (*Bacteroides cellulosilyticus*) DSM 14838), (YP\_097761.1; 박테로이데스 프라길리스 (*Bacteroides fragilis*) YCH46), (ZP\_04545825.1; 박테로이데스 sp. DI), (ZP\_03643544.1; 박테로이데스 코프로필러스 (*Bacteroides coprophilus*) DSM 18228), (ZP\_03207078.1; 박테로이데스 플레바이우스 (*Bacteroides plebeius*) DSM 17135), (YP\_001297855.1; 박테로이데스 볼라쿠스 (*Bacteroides vulgatus*) ATCC 8482), (ZP\_05736702.1; 프레보텔라 타너리 (*Prevotella tanneriae*) ATCC 51259), (ZP\_06007587.1; 프레보텔라 버젠시스 (*Prevotella bergensis*) DSM 17361), (ZP\_05858935.1; 프레보텔라 베로랄리스 (*Prevotella veroralis*) F0319), (ZP\_05916997.1; 프레보텔라 sp. 오랄 탁손 (*Prevotella sp. oral taxon*) 472 str. F0295), (YP\_002308782.1; 칸디다투스 아조 박테로이데스 슈도트리콘 임피 게노모바르 (*Candidatus Azo bacteroides pseudotrichon ymphae genomovar*). CFP2), (YP\_753459.1; 신티트로포모나스 울페이 아종 str. 고이팅겐 (*Syntrophomonas wolfei subsp. wolfei str. Goettingen*)), (ZP\_01771389.1; 콜린셀라 에어로팍시엔스 (*Collinsella aerofaciens*) ATCC 25986), (ZP\_03296849.1; 콜린셀라 스테코리스 (*Collinsella stercoris*) DSM 13279), (ZP\_04445308.1; 콜린셀라 엔테스티날리스 (*Collinsella ntestinalis*) DSM 13280), (ZP\_03567515.1; 아토포비움 리미 (*Atopobium rimaie*) ATCC 49626), (YP\_003179667.1; 아토포비움 파블럼 (*Atopobium parvulum*) DSM 20469), (ZP\_03946133.1; 아토포비움 바기니 (*Atopobium vaginae*) DSM 15829), (ZP\_03990654.1; 오리박테리움 시누스 (*Oribacterium sinus*) F0268), (ZP\_04450849.1; 아비오토르피아 데펙티브 (*Abiotrophia defectiva*) ATCC 49176), (ZP\_05797601.1; 오리박테리움 sp. 오랄 탁손 (*Oribacterium sp. oral taxon*) 078 str. F0262), (ZP\_03730247.1; 클로스트리디움 sp.

M62/1), (ZP\_04856252.1; 루미노코커스 (*Ruminococcus*) sp. 5\_1\_39BFAA), (ZP\_01966332.1; 루미노코커스 오베움 (*Ruminococcus obeum*) ATCC 29174), (ZP\_05345616.1; 브리안텔라 포르마텍시겐스 (*Bryantella formatexigens*) DSM 14469), (ZP\_03780829.1; 블라우티아 하이드로게노트로 피카 (*Blautia hydrogenotrophica*) DSM 10507), (ZP\_03289360.1; 클로스트리듐 넥실 (*Clostridium nexile*) DSM 1787), (ZP\_02042092.1; 루미노코커스 그나부스 (*Ruminococcus gnavus*) ATCC 29149), (ZP\_03168112.1; 루미노코커스 락타리스 (*Ruminococcus lactaris*) ATCC 29176), (ZP\_01968837.1; 루미노코커스 토르퀘스 (*Ruminococcus torques*) ATCC 27756), (ZP\_02430426.1; 클로스트리듐 신덴스 (*Clostridium scindens*) ATCC 35704), (ZP\_03779744.1; 클로스트리듐 힐레모니 (*Clostridium hylemonae*) DSM 15053), (ZP\_02234595.1; 도레아 포르미시게네란스 (*Dorea formicigenerans*) ATCC 27755), (ZP\_01994673.1; 도레아 롱기카테나 (*Dorea longicatena*) DSM 13814), (YP\_001558442.1; 클로스트리듐 피토퍼멘탄스 (*Clostridium phytofermentans*) ISDg), (ZP\_04667085.1; 클로스트리디알레스 박테리움 (*Clostridiales bacterium*) 1\_7\_47FAA), (ZP\_02085391.1; 클로스트리듐 볼테이 (*Clostridium bolteae*) ATCC BAA-613), (ZP\_05790853.1; 부티리비브리오 크로스투스 (*Butyrivibrio crossotus*) DSM 2876), (ZP\_02026034.1; 유박테리움 벤트리오섬 (*Eubacterium ventriosum*) ATCC 27560), (YP\_002930513.1; 유박테리움 엘리겐스 (*Eubacterium eligens*) ATCC 27750), (ZP\_04808213.1; 헬리코박터 풀로럼 (*Helicobacter pullorum*) MIT 98-5489), (ZP\_03656120.1; 헬리코박터 캐나다렌시스 (*Helicobacter Canadensis*) MIT 98-5491), (ZP\_04583217.1; 헬리코박터 윙하멘시스 (*Helicobacter winhamensis*) ATCCBAA-430), (NP\_860840.1; 헬리코박터 헤바티쿠스 (*Helicobacter hepaticus*) ATCC 51449), (ZP\_03657896.1; 헬리코박터 시나디 (*Helicobacter cinaedi*) CCUG 18818), (ZP\_02417779.1; 안에어로스티페스 카키 (*Anaerostipes caccae*) DSM 14662), (ZP\_02437622.1; 클로스트리듐 sp. SS211), (ZP\_02205430.1; 코프로코커스 유타투스 (*Coprococcus eutactus*) ATCC 27759), (ZP\_02692616.1; 에풀로피시움 sp. N.t 모르포타입 B' (*Epulopiscium* sp. 'N.t. morphotype B')), (YP\_003182082.1; 에거텔라 렌타 (*Eggerthella lenta*) DSM 2243), (YP\_003151027.1; 크립토박테리움 커툼 (*Cryptobacterium curtum*) DSM 15641), (YP\_003143601.1; 슬랙키아 헬리오트리니레두센스 (*Slackia heliotrinireducens*) DSM 20476), (ZP\_05498135.1; 클로스트리듐 파피로솔벤스 (*Clostridium papyrosolvens*) DSM 2782), (ZP\_03152606.1; 클로스트리듐 서모셀럼 (*Clostridium thermocellum*) JW20), (YP\_001180817.1; 칼디셀룰로시럽터 사카롤라이티쿠스 (*Caldicellulosiruptor saccharolyticus*) DSM 8903), (AAA72041.1; 메타노사르시나 서모필라 (*Methanosarcina thermophila*), (NP\_618482.1; 메타노사르시나 아세티보란스 (*Methanosarcina acetivorans*) C24), (YP\_305342.1; 메타노사르시나 바르케리 str. 푸사로 (*Methanosarcina barkeri* str. Fusaro)), (ZP\_02142278.1; 로세박터 리토랄리스 (*Roseobacter litoralis*) Och 149), (YP\_681184.1; 로세오박터 데니트리피칸스 (*Roseobacter denitrificans*) Och 114), (YP\_001533168.1; 디노로세오 박터 시비 (*Dinoroseo bacter shibae*) DFL 12), (ZP\_05124935.1; 로도박테레이시이 박테리움 (*Rhodobacteraceae bacterium*) KLH11), (ZP\_05786337.1; 실리시박터 라쿠스카롤렌시스 (*Silicibacter lacuscaerulensis*) ITI-1157), (YP\_001313586.1; 시노리조비움 메디키 (*Sinorhizobium medicae*) WSM419), (NP\_437512.1; 시노리조비움 멜리로티 (*Sinorhizobium meliloti*) 1021), (ZP\_04682129.1; 오크로박트럼 인터메디움 (*Ochrobactrum intermedium*) LMG 3301), (YP\_001372036.1; 오크로박트럼 안트로픽 (*Ochrobactrum anthropic*) ATCC 49188), (YP\_001888115.1; 버크홀데리아 피토퍼만스 (*Burkholderia phytofirmans*) PsJN), (YP\_554613.1; 버크홀데리아 제노보란스 (*Burkholderia xenovorans*) LB400), (YP\_001862297.1; 버크홀데리아 피마툼 (*Burkholderia phymatum*) STM815), (YP\_297974.1; 랄스토니아 유티트로파 (*Ralstonia eutropha*) JMP134), (YP\_002008219.1; 쿠프리아비두스 타이완렌시스 (*Cupriavidus taiwanensis*)), (YP\_001584488.1; 버크홀데리아 멀티보란스 (*Burkholderia multivorans*), (YP\_002233797.1; 버크홀데리아 세노세팍시아 (*Burkholderia cenocepacia*) J2315), (ZP\_01220235.1; 포토박테리움 프로푼덤 (*Photobacterium profundum*) 3TCK), (ZP\_03698361.1; 루티엘라 니트로페럼 (*Lutiella nitroferum*) 2002), (ZP\_01811515.1; 비브리오날리스 박테리움 (*Vibrionales bacterium*) SWAT-3), (ZP\_00988349.1; 비브리오 스피렌디두스 (*Vibrio splendidus*) 12B01), (ZP\_01866234.1; 비브리오 실로니이 (*Vibrio shilonii*) AK1), (ZP\_05885163.1; 비브리오 코랄리아라이티쿠스 (*Vibrio coralliilyticus*) ATCCBAA-450), (AAS78789.1; 파라코커스 데니트리피칸스 (*Paracoccus denitrificans*)), (YP\_345196.1; 로도박터 스페어로이데스 (*Rhodobacter sphaeroides*) 2.4.1), (AAN08490.1; 카스텔라니엘라 데프라그란스 (*Castellaniella defragrans*)), (ZP\_00961345.1; 로세오바리우스 누빈히벤스 (*Roseovarius nubinhibens*) ISM), (YP\_168755.1; 루에게리아 포메로이 (*Ruegeria pomeroyi*) DSS-3), (ZP\_01901193.1; 로세오박터 sp. AzwK-3b), (ZP\_01752570.1; 로세오박터 sp. SK209-2-6), (ZP\_02140073.1; 로세오박터 리토랄리스 (*Roseobacter litoralis*) Och 149), (YP\_510789.1; 야나시아 (*Jannaschia*) sp. CCS1), (ZP\_05073153.1; 로도박테라이 에스 박테리움 (*Rhodobacterial es bacterium*) HTCC2083), (YP\_822367.1; 칸디다투스 솔리박터 우시타투스 (*Candidatus Solibacter usitatus*) Ellin6076),

(ZP\_01313101.1; 테셀푸로몬 에즈 아세톡시단스 (*Desulfuromon as acetoxidans*) DSM 684), (YP\_357950.1; 펠로박터 카르비놀리쿠스 (*Pelobacter carbinolicus*) DSM 2380), (YP\_002537084.1; 지오박터 (*Geobacter*) sp. FRC-32), (YP\_001232124.1; 지오박터 우라니이레두센스 (*Geobacter uraniireducens*) Rf4), (NP\_953751.1; 지오박터 설푸레두센스 (*Geobacter sulfurreducens*) PCA), (YP\_384000.1; 지오박터 메탈리레두센스 (*Geobacter metallireducens*) GS-15), (YP\_900968.1; 펠로박터 프로피오니쿠스 (*Pelobacter propionicus*) DSM 2379), (YP\_001951452.1; 지오박터 로블레이 (*Geobacter lovleyi*) SZ), (ZP\_05311922.1; 지오박터 sp. M18), (YP\_003021758.1; 지오박터 sp. M21), (YP\_358255.1; 펠로박터 카르비놀리쿠스 (*Pelobacter carbinolicus*) DSM 2380), (ZP\_03906856.1; 데니트로비브리오 아세티필러스 (*Denitrovibrio acetiphilus*) DSM 12809), (YP\_001997093.1; 클로로헤르페톤 칼라시움 (*Chloroherpeton thalassium*) ATCC 35110), (ZP\_01924858.1; 빅티발리스 바덴시스 (*Victivallis vadensis*) ATCCBAA-548), (ZP\_03439825.1; 헬리코박터 필로리 (*Helicobacter pylori*) 98-10), (YP\_003057614.1; 헬리코박터 필로리 B38), (YP\_001910417.1; 헬리코박터 필로리 Shi470), (NP\_223559.1; 헬리코박터 필로리 J99), (YP\_665033.1; 헬리코박터 액시노니키스 str. 시바 (*Helicobacter acinonychis* str. Sheeba)), (ZP\_01810337.1; 캄필로박터 제주니 아종 제주니 (*Campylobacter jejuni* subsp. *jejuni*) CG8486), (ZP\_00366840.1; 캄필로박터 콜라이 (*Campylobacter coli*) RM2228), (ZP\_00370527.1; 캄필로박터 옵살리엔시스 (*Campylobacter upsaliensis*) RM3195), (YP\_002575219.1; 캄필로박터 라리 (*Campylobacter lari*) RM2100), (YP\_001406718.1; 캄필로박터 호미니스 (*Campylobacter hominis*) ATCCBAA- 381), (ZP\_05624820.1; 캄필로박터 그라실리스 (*Campylobacter gracilis*) RM3268), (YP\_891988.1; 캄필로박터 페투스 아종 페투스 (*Campylobacter fetus* subsp. *fetus*) 82-40), (YP\_001466901.1; 캄필로박터 콘시수스 (*Campylobacter concisus*) 13826), (YP\_001408221.1; 캄필로박터 커부스 (*Campylobacter curvus*) 525.92), (ZP\_05363348.1; 캄필로박터 쇼애 (*Campylobacter showaeshowae* RM3277), (ZP\_03742933.1; 비피도박테리움 슈도카테놀라툼 (*Bifidobacterium pseudocatenulatum*) DSM 20438), (ZP\_02918887.1; 비피도박테리움 덴티움 (*Bifidobacterium dentium*) ATCC 27678), (ZP\_02028883.1; 비피도박테리움 아돌레센티스 (*Bifidobacterium adolescentis*) L2-32), (ZP\_04448100.1; 비피도박테리움 안굴라툼 (*Bifidobacterium angulatum*) DSM 20098), (ZP\_03618886.1; 비피도박테리움 브레베 (*Bifidobacterium breve*) DSM 20213), (ZP\_03976084.1; 비피도박테리움 롱검 아종 인판티스 (*Bifidobacterium longum* subsp. *infantis*) ATCC 55813), (YP\_002323183.1; 비피도박테리움 롱검 아종 인판티스 ATCC 15697), (ZP\_03646187.1; 비피도박테리움 비피덤 (*Bifidobacterium bifidum*) NCIMB 41171), (ZP\_03937611.1; 가드네렐라 바기날리스 (*Gardnerella vaginalis*) ATCC 14019), (ZP\_02962869.1; 비피도박테리움 애니말리스 아종 락티스 (*Bifidobacterium animalis* subsp. *lactis*) HNO19), (ZP\_05965185.1; 비피도박테리움 갈리컴 (*Bifidobacterium gallicum*) DSM 20093), (ZP\_02043408.1; 액티노마이세스 오돈톨리티쿠스 (*Actinomyces odontolyticus*) ATCC 17982), (ZP\_03925176.1; 액티노마이세스 콜레오카니스 (*Actinomyces coleocanis*) DSM 15436), (NP\_601948.1; 코리네박테리움 글루타미컴 (*Corynebacterium glutamicum*) ATCC 13032), (NP\_739201.1; 코리네박테리움 이피시엔스 (*Corynebacterium efficiens*) YS-314), (NP\_940379.1; 코리네박테리움 디프테리아 (*Corynebacterium diphtheria*) NCTC 13129), (ZP\_04835255.1; 코리네박테리움 글루쿠로놀리티컴 (*Corynebacterium glucuronolyticum*) ATCC 51867), (ZP\_05708623.1; 코리네박테리움 게니탈리움 (*Corynebacterium genitalium*) ATCC 33030), (ZP\_03977910.1; 코리네박테리움 리포필로플라븀 (*Corynebacterium lipophiloflavum*) DSM 44291), (ZP\_03932064.1; 코리네박테리움 아콜렌스 (*Corynebacterium accolens*) ATCC 49725), (ZP\_05366890.1; 코리네박테리움 투베르쿨로스테아리컴 (*Corynebacterium tuberculostearicum*) SK141), (YP\_002835817.1; 코리네박테리움 아우누코숨 (*Corynebacterium aurumucosum*) ATCC 700975), (YP\_250020.1; 코리네박테리움 예이케이움 (*Corynebacterium jeikeium*) K411), (YP\_001801132.1; 코리네박테리움 우레알리티컴 (*Corynebacterium urealyticum*) DSM 7109), (YP\_002906954.1; 코리네박테리움 크로펜스테티이 (*Corynebacterium kroppenstedtii*) DSM 44385), (ZP\_03393297.1; 코리네박테리움 아미콜라툼 (*Corynebacterium amycolatum*) SK46), (ZP\_03718987.1; 네이세리아 플라베센스 (*Neisseria flavescens*) NRL30031/H 210), (ZP\_05318956.1; 네이세리아 시카 (*Neisseria sicca*) ATCC 29256), (YP\_001598731.1; 네이세리아 메닝기티데스 (*Neisseria meningitides*) 053442), (ZP\_04602977.1; 킹엘라 오랄리스 (*Kingella oralis*) ATCC 51147), (YP\_426466.1; 로도스피릴룸 루브럼 (*Rhodospirillum rubrum*) ATCC 11170), (NP\_871183.1; 글로시나 브레비팔피스의 비글레스워티아 글로시니디아 엔도심비온트 (*Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis*)), (NP\_777793.1; 부크네라 아피디콜라 str. Bp (*Buchnera aphidicola* str. Bp (바이즌지아 피스타시아 (*Baizongia pistaciae*))), (YP\_003249406.1; 피브로박터 숙모게네스 아종 숙모게네스 (*Fibrobacter succmogenes* subsp. *succmogenes*) S85), (ZP\_03535302.1; 미코박테리움 투베르쿨로시스 (*Mycobacterium tuberculosis*) T17), (ZP\_04056438.1; 카프노시토파가 킹기발리스 (*Capnocytophaga gingivalis*) ATCC 33624), (YP\_003108500.1; 칸디다투스 술시아 무엘레리

(*Candidatus Sulcia muelleri*) SMDSEM), (P77844; 코리네박테리움 글루타미컴 (*Corynebacterium glutamicum*)), (ZP\_03994160.1; 모빌런쿠스 물리에리스 (*Mobiluncus mulieris*) ATCC 35243), (ZP\_03922640.1; 모빌런쿠스 커티시 (*Mobiluncus curtisii*) ATCC 43063), (ZP\_03716209.1; 유박테리움 할리 (*Eubacterium hallii*) DSM 3353), (ZP\_03718143.1; 유박테리움 할리 (*DSM 3353*), (ZP\_05614434.1; 피칼리박테리움 프라우스니치 (*Faecalibacterium prausnitzii*) A2-165), (ZP\_02034852.1; 박테로이데스 카필로수스 (*Bacteroides capillosus*) ATCC 29799), (ZP\_03753543.1; 로세부리아 이눌리니보란스 (*Roseburia inulinivorans*) DSM 16841), (ZP\_04745275.2; 로세부리아 인테스티날리스 (*Roseburia intestinalis*) L1-82), (YP\_002937332.1; 유박테리움 렉탈 (*Eubacterium rectale*) ATCC 33656), (ZP\_02074244.1; 클로스트리듐 sp. L2-50), (ZP\_04455374.1; 셔클레워티아 사텔리스 (*Shuttleworthia satelles*) DSM 14600), (ZP\_03488480.1; 유박테리움 비포메 (*Eubacterium bifforme*) DSM 3989), (ZP\_02078327.1; 유박테리움 돌리컴 (*Eubacterium dolichum*) DSM 3991), (ZP\_02077559.1; 유박테리움 돌리컴 (*DSM 3991*), (ZP\_03305532.1; 안에어로코쿠스 하이드로게날리스 (*Anaerococcus hydrogenalis*) DSM 7454), (ZP\_05473291.1; 안에어로코쿠스 바기날리스 (*Anaerococcus vaginalis*) ATCC 51170), (ZP\_03931050.1; 안에어로코쿠스 테트라디우스 (*Anaerococcus tetradius*) ATCC 35098), (YP\_003153463.1; 안에어로코쿠스 프레보티 (*Anaerococcus prevotii*) DSM 20548), (ZP\_03916048.1; 안에어로코쿠스 락토라이티쿠스 (*Anaerococcus lactolyticus*) ATCC 51172), (NP\_607213.1; 스트렙토코쿠스 피오게네스 (*Streptococcus pyogenes*) MGAS8232), (AAK34003.1; 스트렙토코쿠스 피오게네스 (*Streptococcus pyogenes*) MIGAS), (YP\_002562185.1; 스트렙토코쿠스 우베리스 (*Streptococcus uberis*) 01401), (YP\_002744451.1; 스트렙토코쿠스 에퀴 아종 주에피데미쿠스 (*Streptococcus equi* subsp. *Zooepidemicus*)), (BAH88016.1; 스트렙토코쿠스 무탄스 (*Streptococcus mutans*) MN2025), (ZP\_02920305.1; 스트렙토코쿠스 인판타리우스 아종 인판타리우스 (*Streptococcus infantarius* subsp. *infantarius*) ATCCBAA-102), (YP\_329798.1; 스트렙토코쿠스 아갈락티 (*Streptococcus agalactiae*) A909), (ZP\_04061789.1; 스트렙토코쿠스 살리바리우스 (*Streptococcus salivarius*) SK126), (YP\_139881.1; 스트렙토코쿠스 서모필스 (*Streptococcus thermophiles* LMG 18311)), (ZP\_04525024.1; 스트렙토코쿠스 뉴모미 (*Streptococcus pneumoniae*) CCRI 1974), (ZP\_06060573.1; 스트렙토코쿠스 sp. 2\_1\_36FAA), (YP\_001198423.1; 스트렙토코쿠스 수이스 (*Streptococcus suis*) 05ZYH33), (NP\_964739.1; 락토바실러스 존소니 (*Lactobacillus johnsonii*) NCC 533), (YP\_193610.1; 락토바실러스 액시도필러스 (*Lactobacillus acidophilus*) NCFM), (ZP\_04011019.1; 락토바실러스 울투넨시스 (*Lactobacillus ultunensis*) DSM 16047), (ZP\_03995297.1; 락토바실러스 크리스파투스 (*Lactobacillus crispatus*) JV- VO1), (ZP\_05752753.1; 락토바실러스 헬베티쿠스 (*Lactobacillus helveticus*) DSM 20075), (ZP\_03956024.1; 락토바실러스 옌세니 (*Lactobacillus jensenii*) JV-V16), (ZP\_04645187.1; 락토바실러스 옌세니 (*Lactobacillus jensenii*) 269-3), (YP\_618719.1; 락토바실러스 델브루이키 (*Lactobacillus delbrueckii* subsp. *bulgaricus*) ATCC 11842), (ZP\_05744366.1; 락토바실러스 이너스 (*Lactobacillus iners*) DSM 13335), (NP\_391646.1; 바실러스 셉틸리스 아종 셉틸리스 (*Bacillus subtilis* subsp. *subtilis*) str. 168), (YP\_001423045.1; 바실러스 아밀롤리퀘팩시엔스 (*Bacillus amyloliquefaciens*) FZB42), (YP\_081073.1; 바실러스 리케니포르미스 (*Bacillus licheniformis*) ATCC 14580), (ZP\_03055101.1; 바실러스 푸밀러스 (*Bacillus pumilus*) ATCC 7061), (YP\_002317098.1; 애녹시바실러스 플라비서무스 (*Anoxybacillus flavithermus*) WKI), (YP\_002951270.1; 지오바실러스 sp. WCH70), (YP\_001127443.1; 지오바실러스 서모테니트리피칸스 (*Geobacillus thermodenitrificans*) NG80-2), (YP\_149268.1; 지오바실러스 카우스토피러스 (*Geobacillus kaustophilus*) HTA426), (ZP\_01861251.1; 바실러스 (*Bacillus*) sp. SG-1), (ZP\_03228176.1; 바실러스 코아후일렌시스 (*Bacillus coahuilensis*) m4-4), (ZP\_01173945.1; 바실러스 sp. NRRLB-14911), (NP\_693944.1; 오셔노바실러스 이헤이엔시스 (*Oceanobacillus iheyensis*) HTE831), (ZP\_04314753.1; 바실러스 세레우스 (*Bacillus cereus*) BGSC 6E1), (YP\_014727.1; 리스테리아 모노사이토젠 에스 str. (*Listeria monocytogenes* str.) 4b F2365), (ZP\_04443757.1; 리스테리아 그라이 (*Listeria grayi*) DSM 20601), (NP\_244690.1; 바실러스 할로두란스 (*Bacillus halodurans*) C-125), (YP\_177402.1; 바실러스 클라우시 (*Bacillus clausii*) KSM-K16), (YP\_002885816.1; 엑시구오 박테리움 (*Exiguobacterium*) sp. AT1b), (YP\_001812721.1; 엑시구오박테리움 시비리컴 (*Exiguobacterium sibiricum*) 255-15), (ZP\_02169346.1; 바실러스 셀레니티레듀센스 (*Bacillus selenitireducens*) MLS10), (ZP\_04818386.1; 스태필로코쿠스 에피데미디스 (*Staphylococcus epidermidis*) M23864:W1), (ZP\_03612973.1; 스태필로코쿠스 카비티스 (*Staphylococcus capitis*) SK14), (ZP\_04677798.1; 스태필로코쿠스 와이메리 (*Staphylococcus wameryi*) L37603), (NP\_763914.1; 스태필로코쿠스 에피데미디스 (*Staphylococcus epidermidis*) ATCC 12228), (ZP\_05685678.1; 스태필로코쿠스 아우레우스 (*Staphylococcus aureus*) A9635), (YP\_254319.1; 스태필로코쿠스 헤몰리티쿠스 (*Staphylococcus haemolyticus*) JCSC1435), (ZP\_04059818.1; 스태필로코쿠스 호미니스 (*Staphylococcus hominis*) SK119),

(ABR57177.1; 스타필로코커스 크실로수스 (*Staphylococcus xylosus*)), (YP\_302214.1; 스타필로코커스 사프로피티쿠스 아종 사프로피티쿠스 (*Staphylococcus saprophyticus* subsp. *saprophyticus*) ATCC 15305), (YP\_002633340.1; 스타필로코커스 카모수스 아종 카모수스 (*Staphylococcus camosus* subsp. *camosus*) TM300), (YP\_002561236.1; 마크로코커스 카세오라이티쿠스 (*Macrococcus caseolyticus*) JCSC5402), (ZP\_03944466.1; 락토바실러스 퍼멘텀 (*Lactobacillus fermentum*) ATCC 14931), (ZP\_05553502.1; 락토바실러스 콜레오호미니스 (*Lactobacillus coleohominis*) 101-4-CHN), (ZP\_03959629.1; 락토바실러스 바기날리스 (*Lactobacillus vaginalis*) ATCC 49540), (YP\_001271004.1; 락토바실러스 루테리 (*Lactobacillus reuteri*) DSM 20016), (ZP\_05745668.1; 락토바실러스 안트리 (*Lactobacillus antri*) DSM 16041), (YP\_818931.1; 루코노스톡 메센테로이데스 아종 메센테로이데스 (*Leuconostoc mesenteroides*) subsp. *mesenteroides* ATCC 8293), (YP\_001727831.1; 루코노스톡 시트레움 (*Leuconostoc citreum*) KM20), (ZP\_04782044.1; 바이셀라 파라메센테로이데스 (*Weissella paramesenteroides*) ATCC 33313), (ZP\_01544468.1; 오에노코커스 오에니 (*Oenococcus oeni*) ATCC BAA-1163), (ZP\_05737294.1; 그라눌리카텔라 아디아센스 (*Granulicatella adiacens*) ATCC 49175), (ZP\_05851915.1; 그라눌리카텔라 엘레간스 (*Granulicatella elegans*) ATCC 700633), (ZP\_02183965.1; 카모박테리움 sp. AT7), (ZP\_05649755.1; 엔테로코커스 갈리나룸 (*Enterococcus gallinarum*) EG2), (ZP\_03947918.1; 엔테로코커스 패칼리스 (*Enterococcus faecalis*) TX0104), (ZP\_03982224.1; 엔테로코커스 패슘 (*Enterococcus faecium*) TX1330), (YP\_395954.1; 락토바실러스 사케이 아종 사케이 (*Lactobacillus sakei* subsp. *sakei*) 23K), (ZP\_04449762.1; 카토넬라 모르비 (*Catonella morbi*) ATCC 51271), (YP\_001032100.1; 락토코커스 락티스 아종 크레몬스 (*Lactococcus lactis* subsp. *cremons*) MG1363), (YP\_806234.1; 락토바실러스 카세이 (*Lactobacillus casei*) ATCC 334), (NP\_784550.1; 락토바실러스 플라타룸 (*Lactobacillus plantarum*) WCFS1), (YP\_794848.1; 락토바실러스 브레비스 (*Lactobacillus brevis*) ATCC 367), (ZP\_03954831.1; 락토바실러스 힐가르디 (*Lactobacillus hilgardii*) ATCC 8290), (BABI9267.1; 락토바실러스 산프란시센시스 (*Lactobacillus sanfranciscensis*)), (ZP\_03958288.1; 락토바실러스 루미니스 (*Lactobacillus ruminis*) ATCC 25644), (YP\_536042.1; 락토바실러스 살리바리우스 (*Lactobacillus salivarius*) UCC118), (ZP\_05747635.1; 에리시펠로트릭스 루시오파티 (*Erysipelothrix rhusiopathiae*) ATCC 19414), (YP\_803875.1; 페디오코커스 펜토사세우스 (*Pediococcus pentosaceus*) ATCC 25745), (ZP\_02093784.1; 파르비모나스 마이크로 (*Parvimonas micra*) ATCC 33270), (YP\_001692923.1; 피네골디아 마그나 (*Fingoldia magna*) ATCC 29328), (ZP\_04431499.1; 바실러스 코아굴란스 (*Bacillus coagulans*) 36DJ), (ZP\_04775813.1; 게멜라 헤몰리산스 (*Gemella haemolysans*) ATCC 10379), (YP\_001360609.1; 키네오코커스 라디오톨레란스 (*Kineococcus radiotolerans*) SRS30216), (ZP\_01115869.1; 레이네키아 블란덴시스 (*Reinekea blandensis*) MED297), (YP\_003074238.1; 테레디니박 턴테리 (*Teredinibac turnterrae*) T7901), (YP\_958411.1; 마리노박터 퀴올레이 (*Marinobacter quaeolei*) VT8), (YP\_435580.1; 하헬라 케유엔시스 (*Hahella chejuensis*) KCTC 2396), (YP\_001189125.1; 슈도모나스 멘도시나 (*Pseudomonas mendocina*) ymp), (YP\_792443.1; 슈도모나스 메우로그모사 (*Pseudomonas aerugmosa*) UCBPP-PA14), (NP\_791001.1; 슈도모나스 신기 pv. 토마토 (*Pseudomonas synngae* pv. *tomato*) str. DC3000), (YP\_258069.1; 슈도모나스 플루오레센스 (*Pseudomonas fluorescens*) Pf-5), (YP\_606637.1; 슈도모나스 엔토모필라 (*Pseudomonas entomophila*) L48), (YP\_002800579.1; 아조토박터 비넬란디 (*Azotobacter vinelandii*) DJ), (YP\_001171663.1; 슈도모나스 스투체리 (*Pseudomonas stutzeri*) A1501), (NP\_840385.1; 니트로소모나스 유로피아 (*Nitrosomonas europaea*) ATCC 19718), (YP\_002801221.1; 아조토박터 비넬란디 (*Azotobacter vinelandii*) DJ), (YP\_002787111.1; 데니코커스 데세르티 (*Deinococcus deserti*) VCD115), (YP\_603523.1; 데니코커스 지오써말리스 (*Deinococcus geothermalis*) DSM 11300), (NP\_293799.1; 데니코커스 라디오두란스 (*Deinococcus radiodurans*) R1), (YP\_521550.1; 로도페락스 패리레두센스 (*Rhodoferax ferrireducens*) T118), (YP\_530962.1; 로도슈도모나스 팔루스트리스 (*Rhodopseudomonas palustris*) BisB18), (YP\_531882.1; 로도슈도모나스 팔루스트리스 BisA53), (ZP\_02367347.1; 버크홀데리아 오클라호헨시스 (*Burkholderia oklahomensis*) C6786), (YP\_428079.1; 로도스피릴룸 루브럼 (*Rhodospirillum rubrum*) ATCC 11170), (YP\_530535.1; 로도슈도모나스 팔루스트리스 BisB18), (NP\_901200.1; 크로모박테리움 비올라세움 (*Chromobacterium violaceum*) ATCC 12472), (ZP\_03698345.1; 루티엘라 니트로페룸 (*Lutiella nitroferrum*) 2002), (YP\_001279250.1; 프사이크로박터 (*Psychrobacter*) sp. PRwf-1), (YP\_579484.1; 프사이크로박터 크리오할로렌티스 (*Psychrobacter cryohalolentis*) K5), (ZP\_05618978.1; 엔하이드로박터 에어로사쿠스 (*Enhydrobacter aerosaccus*) SK60), (ZP\_05362319.1; 액시네토박터 라디오레지스턴스 (*Acinetobacter radioresistens*) SK82), (YP\_045288.1; 액시네토박터 sp. ADP1), (ZP\_05823314.1; 액시네토박터 sp. RUH2624), (ZP\_03824416.1; 액시네토박터 sp. ATCC 27244), (YP\_001380280.1; 안에어로믹소박터 (*Anaeromyxobacter*) sp. Fw109-5), (YP\_466103.1; 안에어로믹소박터 데할로게난스 (*Anaeromyxobacter dehalogenans*) 2CP-C), (YP\_088190.1; 만하이미아 속시니시프로두센스

(*Mannheimia succiniciproducens*) MBEL55E), (YP\_001344949.1; 액티노바실러스 속모게네스 (*Actinobacillus succmogenes*) 130Z), (YP\_003007411.1; 아그레가티박터 아프로필러스 (*Aggregatibacter aphrophilus*) NJ8700), (ZP\_01788798.1; 헤모필러스 인플루엔자 (*Haemophilus influenzae*) 3655), (YP\_719012.1; 헤모필러스 쉘누스 (*Haemophilus somnus*) 129PT), (NP\_245642.1; 파스테우렐라 멀토시다 아종 멀토시다 (*Pasteurella multocida* subsp. *multocida*) str. Pm70), (ZP\_05920444.1; 파스테우렐라 다그마티스 (*Pasteurella dagmatis*) ATCC 43325), (ZP\_00133992.2; 액티노바실러스 플레우로뉴모니아 (*Actinobacillus pleuropneumoniae*) serovar 1 str. 4074), (ZP\_04753547.1; 액티노바실러스 마이너 (*Actinobacillus minor*) NM305), (NP\_873873.1; 헤모필러스 두크레이 (*Haemophilus ducreyi*) 35000HP), (ZP\_04978908.1; 만하이미아 헤몰리티카 (*Mannheimia haemolytica*) PHL213), (YP\_002475022.1; 헤모필러스 파라수이스 (*Haemophilus parasuis*) SH0165), (ZP\_05730581.1; 판토에아 (*Pantoea*) sp. At-9b), (YP\_001907133.1; 에르위니아 (*Erwinia*) tasmaniensis Et1/99), (YP\_455287.1; 소달리스 글로시니디우스 (*Sodalis glossinidius*) str. '모르시탄스 (morsitans)'), (ZP\_05723922.1; 디케이아 다단티이 (*Dickeya dadantii*) Ech586), (YP\_003258889.1; 펙토박테리움 바사비이 (*Pectobacterium wasabiae*) WPP163), (YP\_002988159.1; 디케이아 다단티이 Ech703), (NP\_668938.1; 예르시니아 페스티스 (*Yersinia pestis*) KIM 10), (YP\_001479543.1; 세라티아 프로테아마쿨란스 (*Serratia proteamaculans*) 568), (YP\_002934098.1; 에드워드시엘라 익탈루리 (*Edwardsiella ictaluri*) 93-146), (YP\_002151502.1; 프로테우스 미라빌리스 (*Proteus mirabilis*) HI4320), (NP\_930328.1; 포토랍두스 루미네센스 아종 라우몬다이 (*Photorhabdus luminescens* subsp. *laumondii*) TT01), (YP\_002920553.1; 클렙시엘라 뉴모니 (*Klebsiella pneumomae*) NTUH-K2044), (YP\_001177557.1; 엔테로박터 sp. 638), (YP\_003211286.1; 크로노박터 투리센시스 (*Cronobacter turicensis*)), (BAA04663.1; 에스케리키아 콜라이 (*Escherichia coli*)), (YP\_002924403.1; 칸디다투스 하밀토넬라 데펜사 (*Candidatus Hamiltonella defensa*) 5AT (아시르토시폰 피숨 (*Acyrtosiphon Pisum*))), (ZP\_03827735.1; 펙토박테리움 카로토보럼 아종 브라실리엔시스 (*Pectobacterium carotovorum* subsp. *brasiliensis*) PBR1692), (ZP\_01159282.1; 포토박테리움 (*Photobacterium*) sp. SKA34), (YP\_130973.1; 포토박테리움 프로퐁덤 (*Photobacterium profundum*) SS9), (ZP\_06052481.1; 크리몬티아 홀리시 (*Grimontia hollisae*) CIP 101886), (ZP\_05877035.1; 비브리오 푸미시이 (*Vibrio fumiissii*) CIP 102972), (ZP\_05881960.1; 비브리오 메트스크니코이이 (*Vibrio metschnikoyii*) CIP 69.14), (ZP\_05881960.1; 비브리오 메트스크니코이이 (*Vibrio metschnikoyii*) CIP 69.14), (ZP\_02196748.1; 비브리오 sp. AND4), (NP\_934927.1; 비브리오 불니피쿠스 (*Vibrio vulnificus*) YJ016), (ZP\_01866446.1; 비브리오 실로니이 (*Vibrio shilonii*) AKI), (YP\_002416612.1; 비브리오 스피렌디두스 (*Vibrio splendidus*) LGP32), (YP\_002263486.1; 알리이브리오 살모니시다 (*Aliiyibrio salmonicida*) LFI1238), (ZP\_04415114.1; 비브리오 콜레라 by. 알벤시스 (*Vibrio cholerae* by. *albensis*) VL426), (YP\_001143125.1; 에어로모나스 살모니시다 아종 살모니시다 (*Aeromonas salmonicida* subsp. *salmonicida*) A449), (YP\_002892091.1; 톨루모나스 아우엔시스 (*Tolomonas auensis*) DSM 9187), (ZP\_01215350.1; 프사이크로모나스 (*Psychromonas*) sp. CNPT3), (YP\_944598.1; 프사이크로모나스 잉그라함미이 (*Psychromonas ingrahamii*) 37), (YP\_001473443.1; 슈와넬라 세디미니스 (*Shewanella sediminis*) HAW-EB3), (YP\_001761257.1; 슈와넬라 우디이 (*Shewanella woodyi*) ATCC 51908), (YP\_001094519.1; 슈와넬라 로이히카 (*Shewanella loihica*) PV -4), (YP\_001674811.1; 슈와넬라 할리팩센시스 (*Shewanella halifaxensis*) HAW-EB4), (YP\_869191.1; 슈와넬라 (*Shewanella*) sp. ANA-3), (YP\_927371.1; 슈와넬라 아마조넨시스 (*Shewanella amazonensis*) SB2B), (YP\_751160.1; 슈와넬라 프리지디마리나 (*Shewanella frigidimarina*) NCIMB 400), (YP\_563413.1; 슈와넬라 데니트리피칸스 (*Shewanella denitrificans*) OS217), (YP\_001475272.1; 슈와넬라 세디미니스 (*Shewanella sediminis*) HAW-EB3), (YP\_001674949.1; 슈와넬라 할리팩센시스 (*Shewanella halifaxensis*) HAW-EB4), (ZP\_04716660.1; 알테로모나스 마클레오디이 (*Alteromonas macleodii*) ATCC 27126), (YP\_662160.1; 슈도알테로모나스 아틀란티카 (*Pseudoalteromonas atlantica*) T6c), (ZP\_01612225.1; 알테로모나달레스 박테리움 (*Alteromonadales bacterium*) TW-7), (ZP\_01134640.1; 슈도알테로모나스 투니케이트 (*Pseudoalteromonas tunicate*) D2), (YP\_269873.1; 콜웰리아 프사이크레리트리 (*Colwellia psychrerythrae*) a 34H), (YP\_001341167.1; 마리노모나스 (*Marinomonas*) sp. MWYL1), (ZP\_01077352.1; 마리노모나스 sp. MED121), (YP\_001209362.1; 디켈로박터 노도서스 (*Dichelobacter nodosus*) VCS1703A), (ZP\_05705193.1; 카디오박테리움 호미니스 (*Cardiobacterium hominis*) ATCC 15826), (EEY62817.1; 피토프토라 인페스탄스 (*Phytophthora infestans*) T30-4), (EEY62816.1; 피토프토라 인페스탄스 (*Phytophthora infestans*) T30-4), (XP 001694504.1; 클라미도모나스 라인하르티이 (*Chlamydomonas reinhardtii*)), (XP 001753120.1; 피스코미트렐라 파텐스 아종 파텐스 (*Physcomitrella patens* subsp. *Patens*)), (YP\_001804510.1; 시아노테이스 (*Cyanothecae*) sp. ATCC 51142), (ZP\_01729220.1; 시아노테이스 sp. CCY0110), (YP\_003138337.1; 시아노테이스 sp. PCC 8802), (YP\_002380034.1; 시아노테이스 sp. PCC 7424), (YP\_001661110.1; 마이크로시스티스 에어루그모사

(*Microcystis aeruginosa*) NIES-843), (YP\_002485151.1; 시아노테이스 *sp.* PCC 7425), (NP\_441027.1; 시네코시스티스 (*Synechocystis*) *sp.* PCC 6803), (ZP\_01061171.1; 리우벤호에키 엘라 블란덴시스 (*Leeuwenhoekia ella blandensis*) MED217), (YP\_001195862.1; 플라보박테리움 존소니아 (*Flavobacterium johnsoniae*) UW101), (YP\_003194927.1; 로비기니탈레아 비포르마타 (*Robiginitalea biformata*) HTCC2501), (ZP\_01107792.1; 플라보박테리알리스 박테리움 (*Flavobacteriales bacterium*) HTCC2170), (ZP\_01051731.1; 폴라리박터 (*Polaribacter*) *sp.* MED152), (ZP\_01119204.1; 폴라리박터 이르겐시이 (*Polaribacter irgensii*) 23-P), (ZP\_03390929.1; 카프노사이토파가 스푸티게나 (*Capnocytophaga sputigena*) ATCC 33612), (YP\_003141977.1; 카프노사이토파가 오크레이시아 (*Capnocytophaga ochracea*) DSM 7271), (YP\_012240.1; 데설포비브리오 불가리스 str. 힐덴보로우 (*Desulfovibrio vulgaris str. Hildenborough*)), (YP\_002436276.1; 데설포비브리오 불가리스 str. '미야자키 F' (*Desulfovibrio vulgaris str. 'Miyazaki F'*)), (YP\_389730.1; 데설포비브리오 데설포푸리칸 아종 데설포푸리칸스 (*Desulfovibrio desulfuricans subsp. desulfuricans*) str. G20), (YP\_002992165.1; 데설포비브리오 살렉시겐스 (*Desulfovibrio salexigens*) DSM 2638), (YP\_003197901.1; 데설포할로비움 레트빈스 (*Desulfohalobium retbaense*) DSM 5692), (YP\_003157577.1; 데설포마이크로비움 바쿨라툼 (*Desulfomicrobium baculatum*) DSM 4028), (ZP\_03737911.1; 데설포나트로노스피라 티오디스무탄스 (*Desulfonatronospira thiodismutans*) AS03-1), (YP\_002990332.1; 데설포비브리오 살렉시겐스 (*Desulfovibrio salexigens*) DSM 2638), (ZP\_03312237.1; 데설포비브리오 피거 (*Desulfovibrio piger*) ATCC 29098), (YP\_002478890.1; 데설포비브리오 데설포푸리칸스 아종 데설포푸리칸스 (*Desulfovibrio desulfuricans subsp. desulfuricans*) str. ATCC 27774), (YP\_064294.1; 데설포탈레아 사이크로필라 (*Desulfotalea psychrophila*) LSv54), (YP\_594656.1; 로소니아 인트라셀룰라리스 (*Lawsonia intracellularis*) PHE/MN1-00), (ZP\_01621820.1; 링바야 (*Lyngbya*) *sp.* PCC 8106), (ZP\_03272899.1; 아르트로스피라 맥시마 (*Arthrospira maxima*) CS-328), (YP\_845596.1; 신티로포박터 푸마록시단스 (*Syntrophobacter fumaroxidans*) MPOB), (ZP\_04773932.1; 알로크로마티움 비노숨 (*Allochromatium vinosum*) DSM 180), (NP\_869002.1; 로도피렐룰라 발티카 (*Rhodopirellula baltica*) SH 1), (YP\_392571.1; 설푸리모나스 테니트리피칸스 (*Sulfurimonas denitrificans*) DSM 1251), (ZP\_05071717.1; 캄필로박테랄리스 박테리움 (*Campylobacteriales bacterium*) GD 1), (ZP\_04421899.1; 설푸로스피릴룸 딜레이리아눔 (*Sulfurospirillum deleyianum*) DSM 6946), (YP\_001359295.1; 설푸로부름 (*Sulfurovum*) *sp.* NBC37-1), (YP\_951544.1; 미코박테리움 반바알레나이 (*Mycobacterium vanbaalenii*) PYR-1), (YP\_001131488.1; 미코박테리움 길봄 (*Mycobacterium gilvum*) PYR-GCK), (YP\_637714.1; 미코박테리움 *sp.* MCS), (YP\_885188.1; 미코박테리움 스메그마티스 (*Mycobacterium smegmatis*) str. MC2 155), (YP\_001704953.1; 미코박테리움 앵세스스 (*Mycobacterium abscessus*)), (ZP\_04747529.1; 미코박테리움 칸사시이 (*Mycobacterium kansasii*) ATCC 12478), (YP\_001849024.1; 마이코박테리움 마리눔 (*Mycobacterium marinum*) M), (NP\_214922.1; 미코박테리움 투베르쿨로시스 (*Mycobacterium tuberculosis*) H37Rv), (NP\_962819.1; 미코박테리움 아비움 아종 파라투베르쿨로시스 (*Mycobacterium avium subsp. paratuberculosis*) K-10), (ZP\_05223872.1; 미코박테리움 인트라셀룰라이 (*Mycobacterium intracellulare*) ATCC 13950), (YP\_002764919.1; 로도코커스 에리트르폴리스 (*Rhodococcus erythropolis*) PR4), (YP\_702162.1; 로도코커스 요스티이 (*Rhodococcus jostii*) RHA1), (YP\_121562.1; 노카르디아 파르시니카 (*Nocardia farcinica*) IFM 10152), (ZP\_04025361.1; 추카무렐라 파우로메타볼라 (*Tsukamurella paurometabola*) DSM 20162), (YP\_003275431.1; 고르도니아 브론키알리스 (*Gordonia bronchialis*) DSM 43247), (YP\_003160610.1; 존시아 테니트리피칸스 (*Jonesia denitrificans*) DSM 20603), (ZP\_05816650.1; 상구이박터 케디에이 (*Sanguibacter keddiei*) DSM 10542), (ZP\_04368027.1; 셀룰로모나스 플라비게나 (*Cellulomonas flavigena*) DSM 20109), (YP\_002883054.1; 뷰텐버지아 카베미 (*Beutenbergia cavemae*) DSM 12333), (ZP\_03911481.1; 크실라니모나스 셀룰로실리티카 (*Xylanimonas cellulositytica*) DSM 15894), (YP\_924143.1; 노카르디오이데스 (*Nocardioides*) *sp.* IS614), (ZP\_03864789.1; 크리벨라 플라비다 (*Kribbella flavida*) DSM 17836), (ZP\_01131057.1; 마린 액티노박테리움 (*marine actinobacterium*) PHSC20C1), (YP\_001708941.1; 클라비박터 미키가넨시스 아종 세페도니쿠스 (*Clavibacter michiganensis subsp. Sepedonicus*)), (YP\_061462.1; 라이프소니아 크실리 아종 크실리 (*Leifsonia xyli subsp. xyli*) str. CTCB07), (YP\_748183.1; 니트로소모나스 유티트로파 (*Nitrosomonas eutropha*) C91), (YP\_003116892.1; 카테눌리스포라 액시디필라 (*Catenulispora acidiphila*) DSM 44928), (YP\_003199983.1; 나카무렐라 무이티파트타 (*Nakamurella muitipartita*) DSM 44233), (YP\_003154321.1; 브라키박테리움 피시움 (*Brachybacterium faecium*) DSM 4810), (ZP\_03927492.1; 액티노마이세스 우로게니탈리스 (*Actinomyces urogenitalis*) DSM 15434), (YP\_003148931.1; 키토코커스 세덴타리우스 (*Kytococcus sedentarius*) DSM 20547), (ZP\_05803950.1; 스트렙토마이세스 플라보그리세우스 (*Streptomyces flavogriseus*) ATCC 33331), (YP\_001823623.1; 스트렙토마이세스 그리세우스 아종 그리세우스 (*Streptomyces griseus subsp. griseus*) NBRC 13350), (ZP\_05002693.1; 스트렙토마이세스 클라블리케

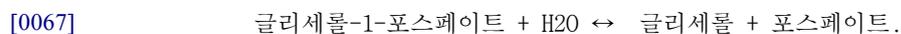
루스 (*Streptomyces clavuligerus*) ATCC 27064), (ZP\_05015493.1; 스트렙토마이세스 스비세우스 (*Streptomyces sviveus*) ATCC 29083), (ZP\_05538660.1; 스트렙토마이세스 그리세오플라부스 (*Streptomyces griseoflavus*) Tu4000), (ZP\_04685789.1; 스트렙토마이세스 가니엔시스 (*Streptomyces ghanaensis*) ATCC 14672), (ZP\_05534308.1; 스트렙토마이세스 비리도크로모게네스 (*Streptomyces viridochromogenes*) DSM 40736), (ZP\_05523554.1; 스트렙토마이세스 리비단스 (*Streptomyces lividans*) TK24), (NP\_823999.1; 스트렙토마이세스 아베르미틸리스 (*Streptomyces avermitilis*) MA-4680), (CBG69921.1; 스트렙토마이세스 스카비에이 (*Streptomyces scabiei*) 87.22), (ZP\_04704905.1; 스트렙토마이세스 알부스 (*Streptomyces albus*) 11074), (ZP\_04997745.1; 스트렙토마이세스 *sp. Mgl*), (ZP\_05509147.1; 스트렙토마이세스 *sp. C*), (ZP\_05514718.1; 스트렙토마이세스 히그로스코피쿠스 (*Streptomyces hygroscopicus*) ATCC 53653), (ZP\_04994290.1; 스트렙토마이세스 *sp. SPB74*), (ZP\_04474082.1; 스트렙토스포랑기움 로세움 (*Streptosporangium roseum*) DSM 43021), (YP\_001160501.1; 살리니스포라 트로피카 (*Salinispora tropica*) CNB-440), (YP\_001538853.1; 살리니스포라 아레니콜라 (*Salinispora arenicola*) CNS-205), (ZP\_04605575.1; 마이크로모노스포라 (*Micromonospora*) *sp.* ATCC 39149), (YP\_832716.1; 아르트로박터 (*Arthrobacter*) *sp.* FB24), (ABR13603.1; 아르트로박터 옥시단스 (*Arthrobacter oxydans*)), (YP\_002956296.1; 마이크로코커스 루테우스 (*Micrococcus luteus*) NCTC 2665), (ZP\_05367249.1; 로티아 무실라지노사 (*Rothia mucilaginosa*) ATCC 25296), (YP\_001854004.1; 코쿠리아 리조필라 (*Kocuria rhizophila*) DC2201), (ZP\_04984463.1; 프란시셀라 툴라렌시스 아종 홀라르크티카 (*Francisella tularensis* subsp. *holarctica*) FSC022), (YP\_001677422.1; 프란시셀라 필로미라지아 아종 필로미라지아 (*Francisella philomiragia* subsp. *philomiragia*) ATCC 25017), (YP\_588827.1; 바우마니아 시카텔리니콜라 (*Baumannia cicadellinicola*) str. Hc (호말로디스카 오아굴라타 (*Homalodisca oagulata*))), (NP\_240007.1; 부크네라 아피디콜라 (*Buchnera aphidicola*) str. APS (아시르코시폰피숨 (*Acyrtosiphonpisum*))), (ZP\_05057494.1; 베루코마이ক্র로비이 박테리움 (*Verrucomicrobiae bacterium*) DG1235), (ZP\_02930252.1; 베루코마이ক্র로비움 스피노숨 (*Verrucomicrobium spinosum*) DSM 4136), (ZP\_01452386.1; 마리프로퓌두스 페로옥시단스 (*Mariprofundus ferrooxydans*) PV-1), 및 (ZP\_01307392.1; 버마넬라 마리스루브리 (*Bermanella marisrubri*)).

[0063] 또한, 본원에 제시된 조성물 및 방법에 사용가능한 포스포트랜스아세틸라제로는 본원에 기술된 임의의 포스포트랜스아세틸라제의 "유도체"로 지칭되는 분자를 포함한다. 상기한 "유도체"는, (1) 본원에 기술된 임의의 포스포트랜스아세틸라제와 상당한 상동성을 공유하며, (2) 아세틸 포스페이트를 아세틸-CoA로 변환하는 반응을 촉매할 수 있는, 특징들을 가진다. 유도체의 아미노산 서열이 포스포트랜스아세틸라제의 서열과 80% 이상, 더 바람직하게는 90% 이상, 가장 바람직하게는 95% 이상 동일한 경우, 포스포트랜스아세틸라제와 "상당한 상동성"을 공유하는 것으로 언급된다.

[0064] **5.4 아세틸 포스파타제 활성의 기능적 파괴**

[0065] 일부 구현예들에서, 본원에 제시된 유전자 변형된 숙주 세포는 아세틸 포스페이트를 아세테이트로 변환하는 효소에 기능적 파괴를 포함한다. 일부 구현예들에서, 효소는 숙주 세포에 천연적이다.

[0066] 일부 구현예들에서, 아세틸 포스페이트를 아세테이트로 변환하는 효소는 글리세롤-1-포스파타제 (EC 3.1.3.21)이다. 일부 구현예들에서, 글리세롤-1-포스파타제 활성을 가진 효소는 RHR2 (GPP1/RHR2; 계통적 명칭: YIL053W), 또는 이의 상동체 또는 변이체이다. GPP1/RHR2는 글리세롤 생합성에 참여하는 구성적으로 발현되는 글리세롤-1-포스파타제이며, 이는 혐기성과 삼투 스트레스 모두에 반응하여 유도된다. 예로, Norbeck *et al.*, *J Biol Chem* 271(23): 13875-13881 (1996); Norbeck *et al.*, *J Biol Chem* 272(9): 13875-13881 (1996); Pahlman *et al.*, *J Biol Chem* 276(5): 3555-3563 (2001); Nevoigt and Stahl, *FEMS Microbiol Rev* 21(3):231-41 (1997); Byrne and Wolf, *Genome Res* 15(10):1456-61; 및 Hirayama *et al.*, *Mol Gen Genet* 249(2):127-38을 참조하며, 이들 문헌의 내용은 그 전체가 원용에 의해 본 명세서에 포함된다. 사카로마이세스 세레비지아의 GPP1/RHR2 유전자 서열은 이전에 개시되었다. 예로, Norbeck *et al.*, *J Biol Chem* 271(23): 13875-13881 (1996); 및 Pahlman *et al.*, *J Biol Chem* 276(5): 3555-3563 (2001)을 참조한다. Gpp1/Rhr2는 이전에 하기 반응을 촉매하는 것으로 기술된 바 있다:



[0068] 사카로마이세스 세레비지아의 GPP1/RHR2 뉴클레오티드 서열의 예로는 등재 번호 NM\_001179403.1 및 본원에 제시된 서열번호 5를 포함한다. 사카로마이세스 세레비지아의 Gpp1/Rhr2 단백질 서열의 예로는 등재 번호 NP\_012211 및 본원에 제시된 서열번호 6을 포함한다.

- [0069] 아세틸 포스페이트를 아세테이트로 가수분해하는 반응을 또한 촉매하는 GPP1/RHR2의 매우 관련성 높은 상동체는 HOR2 (GPP2/HOR2; 계통학적 명칭: YER062C)이다. Gpp2/Hor2 역시 다음과 같은 반응을 촉매할 수 있는 글리세롤-1-포스파타제로서 이전에 기술된 바 있다: 글리세롤-1-포스페이트 + H<sub>2</sub>O ↔ 글리세롤 + 포스페이트. 따라서, GPP2/HOR2의 기능적 파괴 역시 본원에 제시된 조성물 및 방법에 이용된다. 사카로마이세스 세레비지에의 GPP2/HOR2 유전자 서열은 종래에 개시되었다. 예로, Norbeck *et al.*, *J. of Biological Chemistry* 271(23): 13875-13881 (1996); 및 Pahlman *et al.*, *J. of Biological Chemistry* 276(5): 3555-3563 (2001)을 참조한다. 사카로마이세스 세레비지에의 GPP2/HOR2 뉴클레오티드 서열의 예는 등재 번호 NM\_001178953.3 및 본원에 제시된 서열번호 7을 포함한다. 사카로마이세스 세레비지에의 Gpp1/Rhr2 단백질 서열의 예는 등재 번호 NP\_010984 및 본원에 제시된 서열번호 8을 포함한다.
- [0070] 당해 기술 분야에서 이해되는 바와 같이, 사카로마이세스 세레비지에 이외의 효모에서 GPP1/RHR2 및/또는 GPP2/HOR2에 대한 천연 상동체들을 본원에 기술된 방법을 이용하여 불활화할 수 있다. 아울러, 아세틸-포스파타제 활성 (예, RHR2 및/또는 HOR2)을 코딩하는 폴리뉴클레오티드, 유전자 및/또는 폴리펩타이드를 이용하여, 다른 숙주 세포에서 다른 폴리뉴클레오티드, 유전자 및/또는 폴리펩타이드를 동정하거나 또는 아세틸-포스파타제 활성을 가진 상동체를 동정할 수 있다. 상기한 서열은, 예를 들어, 당업자들에게 잘 알려진 문헌 및/또는 생물정보 데이터베이스에서 찾을 수 있다. 예컨대, 생물정보를 이용하여 다른 세포 타입들에서 아세틸-포스파타제 활성을 코딩하는 서열을 동정하는 것은, 본원에 제시된 서열과 같은 아세틸-포스파타제 및/또는 글리세롤-1-포스파타제 활성을 코딩하는 공지된 DNA 및 폴리펩타이드 서열을 이용하여 공공의 데이터베이스를 (전술한 바와 같이) BLAST 검색함으로써, 달성할 수 있다. 동정은 Clustal W method of alignment using the default parameters of 갭 패널티 (GAP PENALTY) = 10, 갭 길이 패널티 (GAP LENGTH PENALTY) = 0.1 및 단백질 웨이트 매트릭스의 Gonnet 250 시리즈로 구성된 디폴트 파라미터를 이용하여, Clustal W 정렬 방법을 토대로 할 수 있다.
- [0071] 일부 구현예들에서, 내인성 아세틸 포스페이트를 아세테이트로 변환하는 효소 (예, RHR2 또는 HOR2)의 활성 또는 발현이 적어도 약 50%까지 감소된다. 다른 구현예에서, 내인성 아세틸 포스페이트를 아세테이트로 변환하는 효소의 활성 또는 발현은, 내인성 아세틸 포스페이트를 아세테이트로 변환하는 효소의 활성 또는 발현의 감소 또는 결손을 포함하지 않는 제조할 미생물과 비교해, 적어도 약 60%까지, 적어도 약 65%까지, 적어도 약 70%까지, 적어도 약 75%까지, 적어도 약 80%까지, 적어도 약 85%까지, 적어도 약 90%까지, 적어도 약 95%까지, 또는 적어도 약 99%까지 감소된다. 일부 구현예들에서, 내인성 아세틸 포스페이트를 아세테이트로 변환하는 효소는 RHR2 또는 이의 상동체이다. 일부 구현예들에서, 내인성 아세틸 포스페이트를 아세테이트로 변환하는 효소는 HOR2 또는 이의 상동체이다.
- [0072] 당해 기술 분야의 당업자에게 이해되는 바와 같이, 아세틸 포스페이트를 아세테이트로 변환하는 단백질, 예컨대 글리세롤-1-포스파타제 (예, RHR2 및/또는 HOR2)의 활성을 감소 또는 교란시키는데 이용가능한 기전들이 몇가지 있으며, 비제한적인 예로, 제어형 프로모터의 사용, 구성적인 약한 프로모터의 사용, 이배체 효모에서 이 단백질을 코딩하는 유전자의 2카피 중 1카피의 파괴, 이배체 효모에서 유전자 2 카피 둘다의 파괴, 안티센스 핵산의 발현, siRNA의 발현, 내인성 프로모터의 네거티브 조절인자의 과발현, 내인성 또는 이종의 유전자의 활성 변형, 특히 활성이 보다 낮은 이종의 유전자의 사용 등 또는 이들의 조합을 포함한다.
- [0073] 일부 구현예들에서, 유전자 변형된 숙주 세포는, 아세틸-포스파타제 활성 (예, RHR2, HOR2 또는 이의 상동체 또는 변이체)을 코딩하는 하나 이상의 유전자에, 이 유전자에 의해 코딩되는 폴리펩타이드의 활성을 감소시키는, 돌연변이를 포함한다. 다른 구현예에서, 유전자 변형된 숙주 세포는, 유전자에 의해 코딩되는 폴리펩타이드의 활성을 감소시키는, 아세틸-포스파타제 활성 (예, RHR2, HOR2 또는 이의 상동체 또는 변이체)을 코딩하는 유전자의 일부 결손을 포함한다. 다른 구현예에서, 유전자 변형된 숙주 세포는, 유전자에 의해 코딩되는 폴리펩타이드의 활성을 감소시키는, 아세틸-포스파타제 활성 (예, RHR2, HOR2 또는 이의 상동체 또는 변이체)을 코딩하는 유전자의 전체 결손을 포함한다. 또 다른 구현예에서, 유전자 변형된 숙주 세포는, 유전자에 의해 코딩되는 폴리펩타이드의 활성을 감소시키는, 아세틸-포스파타제 활성 (예, RHR2, HOR2 또는 이의 상동체 또는 변이체)을 코딩하는 유전자와 관련된 조절 영역의 변형을 포함한다. 또 다른 구현예에서, 유전자 변형된 숙주 세포는, 아세틸-포스파타제 (예, RHR2, HOR2 또는 이의 상동체 또는 변이체)를 코딩하는 유전자의 전사를 감소시키는, 전사 조절인자의 변형을 포함한다.
- [0074] 일부 구현예들에서, 아세틸 포스페이트의 아세테이트로의 변환을 촉매할 수 있는 단백질을 코딩하는 하나 이상의 유전자의 파괴는, 미생물 세포로 구조체의 도입시, 그러한 유전자 (예, RHR2 또는 HOR2)를 특이적으로 파괴할 수 있는 "파괴 구조체"를 이용하고, 그에 따라, 파괴된 유전자를 비-기능성으로 만듦으로써, 달성된다. 일

부 구현예들에서, 타겟 유전자의 파괴는 기능성 단백질의 발현을 방지한다. 일부 구현예들에서, 타겟 유전자의 파괴로 파괴된 유전자로부터 비-기능성 단백질이 발현된다. 일부 구현예들에서, 아세틸 포스페이트를 아세테이트로 변환할 수 있는 단백질을 코딩하는 유전자의 파괴는 상동적인 재조합에 의한 타겟 유전자 좌내 "파괴 서열"의 삽입에 의해 달성된다. 이러한 구현예에서, 파괴 구조체는, 타겟 유전자 좌의 한쌍의 뉴클레오티드 서열과 상동적인 한쌍의 뉴클레오티드 서열 (상동적인 서열)이 측면에 배치된 파괴 서열을 포함한다. 타겟 유전자의 타겟 영역이 파괴 구조체로 치환되면, 파괴 서열은 기능성 단백질의 발현을 방지하거나, 또는 타겟 유전자로부터 비-기능성 단백질이 발현되게 한다.

[0075] 유전자를 파괴할 수 있는 파괴 구조체는 당해 기술 분야에 널리 공지된 표준 분자 생물 기법을 이용하여 구축할 수 있다. 예로, ambrook *et al.*, 2001, *Molecular Cloning -- A Laboratory Manual*, 3<sup>rd</sup> edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 및 Ausubel *et al.*, eds., Current Edition, *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley Interscience, NY를 참조한다. 본 발명의 방법의 실시에서 변경될 수 있는 파괴 구조체에 대한 파라미터로는, 비제한적인 예로, 상동적인 서열의 길이; 상동적인 서열의 뉴클레오티드 서열; 파괴 서열의 길이; 파괴 서열의 뉴클레오티드 서열; 및 타겟 유전자의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 각 상동적인 서열의 유효 길이 범위는 50 - 5,000 bp (base pair)를 포함한다. 특정 구현예에서, 각 상동적인 서열의 길이는 약 500 bp이다. 유전자 타겟팅에 필요한 상동체의 길이에 대한 내용은 Hasty *et al.*, *Mol Cell Biol* 11:5586-91 (1991)을 참조한다. 일부 구현예들에서, 상동적인 서열은 타겟 서열의 코딩 서열을 포함한다. 다른 구현예들에서, 상동적인 서열은 타겟 유전자의 상류 또는 하류 서열을 포함한다. 일부 구현예들에서, 하나의 상동적인 서열은 타겟 유전자의 코딩 서열의 5'에 위치한 뉴클레오티드 서열에 상동적인 뉴클레오티드 서열을 포함하며, 다른 상동적인 서열은 타겟 유전자의 코딩 서열의 3'에 위치한 뉴클레오티드 서열에 상동적인 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 파괴 서열은 파괴 서열을 포함하는 미생물을 선별할 수 있는 선별 마커를 코딩하는 뉴클레오티드 서열을 포함한다. 즉, 이러한 구현예들에서, 파괴 구조체는 듀얼 기능, 즉, 타겟 유전자를 기능적으로 파괴하는 기능과 타겟 유전자가 기능적으로 파괴된 세포를 동정하기 위한 선별 마커를 제공하는 기능을 가진다. 일부 구현예들에서, 종결 코돈 (termination codon)은, 타겟 유전자에 의해 코딩된 야생형 단백질의 활성을 일부 가지는 융합 단백질을 만들 수 있는, 번역 리드-스로우 (translational read-through)를 방지하기 위해, 선별 마커를 코딩하는 뉴클레오티드 서열과 인-프레임으로, 그리고 하류에 위치된다. 일부 구현예들에서, 파괴 서열의 길이는 1 bp이다. 1 bp의 삽입은, 코딩 서열내 1 bp의 삽입이 기능성 단백질의 발현을 방지할 수 있는 프레임 쉬프트 돌연변이를 만들 수 있기 때문에, 타겟 유전자를 파괴하는데 충분할 수 있다. 일부 구현예들에서, 파괴 서열의 서열은 상동적인 서열들 사이에 위치한 타겟 유전자의 뉴클레오티드 서열과 1 bp 차이를 가진다. 타겟 유전자내 뉴클레오티드 서열을 파괴 서열로 치환하면, 도입되는 1 bp 치환으로 단백질의 주요 부위에 하나의 아미노산 치환과 비-기능성 단백질의 발현이 이루어질 수 있다. 그러나, 매우 짧은 파괴 서열을 이용하여 수행되는 파괴는 자연적인 돌연변이를 통해 야생형 서열로 복구되기 쉬워, 아세틸-포스파타제 기능이 속주 균주로 복원됨을, 인지하여야 한다. 즉, 특정 구현예에서, 파괴 서열은 1 - 수개 bp 보다 길다. 다른 극단적인 예로, 과도하게 긴 길이의 파괴 서열은 보통 길이의 파괴 서열에 비해 임의의 이점을 부여할 가능성은 없으며, 형질감염 또는 타겟팅 효과가 사라질 수도 있다. 이 문맥에서, 과도하게 긴 길이는 타겟 유전자에서 선택된 상동적인 서열들 간의 거리 보다 수배 더 긴 것을 의미한다. 즉, 특정 구현예들에서, 파괴 서열의 길이는 2 - 2,000 bp일 수 있다. 다른 구현예들에서, 파괴 서열의 길이는, 파괴 구조체내 상동적인 서열과 부합되는 타겟 유전자 좌의 영역들 사이의 간격과 대략 동일한 길이이다.

[0076] 일부 구현예들에서, 파괴 구조체는 선형 DNA 분자이다. 다른 구현예들에서, 파괴 구조체는 환형의 DNA 분자이다. 일부 구현예들에서, 환형의 파괴 구조체는 전술한 바와 같이 파괴 서열에 의해 분리되어 있는 한쌍의 상동적인 서열을 포함한다. 일부 구현예들에서, 환형의 파괴 구조체는 하나의 상동적인 서열을 포함한다. 이러한 환형의 파괴 구조체는, 타겟 유전자 좌로의 도입시, 양 말단에 상동적인 서열의 영역이 위치되어 선형화될 것이며, 파괴 구조체의 나머지 세그먼트들은 임의의 타겟 유전자 뉴클레오티드 서열을 치환하지 않고도 타겟 유전자에 삽입되어 이를 파괴한다. 특정 구현예에서, 환형의 파괴 구조체의 단일한 상동적인 서열은 타겟 유전자의 코딩 서열 내에 위치한 서열과 상동적이다.

[0077] 파괴 구조체는 당해 기술 분야의 당업자들에게 공지된 임의 방법으로 미생물 세포에 도입할 수 있으며, 이로 한정되지 않는다. 이러한 방법으로는, 비제한적으로, 세포에 의한 용액으로부터의 분자의 직접 흡수, 또는 예를 들어 리포솜 또는 면역리포솜을 이용한 리포펙션을 통한 촉매화된 흡수; 입자 매개 형질감염 등을 포함한다. 예로, 미국 특허 5,272,065; Goeddel *et al.*, eds, 1990, *Methods in Enzymology*, vol. 185, Academic Press,

Inc., CA; Krieger, 1990, Gene Transfer and Expression -- A Laboratory Manual, Stockton Press, NY; Sambrook et al., 1989, Molecular Cloning -- A Laboratory Manual, Cold Spring Harbor Laboratory, NY; 및 Ausubel et al., eds., Current Edition, Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley Interscience, NY를 참조한다. 효모 세포를 형질전환하는 구체적인 방법들이 당해 기술 분야에 잘 공지되어 있다. Hinnen *et al.*, Proc. Natl. Acad. Sci. USA 75:1292-3 (1978); Cregg *et al.*, Mol. Cell. Biol. 5:3376-3385 (1985)를 참조한다. 기법의 예로는, 비제한적으로, 스페로플라스팅 (spheroplasting), 전기충격 (electroporation), PEG 1000 매개 형질전환 및 리튬 아세테이트 또는 리튬 클로라이드 매개의 형질전환을 포함한다.

[0078] 5.5 아세틸-CoA 생산을 개선하기 위한 추가적인 변형

[0079] 5.5.1 ADA

[0080] 일부 구현예들에서, 본원에 제공되는 유전자 변형된 숙주 세포는, 아실화 아세트알데하이드 데하이드로게나제 ("아세틸알데하이드 데하이드로게나제, 아세틸화, " "아세틸알데하이드 데하이드로게나제, 아실화" 또는 ADA (EC 1.2.1.10)로도 지칭됨)를 코딩하는, 하나 이상의 이종의 뉴클레오티드 서열을 더 포함한다.

[0081] 본원에 제시된 조성물 및 방법에 사용가능한 이러한 반응을 촉매할 수 있는 단백질은 하기 4가지 타입의 단백질을 포함한다:

[0082] (1) 아세틸-CoA의 아세트알데하이드로의 가역적인 변환과, 후속적인 아세트알데하이드의 에탄올로의 가역적인 변환을 촉매하는, 2가지 기능의 단백질. 이런 유형의 단백질의 예는 *E. coli*의 AdhE 단백질 (Gen Bank No: NP\_415757)이다. AdhE는 유전자 융합체의 진화 산물인 것으로 보인다. AdhE 단백질의 NH<sub>2</sub>-말단 영역은 알데하이드:NAD<sup>+</sup> 옥시도리덕타제에 대해 상동성이 높은 반면, COOH-말단 영역은 Fe<sup>2+</sup>-의존성 에탄올:NAD<sup>+</sup> 옥시도리덕타제 패밀리와 상동적이다 (Membrillo-Hernandez *et al.*, (2000) *J. Biol. Chem.* 275: 33869-33875). *E. coli* AdhE는 금속-촉매화된 산화를 수행하며, 따라서, 산소에 민감하다 (Tamarit *et al.* (1998) *J. Biol. Chem.* 273:3027-32).

[0083] (2) 엄격 또는 선택적인 혐기성 미생물에서 아세틸-CoA의 아세트알데하이드로의 가역적인 변환을 촉매하지만, 알코올 데하이드로게나제 활성은 가지지 않는, 단백질. 이 유형의 단백질에 대한 예가 클로스트리듐 클루이베리 (*Clostridium kluverii*)에서 보고된 바 있다 (Smith *et al.* (1980) *Arch. Biochem. Biophys.* 203: 663-675). ADA는 클로스트리듐 클루이베리 (*Clostridium kluverii*) DSM 555의 게놈 (등재번호: EDK33116)에 주석으로 달려 있다. 상보적인 단백질 AcdH는 락토바실러스 플라타룸 (*Lactobacillus plantarum*)의 게놈에서 동정된다 (등재 번호: NP\_784141). 이 유형의 단백질에 대한 다른 예는 클로스트리듐 베이예린키이 (*Clostridium beijerinckii*) NRRL B593의 *ald* 유전자 산물이다 (Toth *et al.* (1999) *Appl. Environ. Microbiol.* 65: 4973-4980, 등재 번호: AAD31841).

[0084] (3) 에탄올아민 이화 작용에 참여하는 단백질. 에탄올아민은 다수의 장내세균들이 탄소원 및 질소원으로 이용할 수 있다 (Stojiljkovic *et al.* (1995) *J. Bacteriol.* 177: 1357-1366). 에탄올아민은 먼저 에탄올아민 암모니아 리아제에 의해 암모니아와 아세트알데하이드로 변환된 다음, 아세트알데하이드는 ADA에 의해 아세틸-CoA로 변환된다. 이 유형의 ADA의 예가 살모넬라 티피무리움 (*Salmonella typhimurium*)의 EutE 단백질이다 (Stojiljkovic *et al.* (1995) *J. Bacteriol.* 177: 1357-1366, 등재 번호: AAL21357; U18560.1도 참조함). 또한, *E. coli*는 에탄올아민을 이용할 수 있으며 (Scarlett *et al.* (1976) *J. Gen. Microbiol.* 95:173-176), 살모넬라 티피무리움 (*S. typhimurium*)의 EutE 단백질과 상동적인 EutE 단백질을 가지고 있다 (등재 번호: AAG57564; EU897722.1).

[0085] (4) 4-하이드록시-2-케토발레레이트 이화작용에 관여하는 2가지 기능의 알돌라제-데하이드로게나제 복합체의 일부인 단백질. 상기한 2가지 기능의 효소는, 다수 박테리아 종들에서 페놀, 톨레이트, 나프탈렌, 바이페닐 및 그외 방향족 화합물의 분해의 중간산물인, 카테콜의 메타-절단 경로의 마지막 2 단계를 촉매한다 (Powlowski and Shingler (1994) *Biodegradation* 5, 219-236). 4-하이드록시-2-케토발레레이트는 4-하이드록시-2-케토발레레이트 알돌라제에 의해 먼저 피루베이트와 아세트알데하이드로 변환된 다음, ADA에 의해 아세트알데하이드는 아세틸-CoA로 변환된다. 이 유형의 ADA의 예가 슈도모나스 (*Pseudomonas*) sp CF600의 DmpF 단백질 (등재 번호: CAA43226)이다 (Shingler *et al.* (1992) *J. Bacteriol.* 174:71 1-24). *E. coli*는 슈도모나스 (*Pseudomonas*) sp. CF600의 DmpF 단백질에 상동적인 MphF 단백질을 가지고 있다 (Ferrandez *et al.* (1997) *J. Bacteriol.* 179: 2573-2581, 등재 번호: NP\_414885).

[0086] 일부 구현예들에서, 본원에 기술된 조성물 및 방법에 사용가능한 ADA (또는 이러한 활성을 코딩하는 핵산 서열)는, 국제 출원 공개번호 WO 2009/013159에 기술된 바와 같이, 에스케리키아 콜라이 (*Escherichia coli*) adhE, 엔타모에바 히스톨리티카 (*Entamoeba histolytica*) adh2, 스타필로코쿠스 아우레우스 (*Staphylococcus aureus*) adhE, 피로마이세스 (*Piromyces*) sp.E2 adhE, 클로스트리듐 클루이베리 (*Clostridium kluyveri*) (EDK33116), 락토바실러스 플라타룸 (*Lactobacillus plantarum*) acdH 및 슈도모나스 푸티타 (*Pseudomonas putida*) (YP 001268189)로 이루어진 군으로부터 선택되며, 이들은 그 전체가 원용에 의해 본 명세서에 포함된다. 일부 구현예들에서, ADA는 클로스트리듐 보툴리눔 (*Clostridium botulinum*) eutE (FR745875.1), 데설포탈리에 피크로필라 (*Desulphotalea psychrophila*) eutE (CR522870.1), 액시네토박터 (*Acinetobacter*) sp. HBS-2 eutE (ABQ44511.2), 칼디트릭스 아비시 (*Caldithrix abyssi*) eutE (ZP\_09549576) 및 할로루브룸 라쿠스프로퐁디 (*Halorubrum lacusprofundi*) ATCC 49239 (YP\_002565337.1)로 이루어진 군으로부터 선택된다.

[0087] 특정 구현예에서, 본원에 제공되는 조성물 및 방법에 사용가능한 ADA는 디케이야 지아 (*Dickeya zeae*)의 eutE이다. 디케이야 지아 (*Dickeya zeae*)의 대표적인 eutE 뉴클레오티드 서열은 등재 번호 NC\_012912.1:1110476..1111855와 본원에 제시된 서열번호 9를 포함한다. 디케이야 지아 (*Dickeya zeae*)의 대표적인 eutE 단백질 서열은 등재 번호 YP\_003003316 및 본원에 제공된 서열번호 10을 포함한다.

[0088] 또한, 본원에 제공된 조성물 및 방법에 유용한 ADA는 본원에 기술된 임의의 ADA의 "유도체"로 지칭될 수 있는 분자를 포함한다. 이러한 "유도체"는 다음과 같은 특징을 가진다: (1) 본원에 기술된 임의의 ADA와 상당한 상동성을 보유함, (2) 아세트알데하이드의 아세틸-CoA로의 변환을 촉매할 수 있음. ADA의 유도체는, 이 유도체의 아미노산 서열이 본원에 기술된 임의의 ADA와 80% 이상, 85% 이상, 더 바람직하게는 90% 이상, 가장 바람직하게는 95% 이상 동일하다면, ADA와 "실질적인 상동성"을 공유하는 것으로 지칭된다.

[0089] **5.5.2 PDH-바이패스의 기능적 파괴**

[0090] 아세틸-CoA는 PDH 복합체에 의해 촉매되는 피루베이트의 산화적 탈카르복시화에 의해 미토콘드리아에서 합성될 수 있다. 그러나, 사카로마이세스 세레비지에 (*S. cerevisiae*)는 미토콘드리아 밖으로 아세틸-CoA를 이동시키지 못하기 때문에, PDH 바이패스는 세포질 구획내로 아세틸-CoA를 제공하는데 필수적인 역할을 하며, 피루베이트를 아세틸-CoA로 변환하는 PDH 반응에 대한 대안 경로를 제공한다. PDH 바이패스에는, 효소 피루베이트 데카르복실라제 (PDC; EC 4.1.1.1), 아세트알데하이드 데하이드로게나제 (ACDH; EC 1.2.1.5 및 EC 1.2.1.4), 및 아세틸-CoA 신테타제 (ACS; EC 6.2.1.1)가 참여한다. 피루베이트 데카르복실라제는 피루베이트의 아세트알데하이드와 이산화탄소로의 탈카르복시화를 촉매한다. 아세트알데하이드 데하이드로게나제는 아세트알데하이드를 아세트산으로 산화한다. 사카로마이세스 세레비지에 (*S. cerevisiae*)에서, 알데하이드 데하이드로게나제 패밀리는 5종의 멤버를 포함한다. *ALD2* (YMR170c), *ALD3* (YMR169c) 및 *ALD6* (YPL061w)는 세포질 이소형이며, *ALD4* (YOR374w) 및 *ALD5* (YER073w)는 미토콘드리아 효소를 코딩한다. 주요 세포질 아세트알데하이드 데하이드로게나제 이소형은 *ALD6*에 의해 코딩된다. 아세테이트로부터 아세틸-CoA의 형성은 ACS에 의해 촉매되며, ATP의 가수분해를 수반한다. 2종의 구조 유전자 *ACSI*과 *ACS2*는 ACS를 코딩한다.

[0091] 일부 구현예들에서, 본원에 제시된 유전자 변형된 숙주 세포는 PDH-바이패스 경로의 하나 이상의 유전자에 대한 기능적 파괴를 포함한다. 일부 구현예들에서, 숙주 세포의 PDH-바이패스의 하나 이상의 유전자에 대한 파괴로, 다음과 같은 반응들 중 한가지 이상의 반응을 촉매하는 능력이 손상된 유전자 변형된 미생물 세포가 만들어진다: (1) 피루베이트 데카르복실라제에 의한 피루베이트의 아세트알데하이드로의 탈카르복시화; (2) 아세트알데하이드 데하이드로게나제에 의한 아세트알데하이드의 아세테이트로의 변환; 및 (3) 아세틸-CoA 신테타제에 의한 아세테이트 및 CoA로부터 아세틸-CoA의 합성.

[0092] 일부 구현예들에서, 모 세포와 비교하여, 숙주 세포는 PDH-바이패스 경로의 하나 이상의 유전자에 대한 기능적 파괴를 포함하며, 이때 단독으로 또는 약한 ADA와 조합하여 감소된-기능적 또는 비-기능적 PDH-바이패스 경로의 활성은 숙주 세포의 증식, 생존 및/또는 건강을 뒷받침하는데 충분하지 않다.

[0093] 일부 구현예들에서, PDH-바이패스의 하나 이상의 내인성 단백질의 활성 또는 발현이 약 50% 이상까지 감소된다. 다른 구현예에서, PDH-바이패스의 하나 이상의 내인성 단백질의 활성 또는 발현이, PDH-바이패스의 하나 이상의 내인성 단백질의 활성 또는 발현에 대한 감소 또는 결손을 포함하지 않는 재조합 미생물과 비교해, 약 60% 이상까지, 약 65% 이상까지, 약 70% 이상까지, 약 75% 이상까지, 약 80% 이상까지, 약 85% 이상까지, 약 90% 이상까지, 약 95% 이상까지, 또는 약 99% 이상까지 감소된다.

[0094] **5.5.2.1 ALD4 및 ALD6**

- [0095] 일부 구현예들에서, 알데하이드 데하이드로게나제 (ACDH) 활성을 코딩하는 하나 이상의 유전자가 숙주 세포에서 기능적으로 파괴된다. 일부 구현예들에서, 알데하이드 데하이드로게나제는 ALD2, ALD3, ALD4, ALD5, ALD6 및 이들의 상동체와 변이체로 이루어진 군으로부터 선택되는 유전자에 의해 코딩된다.
- [0096] 일부 구현예들에서, 유전자 변형된 숙주 세포는 ALD4의 기능적 파괴를 포함한다. 사카로마이세스 세레비지에 (*Saccharomyces cerevisiae*)의 대표적인 *ALD4* 뉴클레오티드 서열은 등재번호 NM\_001183794 및 본원에 제시된 서열번호 11을 포함한다. 사카로마이세스 세레비지에 (*Saccharomyces cerevisiae*)의 대표적인 Ald4 단백질 서열은 등재번호 NP\_015019.1 및 본원에 제시된 서열번호 12를 포함한다.
- [0097] 일부 구현예들에서, 유전자 변형된 숙주 세포는 세포질 알데하이드 데하이드로게나제 (ALD6)의 기능적 파괴를 포함한다. Ald6p는 천연적인 PDH-바이패스에서 아세트알데하이드를 아세테이트로 변환하는 기능을 한다. 사카로마이세스 세레비지에 (*Saccharomyces cerevisiae*)의 대표적인 *ALD6* 뉴클레오티드 서열은 등재번호 SCU56604 및 본원에 제시된 서열번호 13을 포함한다. 사카로마이세스 세레비지에 (*Saccharomyces cerevisiae*)의 대표적인 Ald6 단백질 서열은 등재번호 AAB01219 및 본원에 제시된 서열번호 14를 포함한다.
- [0098] 당해 기술 분야에서 이해되는 바와 같이, 사카로마이세스 세레비지에 (*S. cerevisiae*) 이외의 효소에서 유래되는 알데하이드 데하이드로게나제에 대한 천연 상동체는 본원에 기술된 방법을 이용하여 유사하게 불활화할 수 있다.
- [0099] 당해 기술 분야의 당업자가 이해하는 바와 같이, 2 이상의 알데하이드 데하이드로게나제의 활성 또는 발현이 감소 또는 없어질 수 있다. 일 특정 구현예에서, ALD4 및 ALD6, 또는 이의 상동체 또는 변이체의 활성 또는 발현이 감소 또는 소실된다. 다른 특정 구현예에서, ALD5 및 ALD6, 또는 이의 상동체 또는 변이체의 활성 또는 발현이 감소 또는 소실된다. 또 다른 특정 구현예에서, ALD4, ALD5 및 ALD6, 또는 이의 상동체 또는 변이체의 활성 또는 발현이 감소 또는 소실된다. 또 다른 특정 구현예에서, 세포질에 위치한 알데하이드 데하이드로게나제, ALD2, ALD3 및 ALD6, 또는 이들의 상동체 또는 변이체의 활성 또는 발현이, 감소 또는 소실된다. 또 다른 특정 구현예에서, 미토콘드리아에 위치한 알데하이드 데하이드로게나제, ALD4 및 ALD5, 또는 이들의 상동체 또는 변이체의 활성 또는 발현이, 감소 또는 소실된다.
- [0100] **5.5.2.2 ACS1 및 ACS2**
- [0101] 일부 구현예들에서, 아세틸-CoA 신테타제 (ACS) 활성을 코딩하는 하나 이상의 유전자가 숙주 세포에서 기능적으로 파괴된다. 일부 구현예들에서, 아세틸-CoA 신테타제는 ACS1, ACS2 및 상동체로 이루어진 군으로부터 선택되는 유전자, 및 이들의 상동체 또는 변이체에 의해 코딩된다.
- [0102] 일부 구현예들에서, 아세틸-CoA 신테타제 (ACS) 활성을 코딩하는 하나 이상의 유전자가 숙주 세포에서 기능적으로 파괴된다. ACS1과 ACS2는, 둘다, 아세테이트를 아세틸-CoA로 변환할 수 있는, 아세틸-CoA 신테타제이다. ACS1은 호흡 조건 (respiratory condition)에서만 발현되는 반면, ACS2는 구성적으로 발현된다. ACS2가 낮아지면, 균주는 호흡 조건 (예, 에탄올, 글리세롤, 또는 아세테이트 매질)에서 증식할 수 있지만, 발효가능한 탄소원 (예, 슈크로스, 글루코스)에서는 사멸한다.
- [0103] 일부 구현예들에서, 유전자 변형된 숙주 세포는 ACS1의 기능적인 파괴를 포함한다. 사카로마이세스 세레비지에 (*S. cerevisiae*)의 *ACS1* 유전자 서열은 기존에 공지되었다. 예로, Nagasu *et al.*, *Gene* 37 (1-3):247-253 (1985)을 참조한다. 사카로마이세스 세레비지에 (*Saccharomyces cerevisiae*)의 대표적인 *ACS1* 뉴클레오티드 서열은 등재번호 X66425와 본원에 제시된 서열번호 15를 포함한다. 사카로마이세스 세레비지에 (*Saccharomyces cerevisiae*)의 대표적인 Acs1 단백질 서열은 등재번호 AAC04979와 본원에 제시된 서열번호 16을 포함한다.
- [0104] 일부 구현예들에서, 유전자 변형된 숙주 세포는 ACS2의 기능적인 파괴를 포함한다. 사카로마이세스 세레비지에 (*S. cerevisiae*)의 *ACS2* 유전자 서열은 기존에 공지된 바 있다. 예로, Van den Berg *et al.*, *Eur. J. Biochem.* 231(3):704-713 (1995)을 참조한다. 사카로마이세스 세레비지에 (*Saccharomyces cerevisiae*)의 대표적인 *ACS2* 뉴클레오티드 서열은 등재번호 S79456과 본원에 제시된 서열번호 17을 포함한다. 사카로마이세스 세레비지에 (*Saccharomyces cerevisiae*)의 대표적인 Acs2 단백질 서열은 등재번호 CAA97725와 본원에 제시된 서열번호 18을 포함한다.
- [0105] 당해 기술 분야에서 이해되는 바와 같이, 사카로마이세스 세레비지에 (*S. cerevisiae*) 이외의 효소에서, 아세틸-CoA 신테타제에 대한 천연 상동체를 본원에 기술된 방법을 이용하여 유사하게 불활화할 수 있다.
- [0106] 일부 구현예들에서, 숙주 세포는 호흡 조건에서 (즉, 숙주 세포를 예컨대, 에탄올, 글리세롤 또는 아세테이트의

존재 하에 배양하였을 때) 아세테이트를 아세틸-CoA로 변환할 수 있는, 세포질 아세틸-coA 신테타제 활성을 포함한다. 일부 이러한 구현예에서, 숙주 세포는 ACS1 활성을 포함하는 효모 세포이다. 다른 구현예들에서, 숙주 세포는, 모 세포와 비교하여, 호흡 조건에서 내인성 아세틸-CoA 신테타제 활성을 포함하지 않거나, 감소된다. 이러한 일부 구현예들에서, 숙주 세포는, 모 세포와 비교하여, ACS1 활성이 없거나 감소된, 효모 세포이다.

[0107] 일부 구현예들에서, 숙주 세포는 무-호흡 조건에서 (즉, 숙주 세포를 발효가능한 탄소원 (예, 슈크로스, 글루코스)의 존재 하에 배양하였을 때) 아세테이트를 아세틸-CoA로 변환할 수 있는, 세포질 아세틸-coA 신테타제 활성을 포함한다. 일부 이러한 구현예들에서, 숙주 세포는 ACS2 활성을 포함하는 효모 세포이다. 다른 구현예들에서, 숙주 세포는, 모 세포와 비교하여, 무-호흡 조건에서는 내인성 아세틸-CoA 신테타제 활성이 없거나 감소된다. 일부 이러한 구현예들에서, 숙주 세포는, 모 세포와 비교하여, ACS2 활성이 없거나 감소된, 효모 세포이다.

[0108] 일부 구현예들에서, 숙주 세포는 이중의 PK 및 세포질의 아세틸-CoA 신테타제 활성 (예, ACS1 및/또는 ACS2)을 포함한다. 상기한 구현예에서, PK는 숙주 세포에서 아세틸 포스페이트를 생산한다. 온전한 세포질 ACS 활성은, RHR2 및/또는 HOR2-축매화된 아세틸 포스페이트의 가수분해의 결과로서 축적되는, 아세테이트를 아세틸-CoA로 변환할 수 있다.

[0109] **5.6 이소프레노이드를 생산하기 위한 MEV 경로**

[0110] 일부 구현예들에서, 본원에 제시된 유전자 변형된 숙주 세포는 MEV 경로의 하나 이상의 이중의 효소를 포함한다. 일부 구현예들에서, MEV 경로의 하나 이상의 효소는, 아세틸-CoA를 말로닐-CoA와 축합하여 아세토아세틸-CoA를 형성하는 효소를 포함한다. 일부 구현예들에서, MEV 경로의 하나 이상의 효소는 아세틸-CoA 2분자를 축합하여 아세토아세틸-CoA를 형성하는 효소를 포함한다. 일부 구현예들에서, MEV 경로의 하나 이상의 효소는 아세토아세틸-CoA를 아세틸-CoA와 축합하여 HMG-CoA를 형성하는 효소를 포함한다. 일부 구현예들에서, MEV 경로의 하나 이상의 효소는 HMG-CoA를 메발로네이트로 변환하는 효소를 포함한다. 일부 구현예들에서, MEV 경로의 하나 이상의 효소는 메발로네이트를 메발로네이트 5-포스페이트로 인산화하는 효소를 포함한다. 일부 구현예들에서, MEV 경로의 하나 이상의 효소는 메발로네이트 5-포스페이트를 메발로네이트 5-피로포스페이트로 변환하는 효소를 포함한다. 일부 구현예들에서, MEV 경로의 하나 이상의 효소는 메발로네이트 5-피로포스페이트를 이소펜테닐 피로포스페이트로 변환하는 효소를 포함한다.

[0111] 일부 구현예들에서, MEV 경로의 하나 이상의 효소는 아세틸-CoA 티올라제, 아세토아세틸-CoA 신테타제, HMG-CoA 신타제, HMG-CoA 리덕타제, 메발로네이트 키나제, 포스포메발로네이트 키나제 및 메발로네이트 피로포스페이트 데카르복실라제로 이루어진 군으로부터 선택된다. 일부 구현예들에서, 아세토아세틸-CoA의 형성을 촉매할 수 있는 MEV 경로 효소와 관련하여, 유전자 변형된 숙주 세포는, 아세틸-CoA 2 분자를 축합하여 아세토아세틸-CoA를 형성하는 효소, 예컨대 아세틸-CoA 티올라제; 또는 아세틸-CoA를 말로닐-CoA와 축합하여 아세토아세틸-CoA를 형성하는 효소, 예컨대 아세토아세틸-CoA 신타제 중 하나를 포함한다. 일부 구현예들에서, 유전자 변형된 숙주 세포는 아세틸-CoA 2 분자를 축합하여 아세토아세틸-CoA를 형성하는 효소, 예로, 아세틸-CoA 티올라제; 및 아세틸-CoA를 말로닐-CoA와 축합하여 아세토아세틸-CoA를 형성하는 효소, 예컨대 아세토아세틸-CoA 신타제를 둘다 포함한다.

[0112] 일부 구현예들에서, 숙주 세포는 MEV 경로의 1개 보다 많은 수의 효소를 코딩하는 하나 이상의 이중의 뉴클레오티드를 포함한다. 일부 구현예들에서, 숙주 세포는 MEV 경로의 2종의 효소를 코딩하는 하나 이상의 이중의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 숙주 세포는 HMG-CoA를 메발로네이트로 변환할 수 있는 효소 및 메발로네이트를 메발로네이트 5-포스페이트로 변환할 수 있는 효소를 코딩하는 하나 이상의 이중의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 숙주 세포는 MEV 경로의 3종의 효소를 코딩하는 하나 이상의 이중의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 숙주 세포는 MEV 경로의 4종의 효소를 코딩하는 하나 이상의 이중의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 숙주 세포는 MEV 경로의 5종의 효소를 코딩하는 하나 이상의 이중의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 숙주 세포는 MEV 경로의 6종의 효소를 코딩하는 하나 이상의 이중의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 숙주 세포는 MEV 경로의 7종의 효소를 코딩하는 하나 이상의 이중의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 숙주 세포는 MEV 경로의 모든 효소를 코딩하는 복수의 이중의 뉴클레오티드 서열을 포함한다.

[0113] 일부 구현예들에서, 유전자 변형된 숙주 세포는 이소펜테닐 피로포스페이트 (IPP)를 다이메틸알릴 피로포스페이트 (DMAPP)로 변환할 수 있는 효소를 코딩하는 이중의 핵산을 더 포함한다. 일부 구현예에서, 유전자 변형된

숙주 세포는 IPP 및/또는 DMAPP 분자를 축합하여 폴리프레닐 화합물을 형성할 수 있는 효소를 코딩하는 이중의 핵산을 더 포함한다. 일부 구현예들에서, 유전자 변형된 숙주 세포는 IPP 또는 폴리프레닐을 변형시켜 이소프레노이드 화합물을 형성할 수 있는 효소를 코딩하는 이중의 핵산을 추가로 포함한다.

[0114] **5.6.1 아세틸-CoA의 아세토아세틸-CoA로의 변환**

[0115] 일부 구현예들에서, 유전자 변형된 숙주 세포는 아세틸-코엔자임 A 2분자를 축합하여 아세토아세틸-CoA를 형성하는 효소, 예컨대 아세틸-CoA 티올라제를 코딩하는 이중의 뉴클레오티드 서열을 포함한다. 이러한 효소를 코딩하는 뉴클레오티드 서열의 예시적인 예로는, 비제한적으로, (NC\_000913 REGION: 2324131.2325315; 에스케리키아 콜라이 (*Escherichia coli*)), (D49362; 파라코쿠스 데니트리피칸스 (*Paracoccus denitrificans*)), 및 (L20428; 사카로마이세스 세레비지에 (*Saccharomyces cerevisiae*))를 포함한다.

[0116] 아세틸-CoA 티올라제는 아세틸-CoA 2 분자의 아세토아세틸-CoA로의 가역적인 축합을 촉매하지만, 이 반응은 열역학적으로는 비우호적이며; 아세토아세틸-CoA 티올라이시스(thiolysis)가 아세토아세틸-CoA 합성 보다 우호적이다. 아세토아세틸-CoA 신타제 (AACS) (이명, 아세틸-CoA:말로닐-CoA 아세틸트랜스퍼라제; EC 2.3.1.194)는 아세틸-CoA를 말로닐-CoA와 축합하여 아세토아세틸-CoA를 형성한다. 아세틸-CoA 티올라제와는 반대로, AACS-촉매화된 아세토아세틸-CoA 합성은, 조합된 말로닐-CoA의 탈카르복시화로 인해, 기본적으로 에너지-우호적인 반응이다. 아울러, AACS는 또한 아세토아세틸-CoA에 대해서는 티올라이시스 활성을 나타내지 않으므로, 이 반응은 비가역적이다.

[0117] 아세틸-CoA 티올라제와 이중의 ADA 및/또는 포스포트랜스아세틸라제 (PTA)를 포함하는 숙주 세포에서, 아세토아세틸-CoA 티올라이시스에 우호적인, 아세틸-CoA 티올라제에 의해 촉매되는 가역적인 반응으로, 대규모 아세틸-CoA 풀이 만들어질 수 있다. 가역적인 ADA 활성 측면에서, 이러한 아세틸-CoA 풀은 아세틸-CoA를 아세트알데하이드로 변환하는 가역적인 반응쪽으로 ADA를 향하게 할 수 있으며, 이로써 ADA에 의해 제공되는 아세틸-CoA가 생산되는 방향으로 가게 하는 이점이 없어질 수 있다. 마찬가지로, PTA의 활성화도 가역적이며, 따라서 대규모 아세틸-CoA 풀은 아세틸-CoA를 아세틸 포스페이트로 변환하는 역 반응쪽으로 PTA를 향하게 할 수 있다. 따라서, 일부 구현예에서, 아세틸-CoA를 ADA 및 PTA의 정방향 반응으로 가게 하는 강력한 견인력을 제공하기 위해, 본원에 제시된 유전자 변형된 숙주 세포의 MEV 경로는 아세토아세틸-CoA 신타제를 이용하여 아세틸-CoA와 말로닐-CoA로부터 아세토아세틸-CoA를 합성한다.

[0118] 일부 구현예들에서, AACS는 스트렙토마이세스 sp. strain CL190로부터 유래된다 (Okamura *et al.*, *Proc Natl Acad Sci USA* 107(25):11265-70 (2010)). 스트렙토마이세스 sp. strain CL190의 대표적인 AACS 뉴클레오티드 서열은 등재번호 AB540131.1과 본원에 제시된 서열번호 19를 포함한다. 스트렙토마이세스 sp. strain CL190의 대표적인 AACS 단백질 서열은 등재번호 D7URV0, BAJ10048 및 본원에 제시된 서열번호 20을 포함한다. 본원에 제공되는 조성물 및 방법에 이용가능한 다른 아세토아세틸-CoA 신타제로는, 비제한적으로, 스트렙토마이세스 sp. (AB183750; KO-3988 BAD86806); 스트렙토마이세스 아놀라쿠스 (*S. anulatus*) strain 9663 (FN178498; CAX48662); 스트렙토마이세스 sp. KO-3988 (AB212624; BAE78983); 액티노플란스 (*Actinoplanes*) sp. A40644 (AB113568; BAD07381); 스트렙토마이세스 sp. C (NZ\_ACEW010000640; ZP\_05511702); 노카르디오피스 다손빌레이 (*Nocardiosis dassonvillei*) DSM 43111 (NZ\_ABUI01000023; ZP\_04335288); 마이코박테리움 얼세란스 (*Mycobacterium ulcerans*) Agv99 (NC\_008611; YP\_907152); 마이코박테리움 마리눔 (*Mycobacterium marinum*) M (NC\_010612; YP\_001851502); 스트렙토마이세스 sp. Mg1 (NZ\_DS570501; ZP\_05002626); 스트렙토마이세스 sp. AA4 (NZ\_ACEV01000037; ZP\_05478992); 스트렙토마이세스 로세오스포루스 (*S. roseosporus*) NRRL 15998 (NZ\_ABYB01000295; ZP\_04696763); 스트렙토마이세스 sp. ACTE (NZ\_ADFD01000030; ZP\_06275834); 스크렙토마이세스 비리도크로모게네스 (*S. viridochromogenes*) DSM 40736 (NZ\_ACEZ01000031; ZP\_05529691); 프랭키아 (*Frankia*) sp. Cc13 (NC\_007777; YP\_480101); 노카르디아 브라질렌시스 (*Nocardia brasiliensis*) (NC\_018681; YP\_006812440.1); 및 오스트윅키아 첼로네 (*Austwickia chelonae*) (NZ\_BAGZ01000005; ZP\_10950493.1)를 포함한다. 다른 적절한 아세토아세틸-CoA 신타제로는 미국 특허 공개공보 2010/0285549와 2011/0281315에 기술된 것을 포함하며, 그 내용은 그 전체가 원용에 의해 본 명세서에 포함된다.

[0119] 또한, 본원에 제공되는 조성물 및 방법에 사용가능한 아세토아세틸-CoA 신타제는, 본원에 기술된 임의의 아세토아세틸-CoA 신타제의 유도체로 지칭되는 분자를 포함한다. 이러한 "유도체"는 다음과 같은 특징을 가진다: (1) 이것은 본원에 기술된 임의의 아세토아세틸-CoA 신타제와 상당한 상동성을 공유함; 및 (2) 아세틸-CoA와 말로닐-CoA를 비가역적으로 축합하여 아세토아세틸-CoA를 합성하는 반응을 촉매할 수 있음. 아세토아세틸-CoA 신타제의 유도체는, 유도체의 아미노산 서열이 아세토아세틸-CoA 신타제의 서열과 80% 이상, 더 바람직하게는 90% 이

상, 가장 바람직하게는 95% 이상 동일하다면, 아세토아세틸-CoA 신타제와 "상당한 상동성"을 공유한다고 언급된다.

[0120] 5.6.2 아세토아세틸-CoA의 HMG-CoA로의 변환

[0121] 일부 구현예들에서, 숙주 세포는, 아세토아세틸-CoA를 다른 아세틸-CoA 분자와 축합하여 3-하이드록시-3-메틸글루타릴-CoA (HMG-CoA)를 형성할 수 있는 효소, 예컨대 HMG-CoA 신타제를 코딩하는 이중의 뉴클레오티드 서열을 포함한다. 이러한 효소를 코딩하는 뉴클레오티드 서열의 예시적인 예로는, 비제한적으로, (NC\_001145, complement 19061.20536; 사카로마이세스 세레비지애 (*Saccharomyces cerevisiae*)), (X96617; 사카로마이세스 세레비지애 (*Saccharomyces cerevisiae*)), (X83882; 아라비돕시스 탈리아나 (*Arabidopsis thaliana*)), (AB037907; 키타사토스포라 그리세올라 (*Kitasatospora griseola*)), (BT007302; 호모 사피엔스 (*Homo sapiens*)), 및 (NC\_002758, Locus tag SAV2546, GeneID 1122571; 스태필로코커스 아우레우스 (*Staphylococcus aureus*))를 포함한다.

[0122] 5.6.3 HMG-CoA의 메발로네이트로의 변환

[0123] 일부 구현예들에서, 숙주 세포는 HMG-CoA를 메발로네이트로 변환할 수 있는 효소, 예컨대 HMG-CoA 리덕타제를 코딩하는 이중의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, HMG-CoA 리덕타제는 NADH-이용성 하이드록시메틸글루타릴-CoA 리덕타제-CoA 리덕타제이다. HMG-CoA 리덕타제 (EC 1.1.1.34; EC 1.1.1.88)는 (S)-HMG-CoA의 (R)-메발로네이트로의 환원적인 탈아실화를 촉매하며, 이는 클래스 I과 클래스 II HMGr 2종의 클래스로 분류될 수 있다. 클래스 I은 진핵생물과 대부분의 고세균 유래 효소를 포함하며, 클래스 II는 일부 원핵생물과 고세균의 HMG-CoA 리덕타제를 포함한다. 서열 다양성 외에도, 2가지 클래스의 효소는 또한 이들의 조인자 특이성에 차이가 있다. NADPH만을 이용하는 클래스 I 효소와는 달리, 클래스 II HMG-CoA 리덕타제는 NADPH와 NADH를 구별하는 능력이 다양하다. 예로, Hedl *et al.*, *Journal of Bacteriology* 186 (7): 1927-1932 (2004)를 참조한다. 클래스 II HMG-CoA 리덕타제의 조인자 선택 특이성은 아래에 나타낸다.

[0124] 표 1. 클래스 II HMG-CoA 리덕타제의 조인자 선택 특이성

표 1

소스	코엔자임 특이성	$K_m^{NADPH}$ ( $\mu M$ )	$K_m^{NADH}$ ( $\mu M$ )
슈도모나스 메발로니 ( <i>P. mevalonii</i> )	NADH		80
아르케오글로부스 풀기두스 ( <i>A. fulgidus</i> )	NAD(P)H	500	160
스타필로코커스 아우레우스 ( <i>S. aureus</i> )	NAD(P)H	70	100
엔테로코커스 페칼리스 ( <i>E. faecalis</i> )	NADPH	30	

[0126] 본원에 제시된 조성물 및 방법에서 사용가능한 HMG-CoA 리덕타제는 조인자로서 NADH를 이용할 수 있는 HMG-CoA 리덕타제, 예컨대 슈도모나스 메발로니 (*P. mevalonii*), 아르케오글로부스 풀기두스 (*A. fulgidus*) 또는 스태필로코커스 아우레우스 (*S. aureus*) 유래의 HMG-CoA 리덕타제를 포함한다. 특정 구현예에서, HMG-CoA 리덕타제는 조인자로서 NADH만을 이용하며, 예컨대, 슈도모나스 메발로니 (*P. mevalonii*), 실리시박터 포메로이 (*S. pomeroyi*) 또는 델프티아 액시도보란스 (*D. acidovorans*) 유래의 HMG-CoA 리덕타제이다.

[0127] 일부 구현예들에서, NADH-이용성 HMG-CoA 리덕타제는 슈도모나스 메발로니 (*Pseudomonas mevalonii*)로부터 유래된다. HMG-CoA 리덕타제 (E.C. 1.1.1.88)를 코딩하는 슈도모나스 메발로니 (*Pseudomonas mevalonii*)의 야생형 *mvaA* 유전자 서열은 이전에 개시된 바 있다. Beach 및 Rodwell, *J. Bacteriol.* 171:2994-3001 (1989)을 참조한다. 슈도모나스 메발로니 (*Pseudomonas mevalonii*)의 대표적인 *mvaA* 뉴클레오티드 서열은 등재번호 M24015, 및 본원에 제시된 서열번호 21을 포함한다. 슈도모나스 메발로니 (*Pseudomonas mevalonii*)의 대표적인 HMG-CoA 리덕타제 단백질 서열은 등재번호 AAA25837, P13702, MVAA\_PSEMV 및 본원에 제시된 서열번호 22를 포함한다.

[0128] 일부 구현예들에서, NADH-이용성 HMG-CoA 리덕타제는 실리시박터 포메로이 (*Silicibacter pomeroyi*)로부터 유래된다. 실리시박터 포메로이 (*Silicibacter pomeroyi*)의 대표적인 HMG-CoA 리덕타제 뉴클레오티드 서열은 등재번호 NC\_006569.1, 및 본원에 제시된 서열번호 23을 포함한다. 실리시박터 포메로이 (*Silicibacter pomeroyi*)의 대표적인 HMG-CoA 리덕타제 단백질 서열은 등재번호 YP\_164994 및 본원에 제시된 서열번호 24를 포함한다.

[0129] 일부 구현예들에서, NADH-이용성 HMG-CoA 리덕타제는 델프티아 액시도보란스 (*Delftia acidovorans*)로부터 유래

된다. 델프트리아 액시도보란스 (*Delftia acidovorans*)의 대표적인 HMG-CoA 리덕타제 뉴클레오티드 서열은 NC\_010002 REGION: complement(319980..321269), 및 본원에 제시된 서열번호 25를 포함한다. 델프트리아 액시도보란스 (*Delftia acidovorans*)의 대표적인 HMG-CoA 리덕타제 단백질 서열은 등재번호 YP\_001561318 및 본원에 제시된 서열번호 26을 포함한다.

[0130] 일부 구현예들에서, NADH-이용성 HMG-CoA 리덕타제는 솔라눔 투베로숨 (*Solanum tuberosum*)으로부터 유래된다 (Crane et al., *J. Plant Physiol.* 159:1301-1307 (2002)).

[0131] 또한, 본원에 제공되는 조성물 및 방법에 사용가능한 NADH-이용성 HMG-CoA 리덕타제는, 본원에 언급된 임의의 NADH-이용성 HMG-CoA 리덕타제, 예컨대 슈도모나스 메발로니 (*P. mevalonii*), 실리시박터 포메로이 (*S. pomeroyi*) 및 델프트리아 액시도보란스 (*D. acidovorans*)로부터 유래된 NADH-이용성 HMG-CoA 리덕타제에 대한 "유도체"로 지칭되는 분자를 포함한다. 이러한 "유도체"는 다음과 같은 특징을 가진다: (1) 이는 본원에 기술된 임의의 NADH-이용성 HMG-CoA 리덕타제와 상당한 상동성을 공유함; 및 (2) 조인자로서 선호적으로 NADH를 이용하면서 (S)-HMG-CoA의 (R)-메발로네이트로의 환원적인 탈아실화를 촉매할 수 있음. NADH-이용성 HMG-CoA 리덕타제의 유도체는, 유도체의 아미노산 서열이 NADH-이용성 HMG-CoA 리덕타제의 서열과 80% 이상, 더 바람직하게는 90% 이상, 가장 바람직하게는 95% 이상 동일하다면, NADH-이용성 HMG-CoA 리덕타제와 "상당한 상동성"을 공유한다고 지칭된다.

[0132] 본원에서, "NADH-이용성"이라는 표현은, 예를 들어, NADPH 보다 NADH에 대해 높은 비활성 (specific activity)을 나타냄으로써, NADH-이용성 HMG-CoA 리덕타제가 조인자로서 NADPH에 비해 NADH에 대해 선택적이라는 것을 의미한다. 일부 구현예들에서, 조인자로서 NADH에 대한 선택성은  $k_{cat}^{(NADH)} / k_{cat}^{(NADPH)}$  비로 표시된다. 일부 구현예들에서, NADH-이용성 HMG-CoA 리덕타제는  $k_{cat}^{(NADH)} / k_{cat}^{(NADPH)}$  비가 적어도 5, 10, 15, 20, 25이거나, 또는 25보다 높다. 일부 구현예들에서, NADH-이용성 HMG-CoA 리덕타제는 NADH만을 이용한다. 예를 들어, NADH만을 사용하는 NADH-이용성 HMG-CoA 리덕타제는 시험관내에서 유일한 조인자로서 제공되는 NADH를 이용하여 일부 활성을 나타내며, NADPH가 유일한 조인자로서 제공되는 경우 검출가능한 활성은 나타내지 않는다. Kim et al., *Protein Science* 9:1226-1234 (2000); 및 Wilding et al., *J. Bacteriol.* 182(18):5147-52 (2000)에 개시된 방법을 비롯하여, 당해 기술 분야에 공지된 조인자 특이성을 확인하기 위한 임의의 방법을 이용하여, 조인자로서 NADH에 대해 선호성을 가지는 HMG-CoA 리덕타제를 동정할 수 있으며, 이들의 내용은 그 전체가 원용에 의해 본 명세서에 포함된다.

[0133] 일부 구현예들에서, NADH-이용성 HMG-CoA 리덕타제는, 예를 들어 조인자-결합 포켓에 부위 특이적인 돌연변이를 유발함으로써, NADPH에 비해 NADH에 선택적일도록 조작된다. NADH-선택성을 조작하는 방법은 Watanabe et al., *Microbiology* 153:3044-3054 (2007)에 기술되어 있으며, HMG-CoA 리덕타제의 조인자 특이성을 확인하는 방법은 Kim et al., *Protein Sci.* 9:1226-1234 (2000)에 기술되어 있으며, 그 내용은 전체가 원용에 의해 본 명세서에 포함된다.

[0134] 일부 구현예들에서, NADH-이용성 HMG-CoA 리덕타제는, 메발로네이트 분해 경로를 천연적으로 포함하는 숙주 종, 예컨대, 이의 단독 탄소원으로서 메발로네이트를 이화하는 숙주 종으로부터 유래된다. 이들 구현예들에서, 천연 숙주 세포에서 내재화된 (R)-메발로네이트의 (S)-HMG-CoA로의 산화적인 아실화를 정상적으로 촉매하는, NADH-이용성 HMG-CoA 리덕타제를 이용하여, 역 반응 (reverse reaction), 즉, 메발로네이트 생합성 경로를 포함하는 유전자 변형된 숙주 세포에서 (S)-HMG-CoA의 (R)-메발로네이트로의 환원적인 탈아실화를 촉매한다. 단독 탄소원으로서 메발로네이트에서 증식할 수 있는 원핵생물은, Anderson et al., *J. Bacteriol.* 171(12):6468-6472 (1989); Beach et al., *J. Bacteriol.* 171:2994-3001 (1989); Bensch et al., *J. Biol. Chem.* 245:3755-3762; Fimongnari et al., *Biochemistry* 4:2086-2090 (1965); Siddiqi et al., *Biochem. Biophys. Res. Commun.* 8:110-113 (1962); Siddiqi et al., *J. Bacteriol.* 93:207-214 (1967); 및 Takatsuji et al., *Biochem. Biophys. Res. Commun.* 110:187-193 (1983)에 기술되어 있으며, 그 전체 내용이 원용에 의해 본 명세서에 포함된다.

[0135] 본원에 제시된 조성물 및 방법에 대한 일부 구현예들에서, 숙주 세포는 NADH-이용성 HMG-CoA 리덕타제 둘다를 포함한다. NADPH-이용성 HMG-CoA 리덕타제를 코딩하는 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로: (NM\_206548; 드로소필라 멜라노가스터 (*Drosophila melanogaster*)), (NC\_002758, Locus tag SAV2545, GeneID 1122570; 스태필로코커스 아우레우스 (*Staphylococcus aureus*)), (AB015627; 스트렙토마이세스 sp. KO 3988), (AX128213, 절단된 HMG-CoA 리덕타제를 코딩하는 서열 제공; 사카로마이세스 세레

비지아 (*Saccharomyces cerevisiae*)), 및 (NC\_001145: complement (115734.118898; 사카로마이세스 세레비지아 (*Saccharomyces cerevisiae*))를 포함한다.

[0136] **5.6.4 메발로네이트의 메발로네이트-5-포스페이트로의 변환**

[0137] 일부 구현예들에서, 숙주 세포는 메발로네이트를 메발로네이트 5-포스페이트로 변환할 수 있는 효소, 예컨대 메발로네이트 키나제를 코딩하는 이종의 뉴클레오티드 서열을 포함한다. 이러한 효소를 코딩하는 뉴클레오티드 서열의 예시적인 예로는, 비제한적으로: (L77688; 아라비돕시스 탈리아나 (*Arabidopsis thaliana*)) 및 (X55875; 사카로마이세스 세레비지아 (*Saccharomyces cerevisiae*))를 포함한다.

[0138] **5.6.5 메발로네이트-5-포스페이트의 메발로네이트-5-피로포스페이트로의 변환**

[0139] 일부 구현예들에서, 숙주 세포는 메발로네이트 5-포스페이트를 메발로네이트 5-피로포스페이트로 변환할 수 있는 효소, 예컨대 포스포메발로네이트 키나제를 코딩하는 이종의 뉴클레오티드 서열을 포함한다. 이러한 효소를 코딩하는 뉴클레오티드 서열의 예시적인 예로는, 비제한적으로: (AF429385; 헤베아 브라질렌시스 (*Hevea brasiliensis*)), (NM\_006556; 호모 사피엔스 (*Homo sapiens*)), 및 (NC\_001145. complement 712315.713670; 사카로마이세스 세레비지아 (*Saccharomyces cerevisiae*))를 포함한다.

[0140] **5.6.6 메발로네이트-5-피로포스페이트의 IPP로의 변환**

[0141] 일부 구현예들에서, 숙주 세포는 메발로네이트 5-피로포스페이트를 이소펜테닐 다이포스페이트 (IPP)로 변환할 수 있는 효소, 예컨대 메발로네이트 피로포스페이트 데카르복실라제를 코딩하는 이종의 뉴클레오티드 서열을 포함한다. 이러한 효소를 코딩하는 뉴클레오티드 서열의 예시적인 예로는, 비제한적으로, (X97557; 사카로마이세스 세레비지아 (*Saccharomyces cerevisiae*)), (AF290095; 엔테로코커스 패숨 (*Enterococcus faecium*)), 및 (U49260; 호모 사피엔스 (*Homo sapiens*))를 포함한다.

[0142] **5.6.7 IPP의 DMAPP로의 변환**

[0143] 일부 구현예들에서, 숙주 세포는 MEV 경로를 통해 만들어진 IPP를 다이메틸알릴 피로포스페이트 (DMAPP)로 변환할 수 있는 효소, 예컨대 IPP 이소머라제를 코딩하는 이종의 뉴클레오티드 서열을 추가로 포함한다. 이러한 효소를 코딩하는 뉴클레오티드 서열의 예시적인 예로는, 비제한적으로, (NC\_000913, 3031087.3031635; 에스케리키아 콜라이 (*Escherichia coli*)), 및 (AF082326; 헤마토코커스 플루비알리스 (*Haematococcus pluvialis*))를 포함한다.

[0144] **5.6.8 폴리프레닐 신타제**

[0145] 일부 구현예들에서, 숙주 세포는 IPP 및/또는 DMAPP 분자를 축합하여 탄소를 5개 보다 많이 포함하는 폴리프레닐 화합물을 형성할 수 있는 폴리프레닐 신타제를 코딩하는 이종의 뉴클레오티드 서열을 추가로 포함한다.

[0146] 일부 구현예들에서, 숙주 세포는 IPP 한 분자와 DMAPP 1 분자를 축합하여 게라닐 피로포스페이트 ("GPP") 1 분자를 형성할 수 있는 효소, 예로 GPP 신타제를 코딩하는 이종의 뉴클레오티드 서열을 포함한다. 이러한 효소를 코딩하는 뉴클레오티드 서열의 예시적인 예로는, 비제한적으로, (AF513111; 아비에스 그란디스 (*Abies grandis*)), (AF513112; 아비에스 그란디스 (*Abies grandis*)), (AF513113; 아비에스 그란디스 (*Abies grandis*)), (AY534686; 안티리눔 마주스 (*Antirrhinum majus*)), (AY534687; 안티리눔 마주스 (*Antirrhinum majus*)), (Y17376; 아라비돕시스 탈리아나 (*Arabidopsis thaliana*)), (AE016877, Locus AP11092; 바실러스 세레우스 (*Bacillus cereus*); ATCC 14579), (AJ243739; 시트러스 시넨시스 (*Citrus sinensis*)), (AY534745; 클라르키아 브루어리 (*Clarkia breweri*)), (AY953508; 입스 피니 (*Ips pini*)), (DQ286930; 라이코퍼시콘 에스쿨렌툼 (*Lycopersicon esculentum*)), (AF182828; 멘타 x 피페리타 (*Mentha x piperita*)), (AF182827; 멘타 x 피페리타 (*Mentha x piperita*)), (MPI249453; 멘타 x 피페리타 (*Mentha x piperita*)), (PZE431697, Locus CAD24425; 파라코커스 제아산티니팍시엔스 (*Paracoccus zeaxanthinifaciens*)), (AY866498; 피크로리자 쿠로아 (*Picrorhiza kurrooa*)), (AY351862; 비티스 비니페라 (*Vitis vinifera*)), 및 (AF203881, Locus AAF12843; 자이모모나스 모빌리스 (*Zymomonas mobilis*))를 포함한다.

[0147] 일부 구현예들에서, 숙주 세포는 IPP 2 분자를 DMAPP 1 분자와 축합하거나, 또는 IPP 분자를 GPP 분자에 부가하여, 파르네실 피로포스페이트 ("FPP") 분자를 형성할 수 있는 효소, 예로, FPP 신타제를 코딩하는 이종의 뉴클레오티드 서열을 포함한다. 이러한 효소를 코딩하는 뉴클레오티드 서열의 예시적인 예로는, 비제한적으로, (ATU80605; 아라비돕시스 탈리아나 (*Arabidopsis thaliana*)), (ATHFPS2R; 아라비돕시스 탈리아나 (*Arabidopsis thaliana*)), (AAU36376; 아르테미시아 아누아 (*Artemisia annua*)), (AF461050; *Bos taurus*), (D00694; 에스케

리키아 콜라이 (*Escherichia coli*) K-12), (AE009951, Locus AAL95523; 푸조박테리움 뉴클레아툼 아종 뉴클레아툼 (*Fusobacterium nucleatum subsp. nucleatum*) ATCC 25586), (GFFPPSGEN; 지베렐라 푸이쿠로이 (*Gibberella fujikuroi*)), (CP000009, Locus AAW60034; 글루코노박터 옥시단스 (*Gluconobacter oxydans*) 621H), (AF019892; 헬리안투스 안누스 (*Helianthus annuus*)), (HUMFAPS; 호모 사피엔스 (*Homo sapiens*)), (KLPFPPSQCR; 클루이베로마이세스 락티스 (*Cluyveromyces lactis*)), (LAU15777; 루피누스 알부스 (*Lupinus albus*)), (LAU20771; 루피누스 알부스 (*Lupinus albus*)), (AF309508; 무스 무스쿨러스 (*Mus musculus*)), (NCFPPSGEN; 뉴로스포라 크라사 (*Neurospora crassa*)), (PAFPS1; 파르테눔 아르겐타툼 (*Parthenium argentatum*)), (PAFPS2; 파르테눔 아르겐타툼 (*Parthenium argentatum*)), (RATFAPS; 라투스 노르베기쿠스 (*Rattus norvegicus*)), (YSCFPP; 사카로마이세스 세레비지애 (*Saccharomyces cerevisiae*)), (D89104; 스킴조사 카로마이세스 폼베 (*Schizosaccharomyces pombe*)), (CP000003, Locus AAT87386; 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*)), (CP000017, Locus AAZ51849; 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*)), (NC\_008022, Locus YP\_598856; 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) MGAS10270), (NC\_008023, Locus YP\_600845; 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) MGAS2096), (NC\_008024, Locus YP\_602832; 스트렙토코커스 피오게네스 (*Streptococcus pyogenes*) MGAS10750), (MZEFP; 지 메이스 (*Zea mays*)), (AE000657, Locus AAC06913; 아퀴팩스 에어올리쿠스 (*Aquifex aeolicus*) VF5), (NM\_202836; 아라비돕시스 탈리아나 (*Arabidopsis thaliana*)), (D84432, Locus BAA12575; 바실러스 셉틸리스 (*Bacillus subtilis*)), (U12678, Locus AAC28894; 브래디리조븀 자포니쿰 (*Bradyrhizobium japonicum*) USDA 110), (BACFDPS; 지오바실러스 스테아로서모필러스 (*Geobacillus stearothermophilus*)), (NC\_002940, Locus NP\_873754; 헤모필러스 두크레이 (*Haemophilus ducreyi*) 35000HP), (L42023, Locus AAC23087; 헤모필러스 인플루엔자 (*Haemophilus influenzae*) Rd KW20), (J05262; 호모 사피엔스 (*Homo sapiens*)), (YP\_395294; 락토바실러스 사케이 아종 사케이 (*Lactobacillus sakei subsp. sakei*) 23K), (NC\_005823, Locus YP\_000273; 렙토스피라 인테로간스 혈청형 코펜하게니 (*Leptospira interrogans serovar Copenhageni*) str. Fiocruz L1-130), (AB003187; 마이크로코커스 루테우스 (*Micrococcus luteus*)), (NC\_002946, Locus YP\_208768; 네이세리아 고노로이애 (*Neisseria gonorrhoeae*) FA 1090), (U00090, Locus AAB91752; 리조븀 (*Rhizobium*) sp. NGR234), (J05091; 사카로마이세스 세레비지애 (*Saccharomyces cerevisiae*)), (CP000031, Locus AAV93568; 실리시박터 포메로이 (*Silicibacter pomeroyi*) DSS-3), (AE008481, Locus AAK99890; 스트렙토코커스 뉴모니아 (*Streptococcus pneumoniae*) R6), 및 (NC\_004556, Locus NP\_779706; 크실렐라 파스티도사 (*Xylella fastidiosa*) Temecula1)를 포함한다.

[0148]

일부 구현예들에서, 숙주 세포는 IPP를 DMAPP와, 또는 IPP를 FPP와 조합하여 게라닐게라닐 피로포스페이트 ("GGPP")를 형성할 수 있는 효소를 코딩하는 이중의 뉴클레오티드 서열을 추가로 포함한다. 이러한 효소를 코딩하는 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로, (ATHGERPYRS; 아라비돕시스 탈리아나 (*Arabidopsis thaliana*)), (BT005328; 아라비돕시스 탈리아나 (*Arabidopsis thaliana*)), (NM\_119845; 아라비돕시스 탈리아나 (*Arabidopsis thaliana*)), (NZ\_AAJM01000380, Locus ZP\_00743052; 바실러스 투린지엔시스 혈청형 이스라엘렌시스 (*Bacillus thuringiensis serovar israelensis*), ATCC 35646 sq1563), (CRGGPPS; 카타란투스 로세우스 (*Catharanthus roseus*)), (NZ\_AABF02000074, Locus ZP\_00144509; 푸조박테리움 뉴클레아툼 아종 빈센티 (*Fusobacterium nucleatum subsp. vincentii*), ATCC 49256), (GFGGPPSGN; 지베렐라 푸이쿠로이 (*Gibberella fujikuroi*)), (AY371321; 장코 빌로마 (*Ginkgo biloba*)), (AB055496; 헤베아 브라질렌시스 (*Hevea brasiliensis*)), (AB017971; 호모 사피엔스 (*Homo sapiens*)), (MCI276129; 무코르 시르시넬로이데스 f. 루시타니쿠스 (*Mucor circinelloides f. lusitanicus*)), (AB016044; 무스 무스쿨러스 (*Mus musculus*)), (AABX01000298, Locus NCU01427; 뉴로스포라 크라사 (*Neurospora crassa*)), (NCU20940; 뉴로스포라 크라사 (*Neurospora crassa*)), (NZ\_AAKL01000008, Locus ZP\_00943566; 랄스토니아 솔라나세아룸 (*Ralstonia solanacearum*) UW551), (AB118238; 라투스 노르베기쿠스 (*Rattus norvegicus*)), (SCU31632; 사카로마이세스 세레비지애 (*Saccharomyces cerevisiae*)), (AB016095; 시네코코쿠스 엘롱가테스 (*Synechococcus elongates*)), (SAGGPPS; 시나피스 알바 (*Sinapis alba*)), (SSOGDS; 설폴로부스 액시도칼다리우스 (*Sulfolobus acidocaldarius*)), (NC\_007759, Locus YP\_461832; *Syntrophus aciditrophicus* SB), (NC\_006840, Locus YP\_204095; 비브리오 피세리 (*Vibrio fischeri*) ES114), (NM\_112315; 아라비돕시스 탈리아나 (*Arabidopsis thaliana*)), (ERWCRTE; 판토에아 어글로머란스 (*Pantoea agglomerans*)), (D90087, Locus BAA14124; 판토에아 아나나티스 (*Pantoea ananatis*)), (X52291, Locus CAA36538; 로도박터 캡슐라투스 (*Rhodobacter capsulatus*)), (AF195122, Locus AAF24294; 로도박터 스페어로이데스 (*Rhodobacter sphaeroides*)), 및 (NC\_004350, Locus NP\_721015; 스트렙토코커스 무탄스 (*Streptococcus mutans*) UA159)를 포함한다.

[0149]

### 5.6.9 테르펜 신타제

- [0150] 일부 구현예들에서, 숙주 세포는 폴리프레닐을 변형시켜, 헤미테르펜, 모노테르펜, 세스퀴테르펜, 다이테르펜, 트리테르펜, 테트라테르펜, 폴리테르펜, 스테로이드 화합물, 카로테노이드 또는 변형된 이소프레노이드 화합물을 형성할 수 있는 효소를 코딩하는 이종의 뉴클레오티드 서열을 추가로 포함한다.
- [0151] 일부 구현예들에서, 이종의 뉴클레오티드는 카렌 신타제를 코딩한다. 적합한 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로, (AF461460, REGION 43.1926; 피세아 아비에스 (*Picea abies*)) 및 (AF527416, REGION: 78.1871; 살비아 스테노필라 (*Salvia stenophylla*))를 포함한다.
- [0152] 일부 구현예들에서, 이종의 뉴클레오티드는 케라니올 신타제를 코딩한다. 적합한 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로, (AJ457070; 신나모뎀 테누이필럼 (*Cinnamomum tenuipilum*)), (AY362553; 오시뎀 바실리쿰 (*Ocimum basilicum*)), (DQ234300; 펠릴라 프루테센스 (*Perilla frutescens*) strain 1864), (DQ234299; 펠릴라 시트리오도라 (*Perilla citriodora*) strain 1861), (DQ234298; 펠릴라 시트리오도라 (*Perilla citriodora*) strain 4935), 및 (DQ088667; 펠릴라 시트리오도라 (*Perilla citriodora*))를 포함한다.
- [0153] 일부 구현예들에서, 이종의 뉴클레오티드는 리날롤 신타제를 코딩한다. 적합한 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로, (AF497485; 아라비둡시스 탈리아나 (*Arabidopsis thaliana*)), (AC002294, Locus AAB71482; 아라비둡시스 탈리아나 (*Arabidopsis thaliana*)), (AY059757; 아라비둡시스 탈리아나 (*Arabidopsis thaliana*)), (NM\_104793; 아라비둡시스 탈리아나 (*Arabidopsis thaliana*)), (AF154124; 아르테미시아 아누아 (*Artemisia annua*)), (AF067603; 클라르키아 브루어리 (*Clarkia breweri*)), (AF067602; 클라르키아 콘시나 (*Clarkia concinna*)), (AF067601; 클라르키아 브루어리 (*Clarkia breweri*)), (U58314; 클라르키아 브루어리 (*Clarkia breweri*)), (AY840091; 라이코퍼시콘 에스쿨렌툼 (*Lycopersicon esculentum*)), (DQ263741; 라반둘라 안구스티폴리아 (*Lavandula angustifolia*)), (AY083653; 멘타 사이트레이트 (*Mentha citrate*)), (AY693647; 오시뎀 바실리쿰 (*Ocimum basilicum*)), (XM\_463918; 오리자 사티바 (*Oryza sativa*)), (AP004078, Locus BAD07605; 오리자 사티바 (*Oryza sativa*)), (XM\_463918, Locus XP\_463918; 오리자 사티바 (*Oryza sativa*)), (AY917193; 펠릴라 시트리오도라 (*Perilla citriodora*)), (AF271259; 펠릴라 프루테센스 (*Perilla frutescens*)), (AY473623; 피세아 아비에스 (*Picea abies*)), (DQ195274; 피세아 시첸시스 (*Picea sitchensis*)), 및 (AF444798; 펠릴라 프루테센스 var. 크리스파 (*Perilla frutescens* var. *crispa*) cultivar No. 79)를 포함한다.
- [0154] 일부 구현예들에서, 이종의 뉴클레오티드는 리모넨 신타제를 코딩한다. 적합한 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로, (+)-리모넨 신타제 (AF514287, REGION: 47.1867; 시트러스 리몬 (*Citrus limon*)) 및 AY055214 (REGION: 48.1889; 아가스타케 루고사 (*Agastache rugosa*)) 및 (-)-리모넨 신타제 (DQ195275, REGION: 1.1905; 피세아 시첸시스 (*Picea sitchensis*)), (AF006193, REGION: 73.1986; 아비에스 그란디스 (*Abies grandis*)), 및 (MHC4SLSP, REGION: 29.1828; 멘타 스피카타 (*Mentha spicata*))를 포함한다.
- [0155] 일부 구현예들에서, 이종의 뉴클레오티드는 미르센 신타제를 코딩한다. 적합한 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로, (U87908; 아비에스 그란디스 (*Abies grandis*)), (AY195609; 안티리눔 마주스 (*Antirrhinum majus*)), (AY195608; 안티리눔 마주스 (*Antirrhinum majus*)), (NM\_127982; 아라비둡시스 탈리아나 (*Arabidopsis thaliana*) TPS10), (NM\_113485; 아라비둡시스 탈리아나 (*Arabidopsis thaliana*) ATTPS-CIN), (NM\_113483; 아라비둡시스 탈리아나 (*Arabidopsis thaliana*) ATTPS-CIN), (AF271259; 펠릴라 프루테센스 (*Perilla frutescens*)), (AY473626; 피세아 아비에스 (*Picea abies*)), (AF369919; 피세아 아비에스 (*Picea abies*)), 및 (AJ304839; 퀘르쿠스 일렉스 (*Quercus ilex*))를 포함한다.
- [0156] 일부 구현예들에서, 이종의 뉴클레오티드는 옥시멘 신타제를 코딩한다. 적합한 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로, (AY195607; 안티리눔 마주스 (*Antirrhinum majus*)), (AY195609; 안티리눔 마주스 (*Antirrhinum majus*)), (AY195608; 안티리눔 마주스 (*Antirrhinum majus*)), (AK221024; 아라비둡시스 탈리아나 (*Arabidopsis thaliana*)), (NM\_113485; 아라비둡시스 탈리아나 (*Arabidopsis thaliana*) ATTPS-CIN), (NM\_113483; 아라비둡시스 탈리아나 (*Arabidopsis thaliana*) ATTPS-CIN), (NM\_117775; 아라비둡시스 탈리아나 (*Arabidopsis thaliana*) ATTPS03), (NM\_001036574; 아라비둡시스 탈리아나 (*Arabidopsis thaliana*) ATTPS03), (NM\_127982; 아라비둡시스 탈리아나 (*Arabidopsis thaliana*) TPS10), (AB110642; 시트러스 운시우 (*Citrus unshiu*) CitMTSL4) 및 (AY575970; 로터스 코르니쿨라투스 var. 자포니쿠스 (*Lotus corniculatus* var. *japonicus*))를 포함한다.
- [0157] 일부 구현예들에서, 이종의 뉴클레오티드는  $\alpha$ -피넨 신타제를 코딩한다. 적합한 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로, (+)  $\alpha$ -피넨 신타제 (AF543530, REGION: 1.1887; 피누스 타에다 (*Pinus taeda*))를 포함한다.

taeda)), (-) α-피넨 신타제 (AF543527, REGION: 32.1921; 피누스 타에다 (*Pinus taeda*)), 및 (+)/(-) α-피넨 신타제 (AGU87909, REGION: 6111892; 아비에스 그란디스 (*Abies grandis*))를 포함한다.

- [0158] 일부 구현예들에서, 이종의 뉴클레오티드는 β-피넨 신타제를 코딩한다. 적합한 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로, (-) β-피넨 신타제 (AF276072, REGION: 1.1749; 아르테미시아 아누아 (*Artemisia annua*)) 및 (AF514288, REGION: 26.1834; 시트러스 리몬 (*Citrus limon*))을 포함한다.
- [0159] 일부 구현예들에서, 이종의 뉴클레오티드는 사비넨 신타제를 코딩한다. 적합한 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로, 살비아 오피시날리스 (*Salvia officinalis*) 유래 AF051901, REGION: 26.1798을 포함한다.
- [0160] 일부 구현예들에서, 이종의 뉴클레오티드는 γ-테르피넨 신타제를 코딩한다. 적합한 뉴클레오티드 서열에 대한 예시적인 예로는: (AF514286, REGION: 30.1832, 시트러스 리몬 (*Citrus limon*)) 및 (AB110640, REGION 1.1803, 시트러스 운시우 (*Citrus unshiu*))를 포함한다.
- [0161] 일부 구현예들에서, 이종의 뉴클레오티드는 테르피놀렌 신타제를 코딩한다. 적합한 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로, (AY693650, 오스시뭄 바실리컴 (*Oscimum basilicum*)) 및 (AY906866, REGION: 10.1887, 슈도즈가 멘지에서 (*Pseudotsuga menziesii*))를 포함한다.
- [0162] 일부 구현예들에서, 이종의 뉴클레오티드는 아모르파디엔 신타제를 코딩한다. 적정 뉴클레오티드 서열에 대한 예시적인 예는, 미국 특허 공개공보 2004/0005678의 서열번호 37이다.
- [0163] 일부 구현예들에서, 이종의 뉴클레오티드는 α-파르네센 신타제를 코딩한다. 적합한 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로, 피루스 코무니스 (*Pyrus communis*) cultivar d'Anjou의 DQ309034 (배; 유전자 명 AFS1) 및 말루스 도메스티카 (*Malus domestica*)의 AY182241 (사과; 유전자 AFS1)를 포함한다. Pechouus et al., *Planta* 219(1):84-94 (2004).
- [0164] 일부 구현예들에서, 이종의 뉴클레오티드는 β-파르네센 신타제를 코딩한다. 적합한 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로, 멘타 x 피페리타 (*Mentha x piperita*)의 등재번호 AF024615 (페퍼민트; 유전자 Tspa11), 및 아르테미시아 아누아 (*Artemisia annua*)의 AY835398을 포함한다. Picaud et al., *Phytochemistry* 66(9): 961-967 (2005).
- [0165] 일부 구현예들에서, 이종의 뉴클레오티드는 파르네솔 신타제를 코딩한다. 적합한 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로, 지 메이스 (*Zea mays*)의 등재번호 AF529266와 사카로마이세스 세레비지애 (*Saccharomyces cerevisiae*)의 YDR481C (유전자 Pho8)를 포함한다. Song, L., *Applied Biochemistry and Biotechnology* 128:149-158 (2006).
- [0166] 일부 구현예들에서, 이종의 뉴클레오티드는 네롤리돌 신타제를 코딩한다. 적합한 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로, 지 메이스 (*Zea mays*)의 AF529266 (옥수수; 유전자 tps1)을 포함한다.
- [0167] 일부 구현예들에서, 이종의 뉴클레오티드는 폐출리올 신타제를 코딩한다. 적합한 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로, AY508730 (REGION: 포고스테몬 카블린 (*Pogostemon cablin*)의 1.1659)을 포함한다.
- [0168] 일부 구현예들에서, 이종의 뉴클레오티드는 노트카톤 (nootkatone) 신타제를 코딩한다. 적합한 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로, AF441124 (REGION: 1.1647, 시트러스 시넨시스 (*Citrus sinensis*)) 및 AY917195 (REGION: 1.1653, 펄릴라 프루테센스 (*Perilla frutescens*))를 포함한다.
- [0169] 일부 구현예들에서, 이종의 뉴클레오티드는 아비에타다이엔 (abietadiene) 신타제를 코딩한다. 적합한 뉴클레오티드 서열에 대한 예시적인 예로는, 비제한적으로, (U50768; 아비에스 그란디스 (*Abies grandis*)) 및 (AY473621; 피세아 아비에스 (*Picea abies*))을 포함한다.
- [0170] 일부 구현예들에서, 숙주 세포는 C<sub>5</sub> 이소프레노이드를 생산한다. 이 화합물은 하나의 이소프렌 유닛으로부터 파생되며, 또한 헤미테르펜으로도 불린다. 헤미테르펜에 대한 예시적인 예가 이소프렌이다. 다른 구현예들에서, 이소프레노이드는 C<sub>10</sub> 이소프레노이드이다. 이 화합물은 이소프렌 2개의 유닛으로부터 파생되며, 모노테르펜으로도 지칭된다. 모노테르펜에 대한 예시적인 예로는 리모넨, 시트라넬롤, 게라니올, 멘톨, 페릴릴 알코올, 리날롤, 튜온 (thujone) 및 미르센이 있다. 다른 구현예들에서, 이소프레노이드는 C<sub>15</sub> 이소프레노이드이다. 이 화합물은 이소프렌 3개의 유닛으로부터 파생되며, 세스퀴테르펜으로도 지칭된다. 세스퀴테르펜에 대한 예시적

인 예로는 페리플라논 (periplanone) B, 징콜라이드 (gingkolide) B, 아모르파디엔, 아르테미시닌 (artemisinin), 아르테미신산 (artemisinic acid), 발렌센 (valencene), 노트카톤 (nootkatone), 에피-세드롤 (epi-cedrol), 에피-아리스톨로켄 (epi-aristolochene), 파르네솔 (farnesol), 고시폴 (gossypol), 사노닌 (sanonin), 페리플라논, 포르스콜린 (forskolin), 및 패출롤 (패출리 알코올이라고도 함)이 있다. 다른 구현예들에서, 이소프레노이드는 C<sub>20</sub> 이소프레노이드이다. 이 화합물은 이소프렌 4개의 유닛으로부터 파생되며, 또한 다이테르펜으로도 지칭된다. 다이테르펜에 대한 예시적인 예로는 카스벤 (casbene), 엘루테로빈 (eleutherobin), 파클리탁셀, 프로스트라틴 (prostratin), 슈도프테로신 (pseudopterosin) 및 탁사다이엔 (taxadiene)이 있다. 또 다른 예로, 이소프레노이드는 C<sub>20+</sub> 이소프레노이드이다. 이 화합물은 4개 보다 많은 수의 이소프렌 유닛들로부터 파생되며, 트리테르펜 (이소프렌 6개의 유닛으로부터 파생된 C<sub>30</sub> 이소프레노이드 화합물), 예컨대 아르브루시드 (arbruside) E, 브루세안틴 (bruceantin), 테스토스테론, 프로게스테론, 코르티손, 디지톡신 (digitoxin) 및 스쿠알렌; 테트라테르펜 (이소프레노이드 8개로부터 파생된 C<sub>40</sub> 이소프레노이드 화합물), 예컨대 β-카로텐; 및 폴리테르펜 (8개 보다 많은 수의 이소프렌 유닛으로부터 파생된 C<sub>40+</sub> 이소프레노이드 화합물), 예컨대 폴리이소프렌을 포함한다. 일부 구현예들에서, 이소프레노이드는 아비에타다이엔, 아모르파다이엔, 카렌, α-파르네센, β-파르네센, 파르네솔, 게라니올, 게라닐게라니올, 이소프렌, 리날롤, 리모넨, 미르센, 네롤리돌, 옥시멘, 패출롤, β-피넨, 사비넨, γ-테르피넨, 테르피놀렌 및 발렌센으로 이루어진 군으로부터 선택된다. 이소프레노이드 화합물은, 또한, 비제한적으로, 카로테노이드 (예, 라이코펜, α- 및 β-카로텐, α- 및 β-크립토잔틴, 박신 (bixin), 제아잔틴 (zeaxanthin), 아스타잔틴 (astaxanthin) 및 루테인), 스테로이드 화합물, 및 다른 화학 기로 변형된 이소프레노이드로 구성된 화합물들, 예컨대 혼성 테르펜-알카로이드 및 코엔자임 Q-10을 포함한다.

[0171] 5.6.10 이소프레노이드의 생산 방법

[0172] 다른 측면에서, 본 발명은, (a) 이소프레노이드를 생산할 수 있는 본원에 기술된 임의의 유전자 변형된 숙주 세포의 개체군을 이소프레노이드 화합물을 제조하는데 적합한 조건 하에 탄소원이 첨가된 배지에서 배양하는 단계; 및 (b) 상기 배지로부터 상기 이소프레노이드 화합물을 회수하는 단계를 포함하는, 이소프레노이드의 생산 방법을 제공한다.

[0173] 일부 구현예들에서, 유전자 변형된 숙주 세포는 이중적인 포스포케톨라제 발현, 이중적인 포스포프렌스아세틸라제 발현, 메발로네이트 경로의 하나 이상의 효소의 이중 발현; 및 선택적으로, ADA의 이중 발현, NADH-이용성 HMG-CoA 리덕타제의 이중 발현 및 AACs의 이중 발현으로 이루어진 군으로부터 선택되는 한가지 이상의 변형을 포함하며; 유전자 변형된 숙주 세포는 상기 하나 이상의 변형을 포함하지 않는 모 세포 또는 유전자 변형된 숙주 세포의 상기 하나 이상의 변형의 서브세트만을 포함하지만 유전적으로 동일한 모 세포에 비해, 증가된 양으로 이소프레노이드 화합물을 생산한다. 일부 구현예들에서, 증가량은, 예를 들어 수율, 생산, 생산성을, 세포 배양물 L 당 g, 건조 세포 중량 g 당 mg, 세포 배양물 기준 (basis) 단위 체적 당, 단위 건조 세포 중량 기준 당, 세포 배양물 단위 부피/시간 기준 또는 건조 세포 중량 단위/단위 시간 기준으로 측정시, 적어도 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100% 이상이거나 또는 100% 보다 높다.

[0174] 일부 구현예들에서, 숙주 세포는 발효 배지 L 당 약 10 g 이상의 증가된 수준으로 이소프레노이드를 생산한다. 이러한 일부 구현예에서, 이소프레노이드는 세포 배양물 L 당 약 10 - 약 50 g, > 약 15 g, > 약 20 g, > 약 25 g 또는 > 약 30 g으로 생산된다.

[0175] 일부 구현예들에서, 숙주 세포는 건조 세포 중량 g 당 약 50 mg 이상의 증가된 수준으로 이소프레노이드를 생산한다. 일부 이러한 구현예에서, 이소프레노이드는 건조 세포 중량 g 당 약 50 - 약 1500 mg, > 약 100 mg, > 약 150 mg, > 약 200 mg, > 약 250 mg, > 약 500 mg, > 약 750 mg, 또는 > 약 1000 mg으로 생산된다.

[0176] 일부 구현예들에서, 숙주 세포는, 세포 배양물 기준 단위 부피 당, 모 세포에서 생산되는 이소프레노이드 수준의, 적어도 약 10%, 적어도 약 15%, 적어도 약 20%, 적어도 약 25%, 적어도 약 30%, 적어도 약 35%, 적어도 약 40%, 적어도 약 45%, 적어도 약 50%, 적어도 약 60%, 적어도 약 70%, 적어도 약 80%, 적어도 약 90%, 적어도 약 2배, 적어도 약 2.5배, 적어도 약 5배, 적어도 약 10배, 적어도 약 20배, 적어도 약 30배, 적어도 약 40배, 적어도 약 50배, 적어도 약 75배, 적어도 약 100배, 적어도 약 200배, 적어도 약 300배, 적어도 약 400배, 적어도 약 500배, 또는 적어도 약 1,000배이거나 또는 그 보다 높은, 증가된 수준으로 이소프레노이드를 생산한다.

[0177] 일부 구현예들에서, 숙주 세포는, 건조 세포 중량 기준 단위 당, 모 세포에서 생산되는 이소프레노이드 수준의,

적어도 약 10%, 적어도 약 15%, 적어도 약 20%, 적어도 약 25%, 적어도 약 30%, 적어도 약 35%, 적어도 약 40%, 적어도 약 45%, 적어도 약 50%, 적어도 약 60%, 적어도 약 70%, 적어도 약 80%, 적어도 약 90%, 적어도 약 2배, 적어도 약 2.5배, 적어도 약 5배, 적어도 약 10배, 적어도 약 20배, 적어도 약 30배, 적어도 약 40배, 적어도 약 50배, 적어도 약 75배, 적어도 약 100배, 적어도 약 200배, 적어도 약 300배, 적어도 약 400배, 적어도 약 500배 또는 적어도 약 1,000배이거나 또는 그 보다 높은, 증가된 수준으로 이소프레노이드를 생산한다.

[0178] 일부 구현예들에서, 숙주 세포는, 세포 배양물 단위 부피/단위 시간 기준 당, 모 세포에서 생산되는 이소프레노이드 수준의, 적어도 약 10%, 적어도 약 15%, 적어도 약 20%, 적어도 약 25%, 적어도 약 30%, 적어도 약 35%, 적어도 약 40%, 적어도 약 45%, 적어도 약 50%, 적어도 약 60%, 적어도 약 70%, 적어도 약 80%, 적어도 약 90%, 적어도 약 2배, 적어도 약 2.5배, 적어도 약 5배, 적어도 약 10배, 적어도 약 20배, 적어도 약 30배, 적어도 약 40배, 적어도 약 50배, 적어도 약 75배, 적어도 약 100배, 적어도 약 200배, 적어도 약 300배, 적어도 약 400배, 적어도 약 500배 또는 적어도 약 1,000배이거나 또는 그 보다 높은, 증가된 수준으로 이소프레노이드를 생산한다.

[0179] 일부 구현예들에서, 숙주 세포는, 건조 세포 중량 단위/단위 시간 기준 당, 모 세포에서 생산되는 이소프레노이드 수준의, 적어도 약 10%, 적어도 약 15%, 적어도 약 20%, 적어도 약 25%, 적어도 약 30%, 적어도 약 35%, 적어도 약 40%, 적어도 약 45%, 적어도 약 50%, 적어도 약 60%, 적어도 약 70%, 적어도 약 80%, 적어도 약 90%, 적어도 약 2배, 적어도 약 2.5배, 적어도 약 5배, 적어도 약 10배, 적어도 약 20배, 적어도 약 30배, 적어도 약 40배, 적어도 약 50배, 적어도 약 75배, 적어도 약 100배, 적어도 약 200배, 적어도 약 300배, 적어도 약 400배, 적어도 약 500배 또는 적어도 약 1,000배이거나 또는 그 보다 높은, 증가된 수준으로 이소프레노이드를 생산한다.

[0180] 대부분의 구현예들에서, 숙주 세포에 의한 이소프레노이드의 증가된 수준으로의 생산은 유도성 화합물에 의해 유도가능하다. 이러한 숙주 세포는 상기 유도성 화합물의 부재시에 쉽게 조작할 수 있다. 유도성 화합물을 그 후 첨가하여, 숙주 세포에서 증가된 수준으로 이소프레노이드를 생산하도록 유도한다. 다른 구현예들에서, 숙주 세포에 의한 증가된 수준으로의 이소프레노이드 생산은, 예를 들어, 배양 온도, 배지 조성 등의 배양 조건에 변화를 줌으로써 유도가능하다.

[0181] **5.6.11 배양 배지 및 조건**

[0182] 미생물 배양물을 유지 및 증식시키기 위한 재료와 방법들은 미생물학 또는 발효 과학 분야의 당업자들에게 잘 알려져 있다 (예를 들어, Bailey *et al.*, *Biochemical Engineering Fundamentals*, second edition, McGraw Hill, New York, 1986). 숙주 세포의 특이적인 요구성, 발효 및 공정에 따라, 적절한 배양 배지, pH, 온도 및 호기, 미세호기 또는 혐기성 조건 등의 요건에 대한 고려가 이루어져야 한다.

[0183] 본원에 제공되는 이소프레노이드 생산 방법은, 비제한적인 예로, 세포 배양 플레이트, 플라스크 또는 발효기 등의 적정 용기에서, 적정 배양 배지 (예, 판토테네이트 보충물이 첨가되거나 첨가되지 않은 배지)에서 수행될 수 있다. 또한, 본 방법은 미생물 산물의 산업적인 생산을 뒷받침하기 위해 당해 기술 분야에 공지된 임의의 발효 규모로 수행할 수 있다. 교반형 탱크 발효기, 에어리프트 발효기, 버블 발효기 또는 이들의 임의 조합 등의 임의의 적정 발효기가 사용될 수 있다. 숙주 세포로서 사카로마이세스 세레비지에 (*Saccharomyces cerevisiae*)가 사용되는 특정 구현예에서, 균주는 Kosaric, *et al.*, in Ullmann's *Encyclopedia of Industrial Chemistry*, Sixth Edition, Volume 12, pages 398-473, Wiley-VCH Verlag GmbH & Co. KDaA, Weinheim, Germany에 상세하게 기술된 발효기에서 배양할 수 있다.

[0184] 일부 구현예들에서, 배양 배지는 이소프레노이드를 생산할 수 있는 유전자 변형된 미생물을 존속시킬 수 있는, 즉 증식과 생존을 유지시킬 수 있는, 모든 배양 배지이다. 일부 구현예들에서, 배양 배지는 동화가능한 탄소원, 질소원 및 인원을 포함하는 수성 배지이다. 또한, 이러한 배지는 적정 염, 미네랄, 금속 및 기타 영양분을 포함할 수 있다. 일부 구현예들에서, 탄소원과 각각의 세포 필수 영양분은 발효 배지에 증가시키는 방식으로 또는 계속적으로 첨가하며, 각각의 필요 영양분은, 예를 들어, 탄소원을 바이오메스로 변환하는 세포의 대사 기능 또는 호흡 작용을 기초로 사전 결정된 세포 성장 곡선에 따라, 기본적으로 증식 세포에 의한 효율적인 동화에 필요한 최소 수준으로 유지된다.

[0185] 미생물을 배양하기 위한 적정 조건과 적정 배지는 당해 기술 분야에 잘 공지되어 있다. 일부 구현예들에서, 적정 배지에는, 예를 들어, 유도제 (예, 유전자 산물을 코딩하는 하나 이상의 뉴클레오티드 서열이 유도 프로모터의 통제를 받는 경우), 억제제 (예, 유전자 산물을 코딩하는 하나 이상의 뉴클레오티드 서열이 억제 프로모터의

통제를 받는 경우), 또는 선별 물질 (예, 유전자 변형을 포함하는 미생물을 선별하기 위한 항생제) 등의 한가지 이상의 추가적인 물질이 첨가된다.

- [0186] 일부 구현예들에서, 탄소원은 단당류 (단순 당), 이당류, 다당류, 비-발효성 탄소원 또는 이들의 하나 이상의 조합이다. 적합한 단당류에 대한 비제한적인 예로는 글루코스, 갈락토스, 만노스, 프럭토스, 자일로스, 리보스 및 이들의 조합을 포함한다. 적합한 이당류에 대한 비제한적인 예로는 슈크로스, 락토스, 말토스, 트레할로스, 셀로바이오스 (cellobiose) 및 이들의 조합을 포함한다. 적합한 다당류에 대한 비제한적인 예로는 스타치, 글리코겐, 셀룰로스, 키틴 및 이들의 조합을 포함한다. 적합한 비-발효성 탄소원에 대한 비제한적인 예로는 아세테이트와 글리세롤을 포함한다.
- [0187] 배양 배지내 글루코스와 같은 탄소원의 농도는 세포의 증식을 촉진시키지만, 사용되는 미생물의 증식을 억제할 만큼 높진 않아야 한다. 전형적으로, 배양은, 원하는 증식 및 바이오메스 수준을 달성하기 위한 수준이지만 미미한 수준 (검출 한계가 약 < 0.1 g/l임)으로 첨가하면서, 글루코스와 같은 탄소원과 함께 진행된다. 다른 구현예들에서, 글루코스와 같은 탄소원의 배양 배지내 농도는 > 약 1 g/L, 바람직하게는 > 약 2 g/L, 더 바람직하게는 > 약 5 g/L이다. 또한, 글루코스와 같은 탄소원의 배양 배지내 농도는 전형적으로 < 약 100 g/L, 바람직하게는 < 약 50 g/L, 더 바람직하게는 < 약 20 g/L이다. 배지 성분 농도에 대한 언급은 처음 및/또는 진행 중인 성분의 농도 둘다를 지칭할 수 있는 것임에 유념하여야 한다. 어떤 경우에는, 배양하는 중에 탄소원이 고갈 되도록 배양 배지를 두는 것이 바람직할 수도 있다.
- [0188] 적정 배양 배지에 사용될 수 있는 동화가능한 질소원으로는, 비제한적으로, 단순 질소원, 유기 질소원 및 복합 질소원을 포함한다. 이러한 질소원으로는 동물, 식물 및/또는 미생물 기원의 무수 암모니아, 암모늄 염 및 물질을 포함한다. 적합한 질소원으로는, 비제한적으로, 단백질 가수분해물, 미생물 바이오메스 가수분해물, 펩톤, 효모 추출물, 암모늄 설페이트, 우레아 및 아미노산을 포함한다. 전형적으로, 배양 배지내 질소원의 농도는 > 약 0.1 g/L, 바람직하게는 > 약 0.25 g/L, 더 바람직하게는 > 약 1.0 g/L이다. 그러나, 특정 농도를 초과하여, 배양 배지에 질소원을 첨가하는 것은 미생물의 배양에 유익하지 않다. 그 결과, 배양 배지내 질소원의 농도는 < 약 20 g/L, 바람직하게는 < 약 10 g/L, 더 바람직하게는 < 약 5 g/L이다. 또한, 일부 경우들에서는, 배양하는 중에 배양 배지에서 질소원이 고갈되게 하는 것이 바람직할 수 있다.
- [0189] 유효 배양 배지는 무기염, 비타민, 미량 금속 및 증식 촉진제 등의 다른 화합물도 포함할 수 있다. 또한, 이러한 다른 화합물들은 유효 배지에 탄소원, 질소원 또는 미네랄원으로서 존재하거나, 또는 배지에 특이적으로 첨가될 수 있다.
- [0190] 또한, 배양 배지는 적정 인원을 포함할 수 있다. 이러한 인원으로는 무기 인원과 유기 인원을 둘다 포함한다. 바람직한 인원으로는, 비제한적으로, 모노 또는 다이베이직 소듐 및 포타슘 포스페이트 등의 포스페이트 염, 암모늄 포스페이트 및 이의 혼합물을 포함한다. 전형적으로, 배양 배지내 인 농도는 > 약 1.0 g/L, 바람직하게는 > 약 2.0 g/L, 더 바람직하게는 > 약 5.0 g/L이다. 그러나, 특정 농도를 초과하여, 배양 배지에 인을 첨가하는 것은 미생물의 배양에 유익하지 않다. 즉, 배양 배지내 인 농도는 전형적으로 < 약 20 g/L, 바람직하게는 < 약 15 g/L, 더 바람직하게는 < 약 10 g/L이다.
- [0191] 또한, 비슷한 양의 마그네슘을 부여하는 농도로 다른 마그네슘 소스가 사용될 수도 있지만, 적정 배양 배지는 마그네슘의 소스를, 바람직하게는 마그네슘 설페이트 헵타하이드레이트 등의 생리적으로 허용가능한 염 형태로, 포함할 수 있다. 전형적으로, 배양 배지내 마그네슘의 농도는 > 약 0.5 g/L, 바람직하게는 > 약 1.0 g/L, 더 바람직하게는 > 약 2.0 g/L이다. 그러나, 특정 농도를 초과하여, 배양 배지에 마그네슘을 첨가하는 것은 미생물의 배양에 유익하지 않다. 즉, 배양 배지내 마그네슘의 농도는 전형적으로 < 약 10 g/L, 바람직하게는 < 약 5 g/L, 더 바람직하게는 < 약 3 g/L이다. 또한, 어떤 경우에는, 배양하는 중에 배양 배지에서 마그네슘 소스를 고갈되게 하는 것이 바람직할 수 있다.
- [0192] 일부 구현예들에서, 배양 배지는 트리소듐 사이트레이트의 이수화물 (dihydrate) 등의 생물학적으로 허용가능한 킬레이트제를 또한 포함할 수 있다. 이 경우, 배양 배지내 킬레이트의 농도는 > 약 0.2 g/L, 바람직하게는 > 약 0.5 g/L, 더 바람직하게는 > 약 1 g/L이다. 그러나, 특정 농도를 초과하여, 배양 배지에 킬레이트제를 첨가하는 것은 미생물의 배양에 유익하지 않다. 즉, 배양 배지내 킬레이트제의 농도는 전형적으로 < 약 10 g/L, 바람직하게는 < 약 5 g/L, 더 바람직하게는 < 약 2 g/L이다.
- [0193] 또한, 배양 배지는 배양 배지의 바람직한 pH를 유지하기 위해 생물학적으로 허용가능한 산 또는 염기를 처음에 포함할 수 있다. 생물학적으로 허용가능한 산으로는, 비제한적으로, 염산, 황산, 질산, 인산 및 이의 혼합물을

포함한다. 생물학적으로 허용가능한 염기로는, 비제한적으로, 암모늄 하이드록사이드, 소듐 하이드록사이드, 포타슘 하이드록사이드 및 이의 혼합물을 포함한다. 일부 구현예들에서, 사용되는 염기는 암모늄 하이드록사이드이다.

- [0194] 또한, 배양 배지는, 비제한적인 예로, 칼슘 클로라이드 등의 생물학적으로 허용가능한 칼슘 소스를 포함할 수 있다. 전형적으로, 칼슘 클로라이드, 이수화물 등의 칼슘 소스의 배양 배지내 농도는 약 5 mg/L - 약 2000 mg/L, 바람직하게는 약 20 mg/L - 약 1000 mg/L, 더 바람직하게는 약 50 mg/L - 약 500 mg/L의 범위이다.
- [0195] 또한, 배양 배지는 소듐 클로라이드를 포함한다. 전형적으로, 배양 배지내 소듐 클로라이드의 농도는 약 0.1 g/L - 약 5 g/L, 바람직하게는 약 1 g/L - 약 4 g/L, 더 바람직하게는 약 2 g/L - 약 4 g/L의 범위이다.
- [0196] 일부 구현예들에서, 배양 배지는 또한 미량 금속을 포함할 수 있다. 이러한 미량 금속은, 편의상, 배양 배지의 나머지 부분과 별도로 준비될 수 있는, 스탁 용액으로서 배양 배지에 첨가할 수 있다. 전형적으로, 배양 배지에 첨가되는 이러한 미량 금속 용액의 양은 > 약 1 mL/L, 바람직하게는 > 약 5 mL/L, 더 바람직하게는 > 약 10 mL/L이다. 그러나, 특정 농도를 초과하여, 배양 배지에 미량 금속을 첨가하는 것은 미생물의 배양에 유익하지 않다. 즉, 배양 배지에 첨가되는 미량 금속 용액의 양은 전형적으로 < 약 100 mL/L, 바람직하게는 < 약 50 mL/L, 더 바람직하게는 < 약 30 mL/L이다. 아울러, 스탁 용액에 미량 금속을 첨가하는 것 외에도, 개개 성분들을, 미량 금속 용액의 상기 범위로 표시된 양의 성분들에 독립적으로 해당되는 범위로 각각 개별적으로 첨가할 수 있다는 것에 유념하여야 한다.
- [0197] 배양 배지는 판토테네이트, 바이오틴, 칼슘, 판토테네이트, 이노시톨, 피리독신-HCl 및 티아민-HCl 등의 다른 비타민을 포함할 수 있다. 이들 비타민을, 편의상, 배양 배지의 나머지 부분과 별도로 준비할 수 있는, 스탁 용액으로서 배양 배지에 첨가할 수 있다. 그러나, 특정 농도를 초과하여, 배양 배지에 비타민을 첨가하는 것은 미생물의 배양에 유익하지 않다.
- [0198] 본원에 기술된 발효 방법은, 비제한적으로, 배치식, 피드-배치식, 세포 리사이클, 연속 및 반-연속식을 포함하는, 통상적인 배양 방식으로 수행할 수 있다. 일부 구현예들에서, 발효는 피드-배치 모드로 수행한다. 이러한 예에서, 배지 성분들의 일부는 배양 중에, 예컨대 판토테네이트는 생산 발효 단계 중에 고갈된다. 일부 구현예들에서, 첨가가 필요한 시점 전까지 일정 기간 동안 증식 및/또는 이소프레노이드 생산이 유지되도록, 배양물에, 예를 들어, 생산 단계의 착수시 상기한 성분을 상대적으로 고 농도로 첨가할 수 있다. 이들 성분들의 바람직한 범위는 배양시 고갈되는 수준으로 첨가함으로써 배양 전 기간 동안 유지된다. 배양 배지내 성분들의 수준은, 예를 들어, 배양 배지를 주기적으로 샘플링하여 농도를 분석함으로써, 모니터링할 수 있다. 다른 예로, 일단 표준 배양 공정이 전개되면, 배양하는 내내 특정 시점에 공지된 수준에 부합되는 시간 조절된 간격으로 첨가를 행할 수 있다. 당해 기술 분야의 당업자에게 공지되는 바와 같이, 배지의 세포 밀도가 증가됨에 따라 배양하는 동안 영양분의 소비 속도는 빨라진다. 아울러, 외래 미생물이 배양 배지에 도입되는 것을 방지하기 위해, 당해 기술 분야에 공지된 바와 같이, 무균 첨가 방법으로 첨가를 수행한다. 아울러, 소포제 (anti-foaming agent)도 소량 배양 중에 첨가할 수 있다.
- [0199] 배양 배지의 온도는 유전자 변형된 세포의 배양 및/또는 이소프레노이드의 생산에 적합한 모든 온도일 수 있다. 예를 들어, 배양 배지에 접종원을 접종하기 전에, 배양 배지를 약 20°C - 약 45°C, 바람직하게는 약 25°C - 약 40°C, 더 바람직하게는 약 28°C - 약 32°C 범위의 온도로 만들어 유지할 수 있다.
- [0200] 배양 배지의 pH는 배양 배지에 산 또는 염기를 첨가함으로써 조절할 수 있다. pH를 조절하기 위해 암모니아를 사용하는 상기한 경우, 이는 배양 배지에 질소원으로서 편리하게 사용된다. 바람직하게는, pH는 약 3.0 - 약 8.0, 더 바람직하게는 약 3.5 - 약 7.0, 가장 바람직하게는 약 4.0 - 약 6.5로 유지된다.
- [0201] 일부 구현예들에서, 배양 배지의, 글루코스 농도 등의 탄소원 농도는, 배양하는 중에 모니터링한다. 배양 배지의 글루코스 농도는, 예를 들어, 상층액, 예컨대 배양 배지의 세포-무함유 성분에서 글루코스 농도를 모니터링하는데 사용할 수 있는, 글루코스 옥시다제 효소 테스트 또는 고압 액체 크로마토그래피를 이용하는 등의, 공지 기법을 이용하여 모니터링할 수 있다. 이전에 언급된 바와 같이, 탄소원의 농도는 세포 증식의 저해가 이루어지는 수준 보다 낮게 유지되어야 한다. 그런 농도는 유기체에 따라 달라질 수 있지만, 탄소원으로서 글루코스의 경우, 세포 증식 저해는 > 약 60 g/L의 글루코스 농도에서 이루어지며, 실험을 통해 쉽게 측정할 수 있다. 즉, 탄소원으로서 글루코스를 사용하는 경우, 글루코스는 바람직하게는 발효기에 검출 한계 이하로 공급되며 유지된다. 다른 예로, 배양 배지내 글루코스 농도는 약 1 g/L - 약 100 g/L, 더 바람직하게는 약 2 g/L - 약 50 g/L, 더 더 바람직하게는 약 5 g/L - 약 20 g/L의 범위로 유지된다. 탄소원의 농도는, 예를 들어, 실질적으로

순수한 글루코스 용액을 첨가함으로써 바람직한 수준으로 유지시킬 수 있지만, 오리지날 배양 배지의 분액을 첨가함으로써 배양 배지의 탄소원 농도를 유지하는 것 역시 허용되며, 바람직할 수 있다. 오리지날 배양 배지의 분액의 사용은, 배지내 다른 영양분의 농도 (예, 질소원 및 인원)를 동시에 유지시킬 수 있기 때문에, 바람직할 수 있다. 이와 같이, 미량 금속의 농도도 미량 금속 용액의 분액을 첨가함으로써 배양 배지에서 유지시킬 수 있다.

[0202] **5.6.12 이소프레노이드의 회수**

[0203] 숙주 세포에 의해 이소프레노이드가 생산되면, 이는 당해 기술 분야에 공지된 방법으로 임의의 적절한 분리 및 정제를 이용하여 후속적인 사용을 위해 회수 또는 분리할 수 있다. 일부 구현예들에서, 원심분리에 의해 발효물로부터 이소프레노이드를 포함하는 유기상을 분리한다. 다른 구현예들에서, 이소프레노이드를 포함하는 유기상을 발효물로부터 자연적으로 분리한다. 다른 구현예들에서, 이소프레노이드를 포함하는 유기상은 발효 반응물에 탈유화제 (deemulsifier) 및/또는 핵제 (nucleating agent)를 첨가함으로써, 발효물로부터 분리한다. 탈유화제에 대한 예시적인 예로는 응결제 및 응집제를 포함한다. 핵제에 대한 예시적인 예로는 이소프레노이드 자체와 도데칸, 이소프로필 미리스테이트 및 메틸 올리에이트 등의 유기 용매로 구성된 액적 (droplet)을 포함한다.

[0204] 이들 세포에서 생산되는 이소프레노이드는 배양 상층물에 존재하거나 및/또는 숙주 세포와 조합하여 존재할 수 있다. 이소프레노이드가 숙주 세포와 조합하여 존재하는 구현예의 경우, 이소프레노이드의 회수는 세포의 투과화 또는 세포 용혈화를 포함한다. 다른 예로 또는 동시에, 배양 배지내 이소프레노이드는, 비제한적인 예로, 크로마토그래피, 추출, 용매 추출, 막 분리, 전기투석, 역 삼투성, 증류, 화학적 유도체화 및 결정화 등의, 회수 공정을 이용하여 회수할 수 있다.

[0205] 일부 구현예들에서, 이소프레노이드는 유기 상에 존재할 수 있는 다른 산물들로부터 분리한다. 일부 구현예들에서, 분리는 흡착, 증류, 가스-액체 추출 (스트립핑), 액체-액체 추출 (용매 추출), 한외여과 및 표준 크로마토그래피 기법을 이용하여 달성한다.

[0206] **5.7 폴리케티드**

[0207] 일부 구현예들에서, 본원에 제시된 유전자 변형된 숙주 세포는 아세틸-CoA로부터 폴리케티드를 생산할 수 있다. 폴리케티드는 폴리케티드 신타제 (PKS)로 지칭되는 효소 활성들의 집합에 의해 촉매화되는 순차적인 반응을 통해 합성되며, 폴리케티드 신타제는 활성 부위들의 동조된 그룹 (coordinated group)을 포함하는, 거대 멀티-효소 단백질 복합체이다. 폴리케티드 생합성은, 주로 클라이센 축합 반응들을 통한 말로닐-CoA-유래 유닛들의 탈카르복시화 축합을 통해, 아세틸-CoA, 프로피오닐 CoA, 부티릴-CoA 및 이의 활성화된 유도체, 말로닐-, 메틸말로닐- 및 에틸말로닐-CoA 등의 단순 2-, 3-, 4-탄소 빌딩 블럭들로부터 출발하여 단계적으로 진행된다. PKS 유전자는 통상적으로 박테리아의 하나의 오픈로과 진핵생물의 유전자 클러스터에서 조직된다. 3가지 유형의 폴리케티드 신타제가 규명되어 있다: I형 폴리케티드 신타제는 2가지 클래스로 하위 분류되는 거대한 고도의 모듈식 단백질로서, 1) 반복 PKS, 이는 순환되는 방식으로 도메인들을 재사용함, 2) 모듈형 PKS, 이는 별개의 모듈들로 구성된 배열을 포함하며, 반복 도메인을 포함하지 않음. II형 폴리케티드 신타제는 1관능성 단백질의 응집체이고, III형 폴리케티드 신타제는 아실 캐리어 단백질 도메인을 사용하지 않는다.

[0208] 각 연속적인 체인 연장 단계 다음으로, 후술한 바와 같이, 고정된 배열의 케토환원, 탈수 및 에노일, 환원이 진행되는, 지방산 생합성과는 달리, 폴리케티드 생합성의 개개 체인 연장 중간산물은, 관능기 변형을 전체, 일부 진행하거나 또는 전혀 진행하지 않으며, 이로써 많은 수의 화학적으로 다양한 생성물들이 형성된다. 여러가지 스타터 유닛과 체인 연장 유닛의 사용 뿐만 아니라 새로운 스테레오-이성질체의 형성으로, 여분의 복잡성이 형성된다.

[0209] 폴리케티드 신타제에 의해 인가되는 전체 폴리케티드-합성 순서는 다음과 같다 (N-말단에서 C-말단 순서임): 첫 탄소 빌딩 블럭을 아실 캐리어 단백질 상으로 착수 또는 주입, 성장하는 마크롤라이드 체인의 연장을 촉매하는 연장 모듈 및 합성된 마크롤라이드의 분비를 촉매하는 종결 모듈. 이러한 생합성에서 활성을 나타내는 컴포넌트 도메인 또는 개별 효소 기능으로는, 스타터, 연장 및 중간매개 아실 유닛들을 로딩하기 위한 아실-트랜스퍼라제; 티올 에스테르로서 자라는 마크롤라이드를 홀딩하는 아실 캐리어 단백질; 체인 연장을 촉매하는 β-케토-아실 신타제; 알코올 관능기로 일차 환원시키는 β-케토-리덕타제; 물을 제거하여 불포화 티올에스테르를 만드는 데하이드라타제; 최종 환원하여 모두 포화시키는 반응을 촉매하는 에노일 리덕타제; 및 마크롤라이드 방출 및 고리화를 촉매하는 티올에스테라제를 포함한다.

- [0210] 일부 구현예들에서, 본원에 기술된 유전자 변형된 미생물은 아세틸-CoA 및 말로닐-CoA 중 적어도 하나를 아실 캐리어 단백질과 축합할 수 있는 효소, 예컨대 아실-트랜스퍼라제를 코딩하는, 이종의 뉴클레오티드 서열을 포함한다.
- [0211] 일부 구현예들에서, 본원에 기술된 유전자 변형된 미생물은 아세틸-CoA 및 말로닐-CoA로 이루어진 군으로부터 선택되는 1차 반응물을 말로닐-CoA 또는 메틸말로닐-CoA로 이루어진 군으로부터 선택되는 2차 반응물과 축합하여 폴리케티드 생성물을 형성하는 효소, 예를 들어  $\beta$ -케토-아실 신타제를 코딩하는 이종의 뉴클레오티드 서열을 포함한다.
- [0212] 일부 구현예들에서, 본원에 기술된 유전자 변형된 미생물은 폴리케티드 화합물의  $\beta$ -케토 화학 기를  $\beta$ -하이드록시 기로 환원할 수 있는 효소, 예컨대  $\beta$ -케토 리덕타제를 코딩하는, 이종의 뉴클레오티드 서열을 포함한다.
- [0213] 일부 구현예들에서, 본원에 기술된 유전자 변형된 미생물은 폴리케티드 화합물내 알칸 화학 기를 탈수시켜,  $\alpha$ - $\beta$ -불포화 알켄을 만들 수 있는 효소, 예컨대 데하이드라타제를 코딩하는 이종의 뉴클레오티드 서열을 포함한다.
- [0214] 일부 구현예들에서, 본원에 기술된 유전자 변형된 미생물은 폴리케티드 화합물내  $\alpha$ - $\beta$ -이중 결합을 포화 알칸으로 환원할 수 있는 효소, 예컨대 에노일-리덕타제를 코딩하는 이종의 뉴클레오티드 서열을 포함한다.
- [0215] 일부 구현예들에서, 본원에 기술된 유전자 변형된 미생물은 폴리케티드 화합물을 아실 캐리어 담체로부터 가수분해할 수 있는 효소, 예컨대 티오에스테라제를 코딩하는 이종의 뉴클레오티드 서열을 포함한다.
- [0216] 일부 구현예들에서, 폴리케티드 생산 세포는 KS 촉매 영역을 포함하는 효소를 코딩하는 하나 이상의 이종의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 폴리케티드 생산 세포는 AT 촉매 영역을 포함하는 효소를 코딩하는 하나 이상의 이종의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 폴리케티드 생산 세포는 AT 촉매 영역을 포함하는 효소를 코딩하는 2 이상의 이종의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 폴리케티드 생산 세포는 CLF 촉매 영역을 포함하는 효소를 코딩하는 하나 이상의 이종의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 폴리케티드 생산 세포는 ACP 활성을 포함하는 효소를 코딩하는 하나 이상의 이종의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 폴리케티드 생산 세포는 ACP 활성을 포함하는 효소를 코딩하는 2 이상의 이종의 뉴클레오티드 서열을 포함한다.
- [0217] 구체적인 구현예에서, 폴리케티드 생산 세포는 최소한의 방향족 PKS 시스템, 예를 들어, KS 촉매 영역을 포함하는 효소, AT 촉매 영역을 포함하는 효소, CLF 촉매 영역을 포함하는 효소 및 ACP 활성을 포함하는 효소를 각각 코딩하는 이종의 뉴클레오티드 서열을 포함한다. 구체적인 구현예에서, 폴리케티드 생산 세포는 최소의 모듈식 PKS 시스템, 예를 들어, KS 촉매 영역을 포함하는 효소, AT 촉매 영역을 포함하는 효소 및 ACP 활성을 포함하는 효소 각각을 코딩하는 이종의 뉴클레오티드 서열을 포함한다. 또 다른 구체적인 구현예에서, 폴리케티드 생산 세포는 새로운 폴리케티드 합성을 위한 모듈식 방향족 PKS 시스템, 예를 들어, KS 촉매 영역을 포함하는 효소, AT 촉매 영역을 포함하는 하나 이상의 효소 및 ACP 활성을 포함하는 하나 이상의 효소를 각각 코딩하는 이종의 뉴클레오티드 서열을 포함한다.
- [0218] 일부 구현예들에서, 최소의 PKS 시스템, 예를 들어, 최소의 방향족 PKS 시스템 또는 최소의 모듈식 PKS 시스템을 포함하는 폴리케티드 생산 세포는, 최종 산물 폴리케티드의 생산에 기여할 수 있는, 추가적인 촉매 활성을 더 포함한다. 일부 구현예들에서, 폴리케티드 생산 세포는, 초기 (nascent) 폴리케티드 백본의 고리화를 촉매하는, 사이클라제 (CYC) 촉매 영역을 포함하는 효소를 코딩하는 하나 이상의 이종의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 폴리케티드 생산 세포는 케토리덕타제 (KR) 촉매 영역을 포함하는 효소를 코딩하는 하나 이상의 이종의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 폴리케티드 생산 세포는 아로마타제 (ARO) 촉매 영역을 포함하는 효소를 코딩하는 하나 이상의 이종의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 폴리케티드 생산 세포는 에노일리덕타제 (ER) 촉매 영역을 포함하는 효소를 코딩하는 하나 이상의 이종의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 폴리케티드 생산 세포는 티오에스테라제 (TE) 촉매 영역을 포함하는 효소를 코딩하는 하나 이상의 이종의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, 폴리케티드 생산 세포는, ACP의 판테테인닐화에 작용하는, hoI<sub>o</sub> ACP 신타제 활성을 포함하는 효소를 코딩하는 하나 이상의 이종의 뉴클레오티드 서열을 더 포함한다.
- [0219] 일부 구현예들에서, 폴리케티드 생산 세포는 합성 후 폴리케티드 변형 활성을 부여하는 하나 이상의 이종의 뉴클레오티드 서열을 더 포함한다. 일부 구현예들에서, 폴리케티드 생산 세포는, 예를 들어, 항생제 활성을 가진 폴리케티드가 바람직한 경우, 폴리케티드의 합성 후 변형을 구현하는, 글리코실라제 활성을 포함하는 효소를 코

당하는 하나 이상의 이중의 뉴클레오티드 서열을 더 포함한다. 일부 구현예들에서, 폴리케티드 생산 세포는 하이드록실라제 활성을 포함하는 효소를 코딩하는 하나 이상의 이중의 뉴클레오티드 서열을 더 포함한다. 일부 구현예들에서, 폴리케티드 생산 세포 에폭시다제 활성을 포함하는 효소를 코딩하는 하나 이상의 이중의 뉴클레오티드 서열을 더 포함한다. 일부 구현예들에서, 폴리케티드 생산 세포 메틸라제 활성을 포함하는 효소를 코딩하는 하나 이상의 이중의 뉴클레오티드 서열을 더 포함한다.

[0220] 일부 구현예들에서, 폴리케티드 생산 세포는, 비제한적인 예로, 1종 이상의 폴리케티드 합성 경로 효소 및 아세틸-CoA 화합물을 변형하여, 마크롤라이드 (macrolide), 항생제, 항진균, 세포분열억제 화합물, 항-콜레스테롤혈증 화합물, 항-기생충 화합물, 항-콕시듐증 화합물(coccidiostatic compound), 동물 성장 촉진제 또는 살충제 등과 같은 폴리케티드 생성물을 형성할 수 있는 효소를 포함하여, 생합성 효소를 코딩하는 하나 이상의 이중의 뉴클레오티드 서열을 더 포함한다. 일부 구현예들에서, HACD 화합물은 폴리엔 (polyene)이다. 일부 구현예들에서, HACD 화합물은 환형 락톤이다. 일부 구현예들에서, HACD 화합물은 14, 15 또는 16원성의 락톤 고리를 포함한다. 일부 구현예들에서, HACD 화합물은 마크롤라이드 (macrolide), 항생제, 항진균, 세포분열억제 화합물, 항-콜레스테롤혈증 화합물, 항-기생충 화합물, 항-콕시듐증 화합물, 동물 성장 촉진제 또는 살충제로 이루어진 군으로부터 선택되는 폴리케티드이다.

[0221] 일부 구현예들에서, 폴리케티드 생산 세포는, 비제한적으로, 하기 폴리케티드로부터 선택되는 폴리케티드를 생산할 수 있는, 이중의 뉴클레오티드 서열, 예를 들어, PKS 효소 및 폴리케티드 변형 효소를 코딩하는 서열을 포함한다: 아버멕틴 (Avermectin) (예, 미국 특허 5,252,474; 미국 특허 4,703,009; EP 공개번호 118,367; MacNeil *et al.*, 1993, "Industrial Microorganisms: Basic and Applied Molecular Genetics"; Baltz, Hegeman, & Skatrud, eds. (ASM), pp. 245-256, "A Comparison of the Genes Encoding the Polyketide Synthases for Avermectin, Erythromycin, and Nemalectin"; MacNeil *et al.*, 1992, *Gene* 115: 119-125; 및 Ikeda and Omura, 1997, *Chem. Res.* 97: 2599-2609); 칸디시딘 (Candicidin) (FR008) (예, Hu *et al.*, 1994, *Mol. Microbiol.* 14: 163-172); 카보마이신 (Carbomycin), 쿠라마이신 (Curamycin) (예, Bergh *et al.*, *Biotechnol Appl Biochem.* 1992 Feb;15(1):80-9); 다우노루비신 (Daunorubicin) (예, *J Bacteriol.* 1994 Oct;176(20):6270-80); 에포틸론 (Epothilone) (예, PCT 공개번호 99/66028; 및 PCT 공개번호 00/031247); 에리트로마이신 (Erythromycin) (예, PCT 공개번호 93/13663; 미국 특허 6,004,787; 미국 특허 5,824,513; Donadio *et al.*, 1991, *Science* 252:675-9; 및 Cortes *et al.*, Nov. 8, 1990, *Nature* 348:176-8); FK-506 (예, Motamedi *et al.*, 1998; *Eur. J Biochem.* 256: 528-534; 및 Motamedi *et al.*, 1997, *Eur. J Biochem.* 244: 74-80); FK-520 (예, PCT 공개번호 00/020601; 및 Nielsen *et al.*, 1991, *Biochem.* 30:5789-96); 그리세우신 (Griseusin) (예, Yu *et al.*, *J Bacteriol.* 1994 May;176(9):2627-34); 로바스타틴 (Lovastatin) (예, 미국 특허 5,744,350); 프레놀리신 (Frenolycin) (예, Khosla *et al.*, *Bacteriol.* 1993 Apr;175(8):2197-204; 및 Bibb *et al.*, *Gene* 1994 May 3;142(1):31-9); 그라나티신 (Granaticin) (예, Sherman *et al.*, *EMBO J.* 1989 Sep;8(9):2717-25; 및 Bechtold *et al.*, *Mol Gen Genet.* 1995 Sep 20;248(5):610-20); 메더마이신 (Medermycin) (예, Ichinose *et al.*, *Microbiology* 2003 Jul;149(Pt 7):1633-45); 모넨신 (Monensin) (예, Arrowsmith *et al.*, *Mol Gen Genet.* 1992 Aug;234(2):254-64); 노낙틴 (Nonactin) (예, *FEMS Microbiol Lett.* 2000 Feb 1;183(1):171-5); 나나오마이신 (Nanaomycin) (예, Kitao *et al.*, *J Antibiot (Tokyo).* 1980 Jul;33(7):711-6); 네마텍틴 (Nemalectin) (예, MacNeil *et al.*, 1993, *supra*); 니다마이신 (Niddamycin) (예, PCT 공개번호 98/51695; 및 Kakavas *et al.*, 1997, *J. Bacteriol.* 179: 7515-7522); 올린도마이신 (Oleandomycin) (예, Swan *et al.*, 1994, *Mol. Gen. Genet.* 242: 358-362; PCT 공개번호 00/026349; Olano *et al.*, 1998, *Mol. Gen. Genet.* 259(3): 299-308; 및 PCT Pat. App. 공개번호 WO 99/05283); 옥시테트라사이클린 (Oxytetracycline) (예, Kim *et al.*, *Gene.* 1994 Apr 8;141(1):141-2); 피크로마이신 (Picromycin) (예, PCT 공개번호 99/61599; PCT 공개번호 00/00620; Xue *et al.*, 1998, *Chemistry & Biology* 5(11): 661-667; Xue *et al.*, October 1998, *Proc. Natl. Acad. Sci. USA* 95: 12111 12116); 플라테놀라이드 (Platenolide) (예, EP 공개번호 791,656; 및 미국 특허 5,945,320); 라파마이신 (Rapamycin) (예, Schwecke *et al.*, August 1995, *Proc. Natl. Acad. Sci. USA* 92:7839-7843; 및 Aparicio *et al.* 1996, *Gene* 169: 9-16); 리파마이신 (Rifamycin) (예, PCT 공개번호 WO 98/07868; 및 August *et al.*, Feb. 13, 1998, *Chemistry & Biology*, 5(2): 69-79); 소랑기움 (Sorangium) (예, 미국 특허 6,090,601); 소라펜 (Soraphen) (예, 미국 특허 5,716,849; Schupp *et al.*, 1995, *J. Bacteriology* 177: 3673-3679); 스피노신 (Spinocyn) (예, PCT 공개번호 99/46387); 스피라마이신 (Spiramycin) (예, 미국 특허 5,098,837); 테트라세노마이신 (Tetracenomycin) (예, Summers *et al.*, *J Bacteriol.* 1992 Mar;174(6):1810-20; 및 Shen *et al.*, *J Bacteriol.* 1992 Jun;174(11):3818-21); 테트라사이클린 (Tetracycline) (예, *J Am Chem Soc.* 2009 Dec 9;131(48):17677-89); 티로신 (Tylosin) (예, 미국 특허

5,876,991; 미국 특허 5,672,497; 미국 특허 5,149,638; EP 공개번호 791,655; EP 공개번호 238,323; Kuhstoss *et al.*, 1996, *Gene* 183:231-6; 및 Merson-Davies and Cundliffe, 1994, *Mol. Microbiol.* 13: 349-355); 및 6-메틸살리실산 (예, Richardson *et al.*, *Metab Eng.* 1999 Apr;1(2):180-7; 및 Shao *et al.*, *Biochem Biophys Res Commun.* 2006 Jun 23;345(1):133-9).

[0222] **5.8 지방산**

[0223] 일부 구현예들에서, 본원에 제시된 유전자 변형된 숙주 세포는 아세틸-CoA로부터 지방산을 생산할 수 있다. 지방산은 지방산 신타제에 의해 아세틸-CoA 및 말로닐-CoA로부터 일련의 탈카르복시 클라이센 축합 반응을 통해 합성된다. 폴리케티드 신타제와 마찬가지로, 지방산 신타제는 단독 효소가 아니라, 기질이 하나의 기능성 도메인에서 다음 순서의 도메인으로 넘어가는 272 kDa 다기능성 폴리펩타이드로 구성된 효소 시스템이다. 지방산 신타제의 기본적인 2가지 클래스는 규명되어 있다: I형 지방산 신타제는 포유류와 진균류 (포유류와 진균류의 신타제는 구조 배열에서 차이가 있지만) 그리고 박테리아의 CMN 그룹 (코리네박테리아, 마이코박테리아 및 노카르디아)에 공통적인 단일한 다기능성 폴리펩타이드이다. II형 신타제는 고세균 및 유박테리아에서 발견된 것으로, 지방산 합성에 참여하는 별개의 한가지 기능의 효소들의 시리즈이다. 2가지 클래스의 신타제에서 지방산 연장 기전과 환원 기전은, 축합 반응을 담당하는 효소 도메인들은 2가지 클래스들에서 주로 상동성이기 때문에, 동일하다.

[0224] 탈카르복시 클라이센 축합 반응에서 지방산 체인을 연장하는 각 라운드 다음으로, β-케토 기는 케토리덕타제, 테하이드라타제 및 에놀 리덕타제가 순차적으로 작용하여 완전 포화된 탄소 체인으로 환원된다. 연장되는 지방산 체인은 아실 캐리어 단백질에 부착된 이들 활성 부위들 간을 이동하며, 궁극적으로 탄소 16개 길이의 체인 (팔미티드 산)에 도달하면 티오에스테라제의 작용에 의해 방출된다.

[0225] 일부 구현예들에서, 본원에 기술된 유전자 변형된 미생물은, 비제한적으로, 1종 이상의 지방산 합성 경로 효소 및 아세틸-CoA 화합물을 변형하여 팔미테이트, 팔미토일 CoA, 팔미톨레산, 사피엔산, 올레산, 리놀레산, α-리놀렌산, 아라키돈산, 에이코사헵타에노익산, 에루식산 및 도코사헵사에노익산과 같은 지방산 생성물을 형성할 수 있는 효소 등의, 생합성 효소를 코딩하는 이종의 뉴클레오티드 서열을 포함한다. 일부 구현예들에서, HACD 화합물은 팔미테이트, 팔미토일 CoA, 팔미톨레산, 사피엔산, 올레산, 리놀레산, α-리놀렌산, 아라키돈산, 에이코사헵타에노익산, 에루식산 및 도코사헵사에노익산으로 이루어진 군으로부터 선택되는 지방산이다.

[0226] 일부 구현예들에서, 본원에 기술된 유전자 변형된 미생물은, 아세틸-CoA 및 말로닐-CoA 중 하나 이상을 아실 캐리어 단백질과 공유 결합시킬 수 있는 효소, 예컨대 아실-트랜스퍼라제를 코딩하는 이종의 뉴클레오티드 서열을 포함한다.

[0227] 일부 구현예들에서, 본원에 기술된 유전자 변형된 미생물은, 각각 아실 캐리어 단백질 (ACP)에 결합된 아세틸 화학 모이어티와 말로닐 화학 모이어티를 축합하여, 아세토아세틸-ACP를 형성할 수 있는 효소, 예컨대 β-케토 아실-ACP 신타제를 코딩하는 이종의 뉴클레오티드 서열을 포함한다.

[0228] 일부 구현예들에서, 본원에 기술된 유전자 변형된 미생물은, 아세토아세틸-ACP의 이중 결합을 NADPH로 환원하여 D-3-하이드록시부티릴 하이드록실라제-ACP에 하이드록시 기를 형성할 수 있는 효소, 예컨대 β-케토아실-ACP 리덕타제를 코딩하는 이종의 뉴클레오티드 서열을 포함한다.

[0229] 일부 구현예들에서, 본원에 기술된 유전자 변형된 미생물은, D-3-하이드록시부티릴 하이드록실라제-ACP를 탈수하여 베타- 및 감마-탄소 간에 이중 결합을 형성하여 크로토닐-ACP를 형성할 수 있는 효소, 예컨대 β-하이드록시아실-ACP 데하이드라제를 코딩하는 이종의 뉴클레오티드 서열을 포함한다.

[0230] 일부 구현예들에서, 본원에 기술된 유전자 변형된 미생물은, 크로토닐 ACP를 NADPH를 사용해 환원하여 부티릴-ACP를 형성할 수 있는 효소, 예컨대 에노일 ACP 리덕타제를 코딩하는 이종의 뉴클레오티드 서열을 포함한다.

[0231] 일부 구현예들에서, 본원에 기술된 유전자 변형된 미생물은, 아실 캐리어 단백질 유래 C16 아실 화합물을 가수분해하여 팔미테이트를 형성할 수 있는 효소, 예컨대 티오에스테라제를 코딩하는 이종의 뉴클레오티드 서열을 포함한다.

[0232] 일부 구현예들에서, 지방산 생산 세포는, 유전자 비-변형된 모 세포와 비교해, 한가지 이상의 지방산의 생산을 높이기 위해, 아세틸-CoA 신타제 및/또는 말로닐-CoA 신타제를 코딩하는 하나 이상의 이종의 뉴클레오티드 서열을 포함한다.

[0233] 예를 들어, 아세틸-CoA 생산을 높이기 위해, 다음과 같은 유전자들 중 하나 이상을 세포에서 발현시킬 수 있다:

*pdh*, *panK*, *aceEF* (피루베이트 및 2-옥소글루타레이트 데하이드로게나제 복합체의 E1p 데하이드로게나제 구성 요소와 E2p 다이하이드로리포아미드 아세틸트랜스퍼라제 구성 요소를 코딩함), *fabH*, *fabD*, *fabG*, *acpP* 및 *fabF*. 이러한 효소를 코딩하는 뉴클레오티드 서열에 대한 예로는, 비제한적으로 *pdh* (BAB34380, AAC73227, AAC73226), *panK* (coaA이라고도 함, AAC76952), *aceEF* (AAC73227, AAC73226), *fabH* (AAC74175), *fabD* (AAC74176), *fabG* (AAC74177), *acpP* (AAC74178), *fabF* (AAC74179)를 포함한다.

[0234] 일부 구현예들에서, 지방산 수준 증가는 지방산 분해에 참여하는 단백질을 코딩하는 유전자의 약화 또는 녹아웃에 의해 세포에서 달성할 수 있다. 예를 들어, *fadE*, *gpsA*, *idhA*, *pf1b*, *adhE*, *pta*, *poxB*, *ackA*, 및/또는 *ackB*의 발현 수준을 당해 기술 분야에 공지된 기법을 이용해 조작된 숙주 세포에서 약화 또는 녹아웃할 수 있다. 이들 단백질을 코딩하는 뉴클레오티드 서열의 예로는, 비제한적으로 *fadE* (AAC73325), *gpsA* (AAC76632), *IdhA* (AAC74462), *pf1b* (AAC73989), *adhE* (AAC74323), *pta* (AAC75357), *poxB* (AAC73958), *ackA* (AAC75356) 및 *ackB* (BAB81430)를 포함한다. 제조되는 숙주 세포는 적절한 환경에서 배양하였을 때 아세틸-CoA 생산 수준이 증가될 것이다.

[0235] 일부 구현예들에서, 지방산 생산 세포는 효소 아세틸-CoA를 말로닐-CoA로 변환할 수 있는, 예로, 멀티서브유닛 AccABCD 단백질을 코딩하는 이중의 뉴클레오티드 서열을 포함한다. AccABCD를 코딩하는 적합한 뉴클레오티드 서열의 예는 등재 번호 AAC73296, EC 6.4.1.2.를 포함하나, 이들로 한정되지 않는다.

[0236] 일부 구현예들에서, 지방산 생산 세포는 리파제를 코딩하는 이중의 뉴클레오티드 서열을 포함한다. 리파제를 코딩하는 적합한 뉴클레오티드 서열의 예로는 등재 번호 CAA89087 및 CAA98876를 포함하나, 이들로 한정되지 않는다.

[0237] 일부 구현예들에서, 지방산의 수준 증가는 장쇄 아실-ACP의 수준 증가를 유도할 수 있는 P1sB를 저해함으로써 세포에서 달성할 수 있으며, 이는 지방산 생합성 경로에서 초기 단계를 저해할 것이다 (예, *accABCD*, *fabH*, 및 *fabI*). P1sB의 발현 수준은 당해 기술 분야에 공지된 기법을 이용하여 조작된 숙주 세포에서 약화 또는 녹아웃시킬 수 있다. P1sB를 코딩하는 적합한 뉴클레오티드 서열의 예로는 등재 번호 AAC77011을 포함하나 이들로 한정되지 않는다. 특정 구현예에서, *plsB* D31 IE 돌연변이를 이용하여, 세포내 이용가능한 아실-CoA의 양을 증가시킬 수 있다.

[0238] 일부 구현예들에서, 모노불포화된 지방산의 생산 증가는, *fabA*를 억제시키는, *sfa* 유전자를 과발현시킴으로써, 세포에서 달성할 수 있다. *sfa*를 코딩하는 적합한 뉴클레오티드 서열의 예로는 등재 번호 AAN79592를 포함하나, 이들로 한정되지 않는다.

[0239] 일부 구현예들에서, 지방산 수준 증가는 지방산 기질의 체인 길이를 통제하는 효소, 예로, 티오에스테라제의 발현을 조정함으로써, 세포에서 달성할 수 있다. 일부 구현예에서, 지방산 생산 세포는 *tes* 또는 *fat* 유전자를 과발현하도록 변형된다. 적합한 *tes* 뉴클레오티드 서열의 예로는, 비제한적으로, 등재 번호: (*tesA*: *E. coli* 유래 AAC73596, C<sub>18:1</sub> 지방산을 생산할 수 있음) 및 (*tesB*: *E. coli* 유래 AAC73555)를 포함한다. 적합한 *fat* 뉴클레오티드 서열의 예로는, 비제한적으로, (*fatB*: Q41635 및 AAA34215, 움벨루라리아 캘리포니아로부터 유래됨, C<sub>12:0</sub> 지방산을 생산할 수 있음), (*fatB2*: Q39513 및 AAC49269, 쿠페아 후케리아나 (*Cuphea hookeriana*)에서 유래됨, C<sub>8:0</sub> - C<sub>10:0</sub> 지방산을 생산할 수 있음), (*fatB3*: AAC49269 및 AAC72881, 쿠페아 후케리아나에서 유래됨, C<sub>14:0</sub> - C<sub>16:0</sub> 지방산을 생산할 수 있음), (*fatB*: Q39473 및 AAC49151, 신나모넨 캄포럼으로부터 유래됨, C<sub>14:0</sub> 지방산을 생산할 수 있음), (*fatB* [*M141T*]: CAA85388, 아라비둡시스 탈리아나로부터 유래됨, C<sub>16:1</sub> 지방산을 생산할 수 있음), (*fatA*: NP 189147 및 NP 193041, 아라비둡시스 탈리아나로부터 유래됨, C<sub>18:1</sub> 지방산을 생산할 수 있음), (*fatA*: CAC39106, 브라드브리조비움 자포니쿰 (*Bradvrhiizobium japonicum*)으로부터 유래됨, C<sub>18:1</sub> 지방산을 생산할 수 있음), (*fatA*: AAC72883, 쿠페아 후케리아나로부터 유래됨, C<sub>18:1</sub> 지방산을 생산할 수 있음) 및 (*fatA1*, AAL79361, 헬리안투스 아누스(*Helianthus annuus*)로부터 유래됨)을 포함한다.

[0240] 일부 구현예들에서, C<sub>10</sub> 지방산의 수준 증가는 당해 기술 분야에 공지된 기법을 이용하여 티오에스테라제 C<sub>18</sub>의 발현 또는 활성을 약화시킴으로써, 세포에서 달성할 수 있다. 티오에스테라제 C<sub>18</sub>를 코딩하는 적합한 뉴클레오티드 서열에 대한 예로는, 비제한적으로, 등재 번호 AAC73596 및 POADA1을 포함한다. 다른 구현예에서, C<sub>10</sub> 지방산의 수준 증가는 당해 기술 분야에 공지된 기법을 이용하여 티오에스테라제 C<sub>10</sub>의 발현 또는 활성을 약화시킴

으로써, 세포에서 달성할 수 있다. 티오에스테라제 C<sub>10</sub>을 코딩하는 적합한 뉴클레오티드 서열에 대한 예로는, 비제한적으로, 등재 번호 Q39513을 포함한다.

[0241] 일부 구현예들에서, C<sub>14</sub> 지방산의 수준 증가는 당해 기술 분야에 공지된 기법을 이용하여 비-C<sub>14</sub> 지방산을 생산하는 내인성 티오에스테라제의 발현 또는 활성을 약화시킴으로써, 세포에서 달성할 수 있다. 다른 구현예에서, C<sub>14</sub> 지방산의 수준 증가는 당해 기술 분야에 공지된 기법을 이용하여 기질 C<sub>14</sub>-ACP를 이용하는 티오에스테라제의 발현 또는 활성을 증가시킴으로써, 세포에서 달성할 수 있다. 티오에스테라제를 코딩하는 적합한 뉴클레오티드 서열에 대한 예로는, 비제한적으로, 등재 번호 Q39473을 포함한다.

[0242] 일부 구현예들에서, C<sub>12</sub> 지방산의 수준 증가는 당해 기술 분야에 공지된 기법을 이용하여 비-C<sub>12</sub> 지방산을 생산하는 내인성 티오에스테라제의 발현 또는 활성을 약화시킴으로써, 세포에서 달성할 수 있다. 다른 구현예에서, C<sub>12</sub> 지방산의 수준 증가는 당해 기술 분야에 공지된 기법을 이용하여 기질 C<sub>12</sub>-ACP를 이용하는 티오에스테라제의 발현 또는 활성을 증가시킴으로써, 세포에서 달성할 수 있다. 티오에스테라제를 코딩하는 적합한 뉴클레오티드 서열에 대한 예로는, 비제한적으로, 등재 번호 Q41635를 포함한다.

[0243] **5.9 기타 화합물을 생산하기 위한 PK/PTA**

[0244] 일부 구현예들에서, 본원에 제시된 유전자 변형된 숙주 세포 (예, PK/PTA 및 아세틸 포스파타제 활성을 코딩하는 폴리펩타이드, 예컨대 RHR2, OR2 또는 이의 상동체의 기능적 파괴를 포함하는 숙주 세포)는, 세포의 피루베이트를 이용하여 착수하여, 예를 들어, 미국 특허 공개번호 20120156735에 언급된 바와 같이, 2,3-부탄다이올, 2-부탄올, 2-부타논, 발린, 루신, 락트산, 말레이트, 이소아밀 알코올 및 이소부탄올을 생산하는 생합성 경로의 발현이 조작된다. 피루베이트-이용 생합성 경로를 발현하도록 조작된 재조합 숙주 세포에서 효소 피루베이트 데카르복실라제 (PDC)의 파괴를 이용하여, 생합성 경로를 통한 생성물 형성에 있어 피루베이트의 이용성을 증가시킨다. PDC-KO 재조합 숙주 세포는 피루베이트-이용 생합성 경로의 생성물을 생산하는데 이용될 수 있지만, PDC-KO 재조합 숙주 세포는 증식을 위해 외인성 탄소 기질 보충 (예, 에탄올 또는 아세테이트)을 필요로 한다. 특히, 외인성 탄소 기질 2종이 필요하며, 그 중 하나는 바람직한 생성물로 변환되며, 나머지 하나는 증식을 위해 상기한 보충을 필요로 하는 재조합 숙주 세포에 의해 완전히 또는 일부가 아세틸-CoA로 변환된다. 그러나, 이종의 포스포케톨라제 경로의 발현은, PK/PTA에 이종이지 않은 PDC-KO 세포와 비교하여, 증식을 위해 이들 외인성 탄소 기질 공급 필요성을 낮추거나 일소시킨다. 따라서, 아세틸 포스페이트를 아세테이트로 가수분해하는 반응을 촉매할 수 있는 RHR2, HOR2 또는 이의 상동체의 추가적인 기능적 파괴는, PK/PTA가 이들 세포에서 세포 증식을 위한 기질로서 이용가능한 아세틸-CoA의 공급을 증가시키는 능력을 추가로 개선시키는 것으로 예상된다.

[0245] **5.10 유전자 변형된 세포의 제조 방법**

[0246] 또한, 본 발명은, 전술한 한가지 이상의 변형, 예컨대 PK, PTA 및/또는 생합성 경로 효소들, 예로, 세틸-CoA 유래 화합물을 위한 생합성 경로 효소들을 코딩하는 하나 이상의 이종의 핵산을 포함하도록 유전자 변형된 숙주 세포의 제조 방법을 제공한다. 숙주 세포에서 이종의 효소의 발현은, 숙주 세포에, 숙주 세포에서의 발현을 허용하는 조절 인자의 통제 하에 상기 효소를 코딩하는 뉴클레오티드 서열을 포함하는 핵산을 도입함으로써, 달성될 수 있다. 일부 구현예들에서, 핵산은 염색체외 (extrachromosomal) 플라스미드이다. 다른 구현예들에서, 핵산은 숙주 세포의 염색체로 뉴클레오티드 서열을 삽입할 수 있는 염색체 삽입 벡터 (chromosomal integration vector)이다.

[0247] 이들 단백질을 코딩하는 핵산은, 비제한적인 예로, 당해 기술 분야의 당업자에게 공지된 임의의 방법을 이용하여 숙주 세포로 도입할 수 있다 (예를 들어, Hinne et al. (1978) *Proc. Natl. Acad. Sci. USA* 75:1292-3; Cregg et al. (1985) *Mol. Cell. Biol.* 5:3376-3385; Goeddel et al. eds, 1990, *Methods in Enzymology*, vol. 185, Academic Press, Inc., CA; Krieger, 1990, *Gene Transfer and Expression -- A Laboratory Manual*, Stockton Press, NY; Sambrook et al., 1989, *Molecular Cloning -- A Laboratory Manual*, Cold Spring Harbor Laboratory, NY; and Ausubel et al., eds., *Current Edition, Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley Interscience, NY를 참조함). 기법의 예로는, 비제한적으로, 스페로플라스팅 (spheroplasting), 전기충격, PEG 1000 매개 형질전환 및 리튬 아세테이트 또는 리튬 클로라이드 매개의 형질전환을 포함한다.

[0248] 숙주 세포내 효소의 카피 수는, 이 효소를 코딩하는 유전자의 전사를 변형함으로써 변형시킬 수 있다. 이는, 예를 들어, 이 효소를 코딩하는 뉴클레오티드 서열의 카피 수를 변형시킴으로써 (예로, 뉴클레오티드 서열을 포

합하는 고카피수 또는 저카피수 발현 벡터를 이용하거나, 또는 숙주 세포의 게놈에 뉴클레오티드 서열의 부가 카피를 도입하거나 숙주 세포의 게놈에서 뉴클레오티드 서열을 결손 또는 파괴함으로써), 또는 오페론의 폴리시스트론 mRNA에서 코딩 서열의 순서를 바꾸거나, 또는 오페론을 자신의 조절 인자를 가진 각각의 개개 유전자로 분할함으로써, 또는 뉴클레오티드 서열에 작동가능하게 연결된 프로모터나 오퍼레이터의 세기를 높임으로써, 달성할 수 있다. 다른 예로 또는 아울러, 숙주 세포에서 효소의 카피 수는, 이 효소를 코딩하는 mRNA의 번역 수준을 변형함으로써 변형시킬 수 있다. 이는, 예를 들어, mRNA의 안정성을 변형시키거나, 리보솜 결합부의 서열을 변형시키거나, 리보솜 결합부와 효소 코딩 서열의 개시 코돈 사이의 거리나 서열을 변형시킴으로써, 효소 코딩 영역의 개시 코돈의 5' 측의 "상류" 또는 인접하게 위치한 전체 인터시스트론 영역을 변형시킴으로써, mRNA 전사체의 3'-말단을 헤어핀 또는 특수 서열을 이용하거나 안정화하거나, 효소의 코돈 용법을 변형시키거나, 효소의 생합성에 사용되는 희귀 코돈 tRNA의 발현을 변형시키거나, 및/또는 효소의 안정성을, 예를 들어, 이의 코딩 서열의 돌연변이를 통해 높이므로써, 달성할 수 있다.

[0249] 숙주 세포내 효소의 활성은, 비제한적인 예로, 숙주 세포에서 증가된 또는 감소된 용해성을 나타내는 변형된 형태의 효소를 발현하거나, 효소의 활성을 저해하는 도메인이 결합된 변형된 형태의 효소를 발현하거나, 기질에 대한 Kcat가 높거나 낮은, 또는 Km이 낮거나 높은 변형된 형태의 효소를 발현하거나, 또는 경로에서 다른 분자에 의한 피드-백 또는 피드-포워드 조절에 의해 영향을 많이 또는 적게 받는 변형된 형태의 효소를 발현하는 등의, 다수 방식으로 변형시킬 수 있다.

[0250] 일부 구현예들에서, 숙주 세포의 유전자 변형을 위해 사용되는 핵산은 형질전환된 숙주 세포의 선별과 외래 DNA를 유지하도록 숙주 세포에 선택압을 부여하는데 유용한 하나 이상의 선별 마커를 포함한다.

[0251] 일부 구현예들에서, 선별 마커는 항생제 내성 마커이다. 항생제 내성 마커에 대한 예시적인 예로는, 비제한적으로, *BLA*, *NAT1*, *PAT*, *AURI-C*, *PDR4*, *SMR1*, *CAT*, 마우스 dhfr, *HPH*, *DSDA*, *KAN<sup>R</sup>* 및 *SH BLE* 유전자 산물을 포함한다. *E. coli* 유래 *BLA* 유전자 산물은 베타-락탐 항생제 (예, 스펙트럼이 좁은 세팔로스포린, 세파마이신 및 카르바페넴 (에르타페넴), 세파만돌 및 세포페라존)와, 테모실린을 제외한 모든 항-그람-음성 박테리아 페니실린에 대해 내성을 부여하며; 스트렙토마이세스 노우르세이 (*S. noursei*) 유래의 *NAT1* 유전자 산물은 노르세오트리신 (nourseothricin)에 대해 내성을 부여하며; 스크렙토마이세스 비리도크로모게네스 (*S. viridochromogenes*) Tu94 유래 *PAT* 유전자 산물은 비알로포스 (bialophos)에 대해 내성을 부여하며; 사카로마이세스 세레비지에 (*Saccharomyces cerevisiae*) 유래 *AURI-C* 유전자 산물은 아우에로바시딘 A (Auerobasidin A) (AbA)에 대해 내성을 부여하며; *PDR4* 유전자산물은 세룰레닌 (cerulenin)에 대해 내성을 부여하며; *SMR1* 유전자 산물은 설포메투론 메틸에 대해 내성을 부여하며; Tn9 트랜스포손 유래 *CAT* 유전자 산물은 클로람페니콜에 대해 내성을 부여하며; 마우스 dhfr 유전자 산물은 메토티렉세이트에 대해 내성을 부여하며; 클렙시엘라 뉴모니아 (*Klebsiella pneumonia*)의 *HPH* 유전자 산물은 히그로마이신 B에 대해 내성을 부여하며; *E. coli* 유래 *DSDA* 유전자 산물은 단일 질소원으로 D-세린이 첨가된 플레이트에서 세포 생육을 허용하며; Tn903 트랜스포손의 *KAN<sup>R</sup>* 유전자는 G418에 대해 내성을 부여하며; 스트렙토알로테이쿠스 힌더스타누스 (*Streptoalloteichus hindustanus*) 유래 *SH BLE* 유전자 산물은 제오신 (Zeocin) (블레오마이신)에 대해 내성을 부여한다. 일부 구현예들에서, 항생제 내성 마커는 본원에 기술된 유전자 변형된 숙주 세포를 분리한 후에 제거된다.

[0252] 일부 구현예들에서, 선별 마커는 유전자 변형된 미생물에서 영양요구성 (예, 영양분 영양요구성)을 복원한다. 이러한 구현예에서, 모 미생물은, 아미노산 또는 뉴클레오티드 생합성 경로에서 기능하지만 모 세포에게 비-기능성을 부여하였을 때 한가지 이상의 영양분의 보충없이 배지에서 생육할 수 없게 하는, 하나 이상의 유전자 산물에 대한 기능적 파괴를 포함한다. 이러한 유전자 산물로는, 비제한적으로, 효모에서의 *HIS3*, *LEU2*, *LYS1*, *LYS2*, *MET15*, *TRP1*, *ADE2* 및 *URA3* 유전자 산물을 포함한다. 이후, 영양요구성 표현형은 파괴된 유전자 산물의 기능적 카피를 코딩하는 발현 벡터 또는 염색체 삽입 구조체를 모 세포에 형질전환함으로써, 복원할 수 있으며, 제조되는 유전자 변형된 숙주 세포는 모 세포의 영양 요구성 표현형의 소실을 기초로 선별할 수 있다. *URA3*, *TRP1* 및 *LYS2* 유전자의 선별 마커로서의 이용은 파지티브 및 네거티브 선별이 모두 가능하기 때문에 현저한 이점을 가진다. 파지티브 선별은 *URA3*, *TRP1* 및 *LYS2* 돌연변이의 영양요구성 보완에 의해 수행되는 반면, 네거티브 선별은 특정 저해제, 즉, 프로토티로픽 균주의 생육은 방지하지만 각각 *URA3*, *TRP1* 및 *LYS2* 돌연변이의 생육을 허용하는, 5-플루오로-오로트산 (FOA), 5-플루오로안트라닐산 및 아미노아디프산 (aAA)을 이용한다. 다른 구현예들에서, 선별 마커는 공지 선별 방법으로 동정할 수 있는 그외 비-치사성 결핍 또는 표현형을 복원한다.

[0253] 본원은 본원의 방법, 조성물 및 유기체에 이용가능한 특정 유전자와 단백질을 기술하지만; 이러한 유전자에 대한 완전한 동정이 필수적이지 않은 것으로 이해될 것이다. 예를 들어, 폴리펩타이드 또는 효소를 코딩하는 서

열을 포함하는 특정 유전자 또는 폴리뉴클레오티드에 변형을 가하여, 활성을 스크리닝할 수 있다. 전형적으로, 이러한 변형은 보존적인 돌연변이와 침묵 돌연변이를 포함한다. 이러한 변형된 또는 돌연변이된 폴리뉴클레오티드 및 폴리펩타이드는 당해 기술 분야에 공지된 방법을 이용하여 기능성 효소의 발현을 스크리닝할 수 있다.

[0254] 유전자 코드의 고유한 축중 (degeneracy)으로 인해, 기능적으로 동일하거나 등가의 폴리펩타이드를 실질적으로 코딩하는 다른 폴리뉴클레오티드를 또한 사용하여, 이러한 효소를 코딩하는 폴리뉴클레오티드를 클로닝하고 발현할 수 있다.

[0255] 당해 기술 분야의 당업자에게 이해되는 바와 같이, 특정 숙주에서 이의 발현을 강화하도록 코딩 서열을 변형시키는 것이 유익할 수 있다. 유전자 코드는 가능한 코돈 64종으로 많지만, 대부분의 유기체는 전형적으로 이들 코돈의 서브세트를 이용한다. 종들에서 가장 흔히 사용되는 코돈을 최적 코돈이라고 하며, 이용하지 않는 코돈은 희귀 또는 저-이용 코돈으로 분류된다. 코돈은 "코돈 최적화" 또는 "종에서 코돈 편향 조절"로 종종 치장되는 프로세스에서, 숙주의 선호적인 코돈 사용을 반영하여 치환할 수 있다.

[0256] 특정 원핵생물 또는 진핵생물 숙주에 선호되는 코돈을 포함하는 최적화된 코딩 서열 (Murray *et al.*, 1989, *Nucl Acids Res.* 17: 477-508)을, 예를 들어, 비-최적화된 서열로부터 만들어지는 전사체와 비교하여, 긴 반감기 등의 바람직한 특성을 가진 재조합 RNA 전사체를 생산하거나 또는 번역 속도를 높이도록, 제조할 수 있다. 또한, 번역 정지 코돈은 숙주 선호성을 반영하도록 변형될 수 있다. 예를 들어, 사카로마이세스 세레비지에 (*S. cerevisiae*)와 포유류의 전형적인 정지 코돈은 각각 UAA와 UGA이다. 단자엽 식물의 전형적인 정지 코돈은 UGA이고, 곤충과 *E. coli*는 통상 정지 코돈으로 UAA를 사용한다 (Dalphin *et al.*, 1996, *Nucl Acids Res.* 24: 216-8).

[0257] 당해 기술 분야의 당업자는, 유전자 코드의 축중 특성으로 인해, 뉴클레오티드 서열에 차이가 있는 다양한 DNA 분자들을 사용하여 본원의 소정의 효소를 코딩할 수 있다는 것을 알 것이다. 전술한 생합성 효소를 코딩하는 천연 DNA 서열은 단지 본원의 일 구현예를 예시하기 위해 참조하며, 본원은 본원의 방법에서 사용되는 효소의 폴리펩타이드 및 단백질의 아미노산 서열을 코딩하는 임의의 서열의 DNA 분자를 포함한다. 비슷한 방식으로, 폴리펩타이드는 전형적으로 바람직한 활성의 소실 또는 유의한 소실 없이, 아미노산 서열에 하나 이상의 아미노산의 치환, 결손 및 삽입을 허용할 수 있다. 본원은, 변형된 또는 변이체 폴리펩타이드가 기준 폴리펩타이드의 효소적 이화 활성 또는 동화 활성을 가지는 한, 본원에 기술된 특정 단백질과 다른 아미노산 서열을 가진 그러한 폴리펩타이드를 포함한다. 아울러, 본원에 나타낸 DNA 서열에 의해 코딩되는 아미노산 서열은 단지 본원의 구현예들을 예시할 뿐이다.

[0258] 아울러, 본원에 제시된 조성물 및 방법에서 사용가능한 효소의 상동체도 본원에 포함된다. 일부 구현예들에서, 2종의 단백질 (또는 단백질의 영역)은, 아미노산 서열이 적어도 약 30%, 40%, 50% 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 또는 99%의 동일성을 가지는 경우에, 실질적으로 상동하다. 2종의 아미노산 서열 또는 2종의 핵산 서열의 동일성 %를 확인하기 위해, 서열들을 최적 비교 목적으로 정렬한다 (예, 최적 정렬을 위해 제1 및 제2 아미노산 또는 핵산 서열들 중 하나 또는 양자에 갭을 도입할 수 있으며, 비-상동적인 서열은 비교를 위해 무시할 수 있음). 일 구현예에서, 비교 목적으로 정렬한 기준 서열의 길이는 기준 서열의 길이의 적어도 30%, 전형적으로 적어도 40%, 보다 전형적으로 50%, 보다 더 전형적으로 적어도 60%, 보다 더 전형적으로 적어도 70%, 80%, 90%, 100%이다. 그 후 대응되는 아미노산 위치 또는 뉴클레오티드 위치에서 아미노산 잔기와 뉴클레오티드를 비교한다. 제1 서열의 한 위치에 제2 서열의 대응되는 위치와 동일한 아미노산 잔기 또는 뉴클레오티드가 있는 경우, 이 분자는 그 위치에서 동일하다 (본원에서, 아미노산 또는 핵산 "동일성 (identity)"은 아미노산 또는 핵산 "상동성"과 등가이다). 2개의 서열 간의 동일성 %는, 2종의 서열을 최적으로 정렬하기 위해 도입되어야 하는 갭 수와 각 갭의 길이를 고려한, 서열들이 공유하는 동일한 위치의 갯수에 대한 함수이다.

[0259] "상동성"이 단백질 또는 펩타이드에 대해 사용되는 경우, 동일하지 않은 잔기 위치들은 흔히 보존적인 아미노산 치환 차이가 있는 것으로 인지된다. "보존적인 아미노산 치환"은, 아미노산 잔기가 비슷한 화학적 특성 (예, 전하 또는 소수성)을 가진 측쇄 (R 기)를 가진 다른 아미노산 잔기로 치환된 것이다. 일반적으로, 보존적인 아미노산 치환은 단백질의 기능적인 특성을 실질적으로 바꾸지 않을 것이다. 2개 이상의 아미노산 서열이 보존적인 치환에 의해 서로 차이가 있을 경우, 서열의 동일성 % 또는 상동성 정도는 치환의 보존적인 특성에 대한 보정 이상으로 조정될 수 있다. 이러한 조정을 가하는 수단은 당해 기술 분야의 당업자들에게 잘 알려져 있다 (예, Pearson W. R., 1994, *Methods in Mol Biol* 25: 365-89).

[0260] 다음과 같은 6가지 그룹들 각각은 서로에 대한 보존적인 치환인 아미노산을 포함한다: 1) 세린 (S), 트레오닌

(T); 2) 아스파르트산 (D), 글루탐산 (E); 3) 아스파라긴 (N), 글루타민 (Q); 4) 아르기닌 (R), 라이신 (K); 5) 이소루신 (I), 루신 (L), 알라닌 (A), 발린 (V), 및 6) 페닐알라닌 (F), 티로신 (Y), 트립토판 (W).

[0261] 서열 동일성 %로도 언급되는 폴리펩타이드의 서열 상동성은, 전형적으로 서열 분석 소프트웨어를 이용하여 측정한다. 여러가지 유기체로부터 유래된 다수의 서열을 포함하는 데이터베이스에 대해 분자를 비교하는데 사용되는 전형적인 알고리즘은 컴퓨터 프로그램 BLAST이다. 여러가지 다수 유기체들로부터 유래된 서열이 포함된 데이터베이스를 검색하는 경우, 아미노산 서열을 비교하는 것이 전형적이다.

[0262] 아울러, 외래 효소를 코딩하는 임의 유전자 (또는 본원에 언급된 임의의 기타 (또는 이의 발현을 통제 또는 조절하는 임의의 조절 인자)는, 당해 기술 분야의 당업자에게 공지된, 방향성 진화 또는 합리적인 돌연변이 유발 등의 유전자/단백질 조작 기법에 의해 최적화할 수 있다. 이러한 행위로 당해 기술 분야의 당업자는 효모에서 효소의 발현 및 활성을 최적화할 수 있다.

[0263] 아울러, 이들 효소를 코딩하는 유전자는 다른 진균류 및 박테리아 종들에서 동정할 수 있으며, 이 경로를 조절하도록 발현시킬 수 있다. 다양한 유기체들, 비제한적인 예로, 사카로마이세스 세레비지에 (*S. cerevisiae*) 및 사카로마이세스 우바룸 (*S. uvarum*) 등의 사카로마이세스 (*Saccharomyces*) spp., 클루이베로마이세스 썬모톨러란스 (*K. thermotolerans*), 클루이베로마이세스 락티스 (*K. lactis*) 및 클루이베로마이세스 막시아누스 (*K. marxianus*) 등의 클루이베로마이세스 (*Kluyveromyces*) spp., 피키아 (*Pichia*) spp., 한세놀라 폴리모르파 (*H. polymorpha*) 등의 한세놀라 (*Hansenula*) spp., 칸디다 (*Candida*) spp., 트리코스포론 (*Trichosporon*) spp., 야마다지마 스티피티스 (*Y. spp. stipitis*) 등의 야마다지마 (*Yamadazyma*) spp., 토룰라스포라 프레토리엔시스 (*Torulasporea pretoriensis*), 이사첸키아 오리엔탈리스 (*Issatchenkia orientalis*), 스킴조사카로마이세스 폼베 (*S. pombe*) 등의 스킴조사카로마이세스 (*Schizosaccharomyces*) spp., 크립토크커스 (*Cryptococcus*) spp., 아스퍼질러스 (*Aspergillus*) spp., 뉴로스포라 (*Neurospora*) spp. 또는 우스틸라고 (*Ustilago*) spp. 등이 이들 효소의 소스로서 사용될 수 있다. 혐기성 진균류 유래 유전자의 소스로는, 비제한적으로, 피로마이세스 (*Piromyces*) spp., 오르피노마이세스 (*Orpinomyces*) spp., 또는 네오칼리마스틱스 (*Neocallimastix*) spp.를 포함한다. 사용가능한 원핵생물 효소의 소스로는, 비제한적으로, 에스케리키아 콜라이 (*Escherichia coli*), 자이모모나스 모빌리스 (*Zymomonas mobilis*), 스태필로코커스 아우레우스 (*Staphylococcus aureus*), 바실러스 (*Bacillus*) spp., 클로스트리듐 (*Clostridium*) spp., 코리네박테리움 (*Corynebacterium*) spp., 슈도모나스 (*Pseudomonas*) spp., 락토코커스 (*Lactococcus*) spp., 엔테로박터 (*Enterobacter*) spp., 및 살모넬라 (*Salmonella*) spp.를 포함한다.

[0264] 당해 기술 분야에 공지된 기법들은 추가적인 상동성 유전자와 상동성 효소를 동정하는데 적합할 수 있다. 일반적으로, 유전자 상동체 및/또는 효소 상동체는 기능적 분석을 통해 동정할 수 있으며, 기능적 유사성을 가질 것이다. 당해 기술 분야의 당업자에게 공지된 기법들은 유전자 유사체와 효소 유사체를 동정하는데 적합할 수 있다. 예를 들어, PK, PTA, RHR2 또는 HOR2 유전자, 단백질 또는 효소의 상동체 또는 유사체를 동정하기 위해, 비제한적으로, 대상 유전자/효소의 공개 서열을 토대로 한 프라이머를 이용한 PCR에 의한, 또는 대상 유전자들에서 보존된 영역을 증폭하도록 설계된 축중 프라이머를 이용한 축중 PCR에 의한, 유전자 클로닝 등의, 기법들을 포함할 수 있다. 아울러, 당해 기술 분야의 당업자는 기법들을 이용하여, 기능성 상동성 또는 유사성을 가진 유전자, 단백질 또는 효소의 상동체 또는 유사체를 동정할 수 있다. 기법으로는, (예, 본원에 기술된 바와 같이, 또는 Kiritani, K., *Branched-Chain Amino Acids Methods Enzymology*, 1970에 기술된 바와 같이) 활성에 대한 시험관내 효소 분석을 통해 효소의 촉매 활성을 세포 또는 세포 배양물에서 검사한 다음, 정제를 통해 상기 활성을 가진 효소를 분리하고, 에드만 분해, 가능성 있는 핵산 서열에 대한 PCR 프라이머 설계, PCR을 통한 상기 DNA 서열의 증폭 및 상기 핵산 서열의 클로닝 등의 기법을 통해 효소의 단백질 서열을 결정하는 것을 포함한다. 상동적이거나 유사한 유전자 및/또는 상동적이거나 유사한 효소, 유사한 유전자 및/또는 유사한 효소 또는 단백질을 동정하기 위해, 기법으로 또한 BRENDA, KEGG 또는 MetaCYC 등의 데이터베이스를 이용하여 후보 유전자나 효소에 대한 데이터를 비교하는 것을 포함한다. 후보 유전자 또는 효소는 본원의 교시 내용에 따라 진출한 데이터베이스에서 동정할 수 있다.

[0265] 6. 실시예

[0266] 6.1 실시예 1:

[0267] **PK 및 PTA를 발현하는 숙주 세포에서의 아세테이트 생산**

[0268] 본 실시예는 포스포케톨라제 및 포스포트랜스아세틸라제를 이중적으로 발현하는 효모 균주에서의 아세테이트 생

산에 대해 기술한다.

[0269] 6.1.1 재료 및 방법

[0270] 6.1.1.1 균주 조작

[0271] 6.1.1.1.1 Y967 및 Y968

[0272] Y967과 Y968은 야생형 독립영양균 사카로마이세스 세레비지에 CEN.PK2이고, Y967은 MatA, Y968은 Matalpha이다. Y12869, Y12746 및 이들의 모든 유사체의 출발 균주는 사카로마이세스 세레비지에 균주 Y003 (CEN.PK2, Mat alpha, ura3-52, trp1-289, leu2-3,122, his3<sup>+</sup>)이다. 사카로마이세스 세레비지에로의 모든 DNA-매개 형질전환은 Gietz RW and Woods RA, *Guide to Yeast Genetics and Molecular and Cell Biology. Part B.* San Diego, CA: Academic Press Inc. pp. 87-96 (2002)에 기술된 바와 같이, 리튬 아세테이트 표준 절차를 이용하여 수행하였으며, 모든 사례들에서 구조체의 병합은 게놈 DNA를 PCR 증폭하여 검증하였다.

[0273] 6.1.1.1.2 Y12869

[0274] Y003에 3번의 연속적인 삽입을 통해 Y12869를 구축하였다. 먼저, ACS2 유전자 좌의 상류와 하류 뉴클레오티드 서열로 구성된 서열이 측면에 위치하고 천연 사카로마이세스 세레비지에 (*S. cerevisiae*) LEU2 유전자로 구성된, 삽입 구조체 (i2235; 서열번호 45)를 도입함으로써, 유전자 ACS2를 제거하였다. 사카로마이세스 세레비지에 (*S. cerevisiae*) 숙주 세포로 도입되면, 이 구조체는 게놈의 ACS2 유전자 좌로 상동적인 재조합에 의해 삽입되어, ACS2 코딩 서열이 이의 삽입 서열로 치환됨으로써, ACS2가 기능적으로 파괴될 수 있다. 형질전환체를 유일한 탄소원으로서 2% EtOH가 포함된 CSM-leu 플레이트에 도말하고, PCR 증폭에 의해 검증하였다. 제조되는 균주는 Y4940이었다.

[0275] 다음으로, ALD6를 제거하고, 디케이야 지아 (*Dickeya zeae*) eutE를 Y4940에 하기 도시된 삽입 구조체 (i74804; 서열번호 28)를 이용하여 도입하였다.

[0276]

ALD5US	pTDH3	Dz.eutE	tTEF2	TRP1	Z311	DzeutE	pTDH3	ALD6DS
--------	-------	---------	-------	------	------	--------	-------	--------

[0277] 이 삽입 구조체는, 선별 마커 (TRP1)와 디케이야 지아 (*Dickeya zeae*) 유래 유전자 eutE (NCBI 참조 서열: YP\_003003316.1)를 코딩하는 효모-코돈-최적화된 서열 2카피를, TDH3 프로모터 (천연적인 사카로마이세스 세레비지에 (*S. cerevisiae*) TDH3 코딩 영역의 상류 840 bp)와 TEF2 종결인자 (천연적인 사카로마이세스 세레비지에 (*S. cerevisiae*) TEF2 코딩 영역의 하류 508 bp)의 통제 하에 포함한다. 이들 컴포넌트들에는 측면에 ALD6 유전자 좌의 상류 및 하류 뉴클레오티드들이 위치한다. 숙주 세포로 도입되면, 이 구조체는 숙주 세포의 게놈으로 상동적인 재조합에 의해 삽입되어, ALD6 코딩 서열이 이의 삽입 서열로 치환됨으로써, ALD6가 기능적으로 파괴된다. 이 구조체는 미국 특허 8,221,982에 기술된 방법을 이용하여 조립하였다. 이 구조체를 Y4940에 형질전환하고, 형질전환체를 2% 글루코스가 첨가된 CSM-TRP 플레이트에서 선별하여, PCR 증폭에 의해 검증하였다. 제조되는 균주는 y12602이었다.

[0278] 다음으로, 자신의 프로모터와 종결인자의 통제 하에 위치한 천연적인 사카로마이세스 세레비지에 (*S. cerevisiae*) HIS3 유전자와 이의 측면에 위치한 ACS1의 상류 및 하류 뉴클레오티드 서열로 구성된 삽입 구조체 (i76220; 서열번호 29)를 도입함으로써, Y12602에서 ACS1을 제거하였다. 형질전환체를 단독 탄소원으로서 2% 글루코스가 첨가된 CSM-his 플레이트에 도말하고, PCR 증폭에 의해 검증하였다. 제조되는 균주는 Y12747이었다.

[0279] 다음으로, Y12747에 천연 URA3 서열에서 증폭시킨 PCR 산물을 형질전환하였다. 이 서열은 ura3-52 돌연변이를 회복시킨다. Rose and Winston, *Mol Gen Genet* 193:557-560 (1984)을 참조한다. 형질전환체를 단독 탄소원으로서 2% 글루코스가 첨가된 CSM-ura 플레이트에 도말하고, PCR 증폭에 의해 검증하였다. 제조되는 균주는 Y12869이었다.

[0280] 6.1.1.1.3 Y12745

[0281] Y4940에 3번의 연속적인 삽입을 통해 Y12745를 구축하였다. 먼저, Y4940에 아래에 도시된 삽입 구조체 (i73830; 서열번호 30)를 형질전환하였다.

[0282]

BUD9US	pTDH3	Lm.PK	tTDH3	URA3	PK1	CK.PTA	pTDH3	BUD9DS
--------	-------	-------	-------	------	-----	--------	-------	--------

[0283] 이 삽입 구조체는, 선별 마커 (URA3); *TDH3* 프로모터 (*TDH3* 코딩 서열의 상류 870 bp) 및 *TDH3* 종결인자 (*TDH3* 코딩 서열의 하류 259 bp)의 통제 하에 위치한, 루코노스톡 메센테로이데스 (*Leuconostoc mesenteroides*) (NCBI 참조 서열 YP\_819405.1) 유래 효모-코돈 최적화된 포스포케톨라제 버전; *TDH3* 프로모터 (*TDH3* 코딩 서열의 상류 870 bp) 및 *PGK1* 종결인자 (*PGK1* 코딩 서열의 하류 259 bp)의 통제 하에 위치한, 효모 코돈-최적화된 버전의, 클로스트리듐 클루이베리 (*Clostridium kluyveri*) 포스포트랜스아세틸라제 (NCBI 참조 서열: YP\_001394780.1); 측면에 위치한, 사카로마이세스 세레비지에 (*S. cerevisiae*) *BUD9* 유전자 좌의 상류 및 하류 뉴클레오티드 서열들로 구성된 상동적인 서열을 포함한다. 숙주 세포로 도입되면, 이 구조체는 숙주 세포의 게놈으로 상동적인 재조합에 의해 삽입되고, *BUD9* 코딩 서열이 이의 삽입 서열로 치환됨으로써 *BUD9*은 기능적으로 파괴된다. 이 구조체는 미국 특허 8,221,982에 기술된 방법을 이용하여 조립하였다. 형질전환체는 2% 글루코스가 첨가된 CSM-URA 플레이트에서 선별하였다. 제조되는 균주에, 하기 도시된 삽입 구조체 (i74810; 서열번호 31)를 형질전환하였다.

ALD6US	pTDH3	Lm.PK	tTDH3	TRP1	EHQ13	κd'w1	EHQ1d	ALD6DS
--------	-------	-------	-------	------	-------	-------	-------	--------

[0284]

[0285] 이 구조체는 선별 마커 (TRP1); *TDH3* 프로모터 (*TDH3* 코딩 서열의 상류 870 bp) 및 *TDH3* 종결인자 (*TDH3* 코딩 서열의 하류 259 bp)의 통제 하에 위치한 루코노스톡 메센테로이데스 (*Leuconostoc mesenteroides*) 유래 포스포케톨라제 2 카피; 이들의 양 측면에 위치한, *ALD6* 유전자 좌의 상류 및 하류 뉴클레오티드 서열들로 구성된 상동적인 서열을 포함한다. 숙주 세포로 도입되면, 이 구조체는 숙주 세포의 게놈으로 상동적인 재조합에 의해 삽입되어, *ALD6* 코딩 서열이 이의 삽입 서열로 치환됨으로써 *ALD6*가 기능적으로 파괴된다. 이 구조체는 미국 특허 8,221,982에 기술된 방법을 이용하여 조립하였다. 형질전환체를 2% 글루코스가 첨가된 CSM-URA 플레이트에서 선별하여, PCR 증폭에 의해 검증하였다.

[0286] 다음으로, 천연 사카로마이세스 세레비지에 *HIS3* 유전자와 이의 측면에 존재하는 *ACS1*의 상류 및 하류 뉴클레오티드 서열들로 구성되는 삽입 구조체(i76220; 서열번호 29)를, 자신의 프로모터와 종결인자의 통제 하에 삽입함으로써, *ACS1*을 제거하였다. 형질전환체를 단독 탄소원으로서 2% 글루코스가 첨가된 CSM-his 플레이트에 접종하고, PCR 증폭에 의해 검증하였다.

[0287] 6.1.1.1.4 Y12746

[0288] Y4940에 3번의 연속적인 삽입을 통해 Y12746을 구축하였다. 먼저, Y4940에 아래에 도시된 삽입 구조체 (i73830; 서열번호 30)를 형질전환하였다.

BUD9US	pTDH3	Lm.PK	tTDH3	URA3	TKGd3	CK.PTA	pTDH3	BUD9DS
--------	-------	-------	-------	------	-------	--------	-------	--------

[0289]

[0290] 이 삽입 구조체는, 선별 마커 (URA3); *TDH3* 프로모터 (*TDH3* 코딩 서열의 상류 870 bp) 및 *TDH3* 종결인자 (*TDH3* 코딩 서열의 하류 259 bp)의 통제 하에 위치한, 루코노스톡 메센테로이데스 (*Leuconostoc mesenteroides*) (NCBI 참조 서열 YP\_819405.1) 유래 효모-코돈 최적화된 포스포케톨라제 버전; *TDH3* 프로모터 (*TDH3* 코딩 서열의 상류 870 bp) 및 *PGK1* 종결인자 (*PGK1* 코딩 서열의 하류 259 bp)의 통제 하에 위치한, 효모 코돈-최적화된 버전의, 클로스트리듐 클루이베리 (*Clostridium kluyveri*) 포스포트랜스아세틸라제 (NCBI 참조 서열: YP\_001394780.1); 측면에 위치한, 사카로마이세스 세레비지에 (*S. cerevisiae*) *BUD9* 유전자 좌의 상류 및 하류 뉴클레오티드 서열들로 구성된 상동적인 서열을 포함한다. 숙주 세포로 도입되면, 이 구조체는 숙주 세포의 게놈으로 상동적인 재조합에 의해 삽입되고, *BUD9* 코딩 서열이 이의 삽입 서열로 치환됨으로써 *BUD9*은 기능적으로 파괴된다. 이 구조체는 미국 특허 8,221,982에 기술된 방법을 이용하여 조립하였다. 형질전환체는 2% 글루코스가 첨가된 CSM-URA 플레이트에서 선별하였다.

[0291] 제조된 균주에, 하기 도시된 삽입 구조체 (i74810; 서열번호 31)를 형질전환하였다.

ALD6US	pTDH3	Lm.PK	tTDH3	TRP1	EHQ13	κd'w1	EHQ1d	ALD6DS
--------	-------	-------	-------	------	-------	-------	-------	--------

[0292]

[0293] 이 구조체는 선별 마커 (TRP1); *TDH3* 프로모터 (*TDH3* 코딩 서열의 상류 870 bp) 및 *TDH3* 종결인자 (*TDH3* 코딩 서열의 하류 259 bp)의 통제 하에 위치한 루코노스톡 메센테로이데스 (*Leuconostoc mesenteroides*) 유래 포스포케톨라제 2 카피; 이들의 양 측면에 위치한, *ALD6* 유전자 좌의 상류 및 하류 뉴클레오티드 서열들로 구성된 상동적인 서열을 포함한다. 숙주 세포로 도입되면, 이 구조체는 숙주 세포의 게놈으로 상동적인 재조합에 의해

삽입되어, *ALD6* 코딩 서열이 이의 삽입 서열로 치환됨으로써 *ALD6*가 기능적으로 파괴된다. 이 구조체는 미국 특허 8,221,982에 기술된 방법을 이용하여 조립하였다. 형질전환체를 2% 글루코스가 첨가된 CSM-URA 플레이트에서 선별하여, PCR 증폭에 의해 검증하였다.

[0294] 마지막으로, 제조된 균주에 하기 구조체 (i76221; 서열번호 32)를 형질전환하였다.

ACS1US	pTDH3	Dz.eutE	tTEF2	HIS3	ZF11	Dz.eutE	pTDH3	ACS1DS
--------	-------	---------	-------	------	------	---------	-------	--------

[0295] 이 구조체는 선별 마커 (*HIS3*); *TDH3* 프로모터 (천연 사카로마이세스 세레비지에 *TDH3* 코딩 영역의 상류 840 bp) 및 *TEF2* 종결인자 (천연 사카로마이세스 세레비지에 *TEF2* 코딩 영역의 하류 508 bp)의 통제 하에 위치한 디케이 자아 (*Dickeya zeae*) 유래 유전자 *eutE*를 코딩하는 효모-코돈-최적화된 서열 (NCBI 참조 서열: YP\_003003316.1) 2 카피를 포함한다. 이들 컴포넌트들은 측면에 *ACS1* 유전자 좌의 상류 및 하류 뉴클레오티드들이 위치한다. 숙주 세포로 도입되면, 이 구조체는 숙주 세포의 게놈으로 상동적인 재조합에 의해 삽입되어, *ACS1* 코딩 서열이 이의 삽입 서열로 치환됨으로써, *ACS1*이 기능적으로 파괴된다. 이 구조체는 미국 특허 8,221,982에 기술된 방법을 이용하여 조립하였다. 형질전환체를 2% 글루코스가 첨가된 CSM-HIS 플레이트에서 선별하여, PCR 증폭에 의해 검증하였다. 제조된 균주는 Y12746이었다.

[0297] 6.1.1.1.5 Y19390

[0298] Y19390은 Y12869의 직계 후손이다. Y12869의 *ura*-영양요구성 유도체를 하기 도시된 삽입 구조체 MS49253 (서열번호 36)으로 형질전환하였다:

BUD9US	pTDH3	Lm.PK	tTDH3	URA3	tTDH3	Lm.PK	pTDH3	BUD9DS
--------	-------	-------	-------	------	-------	-------	-------	--------

[0299] 이 삽입 구조체는 선별 마커 (*URA3*); *TDH3* 프로모터 (*TDH3* 코딩 영역의 상류 870 bp) 및 *TEF3* 종결인자 (*TEF3* 코딩 영역의 하류 259 bp)의 통제 하의, 루코노스톡 메센테로이데스 (*Leuconostoc mesenteroides*) (NCBI 참조 서열 YP\_819405.1) 유래 포스포케톨라제의 효모-코돈 최적화된 버전 2카피; 측면에 위치한, 사카로마이세스 세레비지에 (*S. cerevisiae*) *BUD9* 유전자 좌의 상류 및 하류 뉴클레오티드 서열들로 구성된 상동적인 서열을 포함한다. 숙주 세포로 도입되면, 이 구조체는 숙주 세포의 게놈에 상동적인 재조합에 의해 삽입되어, *BUD9* 코딩 서열이 이의 삽입 서열로 치환됨으로써, *BUD9*이 기능적으로 파괴된다. 이 구조체는 미국 특허 8,221,982에 기술된 방법을 이용하여 조립하였다. 형질전환체를 2% 글루코스가 첨가된 CSM-URA 플레이트에서 선별하였다.

[0301] 6.1.1.1.6 Y19391

[0302] Y19391은 Y12869의 직계 후손이다. Y12869의 *ura*-영양요구성 유도체를 하기 도시된 삽입 구조체 MS49298 (서열번호 37)로 형질전환하였다:

BUD9US	pTDH3	Ck.PTA	tPGK1	URA3	tPGK1	Ck.PTA	pTDH3	BUD9DS
--------	-------	--------	-------	------	-------	--------	-------	--------

[0303] 이 삽입 구조체는 선별 마커 (*URA3*); *TDH3* 프로모터 (*TDH3* 코딩 영역의 상류 870 bp) 및 *PGK1* 종결인자 (*PGK1* 코딩 영역의 하류 259 bp)의 통제 하의, 클로스트리듐 클루이베리 (*Clostridium kluyveri*) (NCBI 참조 서열 YP\_001394780.1) 유래 포스포트랜스아세틸라제의 효모-코돈 최적화된 버전 2카피; 및 측면에 위치한, 사카로마이세스 세레비지에 (*S. cerevisiae*) *BUD9* 유전자 좌의 상류 및 하류 뉴클레오티드 서열들로 구성된 상동적인 서열을 포함한다. 숙주 세포로 도입되면, 이 구조체는 숙주 세포의 게놈에 상동적인 재조합에 의해 삽입되어, *BUD9* 코딩 서열이 이의 삽입 서열로 치환됨으로써, *BUD9*이 기능적으로 파괴된다. 이 구조체는 미국 특허 8,221,982에 기술된 방법을 이용하여 조립하였다. 형질전환체를 2% 글루코스가 첨가된 CSM-URA 플레이트에서 선별하였다.

[0305] 6.1.1.2 배양 조건

[0306] 하나의 클론을, 50 mM 숙시네이트 pH 5.0 및 20 g/L 슈크로스가 첨가된 종균 배지 (15 g/L 암모늄 설페이트, 8 g/L 포타슘 포스페이트, 6.1 g/L 마그네슘 설페이트, 150 mg/L EDTA, 57.5 mg/L 아연 설페이트, 4.8 mg/L 코발트 클로라이드, 3.24 mg/L 망간 클로라이드, 5 mg/L 구리 설페이트, 29.4 mg/L 칼슘 클로라이드, 27.8 mg/L 철 설페이트, 4.8 mg/L 소듐 폴리브레이트, 0.6 mg/L 바이오틴, 12 mg/L 칼슘 판토테네이트, 12 mg/L 니코틴산, 30 mg/L 이노시톨, 12 mg/L 티아민 하이드로클로라이드, 12 mg/L 피리독신 하이드로클로라이드, 0.24 mg/L 파라

-아미노벤조산) 5 ml에서, 밤새 30℃ 및 200 rpm으로 배양하여, Y967, Y12869, Y12745, Y12746, Y19390 및 Y19391의 접종 배양물을 배양하였다. 예비 배양물을, 50 mM 숙시네이트 pH 5.0 및 40 g/L 슈크로스가 첨가된 종균 배지 25 ml이 담긴 125 ml 플라스크에 최고 OD600 0.1로 접종하여, 30℃, 200 rpm으로 배양하였다.

[0307] **6.1.1.3 아세테이트, 프럭토스, 글루코스, 및 슈크로스의 정량**

[0308] 세포 브로스 전체 1 ml을 1.5 ml 에펜도르프 튜브로 옮기고, 테이블탑 원심분리기를 이용하여 1분간 13,000 RPM에서 스프닝하여 상층액을 수득함으로써, 아세테이트 및 당류 (프럭토스, 글루코스, 슈크로스)를 정량하였다. 그런 후, 상층액을 8 mM 황산에 (1:1 v/v) 희석하고, 볼텍싱한 다음, 다시 원심분리하여 1.8 ml 바이얼로 이동시켰다. 샘플을 Agilent 1200 HPLC에서 BioRad Aminex HPX-87H 300mm x 7.8mm 컬럼을 이용하여 가변성 과장 및 굴절 인덱스 검출로 분석하였다. 이동상은 4 mM 황산이었고, 컬럼 온도는 40℃, 유속은 0.5 ml/분이었다.

[0309] **6.1.1.4 결과**

[0310] 도 3B는, 야생형 Cen.PK2, Y967이 배치식 슈크로스 교반플라스크 배양으로 배양하는 동안 아세테이트를 생산함을 보여준다. PDH-바이패스의 결손 (*acs1Δ acs2Δ ald6Δ*)을 포함하며 아세트알데하이드 데하이드로게나제 아실화 (*Dz.eutE*)를 이중적으로 발현하는, Y12869는, 아마도 아세트알데하이드를 아세테이트로 변환하는 세포질 아세트알데하이드 데하이드로게나제인 ALD6의 결손으로 인해, PDH-바이패스를 이용하는 야생형 대조군에 비해 아세테이트를 매우 적게 생산한다. PDH-바이패스의 결손 (*acs1Δ acs2Δ ald6Δ*)을 포함하며 아세트알데하이드 데하이드로게나제 아실화 (*Dz.eutE*) 뿐만 아니라 포스포케톨라제 (Lm.PK) 및 포스포트랜스아세틸라제 (Ck.PTA)를 이중적으로 발현하는, 균주 Y12746에서, 아세테이트의 상당한 증가가 관찰되며, 이는 야생형 Y967에 의해 생산되는 양을 능가한다. Y12869를 이용한 결과는, 아세테이트의 베이스라인 수준이 *acs1Δ acs2Δ ald6Δ*이며 ADA를 이용해 세포질 아세틸-CoA로 플럭스 (flux)를 수송하는 균주에서, 극도로 낮다는 것을 보여준다. 모든 사례들에서, 당 소비율은 비슷하며 (본원에서 당은 배지내 슈크로스, 글루코스 및 프럭토스의 총 합으로서 정의됨), 이는 아세테이트 수준 차이가 공급원료의 차별적인 소비로 인한 것이 아님을 설명해준다 (도 3A). 이들 결과는, Y12746에서의 아세테이트 증가가 포스포케톨라제 및/또는 포스포트랜스아세틸라제의 존재로 인한 것임을 시사해준다. 포스포케톨라제 및 포스포트랜스아세틸라제의 촉매 활성으로 아세틸 포스페이트가 생산된다. 따라서, 아세테이트의 축적은 Y12746에서 아세틸 포스페이트의 자발적인 또는 촉매화된 가수분해로 인해 발생할 수 있다.

[0311] ADA, PK 및 PTA를 발현하는 균주 (Y12746)에서 아세테이트의 소스를 확인하기 위해, ADA만 이용하여 세포질 AcCoA를 제공하는 균주 (Y12869, PDH-바이패스의 결손 (*acs1Δ acs2Δ ald6Δ*)을 포함하며, 아세트알데하이드 데하이드로게나제 아실화 (*Dz.eutE*)를 이중적으로 발현함)를, (1) 강력한 프로모터  $P_{TDH3}$ 에 의해 구동되는 PK의 과다발현된 2카피를 코딩하는 삽입 구조체로 형질전환하여 Y19390을 제조하거나, 또는 (2) 강력한 프로모터  $P_{TDH3}$ 에 의해 구동되는 PTA의 과다발현된 2카피를 코딩하는 삽입 구조체로 형질전환하여 Y19391을 제조하였다.

도 3D에 나타낸 바와 같이, 모 균주에 비해 PK 또는 PTA를 발현하는 균주에서 아세테이트의 축적 증가가 관찰되었다. 당 소비는 도 3C에 도시하며, 아세테이트 수준이 차별적인 당 소비로 인한 것이 아님을 보여준다. PK는 X5P를 아세틸 포스페이트와 G3P로 변환하는 반면, PTA는 아세틸 CoA + Pi를 아세틸 포스페이트 + CoA로 상호변환할 수 있다. 이러한 결과는, PK에 의해 X5P로부터 유래되거나 또는 PTA에 의해 AcCoA로부터 유래되던, 아세틸 포스페이트가 도 1에 나타낸 바와 같이 아세테이트로 가수분해될 수 있음을 시사해준다.

[0312] **6.2 실시예 2:**

[0313] **사카로마이세스 세레비지에에서 주류 아세틸 포스파타제의 동정**

[0314] 본 실시예는 효모에서 아세틸 포스페이트를 가수분해할 수 있는 효소의 동정을 기술한다.

[0315] **6.2.1 재료 및 방법**

[0316] **6.2.1.1 세포 배양**

[0317] 해당 효모 균주의 콜로니 하나를 2% 텍스트로스가 첨가된 효모 추출 펩톤 배지 (YPD) 5 mL에서 밤새 종균 배양으로서 배양하였다. 다음날, 이 종균 배양물을 YPD 50 mL에 OD600 0.2로 접종하였다. 달리 명시되지 않은 한, 200 RPM으로 24시간 교반함으로써, 30℃에서 플라스크 배양을 실시하였다.

[0318] **6.2.1.2 세포 무-함유성 추출물 제조**

[0319] 세포 배양물을 15 mL 팔콘 튜브 3개에 나누어 넣고, 4000 x g에서 5분간 원심분리하여 회수하였다. 그런 후,

상층액은 버리고, 세포는 병행한 완충제 W (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 10% 글리세롤) 10 mL에 재현탁한 후 4000 x g에서 5분간 원심분리하여 행구었다. 상층액은 버리고, 세포를 1 mL 세포용혈 완충제 (10 mL 당, 100 mM Tris-HCl pH 8.0, 150 mM NaCl, 10% 글리세롤, 1 mM DTT, 1 EDTA 무첨가 프로테아제 저해제 정제 (Roche))에 재현탁하였다. 그런 후, 세포를 0 링 캡이 있는 2 mL 플라스틱 스크류 캡 원심분리 튜브 (Fisher Brand 520-GRD)로 옮기고, 파괴 비드 (Disruption beads, 0.5Mm, Fisher)와 비드 비터 (bead beater)를 이용하여 6 M/S에서 1분간 세포 용혈시켰다. 튜브를 즉시 적어도 5분간 빙수조에 넣었다. 그런 후, 튜브를 다시 6 M/S에서 1분간 비드 비터에 둔 후, 다시 5분간 얼음조로 이동시켰다. 튜브를 20분간 최소 16000 x g에서 회전시켜, 세포 파편을 펠렛화하였다. 그 후, 상층액을 새로운 자가운 튜브에 옮겼다. 고전적인 브래드포드 단백질 분석을 이용하여 단백질 농도를 측정하였다 (Bradford MM A rapid and sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem* 72, 248-254 (1976)).

[0320] 6.2.1.3 아세틸 포스페이트의 아세틸 포스파타제 반응 및 정량

[0321] 100 mM Tris-HCl pH 7.5, 150 mM NaCl 및 1 mM MgCl<sup>2+</sup>로 구성된 반응 완충제내에서 30°C에서 아세틸 포스파타제 활성 분석을 수행하였다. 아세틸 포스페이트를 정해진 출발 농도 5 mM 또는 10 mM로 첨가하였다. 반응은, 세포 무-함유 추출물을 지정된 양으로 첨가함으로써, 반응을 개시하였다. 포스파타제 저해를 테스트하기 위해, 소듐 플루오라이드를 선택 웰에 30 mM 농도로 첨가하였다. 반응은 밀폐된 96웰 플레이트에서 수행하였으며, 총 반응 부피는 250 μl이었다. Lipmann 및 Tuttle (Lipmann F, Tuttle LC, J. Biol. Chem. 159, 21-28 (1945))에 의해 개발된 방법을 이용하여 아세틸포스페이트 농도를 측정하였다. 반응 혼합물 50 μl를 2M 하이드록실아민 pH 6.8 50 μl에 첨가하여 잘 혼합하고, 10분 이상 실온에서 인큐베이션하였다. 그런 후, 15% 트리클로로아세트산 34 μl를 첨가하여 혼합한 다음, 34 μl 4N HCl and 34 μl 5% FeCl<sub>3</sub>를 혼합하였으며, 각각의 첨가 후 잘 혼합하였다. 이후, 플레이트는 Beckman 원심분리기 J-E에서 swinging bucket rotor JS-5.3을 이용해 5분간 3000 rpm에서 원심분리하여, 석출된 단백질을 펠렛화하였다. 상층액 150 μl를 새로운 96웰 투명 평면 바닥 플레이트 (Greiner Bio-One Cat.-No.: 655161)로 옮겼다. Molecular Devices SpectraMax M5 플레이트 리더에서 505 nm 파장으로 플레이트를 판독하였다.

[0322] 6.2.1.4 활성 포스파타제 분획 정제

[0323] 소정의 효모 균주의 콜로니 하나를 2% 텍스트로스가 첨가된 효모 추출 펩톤 배지 (YPD) 5 mL에서 밤새 중간 배양으로서 배양하였다. 다음날, 50 mL YPD가 든 2.8 L Fербach 플라스크 2개에 상기 중간 배양물을 OD600 0.2로 접종하였다. 플라스크를 160 RPM으로 24시간 교반하면서 30°C에서 배양하였다. 배양물을 5분간 4000x g에서 원심분리하여 회수하였다. 세포 펠렛을 500 mL 멸균수로 행구고, 5분간 4000x g에서 원심분리하였다. 그런 후, 세포 펠렛을 병행한 세포용혈 완충제 (10 mL 당, 100 mM Tris-HCl pH 8.0, 150 mM NaCl, 10% 글리세롤, 1 mM DTT, 1 EDTA 무첨가 프로테아제 저해제 정제 (Roche))에 재현탁하였다. 세포 현탁물을 5 mL 파괴 비드 (Disruption beads, 0.5Mm, Fisher)가 든 15 mL 팔콘 튜브 6개에 나누어 넣었다. 그런 후, 튜브를 6 M/S에서 45초간 비드 비터에 두었다. 그런 후 즉시 튜브를 5분 이상 얼음조에 두었다. 비드 비팅을 다시 3회 반복하였으며, 각 파괴 처리 사이에는 적어도 5분간 빙수조에 두었다. 튜브를, Beckman centrifuge J-E의 4°C로 냉각시킨 JA-10 로터에서 30분간 16000 rpm (30,966 x g)으로 원심분리하여, 세포 파편을 펠렛화하였다. 세포 용혈물은, 추가적으로, Salt *et al.* (Selective flocculation of cellular contaminants from soluble proteins using polyethyleneimine: A study of several organisms and polymer molecular weights. *Enzyme and Microbial Technology* 17, 107-113(1995))에 기술된 선택적인 응집 방법에 의해 다음과 같이 투명하게 만들었다: 세포 무-함유 용혈물에 5 mM NaOH 원액을 첨가하여 pH 7.4로 조정하였다. 동일 부피의 PEI/Borax 용액 (0.5M NaCl 0.25% PEI, 100mM Borax)을 세포 용혈물에 첨가하여, 잘 혼합하였다. 그런 후, 혼합물을 4°C에서 2,500 x g로 30분간 원심분리하였다. 이후, 일정한 속도로 교반하면서 80% 포화 농도가 달성될 때까지 단백질에 암모늄 설페이트를 서서히 첨가하여, 단백질을 석출시켰다. 10분 이상 계속 교반한 다음, 석출된 단백질을 Beckman JA-20에서 10분간 15,000 rpm으로 4 °C에서 원심분리하여 회수하였다. 상층액을 제거한 다음, 단백질을 완충제 A (20 mM Tris-Cl, pH 7, 10% 글리세롤)에 천천히 재현탁하였다. 이후, 단백질을 3,500 Da 분자량 컷오프 투석 카세트 (Pierce #66300) 0.5 - 3 mL에 첨가하여, 1.5 L 완충제 A 중에서 밤새 4°C에서 투석하였다. 투석한 샘플은 10분간 16000 x g로 원심분리하여, 석출된 단백질을 펠렛화하였다. 단백질 농도를 고전적인 브래드포드 단백질 분석으로 측정하였다 (Bradford MM, A rapid and sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem* 72, 248-254 (1976)). 단백질 20 mg을, GE AKTAexplorer FPLC에서 Source 15Q 4.6/100 PE 음이온 교

환 컬럼에 주입하였다. 완충제 B (20 mM Tris-Cl pH 7, 1M NaCl, 10% 글리세롤) 0 - 100% 농도구배를 30분간 유속 0.5 mL/분으로 수행하여 단백질을 용출시키고, 샘플 1 mL을 수집하였다. 각 분획의 활성 분석을 위해, 각 분획 75  $\mu$ l를 100mM Tris-Cl pH 7, 150 mM NaCl, 1 mM MgCl<sub>2</sub>가 함유된 250  $\mu$ l 반응물 중의 8 mM ACP에 첨가하고, 전술한 바와 같이 분석하였다. 이러한 분리에서 수득한 활성 분획을 밤새 완충제 A에 대해 다시 투석하였다. 그런 후, 샘플 전부를 동일한 Source 15Q 4.6/100 PE 음이온 교환 컬럼에 주입하고, 컬럼 부피의 30배에 해당되는 부피를 사용하여 완충제 B 0 - 45% 농도구배로 유속 0.5 mL/분으로 용출시키고, 샘플 1 mL을 수집하였다. 샘플을 전술한 바와 같이 활성 분석하였다.

[0324] **6.2.1.5 단백질 겔 전기영동**

[0325] 단백질 분획을 Criterion 겔 전기영동 시스템을 사용해 분석하였다. 분획 10  $\mu$ l를 5% v/v 2-머캅토에탄올이 첨가된 2X Laemmli 샘플 완충제 (BioRad Cat # 161-0737) 10  $\mu$ l에 첨가하여, 10분간 끓였다. 그런 후, 샘플을 간단히 원심분리하고, 15  $\mu$ l를 24웰 4-15% Criterion™ TGX™ 프리캐스트 겔에 주입하여, 130볼트에서 50분간 1X Tris-글리신-SDS 완충제 (BioRad 10x Tris/글리신/SDS #161-0732)에서 전기영동하였다. 겔은 각 5분씩 3번 탈이온수 200 mL로 행구었다. SimplyBlue™ SafeStain (Life Technologies Cat # LC6060)을 겔을 완전히 덮도록 겔에 첨가하여, 실온에서 부드럽게 교반하면서 1시간 인큐베이션하였다. 이후, SafeStain은 버리고, 겔을 흔들면서 1시간 동안 탈이온수 200 mL로 행구었다.

[0326] **6.2.1.6 활성 포스파타제 분획에서 단백질 동정**

[0327] 펩타이드의 단백질분해에 의한 절단 및 분리

[0328] 전체 단백질 100  $\mu$ g을 트립신으로 단백질분해하여, LC/MS/MS로 이후 동정하였다. 전체 단백질 100  $\mu$ g에 30분간 37°C에서 Tris-카르복시에틸포스포닌 (4 mM)을 처리하여 환원시킨 후, 30분간 RT 압조건에서 요오도아세트아미드 (15 mM)를 사용하여 알킬화하였다. 분해 혼합물에 트립신 5  $\mu$ g을 첨가하였으며, 전체 분해 과정은 12시간동안 37°C에서 진행하였다. 0.1% 포름산을 첨가하여 반응을 퀸칭하고, 이를 Ascentis Peptide express column (5cmx2.1mm ID, 2.1  $\mu$ m particle size)에 주입한 다음, 변형체로서 0.1% 포름산을 첨가하여 아세트오니트릴 고농도 에서 저 농도의 농도구배를 90분간 실시하여 분리하였다. LC 펌프는 Shimadzu CBM20A LC 제어기로 작동되는 Shimadzu LC20AD 펌프 2개이다.

[0329] 질량 분광측정 파라미터:

[0330] QTRAP 4000 하이브리드 트리플-쿼드러폴 리니어 이온 트랩 질량 분광측정기 (hybrid triple-quadrupole linear ion trap mass spectrometer)를 사용하여, 컬럼으로부터 용출되는 펩타이드를 동정하였다. IDA 파라미터는 다음과 같다: 선택 이온 350 - 1300 da; 하전 상태 측정에 사용된 스캔; 1+ 이온 리젝션 (rejection), 미지의 것 허용됨; 롤링 충돌 에너지: yes (AB SCIEX 표준: qtrap 4000); 각 MS/MS의 최고 충전 시간: 950 ms.

[0331] Mascot에 의한 펩타이드 동정

[0332] Matrix Science의 Mascot를 사용하여, CENPK2 시퀀스 데이터베이스에서 하기 파라미터를 사용해 펩타이드를 동정하였다. 고정 변수: 카르브아미도메틸. 가변 변수: 탈아미드화 (NQ), 산화 (MW). 프리커스 질량 허용: 0.5 da. 생성물 질량 허용: 1.0 da. 생략된 허용 절단 (Missed cleavages allowed): 1.

[0333] **6.2.1.7 가닥 조작**

[0334] 기능성 *URA3* 유전자가 결핍된 Y968 버전을 ms59858으로 형질전환하여 *RHR2*를 녹아웃시키거나, 또는 ms59971으로 형질전환하여 *HOR2*를 녹아웃시켰다. 이 구조체를 미국 특허 8,221,982에 기술된 방법을 이용하여 조립하였다. 형질전환체를 2% 글루코스가 첨가된 CSM-URA 플레이트에서 선별하고, PCR 증폭에 의해 검증하였다. 이 구조체에서 *URA3* 마커 측면에는 다이렉트 리피드들이 존재하여, 이의 재고리화를 촉진시킨다. *URA3* 마커를 재고리화하기 위해, 세포를 YPD에서 밤새 배양한 후 5' FOA에 접종하였다. PCR 증폭 및 CSM-URA 플레이트에서의 배양 불능에 의해 *URA3*의 루프아웃 (loopout)을 검증하였다. 그런 후, Y968.ms59858의 *ura-* 버전에 ms59971을 형질전환하여, *RHR2* 및 *HOR2* 이중 녹아웃 균주 Y968.ms59858.ms59971을 제작하였다.

[0335] **6.2.2 결과**

[0336] **6.2.2.1 아세틸 포스페이트의 가수분해는 효소에 의해 촉매되며, 열과 광범위 포스파타제 저해제에 의해 저해된다**

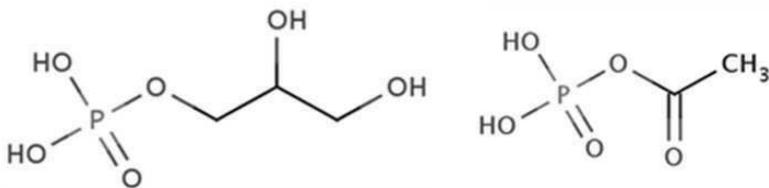
[0337] 도 4에 나타난 바와 같이, 야생형 사카로마이세스 세레비지에 균주 Y967로부터 추출한 세포 무-함유성 추출물의

첨가는 아세틸 포스페이트의 가수분해를 촉매하며, 가수분해 속도는 첨가되는 세포 무-함유 추출물의 양에 따라 결정된다. 세포 무-함유 추출물의 양 증가에 따라 가수분해 속도가 증가한다. 세포 무-함유성 추출물을 끓이면, 세포 무-함유 추출물의 양을 증가시켜 첨가하더라도 더이상 아세틸 포스페이트의 가수분해 속도에 영향이 없으며, 이는 반응가능한 구성성분이 열에 의해 불활화되었다는 것을 의미한다. 마찬가지로, 광범위 포스파타제 저해제로서 30 mM 소듐 플루오라이드를 첨가하면, 세포 무-함유 추출물은 아세틸 포스페이트를 가수분해하는데 효과가 없게 된다. 이러한 결과는, 포스파타제가 아세틸 포스페이트 가수분해의 촉매를 담당할 가능성이 높다는 것을 시사한다.

[0338] 6.2.2.2 단백질 분획화는 한가지 성분이 다량 존재하는 활성 성분을 분리시킨다

[0339] 세포 무-함유 추출물에서 용해성 단백질을 분리하기 위해 음이온 교환 크로마토그래피를 이용하였다. 도 5A는, 포스파타제 활성의 거의 대부분이 하나의 분획에 집중되어 있으며, 잔류 활성이 인접 분획에 존재함을 보여준다. 이는, 세포 무-함유 추출물에서 이러한 활성을 담당하는 효소가 단일 단백질이거나 또는 음이온 교환 크로마토그래피에 의해 분리하였을 때 공동-용출되는 이온 상호작용이 유사한 단백질들이라는 것을, 시사한다.

[0340] FLPC 음이온 교환 정제에서 수득한 활성 분획 #10을 완충액 B를 0-45%의 보다 좁은 농도구배를 이용하여 2차로 정제하였다. 이러한 정제에서 수득한 가장 활성이 높은 분획은 도 6A에 나타난 바와 같이 #14였으며, 이를 질량 분광측정으로 분석하여 분획내 단백질의 정체를 확인하였다. 활성 분획에서 동정된 단백질들 중에서 (도 6B), Rhr2 및 이의 상동체 Hor2는, 현저한 서열 유사성으로 인해 질량 분광측정에서는 구분할 수 없었으며, SGD 데이터베이스에서 포스파타제로 동정된 목록에서 유일한 단백질이었다. Rhr2는 고 농도로 구성적으로 발현되는 글리세롤-1-포스파타제이다. Hor2는 동일한 반응을 촉매하지만 정상 조건에서 낮은 수준으로만 발현되며, 삼투성 스트레스에 의해 유발된다 (Norbeck *et. al.*, Purification and Characterization of Two Isoenzymes of DL-Glycerol-3-phosphatase from *Saccharomyces cerevisiae*, *J. Biol. Chem.*, 271, 13875-13881 (1996)). 아세틸 포스페이트는 효모에 천연적인 대사산물이 아니며, 따라서, 가수분해가 유사한 기질을 타겟팅하는 효소의 혼성 (promiscuous) 반응에 의해 유발되는 것으로 예상된다. Rhr2/Hor2는, 이의 천연 기질인 글리세롤-1-포스페이트가 후술한 바와 같이 아세틸 포스페이트와 유사한 저분자량의 인산화된 화합물이기 때문에, 이런 반응에 최고의 후보이다.



[0341] 글리세롤-1-포스페이트

[0342] 아세틸 포스페이트

[0343] 6.2.2.3 RHR2 및/또는 HOR2의 결손은 포스파타제 활성을 낮춘다

[0344] Rhr2 및/또는 Hor2가 사카로마이세스 세레비지에에서 관찰되는 포스파타제 활성을 담당하는 것인지를 확인하기 위해, RHR2 또는 HOR2 중 어느 한가지가 결여된 새로운 균주들과 RHR2 및 HOR2 둘다 결여된 균주 하나를 구축하였다. 이들 균주는 전술한 바와 같이 배양하고, 세포 무-함유 추출물을 제조하여, 아세틸 포스파타제 활성에 대해 테스트하였다. 도 7에 도시된 바와 같이, RHR2의 결손은 포스파타제 활성을 현저하게 낮추지만, HOR2의 결손은 아세틸 포스페이트의 가수분해율에 영향이 없었다. 그러나, HOR2의 결손은 이미 RHR2가 제거된 균주에서는 아세틸 포스페이트의 가수분해를 낮추지 않는다. 이는, Hor2의 발현이 RHR2의 결손 후 상향 조절됨을 나타낸 공개된 결과와 일치한다 (DeLuna *et. al.*, Need-Based Up-Regulation of Protein Levels in Response to Deletion of Their Duplicate Genes, *PLOS Biol.*, 8, e10000347 (2010)). 이들 포스파타제 2종을 모두 제거하면 도 7에 나타난 바와 같이 아세틸 포스페이트 가수분해율이 거의 백그라운드 수준으로 형성된다. 이러한 결과는, 글리세롤-1-포스파타제 Rhr2 및 Hor2가 사카로마이세스 세레비지에에서 아세틸 포스파타제 활성 대부분을 담당한다는 것을 검증해준다.

[0345] 6.2 실시예 3:

[0346] 아세틸 포스페이트 포스파타제의 결손은 아세테이트 분비를 감소시키며, 아세틸-CoA로부터 유래되는 화합물의 생산을 향상시킨다



[0360] 홀 세포 브로스 (whole cell broth) 1 ml을 1.5 ml 에펜도르프 튜브로 옮기고, 테이블탑 원심분리기를 사용해 1 분간 13,000 RPM으로 원심분리하여 상층액을 투명하게 함으로써, 아세테이트와 글리세롤을 정량하였다. 상층액을 8 mM 황산에 희석 (1:1 v/v)하고, 불택성한 다음, 다시 원심분리한 후 1.8 ml 바이얼로 옮겼다. 샘플을 가변성 파장 및 굴절 인덱스 검출을 이용하여 Agilent 1200 HPLC에서 Aminex HPX-87H 300mm x 7.8mm 컬럼으로 분석하였다. 이동상은 4 mM 황산이고, 컬럼 온도는 40°C이고, 유속은 0.5 ml/분이었다.

[0361] **6.3.1.4 파르네센 정량**

[0362] 34°C에서 2일간 배양한 후, 평판 바닥 96웰 분석 플레이트 (Costar 3916)에서 홀 세포 브로스 98 μl를 나일 레드 용액 (100 μg / DMSO ml) 2 μl와 혼합하고, 30초간 96웰 플레이트 교반기로 혼합하였다. 그런 후, 플레이트를 여기 500 nm 및 방출 550 nm 하에 Beckman M5 플레이트 리더에서 판독하였다.

[0363] **6.3.1.5 광학 밀도 정량**

[0364] 96웰 분석 플레이트에서, 배양물 8 μl를 희석제 (20% PEG 200, 20% 에탄올, 2% Triton X-114)와 혼합하여, 실온에서 30분간 인큐베이션하였다. 이 분석 플레이트를 불택성한 후, Beckman M5 플레이트 리더에서 OD<sub>600</sub>을 측정하였다.

[0365] **6.3.2 결과**

[0366] **도 8A**는, PDH-바이패스의 결손 (*acs1Δ acs2Δ ald6Δ*); 아세트알데하이드 데하이드로게나제 아세틸화 (Dz.eutE) 및 포스포케톨라제 (Lm.PK)와 포스포트랜스아세틸라제 (Ck.PTA)의 이중 발현; 및 파르네센 생산 경로 유전자들의 과다발현을 포함하는, 균주 Y12746.ms63909.ms64472는, *RHR2* 유전자가 결손된 Y12746.ms63909.ms64472 버전 보다 아세테이트를 더 많이 방출한다. **도 8B**에 나타낸 바와 같이, Y12746.ms63909.ms64472 *rhr2<sup>+</sup>*의 글리세롤 수준이 Y12746.ms63909.ms64472에 비해 크게 바뀌지 않기 때문에, *RHR2*의 결손이 글리세롤 생산에는 영향을 주지 않는다. **도 8C**에 도시된 바와 같이, 세포 밀도가 *RHR2<sup>+</sup>* 및 *rhr2<sup>+</sup>* 개체군들에서 비슷하므로, Y12746.ms63909.ms64472 *rhr2<sup>+</sup>*에서의 아세테이트의 실질적으로 감소된 수준은 세포 증식 감소로 인한 것이 아니다. 이러한 결과는, 세포 무-함유 추출물의 아세틸 포스페이트 포스파타제 활성을 담당하는 *Rhr2*가, 또한, 아세틸 포스페이트의 아세테이트로의 생체내 가수분해를 늦추는 일차적인 요인이라는 것을, 입증해준다.

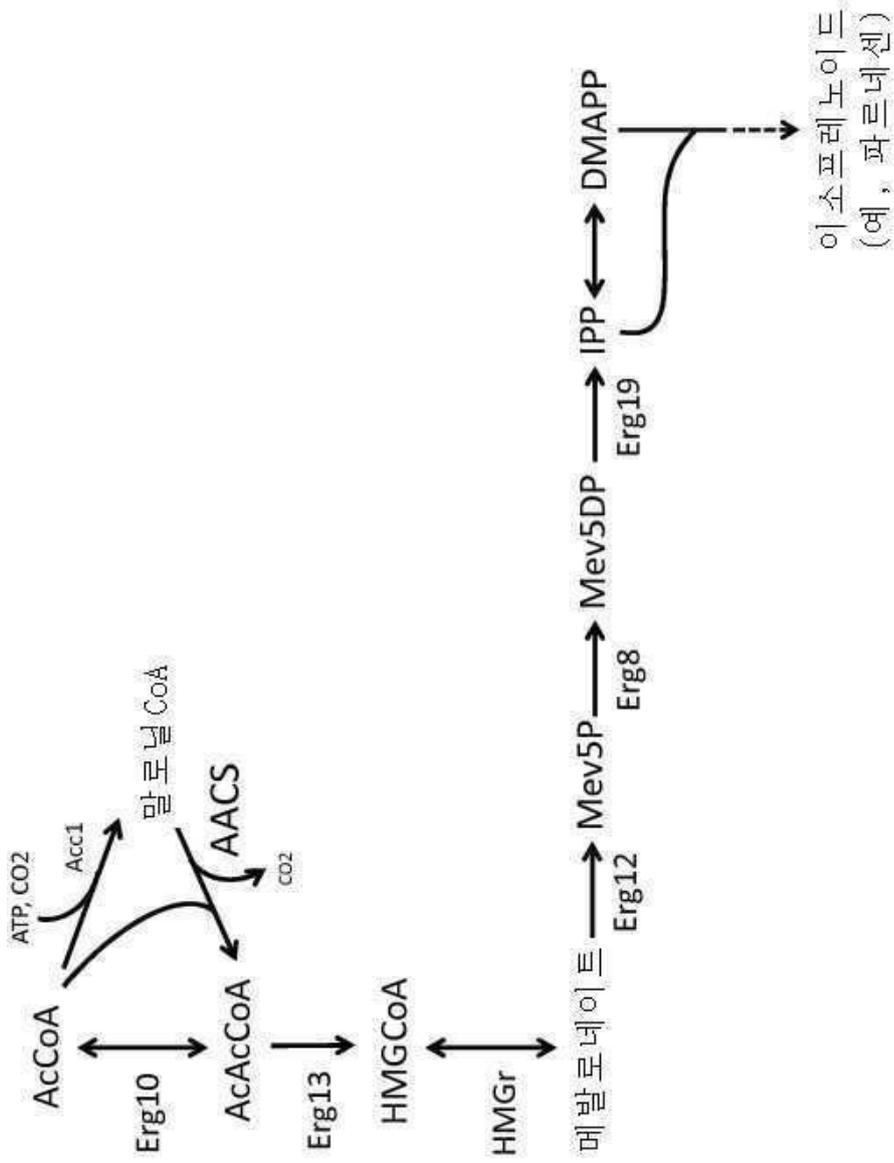
[0367] *RHR2* 결손시 관찰되는 아세테이트의 환원이 파르네센 생산과는 독립적으로 발생하는 지를 확인하기 위해, *RHR2* 유전자가 온전하거나 결손되었지만 파르네센 생산 경로에서 유전자를 발현하지 않는 균주 12746 버전에서 아세테이트 생산을 측정하였다. **도 8D**는, PDH-바이패스의 결손 (*acs1Δ acs2Δ ald6Δ*) 및 아세트알데하이드 데하이드로게나제 아세틸화 (Dz.eutE) 뿐만 아니라 포스포케톨라제 (Lm.PK)와 포스포트랜스아세틸라제 (Ck.PTA)의 이중 발현을 포함하는 균주 Y12746은, *RHR2* 유전자가 결손된 Y12746 버전 보다 아세테이트를 더 많이 방출한다. **도 8E**에 나타낸 바와 같이, *RHR2<sup>+</sup>* 개체군과 *rhr2<sup>+</sup>* 개체군에서 세포 밀도가 비슷하기 때문에, Y12746.ms63909.ms64472 *rhr2<sup>+</sup>*에서 아세테이트의 상당히 감소된 수준은 세포 증식 감소가 원인이 아니다. 이들 데이터는, 아세테이트 감소가 과다발현된 파르네센 생산 경로의 존재와는 무관하게 발생함을 설명해준다.

[0368] **도 9**는, *rhr2* 결손이, 파르네센 생산을 Y12746.ms63907.ms64472에서 2배로, Y12745.ms63907.ms64472에서 1.4배 까지 개선시킴을 보여준다 (각 균주 백그라운드에서, 모 균주의 *RHR2<sup>+</sup>*는 1임). 아울러, *rhr2* 결손은 탄소 고갈 시 Y12746.ms63907.ms64472의 최종 광학 밀도를 개선시킨다. Y12745.ms63907.ms64472와 Y12746.ms63907.ms64472 둘다 포스포케톨라제와 포스포트랜스아세틸라제를 이용하며, 따라서 경로의 중간 산물로서 아세틸 포스페이트를 이용하여 세포질 아세틸-CoA를 생산하며, 이는 파르네센 합성에 사용된다. 균주 Y968.ms63907.ms64472와 균주 Y12869.ms63907.ms64472는 포스포케톨라제 또는 포스포트랜스아세틸라제를 발현하지 않으며, 경로 중간 산물로서 아세틸 포스페이트를 이용하지 않는다. 이들 균주에서 *rhr2* 결손은 이들 균주 백그라운드에서의 파르네센 생산 또는 광학 밀도에 영향을 미치지 않는다. 이는, *rhr2*의 녹아웃 효과가 중간 대사산물로서 아세틸 포스페이트를 이용하는 균주, 예를 들어 이중의 PK 및/또는 PTA를 포함하는 균주에 특이적으로 적용됨을, 시사해준다.

[0369] 본 명세서에 인용된 모든 간행물, 특허 및 특허 출원은, 각 개별 간행물 또는 특허 출원이 구체적이고 개별적으로 원용에 의해 본 명세서에 포함되는 바와 같이, 원용에 의해 본 명세서에 포함된다. 전술한 본 발명은 예시적인 방식으로, 명확한 이해를 위한 예로 기술되어 있지만, 당해 기술 분야의 당업자라면 본 발명의 교시 내용에 비추어, 첨부된 청구항의 사상 및 범위로부터 이탈되지 않으면서 특정한 변형과 수정을 가할 수 있음을 쉽게 알 것이다.

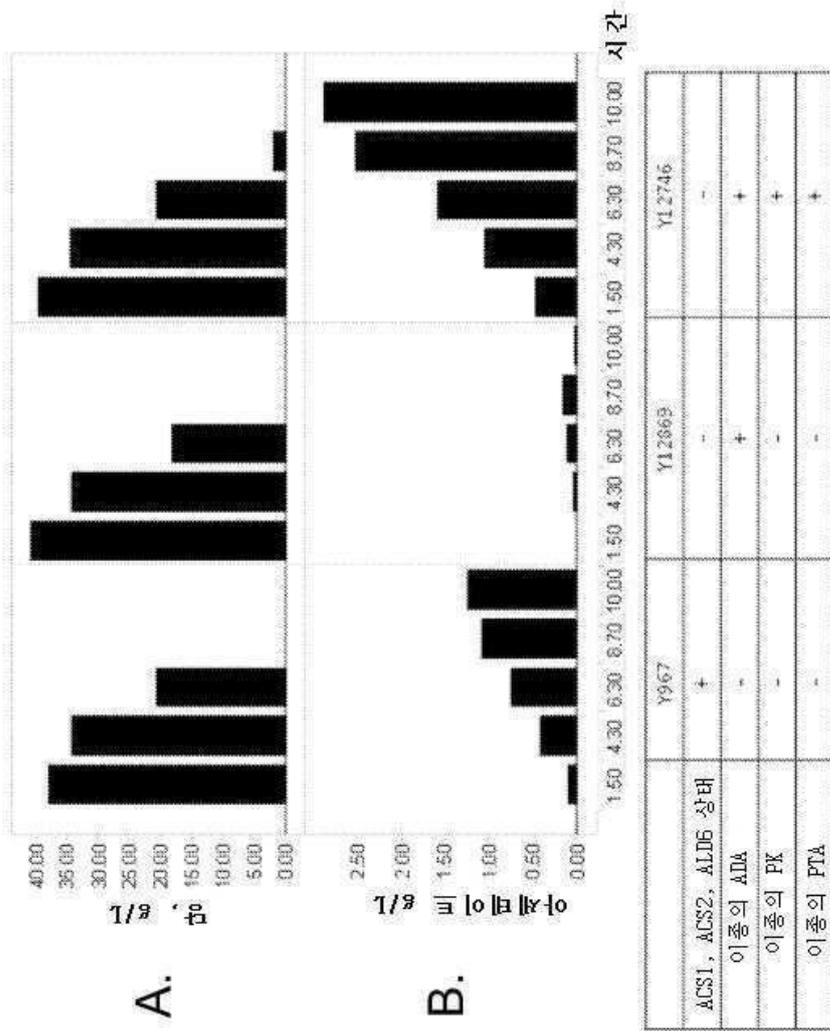


도면2



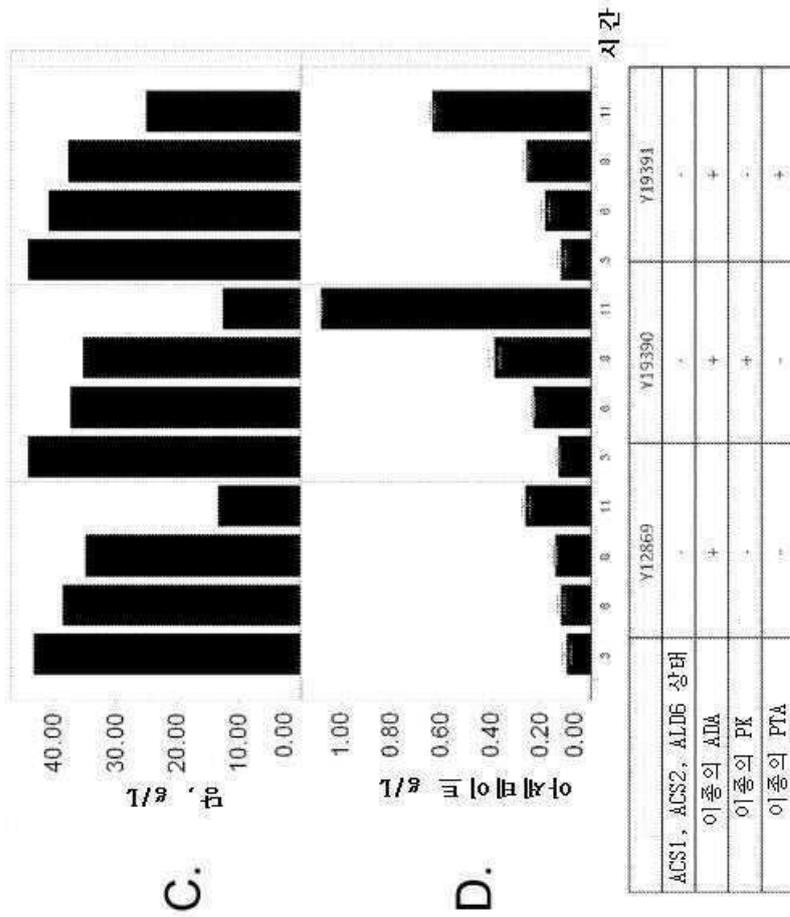
도면3ab

도 3A & 3B

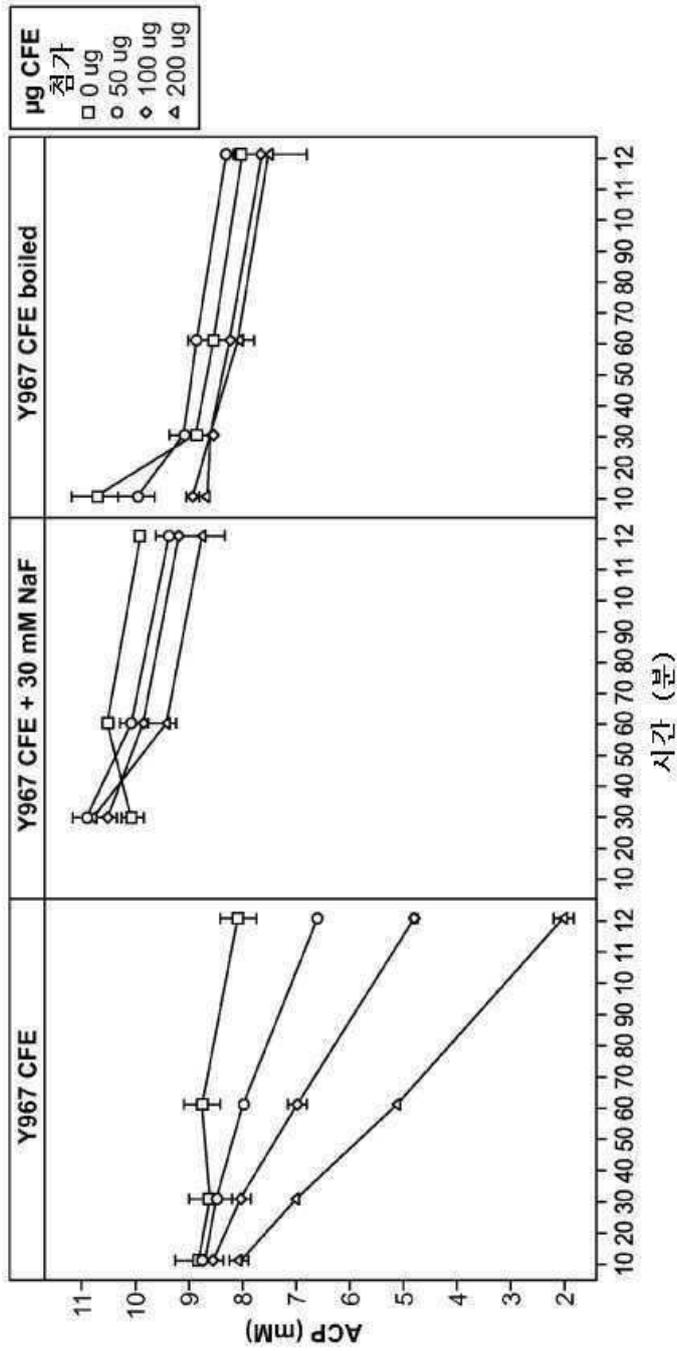


도면3cd

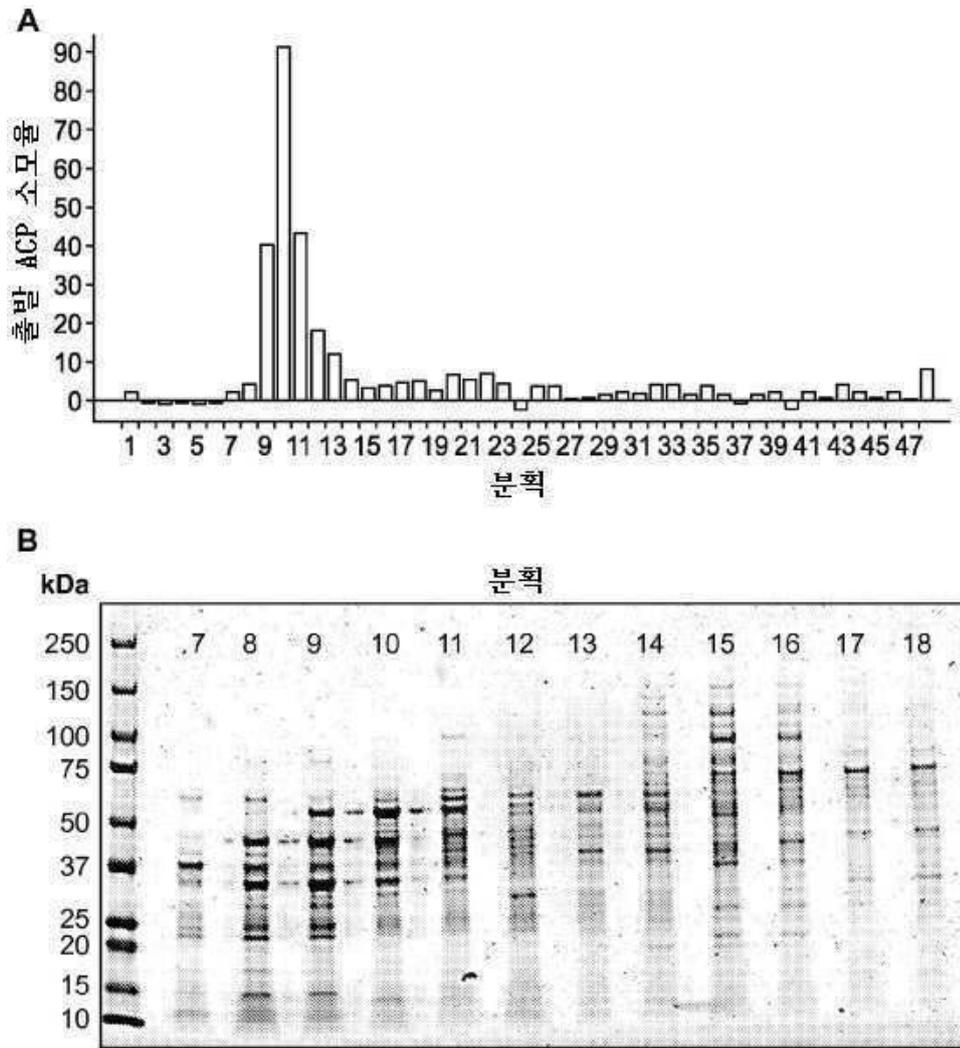
도 3C & 3D



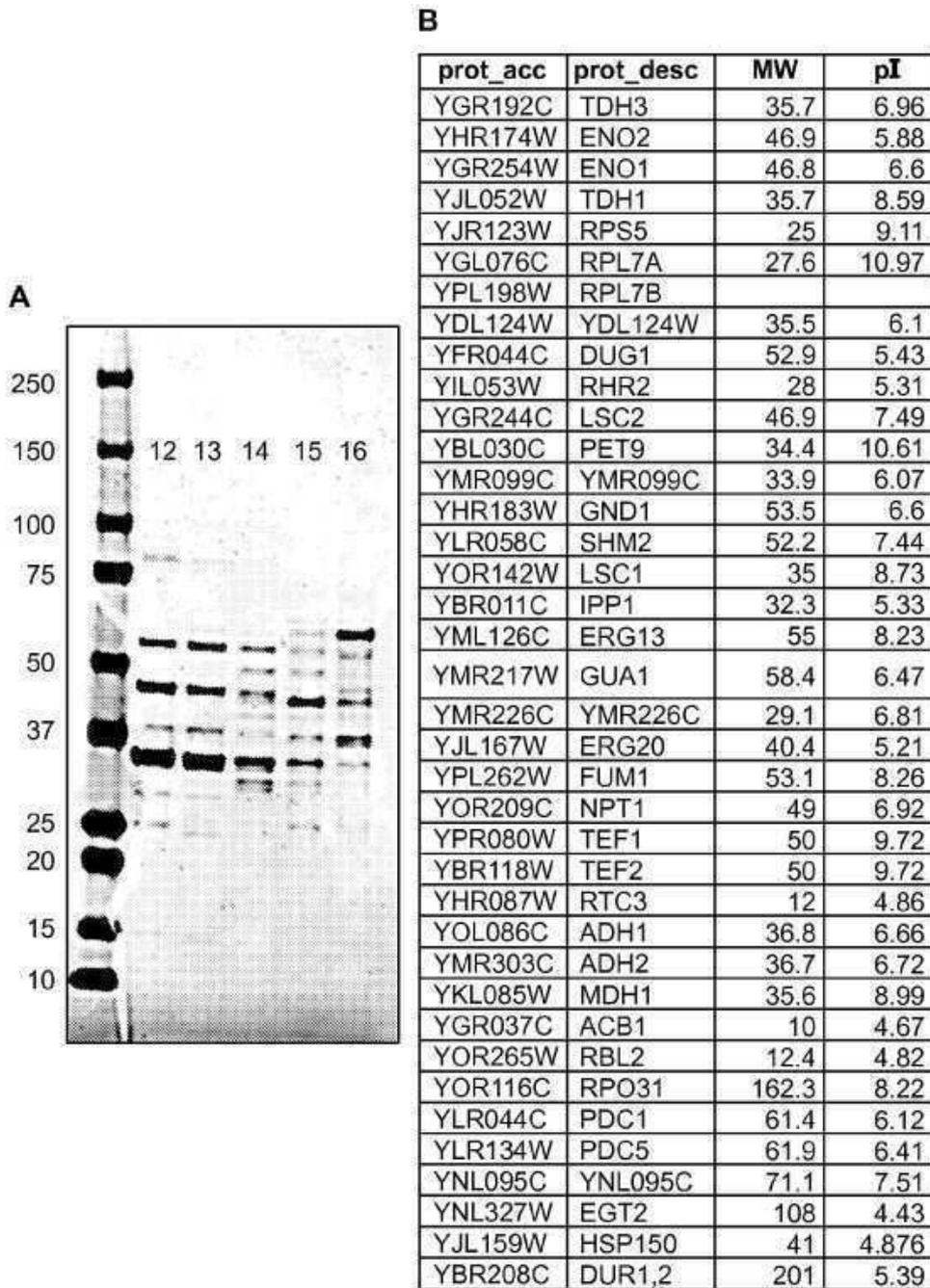
도면4



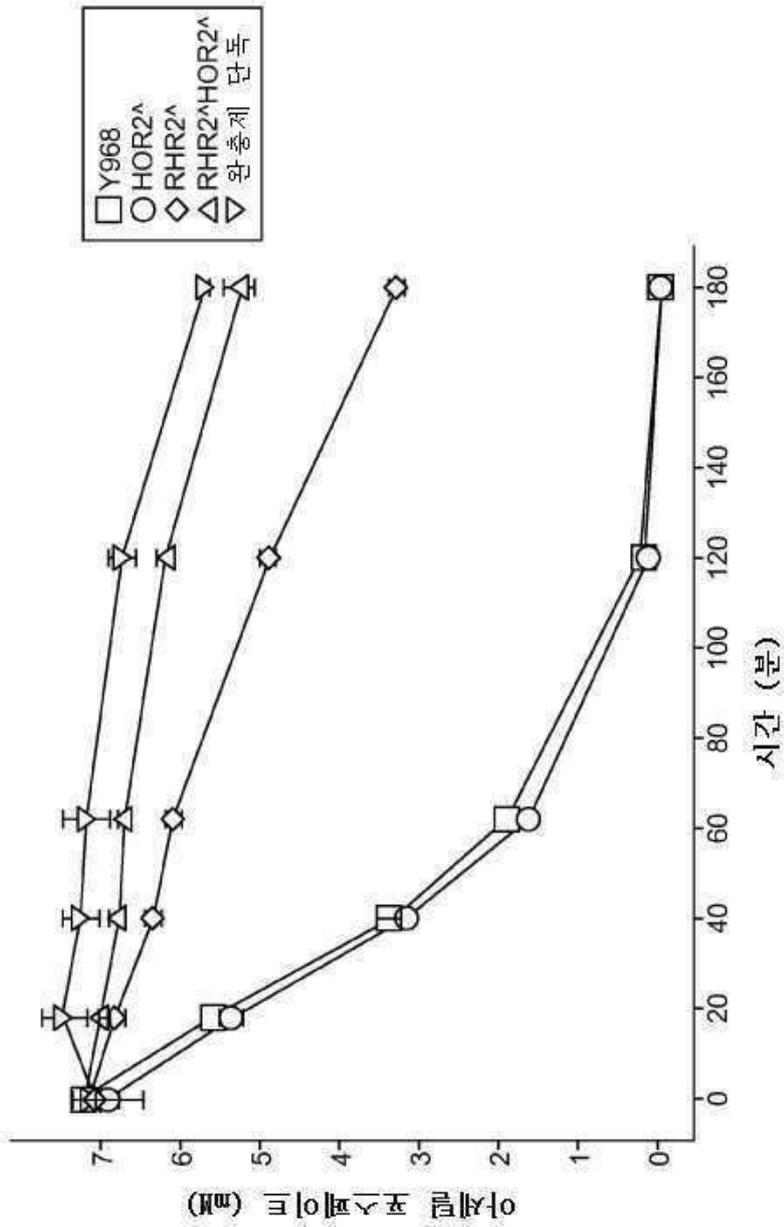
도면5



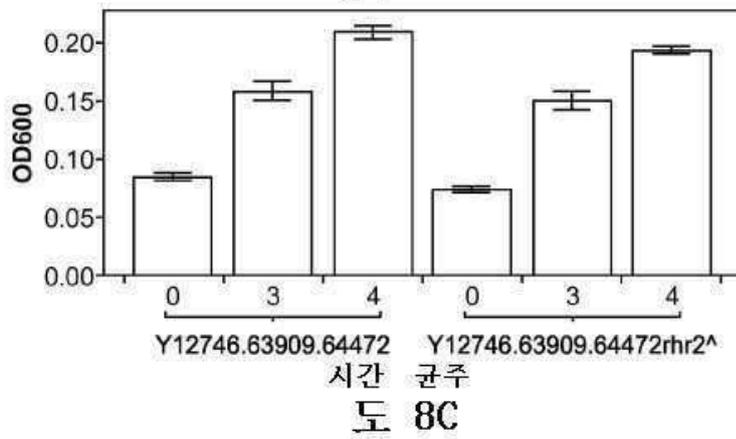
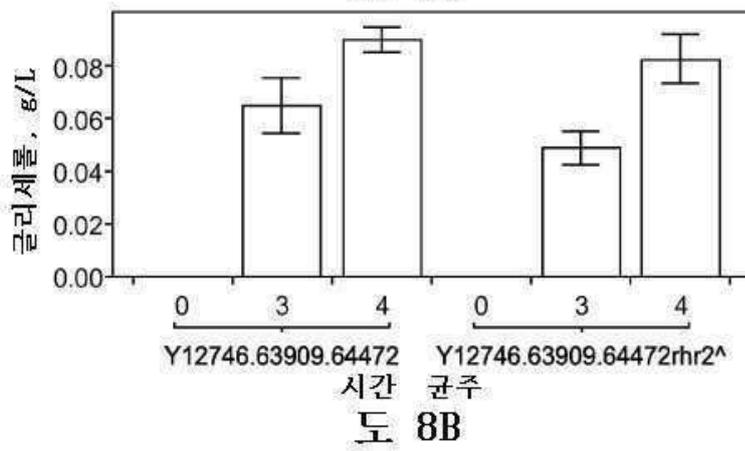
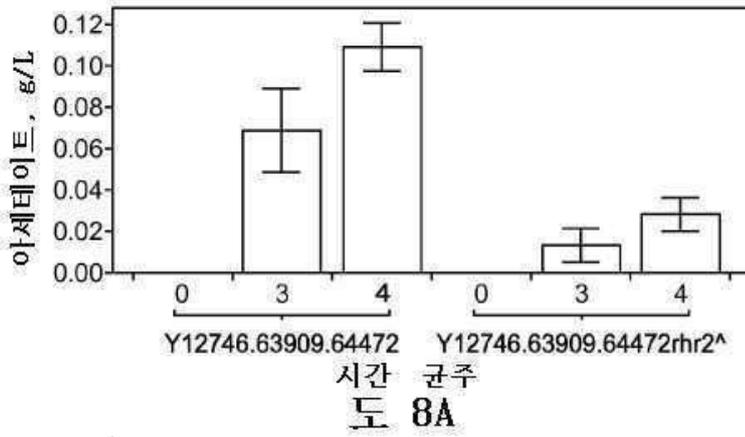
도면6



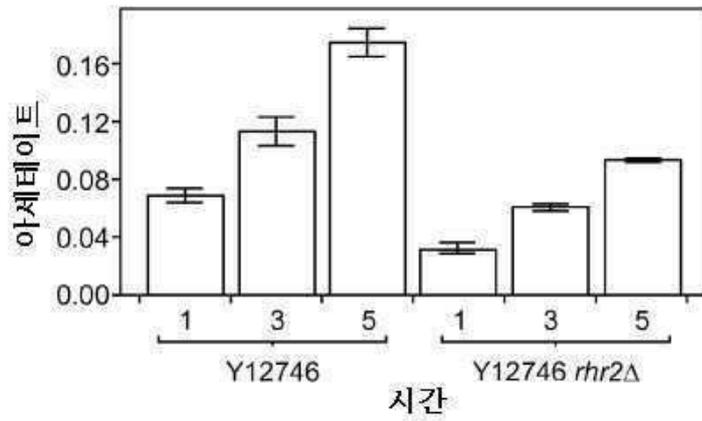
도면7



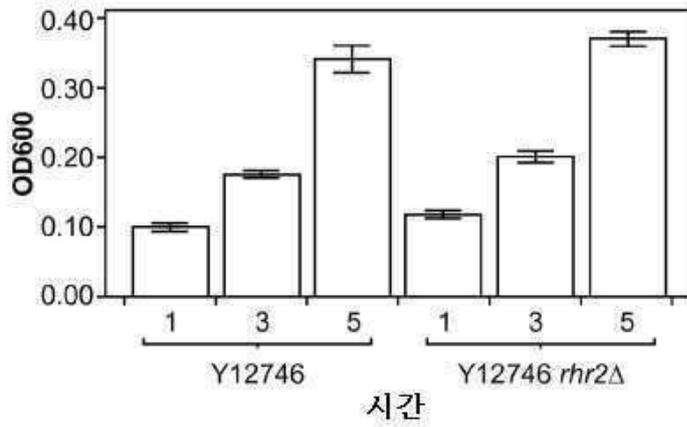
도면8abc



도면 8de

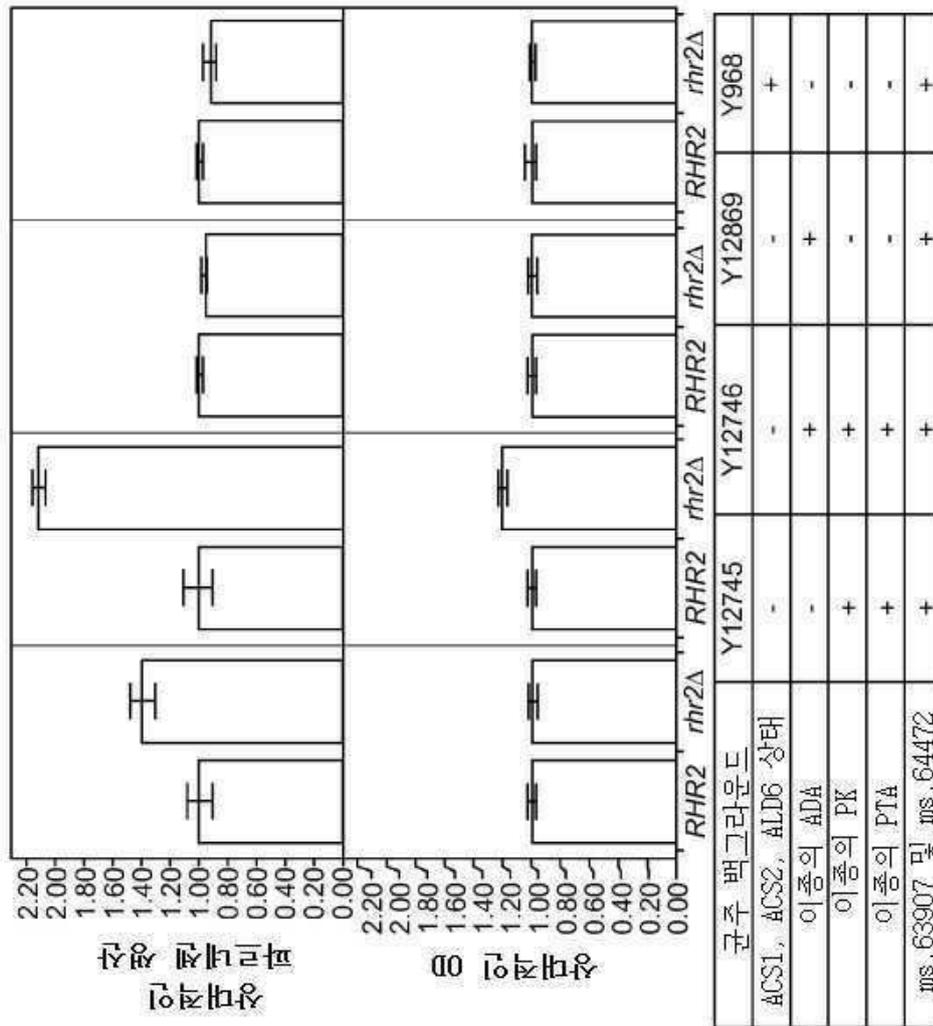


도 8D



도 8E

도면9



서열목록

SEQUENCE LISTING

<110> HAWKINS, Kristy Michelle

MEADOWS, Adam Leon

TSONG, Annie Ening

MAHATDEJKUL-MEADOWS, Tina Tipawan

PICKENS, Lauren Barbara

TAI, Anna

<120> USE OF PHOSPHOKETOLASE AND PHOSPHOTRANSACETYLASE FOR PRODUCTION OF ACETYL-COENZYME A DERIVED COMPOUNDS

<130> 107345.00466

<150> US 61/800,356

<151> 2013-03-15

<160> 37  
 <170> PatentIn version 3.5  
 <210> 1  
 <211> 2749  
 <212> DNA  
 <213> Leuconostoc mesenteroides

<220><221> misc\_feature  
 <222> (1)..(2749)  
 <223> Leuconostoc mesenteroides Phosphoketolase (PK) gene sequence  
 <400> 1

```

gttacggaag aagtctggt ttacggtgt tatgattctt gcaaaaaata aggagtactt      60
aatctcatgg cagatttcga ttcaaaagag tacttggaac ttgttgataa gtggtggcgc      120
gcaactaact atttgtcagc tgggatgata tttttgaaga gcaaccatt gttctcagtt      180
actaatacac ctatcaaggc tgaagatgia aaagttaagc caatcggaca ctggggtact      240
atctcaggtc agacattctt gtatgcacat gctaaccgtt tgatcaacaa gtatggtttg      300

aacatgtttt acgttgggtg tcttggteac ggtggccaag ttatggttac taacgcttac      360
ttagacggcg catatactga agattatcct gaaattactc aagataticga aggtatgagc      420
cacttgttca agcgtttctc attccctggc ggtattggat cacacatgac agctcaaaaa      480
cctggttcat tacacgaagg tggatgaattg ggtatttcat tgagccacgc ttttgggtgcc      540
gttttggaca atctgacca agttgctttc gcagttgttg gtgatggtga agctgaaaca      600
ggtccttcaa tggcttcatg gcaactcaatt aagtttttga atgctaagaa tgatggtgcc      660
gttttgcctg tcttggattt gaacggattc aagatttcaa acccaactat cttctcacgt      720

atgagtgatg aagaaatcac aaagtctttt gaaggtttgg gttattcacc tcgcttcac      780
gaaaacgatg atattcatga ctacgcaaca tatcaccaac ttgcagcaaa cattttggat      840
caagctattg aagatattca agctattcaa aatgatgcac gtgaaaatgg taagtatcaa      900
gatggtgaaa tccttgcagc gccagtaatt attgctcgtc tgccaaaggg ctggggtgga      960
ccaacgcacg atgcaagtaa caatcctatt gaaaactcat tccgtgcgca ccaagtgcca      1020
ttgcctcttg aacaacacga tcttgaacaa ttgcctgaat tcgaagactg gatgaactca      1080
tacaagcctg aagaattatt caatgctgat ggttctttga aggatgaatt gaaagctatc      1140

gctcctaagg gtgacaagcg tatgtcagct aaccctatta caaatggtgg tgctgatcgt      1200
tcagacttga agttgcctaa ctggagagaa ttcgctaacg atatcaatga tgatacacgt      1260
    
```

ggtaaggaat tcgctgatag caagcgcaat atggacatgg caacattgtc aaactacttg 1320  
 ggtgctgttt cacaattgaa cccaactcgt ttccgcttct tcggtcctga tgaacaatg 1380  
 tcaaaccgtt tgtggggatt gttcaatggt acaccacgtc aatggatgga agaaatcaag 1440  
 gaaccacaag atcaattggt gagccctacg ggtcgcatga ttgattcaca attgtctgaa 1500  
 catcaagctg aaggttggtg tgaaggatat actttgactg gtcgtgttgg aatcttcgca 1560  
  
 tcatacagagt cattcttgcg tgtttcgat acaatggtta cgcaacactt caagtggttg 1620  
 cgtcacgctt cagaacaagc atggcgtaat gactatccat cattgaactt gattgcaact 1680  
 tcaactgctt tccaacaaga tcacaatgga tatactcacc aagatccagg tatgttgact 1740  
 cacttggctg aaaagaagtc taactttatt cgtgaatatt tgccagctga tgtaactca 1800  
 ttgttggctg ttcaagaacg tgctttctca gaacgtcata aggttaactt gttgattgct 1860  
 tctaagcaac cacgtcaaca atggtttaca gttgaagaag ctgaagtatt ggctaacgaa 1920  
 ggtttgaaga tcattgattg ggcttctact gcaccttcta gtgatgttga tattacattc 1980  
  
 gcatctgctg gtactgaacc aacaattgaa actttggctg ctttgtggtt gattaaccaa 2040  
 gcattccag atgttaagtt ccgttatggt aacgttgttg aattactacg ttgcaaaag 2100  
 aagtcagaac ctaacatgaa tgatgaacgt gaattatcag ccgaagaatt caacaagtat 2160  
 ttccaagctg atacaccagt tatcttcggt ttccatgctt atgaaaactt gattgaaatca 2220  
 ttcttcttcg aacgtaagtt cacgggtgat gtatacgttc atggatatcg tgaagatggt 2280  
 gacatcacia cgacatatga tatgcgtgta tattcacact tggatcgctt ccatcaagct 2340  
 aaggaagctg ctgaaatctt gtctgcaaat ggtaagattg atcaagctgc tgctgataca 2400  
  
 ttcatcgcta agatggatga tactttggca aagcatttcc aagtactcg taacgaaggt 2460  
 cgtgatatcg aagaattcac tgactggaca tggtcaccac ttaagtaatt taaaattatt 2520  
 ttatcaaaac caactattat ttttaatagt tggttttttt atggctaaat tgactacata 2580  
 ctaaacgaaa ccatgtaaaa gtgccacata gttttactta ataagttcct tttatttttt 2640  
 gatttgaat gcaaaattgt aagcgttaata tgaataataa aaaccccaaa ttagttagct 2700  
 aattgggggt tttgtaaate accatateag ccgctcatag tcttagacg 2749  
  
 <210> 2  
 <211> 813  
 <212> PRT  
 <213> Leuconostoc mesenteroides  
  
 <220><221> misc\_feature  
 <222> (1)..(813)

<223> Leuconostoc mesenteroides Phosphoketolase (PK) protein sequence

<400> 2

Met Ala Asp Phe Asp Ser Lys Glu Tyr Leu Glu Leu Val Asp Lys Trp

1                    5                    10                    15

Trp Arg Ala Thr Asn Tyr Leu Ser Ala Gly Met Ile Phe Leu Lys Ser

                  20                    25                    30

Asn Pro Leu Phe Ser Val Thr Asn Thr Pro Ile Lys Ala Glu Asp Val

                  35                    40                    45

Lys Val Lys Pro Ile Gly His Trp Gly Thr Ile Ser Gly Gln Thr Phe

                  50                    55                    60

Leu Tyr Ala His Ala Asn Arg Leu Ile Asn Lys Tyr Gly Leu Asn Met

65                    70                    75                    80

Phe Tyr Val Gly Gly Pro Gly His Gly Gly Gln Val Met Val Thr Asn

                  85                    90                    95

Ala Tyr Leu Asp Gly Ala Tyr Thr Glu Asp Tyr Pro Glu Ile Thr Gln

                  100                    105                    110

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

                  115                    120                    125

Gly Ile Gly Ser His Met Thr Ala Gln Thr Pro Gly Ser Leu His Glu

                  130                    135                    140

Gly Gly Glu Leu Gly Tyr Ser Leu Ser His Ala Phe Gly Ala Val Leu

145                    150                    155                    160

Asp Asn Pro Asp Gln Val Ala Phe Ala Val Val Gly Asp Gly Glu Ala

                  165                    170                    175

Glu Thr Gly Pro Ser Met Ala Ser Trp His Ser Ile Lys Phe Leu Asn

                  180                    185                    190

Ala Lys Asn Asp Gly Ala Val Leu Pro Val Leu Asp Leu Asn Gly Phe

                  195                    200                    205

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

                  210                    215                    220

Thr Lys Phe Phe Glu Gly Leu Gly Tyr Ser Pro Arg Phe Ile Glu Asn



Ala Glu Gly Trp Leu Glu Gly Tyr Thr Leu Thr Gly Arg Val Gly Ile  
 485 490 495

Phe Ala Ser Tyr Glu Ser Phe Leu Arg Val Val Asp Thr Met Val Thr  
 500 505 510

Gln His Phe Lys Trp Leu Arg His Ala Ser Glu Gln Ala Trp Arg Asn  
 515 520 525

Asp Tyr Pro Ser Leu Asn Leu Ile Ala Thr Ser Thr Ala Phe Gln Gln  
 530 535 540

Asp His Asn Gly Tyr Thr His Gln Asp Pro Gly Met Leu Thr His Leu  
 545 550 555 560

Ala Glu Lys Lys Ser Asn Phe Ile Arg Glu Tyr Leu Pro Ala Asp Gly  
 565 570 575

Asn Ser Leu Leu Ala Val Gln Glu Arg Ala Phe Ser Glu Arg His Lys  
 580 585 590

Val Asn Leu Leu Ile Ala Ser Lys Gln Pro Arg Gln Gln Trp Phe Thr  
 595 600 605

Val Glu Glu Ala Glu Val Leu Ala Asn Glu Gly Leu Lys Ile Ile Asp  
 610 615 620

Trp Ala Ser Thr Ala Pro Ser Ser Asp Val Asp Ile Thr Phe Ala Ser  
 625 630 635 640

Ala Gly Thr Glu Pro Thr Ile Glu Thr Leu Ala Ala Leu Trp Leu Ile  
 645 650 655

Asn Gln Ala Phe Pro Asp Val Lys Phe Arg Tyr Val Asn Val Val Glu  
 660 665 670

Leu Leu Arg Leu Gln Lys Lys Ser Glu Pro Asn Met Asn Asp Glu Arg  
 675 680 685

Glu Leu Ser Ala Glu Glu Phe Asn Lys Tyr Phe Gln Ala Asp Thr Pro  
 690 695 700

Val Ile Phe Gly Phe His Ala Tyr Glu Asn Leu Ile Glu Ser Phe Phe  
 705 710 715 720

Phe Glu Arg Lys Phe Thr Gly Asp Val Tyr Val His Gly Tyr Arg Glu

725 730 735  
 Asp Gly Asp Ile Thr Thr Thr Tyr Asp Met Arg Val Tyr Ser His Leu  
 740 745 750

Asp Arg Phe His Gln Ala Lys Glu Ala Ala Glu Ile Leu Ser Ala Asn  
 755 760 765

Gly Lys Ile Asp Gln Ala Ala Ala Asp Thr Phe Ile Ala Lys Met Asp  
 770 775 780

Asp Thr Leu Ala Lys His Phe Gln Val Thr Arg Asn Glu Gly Arg Asp  
 785 790 795 800

Ile Glu Glu Phe Thr Asp Trp Thr Trp Ser Pro Leu Lys  
 805 810

<210> 3

<211> 1002

<212> DNA

<213> Clostridium kluveri

<220><221> misc\_feature

<222> (1)..(1002)

<223> Clostridium kluveri Phosphotransacetylase (PTA) gene sequence

<400> 3

atgaaattaa tggaaaatat ttttggttta gccaaagcag ataagaaaa aattgttttg 60  
 gcagaaggag aagaagaaag gaacattaga gcttccgaag aaataataag ggatggtatt 120  
 gcagatataa ttttagtagg aagtgaaagt gtaataaaag agaatgcagc taaatttggg 180  
 gtaacttag ctggagtgga aatagtagat cctgaaactt caagtaaaac tgcaggctat 240  
 gccaatgctt tttatgaat tagaagaat aaaggagtta cactggaaaa agcagataaa 300  
  
 atagttagag atcctatata ttttgcaaca atgatggtga aacttgagaga tgcagatggt 360  
 ttagtttcag gtgcaatata tacaacggga gatcttttga gaccaggact tcaaatagtg 420  
 aagacagttc caggtgcttc tgtggtttcc agtgtatfff taatgagtgt accagattgt 480  
 gaatatggag aagatggatt cttgttattt gctgattgtg ctgtaaatgt atgtcctact 540  
 gctgaagaat tatcttcaat tgcaataact acagcagaaa ctgcaaaaaa tttgtgtaaa 600  
 atagaaccaa gagttgccat gctttcattt tctactatgg gaagtgctag tcatgaattg 660  
 gtagataaag ttacaaaagc aacaaaactt gctaaagaag ctagacctga tttggatata 720

gatggagaac ttcaattgga tgcttccctia gtaaaaaaag ttgcagactt aaaagctccg 780  
 ggcagtaaag tggcaggaaa agccaatgta cttatatcc ctgatataca agcaggaaat 840  
 ataggatata agttagtcca aagatttgca aaagctgagg ctataggacc tatatgtcag 900  
 ggatttgcaa agcctataaa tgatttatca agaggctgca gcgttgatga tatagtaaag 960  
 gtagtggctg taactgcagt tcaagcacag gcacagggtt ag 1002

<210> 4

<211> 333

<212> PRT

<213> Clostridium kluveri

<220><221> misc\_feature

<222> (1)..(333)

<223> Clostridium kluveri Phosphotransacetylase (PTA) protein sequence

<400> 4

Met Lys Leu Met Glu Asn Ile Phe Gly Leu Ala Lys Ala Asp Lys Lys

1 5 10 15

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

20 25 30

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

35 40 45

Glu Ser Val Ile Lys Glu Asn Ala Ala Lys Phe Gly Val Asn Leu Ala

50 55 60

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

65 70 75 80

Ala Asn Ala Phe Tyr Glu Ile Arg Lys Asn Lys Gly Val Thr Leu Glu

85 90 95

Lys Ala Asp Lys Ile Val Arg Asp Pro Ile Tyr Phe Ala Thr Met Met

100 105 110

Val Lys Leu Gly Asp Ala Asp Gly Leu Val Ser Gly Ala Ile His Thr

115 120 125

Thr Gly Asp Leu Leu Arg Pro Gly Leu Gln Ile Val Lys Thr Val Pro

130 135 140

Gly Ala Ser Val Val Ser Ser Val Phe Leu Met Ser Val Pro Asp Cys

145                    150                    155                    160  
 Glu Tyr Gly Glu Asp Gly Phe Leu Leu Phe Ala Asp Cys Ala Val Asn  
                                  165                    170                    175  
 Val Cys Pro Thr Ala Glu Glu Leu Ser Ser Ile Ala Ile Thr Thr Ala  
                                  180                    185                    190  
  
 Glu Thr Ala Lys Asn Leu Cys Lys Ile Glu Pro Arg Val Ala Met Leu  
                                  195                    200                    205  
 Ser Phe Ser Thr Met Gly Ser Ala Ser His Glu Leu Val Asp Lys Val  
                                  210                    215                    220  
 Thr Lys Ala Thr Lys Leu Ala Lys Glu Ala Arg Pro Asp Leu Asp Ile  
 225                    230                    235                    240  
 Asp Gly Glu Leu Gln Leu Asp Ala Ser Leu Val Lys Lys Val Ala Asp  
                                  245                    250                    255  
  
 Leu Lys Ala Pro Gly Ser Lys Val Ala Gly Lys Ala Asn Val Leu Ile  
                                  260                    265                    270  
 Phe Pro Asp Ile Gln Ala Gly Asn Ile Gly Tyr Lys Leu Val Gln Arg  
                                  275                    280                    285  
 Phe Ala Lys Ala Glu Ala Ile Gly Pro Ile Cys Gln Gly Phe Ala Lys  
                                  290                    295                    300  
 Pro Ile Asn Asp Leu Ser Arg Gly Cys Ser Val Asp Asp Ile Val Lys  
 305                    310                    315                    320

Val Val Ala Val Thr Ala Val Gln Ala Gln Ala Gln Gly  
                                  325                    330

- <210> 5
- <211> 753
- <212> DNA
- <213> *S. cerevisiae*
- <220><221> misc\_feature
- <222> (1)..(753)
- <223> Nucleotide sequence of GPP1/RHR2 of *S. cerevisiae*
- <400> 5

atgcctttga ccacaaaacc tttatctttg aaaatcaacg ccgctctatt cgatgttgac 60

ggtaccatca tcattctctca accagccatt gctgctttct ggagagattt cggtaaagac 120  
 aagccttact tcgatgccga acacgttatt cacatctctc acggttggag aacttacgat 180

gccattgcc a gttcgetcc agactttgct gatgaagaat acgtaacaa getagaaggt 240  
 gaaatcccag aaaagtacgg tgaacctcc atcgaagttc caggtgctgt caagttgtgt 300  
 aatgctttga acgccttgcc aaaggaaaaa tgggctgtcg ccacctctgg taccctgac 360  
 atggccaaga aatggttga cttttgaag atcaagagac cagaatactt catcacgcc 420  
 aatgatgtca agcaaggtaa gcctcacca gaaccatact taaagggtag aaacggtttg 480  
 ggtttcccaa ttaatgaaca agaccatcc aaatctaagg ttgtgtctt tgaagacgca 540  
 ccagctggta ttgctgctgg taaggctgct ggctgtaaaa tcgttggat tgctaccact 600

ttcgatttgg acttcttgaa ggaaaaggt tgtgacatca ttgtcaagaa ccacgaatct 660  
 atcagagtgc gtgaatacaa cgctgaaacc gatgaagtcg aattgatctt tgatgactac 720  
 ttatacgcta aggatgactt gttgaaatgg taa 753

<210> 6

<211> 250

<212> PRT

<213> S. cerevisiae

<220><221> misc\_feature

<222> (1)..(250)

<223> Protein sequence of Gpp1/Rhr2 of S. cerevisiae

<400> 6

Met Pro Leu Thr Thr Lys Pro Leu Ser Leu Lys Ile Asn Ala Ala Leu

1                    5                    10                    15

Phe Asp Val Asp Gly Thr Ile Ile Ile Ser Gln Pro Ala Ile Ala Ala

20                    25                    30

Phe Trp Arg Asp Phe Gly Lys Asp Lys Pro Tyr Phe Asp Ala Glu His

35                    40                    45

Val Ile His Ile Ser His Gly Trp Arg Thr Tyr Asp Ala Ile Ala Lys

50                    55                    60

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

65                    70                    75                    80

Glu Ile Pro Glu Lys Tyr Gly Glu His Ser Ile Glu Val Pro Gly Ala

85 90 95  
 Val Lys Leu Cys Asn Ala Leu Asn Ala Leu Pro Lys Glu Lys Trp Ala  
 100 105 110  
 Val Ala Thr Ser Gly Thr Arg Asp Met Ala Lys Lys Trp Phe Asp Ile  
 115 120 125  
 Leu Lys Ile Lys Arg Pro Glu Tyr Phe Ile Thr Ala Asn Asp Val Lys  
 130 135 140

Gln Gly Lys Pro His Pro Glu Pro Tyr Leu Lys Gly Arg Asn Gly Leu  
 145 150 155 160  
 Gly Phe Pro Ile Asn Glu Gln Asp Pro Ser Lys Ser Lys Val Val Val  
 165 170 175  
 Phe Glu Asp Ala Pro Ala Gly Ile Ala Ala Gly Lys Ala Ala Gly Cys  
 180 185 190  
 Lys Ile Val Gly Ile Ala Thr Thr Phe Asp Leu Asp Phe Leu Lys Glu  
 195 200 205

Lys Gly Cys Asp Ile Ile Val Lys Asn His Glu Ser Ile Arg Val Gly  
 210 215 220  
 Glu Tyr Asn Ala Glu Thr Asp Glu Val Glu Leu Ile Phe Asp Asp Tyr  
 225 230 235 240  
 Leu Tyr Ala Lys Asp Asp Leu Leu Lys Trp  
 245 250

<210> 7  
 <211> 753  
 <212> DNA  
 <213> *S. cerevisiae*  
 <220><221> misc\_feature  
 <222> (1)..(753)  
 <223> Nucleotide sequence of GPP2/HOR2 of *S. cerevisiae*

<400> 7  
 atgggattga ctactaaacc tctatctttg aaagttaacg ccgctttggt cgacgtcgac 60  
 ggtaccatta tcattctca accagccatt gctgcattct ggagggattt cgtaaggac 120  
 aaaccttatt tcgatgctga acacgttatc caagtctcgc atggttgag aacgtttgat 180

gccattgcta agttcgctcc agactttgcc aatgaagagt atgtaacaa attagaagct 240  
 gaaattccgg tcaagtacgg tgaaaaatcc attgaagtcc caggtgcagt taagctgtgc 300  
 aacgctttga acgctctacc aaaagagaaa tgggctgtgg caacttccgg taccctgat 360  
 atggcacaaa aatggttcga gcatctggga atcaggagac caaagtactt cattaccgct 420

aatgatgtca aacagggtaa gcctcatcca gaaccatata tgaagggcag gaatggctta 480  
 ggatatccga tcaatgagca agacccttcc aatctaagg tagtagtatt tgaagacgct 540  
 ccagcaggta ttccgccggg aaaagccgcc ggttgtaaga tcattggtat tgccactact 600  
 ttcgacttgg acttctctaa ggaaaaaggc tgtgacatca ttgtcaaaaa ccacgaatcc 660  
 atcagagtgg gcggtacaaa tgccgaaaca gacgaagttg aattcatttt tgacgactac 720  
 ttatatgcta aggacgatct gttgaaatgg taa 753

<210> 8

<211> 250

<212> PRT

<213> *S. cerevisiae*

<220><221> misc\_feature

<222> (1)..(250)

<223> Protein sequence of Gpp2/Hor2 of *S. cerevisiae*

<400> 8

Met Gly Leu Thr Thr Lys Pro Leu Ser Leu Lys Val Asn Ala Ala Leu

1                    5                    10                    15

Phe Asp Val Asp Gly Thr Ile Ile Ile Ser Gln Pro Ala Ile Ala Ala

                  20                    25                    30

Phe Trp Arg Asp Phe Gly Lys Asp Lys Pro Tyr Phe Asp Ala Glu His

                  35                    40                    45

Val Ile Gln Val Ser His Gly Trp Arg Thr Phe Asp Ala Ile Ala Lys

                  50                    55                    60

Phe Ala Pro Asp Phe Ala Asn Glu Glu Tyr Val Asn Lys Leu Glu Ala

65                    70                    75                    80

Glu Ile Pro Val Lys Tyr Gly Glu Lys Ser Ile Glu Val Pro Gly Ala

                  85                    90                    95

Val Lys Leu Cys Asn Ala Leu Asn Ala Leu Pro Lys Glu Lys Trp Ala



gataaataca ttaaaaatgt ttcacaggct cgccatacac cggcattga atgtctgagc 300  
 gcggaagttc tgacaggcga caatggcctg aactgattg aaaatgcccc ttggggagtg 360  
 gtggcttccg tgacccaag cacgaacca gccgccacag tcatcaataa tgcaatttcc 420  
 atgattgctg caggaattc agtcgttttt gcaccgcacc catccgcaa aaatgtgtcc 480  
 ttaccacaa tatcgttct taacaaagca attgtggcga cagggtggcc agaaaatctg 540  
 ctggtatccg tcgcaaatcc caacatcga acagctcaac gcctgttccg ttatccaggt 600  
 attggattac tcgtcgtaac aggtggtgag gcggtggtgg aagcggcgcg caaacacact 660

gataaacgtt taattgccgc aggcgccgga aacccccag tagtcttga cgaacagcg 720  
 gatataccga aagccgctcg cgcaatagta aagggcgtt cgtttgaaa caatattatt 780  
 tgtgccgacg agaaagtatt aatcgtggt gatcgcgtag ccgacgcctt attagccgaa 840  
 atgcaacgca acaatgctgt tttactgacg cctgaacaga cagaacgact tctgcccgt 900  
 ttgctgagcg atatagatga gcaggggaag ggacgcgtga accgcgatta tgtggggagg 960  
 gatgccgcta aactagcggc ggccattggt ttagaagtgt cagaacacac aagattatta 1020  
 ctgctgaaa cagatgctga tcatccttt gcagtaaccg aattaatgat gccctattg 1080

cctgttatcc gtgtaaaaa cgttgatgac gccattgcc tcgctgtaa acttgagagt 1140  
 ggttgtcgtc aactgcagc aatgcattcg acaaacatta ggaacctgaa tcggatggca 1200  
 aatgctataa atacatcaat tttgttaaa aatggtccgt gtatcgtgg gctgggcctg 1260  
 ggtggcgagg gctggacgtc gatgactata tctacacca caggggaagg agttacctca 1320  
 gcacgcacct tcgtacgttt acgtagatgt gtattggtg acatgttcag aatcgcgtaa 1380

<210> 10

<211> 459

<212> PRT

<213> Dickeya zeae

<220><221> misc\_feature

<222> (1)..(459)

<223> Dickeya zeae eutE protein sequence

<400> 10

Met Glu His Ser Val Ile Glu Pro Thr Val Pro Met Pro Leu Pro Ala  
 1                    5                    10                    15  
 Met Phe Asp Ala Pro Ser Gly Ile Phe Ser Ser Leu Asp Asp Ala Val  
                   20                    25                    30  
 Gln Ala Ala Thr Leu Ala Gln Gln Gln Leu Ser Ser Val Glu Leu Arg



Leu Thr Pro Glu Gln Thr Glu Arg Leu Leu Pro Ala Leu Leu Ser Asp  
 290 295 300

Ile Asp Glu Gln Gly Lys Gly Arg Val Asn Arg Asp Tyr Val Gly Arg  
 305 310 315 320

Asp Ala Ala Lys Leu Ala Ala Ala Ile Gly Leu Glu Val Ser Glu His  
 325 330 335

Thr Arg Leu Leu Leu Ala Glu Thr Asp Ala Asp His Pro Phe Ala Val  
 340 345 350

Thr Glu Leu Met Met Pro Val Leu Pro Val Ile Arg Val Lys Asn Val  
 355 360 365

Asp Asp Ala Ile Ala Leu Ala Val Lys Leu Glu Ser Gly Cys Arg His  
 370 375 380

Thr Ala Ala Met His Ser Thr Asn Ile Arg Asn Leu Asn Arg Met Ala  
 385 390 395 400

Asn Ala Ile Asn Thr Ser Ile Phe Val Lys Asn Gly Pro Cys Ile Ala  
 405 410 415

Gly Leu Gly Leu Gly Gly Glu Gly Trp Thr Ser Met Thr Ile Ser Thr  
 420 425 430

Pro Thr Gly Glu Gly Val Thr Ser Ala Arg Thr Phe Val Arg Leu Arg  
 435 440 445

Arg Cys Val Leu Val Asp Met Phe Arg Ile Ala  
 450 455

<210> 11

<211> 1798

<212> DNA

<213> Saccharomyces cerevisiae

<220><221> misc\_feature

<222> (1)..(1798)

<223> Saccharomyces cerevisiae ALD4 nucleotide sequence

<400> 11

gcaccaggg acacacagca gccaagtatt ttcagaatgt tcagtagatc tacgctctgc 60

ttaaagacgt ctgcatcctc cattgggaga cttcaattga gatatttctc acaccttct 120

atgacagtgc ctatcaagct gcccaatggg ttggaatatg agcaaccaac ggggttgttc 180

atcaacaaca agtttgttcc ttctaacag aacaagacct tcgaagtcac taacccttcc 240

acggaagaag aaatatgtca tatttatgaa ggtagagagg acgatgtgga agaggccgtg 300

caggccgccg accgtgcctt ctctaattggg tcttggaacg gtatcgacc tattgacagg 360

ggtaaggctt tgtacaggtt agccgaatta attgaacagg acaaggatgt cattgcttcc 420

atcgagactt tggataacgg taaagctatc tcttcctcga gaggagatgt tgatttagtc 480

atcaactatt tgaatcttc tgctggcttt gctgataaaa ttgatgtag aatgattgat 540

actggtagaa cccatttttc ttactaactaag agacagcctt tgggtgtttg tgggcagatt 600

attccttggga atttcccact gttgatgtgg gcctggaaga ttgccctgc tttggtcacc 660

ggtaacaccg tcgtgttgaa gactgccgaa tccaccccat tgtccgcttt gtatgtgtct 720

aaatacatcc cacaggcggg tattccacct ggtgtgatca acattgtatc cgggtttggt 780

aagattgtgg gtgaggccat tacaacacat ccaaaaatca aaaaggttgc cttcacaggg 840

tccacggcta cgggtagaca catttaccag tccgcagccg caggcttgaa aaaagtgact 900

ttggagctgg gtggtaaatc accaaacatt gtcttcgagg acgccgagtt gaaaaaagcc 960

gtgcaaaaca ttatccttgg tatctactac aattctggtg aggtctgttg tgcgggttca 1020

agggtgtatg ttgaagaatc tatttacgac aaattcattg aagagttcaa agccgcttct 1080

gaatccatca aggtgggcga cccattcgat gaatctactt tccaaggtgc acaaacctct 1140

caaatgcaac taaacaaaat cttgaaatac gttgacattg gtaagaatga aggtgctact 1200

ttgattaccg gtggtgaaaag attaggtagc aagggttact tcattaagcc aactgtcttt 1260

ggtgacgtta aggaagacat gagaattgtc aaagaggaaa tctttggccc tgttgcact 1320

gtaaccaaat tcaaatctgc cgacgaagtc attaacatgg cgaacgattc tgaatacggg 1380

ttggtctgtg gtattcacac ctctaataat aataccgctt taaaagtggc tgatagagtt 1440

aatgcgggta cggcttggat aaacacttat aacgatttcc accacgcagt tcctttcggg 1500

gggttcaatg catctggttt gggcagggaa atgtctgttg atgctttaca aaactacttg 1560

caagttaaag cggtcctgtc caaattggac gagtaaggtc atcaataagc ctggtgtcca 1620

atcgatgctt acatacataa aattaaatac tctgtctctg ttatatttcc acatgtcatc 1680

atctcaataa tatgtacttt aaagaaaata aaataaaaaa taaaattttt ttctcccgat 1740

aatcaatttt cttaattaat taattgcggt acgaaacgag atcgccgacg ccgccgat 1798

<210> 12  
 <211> 519  
 <212> PRT

<213> *Saccharomyces cerevisiae*

<220><221> misc\_feature

<222> (1)..(519)

<223> *Saccharomyces cerevisiae* ALD4 protein sequence

<400> 12

Met Phe Ser Arg Ser Thr Leu Cys Leu Lys Thr Ser Ala Ser Ser Ile

1                    5                    10                    15

Gly Arg Leu Gln Leu Arg Tyr Phe Ser His Leu Pro Met Thr Val Pro

                  20                    25                    30

Ile Lys Leu Pro Asn Gly Leu Glu Tyr Glu Gln Pro Thr Gly Leu Phe

                  35                    40                    45

Ile Asn Asn Lys Phe Val Pro Ser Lys Gln Asn Lys Thr Phe Glu Val

50                    55                    60

Ile Asn Pro Ser Thr Glu Glu Glu Ile Cys His Ile Tyr Glu Gly Arg

65                    70                    75                    80

Glu Asp Asp Val Glu Glu Ala Val Gln Ala Ala Asp Arg Ala Phe Ser

                  85                    90                    95

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

                  100                    105                    110

Tyr Arg Leu Ala Glu Leu Ile Glu Gln Asp Lys Asp Val Ile Ala Ser

                  115                    120                    125

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

                  130                    135                    140

Val Asp Leu Val Ile Asn Tyr Leu Lys Ser Ser Ala Gly Phe Ala Asp

145                    150                    155                    160

Lys Ile Asp Gly Arg Met Ile Asp Thr Gly Arg Thr His Phe Ser Tyr

                  165                    170                    175

Thr Lys Arg Gln Pro Leu Gly Val Cys Gly Gln Ile Ile Pro Trp Asn

                  180                    185                    190

Phe Pro Leu Leu Met Trp Ala Trp Lys Ile Ala Pro Ala Leu Val Thr

                  195                    200                    205

Gly Asn Thr Val Val Leu Lys Thr Ala Glu Ser Thr Pro Leu Ser Ala  
 210 215 220  
 Leu Tyr Val Ser Lys Tyr Ile Pro Gln Ala Gly Ile Pro Pro Gly Val  
 225 230 235 240  
  
 Ile Asn Ile Val Ser Gly Phe Gly Lys Ile Val Gly Glu Ala Ile Thr  
 245 250 255  
 Asn His Pro Lys Ile Lys Lys Val Ala Phe Thr Gly Ser Thr Ala Thr  
 260 265 270  
 Gly Arg His Ile Tyr Gln Ser Ala Ala Ala Gly Leu Lys Lys Val Thr  
 275 280 285  
 Leu Glu Leu Gly Gly Lys Ser Pro Asn Ile Val Phe Ala Asp Ala Glu  
 290 295 300  
  
 Leu Lys Lys Ala Val Gln Asn Ile Ile Leu Gly Ile Tyr Tyr Asn Ser  
 305 310 315 320  
 Gly Glu Val Cys Cys Ala Gly Ser Arg Val Tyr Val Glu Glu Ser Ile  
 325 330 335  
 Tyr Asp Lys Phe Ile Glu Glu Phe Lys Ala Ala Ser Glu Ser Ile Lys  
 340 345 350  
 Val Gly Asp Pro Phe Asp Glu Ser Thr Phe Gln Gly Ala Gln Thr Ser  
 355 360 365  
  
 Gln Met Gln Leu Asn Lys Ile Leu Lys Tyr Val Asp Ile Gly Lys Asn  
 370 375 380  
 Glu Gly Ala Thr Leu Ile Thr Gly Gly Glu Arg Leu Gly Ser Lys Gly  
 385 390 395 400  
 Tyr Phe Ile Lys Pro Thr Val Phe Gly Asp Val Lys Glu Asp Met Arg  
 405 410 415  
 Ile Val Lys Glu Glu Ile Phe Gly Pro Val Val Thr Val Thr Lys Phe  
 420 425 430  
  
 Lys Ser Ala Asp Glu Val Ile Asn Met Ala Asn Asp Ser Glu Tyr Gly  
 435 440 445  
 Leu Ala Ala Gly Ile His Thr Ser Asn Ile Asn Thr Ala Leu Lys Val

450                      455                      460  
 Ala Asp Arg Val Asn Ala Gly Thr Val Trp Ile Asn Thr Tyr Asn Asp  
 465                      470                      475                      480  
 Phe His His Ala Val Pro Phe Gly Gly Phe Asn Ala Ser Gly Leu Gly  
                          485                      490                      495  
  
 Arg Glu Met Ser Val Asp Ala Leu Gln Asn Tyr Leu Gln Val Lys Ala  
                          500                      505                      510  
 Val Arg Ala Lys Leu Asp Glu  
                          515  
 <210> 13  
 <211> 2744  
 <212> DNA  
 <213> Saccharomyces cerevisiae  
 <220><221> misc\_feature  
 <222> (1)..(2744)  
 <223> Saccharomyces cerevisiae cytosolic aldehyde dehydrogenase 6  
                          (ALD6) nucleotide sequence  
 <400> 13  
 catatggcgt atccaagccg aaaccctttg cctcatcccc acggaataag gcagccgaca        60  
 aaagaaaaac gaccgaaaag gaaccagaaa gaaaaaagag ggtgggcgcg ccgccggacgt     120  
  
 gtaaaaagat atgcatccag cttctatata gctttaactt taccgttttg ggcacgcca        180  
 acgtatgtaa cattgatctc ctcttgggaa cggtagtgac aacagatgag atatagcacc        240  
 gaccatgtgg gcaaattcgt aataaattcg gggtaggggg gattcaagac aagcaacctt        300  
 gttatgcagc tcaaacagcg atttaacggt tgagtaacac atcaaaacac cgttcgaggt        360  
 caagcctggc gtgtttaaca agttcttgat atcatatata aatgtaataa gaagtttggg        420  
 aatattcaat tcgaagtgtt cagtctttta cttctcttgt tttatagaag aaaaaacatc        480  
 aagaaacatc ttaacatac acaaacacat actatcagaa tacaatgact aagctacact        540  
  
 ttgacactgc tgaaccagtc aagatcacac ttccaaatgg tttgacatac gagcaaccaa        600  
 ccggtctatt cattaacaac aagtttatga aagctcaaga cggtagacc tatcccgtcg        660  
 aagatccttc cactgaaaac accgtttgtg aggtctcttc tgccaccact gaagatgttg        720  
 aatatgctat cgaatgtgcc gaccgtgctt tccacgacac tgaatgggct acccaagacc        780  
 caagagaaag aggccgtcta ctaagtaagt tggtgacga attggaaagc caaattgact        840

tggtttcttc cattgaagct ttggacaatg gtaaaacttt ggcctttaag gcccgtaggg 900  
 atgttaccat tgcaatcaac tgtctaagag atgctgctgc ctatgccgac aagtcaacg 960  
  
 gtagaacaat caacaccggt gacggctaca tgaacttcac caccttagag ccaatcggtg 1020  
 tctgtggtca aattattcca tggaactttc caataatgat gttggcttgg aagatcgccc 1080  
 cagcattggc catgggtaac gtctgtatct tgaaacccgc tgctgtcaca cttttaaatg 1140  
 ccctatactt tgcttcttta tgtaagaagg ttggtattcc agctgggtgc gtcaacatcg 1200  
 ttccaggtcc tggtagaact gttggtgctg ctttgaccaa cgaccaaga atcagaaagc 1260  
 tggcttttac cggttctaca gaagtcggta agagtgttc tgctgactct tctgaatcta 1320  
 acttgaagaa aatcactttg gaactagggtg gtaagtccgc ccatttggtc tttgacgatg 1380  
  
 ctaacattaa gaagacttta ccaaatctag taaacggtat tttcaagaac gctggtcaaa 1440  
 tttgttctct tggttctaga atttacgttc aagaaggtat ttacgacgaa ctattggctg 1500  
 ctttcaaggc ttacttggaa accgaaatca aagttggtaa tccatttgac aaggctaact 1560  
 tccaaggtgc taccactaac cgtcaacaat tcgacacaat tatgaactac atcgatatcg 1620  
 gtaagaaaga aggcgccaag atcttaactg gtggcgaaaa agttggtgac aagggttact 1680  
 tcatcagacc aaccgttttc tacgatgta atgaagacat gagaattggt aaggaagaaa 1740  
 tttttggacc agttgtcact gtcgcaaagt tcaagacttt agaagaaggt gtcgaaatgg 1800  
  
 ctaacagctc tgaattcgggt ctaggttctg gtatcgaac agaacttttg agcacaggtt 1860  
 tgaaggtggc caagatgttg aaggccggtg ccgtctggat caacacatac aacgattttg 1920  
 actccagagt tccattcgggt ggtgtaagc aatctggtta cggtagagaa atgggtgaag 1980  
 aagtctacca tgcatacact gaagtaaaag ctgtcagaat taagttgtaa tgtaccaacc 2040  
 tgcatttctt tccgtcatat acacaaaata ctttcatata aacttacttg gtcttacgtc 2100  
 ataaataaat atgtatacat ataaataaa aaatttgggt ttatattttt acaaaaagaa 2160  
 tcgtttactt catttctccc ttttaagcga tacaatccat gaaaaagag aaaaagagag 2220  
  
 aacaggttg tgccttcttt aaaacatccc acacaaaatc atattgaatt gaattttaca 2280  
 tcttaagcta gtgtacaaca actgctatat ccaaagaaaa ctaacgtgga cgccttttag 2340  
 agttgagaaa aaggtttgaa aaaaatagca atacaaagac ttgtttcata tataaaatac 2400  
 agggagcaca ttgagctaat ataacataaa cactgcgaac caattccaat caaaaggtac 2460  
 acatgagagc attccccga gtactgcat ttcgcatca gagatcatat aataacatcc 2520  
 ttcttcgaac agtaaggctt tttggttcat cactttcttc ttttgatttc tctaggcaaa 2580

tgccctaaggt ggaccctgac aataccgctg caatgctact acagaaaaac ttgatccaaa 2640

gaaacaacat gctctatggg tatggatcag ggacaatagc atgtactttg ctagactcaa 2700

ctggacgagc caaatcacca ttagtagaga taaaacgtga ggat 2744

<210> 14

<211> 500

<212> PRT

<213> Saccharomyces cerevisiae

<220><221> misc\_feature

<222> (1)..(500)

<223> Saccharomyces cerevisiae cytosolic aldehyde dehydrogenase 6  
(ALD6) protein sequence

<400> 14

Met Thr Lys Leu His Phe Asp Thr Ala Glu Pro Val Lys Ile Thr Leu

1 5 10 15

Pro Asn Gly Leu Thr Tyr Glu Gln Pro Thr Gly Leu Phe Ile Asn Asn

20 25 30

Lys Phe Met Lys Ala Gln Asp Gly Lys Thr Tyr Pro Val Glu Asp Pro

35 40 45

Ser Thr Glu Asn Thr Val Cys Glu Val Ser Ser Ala Thr Thr Glu Asp

50 55 60

Val Glu Tyr Ala Ile Glu Cys Ala Asp Arg Ala Phe His Asp Thr Glu

65 70 75 80

Trp Ala Thr Gln Asp Pro Arg Glu Arg Gly Arg Leu Leu Ser Lys Leu

85 90 95

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

100 105 110

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

115 120 125

Ala Ile Asn Cys Leu Arg Asp Ala Ala Ala Tyr Ala Asp Lys Val Asn

130 135 140

Gly Arg Thr Ile Asn Thr Gly Asp Gly Tyr Met Asn Phe Thr Thr Leu



Val Lys Glu Glu Ile Phe Gly Pro Val Val Thr Val Ala Lys Phe Lys  
 405 410 415  
 Thr Leu Glu Glu Gly Val Glu Met Ala Asn Ser Ser Glu Phe Gly Leu  
 420 425 430  
 Gly Ser Gly Ile Glu Thr Glu Ser Leu Ser Thr Gly Leu Lys Val Ala  
 435 440 445  
 Lys Met Leu Lys Ala Gly Thr Val Trp Ile Asn Thr Tyr Asn Asp Phe  
 450 455 460

Asp Ser Arg Val Pro Phe Gly Gly Val Lys Gln Ser Gly Tyr Gly Arg  
 465 470 475 480  
 Glu Met Gly Glu Glu Val Tyr His Ala Tyr Thr Glu Val Lys Ala Val  
 485 490 495  
 Arg Ile Lys Leu  
 500

<210> 15

<211> 2728

<212> DNA

<213> Saccharomyces cerevisiae

<220><221> misc\_feature

<222> (1)..(2728)

<223> Saccharomyces cerevisiae ACS1 nucleotide sequence

<400> 15

acctcccgcg acctccaaaa tcgaactacc ttcaaatgt cgcctctgc cgtacaatca 60  
 tcaaaactag aagaacagtc aagtgaatt gacaagtga aagcaaaaat gtcccagtct 120  
 gcctccactg cgcagcagaa gaaggaacat gagtatgaac atttgacctc ggtcaagatc 180  
 gtgccacaac ggcccatctc agatagactg cagcccgcaa ttgctacca ctattctcca 240  
 cacttggacg ggttgcagga ctatcagcgc ttgcacaagg agtctattga agacctgct 300  
 aagttcttcg gttctaaagc tacceaatit ttaaactggc ctaagccatt cgataaggtg 360  
 ttcatcccag actctaaaac gggtaggccc tccttcaga acaatgcatg gttcctcaac 420  
 ggccaattaa acgcctgtta caactgtgtt gacagacatg ccttgaagac ccctaacaag 480  
 aaagccatta ttttcgaagg tgacgagcct ggccaaggct attccattac ctacaaggaa 540  
 ctacttgaag aagtttgca agtggcacia gtgctgactt actctatggg cgttcgcaag 600

ggcgatactg ttgccgtgta catgcctatg gtcccagaag caatcataac cttgttggcc 660  
 atttcccgta tggcgccat tcaactccgta gtctttgccg ggttttcttc caactccttg 720  
 agagatcgta tcaacgatgg ggactctaaa gttgtcatca ctacagatga atccaacaga 780  
 ggtggtaaag tcattgagac taaaagaatt gttgatgacg cgctaagaga gaccccaggc 840  
 gtgagacacg tcttggttta tagaaagacc aacaatccat ctgttgcttt ccatgcccc 900  
  
 agagatttag attgggcaac agaaaagaag aaatacaaga cctactatcc atgcacaccc 960  
 gttgattctg aggatccatt attcttgttg tatacgtctg gttctactgg tgcccccaag 1020  
 ggtgttcaac attctaccgc aggttacttg ctgggagctt tgttgacat gcgctacact 1080  
 tttgacactc accaagaaga cgttttcttc acagctggag acattggctg gattacaggc 1140  
 cacacttatg tggtttatgg tcccttacta tatggttgtg ccactttggt ctttgaaggg 1200  
 actctcgcgt acccaaatta ctcccgttat tgggatatta ttgatgaaca caaagtcacc 1260  
 caattttatg ttgccccaac tgctttgcgt ttgttgaaaa gagctggtga ttctacatc 1320  
  
 gaaaatcatt ccttaaaatc tttgcgttgc ttgggttcgg tcggtgaacc aattgctgct 1380  
 gaagtttggg agtggctactc tgaaaaataa ggtaaaaatg aatcccat tgtagacacc 1440  
 tactggcaaa cagaatctgg ttcgcatctg gtcacccgc tggctggtgg tgtcacacca 1500  
 atgaaaccgg gttctgcctc attccccttc ttcggtattg atgcagttgt tcttgacct 1560  
 aacactggtg aagaactta taccagccac gcagagggtg tcttgccgt caaagctgca 1620  
 tggccatcat ttgcaagaac tatttggaaa aatcatgata ggtatctaga cacttattg 1680  
 aacccttacc ctggctacta tttcactggt gatggtgctg caaaggataa ggatggttat 1740  
  
 atctggattt tgggtcgtgt agacgatgtg gtgaacgtct ctggtcaccg tctgtctacc 1800  
 gctgaaattg aggtgctat tatcgaagat ccaattgtgg ccgagtgtgc tgttgcgga 1860  
 ttcaacgatg acttgactgg tcaagcagtt gctgcatttg tgggttgaa aaacaaatct 1920  
 aattggtcca ccgcaacaga tgatgaatta caagatatca agaagcattt ggtcttact 1980  
 gttagaaaag acatcgggcc atttgccga ccaaaatga tcattttagt ggatgacttg 2040  
 cccaagacaa gatctggcaa aattatgaga cgtattttaa gaaaaatcct agcaggagaa 2100  
 agtgaccaac taggcgacgt ttctacattg tcaaacctg gcattgtag acatctaatt 2160  
  
 gattcggtea agttgtaatg atgatttctt tctttttat attgacgact ttttttttt 2220  
 cgtgtgtttt tgttctctta taaccgagct gcttacttat tattattca ccttctctt 2280  
 ttatttatac ttataattat ttattcttta catactgtta caagaaactc ttttctacat 2340  
 taattgcata aagtgtcaat cagcacatcc tctatatgc tatcaacaac aaatttgaca 2400  
 aacctgccta tatcttcagg aacaactgcc gcatcgctac caccactact tgtgaagtc 2460

ctggagttta atatgactg aaatttacct agccgtttta cacaagacca taatccatcc 2520  
 atgctatcgc agtatatgat tttgtgttcg tttttcgtct tgcgaaaggc atcctcaatg 2580

gcttgtttca ttgateccatc agtgtggctc gtaggtacca gcaaaaccac ttcacacagc 2640  
 gcgtactcct ccactttat gggcagtcct tgtatcgact tgctcattat aatacatttg 2700  
 ctctatcccc gcgtgcttgg cgggcctg 2728

<210> 16

<211> 713

<212> PRT

<213> *Saccharomyces cerevisiae*

<220><221> misc\_feature

<222> (1)..(713)

<223> *Saccharomyces cerevisiae* ACS1 protein sequence

<400> 16

Met Ser Pro Ser Ala Val Gln Ser Ser Lys Leu Glu Glu Gln Ser Ser  
 1                    5                    10                    15

Glu Ile Asp Lys Leu Lys Ala Lys Met Ser Gln Ser Ala Ser Thr Ala  
                   20                    25                    30

Gln Gln Lys Lys Glu His Glu Tyr Glu His Leu Thr Ser Val Lys Ile  
                   35                    40                    45

Val Pro Gln Arg Pro Ile Ser Asp Arg Leu Gln Pro Ala Ile Ala Thr  
                   50                    55                    60

His Tyr Ser Pro His Leu Asp Gly Leu Gln Asp Tyr Gln Arg Leu His  
 65                    70                    75                    80

Lys Glu Ser Ile Glu Asp Pro Ala Lys Phe Phe Gly Ser Lys Ala Thr  
                   85                    90                    95

Gln Phe Leu Asn Trp Ser Lys Pro Phe Asp Lys Val Phe Ile Pro Asp  
                   100                    105                    110

Ser Lys Thr Gly Arg Pro Ser Phe Gln Asn Asn Ala Trp Phe Leu Asn  
                   115                    120                    125

Gly Gln Leu Asn Ala Cys Tyr Asn Cys Val Asp Arg His Ala Leu Lys  
                   130                    135                    140

Thr Pro Asn Lys Lys Ala Ile Ile Phe Glu Gly Asp Glu Pro Gly Gln  
 145                      150                      155                      160  
 Gly Tyr Ser Ile Thr Tyr Lys Glu Leu Leu Glu Glu Val Cys Gln Val  
                                  165                      170                      175  
 Ala Gln Val Leu Thr Tyr Ser Met Gly Val Arg Lys Gly Asp Thr Val  
                                  180                      185                      190  
 Ala Val Tyr Met Pro Met Val Pro Glu Ala Ile Ile Thr Leu Leu Ala  
                                  195                      200                      205  
  
 Ile Ser Arg Ile Gly Ala Ile His Ser Val Val Phe Ala Gly Phe Ser  
                                  210                      215                      220  
 Ser Asn Ser Leu Arg Asp Arg Ile Asn Asp Gly Asp Ser Lys Val Val  
 225                      230                      235                      240  
 Ile Thr Thr Asp Glu Ser Asn Arg Gly Gly Lys Val Ile Glu Thr Lys  
                                  245                      250                      255  
 Arg Ile Val Asp Asp Ala Leu Arg Glu Thr Pro Gly Val Arg His Val  
                                  260                      265                      270  
  
 Leu Val Tyr Arg Lys Thr Asn Asn Pro Ser Val Ala Phe His Ala Pro  
                                  275                      280                      285  
 Arg Asp Leu Asp Trp Ala Thr Glu Lys Lys Lys Tyr Lys Thr Tyr Tyr  
                                  290                      295                      300  
 Pro Cys Thr Pro Val Asp Ser Glu Asp Pro Leu Phe Leu Leu Tyr Thr  
 305                      310                      315                      320  
 Ser Gly Ser Thr Gly Ala Pro Lys Gly Val Gln His Ser Thr Ala Gly  
                                  325                      330                      335  
  
 Tyr Leu Leu Gly Ala Leu Leu Thr Met Arg Tyr Thr Phe Asp Thr His  
                                  340                      345                      350  
 Gln Glu Asp Val Phe Phe Thr Ala Gly Asp Ile Gly Trp Ile Thr Gly  
                                  355                      360                      365  
 His Thr Tyr Val Val Tyr Gly Pro Leu Leu Tyr Gly Cys Ala Thr Leu  
                                  370                      375                      380  
 Val Phe Glu Gly Thr Pro Ala Tyr Pro Asn Tyr Ser Arg Tyr Trp Asp



Ile Lys Lys His Leu Val Phe Thr Val Arg Lys Asp Ile Gly Pro Phe  
 645 650 655

Ala Ala Pro Lys Leu Ile Ile Leu Val Asp Asp Leu Pro Lys Thr Arg  
 660 665 670

Ser Gly Lys Ile Met Arg Arg Ile Leu Arg Lys Ile Leu Ala Gly Glu  
 675 680 685

Ser Asp Gln Leu Gly Asp Val Ser Thr Leu Ser Asn Pro Gly Ile Val  
 690 695 700

Arg His Leu Ile Asp Ser Val Lys Leu  
 705 710

<210> 17

<211> 2287

<212> DNA

<213> *Saccharomyces cerevisiae*

<220><221> misc\_feature

<222> (1)..(2287)

<223> *Saccharomyces cerevisiae* ACS2 nucleotide sequence

<400> 17

acctcccgcg acctccaaaa tcgaactacc ttcacaatga caatcaagga acataaagta 60  
 gtttatgaag ctcaaacgt aaaggctctt aaggctctc aacattttta caacagccaa 120  
 cccggcaagg gttacgttac tgatatgcaa cattatcaag aaatgtatca acaatctatc 180  
 aatgagccag aaaaattctt tgataagatg gctaaggaat acttgcatg ggatgctcca 240  
 tacaccaaag ttcaatctgg ttcatggaac aatggtgatg ttgcatggtt tttgaacggt 300  
  
 aaattgaatg catcatacaa ttgtgttgac agacatgect ttgctaatec cgacaagcca 360  
 gctttgatct atgaagetga tgacgaatcc gacaacaaaa tcatcacatt tggatgaatta 420  
 ctcaaaaag tttcccaaat cgctggtgtc ttaaaaagct ggggcgtaa gaaaggtgac 480  
 acagtggcta tctatttgcc aatgattcca gaagcggtea ttgctatgtt ggctgtggct 540  
 cgtattggtg ctattcactc tgttctctt gctgggttct ccgctggttc gttgaaagat 600  
 cgtgtcgttg acgctaattc taaagtggc atcacttgtg atgaaggtaa aagaggtggt 660  
 aagaccatca aactaaaaa aattgttgac gaaggttga acggagtcga tttggtttcc 720  
  
 cgtatcttgg ttttcaaag aactggtact gaaggtattc caatgaagc cggtagagat 780

tactgggtggc atgaggaggc cgctaagcag agaacttacc tacctcctgt ttcattgtgac 840  
 gctgaagatc ctctatTTTT attatacact tccggttcca ctggttctcc aaagggtgtc 900  
 gttcacacta caggtgggta tttattaggt gccgctttaa caactagata cgTTTTgat 960  
 attcaccag aagatgttct cttcactgcc ggtgacgtcg gctggatcac gggtcacacc 1020  
 tatgtcttat atggtccatt aaccttgggt accgcctcaa taattttcga atccactcct 1080  
 gcctaccag attatggtag atattggaga attatccaac gtcacaagc tacccatttc 1140

tatgtggctc caactgcttt aagattaatc aaacgtgtag gtgaagccga aattgccaaa 1200  
 tatgacactt cctcattacg tgtcttgggt tccgtcgggtg aaccaatctc tccagactta 1260  
 tgggaatggt atcatgaaaa agtgggtaac aaaaactgtg tcatttgtga cactatgtgg 1320  
 caaacagagt ctggttctca ttttaattgct cctttggcag gtgctgtccc aacaaaacct 1380  
 ggttctgcta ccgtgccatt ctttggatt aacgcttga tcattgacct tgttacaggt 1440  
 gtggaattag aagtaatga tgtcgaaggt gtccttgccg ttaaatcacc atggccatca 1500  
 atggctagat ctgtttggaa ccaccacgac cgttacatgg atacttactt gaaaccttat 1560

cctggtcact atttcacagg tgatgggtgct ggtagagatc atgatggta ctactggatc 1620  
 aggggtagag ttgacgacgt tgtaaatgtt tccggtcata gattatccac atcagaaatt 1680  
 gaagcatcta tctcaaatca cgaaaacgtc tcggaagctg ctgttgtcgg tattccagat 1740  
 gaattgacct gtcaaaccgt cgttgcataat gtttccctaa aagatggta tctacaaaac 1800  
 aacgctactg aaggtgatgc agaacacatc acaccagata atttacgtag agaattgatc 1860  
 ttacaagtta ggggtgagat tggctctttc gcctcaccaa aaaccattat tctagttaga 1920  
 gatctaccaa gaacaaggctc aggaaagatt atgagaagag ttctaagaaa ggttgcttct 1980

aacgaagccg aacagctagg tgacctaaact actttgcca acccagaagt tgtacctgcc 2040  
 atcatttctg ctgtagagaa ccaatttttc tctcaaaaaa agaaataact taaatgagaa 2100  
 aaatttcgta atgagataaa atttcgctcc ttttctgttt tctattttct attttccaa 2160  
 cttttgctct attcagttat aaattactat ttatccatca gttaaaaaac aagatctttt 2220  
 actggtcage taggaaagcg aaaatacaaa gactttatgc actatccccg cgtgcttggc 2280  
 cggccgt 2287

<210> 18

<211> 683

<212> PRT

<213> *Saccharomyces cerevisiae*

<220><221> misc\_feature

<222> (1)..(683)

<223> Saccharomyces cerevisiae ACS2 protein sequence

<400> 18

Met Thr Ile Lys Glu His Lys Val Val Tyr Glu Ala His Asn Val Lys

1                    5                    10                    15

Ala Leu Lys Ala Pro Gln His Phe Tyr Asn Ser Gln Pro Gly Lys Gly

                  20                    25                    30

Tyr Val Thr Asp Met Gln His Tyr Gln Glu Met Tyr Gln Gln Ser Ile

                  35                    40                    45

Asn Glu Pro Glu Lys Phe Phe Asp Lys Met Ala Lys Glu Tyr Leu His

                  50                    55                    60

Trp Asp Ala Pro Tyr Thr Lys Val Gln Ser Gly Ser Leu Asn Asn Gly

65                    70                    75                    80

Asp Val Ala Trp Phe Leu Asn Gly Lys Leu Asn Ala Ser Tyr Asn Cys

                  85                    90                    95

Val Asp Arg His Ala Phe Ala Asn Pro Asp Lys Pro Ala Leu Ile Tyr

                  100                    105                    110

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

                  115                    120                    125

Leu Arg Lys Val Ser Gln Ile Ala Gly Val Leu Lys Ser Trp Gly Val

                  130                    135                    140

Lys Lys Gly Asp Thr Val Ala Ile Tyr Leu Pro Met Ile Pro Glu Ala

145                    150                    155                    160

Val Ile Ala Met Leu Ala Val Ala Arg Ile Gly Ala Ile His Ser Val

                  165                    170                    175

Val Phe Ala Gly Phe Ser Ala Gly Ser Leu Lys Asp Arg Val Val Asp

                  180                    185                    190

Ala Asn Ser Lys Val Val Ile Thr Cys Asp Glu Gly Lys Arg Gly Gly

                  195                    200                    205

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

                  210                    215                    220

Asp Leu Val Ser Arg Ile Leu Val Phe Gln Arg Thr Gly Thr Glu Gly



Ala Val Lys Ser Pro Trp Pro Ser Met Ala Arg Ser Val Trp Asn His  
 485 490 495

His Asp Arg Tyr Met Asp Thr Tyr Leu Lys Pro Tyr Pro Gly His Tyr  
 500 505 510

Phe Thr Gly Asp Gly Ala Gly Arg Asp His Asp Gly Tyr Tyr Trp Ile  
 515 520 525

Arg Gly Arg Val Asp Asp Val Val Asn Val Ser Gly His Arg Leu Ser  
 530 535 540

Thr Ser Glu Ile Glu Ala Ser Ile Ser Asn His Glu Asn Val Ser Glu  
 545 550 555 560

Ala Ala Val Val Gly Ile Pro Asp Glu Leu Thr Gly Gln Thr Val Val  
 565 570 575

Ala Tyr Val Ser Leu Lys Asp Gly Tyr Leu Gln Asn Asn Ala Thr Glu  
 580 585 590

Gly Asp Ala Glu His Ile Thr Pro Asp Asn Leu Arg Arg Glu Leu Ile  
 595 600 605

Leu Gln Val Arg Gly Glu Ile Gly Pro Phe Ala Ser Pro Lys Thr Ile  
 610 615 620

Ile Leu Val Arg Asp Leu Pro Arg Thr Arg Ser Gly Lys Ile Met Arg  
 625 630 635 640

Arg Val Leu Arg Lys Val Ala Ser Asn Glu Ala Glu Gln Leu Gly Asp  
 645 650 655

Leu Thr Thr Leu Ala Asn Pro Glu Val Val Pro Ala Ile Ile Ser Ala  
 660 665 670

Val Glu Asn Gln Phe Phe Ser Gln Lys Lys Lys  
 675 680

<210> 19

<211> 2137

<212> DNA

<213> Streptomyces sp.

<220><221> misc\_feature

<222> (1)..(2137)

<223> Streptomyces sp. CL190 nphT7 gene sequence

<400> 19

```

cctgcaggcc gtcgagggcg cctggaagga ctacgcggag caggacggcc ggtcgctgga      60
ggagttcgcg gcgttcgtct accaccagcc gttcacgaag atggcctaca aggcgcaccg      120
ccacctgctg aacttcaacg gctacgacac cgacaaggac gccatcgagg gcgacctcgg      180
ccagacgacg gcgtacaaca acgtcatcgg caacagctac accgcgtcgg tgtacctggg      240
cctggccgcc ctgctcgacc aggcgggacga cctgacgggc cgttccatcg gcttcctgag      300

ctacggctcg ggcagcgtcg ccgagttctt ctcgggcacc gtcgtcgccg ggtaccgca      360
gcgtctgcgc accgaggcga accaggaggc gatcgcccgg cgcaagagcg tcgactacgc      420
cacctaccgc gagctgcacg agtacacgct cccgtccgac ggcgcgacc acgccacccc      480
ggtgcagacc accggcccct tccggctggc cgggatcaac gaccacaagc gcacttacga      540
ggcgcgctag cgacaccctt cggcaacggg gtgcgccact gttcggcgca ccccgtgccg      600
ggctttcgca cagctattca cgaccatttg aggggcgggc agccgatga ccgacgtccg      660
attccgattt atcggtagcg gtgcctacgt accggaacgg atcgtctcca acgatgaagt      720

cggcgcgccc gccggggtgg acgacgactg gatcaccgcg aagaccggtg tccggcagcg      780
tcgctgggcc gccgacgacc aggeccactc ggacctggcc acggcccgcg ggccggcagc      840
gctgaaagcg gcgggcatca cgcccagca gctgaccgtg atcgcggtcg ccacctccac      900
gccggaccgg ccgacgcccg ccacggcggc ctatgtccag caccacctcg gtgcgaccgg      960
cactgcggcg ttcgacgca acgcggtctg ctccggcacc gtgttcgcg tgcctcggg      1020
ggcgggcacc ctctgtacc ggggcggtta cgcgctggtc atcggcgcg accgtactc      1080
gcgcatctc aaccggcg accgcaagac ggtcgtgctg ttcggggacg gcgcccgcg      1140

aatggtctc gggccgacct cgaccggcac gggccccatc gtccggcgcg tcgccctgca      1200
caccttcgce gccctcaccg acctgatccg tgtgcccgcg ggccgagcc gccagccgt      1260
ggacacggat ggcctcgacg cgggactgca gtacttcgcg atggacgggc gtgaggtgcg      1320
ccgcttcgtc acggagcacc tgccgagct gatcaagggc ttcctgcacg aggccggggt      1380
cgacgccgc gacatcagcc acttcgtgcc gcatcaggcc aacggtgtca tgctcgacga      1440
ggtcttcgce gagctgcac tgcgcgggc gaccatgcac cggacggtcg agacctacgg      1500
caacacggga gcggcctcca tcccgatcac catggacgcg gccgtgcgcg ccggttcctt      1560

ccggccgggc gagctggtcc tgctggccgg gttcggcggc ggcatggcgg cgagcttcgc      1620
cctgatcgag tggtagtcgc ccgtaccacc acagcgtcc ggcgccacct gttccctgcg      1680

```

ccgggccgcc ctctggggcct ttaggccccca caccgccccca gccgacggat tcagtcgcgg 1740  
 cagtacctca gatgtccgct gcgacggcgt cccggagagc ccgggcgaga tcgctgggccc 1800  
 cctttctgctc gtccccggcc cctccccgca gcaccaccgc cgggcgacgg ccgcccctct 1860  
 ccgcgatacg ccgggcgagg tcgcaggcga gcacgccgga cccggagaag cccccagca 1920  
 ccagcgaccg gccgactccg tgcgcggcca gggcaggctg cgcgccgtcg acgtcgggtga 1980

gcagcaccag gagctcctgc ggccccggcgt agaggtcggc cagccggtcg tagcaggctc 2040  
 cgggcgcgcc cggcggcggg atcagacaga tcgtgccgcg ccgctcgtgc ctgcccgcc 2100  
 gcagcgtgac cagcggaatg tccccccag ctccgga 2137

<210> 20

<211> 325

<212> PRT

<213> Streptomyces sp.

<220><221> misc\_feature

<222> (1)..(325)

<223> Streptomyces sp. CL190 acetyl-CoA:malonyl-CoA acyltransferase  
 protein sequence

<400> 20

Arg Phe Arg Ile Ile Gly Thr Gly Ala Tyr Val Pro Glu Arg Ile Val

1 5 10 15

Ser Asn Asp Glu Val Gly Ala Pro Ala Gly Val Asp Asp Asp Trp Ile

20 25 30

Thr Arg Lys Thr Gly Ile Arg Gln Arg Arg Trp Ala Ala Asp Asp Gln

35 40 45

Ala Thr Ser Asp Leu Ala Thr Ala Ala Gly Arg Ala Ala Leu Lys Ala

50 55 60

Ala Gly Ile Thr Pro Glu Gln Leu Thr Val Ile Ala Val Ala Thr Ser

65 70 75 80

Thr Pro Asp Arg Pro Gln Pro Pro Thr Ala Ala Tyr Val Gln His His

85 90 95

Leu Gly Ala Thr Gly Thr Ala Ala Phe Asp Val Asn Ala Val Cys Ser

100 105 110

Gly Thr Val Phe Ala Leu Ser Ser Val Ala Gly Thr Leu Val Tyr Arg



<220><221> misc\_feature

<222> (1)..(1287)

<223> Pseudomonas mevalonii HMG-CoA reductase (mvaA) gene sequence

<400> 21

```

atgagcctcg attccccgct gcccgctttc cgtaacctgt cccctgccgc ggcctggac      60
cacatcggcc agttgctcgg cctgagccac gacgatgtca gcctgctggc caacccgggt      120
gccctgccga tggacatcgc caacggcatg atcgaaaacg tcatcggcac cttcgagctg      180
ccctatgccg tggccagcaa cttccagatc aatggccgtg atgtgctggt gccgctgggt      240
gtggaagagc cctcgatcgt cgccgctgct tcgtacatgg ccaagctggc cegtgccaac      300

ggcggcttca ccacctccag cagcgccttc ctgatgcatg cccaggtaga gatcgtcggc      360
atacaggacc cgtcaatgc acgctgagc ctgctgcgcc gcaaagacga aatcattgaa      420
ctggccaacc gcaaggacca gttgctcaac agcctcggcg gcggctgccg cgacatcgaa      480
gtgcacacct tcgccgatac cccgctggc cggatgctgg tggcgcacct gatcgtcgt      540
gtacgcgatg ccatgggcgc caacaccgtc aataccatgg ccgaggccgt tgcgccctg      600
atggaagcca tcaccggggg ccaggtacgc ctgcgattc tgtccaacct ggccgacctg      660
cgctggcca gggcccaggt gcggattact ccgcagcaac tgaaaacggc cgaattcagt      720

ggcggaggcag tgatcgaagg catcctcgac gcctacgctt tcgctgcggt cgacccttac      780
cgcgcggcca cccacaaca gggcatcatg aatggcatcg acccactgat cgtcgccact      840
ggcaacgact ggcgtgcagt ggaagccggc gcccatgcgt atgcctgccg cagtggtcac      900
tacggctcgc tgaccacctg ggaaaaggac aacaacggcc atttggctcg caccctggaa      960
atgccgatgc ccgtaggcct ggtcggcggc gccacaaaa cccatccgct ggcgcaactg      1020
tcgctgcgca tctcggcgt gaaaacagcc caggcgtcgc ctgagattgc cgtggccgta      1080
ggcctggcgc aaaacctcgg ggccatgcgc gcctggcca ccgaaggcat ccagcgcggc      1140

cacatggccc tgcatgcgcg caatattgcc gtggtggcgg gcgcccagg cgatgaggtg      1200
gactgggttg cccggcagtt ggtggaatac cagcagctgc gcgccgaccg cgccgtagca      1260
ctgctgaaac aaaagcgcgg ccaatga

```

<210> 22

<211> 428

<212> PRT

<213> Pseudomonas mevalonii

<220><221> misc\_feature

<222> (1)..(428)

<223> Pseudomonas mevalonii hydroxymethylglutaryl -CoA reductase  
protein sequence

<400> 22

Met Ser Leu Asp Ser Arg Leu Pro Ala Phe Arg Asn Leu Ser Pro Ala

1                    5                    10                    15  
Ala Arg Leu Asp His Ile Gly Gln Leu Leu Gly Leu Ser His Asp Asp

                  20                    25                    30  
Val Ser Leu Leu Ala Asn Ala Gly Ala Leu Pro Met Asp Ile Ala Asn

                  35                    40                    45  
Gly Met Ile Glu Asn Val Ile Gly Thr Phe Glu Leu Pro Tyr Ala Val

                  50                    55                    60  
Ala Ser Asn Phe Gln Ile Asn Gly Arg Asp Val Leu Val Pro Leu Val

65                    70                    75                    80  
Val Glu Glu Pro Ser Ile Val Ala Ala Ala Ser Tyr Met Ala Lys Leu

                  85                    90                    95  
Ala Arg Ala Asn Gly Gly Phe Thr Thr Ser Ser Ser Ala Pro Leu Met

                  100                    105                    110  
His Ala Gln Val Gln Ile Val Gly Ile Gln Asp Pro Leu Asn Ala Arg

                  115                    120                    125  
Leu Ser Leu Leu Arg Arg Lys Asp Glu Ile Ile Glu Leu Ala Asn Arg

                  130                    135                    140  
Lys Asp Gln Leu Leu Asn Ser Leu Gly Gly Gly Cys Arg Asp Ile Glu

145                    150                    155                    160  
Val His Thr Phe Ala Asp Thr Pro Arg Gly Pro Met Leu Val Ala His

                  165                    170                    175  
Leu Ile Val Asp Val Arg Asp Ala Met Gly Ala Asn Thr Val Asn Thr

                  180                    185                    190  
Met Ala Glu Ala Val Ala Pro Leu Met Glu Ala Ile Thr Gly Gly Gln

                  195                    200                    205  
Val Arg Leu Arg Ile Leu Ser Asn Leu Ala Asp Leu Arg Leu Ala Arg

210                    215                    220

Ala Gln Val Arg Ile Thr Pro Gln Gln Leu Glu Thr Ala Glu Phe Ser  
 225                      230                      235                      240  
 Gly Glu Ala Val Ile Glu Gly Ile Leu Asp Ala Tyr Ala Phe Ala Ala  
                          245                      250                      255  
 Val Asp Pro Tyr Arg Ala Ala Thr His Asn Lys Gly Ile Met Asn Gly  
                          260                      265                      270  
 Ile Asp Pro Leu Ile Val Ala Thr Gly Asn Asp Trp Arg Ala Val Glu  
                          275                      280                      285  
 Ala Gly Ala His Ala Tyr Ala Cys Arg Ser Gly His Tyr Gly Ser Leu  
                          290                      295                      300  
 Thr Thr Trp Glu Lys Asp Asn Asn Gly His Leu Val Gly Thr Leu Glu  
 305                      310                      315                      320  
 Met Pro Met Pro Val Gly Leu Val Gly Gly Ala Thr Lys Thr His Pro  
                          325                      330                      335  
 Leu Ala Gln Leu Ser Leu Arg Ile Leu Gly Val Lys Thr Ala Gln Ala  
                          340                      345                      350  
 Leu Ala Glu Ile Ala Val Ala Val Gly Leu Ala Gln Asn Leu Gly Ala  
                          355                      360                      365  
 Met Arg Ala Leu Ala Thr Glu Gly Ile Gln Arg Gly His Met Ala Leu  
                          370                      375                      380  
 His Ala Arg Asn Ile Ala Val Val Ala Gly Ala Arg Gly Asp Glu Val  
                          385                      390                      395                      400  
 Asp Trp Val Ala Arg Gln Leu Val Glu Tyr His Asp Val Arg Ala Asp  
                          405                      410                      415  
 Arg Ala Val Ala Leu Leu Lys Gln Lys Arg Gly Gln  
                          420                      425

- <210> 23
- <211> 1302
- <212> DNA
- <213> *Silicibacter pomeroyi*
- <220><221> misc\_feature
- <222> (1)..(1302)

<223> *Silicibacter pomeroyi* hydroxymethylglutaryl-CoA reductase gene  
sequence

<400>

23

atgacaggca agacgggtca catcgatggt ttgaactcgc gcattgaaaa gatgcgagat 60  
ctcgaccccc cacaacggct ggtgcgcggt gccgaggcgg cgggcctcga gcccaggcgg 120  
atcagcgcgc tggcgggtaa cggcgccctg ccctctcgc tggccaacgg gatgatcgag 180  
aacgtcatcg gcaaattcga actgccgctg ggcgtggcca cgaatttcac tgtgaacggc 240  
cgcgactatc tgatcccgat ggcggtcgaa gagccctcgg tggcggcggc cgcgtcctat 300  
atggcgcgta tcgcgcgcga gaatggcggg ttaccgcgc atggcaccgc gcccttgatg 360  
cgcgcccaga tccaggtggt cgggttgggt gatcccagg gcgcccggca gcgtctcctc 420

gcccacaagg ccgcttcat ggaggcggcg gacgctgtcg atccggtgct tgtcgggctg 480  
ggtggcggct gcccgatata cgagttcac gtgttccggg ataccgggt gggcgcgatg 540  
gtcgtcctgc acctgatcgt cgatgtgcgc gacgcgatgg gggccaatac ggtcaacacg 600  
atggccgaac ggetggcccc cgaggtcgag cggattgccg gtggcaccgt gcggtcgcgc 660  
atcctgtcga acctcgccga cctgcgattg gtccggggcg gggggaaact ggccccggaa 720  
aactgacaa cgcagggcta tgacggcgcc gacgtggcgc ggggcatggt cgaggcctgc 780  
gcgcttgcca tegtcgacce ctatcgcgcg gcgaccata acaaggggat catgaacggc 840

atcgaccggg tcgtcgtcgc caccggcaat gactggcgcg cgatcgaggc gggtgcccat 900  
gcctatgccg cccgcacggg tcattatacc tcgctgacce gctgggaact ggccaatgac 960  
gggcggcttg tgggcacgat cgaactgcc ctggcgcttg gccttgctcg cggcgcgacc 1020  
aagacgcacc cgaccgcacg ggcggcgctg gcctgatgc aggtagagac tgcaaccgaa 1080  
ctggcccagg tcaccgcgc cgtgggtctg gcgagaaca tggccgcat cgcgcgctg 1140  
gcgaccgaag gcatccagcg cggtcacatg accttcatg cgcgcaacat cgcgatcatg 1200  
gccggcgcaa caggcgcca tatcgaccgc gtcaccggg tcattgtcga agcgggcgac 1260

gtcagcgtgg cccgtgcaaa acaggtgctg gaaaacacct ga 1302

<210> 24

<211> 433

<212> PRT

<213> *Silicibacter pomeroyi*

<220><221> misc\_feature

<222> (1)..(433)

<223> *Silicibacter pomeroyi* hydroxymethylglutaryl-CoA reductase protein  
sequence

<400> 24

Met Thr Gly Lys Thr Gly His Ile Asp Gly Leu Asn Ser Arg Ile Glu  
1                   5                   10                   15

Lys Met Arg Asp Leu Asp Pro Ala Gln Arg Leu Val Arg Val Ala Glu  
                  20                   25                   30

Ala Ala Gly Leu Glu Pro Glu Ala Ile Ser Ala Leu Ala Gly Asn Gly  
                  35                   40                   45

Ala Leu Pro Leu Ser Leu Ala Asn Gly Met Ile Glu Asn Val Ile Gly  
                  50                   55                   60

Lys Phe Glu Leu Pro Leu Gly Val Ala Thr Asn Phe Thr Val Asn Gly  
65                   70                   75                   80

Arg Asp Tyr Leu Ile Pro Met Ala Val Glu Glu Pro Ser Val Val Ala  
                  85                   90                   95

Ala Ala Ser Tyr Met Ala Arg Ile Ala Arg Glu Asn Gly Gly Phe Thr  
                  100                   105                   110

Ala His Gly Thr Ala Pro Leu Met Arg Ala Gln Ile Gln Val Val Gly  
                  115                   120                   125

Leu Gly Asp Pro Glu Gly Ala Arg Gln Arg Leu Leu Ala His Lys Ala  
                  130                   135                   140

Ala Phe Met Glu Ala Ala Asp Ala Val Asp Pro Val Leu Val Gly Leu  
145                   150                   155                   160

Gly Gly Gly Cys Arg Asp Ile Glu Val His Val Phe Arg Asp Thr Pro  
                  165                   170                   175

Val Gly Ala Met Val Val Leu His Leu Ile Val Asp Val Arg Asp Ala  
                  180                   185                   190

Met Gly Ala Asn Thr Val Asn Thr Met Ala Glu Arg Leu Ala Pro Glu  
                  195                   200                   205

Val Glu Arg Ile Ala Gly Gly Thr Val Arg Leu Arg Ile Leu Ser Asn  
                  210                   215                   220

Leu Ala Asp Leu Arg Leu Val Arg Ala Arg Val Glu Leu Ala Pro Glu  
 225                      230                      235                      240  
 Thr Leu Thr Thr Gln Gly Tyr Asp Gly Ala Asp Val Ala Arg Gly Met  
                                  245                      250                      255  
 Val Glu Ala Cys Ala Leu Ala Ile Val Asp Pro Tyr Arg Ala Ala Thr  
                                  260                      265                      270  
 His Asn Lys Gly Ile Met Asn Gly Ile Asp Pro Val Val Val Ala Thr  
                                  275                      280                      285

Gly Asn Asp Trp Arg Ala Ile Glu Ala Gly Ala His Ala Tyr Ala Ala  
                                  290                      295                      300  
 Arg Thr Gly His Tyr Thr Ser Leu Thr Arg Trp Glu Leu Ala Asn Asp  
 305                      310                      315                      320  
 Gly Arg Leu Val Gly Thr Ile Glu Leu Pro Leu Ala Leu Gly Leu Val  
                                  325                      330                      335  
 Gly Gly Ala Thr Lys Thr His Pro Thr Ala Arg Ala Ala Leu Ala Leu  
                                  340                      345                      350

Met Gln Val Glu Thr Ala Thr Glu Leu Ala Gln Val Thr Ala Ala Val  
                                  355                      360                      365  
 Gly Leu Ala Gln Asn Met Ala Ala Ile Arg Ala Leu Ala Thr Glu Gly  
                                  370                      375                      380  
 Ile Gln Arg Gly His Met Thr Leu His Ala Arg Asn Ile Ala Ile Met  
 385                      390                      395                      400  
 Ala Gly Ala Thr Gly Ala Asp Ile Asp Arg Val Thr Arg Val Ile Val  
                                  405                      410                      415

Glu Ala Gly Asp Val Ser Val Ala Arg Ala Lys Gln Val Leu Glu Asn  
                                  420                      425                      430

Thr

<210> 25

<211> 1290

<212> DNA

<213> Delftia acidovorans

<220><221> misc\_feature  
 <222> (1)..(1290)  
 <223> Delftia acidovorans hydroxymethylglutaryl-CoA reductase  
 nucleotide sequence  
 <400> 25  
 atggttgccg attcgcgact gccaatttc cgcgccctca caccggccca gcgcccggat 60  
 ttcttgccg atgcctgcgg cctgtccgat gccgagcgcg ccctgctcgc tgccccggg 120  
  
 gccctgcccc tggcgtggc cgacggcatg atcgagaacg tgttcggcag cttcgagctg 180  
 ccgctgggcg tggccggcaa cttccgcgtc aacggccgcg acgtgctggt gcccatggcg 240  
 gtggaggagc cctcgggtgt ggcccggcc tcgtacatgg ccaagctggc gcgagaggac 300  
 gggggctttc agacctcaag cacgtgccc ctgatgcgcg cccaggtcca ggtgctgggc 360  
 gtgaccgatc cacacggcgc gcgcctggcc gtgctgcagg cgcgtgcgca gatcatcgag 420  
 cgcgccaaca gccgcgacaa ggtgctgatc ggctgggcg gcggtgcaa ggacatcgag 480  
 gtccatgtct tccccgacac gccgcgccc cccatgctgg tggccacct gatcgtggac 540  
  
 gtgcgcgacg ccatgggcg caacaccgtc aacaccatgg ccgaatcggg ggcgccctg 600  
 gtcgagaaga tcacgggagg cagcgtgcgg ctgcgcatcc tgtccaacct ggccgacctg 660  
 cggctggccc gcgcccgcgt gcggetcag ccgagaccc tggccagca ggategcagc 720  
 ggagaggaga tcacgaagg cgtgctggac gcctatacct tcgaggccat cgaccctac 780  
 cgcgcggcca cgcacaaca gggaatcatg aacggcatcg acccgtcat cgtggccag 840  
 ggcaacgact ggcgcgcggt cgagcccggc gcccatgcct atgccagccg cagcggcagc 900  
 tacacctcgc tgacgcgctg ggaaaaggat gccggcgcg ccctggctgg cagcatcgag 960  
  
 ctgccatgc cggctgggct tgtcggcggc gccaccaaga cccatccgct ggcacgctg 1020  
 gcgctgaaga tcatggacct gcagtccgc cagcagctgg gcgagatcgc gccgcccgtg 1080  
 ggctggcgc agaacctggg gcacctgcgc gcctggcca ccgaaggcat tcagcgggc 1140  
 cacatggccc tgcacgccc caacatgcc ctggtggccg gcgccacggg cgacgagtc 1200  
 gatgccgtgg cgcgccagct ggcccggag cagcagctgc gcaccgaccg cgcgctggaa 1260  
 gtgctggccc cgtgcgcgc cagggcctga 1290  
 <210> 26  
 <211> 429  
 <212> PRT  
 <213> Delftia acidovorans

<220><221> misc\_feature

<222> (1)..(429)

<223> Delftia acidovorans hydroxymethylglutaryl-CoA reductase protein  
sequence

<400> 26

Met Val Ala Asp Ser Arg Leu Pro Asn Phe Arg Ala Leu Thr Pro Ala

1                    5                    10                    15

Gln Arg Arg Asp Phe Leu Ala Asp Ala Cys Gly Leu Ser Asp Ala Glu

                  20                    25                    30

Arg Ala Leu Leu Ala Ala Pro Gly Ala Leu Pro Leu Ala Leu Ala Asp

                  35                    40                    45

Gly Met Ile Glu Asn Val Phe Gly Ser Phe Glu Leu Pro Leu Gly Val

                  50                    55                    60

Ala Gly Asn Phe Arg Val Asn Gly Arg Asp Val Leu Val Pro Met Ala

65                    70                    75                    80

Val Glu Glu Pro Ser Val Val Ala Ala Ala Ser Tyr Met Ala Lys Leu

                  85                    90                    95

Ala Arg Glu Asp Gly Gly Phe Gln Thr Ser Ser Thr Leu Pro Leu Met

                  100                    105                    110

Arg Ala Gln Val Gln Val Leu Gly Val Thr Asp Pro His Gly Ala Arg

                  115                    120                    125

Leu Ala Val Leu Gln Ala Arg Ala Gln Ile Ile Glu Arg Ala Asn Ser

                  130                    135                    140

Arg Asp Lys Val Leu Ile Gly Leu Gly Gly Gly Cys Lys Asp Ile Glu

145                    150                    155                    160

Val His Val Phe Pro Asp Thr Pro Arg Gly Pro Met Leu Val Val His

                  165                    170                    175

Leu Ile Val Asp Val Arg Asp Ala Met Gly Ala Asn Thr Val Asn Thr

                  180                    185                    190

Met Ala Glu Ser Val Ala Pro Leu Val Glu Lys Ile Thr Gly Gly Ser

                  195                    200                    205

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



<220><223> Synthetic: i2235 integration construct

<400> 27

gacggcacgg ccacgcgttt aaaccgcctc gatatttctt gtgagaagtt taaatccact 60  
aaggtttttc atgtttgctg cagatgtggt tttccattca tcctgaaata tgcactgcta 120  
ttccgcattc cattccccta gtctttttta gttctttccg ttcgaccttc atcgaaaaat 180  
gacaaaaagc gttaggaaca acaaccaatt gcaacaagc agtgaacaa aaccatcaag 240  
gcccgaaaat acaagtgtgt actaatacag taagtaggtc aaatcgcaa tgaccaaaga 300  
tgccgtgaat ctagatgctt acaccgtgag cttcatgctt ttctataccg agtatcaagg 360

accaaccgaa gagttaaggg attacaaatt cgaagatact atttactttc gtggcaagga 420  
actgaagagg gaaaagtctg cgacgccttc cagtagcgat aacacaacta gtaatacctt 480  
cagtaatggc gccatcctct cgggaaacac aataactggc aagatagttt cagtgaataa 540  
ttacgaaaga gagggcactg atcgcaacga attggcgcgga ttgcaagaat tgatctccct 600  
catcgatgtc ataatcagt aaatataagc tcacacgagg ccagggggag cccgttgagc 660  
cattagtatc aatttgccta cctgtattcc ttactatcc tctttttct ctttttgat 720  
aatgtatgt agattgcgta tatagtttcg tctaccctat gaacatattc cttttgtaa 780

tttcgtgctg tttctattat gaatttcatt tataaagttt atgtacaaat atcataaaaa 840  
aagagaatct ttttaagcaa ggattttctt aacttcttcg gcgacagcat caccgacttc 900  
ggtggtactg ttggaaccac ctaaatcacc agttctgata cctgcatcca aaacctttt 960  
aactgcatct tcaatggcct taccttcttc aggcaagttc aatgacaatt tcaacatcat 1020  
tgcagcagac aagatagttg cgatagggtc aaccttattc tttggcaaat ctggagcaga 1080  
accgtggcat ggttcgtaca aaccaaagtc ggtgttcttg tctggcaaag aggccaagga 1140  
cgcagatggc acaaaacca aggaacctgg gataacggag gcttcatcgg agatgatatc 1200

accaaacatg ttgctgggta ttataatacc atttaggtgg gttgggttct taactaggat 1260  
catggcggca gaatcaatca attgatgttg aaccttcaat gtagggaatt cgttcttgat 1320  
ggtttcctcc acagtttttc tccataatct tgaagaggcc aaaacattag ctttatccaa 1380  
ggaccaaata ggcaatggtg gctcatgttg tagggccatg aaagcggcca ttcttggat 1440  
tctttgcact tctggaacgg tgtattgttc actatcccaa gcgacacat caccatcgtc 1500  
ttcctttctc ttaccaaagt aaatacctcc cactaattct ctgacaacaa cgaagtcagt 1560  
accttagca aattgtggct tgattggaga taagtctaaa agagagtcgg atgcaaagtt 1620

acatggtctt aagtggcgt acaattgaag ttctttacgg atttttagta aacctgttc 1680  
aggctcaaca ctaccggtac cccatttagg accaccaca gcacctaaca aaacggcatc 1740

aaccttcttg gaggcctcca ggcctcatc tggaaatggg acacctgtag catcgaatgc 1800  
agcaccacca attaaatgat ttctgaaatc gaacttgaca ttggaacgaa catcagaaat 1860  
agctttaaga accttaatgg ctctcgctgt gatctcttga ccaacgtggt cacctggcaa 1920  
aacgacgac ttcttagggg cagacatagg ggcagacatt agaatggtat atccttgaaa 1980  
tatatatata tattgctgaa atgtaaaagg taagaaaagt tagaaagtaa gacgattgct 2040  
  
aaccacctat tggaaaaaac aataggctct taataatat tgtcaacttc aagtattgtg 2100  
atgcaagcat ttagtcatga acgcttctct attctatatg aaaagccggg tccggcctct 2160  
cacctttctt tttctccca attttctcagt tgaaaaagg atatgcgtca ggcgacctct 2220  
gaaattaaca aaaaatttcc agtcatcgaa ttgattctg tgcgatagcg cccctgtgtg 2280  
ttctcgttat gttgaggaaa aaaataatgg ttgctaagag attcgaactc ttgcatctta 2340  
cgatactga gtattcccac agttaactgc ggtcaagata tttcttgaat caggcgctc 2400  
gctcgtccaa cgccggcgga cctcttaaat gagaaaaatt tcgtaatgag ataaaattc 2460  
  
gctccttttc tgttttctat tttctatctt cccaactttt gctctattca gttataaatt 2520  
actatattate catcagttaa aaaacaagat ctcttactgg tcagctagga aagcgaaaat 2580  
acaaagactt tatgactta gtgatataata tgtatagata tatccatttt tacgactta 2640  
tcatatatct tagttatcta aatacaatct agttattcgt acacaatcgc cctgttate 2700  
cctatagtgg gaataaagta atgcactgtg acgggggtct tcgccccgga tagggtaaaa 2760  
ggatattgcc gttcaagaa acttcgggga taatcgaata agataccgag aaagctattg 2820  
ttcgttgtgc acgtaggatg tatattgaac aagcatgacc agaactgat gcattacgag 2880  
  
aaggttacgg gatgatatca gacctcgaa gtccatgttg caaaatgtgc cgactttccg 2940  
cggcgctatt tggcacaat ttcaggagaa acatcactgt cgggtttata gaattccatc 3000  
tatattgttt tccccgtagg catacgtcga gcgggtgtta aaccccagcg cctggcggg 3059  
<210> 28  
<211> 8106  
<212> DNA  
<213> Artificial Sequence  
<220><223> Synthetic: i74804 integration construct  
<400> 28  
gacggcacgg ccacgcgttt aaaccgccc gctgcctcat cccacggga ataaggcagc 60  
cgacaaaaga aaaacgaccg aaaaggaacc agaaagaaaa aagaggggtg gcgcgccgcg 120  
gacgtgtaaa aagatatgca tccagcttct atatcgcttt aactttaccg ttttgggcat 180

cgggaacgta tgiaacattg atctcctctt gggaacggtg agtgcaacga atgcgatata 240  
 gcaccgacca tgtgggcaaa ttcgtaataa attcgggggtg aggggggattc aagacaagca 300  
 accttgttag tcagctcaaa cagcgattta acggttgagt aacacatcaa aacaccgttc 360  
 gaggtcaagc ctggcgtggt taacaagttc ttgatatcat atataaatgt aataagaagt 420  
 ttgtaatat tcaattcgaa gtgttcagtc ttttacttct cttgttttat agaagaaaaa 480  
 acatcaagaa acatctttaa catacacaaa cacatactat cagaatacac gctcgtccaa 540  
  
 cgccggcgga ctttcagac gcgactgcct catcagtaag acccgttgaa aagaacttac 600  
 ctgaaaaaaaa cgaatatata cttagcgttga atgttagcgt caacaacaag aagtttaatg 660  
 acgcgagggc caaggcaaaa agattccttg attacgtaag ggagttagaa tcattttgaa 720  
 taaaaaacac gctttttcag ttcgagtta tcattatcaa tactgccatt tcaaagaata 780  
 cgtaaataat taatagtagt gattttccta actttattha gtcaaaaaat tagcctttta 840  
 attctgctgt aaccctgaca tgcccaaaaat agggggcggg ttacacagaa tatataacat 900  
 cgtaggtgtc tgggtgaaca gtttattcct ggcatccact aaatataatg gagcccgtt 960  
  
 tftaagctgg catccagaaa aaaaaagaat cccagcacca aaatattgtt ttcttcacca 1020  
 accatcagtt cataggtcca ttctcttagc gcaactacag agaacagggg cacaaacagg 1080  
 caaaaaacgg gcacaacctc aatggagtga tgcaacctgc ctggagtaaa tgatgacaca 1140  
 aggcaattga cccagcatg tatctatctc attttcttac accttctatt accttctget 1200  
 ctctctgatt tggaaaaagc tgaaaaaaaaa ggttgaacc agttccctga aattattccc 1260  
 ctacttgact aataagtata taaagacggt aggtattgat tgtaattctg taaatctatt 1320  
 tcttaaacct cttaaatctt acttttatag ttagtctttt ttttagtttt aaaacaccaa 1380  
  
 gaacttagtt tgcacctccc gcgacctcca aaatcgaact accttcacaa tggaaacattc 1440  
 tgtaatcgaa ccaactgtgc ccatgccgct accagccatg tttgacgctc catctggtat 1500  
 ttttagctct ttggacgacg ctgtgcaagc agccacctta gccaacaac aactaagttc 1560  
 agttgagttg cgtcagcaag taatcaaagc cataagagtg gccggagaaa ggtatgcaca 1620  
 agttttggct gaaatggcag ttgctgaaac tggatgggt aggggtgggtg ataagtacat 1680  
 taagaatgtc tctcaagctc gtcatacgcc tggatagaa tgtttatcgg ccgaggttct 1740  
 tacgggtgat aatggcctaa cattgattga aaatgccctt tggggagtgc tagcttcagt 1800  
  
 cacgccaagc acaaatccag cagctacggt aattaataat gcaatctcaa tgattgcagc 1860  
 ggggaattca gtcgtgttcg caccacatcc ttctgcaaaa aacgtctcac taaggactat 1920  
 ttctttactc aacaaggcca ttgtcgctac cggcggecca gaaaatttac tagttagtgt 1980  
 ggcaaacctt aacatcgaag ctgcacagag attattcaga tatccgggta ttggattggt 2040

agttgtgaca ggtggtgaag ccgtcgttga agccgctagg aagcatacag ataaaaggtt 2100  
 aattgcagcc ggcgctggta atcctcctgt tgttgtggac gaaactgctg acatacctaa 2160  
 agccgcaaga gcaattgtca aggggtgcttc tttcgacaac aacataatit gtgctgatga 2220  
  
 aaaagttttg atigtggtag acagagttgc agatgcacta ttggcagaaa tgcaaagaaa 2280  
 taacgccgctc ttacttacac ccgaacagac cgaagacta ctaccgctc ttttgtccga 2340  
 tattgacgaa cagggcaaaag gacgtgtgaa tagagattat gttggaagag atgccgctaa 2400  
 attagcagcg gctattggtc tggaaagttag cgaacatact cgtctactcc tggcagagac 2460  
 agacgctgat catccattcg ccgtgacgga gctgatgatg ccagtgttac cagtaataag 2520  
 agtcaagaat gtagatgatg caatcgcait ggcagttaaag ctagagtcag gctgcagaca 2580  
 cacagctgcg atgcactcta ctaataaag aaacttaaat agaatggcta atgcatcaa 2640  
  
 tacctctatc tttgtaaaaa atggccatg tattgcaggt ttgggttttag gcggtgaagg 2700  
 ttggacttca atgactatta gcactccgac cggatgaagg gttacaagcg ctctacctt 2760  
 tgtcagatta agaagggtgt tcttagtcga catgtttcgg attgcttaag cggcccgag 2820  
 taataattat tgettcata taatatitit atatacctct tatttttatg tattagttaa 2880  
 ttaagtattt ttatctatct gcttatcatt ttcttttcat ataggggggg ttggtgtttt 2940  
 cttgcccate agattgatgt cctccaactc ggcactatit tacaagggt ttttttghta 3000  
 gagaaggaga agacagatc taaaccatac gttactcgaa acaaaaaaaaa aaaaaatgga 3060  
  
 aaaagctgct atcaacaaaa gacggcctca tcaaacctaa agaaacctg tcagcgtatg 3120  
 tatafacctt gtaatttacg ttctctttaa tcttctttct actaacgtit tcattattct 3180  
 atactctatg accaataaaa acagactgia ctttcaaaat ttaccagta ggccagcaaa 3240  
 taaagaaaat tataccagat tacttctgaa acacattaat cccaacaaca agtatgcat 3300  
 taatccgtcg ctaccccate cccgctgct tggccggccg tacactgagt aatggtagtt 3360  
 ataagaaaga gaccgagtta gggacagtta gaggcgggtg agatattcct tatggcatgt 3420  
 ctggcgtatg taaaactitit caaacggcag ccccgatcta aaagagctga cagggaatg 3480  
  
 gtcagaaaaa gaaacgtgca cccgcccgtc tggacgcgcc gtcaccccgc acggcagaga 3540  
 ccaatcagta aaaaacaacg gttaacgaca ttactatata tataatatag gaagcattta 3600  
 atagaacagc atcgtaatat atgtgtactt tgcagttatg acgccagatg gcagtagtgg 3660  
 aagatattct ttattgaaaa atagcttgtc accttacgta caatcttgat cggagctitit 3720  
 tctttttttg ccgattaaga attcggtcga aaaaagaaaa ggagagggcc aagaggagg 3780  
 gcattggtga ctattgagca cgtgagtata cgtgattaag cacacaaagg cagcttggag 3840

tatgtctggt attaatttca caggtagttc tgggccattg gtgaaagttt gcggcttgca 3900

gagcacagag gccgcagaat gtgctctaga ttccgatgct gacttgctgg gtattatag 3960

tgtgcccaat agaaagagaa caattgaccg ggttattgca aggaaaattt caagtcttgt 4020

aaaagcatal aaaaatagtt caggcactcc gaaatacttg gttggcgtgt ttcgtaatca 4080

acctaaggag gatgttttgg ctctgggtcaa tgattacggc attgatatcg tccaactgca 4140

tggagatgag tcgtggcaag aataccaaga gttcctcggg ttgccagtta ttaaaagact 4200

cgtatttcca aaagactgca acatactact cagtgcagct tcacagaaac ctcatctggt 4260

tattcccttg tttgattcag aagcaggtgg gacaggtgaa cttttggatt ggaactcgat 4320

tctgactgg gtiggaaggc aagagagccc cgaaagctta cattttatgt tagctggtgg 4380

actgaccca gaaaatgttg gtgatgctct tagattaaat ggcgttattg gtgttgatgt 4440

aagcggagggt gtggagacaa atgggtgtaa agactctaac aaaaatagcaa atttcgtcaa 4500

aaatgctaag aaatagggtta ttactgagta gtattttatt aagtattggt tgtgacttgg 4560

cctgcaggcc ttttgaaaag caagcataaa agatctaacc ataaaatctg taaaataaca 4620

agatgtaaag ataatgctaa atcatttggc tttttgattg attgtacagg aaaaataaca 4680

tcgcaggggg ttgactttta ccatttcacc gcaatggaat caaacttggt gaagagaatg 4740

ttcacaggcg catacgtac aatgacacgg ccggccaagc acgcggggat ggggtagcga 4800

cggattaatg gcatacttgt tgttgggatt aatgtgttc agaagtaatc tggataaatt 4860

ttctttattt gctggcctac tgggtaaat ttgaaagtac agtctgtttt tattggtcat 4920

agagtataga ataataaaaa cgttagtaga aagaagattt aaggaaactg aaattacaag 4980

gtatatacat acgctgacat ggtttcttta ggtttgatga ggccgtcttt tgttgatagc 5040

agctttttcc atttttttt tttttgttc gagtaacgta tggtttagta tctgtcttct 5100

ccttctctta caaaaaaac cttttgtaaaa tagtgccgag ttggaggaca tcaatctgat 5160

gggcaagaaa acaccaacc ccctatatg aaaagaaaat gataagcaga tagataaaaa 5220

tacttaatta actaatacat aaaaataaga ggtatataaa aatattatat ggaagcaata 5280

attattactc gcggccgctt aagcaatccg aaacatgtcg actaagacac accttcttaa 5340

tctgacaaag gtacgagcgc ttgtaacacc tcaccggtc ggagtgctaa tagtcattga 5400

agtccaacct tcaccgcta aacccaacc tgcaatacat ggaccatttt ttacaaagat 5460

agaggattg atggcattag ccattctatt taagtttctt atattagtag agtgcatcgc 5520

agctgtgtgt ctgcagcctg actctagctt aactgccaat gcgattgcat catctacatt 5580

cttgactctt attactggta acactggcat catcagctcc gtcacggcga atggatgatc 5640  
 agcgtctgtc tctgccagga gtagacgagt atgttcgcta acttccagac caatagccgc 5700  
 tgctaattta gccgcatctc ttccaacata atctctattc acacgtcctt tgcctgttc 5760  
 gtcaatatcg gacaaaagag cgggtagtag tctttcggtc tgttcgggtg taagtaagac 5820  
 ggcgttattt ctttgcattt ctgccaatag tgcattctga actctgtcta ccacaatcaa 5880  
 aactttttca tcagcacaaa ttatgttgtt gtcgaaagaa gcacccttga caattgctct 5940  
 tgcggcttta ggiatgtcag cagtttcgtc cacaacaaca ggaggattac cagcgcgggc 6000  
  
 tgcaattaac cttttatctg tatgtttcct agcggcttca acgacggctt caccacctgt 6060  
 cacaactaac aatccaatac ccggatatct gaataatctc tgtgcagttt cgatgttagg 6120  
 gtttgccaca ctaactagta aattttctgg gccgccggtg gcgacaatgg cttgttgag 6180  
 taaagaata gtccttagtg agacgttttt ggcagaagga tgtggtgcga acacgactga 6240  
 attccccgt gcaatcattg agattgcatt attaattacc gtagctgctg gattttgtct 6300  
 tggcgtgact gaagctacga ctccccagg ggcattttca atcaatgta ggccattatc 6360  
 acccgtaaaga acctcgcccg ataaacattc tataccagge gtatgacgag cttgagagac 6420  
  
 attcttaatg tacttatcca ccacctacc cataccagtt tcagcaactg ccatttcagc 6480  
 caaaacttgt gcatacttt ctccggccac tcttatggct ttgattactt gctgacgcaa 6540  
 ctcaactgaa cttagttgtt gttgggctaa ggtggctgct tgcacagcgt cgtccaaaga 6600  
 gctaaaaata ccagatggag cgtcaacat ggctggtagc ggcatgggca cagttggttc 6660  
 gattacagaa tgttccattg tgaaggtagt tcgattttgg aggtcgcggg aggtcgaac 6720  
 taagtctttg gtgttttaa actaaaaaaa agactaacta taaaagtaga atttaagaag 6780  
 tttagaagaat agatttacag aattacaatc aatactacc gtcittatat acttattagt 6840  
  
 caagtagggg aataatttca gggaaactgt ttcaaccttt ttttcagct ttttccaaat 6900  
 cagagagagc agaaggtaat agaaggtgta agaaaatgag atagatacat gcgtgggtca 6960  
 attgccttgt gtcacattt actccaggca ggttgcacat ctccattgag gttgtgcccc 7020  
 tttttgcct gtttgtgcc ctgttctctg tagttgcgct aagagaatgg acctatgaac 7080  
 tgatggttgg tgaagaaaac aatattttgg tgcctgggatt ctttttttt ctggatgcca 7140  
 gcttaaaaag cgggctccat tatatttagt ggatgccagg aataaactgt tcaccagac 7200  
 acctacgatg ttatatattc tgtgtaacc gccccctatt ttggcatgt acgggttaca 7260  
  
 gcagaattaa aaggctaatt ttttgactaa ataaagttag gaaaactact actattaatt 7320  
 atttactgat tctttgaaat ggcagtattg ataataataa actcgaactg aaaaagcgtg 7380  
 tttttattc aaaatgattc taactccctt acgtaataca ggaatctttt tccttggcc 7440

tccgcgtcat taaacttctt gttgttgacg ctaacattca acgctagtat atattcgttt 7500  
 ttttcaggta agttcttttc aacgggtctt actgatgagg cagtcgcgtc tgaaaggtcc 7560  
 gccggcggtg gacgagcgtg taccaacctg cttttctttc cgtcatatac acaaaatact 7620  
 ttcatataaa cttacttggg cttacgtcat aaataaatat gtatacatat aaattaaaaa 7680

atltggtttt atatttttac aaaaagaatc gtttacttca tttctccctt ttaagcgata 7740  
 caatccatga aaaaagagaa aaagagagaa caggcttgtg ctttctttaa aacatcccac 7800  
 acaaaatcat attgaattga attttacatc ttaagctagt gtacaacaac tgctatatcc 7860  
 aaagaaaact aacgtggacc gcttttagag ttgagaaaaa ggtttgaaaa aatagcaat 7920  
 acaaagactt gtttcatata taaaatacag ggagcacatt gagctaatat aacataaaca 7980  
 ctgcgaacca attccaatca aaaggtacac atgagagcat tcccccgagt actgccattt 8040  
 cgccatcaga gatcatataa taacatcctt cttcgaacgg cggtttaaac gcgtggccgt 8100

gccgtc 8106

<210> 29  
 <211> 2404  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic: i76220 integration construct  
 <400> 29

gacggcacgg ccacgcgttt aaaccgccgc acgtgtatgt acggctgtgt aatatgata 60  
 atcatctcgg acgaacggcg tagtactctc catcccctaa aaatgttcac gtgtgactgc 120  
 tccatttcgc cggatgtcga gatgaccccc cccctcaaa aggcactcac ctgctgacat 180  
 gccgtggcaa atgattgggg tcatcctttt tttctgttat ctctaagatc caaagaaaag 240

taaaaaaaa aggttggggg acgaattgcc gccgagcctc cgatgccatt attcaatggg 300  
 tattgcagtt ggggtatagt tctcgggtgg caaatagttc tcccttcatt ttgtatataa 360  
 actgggcggc tattctaagc atatttctcc cttaggttat ctggtagtac gttatatctt 420  
 gttcttatat ttctatcta taagcaaac caaacatata aaaactacta gaaagacatt 480  
 gccccactgt gttcgtcgt ccaacgccgg cggaccttcc tcgacgtggg cttttttctt 540  
 gccatatgga tccgctgcac ggtcctgttc cctagcatgt acgtgagcgt atttcctttt 600  
 aaaccacgac gctttgtctt cattcaacgt ttccattgt tttttctac tattgctttg 660

ctgtgggaaa aacttatcga aagatgacga ctttttctta attctcgttt taagagcttg 720  
 gtgagcgcta ggagtcactg ccaggtatcg ttgaaacacg gcattagtca gggaagtcac 780

aacacagtcc tttcccgcaa ttttcttttt ctattactct tggcctcctc tagtacactc 840  
tatatTTTTT tatgctcgg taatgatttt cattTTTTTT tttccaccta gcggatgact 900  
ctTTTTTTTT cttagcgatt ggcatatca cataatgaat tatacattat ataaagtaat 960  
gtgatttctt cgaagaatat actaaaaaat gaggcaggcaa gataaacgaa ggcaaagatg 1020  
acagagcaga aagccctagt aaagcgtatt acaaatgaaa ccaagattca gattgcatc 1080  
  
tctttaaagg gtggtccctt agcgatagag cactcgatct tcccagaaaa agaggcagaa 1140  
gcagtagcag aacaggccac acaatcgcaa gtgattaacg tccacacagg tatagggttt 1200  
ctggaccata tgatacatgc tctggccaag cattccggct ggtcgctaat cgttgagtgc 1260  
attggtgact tacacataga cgaccatcac accactgaag actgcgggat tgctctcggt 1320  
caagctttta aagaggccct aggggccgtg cgtggagtaa aaaggtttgg atcaggattt 1380  
gcgcctttgg atgaggcact ttccagagcg gtggtagatc tttcgaacag gccgtacgca 1440  
gttgtcgaac ttggtttgca aaggagaaaa gtaggagatc tctcttgca gatgatcccg 1500  
  
cattttcttg aaagctttgc agaggctagc agaattacc tccacgttga ttgtctgca 1560  
ggcaagaatg atcatcaccg tagtgagagt gcgttcaagg ctcttgcggt tgccataaga 1620  
gaagccacct cgccaatgg taccaacgat gttccctcca ccaaggtgt tcttatgtag 1680  
tgacaccgat ttttaaacg tgcagcatac gatatatata catgtgtata tatgtatacc 1740  
tatgaatgtc agtaagtatg tatacgaaca gtatgatact gaagatgaca aggtaatgca 1800  
tcattctata cgtgtcattc tgaacgaggc gcgctttct ttttctttt tgctttttct 1860  
ttttttttct cttgaactcg aggtccgccg gcgttggacg agcgtgatga tttctttct 1920  
  
ttttatattg acgacttttt tttttcgtg tgtttttgtt ctcttataac cgagctgctt 1980  
acttattatt atttcacctt ctctttttat ttatacttat aattatttat tctttacata 2040  
ctgttacaag aaactctttt ctacattaat tgcataaagt gtcaatcagc acatcctcta 2100  
tatcgctatc aacaacaaat ttgacaaacc tgcttatatc ttcaggaaca actgccgat 2160  
cgtaccacc actacttggt aagtccttgg agtttaatat gcaactgaaat ttacctagcc 2220  
gttttacaca agaccataat ccatcatgc tatcgagta tatgattttg tgttcgtttt 2280  
tcgtcttgcg aaaggcatcc tcaatggctt gtttcattga tccatcagtg tggctcgtag 2340  
  
gtaccagcaa aaccacttca tcagcggcgt actcctggcg gtttaaacgc gtggccgtgc 2400  
cgtc 2404  
<210> 30  
<211> 8536  
<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: i73830 integration construct

<400> 30

```

gacggcacgg ccacgcgttt aaaccgccac ccagccaagg tagtctaaaa gctaatttct      60
ctaaaaggga gaaagttggt gattttttat ctgcattat tataatgca agaatagtta      120
aggtatagtt ataaagtttt atcttaattg ccacatacgt acattgacac gtagaaggac      180

tccattattt ttttcattct agcatactat tattccttgt aacgtcccag agtattccat      240
ttaattgtcc tccatttctt aacggtgacg aaggatcacc atacaacaac tactaaagat      300
tatagtacac ttcaccttg caactattta tctgacattt gccttacttt tatctccagc      360
ttcccctcga tttatTTTTT caatttgatt tctaaagctt tttgcttagg cataccaaac      420
catccactca tttaacacct tatttttttt ttcgaagaca gcatccaact ttatacgttc      480
actacctttt tttttacaac aatttcattc ttcacacctat gaacgctcgt ccaacgccgg      540
cggacctttc agacgcgact gcctcatcag taagaccctg tgaagaagaac ttacctgaaa      600

aaaacgaata tatactagcg ttgaatgtta gcgtcaacaa caagaagttt aatgacgcgg      660
aggccaaggc aaaaagattc cttgattacg taaggagatt agaatcattt tgaataaaaa      720
acacgctttt tcagttcgag tttatcatta tcaatactgc catttcaaag aatacgtaaa      780
taattaatag tagtgatttt cctaacttta tttagtcaaa aaattagcct ttaattctg      840
ctgtaaccgg tacatgccca aaataggggg cgggttacac agaatatata acatcgtagg      900
tgtctgggtg aacagtttat tcttggcatc cactaaatat aatggagccc gctttttaag      960
ctggcatcca gaaaaaaaa gaatcccagc accaaaatat tgttttcttc accaaccatc     1020

agttcatagg tccattctct tagcgaact acagagaaca ggggcacaaa caggcaaaaa     1080
acgggcacaa cctcaatgga gtgatgcaac ctgcctggag taaatgatga cacaaggcaa     1140
ttgaccacag catgtatcta tctcattttc ttacaccttc tattaccttc tgctctctct     1200
gatttggaaa aagctgaaaa aaaaggttga aaccagtcc ctgaaattat tcccctactt     1260
gactaataag tatataaaga cggtaggtat tgattgtaat tctgtaaate ttttcttaa     1320
acttcttaaa ttctactttt atagttagtc ttttttttag ttttaaaaca ccaagaactt     1380
agtttcgacc tcccgcgacc tccaaaatcg aactaccttc acaatggctg atttcgattc     1440

taaagaatac ttggagttag ttgacaagtg gtggcgtgcc accaactact tgtccgctgg     1500
tatgattttc ttgaagtcca acccattatt ctctgttact aatacccaaa tcaaggccga     1560
agatgtcaaa gttaaaccaa ttggctactg gggtactatt tccggtcaaa ctttcttata     1620
cgcccacgct aaccgtttga ttaacaagta cggctcctcaac atgttttacg ttggtggtcc     1680

```

aggtcacggt ggicaagtca tggttactaa cgcctactta gacggtgcct acaccgaaga 1740  
 ttaccagaa attactcaag acatcgaagg tatgtctcat ttgttcaagc gtttctcttt 1800  
 ccctggtggt attggttccc atatgaccgc tcaaaactcca ggttccttgc acgaaggtgg 1860  
  
 tgaattgggt tactctttgt cccatgcttt cggtgctggt ttggacaacc cagaccaagt 1920  
 tgcttttctc gtcgttgggt atggtgaagc tgaactggt ccatctatgg cctcttggca 1980  
 ttccattaag ttcttaaag ccaagaacga tggtgccgtt ttgccagttt tggatttaaa 2040  
 cggtttcaag atttccaatc caaccatttt ttctagaatg tctgatgaag aaattactaa 2100  
 gtctctgaa ggtttgggtt attcccctag attcattgaa aatgatgaca ttcacgacta 2160  
 cgccacctac caccaattgg ccgctaacat cttagatcaa gccatcgaag acattcaagc 2220  
 tattcaaaat gacgccagag agaatggtaa atatcaagat ggtgaaatc cagcttggcc 2280  
  
 tgttattatc gctagattgc caaagggttg ggggtgtcca acccagatg cttctaataa 2340  
 tccaattgaa aactctttca gagctacca agttcatta ccatgggaac aacacgattt 2400  
 ggccaccttg ccagaattcg aagattggat gaactcttac aagccagaag aattattcaa 2460  
 cgctgatggt tecttgaagg atgagttgaa agctattgcc ccaaagggtg ataagagaat 2520  
 gtctgctaac ccaatcacca acgggtggtc tgacagatcc gacttgaat tgccaaattg 2580  
 gagagaatc gctaacgaca tcaacgacga taccagaggt aaggaattcg ctgactctaa 2640  
 gagaaacatg gatatggcta ctttatccaa ctatttaggt gccgtttctc aattgaacce 2700  
  
 aaccagatc agattcttcg gtccagatga aaccatgtcc aacagattgt ggggtttggt 2760  
 taatgttacc ccacgtcaat ggatggaaga aatcaaggaa ccacaagatc aattgttctc 2820  
 tccaactggt cgtatcatcg attcccatt gtctgaacac caagctgaag gttggttggga 2880  
 aggttacact ttgactgcta gagttggtat ctttgcctct tacgaatctt tcttgagagt 2940  
 tgttgatacc atggtcactc aacatttcaa gtggttgcgt cacgcttccg aacaagcttg 3000  
 gagaaatgac tatccatcct taaatttgat cgctacctct accgctttcc aacaagatca 3060  
 taacggttat actcaccaag accctggtat gttaaactcat ttggccgaga agaagtctaa 3120  
  
 cttcattaga gaatatttgc cagccgacgg taactctttg ttagccgttc aagagagagc 3180  
 tttctctgaa agacataagg ttaacttatt gatcgcttct aaacaaccaa gacaacaatg 3240  
 gttcactggt gaagaagctg aagtcttagc taacgaaggt ttgaagatta tcgattgggc 3300  
 ttctactgct ccatcttccg atgttgatat tacttttctc tctgccgta ctgaaccaac 3360  
 cattgagact ttggccgctt tatggttgat taatcaagct ttcctgacg ttaagtttag 3420  
 atacgttaac gttgttgaat tgtaagatt gcaaaagaaa tctgaaccaa acatgaacga 3480

cgaaagagaa ttatctgccg aagaatttaa taagtacttc caagccgaca ctccagtat 3540  
  
 ctfcggtttc cacgcttacg aaaacttgat tgaatctttc ttttcgaga gaaagttcac 3600  
 cggatgatgc tatgttcacg gttatagaga agatggatgat atcactacca cctacgatat 3660  
 gagagtctat tcccacttgg atcgtttcca tcaagccaag gaagccgccg aatcttgtc 3720  
 tgctaacggt aaaaatcgacc aagccgctgc cgacaccttt attgctaaga tggacgacac 3780  
 tttggccaaa cacttccaag ttactagaaa tgaaggtaga gatattgaag aattcactga 3840  
 ctggacttgg tctccattga agtaagtga tttactttaa atcttgcat taaataaatt 3900  
 ttttttttat agctttatga cttagtttca atttatatac tttttaatg acattttcga 3960  
  
 ttcattgatt gaaagctttg tgttttttct tgatgcgcta ttgcattgtt ctgtctttt 4020  
 tcgccacatg taatatctgt agtagatacc tgatacattg tggatgctga gtgaaatfff 4080  
 agttaataat ggagcgctc ttaataatff tggggatatt ggcttatccc cgcgtgcttg 4140  
 gccggccgta cgaaaatcgt tattgtcttg aaggtgaaat tctactctt ataatgggtg 4200  
 aacgttaagc tgatgctatg atggaagctg atfggtctta acttgcttgt catcttgcta 4260  
 atggtcattg gctcgtgta ttacttaagt tttttgact cgttttgaac gtaatgctaa 4320  
 tgatcatctt atggaataat agtgagtggf ttcagggfcc ataaagctff tcaattcatc 4380  
  
 tttttttttt ttgttctfff ttttgattcc ggtttctttg aaatffffff gattcggtaa 4440  
 tctccgagca gaaggaagaa cgaaggaagg agcacagact tagattggta tatatacgca 4500  
 tatgtggtgt tgaagaaaca tgaattgcc cagtatctt aaccaactg cacagaacaa 4560  
 aaacctgcag gaaacgaaga taaatcatgt cgaagctac atataaggaa cgtgctgcta 4620  
 ctcatcctag tctgtttgct gccaaagctat ttaatatcat gcacgaaaag caaacaact 4680  
 tgtgtgcttc attggatggt cgtaccacca aggaattact ggagttagtt gaagcattag 4740  
 gtcccaaat ttgtttacta aaaacacatg tggatatctt gactgatttt tccatggagg 4800  
  
 gcacagttaa gccgctaaag gcattatccg ccaagtacaa tttttactc ttcgaagaca 4860  
 gaaaatffgc tgacattggt aatacagtca aatgcagta ctctcgggtg gtatacagaa 4920  
 tagcagaatg ggcagacatt acgaatgcac acggtgtggf gggcccaggt atgttagcg 4980  
 gtttgaagca ggcggcggaa gaagtaacaa aggaacctag aggcctfff atgttagcag 5040  
 aattgcatg caagggctcc ctagctactg gagaatatac taagggtact gttgacattg 5100  
 cgaagagtga caagatfff gttatcggct ttattgctca aagagacatg ggtggaagag 5160  
 atgaaggtta cgattggtg attatgacac ccggtgtggg tttagatgac aaggagacg 5220

cattgggtca acagtataga accgtggatg atgtggcttc tacaggatct gacattatta 5280  
ttgttggaag aggactattt gcaaagggaa gggatgctaa ggtagagggt gaacgttaca 5340  
gaaaagcagg ctgggaagca tatttgagaa gatgcgcca gcaaaactaa aaaactgtat 5400  
tataagtaaa tgcatgtata ctaaaactcac aaattagagc ttcaatttaa ttatatcagt 5460  
tattaccacg aaaatcgtaa ttgtcttgaa ggtgaaatth ctactcttat taatggtgaa 5520  
cgtaaagctg atgctatgat ggaagctgat tggctttaac ttgcttgtca tcttgctaat 5580  
ggtcatatgg ctctgtttat tacttaagtt atttgtactc gttttgaacg taatgctaat 5640  
  
gatcatctta tggaataata gtgaacggcc ggccaagcac gcggggatgg gatgagcttg 5700  
gagcaggaag aatacactat actggatctc aagagtacaa tagatggata agaattttgg 5760  
cagcgcaaaa aggcttcaag cttacacaac acggtttatt tcgaaataat atccttctcg 5820  
aaagctttaa cgaacgcaga attttcgagt tattaaactt aaaatacgtc gaacccgaac 5880  
atagaaatat cgaatgggaa aaaaaactg cataaaggca ttaaagagg agcgaattht 5940  
tttttaataa aaatcttaat aatcattaaa agataaataa tagtctatat atacgtatat 6000  
aaataaaaaa tattcaaaaa ataaaaataa ctattattht agcgtaaaagg atggggaaag 6060  
  
agaaaagaaa aaaattgatc tatcgatttc aattcaattc aatagatctt tatccttgtg 6120  
cttgtcctg aactgcggta acggcaacaa ctttgacgat gtcgtcgact gaacatcccc 6180  
ttgacaaate gttgataggt ttggcaaatc cctgacatat aggaccgatg gcttccgctt 6240  
ttgcgaatct ttggaccaac ttgtatccga tgtttcctgc ctggatgtct gggaagatca 6300  
agacatttg cttaccagcg actttagatc caggggctth caaatctgcg accttcttaa 6360  
ccaatgagge gtctaactgc aattcaccgt cgatgtctaa gtcaggccta gcctccttag 6420  
ccaatthtgt tgcctttgta accttgcga ctaattcatg tgaggctgat cccatggttg 6480  
  
agaatgacaa catggctacc cttggctcga tcttgacaaa attctttgca gtctcagcag 6540  
tggtaattg gattgaagat aactcttcag cgttaggaca aacatttaca gcgcagtcag 6600  
cgaataacaa aaaaccgtcc tctccatact cgcagtcagg tactgacatc aagaagactg 6660  
atgagacgac agatgcacct ggtactgtht tgacaatctg caaaccaggc cttacaagt 6720  
ctcctgtagt atglatagca ccagatacca aaccgtcagc gtcacctaac ttgacatca 6780  
ttgttgcaaa gtagattggg tcctgacga tttgtcagc cttctccaag gtgactcct 6840  
tgthttttct gatctcgtag aaagcgttgg cgtaacggc ggtcttagaa gaagthtctg 6900  
  
ggtcgactat ctctactccg gccaaatth ctcgaattht tgcggcgtth tccttaatga 6960  
cagactctga accgaccaag attatgctgg caatacgtc cctaataatc tctctgaa 7020  
ccctgatgth cctctctcc tcacctctg ccaaacgat thtctcttg tcggccttg 7080

ccaatccgaa gatattctcc atcaatttca ttgtgaaggt agttcgattt tggaggtcgc 7140  
 gggaggtcga aactaagttc ttggtgtttt aaaactaaaa aaaagactaa ctataaaagt 7200  
 agaatttaag aagtttaaga aatagattta cagaattaca atcaatacct accgtcttta 7260  
 tatacttatt agtcaagtag gggaataatt tcaggaact ggtttcaacc tttttttca 7320

gctttttcca aatcagagag agcagaaggt aatagaaggt gtaagaaaat gagatagata 7380  
 catgcgtggg tcaattgctt tgtgcatca ttactccag gcaggttgca tcaactcatt 7440  
 gaggttgtgc ccgttttttg cctgtttgtg ccctgttct ctgtagtgc gctaagagaa 7500  
 tggacctatg aactgatggt tgggaagaa aacaatattt tgggtctggg attctttttt 7560  
 tttctggatg ccagctaaa aagcgggctc cattatattt agtggatgcc aggaataaac 7620  
 tgttcacca gacacctacg atgttatata ttctgtgtaa cccgcccct attttgggca 7680  
 tgtacgggtt acagcagaat taaaaggcta atttttgac taaataaagt taggaaaatc 7740

actactatta attatttacg tattctttga aatggcagta ttgataatga taaactcgaa 7800  
 ctgaaaaagc gtgtttttta ttcaaatga ttctaactcc cttacgtaat caaggaatct 7860  
 ttttgcttg gctcgcgt cattaactt ctgtttgtg acgtaacat tcaacgctag 7920  
 tataatctg ttttttcag gtaagttctt ttcaacgggt cttactgatg aggcagtcgc 7980  
 gtctgaaagg tccgccggc ttggacgagc gctccatgct ggacttactc gtcgaagatt 8040  
 tcctgctact ctctatata ttagacacc atgttataga tttcagaaaa caatgtaata 8100  
 atatatgta gcctcctgaa actaccaagg gaaaaatctc aacaccaaga gctcatattc 8160

gttggaatag cgataatc tctttacctc aatcttatat gcatgttatt tgctcttata 8220  
 attggtctct atttagggaa aaaagtcggt ttgagagctt ctgcgatgt gaaatctcaa 8280  
 ttgaaactgc acgcaaaagc tagccattt cacgaacacc agaaagaaga aatcccaag 8340  
 gatcgcata cagagtatgc tctctcatat cgttgagtat gaatgccaat aactgatca 8400  
 gctttacaag aaacgtaaaa tctggcacga tggtagactg aaatacttc agttaaaca 8460  
 cagattcatg ctttatacgg aaaaggataa cgttttgta gctagtgagg cggtttaaac 8520  
 gcgtggccgt gccgtc 8536

<210> 31

<211> 9734

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: i74810 integration construct

<400> 31

gacggcacgg ccacgcgttt aaaccgcccg ctgcctcat cccacggga ataaggcagc 60  
 cgacaaaaga aaaacgaccg aaaaggaacc agaaagaaaa aagagggtgg gcgcgccgcg 120  
 gacgtgtaaa aagatatgca tccagcttct atatcgcttt aactttaccg ttttgggcat 180  
 cggaacgta tgtaacattg atctcctctt gggaacgggtg agtgcaacga atgcatata 240  
 gcaccgacca tgtgggcaaa ttcgtaataa attcgggggtg agggggattc aagacaagca 300  
  
 accttgtag tcagctcaaa cagcgattta acggttgagt aacacatcaa aacaccgttc 360  
 gaggtcaagc ctggcgtgtt taacaagttc ttgatatcat atataaatgt aataagaagt 420  
 ttggtaatat tcaattcgaa gtgttcagtc ttttacttct cttgttttat agaagaaaaa 480  
 acatcaagaa acatctttaa catacacaaa cacatactat cagaatacac gctcgtccaa 540  
 cgccggcgga ctttcagac gcgactgcct catcagtaag acccgttgaa aagaacttac 600  
 ctgaaaaaaa cgaatatata ctagegttga atgttagcgt caacaacaag aagtttaatg 660  
 acgcgagggc caaggcaaaa agattccttg attacgtaag ggagttagaa tcattttgaa 720  
  
 taaaaaacac gctttttcag ttcgagtta tcattatcaa tactgccatt tcaaagaata 780  
 cgtaataat taatagtagt gattttccta actttattta gtcaaaaaat tagcctttta 840  
 attctgctgt aaccctgaca tgcccaaaaat agggggcggg ttacacagaa tatatacat 900  
 cgtaggtgtc tgggtgaaca gtttattcct ggcatccact aaatataatg gagcccgtt 960  
 tttaaactgg catccagaaa aaaaaagaat cccagcacca aaatattgtt ttcttcacca 1020  
 accatcagtt cataggtcca ttctcttagc gcaactacag agaacagggg cacaaacagg 1080  
 caaaaaacgg gcacaacctc aatggagtga tgcaacctgc ctggagtaaa tgatgacaca 1140  
  
 aggcaattga cccacgcatg tatctatctc attttcttac acctctatt accttctgct 1200  
 ctctctgatt tggaaaaagc tgaaaaaaa ggttgaacc agttccctga aattattccc 1260  
 ctacttgact aataagtata taaagacggt aggtattgat tgtaattctg taaatctatt 1320  
 tcttaaactt cttaaatct acttttatag ttagtctttt ttttagtttt aaaacaccaa 1380  
 gaacttagtt tcgacctccc gcgacctcca aaatcgaact accttcaca tggctgattt 1440  
 cgattctaaa gaatacttgg agttagttga caagtgggtg cgtgccacca actactgtc 1500  
 cgctggtatg atttcttga agtccaacc attattctct gttactaata cccaatcaa 1560  
  
 ggccgaagat gtcaaagtta aaccaattgg tctactgggt actatttccg gtcaaaactt 1620  
 cttatcgcc cacgctaacc gtttgattaa caagtacggt ctcaacatgt ttacgttg 1680  
 tggccaggt cacggtggtc aagtcaggt tactaacgcc tacttagacg gtgcctacac 1740  
 cgaagattac ccagaaatta ctcaagacat cgaaggtatg tctcatttgt tcaagcgttt 1800  
 ctctttccct ggtaggtattg gttccatat gaccgctcaa actccaggtt cttgacga 1860

aggtggtgaa ttgggttact ctttgcceca tgctttcggg gctgttttgg acaaccaga 1920  
 ccaagttgct ttgctgtcg ttggtgatgg tgaagctgaa actggtccat ctatggcctc 1980  
  
 ttggcattcc attaagttct taaatgceaa gaacgatgg gcegttttgc cagttttgga 2040  
 tttaaacggt ttcaagattt ccaatccaac cttttttct agaatgtctg atgaagaaat 2100  
 tactaagttc ttcaaggtt tgggttattc ccctagattc attgaaaatg atgacattca 2160  
 cgactacgcc acctaccacc aattggcgcg taacatctta gatcaagcca tgaagacat 2220  
 tcaagctatt caaatgacg ccagagagaa tggtaaatat caagatggtg aaattccagc 2280  
 ttggcctgtt attatcgcta gattgceaaa gggttggggg ggtccaacc acgatgcttc 2340  
 taataatcca attgaaaact ctttcagagc tcaccaagtt ccattacat tggacaaca 2400  
  
 cgatttggcc accttgcag aattcgaaga ttggatgaac tcttacaagc cagaagaatt 2460  
 attcaacgct gatggttctt tgaaggatga gttgaaagct attgceccaa agggtgataa 2520  
 gagaatgtct gtaacccaa tcaccaacgg tggctgctgac agatccgact tgaattgcc 2580  
 aaattggaga gaattcgcta acgacatcaa cgacgatacc agagtaagg aattcgctga 2640  
 ctctaagaga aacatggata tggctacttt atccaactat ttagtgccg tttctcaatt 2700  
 gaaccaacc agattcagat tcttcggtcc agatgaaacc atgtccaaca gattgtgggg 2760  
 tttgtttaat gttacccac gtcaatgat ggaagaaatc aaggaaccac aagatcaatt 2820  
  
 gttgtctcca actggtcgtt tcatcgattc ccaattgtct gaacaccaag ctgaagttg 2880  
 gttggaaggt tacacttga ctggtagagt tggatcttt gcctcttacg aatctttctt 2940  
 gagagtgtt gataccatgg tcaactaaca tttcaagtgg ttgcgtcacg cttccgaaca 3000  
 agcttggaga aatgactatc catcctaaa tttgatcgct acctctaccg cttccaaca 3060  
 agatcataac ggttatactc accaagacc ttggtatgta actcatttgg ccgagaagaa 3120  
 gtctaacttc attagagaat atttgccagc cgacggtaac tctttgttag ccgttcaaga 3180  
 gagagctttc tctgaaagac ataaggttaa ctattgatc gcttctaac aaccaagaca 3240  
  
 acaatggttc actgttgaag aagctgaagt cttagctaac gaagtttga agattatcga 3300  
 ttgggcttct actgctccat cttccgatgt tgatattact tttgcttctg ccggtactga 3360  
 accaaccatt gagactttgg ccgccttatg gttgattaat caagctttcc ctgacgttaa 3420  
 gtttagatac gttaacgttg ttgaattgtt aagattgcaa aagaaatctg aaccaaacat 3480  
 gaacgacgaa agagaattat ctgccgaaga atttaataag tacttccaag ccgacactcc 3540  
 agttatcttc ggtttccag cttacgaaaa ctfgattgaa tctttcttt tcgagagaaa 3600

gttcaccggt gatgtctatg ttcacggtta tagagaagat ggtgatatca ctaccaccta 3660

cgatatgaga gtctattccc acttggatcg tttccatcaa gccaaaggaag ccgccgaaat 3720

cttgtctgct aacggtaaaa tcgaccaagc cgtgcccgcac acctttattg ctaagatgga 3780

cgacactttg gccaaacact tccaagttaac tagaaatgaa ggtagagata ttgaagaatt 3840

cactgactgg acttggcttc cattgaagta agtgaattta ctttaaatct tgcatttaaa 3900

taaattttct ttttatagct ttatgactta gtttcaattt atatactatt ttaatgacat 3960

tttcgattca ttgattgaaa gctttgtgtt ttttcttgat gcgctattgc attgttcttg 4020

tctttttcgc cacatgtaat atctgtagta gatacctgat acattgtgga tgctgagtga 4080

aattttagtt aataatggag gcgctcttaa taattttggg gatattggct tatccccgcg 4140

tgcttggccg gccgtacact gagtaatggt agttataaga aagagaccga gttagggaca 4200

gttagaggcg gtggagatat tccttatggc atgtctggcg atgataaaac ttttcaaacg 4260

gcagccccga tctaaaagag ctgacaggga aatggtcaga aaaagaaacg tgcaccgcc 4320

cgcttggacg cggcgtcac ccgcacggca gagaccaatc agtaaaaatc aacggttaac 4380

gacattacta tatatataat ataggaagca tttaatagaa cagcatcgta atatatgtgt 4440

actttgcagt tatgacgcca gatggcagta gtggaagata ttcctttattg aaaaatagct 4500

tgtcacetta cgtacaatct tgatccggag cttttctttt tttgccgatt aagaattcgg 4560

tcgaaaaaag aaaaggagag ggccaagagg gagggcattg gtgactattg agcacgtgag 4620

tatacgtgat taagcacaca aaggcagctt ggagtatgtc tgttattaat ttcacaggt 4680

gttctggctc attgggtgaaa gtttccgctc tgcagagcac agaggccgca gaatgtgctc 4740

tagattccga tgctgacttg ctgggtatt 1a tatgtgtgcc caatagaaag agaacaattg 4800

accggttat tgcaaggaaa atttcaagtc ttgtaaaagc atataaaaat agttcaggca 4860

ctccgaaata cttggttggc gtgtttcgta atcaacctaa ggaggatgtt ttggctctgg 4920

tcaatgatta cggcattgat atcgtccaac tgcatggaga tgagtcgtgg caagaatacc 4980

aagagtctct cggtttgcca gttattaaaa gactcgtatt tccaaaagac tgcaacatac 5040

tactcagtgc agcttcacag aaacctcatt cgtttattcc cttgtttgat tcagaagcag 5100

gtgggacagg tgaacttttg gattggaact cgatttctga ctgggttgga aggcaagaga 5160

gccccgaaag cttacatfff atgttagctg gtggactgac gccagaaaat gttggtgatg 5220

cgcttagatt aaatggcgtt attggtgttg atgtaagcgg aggtgtggag acaaatggtg 5280

taaaagactc taacaaaata gcaaatctcg tcaaaaatgc taagaaatag gttattactg 5340

agtagtattt atttaagtat tgtttgtgca cttgcctgca ggccttttga aaagcaagca 5400  
 taaaagatct aaacataaaa tctgtaaaat aacaagatgt aaagataatg ctaaatcatt 5460  
 tggctttttg attgattgta caggaaaata tacatcgag ggggttgact tttaccattt 5520  
 caccgcaatg gaatcaaact tgttgaagag aatgttcaca ggcgcatacg ctacaatgac 5580  
 acggccggcc aagcacgagg ggataagcca atatcccaa aattattaag agcgcctcca 5640  
 ttattaacta aaatttcaact cagcatccac aatgtatcag gtatctacta cagatattac 5700  
 atgtggcgaa aaagacaaga acaatgcaat agcgcacaa gaaaaaacac aaagctttca 5760  
  
 atcaatgaat cgaaaatgac attaaaatag tatataaatt gaaactaagt cataaagcta 5820  
 taaaaagaaa atttatttaa atgcaagatt taaagtaaat tcacttactt caatggagac 5880  
 caagtccagt cagtgaattc ttcaatatct ctacctcat ttctagtaac ttggaagtgt 5940  
 ttggccaaag tctcgtccat cttagcaata aagggtcgg cagcggcttg gtcgatttta 6000  
 ccgtagcag acaagatttc ggcggcttcc ttggcttgat ggaaacgac caagtgggaa 6060  
 tagactctca tctcgtagg gtagtgata tcaccatctt ctctataacc gtgaacatag 6120  
 acatcaccgg tgaactttct ctcgaaaaag aaagattcaa tcaagtttcc gtaagcgtgg 6180  
  
 aaaccgaaga taactggagt gtcggcttgg aagtacttat taaattcttc ggcagataat 6240  
 tctctttcgt cgttcatgtt tgggtcagat ttcttttga atcttaaca ttcaacaacg 6300  
 ttaacgtatc taaactaac gtcagggaaa gcttgattaa tcaaccataa ggcggccaaa 6360  
 gtctcaatgg ttggttcagt accggcagaa gcaaaagtaa tatcaacatc ggaagatgga 6420  
 gcagtagaag cccaatcgat aatcttcaaa ccttcgttag ctaagacttc agcttcttca 6480  
 acagtgaacc attgttgtct tggttgttta gaagcgatca ataagttaac cttatgtctt 6540  
 tcagagaaag ctctctcttg aacggctaac aaagagttac cgtcggctgg caaatattct 6600  
  
 ctaatgaagt tagacttctt ctcgccaaa tgagttaaca taccagggtc ttggtgagta 6660  
 taaccgttat gatcttgttg gaaagcggta gaggtagcga tcaaatttaa ggatggatag 6720  
 tcatttctcc aagcttgttc ggaagcgtga cgcaaccact tgaatgttg agtgaccatg 6780  
 gtatcaaca ctctcaagaa agattcgtaa gaggcaaaga taccaactct accagtcaaa 6840  
 gtgtaacctt ccaaccaacc ttcagcttgg tttcagaca attgggaatc gatgatcga 6900  
 ccagttggag acaacaatg atcttgttgt tccttgattt cttccatcca ttgacgtggg 6960  
 gtaacattaa acaaacceca caatctgttg gacatggttt catctggacc gaagaatctg 7020  
  
 aatctggttg ggttcaattg agaaacggca cctaaatagt tggataaagt agccatatcc 7080  
 atgtttctct tagagtccgc gaattcctta cctctggtat cgtcgttgat gtcgttagcg 7140  
 aattctctcc aatttggcaa tttcaagtcg gatctgtcag caccaccgtt ggtgattggg 7200

ttagcagaca ttctcttacc accctttggg gcaatagctt tcaactcacc cttcaaggaa 7260  
 ccatcagcgt tgaataattc ttctggcttg taagagtcca tccaatcttc gaattctggc 7320  
 aaggtggcca aatcgtgttg ttccaatggt aatggaactt ggtgagctct gaaagagttt 7380  
 tcaattggat tattagaagc atcgtgggtt ggaccacccc aacctttgg caatctagcg 7440  
  
 ataataacag gccaaagctgg aatttcacca tcttgatatt taccattctc tctggcgtca 7500  
 ttttgaatag ctigaatgtc ttcatggtt tgatctaaga tgttagcggc caattggtgg 7560  
 taggtggcgt agtcgtgaat gtcacattt tcaatgaatc taggggaata acccaaact 7620  
 tcgaagaact tagtaatttc ttcatcagac attctagaaa aaatggttgg attggaatc 7680  
 ttgaaaccgt ttaaatccaa aactggcaaa acggcaccat cgttcttggc atttaagaac 7740  
 ttaatggaat gccaaagggc catagatgga ccagtttcag cttcaccatc accaacgaca 7800  
 gcaaaagcaa cttggtctgg gttgtccaaa acagcaccga aagcatggga caaagagtaa 7860  
  
 cccaattcac caccttcgtg caaggaacct ggagtttgag cggtcatatg ggaaccaata 7920  
 ccaccaggga aagagaaacg cttgaacaaa tgagacatac cttcagatgc ttgagtaatt 7980  
 tctgggtaat cttcgggtga ggcaccgtct aagtaggcgt tagtaacat gacttgacca 8040  
 ccgtgacctg gaccaccaac gtaaaacatg ttgagaccgt acttgtaat caaacggtta 8100  
 gcgtggcgt ataagaaagt ttgaccgga atagtagccc agtgaccaat tggtttaact 8160  
 ttgacatctt cggccttgat tggggtatta gtaacagaga ataatgggtt ggacttcaag 8220  
 aaaatcatac cagcggacaa gtagttggtg gcacgccacc acttgtcaac taactccaag 8280  
  
 tattctttag aatcgaaatc agccatttg aaggtagtcc gattttggag gtcgctggag 8340  
 gtcgaaacta agttcttggg gttttaaaac taaaaaaaag actaactata aaagtagaat 8400  
 ttaagaagtt taagaaatag atttacagaa ttacaatcaa tacctaccgt ctttatatac 8460  
 ttattagtca agtaggggaa taatttcagg gaactggttt caaccttttt tttcagcttt 8520  
 ttccaaatca gagagagcag aaggtaatag aaggtgtaag aaaatgagat agatacatgc 8580  
 gtgggtcaat tgccttgtgt catcatttac tccagcagg ttgcatcact ccattgaggt 8640  
 tgtgcccgtt tttgacctgt ttgtcccct gttctctgta gttgcgctaa gagaatggac 8700  
  
 ctatgaactg atggttggg aagaaaacaa tattttggg ctgggattct tttttttct 8760  
 ggatgccagc ttaaaaagcg ggctccatta tatttagtgg atgccaggaa taaactgttc 8820  
 accgagacac ctacgatgtt atatatctg tgtaaccgc cccctatttt gggcatgtac 8880  
 gggttacagc agaattaaaa ggctaatttt ttgactaaat aaagttagga aaatcactac 8940  
 tattaattat ttacgtatc tttgaaatgg cagtattgat aatgataaac tcgaactgaa 9000  
 aaagcgtgtt tttattcaa aatgattcta actcccttac gtaatcaagg aatctttttg 9060

ccttggcctc cgcgtcatta aacttcttgt tgttgacgct aacattcaac gctagtatat 9120  
  
 attcgttttt ttcaggtaag ttcttttcaa cgggtcttac tgatgaggca gtcgcgtctg 9180  
 aaaggtccgc cggcgttga cgagcgtgta ccaacctgca tttctttccg tcatatacac 9240  
 aaaatacttt catataaact tacttggctct tacgtcataa ataaatatgt atacataata 9300  
 attaaaaaat ttggttttat attttttaca aaagaatcgt ttacttcatt tctccctttt 9360  
 aagcgataca atccatgaaa aaagagaaaa agagagaaca ggcttgtgcc ttctttaaaa 9420  
 catcccacac aaaatcatat tgaattgaat ttacatctt aagctagtgt acaacaactg 9480  
 ctatatccaa agaaaactaa cgtggaccgc ttttagagtt gagaaaaagg ttgaaaaaaa 9540  
  
 atagcaatac aaagacttgt ttcatatata aaatacaggg agcacattga gctaataata 9600  
 cataaacact gcgaaccaat tccaatcaaa aggtacacat gagagcattc ccccagtagc 9660  
 tgccatttgc ccatcagaga tcatataata acatccttct tcgaacggcg gtttaaacgc 9720  
 gtggccgtgc cgtc 9734  
 <210> 32  
 <211> 7980  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic: i76221 integration construct  
 <400> 32  
 gacggcacgg ccacgcgttt aaaccgccgc acgtgtatgt acggctgtgt aatatgata 60  
  
 atcatctcgg acgaacggcg tagtactctc catcccctaa aaatgttcac gtgtgactgc 120  
 tccatttcgc cggatgtcga gatgaccccc cccctcaaa aggcactcac ctgctgacat 180  
 gccgtggcaa atgattgggg tcatcctttt tttctgttat ctctaagatc caaagaaaag 240  
 taaaaaaaa aggttggggg acgaattgcc gccgagcctc cgatgccatt attcaatggg 300  
 tattgcagtt ggggtatagt tctcgggtgg caaatagttc tcccttcatt ttgtatataa 360  
 actgggcggc tattctaagc atatttctcc cttaggttat ctggtagtac gttatatctt 420  
 gttcttatat ttctatcta taagcaaac caaacatatac aaaactacta gaaagacatt 480  
  
 gccccactgt gttcgtcgt ccaacgccgg cggaccttcc agacgcgact gcctcatcag 540  
 taagaccctg tgaagaagac ttacctgaaa aaaacgaata tatactagcg ttgaatgtta 600  
 gcgtcaacaa caagaagttt aatgacgcgg aggccaaggc aaaaagattc cttgattacg 660  
 taagggagtt agaatacatt tgaataaaaa acacgctttt tcagttcgag ttatcatta 720  
 tcaatactgc catttcaag aatacgtaaa taattaatag tagtgatttt cctaacttta 780

tttagtcaaa aaattagcct ttttaattctg ctgtaaccg tacatgccca aaataggggg 840  
 cgggttacac agaatatata acatcgtagg tgtctgggtg aacagtttat tcttggcatc 900  
  
 cactaaatat aatggagccc gctttttaag ctggcatcca gaaaaaaaa gaateccagc 960  
 accaaaatat tgttttcttc accaacctac agttcatagg tccattctct tagcgcaact 1020  
 acagagaaca ggggcacaaa caggcaaaaa acgggcacaaa cctcaatgga gtgatgcaac 1080  
 ctgcctggag taaatgatga cacaaggcaa ttgaccacag catgtatcta tetcattttc 1140  
 ttacaccttc tattaccttc tgctctctct gatttggaaa aagctgaaaa aaaaggttga 1200  
 aaccagtacc ctgaaattat tcccctactt gactaataag tatataaaga cggtaggtat 1260  
 tgattgtaat tctgtaaate tatttcttaa acttcttaaa ttctactttt atagttagtc 1320  
  
 ttttttttag ttttaaaaca ccaagaactt agtttcgacc tcccgcgacc tccaaaatcg 1380  
 aactaccttc acaatggaac attctgtaat cgaaccaact gtgcccacgc cgctaccagc 1440  
 catgtttgac gctccatctg gtatttttag ctctttggac gacgctgtgc aagcagccac 1500  
 cttagcccaa caacaactaa gttcagttga gttgcgtcag caagtaatca aagccataag 1560  
 agtggccgga gaaaggtatg cacaagtttt ggctgaaatg gcagttgctg aaactggtat 1620  
 gggtaggggt gtggataagt acattaagaa tgtctctcaa gctcgtcata cgcttggat 1680  
 agaatgttta tccgcccagg ttcttacggg tgataatggc ctaacattga ttgaaaatgc 1740  
  
 cccttgggga gtcgtagctt cagtcacgcc aagcacaaat ccagcagcta cgtaattaa 1800  
 taatgcaate tcaatgattg cagcggggaa ttcagtcgtg ttcgcaccac atccttctgc 1860  
 caaaaacgtc tcaactaagga ctatttcttt actcaacaag gccattgtcg ctaccggcgg 1920  
 cccagaaaat ttactagtta gtgtggcaaa ccctaacatc gaaactgcac agagattatt 1980  
 cagatatccg ggtattggat tgttagttgt gacaggtggt gaagccgtcg ttgaagccgc 2040  
 taggaagcat acagataaaa ggttaattgc agccggcgct ggtaatctc ctgtttgtgt 2100  
 ggacgaaact gctgacatac ctaaagccgc aagagcaatt gtcaagggtg cttctttcga 2160  
  
 caacaacata atttgtgctg atgaaaaagt tttgattgtg gtagacagag ttgcagatgc 2220  
 actattggca gaaatgcaaa gaaataacgc cgtcttactt acaccgaac agaccgaaag 2280  
 actactacce gctcttttgt ccgatattga cgaacagggc aaaggacgtg tgaatagaga 2340  
 ttatgttggga agagatgccc ctaaattagc agcggctatt ggtctggaag tttagcaaca 2400  
 tactcttcta ctctggcag agacagacgc tgatcatcca ttcgccgtga cggagctgat 2460  
 gatgccagtg ttaccagtaa taagagtaa gaatgtagat gatgcaatcg cattggcagt 2520

taagctagag tcaggctgca gacacacagc tgcgatgcac tctactaata taagaaactt 2580  
  
 aaatagaatg gctaagcca tcaatactc tatctttgta aaaaatggtc catgtattgc 2640  
 aggtttgggt ttaggcggtg aaggttggac ttcaatgact attagcactc cgaccggtga 2700  
 aggtgttaca agcgctcgta cctttgtcag attaagaagg tgtgtcttag tcgacatggt 2760  
 tcggattgct taagcggccg cgagtaataa ttattgcttc catataatat ttttatatac 2820  
 ctcttatttt tatgtattag ttaattaagt atttttatct atctgcttat cattttcttt 2880  
 tcatataggg ggggttgggt ttttcttgcc catcagattg atgtcctcca actcggcact 2940  
 attttacaaa ggggtttttt gtaagagaag gagaagacag atactaaacc atacgttact 3000  
  
 cgaaacaaaa aaaaaaaaaa tggaaaaagc tgctatcaac aaaagacggc ctcatcaaac 3060  
 ctaaagaaac catgtcagcg tatgtatata ccttgtaatt tacgtttcct taaatcttct 3120  
 ttctactaac gttttcatta ttctatactc tatgaccaat aaaaacagac tgtactttca 3180  
 aaatttacc agtaggccag caaataaaga aaattatacc agattacttc tgaacacat 3240  
 taatcccaac aacaagtatg ccattaatcc gtcgctacc catccccgcg tgcttggccg 3300  
 gccgtttctc gacgtgggcc ttttcttgc catatggatc cgctgcacgg tectgttccc 3360  
 tagcatgtac gtgagcgtat ttcttttaa accacgacgc tttgtcttca ttcaacgttt 3420  
  
 cccattgttt ttttctacta ttgctttgct gtgggaaaaa cttatcgaaa gatgacgact 3480  
 ttttcttaat tctcgtttta agagcttggg gagcgctagg agtcactgcc aggtatcggt 3540  
 tgaacacggc attagtcagg gaagtcataa cacagtcctt tcccgaatt ttcttttct 3600  
 attactcttg gcctcctcta gtacactcta tttttttta tgccctggta atgattttca 3660  
 ttttttttt tccacctagc ggatgactct tttttttct tagcgattgg cattatcaca 3720  
 taatgaatta tacattatat aaagtaatgt gatttcttcg aagaatatac taaaaatga 3780  
 gcaggcaaga taaacgaagg caaagatgac agagcagaaa gccctagtaa agcgtattac 3840  
  
 aaatgaaacc aagattcaga ttgcgatctc tttaaagggt ggtcccctag cgatagagca 3900  
 ctcgatcttc ccagaaaaag aggcagaagc agtagcagaa caggccacac aatcgcaagt 3960  
 gattaacgct cacacaggtg tagggtttct ggaccatag atacatgctc tggccaagca 4020  
 ttccggctgg tcgctaatcg ttgagtgcac tggtgactta cacatagacg accatcacac 4080  
 cactgaagac tgcgggatg ctctcggta agctttttaa gaggccttag gggccgtgcg 4140  
 tggagtaaaa aggtttggat caggatttgc gcctttggat gaggcacttt ccagagcggg 4200  
 ggtagatctt tcgaacagcg cgtacgcagt tgcgaactt ggtttgcaa gggagaaagt 4260

aggagatctc tcttgcgaga tgatcccgca ttttcttgaa agctttgcag aggctagcag 4320  
aattaccctc cacgttgatt gtctgcgagg caagaatgat catcaccgta gtgagagtgc 4380  
gttcaaggct cttgcggttg ccataagaga agccacctcg cccaatggta ccaacgatgt 4440  
tcctccacc aaaggtgttc ttatgtagtg acaccgatta tttaaagctg cagcatacga 4500  
tataatata tgtgtatata tgtataccta tgaatgtcag taagtatgta tacgaacagt 4560  
atgatactga agatgacaag gtaatgcac atctatcac tgtcattctg aacgaggcgc 4620  
gctttccttt tttctttttg ctttttcttt ttttttctct tgaactcgac ggccggccaa 4680

gcacgcgggg atggggtagc gacggattaa tggcatactt gttgttggga ttaatgtgtt 4740  
tcagaagtaa tctggtataa ttttctttat ttgctggcct actgggtaaa ttttgaaagt 4800  
acagtctgtt tttattggtc atagagtata gaataatgaa aacgttagta gaaagaagat 4860  
ttaaggaaac gtaaattaca aggtatatac atacgctgac atggtttctt taggtttgat 4920  
gaggccgtct tttgttgata gcagcttttt ccattttttt tttttttgtt tcgagtaacg 4980  
tatggtttag tatctgtctt ctcttctct tacaacaaaa ccttttgtaa aatagtccc 5040  
agttggagga catcaatctg atgggcaaga aaacaccaac cccccctata tgaaaagaaa 5100

atgataagca gatagataaa aataactaat taactaatac ataaaaataa gaggtatata 5160  
aaaatattat atggaagcaa taattattac tcgcgccgc ttaagcaatc cgaacatgt 5220  
cgactaagac acaccttctt aatctgacaa aggtacgagc gcttgaaca ccttaccgg 5280  
tcggagtgtc aatagtcatt gaagtccaac cttaccgcc taaacccaaa cctgcaatac 5340  
atggaccatt ttttacaag atagaggat tgatggcatt agccattcta ttttaagttc 5400  
ttatattagt agagtgcac gcagctgtgt gtctgcagcc tgactctagc ttaactgcca 5460  
atgcgattgc atcatctaca ttcttgactc ttattactgg taactggc atcatcagct 5520

ccgtcacgge gaatggatga tcagcgtctg tctctgccag gagtagacga gtatgttcgc 5580  
taacttcag accaatagcc gctgctaatt tagccgcatc tctccaaca taatctctat 5640  
tcacacgtcc tttgccctgt tcgtcaatat cggacaaaag agcgggtagt agtctttcgg 5700  
tctgttcggg tgtaagtaag acggcgttat ttctttgcat ttctgccaat agtgcactg 5760  
caactctgtc taccacaatc aaaacttttt catcagcaca aattatgttg ttgtcgaag 5820  
aagcacctt gacaattgct cttgcggctt taggtatgtc agcagtttcg tccacaaca 5880  
caggaggatt accagcgcg gctgcaattia accttttacc tgtatgcttc ctaggcgtt 5940

caacgacgge ttaccacct gtcacaacta acaatccaat acccgatct ctgaataatc 6000  
tctgtgcagt ttgatgtaa gggtttgcca cactaactag taaattttct gggccgccgg 6060  
tagcgacaat ggccttgttg agtaaagaaa tagtccttag tgagacgttt ttggcagaag 6120

gatgtggtgc gaacacgact gaattccccc ctgcaatcat tgagattgca ttattaatta 6180  
ccgtagctgc tggatttgtg cttggcgtga ctgaagctac gactccccaa ggggcatttt 6240  
caatcaatgt taggccatta tcacccgtaa gaacctcggc cgataaacat tctataaccag 6300  
gcgtatgacg agcttgagag acattcttaa tgtacttata caccacccta cccataaccag 6360  
  
tttcagcaac tgccatttca gccaaaactt gtgcataact ttctccggcc actcttatgg 6420  
ctttgattac ttgctgacgc aactcaactg aacttagttg ttgttgggct aaggtggctg 6480  
cttgcacagc gtcgtccaaa gagctaaaaa taccagatgg agcgtcaaac atggctggta 6540  
gcggcatggg cacagttggt tcgattacag aatgttccat tgtgaaggta gttcgatttt 6600  
ggaggtcgcg ggaggtcgaa actaagttct tgggttttta aaactaaaaa aaagactaac 6660  
tataaaagta gaatttaaga agtttaagaa atagatttac agaattaca tcaataccta 6720  
ccgtctttat atacttatta gtcaagtagg ggaataattt cagggaactg gtttcaacct 6780  
  
ttttttcag ctttttcaa atcagagaga gcagaaggta atagaagggtg taagaaaatg 6840  
agatagatac atgcgtgggt caattgcctt gtgtcatcat ttactccagg caggttgcac 6900  
cactccattg aggttgtgcc cgttttttgc ctgtttgtgc cctgtttctc tgtagtgcg 6960  
ctaagagaat ggacctatga actgatggtt ggtgaagaaa acaatatttt ggtgctggga 7020  
ttcttttttt ttctggatgc cagcttaaaa agcgggctcc attatattta gtggatgcca 7080  
ggaataaact gttcaccag acacctacga tgttatatat tctgtgtaac ccgccccta 7140  
ttttgggcat gtacgggta cagcagaatt aaaaggctaa ttttttgact aaataaagtt 7200  
  
aggaaaatca ctactattaa ttatttacgt attctttgaa atggcagtat tgataatgat 7260  
aaactcgaac tgaaaaagcg tgttttttat tcaaatgat tctaactccc ttacgtaate 7320  
aaggaatctt ttgacctgg cctccgcgtc attaaacttc ttgttgttga cgtaacatt 7380  
caacgetagt atatattcgt ttttttcagg taagttcttt tcaacgggtc ttactgatga 7440  
ggcagtcgcg tcigaaaggt ccgccggcgt tggacgagcg tgatatttc tttccttttt 7500  
atattgacga ctttttttt ttctgtgttt ttgtttctct tataaccgag ctgcttactt 7560  
attattattt caccttctct ttttattat acttataatt atttattctt tacatactgt 7620  
  
tacaagaaac tcttttctac attaattgca taaagtgtca atcagccat cctctatata 7680  
gctatcaaca acaaattga caaacctgcc tatatcttca ggaacaactg ccgcatcgct 7740  
accaccacta ctigtgaagt ccctggagtt taatatgcac tgaatttac ctacccgttt 7800  
tacacaagac cataatccat ccatgctatc gcagtatatg attttgtgtt cgtttttcgt 7860  
cttgcgaaag gcatcctcaa tggcttgttt cattgatcca tcagtgtggc tcgtaggtac 7920  
cagcaaaacc acttcatcag cggcgtactc ctggcggttt aaacgcgtgg ccgtgccgtc 7980

<210> 33

<211> 13266

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: i84022 integration construct

<400> 33

```

gacggcacgg ccacgcgttt aaaccgcaa gtgatgtaac taaatacacg attacatgg      60
aaattaacgt accttttttg tgcgtgtatt gaaatattat gacatattac agaaagggtt    120
cgcaagtccct gtttctatgc ctttctctta gtaattcacg aaataaacct atggtttacg    180
aatgatcca cgaaaatcat gttattattt acatcaacat atcgcgaaaa ttcattgtcat    240
gtccacatta acatcattgc agagcaacaa ttcattttca tagagaaatt tgctactatc    300
accactagt actaccattg gtacctacta ctttgaattg tactaccgct gggcgttatt    360

aggtgtgaaa ccacgaaaag ttcaccataa cttcgaataa agtcgcggaa aaaagtaaac    420
agctattgct actcaaatga gggttgcaga agcttgttga agcatgatga agcgttctaa    480
acgcactatt catcattaaa tatttaaagc tcataaaatt gtattcaatt cctattctaa    540
atggctttta tttctattac aactattage tctaaatcca tatectcata agcagcaatc    600
aattctatct ataactttaa cgctcgtcca acgccggcgg acctgatgtg tattactagt    660
gtcgacgaca gcattcgccc agtatTTTTT ttattctaca aaccttctat aatttcaaag    720
tatttacata attctgtatc agtttaatca ccataatac gttttctttg tttagtgcaa    780

ttaatttttc ctattgttac ttcgggcctt tttctgtttt atgagctatt ttttccgtca    840
tccttccgga tccagatttt cagcttcac cccagattgt gtctacgtaa tgcacgcat    900
cattttaaga gaggacctcc cgcgacctcc aaaatcgaac taccttcaca atgaaacttc    960
ttagtagtat tgagcaggcg tgtgacatct gtagattgaa gaaattgaaa ttagtaagg    1020
agaagcccaa atgtgcgaaa tgccttaaaa ataattggga atgcagatat agtccgaaga    1080
cgaagcgag tccccttacc cgcgcgcacc ttacggagggt cgagagtcgc cttgagcgcc    1140
ttgagcaact ttccttctt atcttcccca gagaggattt ggatatgac cttaatgatg    1200

acagtcttca agacattaag gcgcttctta cggggctttt cgtgcaggac aacgtcaaca    1260
aggacgcggt gacggaccgc cttgccagtg tcgaaaccga catgccctt acgcttcgcc    1320
aacaccgcat ttcgccacg agtagtagtg aggaatctc caataagggg cagcgccaac    1380
ttaccgtgag tatcgatagt gcggcccacc acgacaatag tacgatcccc cttgacttca    1440
tgccgcgca cgccttgcac gggttcgact ggagtgagga agacgatatg agtgacggtc    1500

```

ttccgtttct taagaccgat ccgaataaca acggtttttt cggatgatggg agtttgcttt 1560  
 gcatcttgag aagatcggg ttcaagcccg agaactatac caatagtaat gtcaatcgct 1620  
  
 tgcccacgat gatcaccgac cgctataacc ttgccagtcg cagtagcagc agtagacttt 1680  
 tgcagtccta ctigaacaac ttccatccgt attgtcccat tgtccatagt cccaccctta 1740  
 tgatgcttta caacaatcaa atcgagattg ccagtaaaga ccagtggcag attttgttca 1800  
 attgtattct tgcgatcggg gcgtgggtgca ttgaaggtga gagtaccgac attgacgtct 1860  
 tctattacca gaacccaag agtcacctta cctccaaagt gtttgaaagt gggagtatta 1920  
 tccttgctac ggcgcttac ttgcttagta gatacacgca atggcgccaa aagacgaaca 1980  
 cctectacaa cticcattcc ttcagtattc gcatggcgat tagtcttggg cttaacccg 2040  
  
 atttgccgag tagtttttcc gactcctcca tccttgagca ggcgagaaga atctggtgga 2100  
 gtgtgtagat ttgggaaatt cagcttagtc tttgtacgg gagaagtatt caattgagtc 2160  
 aaaacacgat tagtttttcc agtagtgtgg atgacgtcca aagaacgacg acggggccga 2220  
 cgattacca cggatattac gagacggcgc gcttgcttca ggtctttacg aagatttacg 2280  
 agcttgataa gacggtgacc gcggagaagt cccccattg cggaagaag tgtcttatga 2340  
 tctgcaacga aatcgaagaa gtcagtcgcc aagcgcgcaa attccttcag atggacatca 2400  
 gtacgacggc cttacgaac cttcttaaag agcatccctg gcttagtttc acgcgcttg 2460  
  
 agcttaaatg gaagcaactt agtttgattt tctacgtgct tcgcgacttc tttaccaact 2520  
 tcacgcaaaa gaaaagtcag cttgagcaag accagaacga ccaccagtcc tacgaggtca 2580  
 agagatgtag tattatgctt tccgacggg cgcagcgcac cgtcatgagt gtgtcctcct 2640  
 acatggataa ccacaacgtg acgccgtact tcgctggaa ctgcagttac tatcttttta 2700  
 acgcggtgct tgtgccgatt aaaacccttt tgagtaatag taagagtaac gccgaaaaca 2760  
 atgaaacggc gcagcttctt cagcagatca ataccgtcct tatgcttctt aagaagcttg 2820  
 cgaccttcaa gattcaaacc tgcgagaagt atatccaggt gcttgaggaa gtgtgcgcc 2880  
  
 cttccttct tagtcaatgc gcgattccgc tccccacat ttctacaat aactccaacg 2940  
 ggtcccgat caagaacatc gtggggagtg cgaccattgc gcagtatccc accttgcccg 3000  
 aagagaacgt gaataacatt tccgtcaagt acgtcagtc ccgtagtgtg ggtcccagtc 3060  
 ccgtcccgt taagagtggg gcgtcctttt ccgacctgt gaaacttctt agtaatagac 3120  
 cgccgagtag aaatagtcg gtcacgattc cgcgtccac gcccagtcac agaagtgtga 3180  
 cccccttct tggtcagcaa cagcaacttc agagtctgt cccgcttacg cccagtgcc 3240

ttttcggggg tgcgaacttc aaccagtccg gtaacatcgc cgactccagt cttagtttta 3300  
  
 cctttaccaa ttctccaat gggcccaatt tgattacgac ccagacgaac agtcaggcct 3360  
 tgagttagcc gatcgcgagt agtaatgtcc acgacaattt tatgaacaac gagattaccg 3420  
 cctccaagat cgacgacggg aacaacagta agccgcttag tcccgggtgg accgatcaga 3480  
 ccgcctacaa tgccttcggg attaccacgg gtatgttcaa cacgaccacg atggacgacg 3540  
 tgtacaatta cctttttgac gacgaggaca cgccgccgaa tccgaagaag gaatgagcca 3600  
 attggtgcgg caattgataa taacgaaaat gtcttttaat gatctgggta taatgaggaa 3660  
 ttttccgaac gtttttactt tatatatata tatacatgta acatatattc tatacgtat 3720  
  
 atcgagaaaa cgcatggtg gggtgacttt caactcggcg tatccccgcg tgcttgcccg 3780  
 gccgtagtta tgacaattac aacaacagaa ttctttctat atatgcacga acttgaata 3840  
 tggagaagat tatgacgtac aaactataaa gtaaatattt tacgtaacac atggtgctgt 3900  
 tgtgcttctt tttcaagaga ataccaatga cgtatgacta agtttatgta ttttcaaaa 3960  
 cctgtttage cctggcgaca gatacgtctc cggttcaac gatgaccctg gtgaccctgt 4020  
 caatgtcggc tccggtggca ccggccatga ttgcgatgtt ccttgcgtgt aaggtcatgt 4080  
 gtcccctttg gattccctcg gttgccaagg ccctaattgc ggccatattc tgagccaaac 4140  
  
 caacggcggc agtaacctgg gccaaactcag tagcggtttc gacctgcatt aaggccaaag 4200  
 cggccctagc tgiagggtga gtcttgggtg ctctctctac caaacccaag gccaaaggca 4260  
 attcaatggt accgaccaac ctaccgtcgt tggccaactc ccaccttctc aaagagggtg 4320  
 aatgtccggt cctggcggcg taggcgtggg ctccagcttc gatggcctc cagtcgttac 4380  
 ctgttgcgac gacgactggg tcaattccgt tcataattcc cttgttatgg gttgcggccc 4440  
 tgtaagggtc gactattgct aaggcgcagg ctcaacat tccccttgca acgtcggcac 4500  
 catcgtatcc ctgggtggtc aaagtctcag gggctaactc aacctggtc cttaccaacc 4560  
  
 tcaagtcggc caagttagac aaaatcctca acctgacggt tccaccagcg atcctctcta 4620  
 cctctggagc taacctttca gccatggtgt taactgtgtt ggcacccatg gcgtctctga 4680  
 catcaacaat caagtgcaat acgaccattg caccaacagg ggtgtcccta aaaacatgga 4740  
 cctcaatgtc tctgcaacca ccacctaac caacaaaac tggatctacg gcatctgctg 4800  
 ctccatgaa agcagcctta tgggccaaca acctttgcct agctccttct gggctctcta 4860  
 atccgacaac ttggatttgg gccctcatta aagggtgcagt tccgtgtgcg gtgaatccac 4920  
 cgttctctct agctatcctt gccatataag aggctgcggc aacaacagat ggttcctcga 4980

ctgccatagg tattaagtag tcccttccgt tgacggtgaa gttggtggcg acaccaatg 5040  
 gcaactcaaa ttttccgata acattctcga tcataccggt ggccaatgac aaaggcaaag 5100  
 caccgttacc ggccaatgca gaaatggctt caggttccaa tcttgcggct tcggcaaccc 5160  
 taactaacct ctgagcagga tccaagtccc tcattctctc gatccttgag ttcaatccgt 5220  
 cgatgtgacc tgtctttcca gtcattgtaa agttagtgg ttgcgcgact tcgggtgggg 5280  
 taagtataga ggtatattaa caatTTTTTg ttgatacttt tatgacattt gaataagaag 5340  
 taatacaaac cgaaaatggt gaaagtattt gttaaagtgg ttatgcagct tttgcattta 5400  
  
 tatactctgtt aatagatcaa aaatcatcgc ttcgctgatt aattaccca gaaataaggc 5460  
 taaaaaacta atcgattat tatcctatgg ttgtaattt gattcgttga tttgaaggtt 5520  
 tgtggggcca ggttactgcc aatTTTTcct cttcataacc ataaaagcta gtatttaga 5580  
 atctttattg ttcggagcag tgcggcgcga ggcacatctg cgtttcagga acgcgaccgg 5640  
 tgaagaccag gacgcacgga ggagagtctt ccgtcggagg gctgtcgccc gctcggcggc 5700  
 ttctaaccg tacttcaata tagcaatgag cagttaagcg tattactgaa agttccaaag 5760  
 agaaggtttt tttaggctaa gataatgggg ctctttacat ttccacaaca tataagtaag 5820  
  
 attagatatg gatatgtata tgggtgtatt gccatgtaat atgattatta aacttctttg 5880  
 cgtccatcca aaaaaaagt aacgcacgca cactcccagc agacaactag cttgataatg 5940  
 tctcagaacg tttacattgt atcgactgcc agaaccceaa ttggttcatt ccagggttct 6000  
 ctatctcca agacagcagt ggaattgggt gctgttgctt taaaaggcgc cttggctaag 6060  
 gttccagaat tggatgcatc caaggatttt gacgaaatta tttttgtaa cgttctttct 6120  
 gccaatTTTg gccaaagctcc ggccagacaa gttgctttgg ctgccggttt gagtaatcat 6180  
 atcgttgcaa gcacagttaa caaggtctgt gcatccgcta tgaaggcaat cattttgggt 6240  
  
 gctcaatcca tcaaatgtgg taatgctgat gttgtcgtag ctggtggttg tgaatctatg 6300  
 actaacgcac calactacat gccagcagcc cgtgcgggtg ccaaatttgg ccaaactgtt 6360  
 cttgttgatg gtgtcgaag agatgggttg aacgatgcgt acgatggtct agccatgggt 6420  
 gtacacgcag aaaagtgtgc ccgtgattgg gatattacta gagaacaaca agacaatttt 6480  
 gccatcgaat cctacaaaa atctcaaaaa tctcaaaagg aaggtaaatt cgacaatgaa 6540  
 attgtacctg ttaccattaa gggatttaga ggtaagcctg atactcaagt cacgaaggac 6600  
 gaggaacctg ctagattaca cgttgaaaaa ttgagatctg caaggactgt tttcaaaaa 6660  
  
 gaaaacggta ctgttactgc cgctaacgct tctccaatca acgatggtgc tgcagccgtc 6720  
 atcttgggtt ccgaaaaagt tttgaaggaa aagaatttga agcctttggc tattatcaaa 6780  
 ggttgggggtg aggccgctca tcaaccagct gattttacat gggctccatc tcttgagtt 6840

ccaaaggctt tgaaacatgc tggcatcgaa gacatcaatt ctggtgatta ctttgaattc 6900  
 aatgaagcct ttcggttgt cggtttgggtg aacactaaga ttttgaagct agacccatct 6960  
 aaggttaatg tatatggtgg tgctgttgcct ctaggtcacc cattgggttg ttctggtgct 7020  
 agagtggttg ttacactgct atccatctta cagcaagaag gaggtaagat cgggtgtgcc 7080  
  
 gccatttga atggtgggtg tgggtccttc tctattgtca ttgaaaagat atgattacgt 7140  
 tctgcgattt tctcatgac ttttcataa aatacataa tatataaatg getttatgta 7200  
 taacaggcat aatttaaagt tttatttgcg attcatcgtt tttcaggtac tcaaacgctg 7260  
 aggtgtgcct tttgacttac tttccgctc tggcaagctg gccgaacctg caggcccgca 7320  
 gcgccgatac gaaaatcgtt attgcttga aggtgaaatt tctactctta ttaatggtga 7380  
 acgttaagct gatgctatga tggaaactga ttggtcttaa cttgcttgtc atcttgctaa 7440  
 tggtcattgg ctcgtgttat tacttaagtt attgtactc gttttgaacg taatgctaat 7500  
  
 gatcatctta tggaaataa gtgagtgggt tcagggtcca taaagctttt caattcatct 7560  
 ttttttttt tgttctttt tttgattccg gtttcttga aattttttg attcggtaat 7620  
 ctccgagcag aaggaagaac gaaggaagga gcacagactt agattggtat atatacgc 7680  
 atgtggtgtt gaagaaacat gaaattgcc agtattctta acccaactgc acagaacaaa 7740  
 aacctgcagg aaacgaagat aatcatgtc gaaagctaca tataaggaac gtgctgctac 7800  
 tcatcctagt ccgttgcct ccaagctatt taatatcatg cacgaaaagc aaacaaactt 7860  
 gtgtgcttca ttggatgttc gtaccacaa ggaattactg gagttagttg aagcattagg 7920  
  
 tcccaaaatt tgttactaa aaacacatgt ggatatcttg actgattttt ccatggaggg 7980  
 cacagttaag ccgctaaggt cattatccgc caagtacaat ttttactct tcgaagacag 8040  
 aaaatttgc t gacattgta atacagtcaa attgcagtac tctgcgggtg tatacagaat 8100  
 agcagaatgg gcagacatta cgaatgcaca cgggtgtggtg ggcccaggta ttgtagcgg 8160  
 tttgaagcag gcggcggaag aagtaacaaa ggaacctaga ggccttttga tgttagcaga 8220  
 attgtcatgc aagggtccc tagctactgg agaataact aagggtactg ttgacattgc 8280  
 gaagagtgac aaagattttg ttatcgctt tattgtcaa agagacatgg gtggaagaga 8340  
  
 tgaaggttac gattggttga ttatgacacc cgggtgtgggt ttagatgaca agggagacgc 8400  
 attgggtcaa cagtatagaa ccgtggatga tgtggtctct acaggatctg acattattat 8460  
 tgttgaaga ggactatttg caaaggaag gtagtctaa gtagagggtg aacgttacag 8520  
 aaaagcagc tgggaagcat atttgagaag atgcccagc caaaactaaa aaactgtatt 8580  
 ataagtaaat gcatgtatac taaactcaca aattagagct tcaatttaat tatatcagtt 8640  
 attaccacga aaatcgttat tgtcttgaag gtgaaatttc tactcttatt aatggtgaac 8700

gttaaagctga tgctatgatg gaagctgatt ggtcttaact tgcttgtcat cttgctaattg 8760  
  
 gtcataatggc tcgtgttatt acttaagtta tttgtactcg ttttgaacgt aatgctaattg 8820  
 atcatcttat ggaataatag tgaacggccg gccaagcacg cggggattga atgagaaaaa 8880  
 aaatcggttg ggcttaactt taaagaaaaa agttgagatt agatttattg tgttataaat 8940  
 atagatatac aattctttat aaaaaaata tatatatata tcattgttat taaataaaga 9000  
 gttttcctag tatatagatt aaaaaactac tctattaat gagagctaaa aaaagcaggc 9060  
 tgccaaaaaa ataaagcatt tatgaagggg gttcagcaag atgcaatcga tgggggaaga 9120  
 ttatTTTTTta acatcgtaag atcttctaaa tttgtcatcg atgttggca agtagtaaac 9180  
  
 accactttgc aaatgctcaa tggaaccttg aggtttgaag ttcttcttca aatgggcatt 9240  
 ttctctcaat tcgatggcag cttcgtaatc ctttggagtt tcggtgattc tcttgctaa 9300  
 tttgttagta atatctaatt ctttgataat atgttggacg tcaccaacaa ttttgcaaga 9360  
 atatagatat gcagctaac cggaaccgta agaaaataaa ccaacacgct tccttgtaa 9420  
 gtcgtcagat ccaacatagt ttaatagaga tgcaaaggcg gcataaacag atgcggtgta 9480  
 catgttacct gtgtttgtg gaacaatcaa agattgggca actctctctt tgtggaatgg 9540  
 cttagcaaca ttaacaaaag tttttcaat gttcttatcg gttaaagatt cgtcataatc 9600  
  
 gcgagtagct aattcggcgt caacttctgg gaacaattga ggattggctc tgaatcggt 9660  
 atatagtaat ctaccgatg attttgtgac caatttacag gttggaacat ggaaaacgtt 9720  
 gtagtcgaaa tatttcaaaa cgttcaaac atccgaacca gcgggatcgc taaccaacce 9780  
 tttagaata gccttcttgg aataactctt gtaacttga tcaagagcct tgacgtaaca 9840  
 agttaatgaa aaatgacat cgacgtaagg atattcgtg gtgaaatctg gcttgtaaaa 9900  
 atcgtaggcg tgttccatg aagaagctct tacagagtca aatacaattg gagcatcagg 9960  
 accgatccac atagcaacag taccggcacc accggttggc cttgcggcac cettatcgta 10020  
  
 gatggcaata tcaccgcaa ctacaatggc gtctctacca tcccatgcgt tagattcaat 10080  
 ccagttcaaa gagttgaaca acgcgttgg accaccgtaa caggcattaa gcgtgtcaat 10140  
 accttcgacg tcagtgtttt caccaacaa ttgcatcaag acagacttga cagacttggc 10200  
 cttgtcaatc agagtctcag taccgacttc taatctacca attttgttgg tgtcgtggt 10260  
 gtaactctt atcaacttag acaaaacagt tagggacatc gagtagatat ctctctgtc 10320  
 attgacaaaa gacatgttgg tttggcccag accaattgtg tatttacctt gagaaacgcc 10380  
 atcaaatctc tctagctcag attggttgac acattgagtt gggatgtaaa tttggatacc 10440

tttaataccg acatTTTgag gtctggTTTT ttgttcagcg gtctTTTgtt tTTTtagttc 10500  
 agtcattTgc aagTTTgtat tgtgtaattg ttgttgcttt tgcggcctaa gtcttctttt 10560  
 aataccacac caacaaagtT tagttgagag tttcattgtg aaggtagttc gattttggag 10620  
 gtgcggggag gtiactTTTT ttttgatgg acgcaaagaa gTTTaaataat catattacat 10680  
 ggcaatacca ccatatacat atccatatct aatcttactt atatgtttgtg gaaatgtaaa 10740  
 gagccccatt atcttagcct aaaaaaacct tctctttgga actttcagta atacgcttaa 10800  
 ctgctcattg ctatattgaa gtacggattA gaagccgccg agcgggcgac agccctccga 10860  
  
 cggaagactc tcttccgtgc gtcttggTct tcaccggTcg cgttctgaa acgcagatgt 10920  
 gcctcgcgcc gcactgtctc gaacaataaa gattctacaa tactagcttt tatggttatg 10980  
 aagaggaaaa atTggcagta acctggcccc acaaaccttc aatcaacga atcaaatTaa 11040  
 caacatagg ataataatgc gattagTTTT ttagccttat tTctgggta attaatcagc 11100  
 gaagcgatga tTTTgatct attaacagat atataaatgc aaaagctgca taaccacttt 11160  
 aactaatact ttcaacattt tcggtttgta ttacttctta ttcaaatgtc ataaaagtat 11220  
 caacaaaaaa ttgttaatat acctctatac ttacccacc cgaagtcgcg caaccaacta 11280  
  
 actttacaat gactggaaag acaggtcaca tcgacggatt gaactcaagg atcgagaaga 11340  
 tgagggactt ggatcctgct cagaggttag ttagggtTgc cgaagccgca ggattggaac 11400  
 ctgaagccat tTctgcattg gccgtaacg gtgctttgcc tttgtcattg gccaacgta 11460  
 tgatcgagaa tgttatcgga aaatttgagt tgccattggg tgtcggcacc aacttcaccg 11520  
 tcaacggaag ggactactta atacctatgg cagtcgagga accatctgtt gttgccgcag 11580  
 cctcatatat ggcaaggata gctagagaga acggtggatt caccgcacac ggaactgcac 11640  
 ctTtaatgag ggcccaaatc caagtTgtcg gattaggaga cccagaagga gctaggcaaa 11700  
  
 ggtTgtTgce ccataaggct gctttcatgg aagcagcaga tgccgtagat ccagttttgg 11760  
 ttggttagg tggTggtTgc agagacattg aggtccatgt ttttagggac acccctgttg 11820  
 gtgcaatggt cgtattgcac ttgattgttg atgtcagaga cgccatgggt gccaacacag 11880  
 ttaacacat ggctgaaagg ttagctccag aggtagagag gatcgtgtgt ggaaccgtca 11940  
 ggttgaggat tttgtctaac ttggccgact tgaggttggT aagagccagg gttgagttag 12000  
 cccctgagac tttgaccacc cagggatacg atggtgccga cgttgcaagg ggaatggttg 12060  
 aagcctgcgc cttagcaata gtcgacctt acagggccgc aaccataac aagggaatta 12120  
  
 tgaacggaat tgaccagtc gtcgtcgcaa caggtaacga ctggagggcc atcgaagctg 12180  
 gagcccacgc ctacgccgc aggaccggac attacacctc tttgacaagg tgggagttgg 12240  
 ccaacgacgg taggttgTgc ggtaccattg aattgccttt ggccttgggt ttgtaggag 12300

gagccaccaa gactcacctt acagctaggg ccgctttggc cttaatgcag gtcgaaaccg 12360  
 ctactgagtt ggcccaggtt actgccgccc ttggtttggc tcagaatatg gccgcaatta 12420  
 gggccttggc aaccgagggg atccaaaggg gacacatgac cttacacgca aggaacatcg 12480  
 caatcatggc cggtgccacc ggagccgaca ttgacaggtt caccagggtc atcgttgaag 12540  
  
 ccggagacgt atctgtgcc agggctaac aggttttggg aaatacataa acttagtcat 12600  
 acgtcattgg tattctcttg aaaaagaagc acaacagcac catgtgttac gtaaaatatt 12660  
 tactttatag ttgtacgtc ataatttctt ccatattaca agttcgtgca tatatagaaa 12720  
 gaattctgtt gttgtaattg tcataactag gtcgccggc gttggacgag cgaatgtgta 12780  
 tattagttaa aaaagttgta tgtaataaaa gtaaaattta atattttgga tgaaaaaac 12840  
 ctttttaga ctttttcta actagaatgc tggagtagaa atacccatc tcaagataca 12900  
 aaaagcgtta ccggcactga tttgtttcaa ccagtatata gattattatt gggctttgat 12960  
  
 caactttcct cagacatc agtaacagtt atcaagctaa atatttacgc gaaagaaaa 13020  
 caaatatatt aattgtgata cttgtgaatt ttattttatt aaggatacaa agttaagaga 13080  
 aaacaaaatt tatatacaat ataagtaata ttcatatata tgtgatgaat gcagtcttaa 13140  
 cgagaagaca tggccttggg gacaactctc ttcaaaccac cttcagcctt tctcaattca 13200  
 tcagcagatg ggtcttcgat ttgcaaagca gccaaagcgg cggtttaaac gcgtggccgt 13260  
 gccgtc 13266  
  
 <210> 34  
 <211> 13964  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220><223> Synthetic: i84026 integration construct  
 <400> 34  
 gacggcacgg ccacgcgttt aaaccgcaa gtgatgtaac taaatacacg attaccatgg 60  
 aaattaacgt accttttttg tgcgtgtatt gaaatattat gacatattac agaaagggtt 120  
 cgcaagtccct gtttctatgc ctttcttta gtaattcacg aaataaacct atggtttacg 180  
 aatgatcca cgaaaatcat gttattattt acatcaacat atcgcgaaaa ttcattgtcat 240  
 gtccacatta acatcattgc agagcaacaa ttcattttca tagagaaatt tgctactatc 300  
 acccactagt actaccattg gtacctacta ctttgaattg tactaccgct gggcgttatt 360  
  
 aggtgtgaaa ccacgaaaag ttcaccataa cttcgataaa agtcgaggaa aaaagtaaac 420  
 agctattgct actcaaatga ggtttgcaga agcttgttga agcatgatga agcgttctaa 480

acgcactatt catcattaa tatttaaagc tcataaaatt gtattcaatt cctattctaa 540  
 atggctttta tttctattac aactattagc tctaaatcca tctctcata agcagcaatc 600  
 aattctatct atactttaaa cgctcgtcca acgccggcgg acctgatgtg tattactagt 660  
 gtcgacgaca gcattcgcce agtatTTTT ttattctaca aaccttctat aatttcaaag 720  
 tatttacata attctgtatc agtttaatca ccataatata gttttctttg tttagtcaa 780  
  
 ttaatttttc ctattgttac ttcgggcctt tttctgtttt atgagctatt tttccgtea 840  
 tccttccgga tccagatttt cagcttcatc tccagattgt gtctacgtaa tgcacgccat 900  
 cattttaaga gaggacctcc cgcgacctcc aaaatcgaac taccttcaca atgaaacttc 960  
 ttagtagtat tgagcaggcg tgtgacatct gtagattgaa gaaattgaaa ttagtaagg 1020  
 agaagcccaa atgtgcgaaa tgccttaaaa ataattggga atgcagatat agtccgaaga 1080  
 cgaagcgcag tccccttacc cgcgcgcacc ttacggaggt cgagagtcgc cttgagcgcc 1140  
 ttgagcaact tttccttctt atcttcccca gagaggattt ggatatgata cttaatagg 1200  
  
 acagtcttca agacattaag gcgcttctta cggggctttt cgtgcaggac aacgtcaaca 1260  
 aggacgcggt gacgaccgc cttgccagtg tcgaaaccga catgcccctt acgcttcgcc 1320  
 aacaccgcat ttcgccacg agtagtagtg aggaatctc caataagggg cagcgccaac 1380  
 ttaccgtgag tatcgatagt gcggcccacc acgacaatag tacgatcccc cttgacttca 1440  
 tgccgcgca cgccttgca gggttcgact ggagtgagga agacgatatg agtgacggtc 1500  
 ttcctttct taagaccgat ccgaataaca acggtttttt cggtagtggg agtttgcttt 1560  
 gcatcttgag aagtatcggg ttcaagcccg agaactatac caatagtaat gtcaatcgt 1620  
  
 tgcccacgat gatcaccgac cgctataccc ttgccagtgc cagiacgacg agtagacttt 1680  
 tgcagtctta ctgaaaca ttccatccgt attgtcccat tgtccatagt cccaccctta 1740  
 tgatgcttta caacaatcaa atcgagattg ccagtaaaga ccagtggcag attttgttca 1800  
 attgtattct tgcgatcggg gcgtggtgca ttgaaggatg gatlaccgac attgactct 1860  
 tctattacca gaacccaag agtcacctta cctccaaagt gtttgaaagt gggagtatta 1920  
 tccttgtcac ggcgcttca ttgcttagta gatcacgca atggcgcaa aagacgaaca 1980  
 cctcctaaa ctccattcc ttcagtattc gcatggcgat tagtcttggg cttaacgcg 2040  
  
 atttgccgag tagtttttcc gactctcca tccttgagca gcgcagaaga atctggtgga 2100  
 gtgtgtatag ttgggaaatt cagcttagtc tttgtacgg gagaagtatt caattgagtc 2160  
 aaaacacgat tagttttccc agtagtgtg atgacgtcca aagaacgacg acggggccga 2220  
 cgattacca cggattatc gagacggcgc gcttgcttca ggtctttacg aagatttacg 2280  
 agcttgataa gacggtgacc gcggagaagt cccccattg cgcaagaag tgtcttatga 2340

tctgcaacga aatcgaagaa gtcagtcgcc aagcgccgaa attccttcag atggacatca 2400  
gtacgacggc ctttacgaac cttcttaaag agcatccttg gcttagtttc acgcgctttg 2460

agcttaaatg gaagcaactt agtttgatta tctacgtgct tgcgacttc tttaccaact 2520  
tcacgcaaaa gaaaagtcag cttgagcaag accagaacga ccaccagtcc tacgaggtca 2580  
agagatgtag tattatgctt tccgacgcgg cgcagcgcac cgtcatgagt gtgtcctcct 2640  
acatggataa ccacaacgtg acgcccgtact tgcggtggaa ctgcagttac tatcttttta 2700  
acgcggtgct tgtgccgatt aaaacccttt tgagtaatag taagagtaac gccgaaaaca 2760  
atgaaacggc gcagcttctt cagcagatca ataccgtcct tatgcttctt aagaagcttg 2820  
cgaccttcaa gattcaaacc tgcgagaagt atatccaggt gcttgaggaa gtgtgcgccc 2880

ccttccttct tagtcaatgc gcgattccgc ttccccacat ttcttacaat aactccaacg 2940  
ggtcgcat caagaacatc gtggggagtg cgaccattgc gcagtatccc accttgcccg 3000  
aagagaacgt gaataacatt tccgtcaagt acgtcagtcc cggtagtggt ggtcccagtc 3060  
cgtcccgt taagagtggg gcgtcctttt cgcacctgt gaaacttctt agtaatagac 3120  
cgccgagtag aaatagtccg gtcacgattc cgcgctccac gcccagtcac agaagtgtga 3180  
cccccttctt tggtcagcaa cagcaacttc agagtcttgt cccgcttacg cccagtgccc 3240  
ttttcggggg tgcgaacttc aaccagtcg gtaacatcgc cgactccagt cttagtttta 3300

cctttaccaa ttctccaat gggccaatt tgattacgac ccagacgaac agtcaggcct 3360  
tgagtacgcc gatcgcgagt agtaatgtcc acgacaattt tatgaacaac gagattaccg 3420  
cctccaagat cgacgacggg aacaacagta agccgcttag tcccgggtgg accgatcaga 3480  
ccgcctaaa tgccttcggg attaccacgg gtatgttcaa cacgaccacg atggacgacg 3540  
tgtacaatta ctttttgac gacgaggaca cgcgccgaa tccgaagaag gaatgagcca 3600  
attggtgagg caattgataa taacgaaaat gtcttttaat gatctgggta taatgaggaa 3660  
ttttccgaac gtttttactt tataatata tatacatgta acatatattc tatacgtat 3720

atcgagaaaa cgcgatggtg gggtgacttt caactcggcg tatccccgcg tgcttgccg 3780  
gccgtccgca tgactcaaga gaagcatgig gtttttgagt ttttttcgtt gaattttcag 3840  
gtaaagctca atagttatga caattacaac aacagaattc tttctatata tgcacgaact 3900  
tgtaatatgg aagaattat gacgtacaaa ctataaagta aatatattac gtaacacatg 3960  
gtgctgttgt gcttctttt caagagaata ccaatgacgt atgactaagt ttaggattta 4020  
atgcaggtga cggaccatc tttcaaacga tttatatcag tggcgtcaa attgttaggt 4080

tttgttgggt cagcaggttt cctgttgtgg gtcataatgac tttgaaccaa atggccggct 4140  
  
 gctagggcag cacataagga taattcacct gccaaagacgg cacaggcaac tattcttgct 4200  
 aattgacgtg cgttggfacc aggagcggta gcatgtgggc ctcttacacc taataagtec 4260  
 aacatggcac ctigtggttc tagaacagta ccaccaccga tggfacctac ttcgatggat 4320  
 ggcatggata cggaaattct caaatcaccg tccacttctt tcatcaatgt tatacagttg 4380  
 gaactttcga ctttttggc aggatcttgt cctaatgcca agaaaacagc tgtcactaaa 4440  
 ttagctgcat gtgcgttaaa tccaccaaca gaccagcca ttgcagatcc aaccaaattc 4500  
 ttagcaatgt tcaactcaac caatgcggaa acatcacttt ttaacacttt tctgacaaca 4560  
  
 tcaccaggaa tagtagcttc tgcgacgaca ctcttaccac gaccttcgat ccagttgatg 4620  
 gcagctgggt tttgtcggg acagtagtta ccagaaacgg agacaacctc catatcttcc 4680  
 cagccatact ctctaccat ttgctttaat gattattcga cacccttaga aatcatattc 4740  
 ataccattg cgtcaccagt agttgttcta aatctcatga agagfaaatc tctgctaga 4800  
 caagtttgaa tatgttgacg acgtgcaaat cttgatgtag agtfaaaagc ttttttaatt 4860  
 gcgttttgc cctctctga gtctaaccat atcttacagg caccagatct tttcaaagtt 4920  
 gggaaacgga ctactgggcc tcttgcata ccaccttag tfaaaacagt tgttgacca 4980  
  
 ccgccagcat tgattgcctt acagccacgc atggcagaag ctaccaaaaca accctctgta 5040  
 gttgccattg gtatatgata agatgtacca tcgataacca aggggcctat aacaccaacg 5100  
 ggcaaaggca tgfaacctat aacattttca caacaagcgc caaatacgcg gtcgtagtca 5160  
 taatttttat atggfaaacg atcagatgct aatcacaggag cttctgcca aattgaaaga 5220  
 gccttctac gtaccgcaac cgctctcgta gtatcaccta atttttctc caaagcgtac 5280  
 aaagftaact taccgtgaat aaccaaggca gcgacctctt tgttcttcaa ttgttttgta 5340  
 tttccactac ttaataatgc ttctaattct tctaaaggac gtattttctt atccaagctt 5400  
  
 tcaatatcgc gggaaatcctc ttctcacta gatgatgaag gtcctgatga gctcgattgc 5460  
 gcagatgata aacttttgac tttcgatcca gaaatgactg ttttattgggt taaaactggf 5520  
 glagaagcct ttgtacagg agcagfaaaa gacttcttgg tgacttcagt cttaccaat 5580  
 tggctcgcag ccattgfaaa gttagttggf tgcgcgactt cgggtggggf aagtatagag 5640  
 gtatattaac aattttttg gtatactttt atgacattg aataagaagt aatacaaac 5700  
 gaaaatgttg aaagtattag ttaaagttgt tatgcagctt ttgcatttat atatctgtta 5760  
 atagatcaaa aatcatcgct tcgctgatta attacccag aataaggct aaaaaactaa 5820

tcgcattatt atcctatggt tgttaatttg attcgttgat ttgaaggttt gtggggccag 5880  
 gttactgcca attttctcct tcataacca taaaagctag tattgtagaa tctttattgt 5940  
 tcggagcagt gcggcgcgag gcacatctgc gtttcaggaa cgcgaccggt gaagaccagg 6000  
 acgcacggag gagagtcttc cgtcggaggg ctgtcgccc ctcggcggct tctaaccgt 6060  
 acttcaatat agcaatgagc agttaagcgt attactgaaa gttccaaaga gaaggttttt 6120  
 ttaggctaag ataattgggc tctttacatt tccacaacat ataagtaaga ttagatatgg 6180  
 atatgtatat ggigtattg ccatgtaata tgattattaa acttctttgc gtccatccaa 6240  
  
 aaaaaaagta acgcacgcac actcccgaca gacaactagc ttgataatgt ctcagaacgt 6300  
 ttacattgta tcgactgcca gaacccaat tggttcattc cagggttctc tatccitcaa 6360  
 gacagcagtg gaattgggtg ctgttgcttt aaaaggcgcc ttggctaagg ttccagaatt 6420  
 ggatgcatcc aaggattttg acgaaattat ttttggtaac gttctttctg ccaatttggg 6480  
 ccaagctccg gccagacaag ttgctttggc tgccggtttg agtaatcata tcgttgcaag 6540  
 cacagttaac aaggtctgtg catccgctat gaaggcaatc attttgggtg ctcaatccat 6600  
 caaatgtggt aatgctgatg ttgtcgtagc tgggtggtgt gaatctatga ctaacgcacc 6660  
  
 atactacatg ccagcagccc gtgcgggtgc caaatttggc caaactgttc ttgttgatgg 6720  
 tgtcgaaga gatgggttga acgatgcgta cgatggtcta gccatgggtg tacacgcaga 6780  
 aaagtgtgcc cgtgattggg atattactag agaacaacaa gacaattttg ccatcgaatc 6840  
 ctacaaaaa tctcaaaaat ctcaaagga aggtaaattc gacaatgaaa ttgtacctgt 6900  
 taccattaag ggatttagag gtaagcctga tactcaagtc acgaaggacg aggaacctgc 6960  
 tagattacac gttgaaaaat tgagatctgc aaggactgtt ttccaaaaag aaaacggtac 7020  
 tgttactgcc gctaacgctt ctccaatcaa cgatggtgct gcagccgtca tcttggtttc 7080  
  
 cgaaaaagt tgaaggaaga agaattttaa gcctttggct attatcaaag gttgggtgga 7140  
 ggccgctcat caaccagctg attttacctg ggctccatct cttgcagttc caaaggcttt 7200  
 gaaacatgct ggcacgaag acatcaattc tgttgattac tttgaattca atgaagcctt 7260  
 ttcggttgtc ggtttggtga aactaagat tttgaagcta gaccatcta aggttaatgt 7320  
 atatggtggt gctgttgctc taggtcacc attgggttgt tctggtgcta gagtgttgt 7380  
 tacactgcta tccatcttac agcaagaagg aggttaagatc ggtgttgccg ccatttghtaa 7440  
 tgggtggtgt ggtgcttct ctattgtcat tgaaaagata tgattacgtt ctgcgatttt 7500  
  
 ctcatgatct tttcataaaa atacataaat atataaatgg ctttatgtat aacaggcata 7560  
 atttaaagt ttatttgcga ttcatcgttt ttcaggtaact caaacgtga ggtgtgcctt 7620  
 ttgacttact tttccgctt ggcaagctgg ccgaacctgc aggcccgag cggcgatcag 7680

aaaatcgta ttgtcttgaa ggtgaaattt ctactcttat taatggtgaa cgtaagctg 7740  
 atgctatgat ggaagctgat tggctttaac ttgcttgca tcttgctaag ggtcattggc 7800  
 tcgtgttatt acttaagtta tttgtactcg ttttgaacgt aatgctaag atcatcttat 7860  
 ggaataatag tgagtggttt caggggccat aaagcttttc aattcatctt ttttttttt 7920  
  
 gttctttttt ttgattccgg tttctttgaa atttttttga ttcggtaatc tccgagcaga 7980  
 aggaagaacg aaggaaggag cacagactta gattggtata tatacgcata tgtgggtgtg 8040  
 aagaacatg aaattgccca gtattcttaa cccaactgca cagaacaaaa acctgcagga 8100  
 aacgaagata aatcatgtcg aaagctacat ataaggaacg tgctgctact catcctagtc 8160  
 ctgttgctgc caagctattt aatatcatgc acgaaaagca aacaaacttg tgtgcttcat 8220  
 tggatgttcg taccaccaag gaattactgg agttagtga agcattaggt cccaaaattt 8280  
 gtttactaaa aacacatgtg gatattctga ctgatttttc catggagggc acagttaagc 8340  
  
 cgctaaaggc attatccgcc aagtacaatt ttttactctt cgaagacaga aaatttgctg 8400  
 acattggtaa tacagtcaaa ttgcagtact ctgcgggtgt atacagaata gcagaatggg 8460  
 cagacattac gaatgcacac ggtgtgggtgg gccaggtat tgttagcggg ttgaagcagg 8520  
 cggcgaaga aglaacaaag gaacctagag gccttttgat gtttagcagaa ttgtcatgca 8580  
 agggctccct agctactgga gaatatacta agggactgtg tgacattgcg aagagtgaca 8640  
 aagattttgt tatcggtttt attgctcaaa gagacatggg tggaaagat gaaggttacg 8700  
 attggttgat tatgacacc ggtgtgggtt tagatgacaa gggagacgca ttgggtcaac 8760  
  
 agtatagaac cgtggatgat gtggctctta caggatctga cattattatt gttggaagag 8820  
 gactatttgc aaaggaagg gatgctaagg tagagggtga acgttacaga aaagcaggct 8880  
 gggaagcata tttgagaaga tgcggccagc aaaactaaaa aactgtatta taagtaaatg 8940  
 catgtatact aaactcaaa attagagctt caatttaatt atatcagtta ttaccacgaa 9000  
 aatcgttatt gtcttgaagg tgaaatttct actcttatta atggtgaacg ttaagctgat 9060  
 gctatgatgg aagctgattg gtcttaactt gcttgcacg ttgctaagtg tcatatggct 9120  
 cgtgttatta ctttaagttat ttgtactcgt tttgaacgta atgctaagga tcatcttatg 9180  
  
 gaataatagt gaacggccgg ccaagcacgc ggggattgaa tgagaaaaaa aatcggttgg 9240  
 gcttaacttt aaagaaaaaa gttgagatta gatttattgt gttataaata tagatataca 9300  
 attctttata aaaaaatat atatatatat cattgttatt aaataaagag ttttcctagt 9360  
 atatagatta aaaaactact ctattaaatg agagctaaaa aaagcaggct gccaaaaaaa 9420  
 taaagcattt atgaaggggg ttcagcaaga tgcaatcgat gggggaagat tattttttaa 9480  
 catcgtaaaga tcttctaata ttgtcatcga tgttgggtcaa gtagtaaca ccactttgca 9540

aatgctcaat ggaaccttga ggtttgaagt tcttcttcaa atgggcattt tctctcaatt 9600  
  
cgatggcagc ttcgtaatcc tttggagttt cggtgattct cttggctaata ttgttagtaa 9660  
tatctaatte cttgataata tgttggacgt caccaacaat tttgcaagaa tatagagatg 9720  
cagctaaacc ggaaccgtaa gaaaataaac caacacgctt gccttgtaag tcgtcagatc 9780  
caacatagtt taatagagat gcaaaggcgg cataaacaga tgcggtgtac atgttacctg 9840  
tgtttgttg aacaatcaaa gattgggcaa ctctctcttt gtggaatggc ttagcaacat 9900  
taacaaaagt ttttcaatg ttcttatcgg ttaaagattc gtcataatcg cgagtagcta 9960  
attcggcgtc aacttctggg aacaattgag gattggctct gaaatcgtta tatagtaatc 10020  
  
taccgtatga ttttgtgacc aattfacagg ttggaacatg gaaaacgttg tagtcgaaat 10080  
atttcaaac gttcaaagca tccgaaccag cgggatcgtt aaccaaccct ttagaaatag 10140  
ccttcttggg ataactcttg taaacttgat caagagcctt gacgtaacaa gttaatgaaa 10200  
aatgaccatc gacgtaagga tattcgtctg tgaactctgg cttgtaaaaa tcgtaggcgt 10260  
gttccatgta agaagctctt acagagtcaa atacaattgg agcatcagga ccgatccaca 10320  
tagcaacagt accggcacca ccggttggtc ttgcggcacc cttatcgtag atggcaatat 10380  
caccgcaaac tacaatggcg tctctacat cccatgcgtt agattcaatc cagttcaaag 10440  
  
agttgaacaa cgcgttggta ccaccgtaac aggcatlaag cgtgtcaata ccttcgacgt 10500  
cagtgttttc accaaacaat tgcatcaaga cagacttgac agacttggac ttgtcaatca 10560  
gagtttcagt accgacttct aatctacaa ttttgttggg gtcgatgttg taactcttga 10620  
tcaacttaga caaacagtt agggacatcg agtagatc tctctgtca ttgacaaaag 10680  
acatgttggg ttggcccaga ccaattgtgt atttaccttg agaaacgcca tcaaatttct 10740  
ctagctcaga ttggttgaca cattgagttg ggatgtaaat ttggatacct ttaataccga 10800  
cattttgagg tctggttttt tgttcagegg tcttttgttt ttttagttca gtcatttgea 10860  
  
agtttgtatt gtgtaattgt tgttgctttt gcggcctaag tcttcttta ataccacacc 10920  
aacaagttt agttgagagt ttcatgtga aggtagtctg attttggagg tcgctgggagg 10980  
ttactttttt ttggatgga cgcaagaag ttaataatc atattacatg gcaataccac 11040  
catafacata tccatalcta atcttactta tafgttgg aaatgtaaag agccccatta 11100  
tcttagccta aaaaaacctt ctctttggaa ctctcagtaa tacgcttaac tgctcattgc 11160  
tatattgaag tacggattag aagccgccga gcgggcgaca gccctccgac ggaagactct 11220  
cctccgtgcg tcttggctt caccggctgc gttcctgaaa cgcagatgtg cctcgcgccg 11280

cactgctccg aacaataaag attctacaat actagctttt atggttatga agaggaaaa 11340  
 ttggcagtaa cctggcccca caaaccttca aatcaacgaa tcaaattaac aacctagga 11400  
 taataatgcg attagttttt tagccttatt tctggggtaa ttaatcagcg aagcgatgat 11460  
 ttttgatcta ttaacagata tataaatgca aaagctgcat aaccacttta actaactt 11520  
 tcaacatfff cggtttggat tacttcttat tcaaatgtca taaaagtatc aacaaaaaat 11580  
 tgттаатата cctctatact taceccacc gaagtgcgcg aaccaactaa ctttacaatg 11640  
 gctgcagacc aattggtgaa gactgaagtc accaagaagt cttttactgc tcctgtacaa 11700  
  
 aaggcttcta caccagtttt aaccaataaa acagtcattt ctggatcgaa agtcaaaagt 11760  
 ttatcatctg cgcaatcgag ctcatcagga ccttcatcat ctagtgagga agatgattcc 11820  
 cgcgatattg aaagcttggg taagaaaata cgtcctttag aagaattaga agcattatta 11880  
 agtagtgga atacaaaaca attgaagaac aaagaggtcg ctgccttggg tattcacggt 11940  
 aagttacctt tgiacgcttt ggagaaaaaa ttaggtgata ctacgagagc ggttgcggta 12000  
 cgtaggaagg ctctttcaat tttggcagaa gctcctgtat tagcatctga tcgттtacca 12060  
 tataaaaatt atgactacga ccgcgtattt ggcgcttggт gtgaaaatgt tataggttac 12120  
  
 atgcctttgc ccgttgggtg tataggcccc ttggttatcg atggtacatc ttatcatata 12180  
 ccaatggcaa ctacagaggg ttgtttggta gcttctgcca tgcgtggctg taaggcaatc 12240  
 aatgctggcg gtggtgcaac aactgtttta actaaggatg gtatgacaag aggcccagta 12300  
 gtccgtttcc caactttgaa aagatctggт gcctgtaaga tatggttaga ctcagaagag 12360  
 ggacaaaacg caattaaaa agcttttaac tctacatcaa gatttgcacg tctgcaacat 12420  
 attcaaaact gtctagcagg agatttactic ttcatgagat ttagaacaac tactggtgac 12480  
 gcaatgggta tgaatatgat ttctaagggt gtcgaatact cattaaagca aatggtagaa 12540  
  
 gagtatggct gggaagatat ggaggttgtc tccgtttctg gtaactactg taccgacaaa 12600  
 aaaccagctg ccatcaactg gatcgaaggt cgtggtaaga gtgtcgtcgc agaagctact 12660  
 attcctgggt atgttgtcag aaaagtgtta aaaagtgatg tttccgcatt ggttgagttg 12720  
 aacattgcta agaatttggт tggatctgca atggctgggt ctgttgggtg atttaacgca 12780  
 catgcagcta atttagtgac agctgttttc ttggcattag gacaagatcc tgcacaaaat 12840  
 gtcgaaagt ccaactgtat aacattgatg aaagaagtgg acggtgattt gagaatttcc 12900  
 gtatccatgc catccatcga agtaggtacc atcggtgggt gtactgttct agaaccacaa 12960  
  
 ggtgccatgt tggacttatt aggtgtaaga ggcccacatg ctaccgtcc tggtagaac 13020  
 gcacgtcaat tagcaagaat agttgcctgt gccgtcttgg caggtgaatt atccttatgt 13080  
 gctgccctag cagccggcca tttggttcaa agtcatatga cccacaacag gaaacctgct 13140

gaaccaacaa aacctaacaa tttggacgcc actgatataa atcgtttgaa agatgggtcc 13200  
 gtcacctgca ttaaatccta aacttagtca tacgtcattg gtattctctt gaaaaagaag 13260  
 cacaacagca ccatgtgfta cgtaaaatat ttactttata gtttgtagt cataatttct 13320  
 tccatattac aagttcgtgc atatatagaa agaattctgt tgttgtaatt gtcataacta 13380  
  
 ttgagcttta cctgaaaatt caacgaaaa aactcaaaaa ccacatgctt ctcttgagtc 13440  
 atgaggaggt ccgccggcgt tggacgagcg aatgtgtata ttagtttaaa aagttgtatg 13500  
 taataaaagt aaaatftaat attttggatg aaaaaacca ttttagact tttcttaac 13560  
 tagaatgctg gagtagaagt acgccatctc aagatacaaa aagcgttacc ggcactgatt 13620  
 tgtttcaacc agtatataga ttattattgg gtcttgatca acttctctca gacatatcag 13680  
 taacagttat caagctaat atttacgca aagaaaaaca aatattftaa ttgtgatact 13740  
 tgtgaatftt atftftftaa ggatacaaaag ttaagagaaa acaaaatfta tatacaatat 13800  
  
 aagtaatatt catatatatg tgatgaatgc agtcttaacg agaagacatg gccttggatga 13860  
 caactctctt caaaccaact tcagcctttc tcaattcctc agcagatggg tcttcgattt 13920  
 gcaaagcagc caaagcggcg gtttaaacgc gtggccgtgc cgtc 13964  
  
 <210> 35  
 <211> 13963  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic: i85207 integration construct  
 <400> 35  
  
 gacggcacgg ccacgcgttt aaaccgccag ggcaaggttg gcctctactt actccatcga 60  
 caattcaaga tacagaacct cctccagatg gaatcccttc catagagaga aggagcaagc 120  
  
 aactgacca atattgactg ccaactggacc tgaagacatg caacaaagtg caagcatagt 180  
 ggggccttct tccaatgcta atccggteac tgccactgct gctacggaaa accaacctaa 240  
 aggtattaac ttcttacta taagaaaatc acacgagcgc ccggacgatg tctctgttta 300  
 aatggcgcaa gttttccgct ttgtaataa tatttatacc ctttcttct ctcccctgca 360  
 atataatagt ttaattctaa tattaataat atcctatatt ttcttcattt accggcgcac 420  
 tctcgcccga acgacctcaa aatgctgct acattcataa taacaaaag ctcataactt 480  
 ttttttttga acctgaatat atatacatca catgtcactg ctggctcttg ccgaccagcg 540  
  
 tatacaatct cgatagttgg tttcccgctt tttccaactcc cgtccgctcg tecaacgceg 600  
 gcggaccttc acatgtaggg accgaattgt ttacaagttc tctgtaccac catggagaca 660

tcaaagattg aaaaatctatg gaaagatatg gacggtagca acaagaatat agcacgagcc 720  
gcaaggttca tttcgttact tttgatatcg ctcaacta ttgcgaagcg cttcagttaa 780  
aaaatcataa ggaaaagtg taaatattat tggtagtatt cgtttggtaa agtagagggg 840  
gtaatttttc ccctttatit tgttcataca ttcttaaatt gctttgcctc tccttttgg 900  
aagctatact tcggagcact gttgagcgaa ggctcattag atatatattc tgtcattttc 960

cttaacccaa aaataaggga aagggtccaa aaagcgtcgc gacaactgtt gaccgtgatc 1020  
cgaaggactg gctatacagt gttcacaaaa tagccaagct gaaaataatg ttagctatg 1080  
ttcagttagt ttggctagca aagatataaa agcaggtcgg aatatattat gggcattatt 1140  
atgcagagca tcaacatgat aaaaaaacct cccgcgacct ccaaaatcga actaccttca 1200  
caatgactgc cgacaacaat agtatgccc atggtgcagt atctagttac gccaaattag 1260  
tgcaaaaacca aacacctgaa gacattttgg aagagtttcc tgaaattatt ccattacaac 1320  
aaagacctaa tacccgatct agtgagacgt caaatgacga aagcggagaa acatgttttt 1380

ctggatcatga ttaggagcaa attaagttaa tgaatgaaaa ttgtattgtt ttggattggg 1440  
acgataatgc tattggtgcc ggtaccaaga aagtttgc tttaatggaa aatattgaaa 1500  
agggtttact acatcgtgca ttctcgtct ttattttcaa tgaacaaggt gaattacttt 1560  
tacaacaaag agccactgaa aaaataactt tccctgatct ttggactaac acatgctgct 1620  
ctcatccact atgtattgat gacgaattag gtttgaaggg taagctagac gataagatta 1680  
agggcctat tactgcggcg gtgagaaaac tagatcatga attaggtatt ccagaagatg 1740  
aaactaagac aaggggtaag tttcactttt taaacagaat ccattacatg gcaccaagca 1800

atgaacctg ggtgaacat gaaattgatt acatcctatt ttataagatc aacgctaaag 1860  
aaaacttgac tgtaaccca aacgtcaatg aagttagaga cttcaaatgg gttcaccaa 1920  
atgatttgaa aactatgttt gctgacccaa gttacaagtt tacgccttgg tttagatta 1980  
tttgcgagaa ttacttattc aactggtggg agcaattaga tgacctttct gaagtggaaa 2040  
atgacaggca aattcataga atgctataac aacgcgtcaa taatatagc tacataaaaa 2100  
tcataataac tttgttatca tagcaaatg tgatataaaa cgtttcattt cacctgaaaa 2160  
atagtaaaaa taggcgacaa aaatccttag taatatgtaa actttatit ctttattat 2220

ttacagaact ctgaatatac attgattgtt cacatttttt ttttctctc tcaatttccc 2280  
ttgattatat tcaaaaggtt attggcctct tgaatgttc cactgaatc cccgcgtgct 2340  
tggccggccg tggagcgacc tcatgctata cctgagaaag caacctgacc tacagaaaag 2400  
agttactcaa gaataagaat tttcgtttta aaacctaaaga gtcactttta aatttgtata 2460  
cacttatttt tttataact tatttaataa taaaaatcat aatcataag aaattcgctc 2520

aaacgacat tggatggaca aagaaggact tcatgtaaga tttcatgtca cttcggcgt 2580  
 gagtgaacc atcgtaaca gagtataaaa cttcacacat tctagccaag ttgatagctg 2640  
  
 gcattaacaa agggaatgga acggcggttg gtctcaaaga ttctctgtta ataacctec 2700  
 aggcgtcttc gacttttcta gagatgtatt cacaggcttc ttcttcagaa gcaccggatt 2760  
 ccttagaata acattcgatg gaggaggcaa catgacctct ttcttggtct tctttatgag 2820  
 agacaatate atccatcaat ctaatgataa cacaagaagc ttcaacaata ggtgggtagg 2880  
 aagaaacca tttaaaagt tctctgttaa caatgtcacc tctaccaacg taagatctag 2940  
 cagtgatcaa accgtaggta cggtaacca tggaaacaga catgtactct tccaaagtag 3000  
 gcatgtaacc ttctttcaac catctggcct caaccaagta gtttctgacc aattccttag 3060  
  
 ccatttctt aacgtagigg atttgataag ccttaccttc cttttctaaa gattcttcca 3120  
 ttccaacgtg caagtaacc aattcttggg agatcaactt catgtattct ggcaacatgt 3180  
 ccaaacaaga aatggaccac ttctcaacgg cttgagttaa aatttccaat tcttcgtagg 3240  
 taccgtagtt gtcgaaggta tcatcaaaa cgaccaacca catacaagac ttcatacaaga 3300  
 acattctggt tctggcatgt tgtggttcat agtaaataga caaaatccag aagtaacctt 3360  
 cgacaactct atcacgaacg aatggcaatt tgttttgcaa gtctaaatct ttccaccact 3420  
 tgcagatgtg agacaattct ttcttatgca tggattgcaa aacagagaaa tctaacttag 3480  
  
 ccaacttcaa caaacctcg tctgagaag tttcttgttg gtaattggc ataatgtgta 3540  
 aagcttcgat tctggccaat cttcttctca atggttgctt caaggcttgg tggatttggg 3600  
 ttcttaagga agagtcacaa gatggatcct tggcaataat gtccaagtga accttagaga 3660  
 attccaaagc gttgtccaag atggtttcat cttcgactct catgaaagca gcttcgtaca 3720  
 aggccaagat accttgagcg tcttacaca aagattcctt aaatttacct tttctgtcca 3780  
 taaagtcctt gaaaacacca gaggagacat tgaaccttg ttgacgcaac aaacgaaacc 3840  
 acaaggagat agattgtaaa ttttcttat cgaccattg ttcaccgtaa gtgacatgga 3900  
  
 tatgttgtaa agcttcttcg atttcttctt caaatggta agcaatacct aaacgttgaa 3960  
 cagcattgat taattcgatc aacttaacat gttgcatagg ttcgttagaa cccttaatag 4020  
 taatcaatc cttcttaact tctctctta actcttcgac taattgcttc ttcataacca 4080  
 agtctcttgg ttcacgttaa gtcaaaaatt gatcacecca aatggaagcg ttgaagttag 4140  
 cggtatgtct aataacgtct ggcttggtag aatccttacc atcgacaaca attggggaag 4200  
 tagatggaga ggaagaaaca gaggaatag gcaaagtgga cattgtaag ttagttggtt 4260

gcgcgacttc gggtagggta agtatagagg tatattaaca attttttgtt gatactttta 4320

tgacatttga ataagaagta atacaaccg aaaatgttga aagtattagt taaagtggtt 4380

atgcagcttt tgcatttata tatctgttaa tagatcaaaa atcatcgctt cgctgattaa 4440

ttaccccaga aataaggcta aaaaactaat cgcattatta tcctatggtt gttaatttga 4500

ttcgttgatt tgaaggtttg tggggccagg ttactgccaa tttttcctct tcataacat 4560

aaaagctagt attgtagaat ctttattgtt cggagcagtg cggcgcgagg cacatctgcg 4620

tttcaggaac gcgaccggtg aagaccagga cgcacggagg agagtcctcc gtcggagggc 4680

tgtcgcccgc tcggcggctt ctaatccgta cttcaatata gcaatgagca gttaagcgt 4740

ttactgaaag ttccaaagag aaggttttt taggctaaga taatggggct ctttacattt 4800

ccacaacata taagtaagat tagatatgga tatgtatatg gtggtattgc catgtaatat 4860

gattattaaa cttcttttgc tccatccaaa aaaaaagtaa cgcacgcaca ctcccagacag 4920

acaactagct tgataatggc ttcagaaaaa gaaattagga gagagagatt cttgaacgtt 4980

ttccctaaat tagtagagga attgaacgca tcgcttttgg cttacggtat gcctaaggaa 5040

gcatgtgact ggtatgccca ctcatgtaac tacaacactc caggcggtaa gttaaataga 5100

ggtttgtccg ttgtggacac gtatgctatt ctctcaaca agaccgttga acaattgggg 5160

caagaagaat acgaaaaggt tgctattcta ggttgggtgca ttgagttgtt gcaggcttac 5220

ttcttggctg ccgatgatat gatggacaag tccattacca gaagaggcca accatgttgg 5280

tacaagttc ctgaagttgg ggaaattgcc atcaatgacg cattcatgtt agaggctgct 5340

atctacaagc ttttgaatc tcacttcaga aacgaaaaat actacataga tatcaccgaa 5400

ttgttccatg aagtcacctt ccaaaccgaa ttgggccaat tgatggactt aatcactgca 5460

cctgaagaca aagtcgactt gagtaagttc tcctaaaga agcactcctt catagttact 5520

ttcaagactg cttactatc tttctacttg cctgtcgcac tggetatgta cgttgccggt 5580

atcacagatg aaaaggattt gaaacaagcc agagatgtct tgattccatt gggatgaatat 5640

ttccaaatc aagatgacta cttagactgc ttcggtacce cagaacagat cggttaagatc 5700

ggtacagata tccaagataa caaatgttct tgggtaatca acaaggcatt agaacttgct 5760

tccgcagaac aaagaaagac tttagacgaa aattacggta agaaggactc agtcgcagaa 5820

gccaaatgca aaaagatttt caatgacttg aaaatcgacc agttatacca cgaatatgaa 5880

gagtctgttg ccaaggattt gaaggccaag atctccaag tcgacgagtc tcgtggcttc 5940

aaagccgacg tcttaactgc gtttttgaac aaggtttaca agagaagtaa atagaactaa 6000

cgctaacga taaaacatta gatttcagat tagataagga ccatgtataa gaaatatata 6060  
 cttccactat aatatagat aagcttacag atagtatctc tcgatctacc gttccacgtg 6120  
 actagtccaa gaacctgcag gccgcgagcg ccgatacгаа aatcgttatt gtcttgaagg 6180  
 tgaaatttct aciccttatta atggigaacg ttaagctgat gctatgatgg aagctgattg 6240  
 gtcttaactt gcttgtcatc ttgctaattg tcattggctc gtgttattac ttaagttatt 6300  
 tgtactcgtt ttgaacgtaa tgctaattg catcttatgg aataatagtg agtggtttca 6360  
 gggtcataa agcttttcaa ttcactttt tttttttgt tcttttttt gattccggtt 6420  
  
 tctttgaaat tttttgatt cggtaatctc cgagcagaag gaagaacгаа ggaaggagca 6480  
 cagacttaga ttggtatata tacgcataatg tgggtgttaa gaaacatгаа attgccagat 6540  
 attcttaacc caactgcaca gaacaaaaac ctgcaggaaa cgaagataaa tcatgtcгаа 6600  
 agctacatat aaggaacgtg ctgctactca tcctagtcct gttgctgcaa agctatttaa 6660  
 tatcatgcac gaaaagcaaa caaacttgig tgcttcattg gatgttcgta ccaccaagga 6720  
 attactggag ttagttgaag cattaggtcc caaaattgt ttactaaaa cacatgtgga 6780  
 tatcttgact gatttttcca tggagggcac agttaagccg ctaaaggcat tatccgcaa 6840  
  
 gtacaatfff ttactcttcg aagacagaaa atttgctgac attgtaata cagtcaaatt 6900  
 gcagtactct gcgggtgat acagaatagc agaatgggca gacattacga atgcacacgg 6960  
 tgtggggggc ccaggtattg ttagecggttt gaagcaggcg gcggaagaag taacaaagga 7020  
 acctagaggc cttttgatg tagcagaatt gtcatgcaag ggctccctag ctactggaga 7080  
 atatactaag ggtactgttg acattgcгаа gagtgacaaa gatfffgtta tcggctttat 7140  
 tgctcaaaga gacatgggtg gaagagatga aggttacgat tggttgatta tgacacccgg 7200  
 tgtgggttta gatgacaagg gagacgat tgggtcaacag tatagaaccg tggatgatgt 7260  
  
 ggtctctaca ggatctgaca ttattattgt tggaaagga ctatffgcaa agggaagggga 7320  
 tgctaaggta gagggtgaac gttacagaaa agcaggctgg gaagcatatt tgagaagatg 7380  
 cggccagcaa aactaaaaaa ctgtattata agtaaagca tgtatactaa actcacaat 7440  
 tagagcttca atttaattat atcagttatt accacgaaaa tcgttattgt cttgaaggtg 7500  
 aaatffctac tcttattaat ggtgaacgtt aagctgatgc tatgatgгаа gctgattggt 7560  
 cttacttgc ttgtcatctt gctaattggtc atatggctcg tgttattact taagttattt 7620  
 gtactcgttt tgaacgtaat gctaattgat atcttatgga ataatagtg atcggcgctc 7680  
  
 gcggcctgca ggtffctca tcctagtatg tatagcttgt accattaaa cgaatfftat 7740  
 catgccgccg aaaggaacaa tffcaagtac tatcggaaga tgaatggtta gatgttaagc 7800  
 gcggtcactt caaactcac atffataaag atgtcacatg gaccattat atctacctta 7860

agttatttat caagataagt ttccggatct ttttctttcc taacacccca gtcagcctga 7920  
 gttacatcca gccattgaac cttagaaaat cttttgtcat cagcggtttg agccctaaga 7980  
 tcaacatctt gcttagcaat cactgcaatg gcgtcataac caccagcacc aggtattaag 8040  
 caagtaagaa ctctttttaa ggtctggcaa tcatccaata agctagtttg tacgggaggt 8100  
  
 tcgatatcgg caccagattc tttagttatt tttctaaagg aacgtctaata tgtggcaact 8160  
 gcatctctaa ctctgtgat ctcaggatac ttttgacagg tacagtcatt cctctcaaga 8220  
 gactcaaata tctgatcgt gtaatcgtca tgagtctcgt gtaagcgatc tagtttagat 8280  
 agtccatcca taaatctaga atttgcatga tcgagttctg tatatatttt caagctttcc 8340  
 ggcatatgcg aatcatacca atttttacc tctggacca gttttactgt ttctgaacca 8400  
 ttcttaatat cgcccatcca taaagttaat cccgaaggta aatggttact tttaatcgtt 8460  
 atattccagt ctcttctatt aaccaaagtc gccagtttac tgccgtaagt agcacttcca 8520  
  
 atatctggca aattagagat taatgcgggt gggaatcttc tatactgat agatccatat 8580  
 gctgccgccc ctacatcaaa cccgcttcca atttaccct gagcttgaca atgagcaact 8640  
 tgtgataaat tatgaataac ttctctatat ttgtctacat tattttccag gtccgataca 8700  
 aaaaaggagg ccaaagctgt agttaaact gtgactaac ctgccgagga gccagccct 8760  
 gttttgggaa ctcttcaat tctgtgcgaa tgaaaactca atcttctgtt gccacgatgt 8820  
 tcggtaacgc tgcctcctg agaatggtag gcacatcag agaaaatac aataacgaac 8880  
 aagtttctat tgcagtagtc gtccatgta ggcttaaagt agctaatac gttagcgata 8940  
  
 actttttcaa tgaagggtt cttagatccg cctatcgaaa caggaatgaa gccagtttta 9000  
 ggacttatat ggiacagcca ctccccatct ttaaattgtt tacttttcac acgacttca 9060  
 aacttatcag actcttgcaa tgaaccgtaa ggatgggcta cagcatgcat tcttgccgat 9120  
 aatccgacta caaatgcttc atatttcgga tctaaaacta aatatccacc agctagtaac 9180  
 gctttccctg gggcactgaa ggctctcaac tctgacatta tcaagctagt tgtctgtcgg 9240  
 gagtgtgcgt gcgttttttt atcatgttga tgctctgcat aataatgcc ataatattt 9300  
 ccgacctgct tttatatctt tgctagccaa actaactgaa catagctaca cattattttc 9360  
  
 agcttggcta tttgtgaac actgtatagc cagtccttcg gatcacggtc aacagttgtc 9420  
 cgagcgtttt ttggaccctt tccctatttt ttgggttaag gaaaatgaca gaaaatatat 9480  
 ctaatgagcc ttcgctcaac agtgctccga agtatagctt tccaaaagga gaggcaaagc 9540  
 aatttaagaa tgtatgaaca aaataaagg gaaaaattac cccctctact ttaccaaacg 9600  
 aatactacca ataataattt caacttttcc ttatgatttt tteactgaag cgcttcgcaa 9660  
 tagttgtgag cgatatcaaa agtaacgaaa tgaacttcgc ggctcgtgct atattcttgt 9720

tgctaccgtc catatctttc catagatfff caatctttga tgtctccatg gtggtacaga 9780  
  
 gaacttgtaa acaattcggg ccctacatgt gaacggccgg ccaagcacgc ggggatccga 9840  
 agcatgtagg gaggtcatga tatgaaaaag caaaagagta ggcatcaaaa agttttctcat 9900  
 tcaagtggta acigctggtt aaattaagat atttataaat tgaagcttgg tcgttccgac 9960  
 caataccgta gggaaacgta aattagctat tgtaaaaaaa ggaaaagaaa agaaaagaaa 10020  
 aatgttacat atcgaattga tcttattcct ttggtagacc agtctttgcg tcaatcaaag 10080  
 attcgtttgt ttcttgtggg cctgaaccga cttgagttaa aatcactctg gcaacatcct 10140  
 tttgcaactc aagatccaat tcacgtgcag taaagttaga tgattcaaat tgatggttga 10200  
  
 aagcctcaag ctgctcagta gtaaatttct tgtcccatcc aggaacagag ccaaacaatt 10260  
 tatagataaa tgcaaagagt ttcgactcat tttcagctaa gtagtacaac acagcatttg 10320  
 gacctgcate aaacgtgtat gcaacgattg tttctccgta aaactgatta atgggtgtggc 10380  
 accaactgat galacgcttg gaagtgtcat tcatgtagaa tattggaggg aaagagtcca 10440  
 aacatgtggc atggaaagag ttggaatcca tcattgtttc ctttgcaaag gtggcgaaat 10500  
 ctttttcaac aatggcttta cgcatgactt caaatctctt tggtagcaca tgttcaattc 10560  
 tttctttaa tagttcggag gttgccacgg tcaattgcat accctgagtg gaactcacat 10620  
  
 cttttttaat atcgctgaca actaggacac aagctttcat ctgaggccag tcagagctgt 10680  
 ctgcgatttg tactgccatg gaatcatgac catcttcagc ttttccatt tcccaggcca 10740  
 cgtatccgcc aaacaacgat ctacaagctg aaccagacc ctttcttgct attctagata 10800  
 tttctgaagt tgactgtggt aattggtata acttagcaat tgcagagacc aatgcagcaa 10860  
 agccagcagc ggaggaagct aaaccagctg ctgtaggaaa gttattttcg gagacaatgt 10920  
 ggagtttcca ttgataaat gtgggcaatg aggcgtcctt cgattccatt tcctttctta 10980  
 attggcgtag gtcgcgcaga caattttgag ttctttcatt gtcgatgctg tgtggttctc 11040  
  
 catttaacca caaagtgtcg cgttcaaact caggtgcagt agccgcagag gtcaacgttc 11100  
 tgaggtcacc ttgcgataaa gtcactgata tggacgaatt ggtgggcaga ttcaacttcg 11160  
 tgtccctttt ccccaatac ttaagggttg cgatgttgac gggtagcgta acggatgctg 11220  
 tgtaaacggt cattgtgaag gtagttcgat ttggaggtc gcgggaggtt actttttttt 11280  
 tggatggacg caaagaagtt taataatcat attacatggc aataccacca tatacatatc 11340  
 catatctaata ctiacttata tgtttggaa atgtaaagag cccattatc ttagcctaaa 11400  
 aaaaccttct ctttggaaact ttcagtaata cgcttaactg ctcatgcta tattgaagta 11460

cggattagaa gccgccgagc gggcgacagc cctccgacgg aagactctcc tccgtgcgtc 11520  
 ctggtcttca ccggtcgcgt tcttgaacg cagatgtgcc tcgcccgcga ctgctccgaa 11580  
 caataaagat tctacaatac tagcttttat ggttatgaag aggaaaaatt ggcagtaacc 11640  
 tggccccaca aaccttcaaa tcaacgaatc aaattaacaa ccataggata ataatgcat 11700  
 tagtttttta gccttatttc tgggtaatt aatcagcga gcatgattt ttgatctatt 11760  
 aacagatata taaatgcaaa agctgcataa ccactttaac taatacttcc aacattttcg 11820  
 gtttgattta ctcttattc aaatgtcata aaagtatcaa caaaaaattg ttaatatacc 11880  
  
 tctatactta cccaccgca agtcgcgcaa ccaactaact ttacaatgct attaccgttc 11940  
 ttaacttctg caccgggaaa ggttattatt ttgggtgaac actctgctgt gtacaacaag 12000  
 cctgccgtcg ctgctagtgt gtctgcgttg agaacctacc tgctaataag cgagtcatct 12060  
 gcaccagata ctattgaatt ggacttcccg gacattagct ttaatcataa gtggccatc 12120  
 aatgatttca atgccatcac cgaggatcaa gtaaacctcc aaaaattggc caaggctcaa 12180  
 caagccaccg atggcttctc tcaggaactc gttagtcttt tggatccgtt gttagctcaa 12240  
 ctatccgaat ccttccacta ccatgcagcg tttgtttcc tgtatatgtt tgtttgctta 12300  
  
 tgccccatg ccaagaatat taagtttct ttaaagtcta ctttaccat cgggtgctggg 12360  
 ttgggctcaa gcgctctat ttctgtatca ctggccttag ctatggccta cttggggggg 12420  
 ttaataggat ctaatgactt ggaaaagctg tcagaaaacg ataagcatat agtgaatcaa 12480  
 tgggccttca tagtgaaaa gtgtattcac ggtaccctt caggaataga taacgctgtg 12540  
 gccacttatg gtaatgccct gctattttaa aaagactcac ataatggaac aataaacaca 12600  
 aacaatttta agttcttaga tgatttccca gccattccaa tgatcctaac ctatactaga 12660  
 attccaaggt ctacaaaaga tcttgttgct cgcgttcgtg tgttggctac cgagaaattt 12720  
  
 cctgaagtta tgaagccaat tctagatgcc atgggtgaat gtgccctaca aggcttagag 12780  
 atcatgacta agttaagtaa atgtaaaggc accgatgacg aggctgtaga aactaataat 12840  
 gaactgtatg aacaactatt ggaattgata agaataaatc atggactgct tgtctcaatc 12900  
 ggtgtttctc atcctggatt agaacttatt aaaaatctga gcatgattt gagaattggc 12960  
 tccacaaaac ttaccggtgc tggtagcggc ggttgctctt tgactttggt acgaagagac 13020  
 attactcaag agcaaatga cagtttcaaa aagaaattgc aagatgattt tagttacgag 13080  
 acatttgaag cagacttggg tgggactggc tgctgtttgt taagcgcaaa aaatttgaat 13140  
  
 aaagatctta aaatcaaatc cctagtattc caattattg aaaataaac taccacaaag 13200  
 caacaaattg acgatctatt attgccagga aacacgaatt taccatggac ttcataagct 13260  
 aatttgcgat aggcattatt tattagttgt ttttaactt aactgtgtat gaagttttat 13320

gtaataaaga tagaaagaga aacaaaaaaaa aatttttcgt agtatcaatt cagctttcga 13380  
 agacagaatg aaatttaagc agaccatagt atccttgata cattgactca ggtccgccgg 13440  
 cgttggacga gcaagcatc ttgccctgtg cttggccccc agtgcagcga acgttataaa 13500  
 aacgaatact gagtataat ctatgtaaaa caaccatatac atttcttgtt ctgaactttg 13560

tttacctaac tagtttttaa tttccctttt tctgcatgc ggggtttctt atttattagc 13620  
 atactacatt tgaatatca aatttcctta gtagaaaagt gagagaaggt gcactgacac 13680  
 aaaaaataaa atgctacgta taactgtcaa aactttgcag cagcgggcat ccttccatca 13740  
 tagcttcaaa catattagcg ttctgatct tcatacccg tctcaaaatg atcaaacaaa 13800  
 ctgttattgc caagaaataa acgcaaggct gccttcaaaa actgatccat tagatcctca 13860  
 tatcaagctt cctcatagaa cgccaatta caataagcat gttttgctgt tatcaccggg 13920  
 tgataggttt gctcaggcgg tttaaacgcg tggccgtgcc gtc 13963

<210> 36

<211> 10209

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: MS49253 integration construct

<400> 36

gacggcacgg ccacgcgttt aaaccgccac ccagccaagg tagtctaaaa gctaatttct 60  
 ctaaaagggg gaaagtgggt gattttttat ctcgattat tatatatgca agaatagtta 120  
 aggtatagtt ataaagtttt atcttaattg ccacatacgt acattgacac gtagaaggac 180  
 tccattatth tttcattct agcatactat tattccttgt aacgtcccag agtattccat 240  
 ttaattgtcc tccatttctt aacggtgacg aaggatcacc atacaacaac tactaaagat 300

tatagtacac tctcaccttg caactattta tctgacattt gccttacttt tatctccagc 360  
 ttcccctcga ttttattttt caatttgatt tctaaagctt tttgcttagg cataccaaac 420  
 catccactca tttaacacct tatttttttt ttcgaagaca gcatccaact ttatacgttc 480  
 actacctttt tttttacaac aatttcattc ttcacatctat gaacgctcgt ccaacgccgg 540  
 cggacctttc agacgcgact gcctcatcag taagaccctg tgaagaagaac ttacctgaaa 600  
 aaaacgaata tatactagcg ttgaatgtta gcgtcaacaa caagaagttt aatgacgcgg 660  
 aggccaagcg aaaaagattc cttgattacg taaggaggtt agaatcattt tgaataaaaa 720

acacgctttt tcagttcgag tttatcatta tcaatactgc catttcaaag aatacgtaaa 780  
 taattaatag tagtgatttt cctaacttta tttagtcaaa aaattagcct ttaattctg 840

ctgtaaccgg tacatgcccc aaataggggg cgggttacac agaatatata acatcgtagg 900  
 tgtctgggtg aacagtttat tcttggcatc cactaaatat aatggagccc gctttttaag 960  
 ctggcatcca gaaaaaaaaa gaatcccagc accaaaatat tgttttcttc accaaccatc 1020  
 agttcatagg tccattctct tagcgcaact acagagaaca ggggcacaaa caggcaaaaa 1080  
 acgggcacaa cctcaatgga gtgatgcaac ctgcctggag taaatgatga cacaaggcaa 1140  
  
 ttgaccacag catgtatcta tctcattttc ttacaccttc tattaccttc tgctctctct 1200  
 gatttggaaa aagctgaaaa aaaaggttga aaccagtcc ctgaaattat tcccctactt 1260  
 gactaataag tatataaaga cggtaggtat tgattgtaat tctgtaaatc tatttcttaa 1320  
 acttcttaaa ttctactttt atagttagtc ttttttttag ttttaaaaca ccaagaactt 1380  
 agtttcgacc tcccgcgacc tccaaaatcg aactaccttc acaatgacat ctccagttat 1440  
 cggaaactct tggaagaagt tgaaccccc agtatctgaa gaggcaatag aggggtgtcg 1500  
 caagtattgg agggctgcta attacttgtc aatcggacaa atctacttga ggtcaaacc 1560  
  
 attaatgaag gaacctttca ccaggggaaga tgttaagcat aggttgggtcg gtcactgggg 1620  
 aaccacacca ggattaaact ttttgattgg acacatcaac agattgattg cagatcacca 1680  
 acagaacacc gtcataatca tgggtccagg acatggaggt ccagctggta cgcgccagtc 1740  
 ttatttggat ggtacctata ccgagtattt cccaaatatt acaaaggacg aggcaggttt 1800  
 acagaagttt ttcagacagt tctcttacc aggaggaate ccttcacact acgctccaga 1860  
 aacaccaggt tctattcatg aggggtggtga attaggatat gccttatcac atgcttacgg 1920  
 agcagtcatg aacaatcctt ctttgttcgt cccagccata gtaggtgatg gagaagccga 1980  
  
 gaccggtcct ttagccacag gttggcaatc taacaagtta gtaaaccta ggactgacgg 2040  
 aattgtcttg cctattttgc atttgaacgg atacaagatc gctaaccbaa ccatcttgtc 2100  
 taggatatct gacgaggagt tgcatgagtt tttccacggt atgggatacg agccttatga 2160  
 gttcgtcgca ggatttgaca acgaagacca tttgtcaatc cacagaaggt tcgccgagtt 2220  
 gtttgagacc gtctttgacg agatctgtga cattaaggca gctgctcaga cagacgacat 2280  
 gaccagacct ttctatccta tgataatctt caggacacct aagggttga cttgcctaa 2340  
 gtttatagac ggaaaaaaga ctgagggatc atggagatct catcaagtac ctttggcatc 2400  
  
 tgcaagagat acagaagctc acttcgaggt tttgaaaaat tggttggagt catataagcc 2460  
 tgaagaattg ttcgatgcaa atggagctgt taagccagag gtaactgctt ttatgcctac 2520  
 cggagagtta agaatcggag agaatcctaa cgctaaggt ggtagaatca gagaggaatt 2580  
 gaatttgcct gcatggagg attacaggt aaaagaggtt gctgaatatg gtcaggtatg 2640  
 gggtcagttg gaagcaacca gaagattagg tgtttacacc agggacatta taaagaacaa 2700

cccagactca ttiaggatct ttggaccaga tgaaccgca tcaaataggt tacaggctgc 2760  
 atatgacggt actaataagc aatgggacgc tggttactta tcagcccaag tagacgaaca 2820  
  
 tatggccggt acaggtcaag ttacagagca attgtctgaa catcaaatgg aaggattctt 2880  
 ggaagcttat ttgttgaccg gaaggcatgg aatttggca tcatatgagt cattcgtaca 2940  
 tgtcatcgac tcaatgttaa atcagcatgc taagtgtta gaagccactg taagagagat 3000  
 cccatggagg aaaccaatit cttcaatgaa cttattgggt tcatctcagc tetggaggca 3060  
 ggatcataat ggattttctc atcaggacc aggtgtcaca tcagttttat tgaacaagtg 3120  
 cttcaacaac gatcacgtta tgggaattta ctttctgtc gattctaaca tgttgttagc 3180  
 tgttgccgag aagtgtaca agtctacaga catgataaac gccatcattg ccggaagca 3240  
  
 gccagccgcc acctggttga ctttggatga ggcaaggct gaattggaga aaggagcagc 3300  
 cgaatgggag tgggctcaa cagccaagtc aaatgatgaa gcacagatag tattggcttc 3360  
 agccggtgat gttcctgtc aagaaatcat ggctgtctcc gataagttag atgctatggg 3420  
 tatcaagttc aaggttgtca acgtagtcga cttgggtcaaa ttgcagtcta ccaaagaaaa 3480  
 tgacgagcc atctctgacg ctgaacttcg agacttattt accgaagaca agccagtatt 3540  
 attcgctac cattcatagc ccagagatgt taggggattg atctatgaca gcctaacca 3600  
 tgacaacttc aacgtccagc gatacgaaga acagggttca accactacc cttatgacat 3660  
  
 ggtcagagtc aacaatattg acaggtacga gttggtcgct gaagcattga gaatgatcga 3720  
 tgcagacaaa tacgcagata aaatcgacga attggaggcc tcagaaaagg aagcattcca 3780  
 gtttgcagtt gataacggtt acgaccatcc tgactacacc gactgggtct attcaggagt 3840  
 aaataccaac aagcagggtg ctgtttcagc taccgctgca actgctggtg acaatgaata 3900  
 aagatctatt gaattgaatt gaaatcgata gatcaatit tttcttttct ctttcccat 3960  
 cctttacgct aaaataatag tttatitit tttttgaata tttttatit atatacgtat 4020  
 atatagacta ttatitct tttaatgatt attaagattt ttattaaaa aaaattcgct 4080  
  
 cctcttttaa tgcctttatg cagttttit tcccattcg atatitctat gttcgggttc 4140  
 agcgtatit aagttaata actcgaat tctgcgttcg ttaaagctit cgagaaggat 4200  
 attatitcga aataaacgt gttgtgtaag cttgaagcct ttttgcctg ccaatitct 4260  
 tatccatceta ttgtactct tagatccagt atagtgtatt cttctgtc caagctcctc 4320  
 ccatccccgc gtgcttgcc ggccgtacga aaatcgtat tgtctgaag gtgaaatitc 4380  
 tactcttatt aatggtgaac gttaaagctga tgctatgat gaagctgatt ggtcttaact 4440

tgcttgtcat cttgctaata gtcattggct cgtgttatta ctttaagtat ttgtactcgt 4500  
  
 ttggaacgta atgctaata gtcattggct cgtgttatta ctttaagtat ttgtactcgt 4560  
 aagcttttca attcatcttt tttttttttg ttcttttttt tgattccggg ttctttgaaa 4620  
 tttttttgat tcggtaactc ccgagcagaa ggaagaacga aggaaggagc acagacttag 4680  
 attggtatata atacgcata gttggtgtga agaacaatga aattgccagc tattcttaac 4740  
 ccaactgcac agaacaataa cctgcaggaa acgaagataa atcatgtcga aagctacata 4800  
 taaggaacgt gctgctactc atcctagtcc tgttgctgcc aagctattta atatcatgca 4860  
 cgaaaagcaa acaaaactgt gtgcttcatt ggatgttcgt accaccaagg aattactgga 4920  
  
 gttagttaa gcattagtc ccaaaatttg ttactaaaa acacatgtgg atatcttgac 4980  
 tgatttttcc atggagggca cagttaagcc gctaaaaggca ttatccgcca agtacaattt 5040  
 ttactcttc gaagacagaa aatttgctga cattggtaat acagtcaaat tgcagtactc 5100  
 tgcgggtgta tacagaatag cagaatgggc agacattacg aatgcacacg gtgtggtggg 5160  
 cccaggtatt gttagcggtt tgaagcaggc ggcggaagaa gtaacaaaagg aacctagagg 5220  
 ccttttgatg ttagcagaat tgtcatgcaa gggctccta gctactggag aatatactaa 5280  
 ggtactgtt gacattgcga agagtacaa agattttgtt atcggcttta ttgctcaaag 5340  
  
 agacatgggt ggaagagatg aaggttacga ttggttgatt atgacaccg gtgtgggttt 5400  
 agatgacaag ggagacgcat tgggtcaaca gtatagaacc gtggatgatg tggctctac 5460  
 aggatctgac attattatg ttggaagagg actatttga aaggaaggg atgctaaggt 5520  
 agagggtgaa cgttacagaa aagcaggctg ggaagcatal ttgagaagat gcggccagca 5580  
 aaactaaaa actgtattat aagtaaatgc atgtatactaa aactcacaaa ttagagcttc 5640  
 aatttaatta tatcagttat taccacgaaa atcgttatg tcttgaaggt gaaatttcta 5700  
 ctcttattaa tggatgaact taagctgatg ctatgatgga agctgattgg tettaacttg 5760  
  
 cttgtcatct tgctaata gtcattggct cgtgttatta ctttaagtat ttgtactcgt 5820  
 ttgaacgtaa tgctaata gtcattggct cgtgttatta ctttaagtat ttgtactcgt 5880  
 gggatgggat gagcttgag caggaagaat aactatact ggatctaaag agtacaatag 5940  
 atggataaga atattggcag cgcaaaaagg cttcaagctt acacaacacg gtttatttcg 6000  
 aaataatata cttctcga gctttaacga acgcagaatt ttcagattat taaacttaa 6060  
 atacgtgaa cccgaacata gaaatcga atgggaaaa aaaactgcat aaaggcatta 6120  
 aaagaggagc gaatttttt ttaataaaaa tcttaataat cattaaaaga taaataatag 6180

tctatatata cgtatataaa taaaaaatat tcaaaaaata aaataaacta ttattttagc 6240  
gtaaaggatg gggaaagaga aaagaaaaaa attgatctat cgatttcaat tcaattcaat 6300  
agatctttat tcattgtcac cagcagttgc agcggtagct gaaacagcac cctgcttggt 6360  
ggtatftact ccigaataga cccagtcggt gtagtcagga tggtcgtaac cgttatcaac 6420  
tgcaactgg aatgcttctt ttctgaaggc ctccaattcg tcgattttat ctgcgtatft 6480  
gtctgcatcg atcatttca atgcttcagc gaccaactcg tacctgtcaa tattgttgac 6540  
tctgaccatg tcataagggg tagtggttga accctgttct tcglatccgt ggacgttgaa 6600  
  
gttgtcatgg ttaggcctgt catagatcaa tcccctaaca tctctggcgt atgaatggta 6660  
ggcgaataat actggcttgt cttcggtaaa taagtctgcg aagtcagcgt cagagatggc 6720  
ctcgtcattt tctttggtag actgcaattt gaccaagtcg actacgttga caaccttgaa 6780  
cttgatacc atagcatcta acttatcggc agcagccatg atttcttgag caggaacatc 6840  
accggctgaa gccaatacta tctgtgcttc atcatttgac ttggctgttg aggcccactc 6900  
ccattcggct gtcctttct ccaattcagc ccttgcctca tccaaggtca accaggtggc 6960  
ggctggctgc tttccggcaa tgatggcgtt tatcatgtct gtagacttgt agcacttctc 7020  
  
ggcaacagct aacaacatgt tagaatcgac aggaaagtaa attccgataa cgtgatcgtt 7080  
gttgaagcac ttgttcaata aaactgatgt gacacctggg tctgatgag aaaatccatt 7140  
atgatcctgc ctccagacgt gagatgaaac caataagttc attgaagaaa ttggtttctc 7200  
ccatgggac tctcttacag tggcttctaa ccacttagca tgctgattta acattgagtc 7260  
gatgacatgt acgaatgact catatgatga ccaaatcca tgccttccgg tcaacaata 7320  
agcttccaag aatccttcca tttgatgttc agacaattgc tctgtaactt gacctgtaac 7380  
ggccatattg tcgtctactt gggctgataa gtaaccagcg tcccattgct tattagtaac 7440  
  
gtcatatgca gctgtaacc tatttgatgc ggtttcatct ggtccaaaga tectaaatga 7500  
gtctgggttg ttctttataa tgtccctggf gtaaacacct aatcttctgg ttgcttccaa 7560  
ctgaccccat ccatgaccat attcagcaac ctcttttacc tcgtaatect ccaatgcagg 7620  
caaattcaat tectctctga ttctaccacc attagcgtta ggattctctc cgattcttaa 7680  
ctctccggtg ggcataaaag cagttacctc tggcttaaca gctccatttg catcgaacaa 7740  
ttcttcagge ttatatgact ccaaccaatt tttcaaaacc tcgaagtgag cttctgtatc 7800  
tcttgcatg gccaaaggta cttgatgaga tctccatgat cctcagttct tttttcctc 7860  
  
tataaactta gggcaagtcc aacccttagg tgcctgaag attatcatag gatagaaagg 7920  
tctggctcatg tcgtctgtct gacgagctgc cttaattgtca cagatctcgt caaagacggf 7980  
ctcaaacaaac tcggcgaacc ttctgtggat tgacaaatgg tcttctgtgt caaatctgc 8040

gacgaactca taaggctcgt atcccatacc gtggaaaaac tcatgcaact cctcgtcaga 8100  
tatcctagac aagatggttg ggttagcgat cttgtatccg ttcaaatgca aaataggcaa 8160  
gacaattccg tcagtcctag ggtttactaa cttgttagat tgccaacctg tggctaaagg 8220  
accggtctcg gcttctccat cacctactat ggctgggacg aacaaagaag gattgttcat 8280  
  
gactgctccg taagcatgtg ataaggcata tcctaattca ccaccctcat gaatagaacc 8340  
tgggttttct ggagcgtagt gtgaagggat tcctcctggg taagagaact gtctgaaaaa 8400  
cttctgtaaa cctgcctcgt cctttgtaat atttgggaaa tactcggat aggtaccatc 8460  
caaataagac tgggcggtac cagctggacc tccatgtcct ggacccatga ttatgacggt 8520  
gttctgttgg tgatctgcaa tcaatctggt gatgtgtcca atcaaaaagt ttaatcctgg 8580  
tgtggttccc cagtgaccga ccaacctatg cttaacatct tccctgggta aaggttcctt 8640  
cattaatggg tttgacctca agtagatttg tccgattgac aagtaattag cagccctcca 8700  
  
atacttgtcg acaccctcta ttgcctcttc agatactggg gcgttcaact tcttccaagg 8760  
agttccgata actggagatg tcattgtgaa ggtagttcga ttttggaggt cgcgggaggt 8820  
cgaaactaag ttcttgggtg tttaaaacta aaaaaagac taactataaa agtagaattt 8880  
aagaagtta agaatagat ttacagaatt acaatcaata cctaccgtct ttataactt 8940  
attagtcaag taggggaata atttcaggga actggtttca acctttttt tcagcttttt 9000  
ccaaatcaga gagagcagaa ggtaatagaa ggtgtaagaa aatgagatag atacatgcgt 9060  
gggtcaattg ccttgtgtca tcatttactc caggcaggtt gcactactcc attgaggttg 9120  
  
tgccccTTTT ttgcctgttt gtgcccctgt tctctgtagt tgcgctaaga gaatggacct 9180  
atgaactgat ggttggtgaa gaaaacaata ttttgggtct gggattcttt ttttttctgg 9240  
atgccagctt aaaaagcggg ctccattata tttagtggat gccaggaata aactgttcac 9300  
ccagacacct acgatgttat atattctgtg taaccgccc cctattttgg gcactgtacgg 9360  
gttacagcag aattaaaagg ctaattttt gactaaataa agttaggaaa atcactacta 9420  
ttaattattt acgtattctt tgaaatggca gtattgataa tgataaactc gaactgaaaa 9480  
agcgtgtttt ttattcaaaa tgattctaac tccttacgt aatcaaggaa tetttttgcc 9540  
  
ttggcctccg cgtcattaaa cttcttgttg ttgacgctaa cattcaacgc tagtatatat 9600  
tcgttttttt caggtaagtt cttttcaacg ggtcttactg atgaggcagt cgcgtctgaa 9660  
aggctccgcc gcgttggacg agcgtccat gctggactta ctcgtcgaag atttctgct 9720  
actctctata taattagaca cccatgttat agatttcaga aaacaatgta ataatatatg 9780  
gtagcctcct gaaactacca agggaaaaat ctcaacacca agagctcata ttcgttggaa 9840  
tagcgataat atctctttac ctcaacttta tatgcatgtt atttgctctt ataattggtc 9900

tctatttagg gaaaaaagtc ggtttgagag ctctcgcga tgtgaaatct caatttgaac 9960

tgacagccaa agctagccca ttccacgaac accagaaaga agaaatcccc aaggatcgca 10020

tgacagagta tgctctctca tategttgag tatgaatgcc aatacactga tcagctttac 10080

aagaaacgta aaatctggca cgatggtaga ctgaaatact ttcagttaaa caacagattc 10140

atgctttata cggaaaagga taacgttttg ttagctagtg aggcggttta aacgcgtggc 10200

cgtgccgctc 10209

<210> 37

<211> 7257

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic: MS49298 integration construct

<400>

37

gacggcacgg ccacgcgttt aaaccgccac ccagccaagg tagtctaaaa gctaatttct 60

ctaaaagggg gaaagtgtgt gattttttat ctgcattat tataatgca agaatagtta 120

aggtatagtt ataaagtttt atcttaattg ccacatacgt acattgacac gtagaaggac 180

tccattatth ttttcattct agcatactat tttccttgt aacgtcccag agtattccat 240

ttaattgtcc tccatttctt aacggtgacg aaggatcacc atacaacaac tactaaagat 300

tatagtacac ttcaccttg caactattta tctgacattt gccttacttt tatctccagc 360

ttcccctcga ttttattttt caatttgatt tctaaagctt tttgcttagg cataccaac 420

catccactca ttiaacacct tatttttttt ttcgaagaca gcatccaact ttatacgttc 480

actacctttt tttttacaac aatttcattc ttcacctat gaacgctcgt ccaacgccgg 540

cggaccttcc agacgcgact gcctcatcag taagaccctg tgaagaagaac ttacctgaaa 600

aaaacgaata taaactagcg ttgaatgtta gcgtcaacaa caagaagttt aatgacgcgg 660

aggccaagcg aaaaagattc cttgattacg taaggagatt agaactcatt tgaataaaaa 720

acacgctttt tcagttcgag tttatcatta tcaatactgc catttcaaag aatacgtaaa 780

taattaatag tagtgatttt cctaacttta tttagtcaaa aaattagcct ttaattctg 840

ctgtaaccgg tacatgcccc aaataggggg cgggttacac agaataata acatcgtagg 900

tgtctgggtg aacagtttat tcctggcacc cactaaatat aatggagccc gctttttaag 960

ctggcatcca gaaaaaaaa gaatcccagc accaaaatat tgttttcttc accaaccatc 1020

agttcatagg tccattctct tagcgcact acagagaaca ggggcacaaa caggcaaaaa 1080

acgggcacaa cctcaatgga gtgatgcaac ctgcctggag taaatgatga cacaaggcaa 1140

ttgacccacg catgatccta tctcattttc ttacaccttc tattaccttc tgctctctct 1200  
 gatttggaaa aagctgaaaa aaaaggttga aaccagttcc ctgaaattat tcccctactt 1260  
  
 gactaataag tatataaaga cggtaggtat tgattgtaat tctgtaaate tatttcttaa 1320  
 acttcttaaa ttctactttt atagttagtc ttttttttag ttttaaaaca ccaagaactt 1380  
 agtttcgacc tcccgcgacc tccaaaatcg aactaccttc acaatgaaat tgatggagaa 1440  
 tatcttcgga ttggccaagg ccgacaagaa gaaaatcggt ttggcagagg gtgaggaaga 1500  
 gaggaacatc agggcttcag aggagattat tagggacggt attgccgaca taatcttggt 1560  
 cggttcagag tctgtcatta aggaaaacgc cgcaaaatc ggagtaaatt tggccggagt 1620  
 agagatagtc gaccagaaaa cttcttctaa gaccgcccgt tacgccaacg ctttctacga 1680  
  
 gatcagaaaa aacaaggag tccacttga gaaggctgac aaaatcgtca gggaccaat 1740  
 ctacttcga acaatgatgg tcaagttagg tgacgctgac ggtttggtat ctggtgctat 1800  
 acatactaca ggagacttgt taaggcctgg ttgacagatt gtcaaaacag taccaggtgc 1860  
 atctgtcgtc tcatcagctt tcttgatgac agtacctgac tgcgagtatg gagaggacgg 1920  
 ttttttgta ttcgctgact gcgctgtaaa tgtttgtcct accgctgaag agttatcttc 1980  
 aatcgcaatt accactgctg agactgcaaa gaatttgtgc aagatcgagc caagggtagc 2040  
 catgttgcata ttctcaacca tgggatcagc ctccatgaa ttagtcgaca aggttacaaa 2100  
  
 ggcaacaaaa ttggctaagg aggctaggcc tgacttagac atcgacggtg aattgcagtt 2160  
 agacgcctca ttggttaaga aggtcgaga tttgaaagcc cctggatcta aagtcgctgg 2220  
 taaggcaaat gtcttgatct tcccagacat ccaggcagga aacatcggat acaagttggt 2280  
 ccaaagattc gcaaaggccg aagccatcgg tcctatatgt cagggatttg ccaaactat 2340  
 caacgatttg tcaagggat gttcagtcga cgacatcgtc aaagtgttg ccgttacgc 2400  
 agttcaggca caagcacaag gataaagatc tattgaattg aattgaaatc gatagatcaa 2460  
 ttttttctt ttctctttcc ccatccttta cgctaaaata atagtttatt ttattttttg 2520  
  
 aatatttttt atttatatac gtatatatag actattatit atcttttaat gattattaag 2580  
 atttttatta aaaaaaatt cgctcctctt ttaatgcctt tatgcagttt ttttttccca 2640  
 ttcgatattt ctatgttcgg gttcagcgtt ttttaagttt aataactcga aaattctgcg 2700  
 ttcgttaaag ctttcgagaa ggatattatt tcgaaataaa ccgtgttggt taagcttgaa 2760  
 gcctttttgc gctgccaata ttcttatcca tctattgtac tctttagatc cagtatagt 2820  
 tattcttctt gctccaagct catcccatcc ccgctgctt ggccggccgt acgaaaaatcg 2880

ttattgtcctt gaagggtgaaa tttctactct tattaatggg gaacgttaag ctgatgctat 2940  
  
 gatggaagct gattggctctt aacttgcttg tcatcttgct aatggtcatt ggctcgtggt 3000  
 attacttaag ttatittgtac tcgttttgaa cgtaatgcta atgatcatct tatggaataa 3060  
 tagtgagtgg tttcagggtc cataaagctt ttcaattcat cttttttttt tttgttcttt 3120  
 tttttgattc cggtttcttt gaaatttttt tgattcggta atctccgagc agaaggaaga 3180  
 acgaaggaag gagcacagac ttagattggg atatatagc atagtgggtg ttgaagaaac 3240  
 atgaaattgc ccagtattct taaccaact gcacagaaca aaaacctgca gaaacgaag 3300  
 ataatcatg tcgaaagcta catataagga acgtgctgct actcacccta gtctgttgc 3360  
  
 tgccaagcta ttiaatatca tgcacgaaaa gcaaacaac ttgtgtgctt cattggatgt 3420  
 tcgtaccacc aaggaattac tggagttagt tgaagcatta ggtcccaaaa tttgtttact 3480  
 aaaaacacat gtggatact tgaactgatt ttccatggag ggcacagtt aagccgctaaa 3540  
 ggcatatcc gccaagtaca attttttact cttcgaagac agaaaatttg ctgacattgg 3600  
 taatacagtc aaattgcagt actctgcggg tgtatacaga atagcagaat gggcagacat 3660  
 tacgaatgca cacggtgtgg tgggccagg tattgttagc ggtttgaagc aggcggcgga 3720  
 agaagtaaca aaggaaccta gaggcctttt gatgttagca gaattgtcat gcaagggtc 3780  
  
 cctagctact ggagaatata ctaagggtac tgttgacatt gcgaagagtg acaaagattt 3840  
 tgttatcggc ttiattgctc aaagagacat ggggtggaaga gatgaaggtt acgattgggt 3900  
 gattatgaca cccggtgtgg gtttagatga caaggagac gcattgggtc aacagtatag 3960  
 aaccgtggat gatgtggtct ctacaggatc tgacattatt attgttggaa gaggactatt 4020  
 tgcaaagga agggatgcta aggtagaggg tgaacgttac agaaaagcag gctgggaagc 4080  
 atatttgaga agatgcggcc agcaaaacta aaaaactgta ttataagtaa atgcatgtat 4140  
 actaaactca caaattagag cttcaattta attatatcag ttattaccac gaaaatcgtt 4200  
  
 attgtcttga aggtgaaatt tctactctta ttaatgggta acgttaagct gatgctatga 4260  
 tggaagctga ttggtcttaa cttgcttgc atcttgctaa tggcatatg gctcgtgtta 4320  
 ttacttaagt tatttgtact cgttttgaac gtaatgctaa tgatcatctt atggaataat 4380  
 agtgaacggc cggccaagca cgcgggatg ggatgagctt ggagcaggaa gaatacacta 4440  
 tactggatct aaagagtaca atagatggat aagaatattg gcagcgcaaa aaggcttcaa 4500  
 gcttacacaa cacggtttat ttcgaataa tacccttctc gaaagcttta acgaacgcag 4560  
 aattttcgag ttattaaact taaaatagc tgaaccgaa catagaaata tcgaatggga 4620

aaaaaaact gcataaaggc attaaaagag gagcgaatth tttttaata aaaatcttaa 4680  
 taatcattaa aagataaata atagtctata tatacgtata taaataaaaa atattcaaaa 4740  
 aataaaataa actattatth tagcgtaaag gatggggaaa gagaaaagaa aaaaattgat 4800  
 ctatcgatth caattcaatt caatagatct ttatccttgt gcttgtgcct gaactgcggt 4860  
 aacggcaaca actttgacga tgcctcgcac tgaacatccc cttgacaaat cgttgatagg 4920  
 tttggcaaat ccttgacata taggaccgat ggcttcggcc tttgcgaatc tttggaccaa 4980  
 cttgatccg atgtttcctg cctggatgic tgggaagatc aagacattg ccttaccagc 5040  
  
 gactttagat ccaggggctt tcaaatctgc gaccttctta accaatgagg cgtctaactg 5100  
 caattcaccg tcgatgtcta agtcaggcct agcctcctta gccaatthtgg ttgcctttht 5160  
 aaccttgtcg actaattcat gtgaggctga tcccatggtt gagaatgaca acatggctac 5220  
 ccttggctcg atcttgaca aattctthtgc agtctcagca gtggtaatg cgattgaaga 5280  
 taactcttca gcggtaggac aaacatthtac agcgcagtca gcgaataaca aaaaaccgtc 5340  
 ctctccatac tcgcagtcag gtactgacat caagaagact gatgagacga cagatgcacc 5400  
 tggtagctgt ttgacaatc gcaaaccagg ccttaacaag tctcctgtag tatgtatagc 5460  
  
 accagatacc aaaccgtcag cgtcacctaa cttgaccatc attgttgcga agtagattgg 5520  
 gtccctgacg atthtgtcag ccttctccaa ggtgactccc ttgtthtthc tgatctcgta 5580  
 gaaagcgttg gcgtaaccgg cggtcttaga agaagthtct gggtcgacta tctctactcc 5640  
 ggccaaatth actccgaatth ttgcggcgtt thccttaatg acagactctg aaccgaccaa 5700  
 gattatgtcg gcaataccgt ccctaataat ctctctgaa gccctgatgt tctctctthc 5760  
 ctcacctct gccaaaacga thtctctctt gtcggccttg gccaatccga agatattctc 5820  
 catcaatthc attgtgaagg tagttcgatth ttggaggtcg cgggaggtcg aactaagth 5880  
  
 cttggtgtht taaaactaaa aaaaagacta actataaaag tagaatthaa gaagthtaag 5940  
 aaatagatth acagaatthc aatcaatacc taccgtctth atatacttht tagtcaagta 6000  
 ggggaataat thcagggaac tggthtcaac cthtthtthc agctthtthc aatcagaga 6060  
 gagcagaagg taatagaagg tgtaagaaaa tgagatagat acatgcgtgg gtcaattgcc 6120  
 ttgtgtcatc attactcca ggcaggthtgc atcactccat tgaggthtgt cccgthttht 6180  
 gcctgthtgt gccctgttc tctgtagtht cgctaagaga atggacctat gaactgatgg 6240  
 ttggtgaaga aaacaattht ttggtgctgg gattctthtth thtctggat gccagctthaa 6300  
  
 aaagcgggct ccattattht tagtggatgc caggaataaa ctgthcacc agacacctac 6360  
 gatgthtatat atthtctgtga acccctcccc thtthtgggc atgtacgggt tacagcagaa 6420  
 ttaaaaggct aatthtthtga ctaaataaag ttaggaaat cactacttht aatthtthc 6480

gtattctttg aaatggcagt attgataatg ataaactcga actgaaaaag cgtgtttttt	6540
attcaaatg attctaactc ctttacgtaa tcaaggaatc tttttgcctt gcctccgcg	6600
tcattaaact tcttggtgtt gacgctaaca ttcaacgcta gtatatattc gtttttttca	6660
ggtaagttct ttcaacggg tcttactgat gaggcagtcg cgtctgaaag gtccgccggc	6720
gttgacgag cgctccatgc tggacttact cgtcgaagat ttctgctac tctctatata	6780
attagacacc catgttatag atttcagaaa acaatgtaat aatatatggt agcctcctga	6840
aactaccaag ggaaaaatct caacaccaag agctcatatt cgttgaata gcgataatat	6900
ctctttacct caatcttata tgcattgtat ttgctcttat aattggtctc tatttaggga	6960
aaaaagtcgg ttgagagct tctcgcgatg tgaatctca attgaaactg cacgccaaag	7020
ctagcccatt tcacgaacac cagaaagaag aaatcccaa ggatcgcgat acagagtatg	7080
ctctctcata tcgttgagta tgaatgcaa tacactgac agctttaca gaaacgtaaa	7140
atctggcacg atggtagact gaaatacttt cagttaaaca acagattcat gctttatag	7200
gaaaaggata acgttttggt agctagttag gcggtttaaa cgcgtggccg tgccgtc	7257