

|                                                                                                                                          |                                                  |                                                                                    |
|------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|------------------------------------------------------------------------------------|
|                                                         | <b>(19) 대한민국특허청(KR)</b><br><b>(12) 공개특허공보(A)</b> | <b>(11) 공개번호</b> 10-2010-0037031<br><b>(43) 공개일자</b> 2010년04월08일                   |
| <b>(51)</b> Int. Cl.<br><i>C12N 15/31</i> (2006.01) <i>C12N 1/21</i> (2006.01)<br><i>C12P 7/06</i> (2006.01) <i>C12R 1/145</i> (2006.01) |                                                  | <b>(71)</b> 출원인<br><b>마스코마 코포레이션</b><br>미국 뉴햄프셔 레바논 스위트 300 에트나 로드 67<br>(우:03766) |
| <b>(21)</b> 출원번호 10-2009-7025740                                                                                                         |                                                  | <b>(72)</b> 발명자<br><b>호그셋, 데이비드, 에이.</b><br>미국 03753 뉴햄프셔 그랜템 식스 윈터 힐              |
| <b>(22)</b> 출원일자 2008년05월09일<br>심사청구일자 없음                                                                                                |                                                  | <b>라이갈하이아, 비니트, 비.</b><br>미국 03766 뉴햄프셔 레바논 #201 스펜서 스트리트 21                       |
| <b>(85)</b> 번역문제출일자 2009년12월09일                                                                                                          |                                                  | <b>(74)</b> 대리인<br><b>남상선</b>                                                      |
| <b>(86)</b> 국제출원번호 PCT/US2008/063237                                                                                                     |                                                  |                                                                                    |
| <b>(87)</b> 국제공개번호 WO 2008/141174<br>국제공개일자 2008년11월20일                                                                                  |                                                  |                                                                                    |
| <b>(30)</b> 우선권주장<br>60/916,978 2007년05월09일 미국(US)                                                                                       |                                                  |                                                                                    |

전체 청구항 수 : 총 92 항

#### (54) 유전자 녹아웃 중온성 및 호열성 생물체, 및 이의 사용 방법

##### (57) 요약

본 발명의 한 가지 일면은 유전자 변형된 호열성(thermophilic) 또는 중온성(mesophilic) 미생물에 관한 것인데, 여기서 유기산 또는 이의 염의 대사적 생산에 관여하는 제 1의 고유(native) 효소를 엔코딩하는 제 1의 고유 유전자가 부분적으로, 실질적으로 또는 완전히 결실되거나 침묵되거나 비활성화되거나 하향조절됨으로써 발효 생산물로서 에탄올을 생산하는 상기 호열성 또는 중온성 미생물의 고유 능력을 증가시킨다. 특정 구체예에서, 상기 미생물은 에탄올의 대사적 생산에 관여하는 제 1의 비-고유(non-native) 효소를 엔코딩하는 제 2의 비-고유 유전자를 추가로 포함한다. 본 발명의 또 다른 일면은 리그노셀룰로오스 바이오매스를 유전자 변형된 호열성 또는 중온성 미생물과 접촉시키는 것을 포함하여 리그노셀룰로오스 바이오매스를 에탄올로 전환시키는 방법에 관한 것이다.

## 특허청구의 범위

### 청구항 1

SEQ ID NO:1-5, 30-31 및 47-61 중 어느 하나의 뉴클레오티드 서열, 또는 이의 상보서열(complement)을 포함하는 분리된 핵산 분자.

### 청구항 2

SEQ ID NO:1-5, 30-31 및 47-61 중 어느 하나의 뉴클레오티드 서열, 또는 이의 상보서열과 80% 이상의 동일성을 공유하는 뉴클레오티드 서열을 포함하는 분리된 핵산 분자.

### 청구항 3

제 2항에 있어서, SEQ ID NO:1-5, 30-31 및 47-61 중 어느 하나의 뉴클레오티드 서열, 또는 이의 상보서열과 약 95% 이상의 서열 동일성을 지닌 핵산 분자.

### 청구항 4

호열성(thermophilic) 또는 중온성(mesophilic) 세균에서 발현가능한 프로모터에 작동적으로 결합된 SEQ ID NO:1-5, 30-31 및 47-61 중 어느 하나를 포함하는 유전자 구성물(genetic construct).

### 청구항 5

제 4항의 유전자 구성물을 포함하는 재조합 호열성 또는 중온성 세균.

### 청구항 6

제 1항 내지 제 3항 중 어느 한 항의 핵산 분자를 포함하는 벡터.

### 청구항 7

제 1항 내지 제 3항 중 어느 한 항의 핵산 분자를 포함하는 숙주 세포.

### 청구항 8

제 7항에 있어서, 상기 숙주 세포가 호열성 또는 중온성 세균 세포인 숙주 세포.

### 청구항 9

유전자 변형된 호열성 또는 중온성 미생물로서, 유기산 또는 이의 염의 대사적 생산에 관여하는 제 1의 고유(native) 효소를 엔코딩하는 제 1의 고유 유전자가 부분적으로, 실질적으로 또는 완전히 결실되거나 침묵(silencing)되거나 비활성화되거나 하향조절(down-regulation)됨으로써, 발효 생산물로서 에탄올을 생산하는 상기 호열성 또는 중온성 미생물의 고유 능력이 증가된 유전자 변형된 호열성 또는 중온성 미생물.

### 청구항 10

제 9항에 있어서, 상기 미생물이 그람-음성 세균 또는 그람-양성 세균인 유전자 변형된 미생물.

### 청구항 11

제 9항에 있어서, 상기 미생물이 썬모안에어로박테리움(*Thermoanaerobacterium*), 썬모안에어로박터(*Thermoanaerobacter*), 클로스트리디움(*Clostridium*), 게오바실루스(*Geobacillus*), 사카로코쿠스(*Saccharococcus*), 페니바실루스(*Paenibacillus*), 바실루스(*Bacillus*), 칼디셀룰로시럽터(*Caldicellulosiruptor*), 안에어로셀럼(*Anaerocellum*), 또는 안옥시바실루스(*Anoxybacillus*) 속(genus)의 종(species)인 유전자 변형된 미생물.

### 청구항 12

제 9항에 있어서, 상기 미생물이 썬모안에어로박테리움 썬모술폴리게네스(*Thermoanaerobacterium thermosulfurigenes*), 썬모안에어로박테리움 아오테아로엔세(*Thermoanaerobacterium aotearoense*), 썬모안에어

로박테리움 폴리사카롤리티쿰(*Thermoanaerobacterium polysaccharolyticum*), 썩모안에어로박테리움 제에(*Thermoanaerobacterium zee*), 썩모안에어로박테리움 자일라노리티쿰(*Thermoanaerobacterium xylanolyticum*), 썩모안에어로박테리움 사카롤리티쿰(*Thermoanaerobacterium saccharolyticum*), 썩모안에어로비움 브록키(*Thermoanaerobium brockii*), 썩모안에어로박테리움 썩모사카롤리티쿰(*Thermoanaerobacterium thermosaccharolyticum*), 썩모안에어로박터 에타놀리쿠스(*Thermoanaerobacter ethanolicus*), 썩모안에어로박터 브록키(*Thermo anaerobacter brocki*), 클로스트리디움 썩모셀룸(*Clostridium thermocellum*), 클로스트리디움 셀룰로리티쿰(*Clostridium cellulolyticum*), 클로스트리디움 피토펜타스(*Clostridium phytofermentans*), 클로스트리디움 스트라미노솔벤스(*Clostridium straminosolvens*), 게오바실루스 썩모글루코시다시우스(*Geobacillus thermoglucosidasius*), 게오바실루스 스테아로썩모필루스(*Geobacillus stearothermophilus*), 사카로코쿠스 칼독실로실리티쿠스(*Saccharococcus caldxylosilyticus*), 사카로쿠스 썩모필루스(*Saccharococcus thermophilus*), 페니바실루스 캄피나센시스(*Paenibacillus campinasensis*), 바실루스 플라보썩무스(*Bacillus flavothermus*), 안옥시바실루스 캄차트켄시스(*Anoxybacillus kamchatkensis*), 안옥시바실루스 고넨시스(*Anoxybacillus gonensis*), 칼디셀룰로시럽터 아세티게누스(*Caldicellulosiruptor acetigenus*), 칼디셀룰로시럽터 사카로리티쿠스(*Caldicellulosiruptor saccharolyticus*), 칼디셀룰로시럽터 크리스티안소니(*Caldicellulosiruptor kristjanssonii*), 칼디셀룰로시럽터 오웬센시스(*Caldicellulosiruptor owensensis*), 칼디셀룰로시럽터 락토아세티쿠스(*Caldicellulosiruptor lactoaceticus*), 및 안에어로셀룸 썩모필룸(*Anaerocellum thermophilum*)로 구성된 군으로부터 선택된 세균인 유전자 변형된 미생물.

### 청구항 13

제 9항에 있어서, 상기 미생물이 썩모안에어로박테리움 사카롤리티쿰(*Thermoanaerobacterium saccharolyticum*)인 유전자 변형된 미생물.

### 청구항 14

제 9항에 있어서, 상기 미생물이 (a) 헥소오스 당을 대사하는 고유 능력을 지닌 호열성 또는 중온성 미생물; (b) 펜토오스 당을 대사하는 고유 능력을 지닌 호열성 또는 중온성 미생물; 및 (c) 헥소오스 당과 펜토오스 당을 대사하는 고유 능력을 지닌 호열성 또는 중온성 미생물로 구성된 군으로부터 선택되는 유전자 변형된 미생물.

### 청구항 15

제 9항에 있어서, 상기 미생물이 헥소오스 당을 대사하는 고유 능력을 지니는 유전자 변형된 미생물.

### 청구항 16

제 15항에 있어서, 상기 미생물이 클로스트리디움 스트라미노솔벤스(*Clostridium straminosolvens*) 또는 클로스트리디움 썩모셀룸(*Clostridium thermocellum*)인 유전자 변형된 미생물.

### 청구항 17

제 9항에 있어서, 상기 미생물이 헥소오스 당과 펜토오스 당을 대사하는 고유 능력을 지니는 유전자 변형된 미생물.

### 청구항 18

제 17항에 있어서, 상기 미생물이 클로스트리디움 셀룰로리티쿰(*Clostridium cellulolyticum*), 클로스트리디움 크리스티안소니(*Clostridium kristjanssonii*), 또는 클로스트리디움 스테르코라리움 아종 렙토사프라르툼(*Clostridium stercorarium subsp. leptosaprartum*)인 유전자 변형된 미생물.

### 청구항 19

제 15항에 있어서, 펜토오스 당을 대사하는 능력을 부여하는 제 1의 비고유(non-native) 효소를 엔코딩하는 제 1의 비고유 유전자가 삽입됨으로써, 상기 호열성 또는 중온성 미생물이 펜토오스 당으로부터 발효 생산물로서 에탄올을 생산할 수 있게 하는 유전자 변형된 미생물.

#### 청구항 20

제 9항에 있어서, 상기 미생물이 펜토오스 당을 대사하는 고유 능력을 지니는 유전자 변형된 미생물.

#### 청구항 21

제 20항에 있어서, 상기 미생물이 썬모안에어로박테리움 사카롤리티쿰(*Thermoanaerobacterium saccharolyticum*), 썬모안에어로박테리움 자일라노리티쿰(*Thermoanaerobacterium xylanolyticum*), 썬모안에어로박테리움 폴리사카롤리티쿰(*Thermoanaerobacterium polysaccharolyticum*), 및 썬모안에어로박테리움 썬모사카롤리티쿰(*Thermoanaerobacterium thermosaccharolyticum*)으로 구성된 군으로부터 선택되는 유전자 변형된 미생물.

#### 청구항 22

제 20항에 있어서, 핵소오스 당을 대사하는 능력을 부여하는 제 1의 비고유 효소를 엔코딩하는 제 1의 비고유 유전자가 삽입됨으로써, 상기 호열성 또는 중온성 미생물이 핵소오스 당으로부터 발효 생산물로서 에탄올을 생산할 수 있게 하는 유전자 변형된 미생물.

#### 청구항 23

제 9항 내지 제 22항 중 어느 한 항에 있어서, 상기 유기산이 락트산 및 아세트산으로 구성된 군으로부터 선택되는 것인 유전자 변형된 미생물.

#### 청구항 24

제 9항 내지 제 22항 중 어느 한 항에 있어서, 상기 유기산이 락트산인 유전자 변형된 미생물.

#### 청구항 25

제 9항 내지 제 22항 중 어느 한 항에 있어서, 상기 유기산이 아세트산인 유전자 변형된 미생물.

#### 청구항 26

제 9항 내지 제 22항 중 어느 한 항에 있어서, 상기 제 1의 고유 효소가 락테이트 데히드로게나아제, 아세테이트 키나아제 및 포스포트랜스아세틸라아제로 구성된 군으로부터 선택되는 것인 유전자 변형된 미생물.

#### 청구항 27

제 9항 내지 제 22항 중 어느 한 항에 있어서, 상기 제 1의 고유 효소가 락테이트 데히드로게나아제인 유전자 변형된 미생물.

#### 청구항 28

제 9항 내지 제 22항 중 어느 한 항에 있어서, 상기 제 1의 고유 효소가 아세테이트 키나아제인 유전자 변형된 미생물.

#### 청구항 29

제 9항 내지 제 22항 중 어느 한 항에 있어서, 상기 제 1의 고유 효소가 포스포트랜스아세틸라아제인 유전자 변형된 미생물.

#### 청구항 30

제 9항 내지 제 29항 중 어느 한 항에 있어서, 유기산 또는 이의 염의 대사적 생산에 관여하는 제 2의 고유 효소를 엔코딩하는 제 2의 고유 유전자가 부분적으로, 실질적으로 또는 완전히 결실되거나 침묵되거나 비활성화되거나 하향조절되는 유전자 변형된 미생물.

#### 청구항 31

제 30항에 있어서, 상기 제 2의 고유 효소가 아세테이트 키나아제 또는 포스포트랜스아세틸라아제인 유전자 변

형된 미생물.

### 청구항 32

제 30항에 있어서, 상기 제 2의 고유 효소가 락테이트 데히드로게나아제인 유전자 변형된 미생물.

### 청구항 33

유전자 변형된 호열성 또는 중온성 미생물로서, (a) 유기산 또는 이의 염의 대사적 생산에 관여하는 제 1의 고유 효소를 엔코딩하는 제 1의 고유 유전자가 부분적으로, 실질적으로 또는 완전히 결실되거나 침묵되거나 비활성화되거나 하향조절되고, (b) 에탄올의 대사적 생산에 관여하는 제 1의 비고유 효소를 엔코딩하는 제 1의 비고유 유전자가 삽입됨으로써, 상기 호열성 또는 중온성 미생물이 발효 생산물로서 에탄올을 생산할 수 있게 하는 유전자 변형된 호열성 또는 중온성 미생물.

### 청구항 34

제 33항에 있어서, 상기 제 1의 비고유 유전자가 헥소오스 당을 대사하는 능력을 부여하는 제 1의 비고유 효소를 엔코딩함으로써, 상기 호열성 또는 중온성 미생물이 헥소오스 당을 대사할 수 있게 하는 유전자 변형된 미생물.

### 청구항 35

제 33항에 있어서, 상기 제 1의 비고유 유전자가 펜토오스 당을 대사하는 능력을 부여하는 제 1의 비고유 효소를 엔코딩함으로써, 상기 호열성 또는 중온성 미생물이 펜토오스 당을 대사할 수 있게 하는 유전자 변형된 미생물.

### 청구항 36

제 33항에 있어서, 상기 제 1의 비고유 유전자가 헥소오스 당을 대사하는 능력을 부여하는 제 1의 비고유 효소를 엔코딩하고; 펜토오스 당을 대사하는 능력을 부여하는 제 2의 비고유 효소를 엔코딩하는 제 2의 비고유 유전자가 삽입됨으로써, 상기 호열성 또는 중온성 미생물이 헥소오스 당과 펜토오스 당을 대사할 수 있게 하는 유전자 변형된 미생물.

### 청구항 37

제 33항 내지 제 36항 중 어느 한 항에 있어서, 상기 유기산이 락트산인 유전자 변형된 미생물.

### 청구항 38

제 33항 내지 제 36항 중 어느 한 항에 있어서, 상기 유기산이 아세트산인 유전자 변형된 미생물.

### 청구항 39

제 33항, 제 34항 또는 제 36항 중 어느 한 항에 있어서, 상기 제 1의 비고유 효소가 피루베이트 데카르복실라아제(PDC) 또는 알코올 데히드로게나아제(ADH)인 유전자 변형된 미생물.

### 청구항 40

제 33항 또는 제 35항에 있어서, 상기 제 2의 비고유 효소가 자일로오스 이소머라아제인 유전자 변형된 미생물.

### 청구항 41

제 40항에 있어서, 상기 제 1의 비고유 유전자가 SEQ ID NO:6, 10 또는 14에 상응하는 것인 유전자 변형된 미생물.

### 청구항 42

제 33항, 제 35항 또는 제 36항 중 어느 한 항에 있어서, 상기 비고유 효소가 자일룰로키나아제(xylulokinase)인 유전자 변형된 미생물.

#### 청구항 43

제 42항에 있어서, 상기 비고유 유전자가 SEQ ID NO:7, 11 또는 15에 상응하는 것인 유전자 변형된 미생물.

#### 청구항 44

제 33항, 제 35항 또는 제 36항 중 어느 한 항에 있어서, 상기 비고유 효소가 L-아라비노오스 이소머라아제인 유전자 변형된 미생물.

#### 청구항 45

제 44항에 있어서, 상기 비고유 유전자가 SEQ ID NO:8 또는 12에 상응하는 것인 유전자 변형된 미생물.

#### 청구항 46

제 33항, 제 35항 또는 제 36항 중 어느 한 항에 있어서, 상기 비고유 유전자가 L-리블로오스-5-포스페이트 4-에피머라아제인 유전자 변형된 미생물.

#### 청구항 47

제 46항에 있어서, 상기 비고유 유전자가 SEQ ID NO:9 또는 13에 상응하는 것인 유전자 변형된 미생물.

#### 청구항 48

제 9항 내지 제 47항 중 어느 한 항에 있어서, 상기 미생물이 대사된 바이오매스로부터의 탄소 중 60% 이상을 에탄올로 전환시킬 수 있는 유전자 변형된 미생물.

#### 청구항 49

제 9항에 있어서, 상기 미생물이, 셀룰로오스를 가수분해하는 고유 능력을 지닌 호열성 또는 중온성 미생물 (a), 자일란(xylan)을 가수분해하는 고유 능력을 지닌 호열성 또는 중온성 미생물 (b) 및 셀룰로오스와 자일란을 가수분해하는 고유 능력을 지닌 호열성 또는 중온성 미생물 (c)로 구성된 군으로부터 선택되는 유전자 변형된 미생물.

#### 청구항 50

제 9항 또는 제 33항에 있어서, 상기 미생물이 셀룰로오스를 가수분해하는 고유 능력을 지니는 유전자 변형된 미생물.

#### 청구항 51

제 9항 또는 제 33항에 있어서, 상기 미생물이 셀룰로오스와 자일란을 가수분해하는 고유 능력을 지니는 유전자 변형된 미생물.

#### 청구항 52

제 50항에 있어서, 자일란을 가수분해하는 능력을 부여하는 제 1의 비고유 효소를 엔코딩하는 제 1의 비고유 유전자가 삽입되어 있는 유전자 변형된 미생물.

#### 청구항 53

제 9항 또는 제 33항에 있어서, 상기 미생물이 자일란을 가수분해하는 고유 능력을 지니는 유전자 변형된 미생물.

#### 청구항 54

제 53항에 있어서, 셀룰로오스를 가수분해하는 능력을 부여하는 제 1의 비고유 효소를 엔코딩하는 제 1의 비고유 유전자가 삽입되어 있는 유전자 변형된 미생물.

#### 청구항 55

제 49항 내지 제 54항 중 어느 한 항에 있어서, 상기 유기산이 락트산 및 아세트산으로 구성된 군으로부터 선택되는 것인 유전자 변형된 미생물.

#### 청구항 56

제 49항 내지 제 54항 중 어느 한 항에 있어서, 상기 유기산이 락트산인 유전자 변형된 미생물.

#### 청구항 57

제 49항 내지 제 54항 중 어느 한 항에 있어서, 상기 유기산이 아세트산인 유전자 변형된 미생물.

#### 청구항 58

제 49항 내지 제 54항 중 어느 한 항에 있어서, 상기 제 1의 고유 효소가 락테이트 데히드로게나아제, 아세테이트 키나아제 및 포스포트랜스아세틸라아제로 구성된 군으로부터 선택되는 것인 유전자 변형된 미생물.

#### 청구항 59

제 49항 내지 제 54항 중 어느 한 항에 있어서, 상기 제 1의 고유 효소가 락테이트 데히드로게나아제인 유전자 변형된 미생물.

#### 청구항 60

제 49항 내지 제 54항 중 어느 한 항에 있어서, 상기 제 1의 고유 효소가 아세테이트 키나아제인 유전자 변형된 미생물.

#### 청구항 61

제 49항 내지 제 54항 중 어느 한 항에 있어서, 상기 제 1의 고유 효소가 포스포트랜스아세틸라아제인 유전자 변형된 미생물.

#### 청구항 62

제 50항 내지 제 61항 중 어느 한 항에 있어서, 유기산 또는 이의 염의 대사적 생산에 관여하는 제 2의 고유 효소를 엔코딩하는 제 2의 고유 유전자가 부분적으로, 실질적으로 또는 완전히 결실되거나 침묵되거나 비활성화되거나 하향조절되는 유전자 변형된 미생물.

#### 청구항 63

제 62항에 있어서, 상기 제 2의 고유 효소가 아세테이트 키나아제 또는 포스포트랜스아세틸라아제인 유전자 변형된 미생물.

#### 청구항 64

제 62항에 있어서, 상기 제 2의 고유 효소가 락테이트 데히드로게나아제인 유전자 변형된 미생물.

#### 청구항 65

유전자 변형된 호열성 또는 중온성 미생물로서, (a) 유기산 또는 이의 염의 대사적 생산에 관여하는 제 1의 고유 효소를 엔코딩하는 제 1의 고유 유전자가 부분적으로, 실질적으로 또는 완전히 결실되거나 침묵되거나 비활성화되거나 하향조절되고, (b) 다당류의 가수분해에 관여하는 제 1의 비고유 효소를 엔코딩하는 제 1의 비고유 유전자가 삽입됨으로써, 상기 호열성 또는 중온성 미생물이 발효 생산물로서 에탄올을 생산할 수 있게 하는 유전자 변형된 호열성 또는 중온성 미생물.

#### 청구항 66

제 65항에 있어서, 상기 제 1의 비고유 유전자가 셀룰로오스를 가수분해하는 능력을 부여하는 제 1의 비고유 효소를 엔코딩함으로써, 상기 호열성 또는 중온성 미생물이 셀룰로오스를 가수분해할 수 있게 하는 유전자 변형된 미생물.

#### 청구항 67

제 65항에 있어서, 상기 제 1의 비고유 유전자가 자일란을 가수분해하는 능력을 부여하는 제 1의 비고유 효소를 엔코딩함으로써, 상기 호열성 또는 중온성 미생물이 자일란을 가수분해할 수 있게 하는 유전자 변형된 미생물.

#### 청구항 68

제 65항에 있어서, 상기 제 1의 비고유 유전자가 셀룰로오스를 가수분해하는 능력을 부여하는 제 1의 비고유 효소를 엔코딩하고; 자일란을 가수분해하는 능력을 부여하는 제 2의 비고유 효소를 엔코딩하는 제 2의 비고유 유전자가 삽입됨으로써, 상기 호열성 또는 중온성 미생물이 셀룰로오스와 자일란을 가수분해할 수 있게 하는 유전자 변형된 미생물.

#### 청구항 69

제 65항 내지 제 68항 중 어느 한 항에 있어서, 상기 유기산이 락트산인 유전자 변형된 미생물.

#### 청구항 70

제 65항 내지 제 68항 중 어느 한 항에 있어서, 상기 유기산이 아세트산인 유전자 변형된 미생물.

#### 청구항 71

제 65항, 제 66항 또는 제 68항 중 어느 한 항에 있어서, 상기 제 1의 비고유 효소가 피루베이트 데카르복실라아제(PDC) 또는 알코올 데히드로게나아제(ADH)인 유전자 변형된 미생물.

#### 청구항 72

제 49항 내지 제 71항 중 어느 한 항에 있어서, 상기 미생물이 대사된 바이오매스로부터의 탄소 중 60% 이상을 에탄올로 전환시킬 수 있는 유전자 변형된 미생물.

#### 청구항 73

제 9항, 제 11항, 제 12항, 제 14항, 제 16항, 제 18항, 제 19항, 제 22항, 제 33항, 제 34항, 제 35항, 제 36항, 제 49항, 제 65항, 제 66항, 제 67항 또는 제 68항 중 어느 한 항에 있어서, 상기 미생물이 중온성 미생물인 유전자 변형된 미생물.

#### 청구항 74

제 9항, 제 11항, 제 12항, 제 13항, 제 14항, 제 19항, 제 21항, 제 22항, 제 33항, 제 34항, 제 35항, 제 36항, 제 49항, 제 65항, 제 66항, 제 67항 또는 제 68항 중 어느 한 항에 있어서, 상기 미생물이 호열성 미생물인 유전자 변형된 미생물.

#### 청구항 75

리그노셀룰로오스(lignocelulosic) 바이오매스를 제 9항 내지 제 48항 중 어느 한 항에 따른 유전자 변형된 호열성 또는 중온성 미생물과 접촉시키는 것을 포함하여, 리그노셀룰로오스 바이오매스를 에탄올로 전환시키는 방법.

#### 청구항 76

제 75항에 있어서, 상기 리그노셀룰로오스 바이오매스가 그래스(grass), 스위치 그래스(switch grass), 코드 그래스(cord grass), 라이 그래스(rye grass), 리드 카나리 그래스(reed canary grass), 혼합형 프레어리 그래스(mixed prairie grass), 미스칸투스(miscanthus), 설탕-가공 잔류물(sugar-processing residue), 사탕수수 버개스(sugarcane bagasse), 사탕수수짚(sugarcane straw), 농업 폐기물, 벃짚(rice straw), 왕겨(rice hull), 보리짚(barley straw), 옥수수 속대(corn cob), 곡물짚(cereal straw), 밀짚(wheat straw), 카놀라짚(canola straw), 귀리짚(oat straw), 귀리껍질(oat hull), 옥수수 섬유, 스토버(stover), 대두 스토버, 옥수수 스토버, 산림 폐기물(forestry waste), 리사이클링된 목재 펄프 섬유, 종이 슬러지(paper sludge), 톱밥(sawdust), 경질목재(hardwood), 연질목재(softwood), 및 이들의 조합물로 구성된 군으로부터 선택되는 것인 방법.



#### 청구항 77

제 75항에 있어서, 상기 리그노셀룰로오스 바이오매스가 옥수수 스토버, 사탕수수 버개스, 스위치그래스 및 포플러 목재(poplar wood)로 구성된 군으로부터 선택되는 것인 방법.

#### 청구항 78

제 75항에 있어서, 상기 리그노셀룰로오스 바이오매스가 옥수수 스토버인 방법.

#### 청구항 79

제 75항에 있어서, 상기 리그노셀룰로오스 바이오매스가 사탕수수 버개스인 방법.

#### 청구항 80

제 75항에 있어서, 상기 리그노셀룰로오스 바이오매스가 스위치그래스인 방법.

#### 청구항 81

제 75항에 있어서, 상기 리그노셀룰로오스 바이오매스가 포플러 목재인 방법.

#### 청구항 82

제 75항에 있어서, 상기 리그노셀룰로오스 바이오매스가 버드나무(willow)인 방법.

#### 청구항 83

제 75항에 있어서, 상기 리그노셀룰로오스 바이오매스가 종이 슬러지인 방법.

#### 청구항 84

리그노셀룰로오스 바이오매스를 제 49항 내지 제 74항 중 어느 한 항에 따른 유전자 변형된 호열성 또는 중온성 미생물과 접촉시키는 것을 포함하여, 리그노셀룰로오스 바이오매스를 에탄올로 전환시키는 방법.

#### 청구항 85

제 84항에 있어서, 상기 리그노셀룰로오스 바이오매스가 그래스, 스위치 그래스, 코드 그래스, 라이 그래스, 리드 카나리 그래스, 혼합형 프레어리 그래스, 미스칸투스, 설탕-가공 잔류물, 사탕수수 버개스, 사탕수수짚, 농업 폐기물, 볏짚, 왕겨, 보리짚, 옥수수 속대, 곡물짚, 밀짚, 카놀라짚, 귀리짚, 귀리껍질, 옥수수 섬유, 스토버, 대두 스토버, 옥수수 스토버, 산림 폐기물, 리사이클링된 목재 펄프 섬유, 종이 슬러지, 톱밥, 경질목재, 연질목재, 및 이들의 조합물로 구성된 군으로부터 선택되는 것인 방법.

#### 청구항 86

제 84항에 있어서, 상기 리그노셀룰로오스 바이오매스가 옥수수 스토버, 사탕수수 버개스, 스위치그래스 및 포플러 목재로 구성된 군으로부터 선택되는 것인 방법.

#### 청구항 87

제 84항에 있어서, 상기 리그노셀룰로오스 바이오매스가 옥수수 스토버인 방법.

#### 청구항 88

제 84항에 있어서, 상기 리그노셀룰로오스 바이오매스가 사탕수수 버개스인 방법.

#### 청구항 89

제 84항에 있어서, 상기 리그노셀룰로오스 바이오매스가 스위치그래스인 방법.

#### 청구항 90

제 84항에 있어서, 상기 리그노셀룰로오스 바이오매스가 포플러 목재인 방법.

## 청구항 91

제 84항에 있어서, 상기 리그노셀룰로오스 바이오매스가 버드나무인 방법.

## 청구항 92

제 84항에 있어서, 상기 리그노셀룰로오스 바이오매스가 종이 슬러지인 방법.

## 명세서

### 발명의 상세한 설명

[0001] 관련 출원

[0002] 본 출원은 2007년 5월 9일에 출원된 미국 가특허출원 일련 번호 60/916,978을 우선권으로 주장하며, 상기 가특허출원은 그 전문이 본원에 참조로 포함된다.

[0003] 발명의 배경

[0004] 에너지 전환, 이용 및 접근이 지속가능성, 환경의 질(environmental quality), 안전 및 빈곤과 관련된 난제를 포함하는 현시대의 커다란 난제들 중 다수의 근거에 존재한다. 신형 기술의 새로운 적용이 이러한 난제들에 대응하기 위해 필요하다. 신형 기술 중 가장 강력한 기술 중의 하나인 생물공학은 중요한 새로운 에너지 전환 공정을 초래할 수 있다. 식물 바이오매스와 이의 유도체는 에너지를 인류에게 유용한 형태로 생물학적으로 전환시키기 위한 자원이다.

[0005] 식물 바이오매스의 형태 중에서, 리그노셀룰로오스(lignocellulosic) 바이오매스 ('바이오매스')가 이의 대규모 이용가능성, 낮은 가격 및 환경친화적(environmentally benign) 생산으로 인해 에너지 적용을 위해 특히 적합하다. 특히, 셀룰로오스 바이오매스를 기반으로 하는 많은 에너지 생산 및 이용 사이클은 라이프-사이클(life-cycle)을 기준으로 하여 온실 가스 배출물이 거의 0이다. 바이오매스 공급원료로부터의 보다 광범위한 에너지 생산을 저해하는 주요 장애요인은 이러한 재료가 유용한 연료로 전환되지 않게 하는 특성을 극복하기 위한 저렴한 기술이 대체로 없다는 것이다. 리그노셀룰로오스 바이오매스는 에탄올로 전환될 수 있는 탄수화물 분획 (예를 들어, 셀룰로오스와 헤미셀룰로오스)를 함유한다. 이러한 분획을 전환하기 위해, 셀룰로오스와 헤미셀룰로오스는 궁극적으로 당당류로 전환되거나 가수분해되어야 하는데, 역사적으로 문제가 되는 것으로 판명되어온 것이 바로 가수분해이다.

[0006] 생물학적으로 매개되는 공정은 에너지 전환을 위해 유망하며, 특히 리그노셀룰로오스 바이오매스를 연료로 전환시키는 데에 있어서 유망하다. 효소 또는 미생물에 의한 가수분해를 포함하는 바이오매스 처리 계획은 일반적으로 다음과 같은 4가지의 생물학적으로 매개되는 변환을 포함한다: (1) 당분해(saccharolytic) 효소 (셀룰라아제 및 헤미셀룰라아제)의 생산; (2) 전처리된 바이오매스에 존재하는 탄수화물 성분의 당(sugar)으로의 가수분해; (3) 핵소오스 당 (예를 들어, 글루코오스, 만노오스 및 갈락토오스)의 발효; 및 (4) 펜토오스 당 (예를 들어, 자일로오스 및 아라비노오스)의 발효. 이러한 4가지 변환은 통합된 생물가공(consolidated bioprocessing, CBP)으로 일컬어지는 공정 구성에서 단일 단계로 일어나는데, 이는 셀룰라아제 및/또는 헤미셀룰라아제 생산을 위한 전용 공정 단계를 포함하지 않는다는 점에서 다른 덜 고도로 집적된 구성과 구별된다.

[0007] CBP는 전용 셀룰라아제 생산을 특징으로 하는 공정에 비해 저렴하고 효율이 높다는 잠재력을 제공한다. 상기 이점은 부분적으로는 자본 비용, 기질 및 다른 원료 물질, 및 셀룰라아제 생산과 관련된 설비(utility)가 회피됨에 따른 것이다. 또한, 효소-미생물 상승작용 및 호열성(thermophilic) 생물체 및/또는 복합 셀룰라아제 시스템의 사용을 포함하는 수 가지 요인들이 높은 가수분해율의 실현, 및 이에 따른 CBP의 사용시의 반응기 부피 및 자본 투자의 감소를 지원한다. 더욱이, 셀룰로오스-부착성(cellulose-adherent) 셀룰로오스분해성(cellulolytic) 미생물은 부착되지 않은 미생물, 예를 들어 오염균(contaminant)과 셀룰로오스 가수분해의 생산물에 대해 성공적으로 경합하는 것으로 여겨지며, 이는 미생물에 의한 셀룰로오스 이용을 기반으로 하는 산업적 공정의 안정성을 증가시킬 수 있다. CBP를 가능하게 하는 미생물을 개발하는 데에 있어서 다음의 2가지 전략을 통해 진전이 이루어지고 있다: 천연 셀룰로오스분해성 미생물을 엔지니어링(engineering)하여, 수율 및 역가와 같은 생산물 관련된 특성을 개선시키려는 전략; 및 높은 생산물 수율 및 역가를 나타내는 비-셀룰로오스분해성(non-cellulolytic) 생물체를 엔지니어링하여, 이종성(heterologous) 셀룰로오스 및 헤미셀룰로오스 이용을 가능하게 하는 셀룰라아제 및 헤미셀룰라아제 시스템을 발현하게 하려는 전략.

- [0008] 많은 세균은 단순 헥소오스 당을 해당과정(glycolysis)을 통해 산성 생산물과 pH-중성 생산물의 혼합물로 발효시키는 능력을 지닌다. 해당과정 경로(glycolytic pathway)는 풍부하고, 일련의 효소에 의한 단계들을 포함하는데, 이에 의해 탄소 6개의 글루코오스 분자가 다수의 중간체들을 통해 탄소 3개의 화합물인 피루베이트 2분자로 분해된다. 이러한 공정은 ATP (생물학적 에너지 공급원) 및 환원된 보조인자인 NADH를 궁극적으로 생성시킨다.
- [0009] 피루베이트는 대사의 중요한 중간 화합물이다. 예를 들어, 호기성 조건하에서 피루베이트는 아세틸 조효소 A (아세틸 CoA)로 산화될 수 있는데, 그 다음 이는 트리카르복실산 사이클(TCA)로 들어가고, 이어서 합성 전구체, CO<sub>2</sub> 및 환원된 보조인자를 생성시킨다. 그 후, 상기 보조인자는 일련의 효소에 의한 단계들을 통해 수소 당량을 산소에게 내줌으로써 산화되어, 물과 ATP를 형성한다. 이러한 에너지 형성 공정은 산화적 인산화로서 공지되어 있다.
- [0010] 혐기성 조건 (산소가 이용가능하지 않음)하에서, 유기 화합물의 분해 생산물이 수소 공여체 및 수용체로서 기능하는 발효가 일어난다. 해당과정으로부터의 과량의 NADH가 유기 기질을 락테이트와 에탄올과 같은 생산물로 환원시키는 것을 포함하는 반응으로 산화된다. 또한, ATP는 기질 수준 인산화로서 공지된 공정으로 아세테이트와 같은 유기산의 생산으로부터 재생된다. 따라서, 해당과정 및 피루베이트 대사의 발효 생산물은 다양한 유기산, 알코올 및 CO<sub>2</sub>를 포함한다.
- [0011] 대다수의 통성 혐기성 세균은 호기성 또는 혐기성 조건하에서 에탄올을 고수율로 생산하지 않는다. 대부분의 통성 혐기성생물(anaerobe)은 피루베이트 데히드로게나아제(PDH)와 트리카르복실산 사이클(TCA)을 통해 혐기적으로 피루베이트를 대사한다. 혐기성 조건하에서, 피루베이트의 대사를 위한 주요 에너지 경로는 포르메이트와 아세틸-CoA를 제공하는 피루베이트-포르메이트-리아제(pyruvate-formate-lyase, PFL)를 통해 이루어진다. 그 후, 아세틸-CoA는, 포스포트랜스아세틸라아제(PTA) 및 아세테이트 키나아제(ACK)를 통해 아세테이트로 전환되며 ATP를 동시에 생산하거나, 아세트알데히드 데히드로게나아제(AcDH)와 알코올 데히드로게나아제(ADH)를 통해 에탄올로 환원된다. 환원 당량(reducing equivalent)의 균형을 유지하기 위해, 해당과정으로부터 생산된 과량의 NADH는, 피루베이트가 락테이트로 환원되는 동안 락테이트 데히드로게나아제(LDH)에 의해 NAD<sup>+</sup>로 재산화된다. 또한, NADH는 아세틸-CoA가 에탄올로 환원되는 동안 AcDH와 ADH에 의해 재산화될 수 있지만, 이는 기능성 LDH를 지닌 세포에서는 경미한 반응이다. 따라서, 에탄올의 이론적 수율이 달성되지 않는데, 이는 ATP를 재생시키기 위해 대부분의 아세틸 CoA가 아세테이트로 전환되고, 해당과정 동안 생산된 과량의 NADH가 LDH에 의해 산화되기 때문이다.
- [0012] 또한, 미생물의 대사적 엔지니어링은 락테이트 데히드로게나아제와 같은 효소의 생산을 엔코딩하는 유전자의 표적화된 녹아웃(targeted knockout)의 발생을 초래할 수 있다. 이러한 경우, 유전자의 "녹 아웃(knock out)"은 부분적, 실질적 또는 완전한, 결실, 침묵(silencing), 비활성화 또는 하향조절(down-regulation)을 의미한다. LDH의 작용에 의한 피루베이트의 락테이트 (락트산의 염 형태)로의 전환이 해당과정 경로의 초기 단계에 이용가능하지 않은 경우, 피루베이트는 피루베이트 데히드로게나아제 또는 피루베이트-페레독신 옥시도리덕타아제(oxidoreductase)의 작용에 의해 보다 효율적으로 아세틸 CoA로 전환될 수 있다. 포스포트랜스아세틸라아제와 아세테이트 키나아제에 의한 아세틸 CoA의 아세테이트 (아세트산의 염 형태)로의 추가의 전환이 또한 이용가능하지 않은 경우, 즉, PTA와 ACK의 생산을 엔코딩하는 유전자가 녹아웃된 경우, 아세틸 CoA는 AcDH 및 ADH에 의해 보다 효율적으로 에탄올로 전환될 수 있다. 따라서, 유기산의 생산을 제거하는 그러한 표적화된 유전자 녹아웃을 지닌 유전자 변형된 미생물 균주(strain)는 발효 생산물로서 에탄올을 생산하는 능력이 증가할 것이다.
- [0013] 에탄올생산성(ethanologenic) 생물체, 예를 들어 자이모모나스 모빌리스(*Zymomonas mobilis*), 자이모박터 팔매(*Zymobacter palmae*), 아세토박터 파스퇴리아누스(*Acetobacter pasteurianus*), 또는 사르시나 벤트리쿨리(*Sarcina ventriculi*), 및 일부 효모 (예를 들어, 사카로마이세스 세레비시아(*Saccharomyces cerevisiae*))는 일반적으로 알코올 발효라고 지칭되는 제2형의 혐기성 발효를 수행할 수 있으며, 여기서 피루베이트가 피루베이트 데카르복실라아제(PDC)에 의해 아세트알데히드와 CO<sub>2</sub>로 대사된다. 그 후, 아세트알데히드는 ADH에 의해 에탄올로 환원되며 NAD<sup>+</sup>를 재생시킨다. 알코올 발효는 글루코오스 1분자를 2분자의 에탄올과 2분자의 CO<sub>2</sub>로 대사시킨다. 상기 상세히 설명된 바와 같이 피루베이트가 요망되지 않는 유기산으로 전환되는 것이 회피될 수 있는 경우, 이러한 유전자 변형된 미생물은 발효 생산물로서 에탄올을 생산하는 능력이 증가할 것이다.

[0014] 발명의 개요

- [0015] 본 발명의 한 가지 일면은 SEQ ID NO:1-5, 30-31 및 47-61 중 어느 하나의 뉴클레오티드 서열, 또는 이의 상보 서열(complement)을 포함하는 분리된 핵산 분자에 관한 것이다. 본 발명의 또 다른 일면은 SEQ ID NO:1-5, 30-31 및 47-61 중 어느 하나의 뉴클레오티드 서열, 또는 이의 상보서열과 80% 이상의 동일성을 공유하는 뉴클레오티드 서열을 포함하는 분리된 핵산 분자에 관한 것이다. 특정 구체예에서, 본 발명은 SEQ ID NO:1-5, 30-31 및 47-61 중 어느 하나의 뉴클레오티드 서열, 또는 이의 상보서열과 약 95% 이상의 서열 동일성을 공유하는 상기 언급된 핵산 분자에 관한 것이다.
- [0016] 본 발명의 또 다른 일면은 호열성 또는 중온성(mesophilic) 세균에서 발현가능한 프로모터에 작동적으로 결합된 SEQ ID NO:1-5, 30-31 및 47-61 중 어느 하나를 포함하는 유전자 구성물(genetic construct)에 관한 것이다. 또한, 본 발명은 상기 언급된 유전자 구성물을 포함하는 재조합 호열성 또는 중온성 세균에 관한 것이다.
- [0017] 또한, 본 발명은 상기 언급된 핵산 분자 중 어느 하나를 포함하는 벡터를 포함한다. 또한, 본 발명은 상기 언급된 핵산 분자 중 어느 하나를 포함하는 숙주 세포를 포함한다. 특정 구체예에서, 본 발명은 호열성 또는 중온성 세균 세포인 상기 언급된 숙주 세포에 관한 것이다.
- [0018] 본 발명의 또 다른 일면은 유기산 또는 이의 염의 대사적 생산에 관여하는 제 1의 고유(native) 효소를 엔코딩하는 제 1의 고유 유전자가 부분적으로, 실질적으로 또는 완전히 결실되거나 침묵되거나 비활성화되거나 하향조절됨으로써, 발효 생산물로서 에탄올을 생산하는 고유 능력이 증가된, 유전자 변형된 호열성 또는 중온성 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 그람-음성 세균 또는 그람-양성 세균인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 써모아나에어로박테리움(*Thermoanaerobacterium*), 써모아나에어로박터(*Thermoanaerobacter*), 클로스트리디움(*Clostridium*), 게오바실루스(*Geobacillus*), 사카로코쿠스(*Saccharococcus*), 페니바실루스(*Paenibacillus*), 바실루스(*Bacillus*), 칼디셀룰로시럽터(*Caldicellulosiruptor*), 아나에어로셀룸(*Anaerocellum*), 또는 안옥시바실루스(*Anoxybacillus*) 속(genus)의 종(species)인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 써모아나에어로박테리움 써모술푸리게네스(*Thermoanaerobacterium thermosulfurigenes*), 써모아나에어로박테리움 아오테아로엔세(*Thermoanaerobacterium aotearoense*), 써모아나에어로박테리움 폴리사카롤리티쿰(*Thermoanaerobacterium polysaccharolyticum*), 써모아나에어로박테리움 제아(*Thermoanaerobacterium zae*), 써모아나에어로박테리움 자일라노리티쿰(*Thermoanaerobacterium xylanolyticum*), 써모아나에어로박테리움 사카롤리티쿰(*Thermoanaerobacterium saccharolyticum*), 써모아나에어로비움 브록키(*Thermoanaerobium brockii*), 써모아나에어로박테리움 써모사카롤리티쿰(*Thermoanaerobacterium thermosaccharolyticum*), 써모아나에어로박터 써모히드로숄푸리쿠스(*Thermoanaerobacter thermohydrosulfuricus*), 써모아나에어로박터 에타놀리쿠스(*Thermoanaerobacter ethanolicus*), 써모아나에어로박터 브록키(*Thermo anaerobacter brocki*), 클로스트리디움 써모셀룸(*Clostridium thermocellum*), 클로스트리디움 셀룰로리티쿰(*Clostridium cellulolyticum*), 클로스트리디움 피토펙페르멘탄스(*Clostridium phytofermentans*), 클로스트리디움 스트라미노솔벤스(*Clostridium straminosolvens*), 게오바실루스 써모글루코시다시우스(*Geobacillus thermoglucosidasius*), 게오바실루스 스테아로써모필루스(*Geobacillus stearothermophilus*), 사카로코쿠스 칼독실로실리티쿠스(*Saccharococcus caldxylosilyticus*), 사카로쿠스 써모필루스(*Saccharoccus thermophilus*), 페니바실루스 캄피나센시스(*Paenibacillus campinasensis*), 바실루스 플라보써무스(*Bacillus flavothermus*), 안옥시바실루스 캄차트켄시스(*Anoxybacillus kamchatkensis*), 안옥시바실루스 고넨시스(*Anoxybacillus gonensis*), 칼디셀룰로시럽터 아세티게누스(*Caldicellulosiruptor acetigenus*), 칼디셀룰로시럽터 사카롤리티쿠스(*Caldicellulosiruptor saccharolyticus*), 칼디셀룰로시럽터 크리스티안소니(*Caldicellulosiruptor kristjanssonii*), 칼디셀룰로시럽터 오웬센시스(*Caldicellulosiruptor owensensis*), 칼디셀룰로시럽터 락토아세티쿠스(*Caldicellulosiruptor lactoaceticus*), 및 아나에어로셀룸 써모필룸(*Anaerocellum thermophilum*)로 구성된 군으로부터 선택된 세균인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 써모아나에어로박테리움 사카롤리티쿰인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 (a) 핵소오스 당을 대사하는 고유 능력을 지닌 호열성 또는 중온성 미생물; (b) 펜토오스 당을 대사하는 고유 능력을 지닌 호열성 또는 중온성 미생물; 및 (c) 핵소오스 당과 펜토오스 당을 대사하는 고유 능력을 지닌 호열성 또는 중온성 미생물로 구성된 군으로부터 선택되는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 핵소오스 당을 대사하는 고유 능력을 지니는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 클로스트리디움 스트라미노솔벤스(*Clostridium straminisolvens*) 또는 클로스트리디움 써모셀룸인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 클로스트리디움 셀룰로리티쿰



(*Clostridium cellulolyticum*), 칼디셀룰로시쿠티 크리스티안소니, 또는 클로스트리디움 스테르코라리움 아종 랩토사프라르툼(*Clostridium stercorarium subsp. leptosaprartum*)인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 펜토오스 당을 대사하는 능력을 부여하는 제 1의 비고유(non-native) 효소를 엔코딩하는 제 1의 비고유 유전자가 삽입됨으로써, 상기 호열성 또는 중온성 미생물이 펜토오스 당으로부터 발효 생산물로서 에탄올을 생산할 수 있게 하는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 펜토오스 당을 대사하는 고유 능력을 지니는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 써모안에어로박테리움 사카롤리티쿰, 써모안에어로박테리움 자일라노리티쿰, 써모안에어로박테리움 폴리스카롤리티쿰, 및 써모안에어로박테리움 써모사카롤리티쿰으로 구성된 군으로부터 선택되는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 헥소오스 당을 대사하는 능력을 부여하는 제 1의 비고유 효소를 엔코딩하는 제 1의 비고유 유전자가 삽입됨으로써, 상기 호열성 또는 중온성 미생물이 헥소오스 당으로부터 발효 생산물로서 에탄올을 생산할 수 있게 하는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 유기산이 락트산 및 아세트산으로 구성된 군으로부터 선택되는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 유기산이 락트산인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 유기산이 아세트산인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 1의 고유 효소가 락테이트 데히드로게나아제, 아세테이트 키나아제 및 포스포트랜스아세틸라아제로 구성된 군으로부터 선택되는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 1의 고유 효소가 락테이트 데히드로게나아제인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 1의 고유 효소가 아세테이트 키나아제인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 1의 고유 효소가 포스포트랜스아세틸라아제인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 유기산 또는 이의 염의 대사적 생산에 관여하는 제 2의 고유 효소를 엔코딩하는 제 2의 고유 유전자가 부분적으로, 실질적으로 또는 완전히 결실되거나 침묵되거나 비활성화되거나 하향조절되는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 제 2의 고유 효소가 아세테이트 키나아제 또는 포스포트랜스아세틸라아제인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 제 2의 고유 효소가 락테이트 데히드로게나아제인 상기 언급된 유전자 변형된 미생물에 관한 것이다.

[0019]

본 발명의 또 다른 일면은 유전자 변형된 호열성 또는 중온성 미생물로서, (a) 유기산 또는 이의 염의 대사적 생산에 관여하는 제 1의 고유 효소를 엔코딩하는 제 1의 고유 유전자가 부분적으로, 실질적으로 또는 완전히 결실되거나 침묵되거나 비활성화되거나 하향조절되고, (b) 에탄올의 대사적 생산에 관여하는 제 1의 비고유 효소를 엔코딩하는 제 1의 비고유 유전자가 삽입됨으로써, 상기 호열성 또는 중온성 미생물이 발효 생산물로서 에탄올을 생산할 수 있게 하는, 유전자 변형된 호열성 또는 중온성 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 1의 비고유 유전자가 헥소오스 당을 대사하는 능력을 부여하는 제 1의 비고유 효소를 엔코딩함으로써, 상기 호열성 또는 중온성 미생물이 헥소오스 당을 대사할 수 있게 하는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 1의 비고유 유전자가 펜토오스 당을 대사하는 능력을 부여하는 제 1의 비고유 효소를 엔코딩함으로써, 상기 호열성 또는 중온성 미생물이 펜토오스 당을 대사할 수 있게 하는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 1의 비고유 유전자가 헥소오스 당을 대사하는 능력을 부여하는 제 1의 비고유 효소를 엔코딩하고; 펜토오스 당을 대사하는 능력을 부여하는 제 2의 비고유 효소를 엔코딩하는 제 2의 비고유 유전자가 삽입됨으로써, 상기 호열성 또는 중온성 미생물이 헥소오스 당과 펜토오스 당을 대사할 수 있게 하는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 유기산이 락트산인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 유기산이 아세트산인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 1의 비고유 효소가 피루베이트 데카르복실라아제(PDC) 또는 알코올 데히드로게나아제(ADH)인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 2의 비고유 효소가 자일로오스 이소머라아제인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 1의 비고유 유전자가 SEQ ID NO:6, 10 또는 14에 상응하는 것인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 비고유 효소가 자일룰로키나아제(xylulokinase)인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 비고유 유전자가 SEQ ID NO:7, 11 또는 15에 상응하는 것인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 비고유 효소가 L-아라비노오스 이소머라아제인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 비고유 유전자가 SEQ ID NO:8 또는 12에 상응하는 것인 상

기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 비고유 효소가 L-리불로오스-5-포스페이트 4-에피머라아제인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 비고유 유전자가 SEQ ID NO:9 또는 13에 상응하는 것인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 대사된 바이오매스로부터의 탄소 중 60% 이상을 에탄올로 전환시킬 수 있는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 셀룰로오스를 가수분해하는 고유 능력을 지닌 호열성 또는 중온성 미생물 (a), 자일란(xylan)을 가수분해하는 고유 능력을 지닌 호열성 또는 중온성 미생물 (b) 및 셀룰로오스와 자일란을 가수분해하는 고유 능력을 지닌 호열성 또는 중온성 미생물 (c)로 구성된 군으로부터 선택되는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 셀룰로오스를 가수분해하는 고유 능력을 지니는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 셀룰로오스 및 자일란을 가수분해하는 고유 능력을 지니는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 자일란을 가수분해하는 능력을 부여하는 제 1의 비고유 효소를 엔코딩하는 제 1의 비고유 유전자가 삽입되어 있는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 자일란을 가수분해하는 고유 능력을 지니는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 셀룰로오스를 가수분해하는 능력을 부여하는 제 1의 비고유 효소를 엔코딩하는 제 1의 비고유 유전자가 삽입되어 있는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 유기산이 락트산 및 아세트산으로 구성된 군으로부터 선택되는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 유기산이 락트산인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 유기산이 아세트산인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 1의 고유 효소가 락테이트 데히드로게나아제, 아세테이트 키나아제 및 포스포트랜스아세틸라아제로 구성된 군으로부터 선택되는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 1의 고유 효소가 락테이트 데히드로게나아제인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 1의 고유 효소가 아세테이트 키나아제인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 1의 고유 효소가 포스포트랜스아세틸라아제인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 유기산 또는 이의 염의 대사적 생산에 관여하는 제 2의 고유 효소를 엔코딩하는 제 2의 고유 유전자가 부분적으로, 실질적으로 또는 완전히 결실되거나 침묵되거나 비활성화되거나 하향조절되는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 2의 고유 효소가 아세테이트 키나아제 또는 포스포트랜스아세틸라아제인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 2의 고유 효소가 락테이트 데히드로게나아제인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 (a) 유기산 또는 이의 염의 대사적 생산에 관여하는 제 1의 고유 효소를 엔코딩하는 제 1의 고유 유전자가 부분적으로, 실질적으로 또는 완전히 결실되거나 침묵되거나 비활성화되거나 하향조절되고, (b) 다당류의 가수분해에 관여하는 제 1의 비고유 효소를 엔코딩하는 제 1의 비고유 유전자가 삽입됨으로써, 상기 호열성 또는 중온성 미생물이 발효 생산물로서 에탄올을 생산할 수 있게 하는, 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 1의 비고유 유전자가 셀룰로오스를 가수분해하는 능력을 부여하는 제 1의 비고유 효소를 엔코딩함으로써, 상기 호열성 또는 중온성 미생물이 셀룰로오스를 가수분해할 수 있게 하는, 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 1의 비고유 유전자가 자일란을 가수분해하는 능력을 부여하는 제 1의 비고유 효소를 엔코딩함으로써, 상기 호열성 또는 중온성 미생물이 자일란을 가수분해할 수 있게 하는 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 1의 비고유 유전자가 셀룰로오스를 가수분해하는 능력을 부여하는 제 1의 비고유 효소를 엔코딩하고; 자일란을 가수분해하는 능력을 부여하는 제 2의 비고유 효소를 엔코딩하는 제 2의 비고유 유전자가 삽입됨으로써, 상기 호열성 또는 중온성 미생물이 셀룰로오스와 자일란을 가수분해할 수 있게 하는, 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 유기산이 락트산인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 유기산이 아세트산인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 상기 제 1의 비고유 효소가 피루베이트 데카르복실라아제(PDC) 또는 알코올 데히드로게나아제(ADH)인 상기 언급된 유전자 변형된 미생물에 관한 것이다. 특정 구체예에서, 본 발명은 대사된 바이오매스로부터의 탄소 중 60% 이상을 에탄올로 전환시킬 수 있는 상기 언급된 유전자 변형된 미생물에 관한 것이다.

[0020] 특정 구체예에서, 본 발명은 중온성인 상기 언급된 유전자 변형된 미생물 중의 어느 하나에 관한 것이다. 특정 구체예에서, 본 발명은 호열성인 상기 언급된 유전자 변형된 미생물 중 어느 하나에 관한 것이다.

[0021] 본 발명의 또 다른 일면은 리그노셀룰로오스 바이오매스를 상기 언급된 유전자 변형된 호열성 또는 중온성 미생

물과 접촉시키는 것을 포함하여, 리그노셀룰로오스 바이오매스를 에탄올로 전환시키는 방법에 관한 것이다. 특정 구체예에서, 본 발명은 상기 리그노셀룰로오스 바이오매스가 그래스(grass), 스위치 그래스(switch grass), 코드 그래스(cord grass), 라이 그래스(rye grass), 리드 카나리 그래스(reed canary grass), 혼합형 프레어리 그래스(mixed prairie grass), 미스칸투스(miscanthus), 설탕-가공 잔존물(sugar-processing residue), 사탕수수 버개스(sugarcane bagasse), 사탕수수짚(sugarcane straw), 농업 폐기물, 벃짚(rice straw), 왕겨(rice hull), 보리짚(barley straw), 옥수수 속대(corn cob), 곡물짚(cereal straw), 밀짚(wheat straw), 카놀라짚(canola straw), 귀리짚(oat straw), 귀리껍질(oat hull), 옥수수 섬유, 스토버(stover), 대두 스토버, 옥수수 스토버, 산림 폐기물(forestry waste), 리사이클링된 목재 펄프 섬유, 종이 슬러지(paper sludge), 톱밥(sawdust), 경질목재(hardwood), 연질목재(softwood), 및 이들의 조합물로 구성된 군으로부터 선택되는 것인 상기 언급된 방법에 관한 것이다. 특정 구체예에서, 본 발명은 상기 리그노셀룰로오스 바이오매스가 옥수수 스토버, 사탕수수 버개스, 스위치그래스 및 포플러 목재(poplar wood)로 구성된 군으로부터 선택되는 것인 상기 언급된 방법에 관한 것이다. 특정 구체예에서, 본 발명은 상기 리그노셀룰로오스 바이오매스가 옥수수 스토버인 상기 언급된 방법에 관한 것이다. 특정 구체예에서, 본 발명은 상기 리그노셀룰로오스 바이오매스가 사탕수수 버개스인 상기 언급된 방법에 관한 것이다. 특정 구체예에서, 본 발명은 상기 리그노셀룰로오스 바이오매스가 스위치그래스인 상기 언급된 방법에 관한 것이다. 특정 구체예에서, 본 발명은 상기 리그노셀룰로오스 바이오매스가 포플러 목재인 상기 언급된 방법에 관한 것이다. 특정 구체예에서, 본 발명은 상기 리그노셀룰로오스 바이오매스가 버드나무(willow)인 상기 언급된 방법에 관한 것이다. 특정 구체예에서, 본 발명은 상기 리그노셀룰로오스 바이오매스가 종이 슬러지인 상기 언급된 방법에 관한 것이다.

[0022] 도면의 간단한 설명

[0023] 도 1은 해당과정 경로를 도시한다.

[0024] 도 2는 펜토오스 및 글루쿠로네이트 상호전환을 도시하며, D-자일로오스 → 에탄올 경로에서 자일로오스 이소머라아제 (XI 또는 5.3.1.5) 및 자일롤로키나아제 (XK 또는 2.7.1.17) 효소를 강조하고 있다.

[0025] 도 3은 펜토오스 및 글루쿠로네이트 상호전환을 도시하며, L-아라비노오스 이용 경로에서 L-아라비노오스 이소머라아제 (5.3.1.4) 및 L-리불로오스-5-포스페이트 4-에피머라아제 (5.1.3.4) 효소를 강조하고 있다.

[0026] 도 4는 펜토오스 및 글루쿠로네이트 상호전환을 도시하며, 클로스트리디움 셀룰로리티쿰에 자일로오스 이소머라아제, 자일롤로키나아제, L-아라비노오스 이소머라아제 및 L-리불로오스-5-포스페이트 4-에피머라아제에 대한 유전자가 존재함을 보여준다.

[0027] 도 5는 펜토오스 및 글루쿠로네이트 상호전환을 도시하며, 클로스트리디움 피토펙르멘탄스에 자일로오스 이소머라아제 및 자일롤로키나아제는 존재하지만 L-아라비노오스 이소머라아제 및 L-리불로오스-5-포스페이트 4-에피머라아제는 존재하지 않음을 보여준다.

[0028] 도 6은 클로스트리디움 씨모셀룸 (SEQ ID NO: 77), 클로스트리디움 셀룰로리티쿰 (SEQ ID NO: 78), 씨모안에어 로박테리움 사카롤리티쿰 (SEQ ID NO: 79), 클로스트리디움 스테르코라리움 (SEQ ID NO: 80), 클로스트리디움 스테르코라리움 II (*C. stercorarium* II) (SEQ ID NO: 81), 칼디셀룰로시룸터 크리스티안소니 (SEQ ID NO: 82), 클로스트리디움 피토펙르멘탄스 (SEQ ID NO: 83)을 정렬한 도면이며, 이는 16S rDNA 유전자의 수준에서 약 73 내지 89% 상동성을 나타낸다.

[0029] 도 7은 플라스미드 pIKM1을 기반으로 하는, 클로스트리디움 씨모셀룸에서의 *ack* 유전자의 비활성화를 위한 이중 크로스오버(crossover) 녹아웃 벡터의 구성을 도시한다.

[0030] 도 8은 복제성(replicative) 플라스미드 pNW33N을 기반으로 하는, 클로스트리디움 씨모셀룸에서의 *ack* 유전자의 비활성화를 위한 이중 크로스오버 녹아웃 벡터의 구성을 도시한다.

[0031] 도 9는 플라스미드 pIKM1을 기반으로 하는, 클로스트리디움 씨모셀룸에서의 *ldh* 유전자의 비활성화를 위한 이중 크로스오버 녹아웃 벡터의 구성을 도시한다.

[0032] 도 10은 복제성 플라스미드 pNW33N을 기반으로 하는, 클로스트리디움 씨모셀룸에서의 *ldh* 유전자의 비활성화를 위한 이중 크로스오버 녹아웃 벡터의 구성을 도시한다.

[0033] 도 11은 플라스미드 pUC19를 기반으로 하는, 클로스트리디움 씨모셀룸에서의 *ldh* 유전자의 비활성화를 위한 이중 크로스오버 녹아웃 벡터의 구성을 도시한다.

- [0034] 도 12A 및 12B는 각각 셀로비오스(cellobiose) 및 아비셀®(Avicel®)에서 증식된 클로스트리디움 스트라미니슐 벤스에 대한 생산물 형성 및 OD<sub>600</sub>을 도시한다.
- [0035] 도 13A 및 13B는 각각 셀로비오스 및 아비셀®에서 증식된 클로스트리디움 씨모셀룸에 대한 생산물 형성 및 OD<sub>600</sub>을 도시한다.
- [0036] 도 14A 및 14B는 각각 셀로비오스 및 아비셀®에서 증식된 클로스트리디움 셀룰로리티쿰에 대한 생산물 형성 및 OD<sub>600</sub>을 도시한다.
- [0037] 도 15A 및 15B는 각각 셀로비오스 및 아비셀®에서 증식된 클로스트리디움 스테르코라리움 아중 랩토사파르툼 (*Clostridium stercorarium subsp. leptosapartum*)에 대한 생산물 형성 및 OD<sub>600</sub>을 도시한다.
- [0038] 도 16A 및 16B는 각각 셀로비오스 및 아비셀®에서 증식된 칼디셀룰로시룸터 크리스티안소니에 대한 생산물 형성 및 OD<sub>600</sub>을 도시한다.
- [0039] 도 17A 및 17B는 각각 셀로비오스 및 아비셀®에서 증식된 클로스트리디움 피토펬르멘탄스에 대한 생산물 형성 및 OD<sub>600</sub>을 도시한다.
- [0040] 도 18은 2.5 g/L의 자일란 및 2.5 g/L의 셀로비오스의 48시간 발효 후의 전체 대사 부산물을 도시한다.
- [0041] 도 19는 *ack* 유전자의 맵(map) 및 유전자 붕괴를 위해 PCR에 의해 증폭되는 영역을 도시한다.
- [0042] 도 20은 *ldh* 2262 유전자의 맵 및 유전자 붕괴를 위해 PCT에 의해 증폭되는 영역을 도시한다.
- [0043] 도 21은 클로스트리디움 셀룰로리티쿰 *ldh* (2262) 이중 크로스오버 녹아웃 단편의 예를 도시한다.
- [0044] 도 22는 클로스트리디움 피토펬르멘탄스의 *ack* 유전자의 맵 및 유전자 붕괴를 위해 PCR에 의해 증폭되는 영역을 도시한다.
- [0045] 도 23은 클로스트리디움 피토펬르멘탄스에서 선별가능 마커로서 mls 유전자를 지닌 추정 이중 크로스오버 녹아웃 구성물의 예를 도시한다.
- [0046] 도 24는 *ldh* 1389 유전자의 맵 및 유전자 붕괴를 위해 PCR에 의해 증폭되는 영역을 도시한다.
- [0047] 도 25는 선별가능 마커로서 mls 유전자를 지닌 추정 이중 크로스오버 녹아웃 구성물의 예를 도시한다.
- [0048] 도 26은 pMOD™-2<MCS>의 bp 250-550 (SEQ ID NO: 84)를 나타내는 도면이다.
- [0049] 도 27은 클로스트리디움 스트라미니슐벤스를 사용한 경우 1% 아비셀®에 대한 생산물 농도 프로파일을 도시한다. 에탄올 대 아세테이트 비는 E/A로 표시되고, 에탄올 대 전체 생산물의 비는 E/T로 표시되어 있다.
- [0050] 도 28은 *C. cell* ACK 유전자 (SEQ ID NO:21)에 삽입되도록 L1.LtrB 인트론을 리타겟팅(retargeting)하기 위한 벡터의 예를 도시한다.
- [0051] 도 29는 *C. cell* LDH2744 유전자 (SEQ ID NO:23)에 삽입되도록 L1.LtrB 인트론을 리타겟팅하기 위한 벡터의 예를 도시한다.
- [0052] 도 30은 16S rDNA 유전자의 수준에서 씨모안에어로박터 슈도에타놀리쿠스 39E (SEQ ID NO: 85), 씨모안에어로박테리움 중 균주(*T. sp* strain) 59 (SEQ ID NO: 86), 씨모안에어로박테리움 사카롤리티쿰(*T. saccharolyticum*) B6A-RI (SEQ ID NO: 87), 씨모안에어로박테리움 사카롤리티쿰 YS485 (SEQ ID NO: 88) 및 컨센서스(consensus) (SEQ ID NO: 89)를 정렬한 도면이다.
- [0053] 도 31은 pta 유전자의 수준에서 씨모안에어로박테리움 중 균주 59 (SEQ ID NO: 36), 씨모안에어로박터 슈도에타놀리쿠스 (SEQ ID NO: 35), 씨모안에어로박테리움 사카롤리티쿰 B6A-RI (SEQ ID NO: 38), 씨모안에어로박테리움 사카롤리티쿰 YS485 (SEQ ID NO: 32) 및 컨센서스 (SEQ ID NO: 90)를 정렬한 도면이다.
- [0054] 도 32는 *ack* 유전자의 수준에서 씨모안에어로박테리움 중 균주 59 (SEQ ID NO: 37), 씨모안에어로박터 슈도에타놀리쿠스 (SEQ ID NO: 34), 씨모안에어로박테리움 사카롤리티쿰 B6A-RI (SEQ ID NO: 39), 씨모안에어로박테리움 사카롤리티쿰 YS485 (SEQ ID NO: 33) 및 컨센서스 (SEQ ID NO: 91)를 정렬한 도면이다.
- [0055] 도 33은 *ldh* 유전자의 수준에서 씨모안에어로박테리움 중 균주 59 (SEQ ID NO: 41), 씨모안에어로박터 슈도에타



놀리쿠스 (SEQ ID NO: 42), 씨모안에어로박테리움 사카롤리티쿰 B6A-RI (SEQ ID NO: 43), 씨모안에어로박테리움 사카롤리티쿰 YS485 (SEQ ID NO: 40) 및 컨센서스 (SEQ ID NO: 92)를 정렬한 도면이다.

- [0056] 도 34는 해당과정/발효 경로를 개략적으로 도시한다.
- [0057] 도 35는 pMU340 플라스미드의 예를 도시한다.
- [0058] 도 36은 pMU102 자이모모나스 모빌리스 PDC-ADH 플라스미드의 예를 도시한다.
- [0059] 도 37은 pMU102 자이모박터 팔매 PDC, 자이모모나스 모빌리스 ADH 플라스미드의 예를 도시한다.
- [0060] 도 38은 pMU360의 플라스미드 맵을 도시한다. pMU360의 DNA 서열은 SEQ ID NO:61로 제시된다.
- [0061] 도 39는 티아페니콜-내성 형질전환체(thiamphenicol-resistant transformant)의 9개 콜로니에서의 락테이트 수준을 도시한다.
- [0062] 도 40은 *T. sacch.* pfl KO 단일 크로스오버 플라스미드 (SEQ ID NO:47)의 예를 도시한다.
- [0063] 도 41은 *T. sacch.* pfl KO 이중 크로스오버 플라스미드 (SEQ ID NO:48)의 예를 도시한다.
- [0064] 도 42는 *C. therm.* pfl KO 단일 크로스오버 플라스미드 (SEQ ID NO:49)의 예를 도시한다.
- [0065] 도 43은 *C. therm.* pfl KO 이중 크로스오버 플라스미드 (SEQ ID NO:50)의 예를 도시한다.
- [0066] 도 44는 *C. phyto.* pfl KO 단일 크로스오버 플라스미드 (SEQ ID NO:51)의 예를 도시한다.
- [0067] 도 45는 *C. phyto.* pfl KO 이중 크로스오버 플라스미드 (SEQ ID NO:52)의 예를 도시한다.
- [0068] 도 46은 *T. sacch.* #59 L-ldh KO 단일 크로스오버 플라스미드 (SEQ ID NO:53)의 예를 도시한다.
- [0069] 도 47은 *T. sacch.* #59 L-ldh KO 이중 크로스오버 플라스미드 (SEQ ID NO:54)의 예를 도시한다.
- [0070] 도 48은 *T. sacch.* #59 pta/ack KO 단일 크로스오버 플라스미드 (SEQ ID NO:55)의 예를 도시한다.
- [0071] 도 49는 *T. sacch.* #59 pta/ack KO 이중 크로스오버 플라스미드 (SEQ ID NO:56)의 예를 도시한다.
- [0072] 도 50은 *T. pseudo.* L-ldh KO 단일 크로스오버 플라스미드 (SEQ ID NO:57)의 예를 도시한다.
- [0073] 도 51은 *T. pseudo.* L-ldh KO 이중 크로스오버 플라스미드의 예를 도시한다 (SEQ ID NO:58).
- [0074] 도 52는 *T. pseudo.* ack KO 단일 크로스오버 플라스미드 (SEQ ID NO:59)의 예를 도시한다.
- [0075] 도 53은 *T. pseudo.* pta/ack KO 이중 크로스오버 플라스미드 (SEQ ID NO:60)의 예를 도시한다.

#### 표의 간단한 설명

- [0077] 표 1은 대표적인 고도로 셀룰로오스분해성인 생물체들을 요약한 것이다.
- [0078] 표 2는 대표적인 고유한 셀룰로오스분해성 및 자일란분해성(xylanolytic) 생물체들을 요약한 것이다.
- [0079] 표 3은 기질 이용을 기초로 하여 세균 균주를 분류한 것이다.
- [0080] 표 4는 클로스트리디움 셀룰로리티쿰 아세테이트 키나아제에 인트론을 리타겟팅하기 위한 삽입 위치 및 프라이머를 나타낸다.
- [0081] 표 5는 클로스트리디움 셀룰로리티쿰 락테이트 데히드로게나아제에 인트론을 리타겟팅하기 위한 삽입 위치 및 프라이머를 나타낸다.
- [0082] 표 6은 엔지니어링된 씨모안에어로박터 및 씨모안에어로박테리움 균주의 발효 성능을 나타낸다.

#### 발명의 상세한 설명

- [0084] 본 발명의 일면은 리그노셀룰로오스 바이오매스부터 에탄올을 생산하는 데에 사용되는 호열성 또는 중온성 미생물의 엔지니어링에 관한 것이다. 에탄올 생산을 위해 호열성 세균을 사용하는 것은 중온성 에탄올 생산균을 기반으로 하는 종래의 공정에 비해 많은 이점을 제공한다. 예를 들어, 호열성 생물체를 사용하는 것은 적은 에탄올 분리 비용, 외부 효소 첨가를 위한 요건 감소 및 처리 시간 감소로 인해 종래의 공정에 비해 현저한 경제적 절감효과를 제공한다.

- [0085] 본 발명의 일면은 신규한 처리 구성을 이용함으로써 셀룰로오스 바이오매스 함유 물질로부터 에탄올을 생산하는 비용을 감소시킬 수 있는 공정에 관한 것이다. 특히, 본 발명은 유전자 변형된 미생물에서 에탄올 생산을 증가시키기 위한 다수의 방법을 제공한다.
- [0086] 특정한 다른 구체예에서, 본 발명은 발효 생산물로서 유기산을 생산하는 능력을 미생물에게 부여하는 효소를 엔코딩하는 유전자 또는 특정 폴리뉴클레오티드 서열이 부분적으로, 실질적으로 또는 완전히 결실되거나 침묵되거나 비활성화되거나 하향조절됨으로써, 주요 발효 생산물로서 에탄올을 생산하는 능력이 증가된 유전자 변형된 호열성 또는 중온성 미생물에 관한 것이다. 또한, 통합된 생물가공으로 일반적으로 공지된 처리 단계들의 새로운 통합에 의해, 본 발명의 일면은 셀룰로오스 바이오매스를 함유하는 원료 물질로부터의 에탄올의 보다 효율적인 생산을 제공한다. 유전자 변형된 호열성 또는 중온성 미생물을 상기 물질을 처리하는 데에 포함시키면 발효 단계가 보다 높은 온도에서 수행될 수 있게 하여 공정 경제성을 개선시킨다. 예를 들어, 반응 속도론은 전형적으로 온도에 비례하므로, 높은 온도는 전체 생산율의 증가와 일반적으로 관련된다. 또한, 높은 온도는 브로스(broth)로부터 휘발성 생산물이 제거되는 것을 촉진하고, 전처리 후에 냉각시킬 필요성을 감소시킨다.
- [0087] 특정 구체예에서, 본 발명은 발효 생산물로서 에탄올을 생산하는 능력을 부여하는 효소를 생산하는 능력이 증가된 유전자 변형 또는 재조합 호열성 또는 중온성 미생물에 관한 것이며, 상기 효소(들)의 존재는 주요 발효 생산물로서 에탄올을 생산하도록 리그노셀룰로오스 바이오매스 물질을 대사하는 공정을 변화시킨다. 본 발명의 한 가지 일면에서, 하나 이상의 비고유 유전자가 유전자 변형된 호열성 또는 중온성 미생물내로 삽입되며, 여기서 상기 비고유 유전자는 에탄올의 대사적 생산에 관여하는 효소를 엔코딩하는데, 예를 들어 이러한 효소는 펜토오스 당 및/또는 헥소오스 당을 대사하는 능력을 부여할 수 있다. 예를 들어, 한 가지 구체예에서, 상기 효소는 D-자일로오스 또는 L-아라비노오스 경로에 관여함으로써, 미생물이 펜토오스 당, 즉, D-자일로오스 또는 L-아라비노오스를 대사하게 할 수 있다. D-자일로오스 또는 L-아라비노오스의 대사 또는 이용에 관여하는 효소를 엔코딩하는 비고유 유전자를 삽입 (예를 들어, 도입 또는 첨가)함으로써, 상기 미생물은 고유의 생물체에 비해 에탄올을 생산하는 능력이 증가한다.
- [0088] 또한, 본 발명은 본 발명의 미생물내로 통합될 수 있는 신규한 조성물을 제공한다. 한 가지 구체예에서, 본 발명의 분리된 핵산 분자는 SEQ ID NO: 1-76 중 어느 하나에 제시된 뉴클레오티드 서열의 상보서열인 핵산 분자를 포함한다. 또 다른 구체예에서, 본 발명의 분리된 핵산 분자는 SEQ ID NO: 1-76 중 어느 하나에 제시된 뉴클레오티드 서열 또는 이러한 뉴클레오티드 서열 중 어느 하나의 일부의 상보서열인 핵산 분자를 포함한다. SEQ ID NO: 1-76 중 어느 하나에 제시된 뉴클레오티드 서열 또는 이의 코딩 영역에 상보적인 핵산 분자는 SEQ ID NO: 1-76 중 어느 하나에 제시된 뉴클레오티드 서열 또는 이의 코딩 영역에 충분히 상보적인 핵산 분자이어서, SEQ ID NO: 1-76 중 어느 하나에 제시된 뉴클레오티드 서열 또는 이의 코딩 영역에 하이브리드화함으로써 안정한 듀플렉스(duplex)를 형성할 수 있다.
- [0089] 또 다른 바람직한 구체예에서, 본 발명의 분리된 핵산 분자는 SEQ ID NO: 1-76 중 어느 하나에 제시된 뉴클레오티드 서열 (뉴클레오티드 서열의 전장) 또는 이러한 뉴클레오티드 서열 중 어느 하나의 일부에 대해 적어도 약 50%, 54%, 55%, 60%, 62%, 65%, 70%, 75%, 78%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 이를 초과하는 비율로 상동성인 뉴클레오티드 서열을 포함한다.
- [0090] 더욱이, 본 발명의 핵산 분자는 SEQ ID NO: 1-76 중 어느 하나의 핵산 서열 또는 이의 코딩 영역의 일부만을 포함할 수 있는데, 예를 들어 이러한 핵산 분자는 프로브 또는 프라이머로서 사용될 수 있는 단편 또는 단백질의 생물학적으로 활성인 부분을 엔코딩하는 단편일 수 있다. 또 다른 구체예에서, 본 발명의 핵산 분자는 SEQ ID NO: 1-76 중 어느 하나의 적어도 약 12개 또는 15개, 바람직하게는 20개 또는 25개, 더욱 바람직하게는 약 30개, 35개, 40개, 45개, 50개, 55개, 60개, 65개, 또는 75개의 연속 뉴클레오티드를 포함할 수 있다.
- [0091] 정의
- [0092] "이중성 폴리뉴클레오티드 세그먼트"란 용어는 하나 이상의 폴리펩티드 또는 폴리펩티드의 일부 또는 단편을 엔코딩하는 폴리뉴클레오티드 세그먼트를 포함하도록 의도된다. 이중성 폴리뉴클레오티드 세그먼트는 임의의 공급원, 예를 들어 진핵생물, 원핵생물, 바이러스 또는 합성 폴리뉴클레오티드 단편으로부터 유래될 수 있다.
- [0093] "프로모터" 또는 "대용(surrogate) 프로모터"라는 용어는 자연 상태에서는 전사적으로 제어하지 못하는 관심있는 유전자(gene-of-interest)를 전사적으로 제어할 수 있는 폴리뉴클레오티드 세그먼트를 포함하도록 의도된다. 특정 구체예에서, 대용 프로모터의 전사적 제어는 관심있는 유전자의 발현을 증가시킨다. 특정 구체예에서, 대용 프로모터는 관심있는 유전자의 5'에 위치한다. 대용 프로모터는 천연 프로모터를 치환하기 위해 사용되거나

천연 프로모터에 더하여 사용될 수 있다. 대용 프로모터는 이것이 사용되는 숙주 세포에 대해 내생성(endogenous)인 것일 수 있거나, 숙주 세포내로 도입되는 이중성 폴리뉴클레오티드 서열, 예를 들어 이것이 사용되는 숙주 세포에 대해 외생적(exogenous)인 것일 수 있다.

[0094] "유전자(들)" 또는 "폴리뉴클레오티드 세그먼트" 또는 "폴리뉴클레오티드 서열(들)"은 폴리펩티드를 엔코딩하는 오픈 리딩 프레임에 포함하고 추가로 비코딩(non-coding) 조절 서열 및 인트론을 포함할 수 있는 핵산 분자, 예를 들어 폴리뉴클레오티드를 포함하도록 의도된다. 또한, 상기 용어는 기능적 유전자좌(locus)에 대해 맵핑되는 하나 이상의 유전자를 포함하도록 의도된다. 또한, 상기 용어는 선택된 목적을 위해 특정 유전자를 포함하도록 의도된다. 유전자는 숙주 세포에 대해 내생성일 수 있거나, 숙주 세포내로 재조합적으로 도입될 수 있는 데 예를 들어 에피솜으로 유지되는 플라스미드로서 또는 계놈내로 안정하게 통합되는 플라스미드 (또는 이의 단편)로서 도입될 수 있다. 유전자는 플라스미드 형태 뿐만 아니라 예를 들어 선형 DNA의 형태로 존재할 수 있다. 특정 구체예에서, 폴리뉴클레오티드 세그먼트의 유전자는 탄수화물을 에탄올로 생물전환(bioconversion)하는 데에 있어서 적어도 하나의 단계에 관여한다. 따라서, 상기 용어는 폴리펩티드, 예를 들어 효소인 아세테이트 키나아제 (ACK), 포스포트랜스아세틸라아제 (PTA), 및/또는 락테이트 데히드로게나아제 (LDH), D-자일로오스 경로의 효소, 예를 들어 자일로오스 이소머라아제 및 자일룰로키나아제, L-아라비노오스 경로의 효소, 예를 들어 L-아라비노오스 이소머라아제 및 L-리블로오스-5-포스페이트 4-에피머라아제를 엔코딩하는 임의의 유전자를 포함하도록 의도된다. 또한, 유전자라는 용어는 특정 유전자의 모든 카피(copy), 예를 들어 특정 유전자 생산물을 엔코딩하는 세포내의 모든 DNA 서열을 포함하도록 의도된다.

[0095] "전사 제어"라는 용어는 전사 수준에서 유전자 발현을 조절하는 능력을 포함하도록 의도된다. 특정 구체예에서, 전사 및 이에 따른 발현은 관심있는 유전자의 코딩 영역의 5' 단부 근처에서 대용 프로모터로 치환하거나 이러한 프로모터를 첨가하여 변경된 유전자 발현을 일으킴으로써 조절된다. 특정 구체예에서, 하나 이상의 유전자의 전사 제어는 그러한 유전자들의 최적 발현, 예를 들어 요망되는 비를 생성시키도록 엔지니어링된다. 또한, 상기 용어는 당 분야에서 인정되는 유도성 전사 제어를 포함한다.

[0096] "발현"이란 용어는 적어도 mRNA 생산의 수준에서의 유전자의 발현을 포함하도록 의도된다.

[0097] "발현 생산물"이란 용어는 발현된 유전자의 결과로서 생긴 생산물, 예를 들어 폴리펩티드를 포함하도록 의도된다.

[0098] "발현 증가"란 용어는 적어도 증가된 mRNA 생산의 수준, 바람직하게는 폴리펩티드 발현의 수준에서의 유전자 발현의 변화를 포함하도록 의도된다. "생산 증가"란 용어는 폴리펩티드의 효소적 활성의 수준 또는 이들이 조합된 수준에서의 발현된 폴리펩티드의 양의 증가를 포함하도록 의도된다.

[0099] "활성", "활성들", "효소적 활성" 및 "효소적 활성들"이란 용어는 상호교환적으로 사용되며, 유리한 조건하에서 생산되는 경우에 선택된 폴리펩티드에 정상적으로 기인하는 임의의 기능적 활성을 포함하는 것으로 의도된다. 전형적으로, 선택된 폴리펩티드의 활성은 생산된 폴리펩티드와 관련된 전체 효소적 활성을 포함한다. 숙주 세포에 의해 생산되고 효소적 활성을 지니는 폴리펩티드는 세포의 세포내 공간에 위치하거나, 세포와 결합되어 있거나, 세포의 환경내로 분비되거나, 이들 모두가 조합된 경우일 수 있다. 분비된 활성과 비교하여 전체 활성을 결정하기 위한 기술은 본원에 기재되어 있으며, 당 분야에 공지되어 있다.

[0100] "자일란분해 활성"이란 용어는 올리고펜토오스 및 폴리펜토오스에서 글리코시드 연결을 가수분해하는 능력을 포함하도록 의도된다.

[0101] "셀룰로오스분해 활성"이란 용어는 올리고헥소오스 및 폴리헥소오스에서 글리코시드 연결을 가수분해하는 능력을 포함하도록 의도된다. 또한, 셀룰로오스분해 활성은 셀룰로오스 및 헤미셀룰로오스를 탈중합(depolymerize)시키거나 가지제거(debranch)하는 능력을 포함할 수 있다.

[0102] 본원에 사용된 용어 "락테이트 데히드로게나아제" 또는 "LDH"는 피루베이트를 락테이트로 전환시킬 수 있는 효소를 포함하도록 의도된다. LDH는 히드록시부티레이트의 산화를 또한 촉매작용할 수 있는 것으로 이해된다.

[0103] 본원에 사용된 용어 "알코올 데히드로게나아제" 또는 "ADH"는 아세트알데히드를 알코올, 유리하게는 에탄올로 전환시킬 수 있는 효소를 포함하도록 의도된다.

[0104] "피루베이트 데카르복실라아제 활성"이란 용어는 피루베이트를 아세트알데히드로 효소에 의해 전환시키는 폴리펩티드의 능력을 포함하는 것으로 의도된다 (예를 들어, "피루베이트 데카르복실라아제" 또는 "PDC"). 전형적으로, 선택된 폴리펩티드의 활성은, 예를 들어 효소의 우수한 기질 친화성, 열안정성, 다양한 pH에서의 안정성

또는 이들 특성의 조합을 포함하는, 생산된 폴리펩티드와 관련된 전체 효소적 활성을 포함한다.

- [0105] "에탄올생산성"이란 용어는 발효 생산물로서 탄수화물로부터 에탄올을 생산하는 미생물의 능력을 포함하도록 의도된다. 이러한 용어는 비제한적으로 천연 에탄올생산성 생물체, 자연 발생적 돌연변이 또는 유도된 돌연변이를 지닌 에탄올생산성 생물체 및 유전자 변형된 에탄올생산성 생물체를 포함하도록 의도된다.
- [0106] "발효시키는" 및 "발효"라는 용어는 탄수화물로부터 에탄올이, 특히 발효의 생산물로서, 생산되게 하는 효소적 공정 (예를 들어, 세포성 또는 무세포성이고, 예를 들어 용해물(lysate) 또는 정제된 폴리펩티드 혼합물임)을 포함하도록 의도된다.
- [0107] "분비된"이란 용어는 폴리펩티드가 주변세포질 공간 또는 세포외 환경으로 이동하는 것을 포함하도록 의도된다. "분비 증가"라는 용어는 주어진 폴리펩티드가 증가된 수준으로 (즉, "천연 분비량" 보다 과량으로) 분비되는 경우를 포함하도록 의도된다. 특정 구체예에서, "분비 증가"라는 용어는 천연 분비 수준과 비교하여 적어도 약 10% 또는 적어도 약 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000% 또는 이를 초과하는 비율로 주어진 폴리펩티드의 분비가 증가함을 지칭한다.
- [0108] "분비성 폴리펩티드"라는 용어는 또 다른 폴리펩티드를 세포의 세포내 공간으로부터 세포외 환경으로 수송하는 것을 촉진하는 임의의 폴리펩티드(들)를 단독으로 또는 다른 폴리펩티드와 함께 포함하도록 의도된다. 특정 구체예에서, 분비성 폴리펩티드(들)은 분비 활성을 그람-음성 또는 그람-양성 숙주 세포에게 부여하기에 충분한 모든 필요한 분비성 폴리펩티드를 포함한다. 전형적으로, 분비성 단백질은 하나의 숙주 세포로부터 분리되어 유전 공학을 이용하여 또 다른 숙주 세포로 전달될 수 있는 단일 영역 또는 유전자좌에 엔코딩되어 있다. 특정 구체예에서, 분비성 폴리펩티드(들)은 분비 활성을 지닌 임의의 세균 세포로부터 유래된다. 특정 구체예에서, 분비성 폴리펩티드(들)은 타입 II 분비 활성을 지닌 숙주 세포로부터 유래된다. 특정 구체예에서, 숙주 세포는 호열성 세균 세포이다.
- [0109] "로부터 유래된"이란 용어는 지시된 공급원으로부터 폴리뉴클레오티드 세그먼트를 (전체적으로 또는 부분적으로) 분리하거나 지시된 공급원으로부터 폴리펩티드를 정제하는 것을 포함하도록 의도된다. 상기 용어는, 예를 들어 지시된 폴리뉴클레오티드 공급원과 관련된 서열로부터 또는 이러한 서열을 기초로 하여 직접 클로닝하거나, PCR 증폭시키거나 인공적으로 합성하는 것을 포함하도록 의도된다.
- [0110] "호열성"이란 약 45℃ 또는 이 보다 높은 온도에서 잘 증식하는 생물체를 의미한다.
- [0111] "중온성"이란 약 20 내지 45℃의 온도에서 잘 증식하는 생물체를 의미한다.
- [0112] "유기산"이란 용어는 당 분야에 인지되어 있다. "락트산"이란 용어는 유리 산(free acid) 또는 염 형태의 유기산인 2-히드록시프로피온산을 지칭한다. 락트산의 염 형태는 중화제, 즉, 탄산칼슘 또는 수산화암모늄과 무관하게 "락테이트"로 지칭된다. "아세트산"이란 용어는 유리산 또는 염 형태의 에탄산(ethanoic acid)으로도 공지되어 있는 유기산인 메탄카르복실산을 지칭한다. 아세트산의 염 형태는 "아세테이트"로 지칭된다.
- [0113] 본 발명의 특정 구체예는 호열성 또는 중온성 미생물내에서의 소정의 유전자 또는 특정 폴리뉴클레오티드 서열의 "삽입" (예를 들어, 첨가, 통합, 혼입 또는 도입)을 제공하며, 유전자 또는 특정 폴리뉴클레오티드 서열의 이러한 삽입은, 결과로서 생긴 상기 호열성 또는 중온성 미생물의 군주가 "유전자 변형"되거나 "형질전환"된 것으로 이해될 수 있도록 "유전자 변형(들)" 또는 "형질전환(들)"을 포함하는 것으로 이해될 수 있다. 특정 구체예에서, 군주는 세균, 진균 또는 효모 기원일 수 있다.
- [0114] 본 발명의 특정 구체예는 호열성 또는 중온성 미생물내에서의 소정의 유전자 또는 특정 폴리뉴클레오티드 서열의 "비활성화" 또는 "결실"을 제공하며, 유전자 또는 특정 폴리뉴클레오티드 서열의 이러한 "비활성화" 또는 "결실"은, 결과로서 생긴 상기 호열성 또는 중온성 미생물의 군주가 "유전자 변형"되거나 "형질전환"된 것으로 이해될 수 있도록 "유전자 변형(들)" 또는 "형질전환(들)"을 포함하는 것으로 이해될 수 있다. 특정 구체예에서, 군주는 세균, 진균 또는 효모 기원일 수 있다.
- [0115] "CBP 생물체"라는 용어는 본 발명의 미생물, 예를 들어 CBP에 적합한 특성을 지니는 미생물을 포함하도록 의도된다.
- [0116] 본 발명의 한 가지 일면에서, 유전자 또는 특정 폴리뉴클레오티드 서열은 이에 의해 엔코딩되는 활성, 예를 들어 효소의 발현을 활성화시키도록 삽입된다. 특정 구체예에서, 에탄올의 대사적 생산에 관여하는 효소, 예를 들어 펜토오스 및/또는 헥소오스 당을 대사하는 효소를 엔코딩하는 유전자가 중온성 또는 호열성 생물체에 첨가될 수 있다. 본 발명의 특정 구체예에서, 효소는 펜토오스 당을 대사하는 능력을 부여할 수 있고, 예를 들어



D-자일로오스 경로 및/또는 L-아라비노오스 경로에 관여할 수 있다.

[0117] 본 발명의 한 가지 일면에서, 유전자 또는 특정 폴리뉴클레오티드 서열은 이에 의해 엔코딩되는 활성, 예를 들어 효소의 발현을 비활성화시키도록 부분적으로, 실질적으로 또는 완전히 결실되거나 침묵되거나 비활성화되거나 하향조절된다. 결실은 최대 안정성을 제공하는데, 이는 복귀 돌연변이로 하여금 기능을 복구하게 할 기회가 없기 때문이다. 또한, 유전자의 기능 및/또는 발현을 붕괴시키는 핵산 서열의 삽입에 의해 유전자가 부분적으로, 실질적으로 또는 완전히 결실되거나 침묵되거나 비활성화되거나 하향조절될 수 있다 (예를 들어, P1 형질도입 또는 당 분야에 공지된 다른 방법). "제거한다", "제거" 및 "녹아웃"이란 용어는 "결실"이라는 용어와 상호교환적으로 사용된다. 특정 구체예에서, 유기산의 생산을 녹아웃시키기 위해 관심있는 호열성 또는 중온성 미생물의 균주가 부위 지정(site directed) 상동 재조합에 의해 엔지니어링될 수 있다. 다른 구체예에서, 관심있는 특정 유전자를 부분적으로, 실질적으로 또는 완전히 침묵시키거나 비활성화시키거나 하향조절하기 위해 RNAi 또는 안티센스 DNA(asDNA)가 사용될 수 있다.

[0118] 특정 구체예에서, 본원에 기재된 바와 같은 결실 또는 비활성화를 위해 표적화된 유전자는 미생물의 고유한 군주에 내생적일 수 있고, 이에 따라 "고유 유전자(들)" 또는 "내생 유전자(들)"로 지칭되는 것으로 이해될 수 있다. 생물체가 이러한 생물체의 유전자 및/또는 표현형 구성을 의도적으로 변화시키는 방식으로 유전공학처리되거나 다른 방식으로 인위적으로 조작된 것이 아니라면 그러한 생물체는 "고유한 상태(native state)"로 존재하는 것이다. 예를 들어, 야생형 생물체는 고유한 상태로 존재하는 것으로 간주될 수 있다. 다른 구체예에서, 결실 또는 비활성화를 위해 표적화된 유전자(들)은 생물체에 대해 비고유한 것일 수 있다.

[0119] *바이오매스*

[0120] "리그노셀룰로오스 물질", "리그노셀룰로오스 기질" 및 "셀룰로오스 바이오매스"라는 용어는 셀룰로오스, 헤미셀룰로오스, 리그닌 또는 이들의 조합물을 포함하는 임의의 유형의 바이오매스, 예를 들어 비제한적으로 목재 바이오매스(woody biomass), 먹이풀(forage grass), 초본 에너지 작물(herbaceous energy crop), 비목본식물 바이오매스(non-woody-plant biomass), 농업 폐기물 및/또는 농업 잔존물, 산림 잔존물 및/또는 산림 폐기물, 종이-생산 슬러지 및/또는 폐지 슬러지, 폐수-처리 슬러지, 도시 고형 폐기물(municipal solid waste), 습식 및 건식 제분 옥수수 에탄올 플랜트로부터의 옥수수 섬유, 및 설탕-가공 잔존물을 의미한다.

[0121] 비제한적 예에서, 리그노셀룰로오스 물질은 비제한적으로 목재 바이오매스, 예를 들어 리사이클링된 목재 펄프 섬유, 톱밥, 경질목재, 연질목재, 및 이들의 조합물; 그래스, 예를 들어, 스위치 그래스, 코드 그래스, 라이 그래스, 리드 카나리 그래스, 미스칸투스, 또는 이들의 조합물; 설탕-가공 잔존물, 예를 들어 비제한적으로 사탕수수 버캐스; 농업 폐기물, 예를 들어 비제한적으로 볏짚, 왕겨, 보리짚, 옥수수 속대, 곡물짚, 밀짚, 카놀라짚, 귀리짚, 귀리껍질, 및 옥수수섬유; 스토버, 예를 들어 비제한적으로 대두 스토버, 옥수수 스토버; 산림 폐기물, 예를 들어 비제한적으로 리사이클링된 목재 펄프 섬유, 톱밥, 경질목재 (예를 들어, 포플러, 오크(oak), 단풍나무(maple), 자작나무(birch), 버드나무), 연질목재, 또는 이들의 임의의 조합물을 포함할 수 있다. 리그노셀룰로오스 물질은 하나의 섬유 종(species)을 포함할 수 있고; 대안적으로 리그노셀룰로오스 물질은 다양한 리그노셀룰로오스 물질로부터 유래된 섬유들의 혼합물을 포함할 수 있다. 특히 유리한 리그노셀룰로오스 물질은 농업 폐기물, 예를 들어 밀짚, 보리짚, 카놀라짚 및 귀리짚을 포함하는 곡물짚; 옥수수 섬유; 스토버, 예를 들어 옥수수 스토버 및 대두 스토버; 그래스, 예를 들어 스위치 그래스, 리드 카나리 그래스, 코드 그래스, 및 미스칸투스; 또는 이들의 조합물이다.

[0122] 또한, 종이 슬러지는 에탄올 생산을 위한 실용적인 공급원료이다. 종이 슬러지는 펄프화 및 제지 과정으로부터 발생하는 고형 잔존물이고, 이는 전형적으로 1차 정화기에서 공정 폐수로부터 분리된다. 처분 비용이 습윤 톤(wet ton) 당 \$30인 경우, 슬러지 처분 비용은 판매용으로 생산되는 종이 톤 당 \$5와 동일하다. 습윤 슬러지를 처분하는 비용은 에탄올로 전환시키는 것과 같이 물질을 다른 용도로 전환시키는 것에 대한 중요한 유인(incentive)이다. 본 발명에 의해 제공되는 공정은 광범위하게 적용가능하다. 더욱이, 에탄올 또는 고부가가치 화학물질, 예를 들어 유기산, 방향족물질, 에스테르, 아세톤 및 고분자 중간체(polymer intermediate)를 생산하기 위해 당화(saccharification) 및/또는 발효 생산물이 사용될 수 있다.

[0123] *피루베이트 포르메이트 리아제 (PFL)*

[0124] 피루베이트 포르메이트 리아제 (PFL)는 혐기적 글루코오스 대사를 조절하는 것을 보조하는 중요한 효소이다 (이는 대장균 및 다른 생물체에서 발견됨). 라디칼 화학을 이용하는 경우, 상기 효소는 피루베이트 및 조효소-A를 포르메이트 및 에탄올의 전구체인 아세틸-CoA로 가역적으로 전환시키는 것을 촉매작용한다. 피루베이트 포르메

이트 리아제는 85 kDa의 759-잔기 서브유닛으로 이루어진 동형이량체(homodimer)이다. 상기 효소는 10-가닥 베타/알파 배럴 모티프(10-stranded beta/alpha barrel motif)를 지니며, 이러한 모티프내로 주요 촉매 잔기를 함유하는 베타 핑거(beta finger)가 삽입되어 있다. x선 결정학에 의해 밝혀진 상기 효소의 활성 부위는 촉매 작용을 수행하는 3개의 필수 아미노산 (Gly734, Cys418, 및 Cys419), 기질인 피루베이트를 바로 곁에 보유하는 3개의 주요 잔기 (Arg435, Arg176, 및 Ala272), 및 2개의 플랭킹(flanking) 소수성 잔기 (Trp333 및 Phe432)를 보유한다.

[0125] 피루베이트 포르메이트 리아제의 활성 부위와 클래스 I 및 클래스 III 리보뉴클레오티드 리덕타아제 (RNR) 효소의 활성 부위 사이의 구조적 유사성이 연구에 의해 밝혀졌다. 3개의 촉매작용 잔기의 역할은 다음과 같다: Gly734 (글리실 라디칼)은 Cys419를 통해 Cys418로 온 앤드 오프(on and off) 방식으로 라디칼을 전달하고; Cys418 (티일 라디칼(thiyl radical))은 피루베이트 카르보닐의 탄소 원자상에서 아실화 화학을 수행하고; Cys419 (티일 라디칼)은 수소-원자 전달을 수행한다.

[0126] 피루베이트 포르메이트 리아제에 대한 제안된 메커니즘은 Cys419를 통해 Gly734로부터 Cys418로 라디칼을 전달하는 것으로 시작한다. Cys418 티일 라디칼은 피루베이트의 C2 (2번 탄소 원자)에 공유적으로 부가되어, 아세틸-효소 중간체 (이는 이제 라디칼을 함유함)를 생성시킨다. 이러한 아세틸-효소 중간체는 Cys419에 의해 수소-원자를 전달받게 되는 포르밀 라디칼을 방출한다. 이는 포르메이트 및 Cys419 라디칼을 생성시킨다. 조효소-A는 Cys419 라디칼에 의해 수소-원자를 전달받아서 조효소-A 라디칼을 생성시킨다. 그 후, 조효소-A 라디칼은 Cys418로부터 아세틸기를 취하여 아세틸-CoA를 생성시킴으로써, Cys418 라디칼이 남게 한다. 그 후, 피루베이트 포르메이트 리아제는 라디칼을 전달받아서 그 라디칼을 Gly734상에 되돌려 놓을 수 있다. 또한, 상기 언급된 단계들은 모두 가역적이다.

[0127] 2개의 추가 효소인 PFL 활성화효소(activase) (AE) 및 PFL 비활성화효소(deactivase) (DA)는 혐기적 글루코오스 대사를 조절하기 위해 피루베이트 포르메이트 리아제의 "온(on)" 및 "오프(off)" 상태를 조절한다. 활성화된 피루베이트 포르메이트 리아제는, 피루베이트가 이용가능한 경우, 에너지 생산에서 중요한 작은 분자인 아세틸-CoA의 형성을 가능하게 한다. 비활성화된 피루베이트 포르메이트 리아제는 기질이 존재하는 경우에도 반응을 촉매작용하지 않는다. PFL 활성화효소는 라디칼 SAM (S-아데노실메티오닌) 슈퍼패밀리(superfamily)의 일부이다.

[0128] 따라서, PFL 활성화효소 (AE)는 Gly734 (G-H)를 Gly734 라디칼 (G\*)로 전환시킴으로써 피루베이트 포르메이트 리아제를 "온(on)" 상태가 되게 한다. PFL 비활성화효소 (DA)는 Gly734 라디칼을 켄칭(quenching)시킴으로써 피루베이트 포르메이트 리아제를 "오프(off)" 상태가 되게 한다. 또한, 피루베이트 포르메이트 리아제는 분자 산소 (O<sub>2</sub>)에 민감한데, 이러한 분자 산소가 존재하면 상기 효소는 셧오프(shut off)된다.

[0129] *자일로오스 대사*

[0130] 자일로오스는 다양한 생물체에 의해 유용한 생산물로 대사될 수 있는 탄소 5개의 단당류이다. 2개의 주요 자일로오스 대사 경로가 존재하며, 각각의 경로는 사용되는 특징적 효소가 독특하다. 한 가지 경로는 "자일로오스 리덕타아제-자일리톨 데히드로게나아제" 또는 XR-XDH 경로라고 일컬어진다. 자일로오스 리덕타아제 (XR) 및 자일리톨 데히드로게나아제 (XDH)는 이러한 자일로오스 분해 방법에서 사용되는 2개의 주된 효소이다. XYL1 유전자에 의해 엔코딩되는 XR은 자일로오스를 자일리톨로 환원시키는 것을 담당하고, 보조인자인 NADH 또는 NADPH에 의해 보조를 받는다. 그 후, 자일리톨은 XYL2 유전자를 통해 발현되는 XDH에 의해 자일롤로오스로 산화되고, 이는 오로지 보조인자 NAD<sup>+</sup>에 의해서만 달성된다. 이러한 경로에 필요한 다양한 보조인자 및 그러한 보조인자가 사용을 위해 이용가능한 정도로 인해, 불균형은 자일리톨 부산물의 과다생산 및 바람직하지 않은 에탄올의 비효율적 생산을 초래할 수 있다. 자일로오스 대사 경로의 효율을 최적화하기 위해 XR 및 XDH 효소 수준의 발현을 변화시키는 것이 실험실에서 시험되었다.

[0131] 자일로오스 대사를 위한 나머지 경로는 "자일로오스 이소머라아제" (XI) 경로라고 일컬어진다. 효소 XI는 자일로오스를 자일롤로오스로 직접 전환시키는 것을 담당하며, 자일리톨 중간체를 거치지 않는다. 둘 모두의 경로는 자일롤로오스를 생성시키지만, 사용되는 효소는 상이하다. 자일롤로오스의 생산 후, XR-XDH 및 XI 경로는 둘 모두 XKS1 유전자상에서 엔코딩되는 자일롤로키나아제 (XK) 효소를 통해 진행하여 자일롤로오스를 자일롤로오스-5-P로 추가로 변화시키며, 이어서 자일롤로오스-5-P는 추가의 이화(catabolism)를 위한 펜토오스 포스페이트 경로에 진입한다.

[0132] 자일로오스 대사 동안 펜토오스 포스페이트 경로를 통한 플럭스(flux)에 관한 연구는 이러한 단계의 속도를 제



한하는 것이 에탄올 발효 효율에 유리할 수 있음을 밝혀냈다. 에탄올 생산을 개선시킬 수 있는 이러한 플럭스에 대한 변화는 a) 포스포글루코오스 이소머라아제 활성을 낮추고, b) GND1 유전자를 결실시키고, c) ZWF1 유전자를 결실시키는 것을 포함한다 (Jeppsson et al., 2002). 펜토오스 포스페이트 경로가 대사 동안 추가의 NADPH를 생산하므로, 이러한 단계를 제한하는 것은 NAD(P)H 와 NAD<sup>+</sup> 보조인자 사이의 이미 분명한 불균형을 바로잡고 자일리톨 부산물을 감소시키는 것을 도와줄 것이다. 2개의 자일로오스 대사 경로를 비교하는 또 다른 실험은 XI 경로가 최대 에탄올 수율을 수득하도록 자일로오스를 가장 잘 대사할 수 있었으며 XR-XDH 경로는 훨씬 빠른 에탄올 생산 속도에 도달하였음을 밝혀내었다 (Karhumaa et al., 2007).

[0133] 미생물

[0134] 본 발명은 CBP에 필요한 기질-이용 및 생산물-형성 특성을 겸비한 미생물을 개발하기 위한 다수의 전략을 포함한다. "고유 셀룰로오스분해 전략(native cellulolytic strategy)"은 천연 셀룰로오스분해성 미생물을 수율 및 역가와 같은 생산물 관련된 특성을 개선시키도록 엔지니어링하는 것을 포함한다. "재조합 셀룰로오스분해 전략(recombinant cellulolytic strategy)"은 높은 생산물 수율 및 역가를 나타내는 고유하게 비-셀룰로오스분해성인 생물체를 셀룰로오스 이용 또는 헤미셀룰로오스 이용 또는 둘 모두를 가능하게 하는 이중성 셀룰라아제 시스템을 발현하도록 엔지니어링하는 것을 포함한다.

[0135] 셀룰로오스분해성 미생물

[0136] 셀룰로오스분해성이거나 셀룰로오스분해 활성을 지니는 것으로 문헌에 보고된 수 가지 미생물은 미세결정 셀룰로오스 뿐만 아니라 다양한 다른 당에서 증식하는 능력을 포함하는 다양한 수단에 의해 특징화되었다. 또한, 그러한 생물체들은 비제한적으로 셀룰로오스 및 헤미셀룰로오스를 탈중합시키거나 가지제거하는 능력을 포함하는 다른 수단에 의해 특징화될 수 있다. 관심있는 생물체를 벤치마킹(benchmarking)하기 위해 클로스트리디움 써모셀룸 (DSMZ 1237 균주)이 사용되었다. 본원에 사용된 바와 같이, 클로스트리디움 써모셀룸은 비제한적으로 DSMZ 1237, DSMZ 1313, DSMZ 2360, DSMZ 4150, DSMZ 7072, 및 ATCC 31924를 포함하는 다양한 균주를 포함할 수 있다. 본 발명의 특정 구체예에서, 클로스트리디움 써모셀룸은 비제한적으로 DSMZ 1313 또는 DSMZ 1237을 포함할 수 있다. 또 다른 구체예에서, 본 발명에 사용되는 특히 적합한 관심있는 생물체는 클로스트리디움 써모셀룸과 70% 이상의 16S rDNA 상동성을 지닌 셀룰로오스분해성 미생물을 포함한다. 클로스트리디움 써모셀룸, 클로스트리디움 셀룰로리툼, 써모안에어로박테리움 사카롤리툼, 클로스트리디움 스테르코라리움, 클로스트리디움 스테르코라리움 II, 칼디셀룰로시툼터 크리스티안소니, 클로스트리디움 피토펜페르멘탄스가 정렬되면 16S rDNA 유전자의 수준에서 73 내지 85% 상동성을 나타낸다 (도 6).

[0137] 클로스트리디움 스트라미니술벤스는 아비셀®에서 거의 클로스트리디움 써모셀룸 만큼 잘 증식하는 것으로 결정되었다. 표 1에는 특정한 고도로 셀룰로오스분해성인 생물체가 요약되어 있다.

[0138] 표 1

[0139]

| 균주                  | DSMZ<br>번호 | T<br>최적;<br>또는 범위 | pH<br>최적;<br>또는 범위  | 그람<br>염색 | 내기성(Aero-t<br>olerent) | 이용물질            | 생산물                                                       |
|---------------------|------------|-------------------|---------------------|----------|------------------------|-----------------|-----------------------------------------------------------|
| 클로스트리디움<br>써모셀룸     | 1313       | 55-60             | 7                   | 양성       | 아니오                    | 셀로비오스,<br>셀룰로오스 | 아세트산,<br>락트산,<br>에탄올, H <sub>2</sub> ,<br>CO <sub>2</sub> |
| 클로스트리디움<br>스트라미니술벤스 | 16021      | 50-55;<br>45-60   | 6.5-6.8;<br>6.0-8.5 | 양성       | 예                      | 셀로비오스,<br>셀룰로오스 | 아세트산,<br>락트산,<br>에탄올, H <sub>2</sub> ,<br>CO <sub>2</sub> |

[0140] 생물체들을 20 g/L의 셀로비오스 또는 20 g/L의 아비셀®에서 증식시켰다. 클로스트리디움 써모셀룸을 60℃에서 증식시키고, 클로스트리디움 스트라미니술벤스를 55℃에서 증식시켰다. 둘 모두를 50mM MOPS가 함유된 M122에서 -80℃ 냉동기 스톡(freezer stock) (DSMZ 유래)으로부터 예비배양하였다. 중기 내지 말기 대수증식기 동안, 예비배양물을 사용하여 100 mL 혈청 병(serum bottle) 중의 배치(batch) 배양물에 50 mL의 작업 부피로 접

중하였다. 대사 부산물과 당 소비의 HPLC 분석을 위해 주기적으로 액체 샘플을 분리하였다. OD<sub>600</sub>을 이러한 시점 각각에서 수득하였다. 도 12A 및 12B는 각각 셀로비오스 및 아비셀®에서의 클로스트리디움 스트라미니솔벤스에 대한 생산물 형성 및 OD<sub>600</sub>을 도시한다. 48시간내에 상당한 셀로비오스 (37%)가 소비된 후 OD가 떨어지고 생산물 형성이 일정해졌다. 도 13A 및 13B는 각각 셀로비오스 및 아비셀®에서의 클로스트리디움 써모셀룸에 대한 생산물 형성 및 OD<sub>600</sub>을 도시한다. 클로스트리디움 써모셀룸은 48시간내에 셀로비오스의 ~60%를 소비하였고, 이 시점에서 생산물 형성이 일정해졌다. 유기산의 형성으로 인한 역제는 기질의 불완전한 이용을 야기하였다.

[0141] 예를 들어 클로스트리디움 써모셀룸 및 클로스트리디움 스트라미니솔벤스를 포함하는 특정 미생물은 펜토오스당, 예를 들어 D-자일로오스 또는 L-아라비노오스를 대사할 수 없지만, 헥소오스 당은 대사할 수 있다. D-자일로오스 및 L-아라비노오스 둘 모두는 바이오매스에 풍부하게 존재하는 당인데, D-자일로오스는 연질 및 경질 목재의 경우 약 16 내지 20%를 차지하고 L-아라비노오스는 옥수수 섬유의 경우 약 25%를 차지한다. 따라서, 본 발명의 한 가지 목적은 D-자일로오스와 L-아라비노오스와 같은 펜토오스당을 대사하는 능력을 지닌 유전자 변형된 셀룰로오스분해성 미생물을 제공함으로써 바이오매스→에탄올 산업에서 발효를 위한 생물촉매(biocatalyst)로서의 이용 용도를 증진시키는 데에 있다.

[0142] 셀룰로오스분해성 및 자일란분해성 미생물

[0143] 셀룰로오스분해성이면서 자일란분해성인 것으로 문헌으로부터 결정된 수 가지 미생물은 미세결정 셀룰로오스 및 자작나무(birchwood) 자일란 뿐만 아니라 다양한 다른 당에서 증식하는 능력에 의해 특징화되었다. 관심있는 생물체를 벤치마킹하기 위해 클로스트리디움 써모셀룸이 사용되었다. 특징화를 위해 선택된 균주 중에서 클로스트리디움 셀룰로리툼, 클로스트리디움 스테르코라리움 아종 랩토사파르툼, 칼디셀룰로시룸터 크리스티안소니 및 클로스트리디움 피토펜타스가 아비셀®에서는 약하게 증식하였고, 자작나무 자일란에서는 잘 증식하였다. 표 2에는 고유하게 셀룰로오스분해성 및 자일란분해성인 생물체의 일부가 요약되어 있다.

[0144] 표 2

| 균주                                        | 공급원<br>/번호     | T<br>최적;<br>또는 범위 | pH<br>최적;<br>또는 범위 | 그람<br>염색               | 내기성 | 이용물질                                                                                            | 생산물                                                                  |
|-------------------------------------------|----------------|-------------------|--------------------|------------------------|-----|-------------------------------------------------------------------------------------------------|----------------------------------------------------------------------|
| 클로스트리<br>디움 셀룰로<br>리툼                     | DSM<br>5812    | 34                | 7.2                | 음성                     | 아니오 | 셀룰로오스, 자일<br>란,<br>아라비노오스, 만<br>노오스, 갈락토오<br>스, 자일로오스,<br>글루코오스, 셀로<br>비오스                      | 아세트산,<br>락트산,<br>에탄올,<br>H <sub>2</sub> , CO <sub>2</sub>            |
| 클로스트리<br>디움 스테르<br>코라리움 아<br>종 랩토사파<br>르툼 | DSM<br>9219    | 60-65             | 7.0-7.5            | 음성                     | 아니오 | 셀룰로오스, 셀로<br>비오스,<br>락토오스, 자일로<br>오스,<br>멜리비오스, 라피<br>노오스,<br>리보오스, 프룩토<br>오스, 수크로오스            | 아세트산,<br>락트산,<br>에탄올,<br>H <sub>2</sub> , CO <sub>2</sub>            |
| 칼디셀룰로<br>시룸터 크리스<br>티안소니                  | DSM<br>12137   | 78; 45-82         | 7; 5.8-8.0         | 음성                     | 아니오 | 셀로비오스, 글루<br>코오스, 자일로오<br>스, 갈락토오스,<br>만노오스, 셀룰로<br>오스                                          | 아세트산,<br>CO <sub>2</sub> ,<br>락트산,<br>에탄올,<br>포르메이트                  |
| 클로스트리<br>디움 피토펜<br>타스                     | ATCC<br>700394 | 37; 5-45          | 8.5; 6-9           | 음성<br>(그람<br>타입<br>양성) | 아니오 | 셀룰로오스, 자일<br>란, 셀로비오스,<br>프룩토오스, 갈락<br>토오스, 글루코오<br>스, 락토오스, 말<br>토오스,<br>만노오스, 리보오<br>스, 자일로오스 | 아세트산,<br>H <sub>2</sub> , CO <sub>2</sub> ,<br>락트산,<br>에탄올,<br>포르메이트 |

[0146] 생물체들을 20 g/L의 셀로비오스, 20 g/L의 아비셀® 또는 5 g/L의 자작나무 자일란에서 증식시켰다. 클로스트리디움 써모셀룸을 37℃에서 증식시키고, 클로스트리디움 스테르코라리움 아종 렙토사파르툼을 60℃에서 증식시키고, 칼디셀룰로시룸터 크리스티안소니를 75℃에서 증식시키고, 클로스트리디움 피토펜타멘타스를 37℃에서 증식시켰다. 모든 생물체들을 50mM MOPS가 보충된 M122c에서 -80℃ 냉동기 스톱으로부터 예비배양하였다. 증기 내지 말기 대수증식기 동안, 예비배양물을 사용하여 100 mL 혈청 병 중의 배치 배양물에 50 mL의 작업 부피로 접종하였다. 대사 부산물과 당 소비의 HPLC 분석을 위해 주기적으로 액체 샘플을 분리하였다. OD<sub>600</sub>을 이러한 시점 각각에서 측정하였다. 도 14A 내지 17B는 셀로비오스 및 아비셀®에서의 생산물 형성 및 OD<sub>600</sub>을 도시한다.

[0147] 별도의 실험에서, 생물체들을 셀로비오스, 글루코오스, 자일로오스, 갈락토오스, 아라비노오스, 만노오스 및 락토오스를 포함하는 2.5 g/L의 단일 당(single sugar) 뿐만 아니라 5 g/L의 아비셀®과 자작나무 자일란에서 증식시켰다. 도 18에는 2일 후 셀로비오스 및 자작나무 자일란에서의 생산물 형성이 비교되어 있다. 표 3에는 세균 균주가 이의 기질 이용을 기초로 하여 어떻게 분류될 수 있는 지가 요약되어 있다.

표 3

|                           | 셀로비오스 | 글루코오스 | 자일로오스 | 갈락토오스 | 아라비노오스 | 만노오스 | 락토오스 |
|---------------------------|-------|-------|-------|-------|--------|------|------|
| 클로스트리디움 셀룰로리티쿰            | X     | X     | X     | X     | X      |      |      |
| 클로스트리디움 스테르코라리움 아종 렙토사파르툼 | X     | X     | X     | X     | X      | X    |      |
| 칼디셀룰로시룸터 크리스티안소니          | X     | X     | X     | X     |        | X    |      |
| 클로스트리디움 피토펜타멘타스           | X     | X     | X     | X     |        | X    |      |

[0150] 미생물의 트랜스제닉(Transgenic) 전환

[0151] 본 발명은 특정 미생물의 트랜스제닉 전환을 위한 조성물 및 방법을 제공한다. 예를 들어 D-자일로오스 및/또는 L-아라비노오스를 포함하는 에탄올의 대사 경로에 관여하는 효소를 엔코딩하는 유전자가 이러한 유전자 중 하나 이상이 결여된 세균 균주, 예를 들어 클로스트리디움 써모셀룸 또는 클로스트리디움 스트라미니슬벤스내로 도입되는 경우, D-자일로오스에서 증식하거나 L-아라비노오스에서 증식하는 것에 대해 형질전환된 균주를 선별할 수 있다. 다른 클로스트리디알(Clostridial) 종으로부터의 유전자가 클로스트리디움 써모셀룸 및 클로스트리디움 스트라미니슬벤스에서 발현될 수 있는 것으로 예상된다. 표적 유전자 공여체는 헥소오스 및 펜토오스당을 대사하는 능력을 부여하는 미생물, 예를 들어 클로스트리디움 셀룰로리티쿰, 칼디셀룰로시룸터 크리스티안소니, 클로스트리디움 피토펜타멘타스, 클로스트리디움 스테르코라리움, 및 써모안테어로박테리움 사카롤리티쿰을 포함할 수 있다.

[0152] 써모안테어로박테리움 사카롤리티쿰, 클로스트리디움 셀룰로리티쿰 및 클로스트리디움 피토펜타멘타스의 게놈은 이용가능하다. 따라서, 본 발명은 상기 제시된 3개의 숙주 각각에 존재하는 자일로오스 이소머라아제 및 자일룰로키나아제에 상응하는 서열을 제공한다. 특히, 써모안테어로박테리움 사카롤리티쿰으로부터의 자일로오스 이소머라아제 (SEQ ID NO:6), 자일룰로키나아제 (SEQ ID NO:7), L-아라비노오스 이소머라아제 (SEQ ID NO:8), 및 L-리불로오스-5-포스페이트 4-에피머라아제 (SEQ ID NO: 9)에 상응하는 서열이 본원에 제시된다. 유사하게는, 클로스트리디움 셀룰로리티쿰으로부터의 자일로오스 이소머라아제 (SEQ ID NO:10), 자일룰로키나아제 (SEQ ID NO:11), L-아라비노오스 이소머라아제 (SEQ ID NO:12), 및 L-리불로오스-5-포스페이트 4-에피머라아제 (SEQ ID NO: 13)에 상응하는 서열이 본원에 제공된다. 클로스트리디움 피토펜타멘타스는 D-자일로오스 경로를 이용하며 L-아라비노오스를 이용하지 않는다. 따라서, 클로스트리디움 피토펜타멘타스로부터의 자일로오스 이소머

라아제 (SEQ ID NO:14) 및 자일롤로키나아제 (SEQ ID NO:15)에 상응하는 서열이 본원에 제시된다.

- [0153] 칼디셀룰로시럽터 크리스티안소니는 자일로오스를 분명히 대사한다. 이를 위해, 칼디셀룰로시럽터 크리스티안소니의 자일로오스 이소머라아제 (SEQ ID NO:71) 및 자일롤로키나아제 (SEQ ID NO:70) 유전자가 시퀀싱되어 본원에 제공된다. 클로스트리디움 스트라미니솔벤스는 자일로오스에서 증식하는 것으로 밝혀지지 않았지만, 이는 탄소원으로서의 자일로오스에서 적응한 후 기능적일 수 있는 자일로오스 이소머라아제 (SEQ ID NO:73) 및 자일롤로키나아제 (SEQ ID NO:72) 유전자를 분명히 함유한다.
- [0154] 클로스트리디움 써모셀룸 및 클로스트리디움 스트라미니솔벤스는 D-자일로오스→에탄올 경로 및/또는 L-아라비노오스 이용 경로에 관여하는 하나 이상의 공지된 유전자 또는 효소를 결여할 수 있다. 도 2 및 3은 클로스트리디움 써모셀룸의 이들 경로 각각에 존재하지 않는 2개의 핵심 효소를 도시한다. 클로스트리디움 스트라미니솔벤스는 자일로오스 이소머라아제 및 자일롤로키나아제를 지니지만, 이들 효소의 기능성은 알려져 있지 않다. 계놈 시퀀싱은 클로스트리디움 스트라미니솔벤스에서 L-아라비노오스 이소머라아제 또는 L-리불로오스-5-포스페이트 4-에피머라아제의 카피를 밝혀내지 못했다.
- [0155] 클로스트리디움 써모셀룸 및 클로스트리디움 스트라미니솔벤스는 자일롤로오스를 대사할 수 없는데, 이는 D-자일로오스를 D-자일롤로오스로 전환시키는 자일로오스 이소머라아제 (도 2에서 "XI" 또는 5.3.1.5로 지칭됨) 및 D-자일롤로오스를 D-자일롤로오스-5-포스페이트로 전환시키는 자일롤로키나아제 (도 2에서 "XK" 또는 2.7.1.1로도 지칭됨)에 대한 유전자가 없음 (클로스트리디움 써모셀룸) 또는 이러한 유전자의 활성 및/또는 발현의 결여 (클로스트리디움 스트라미니솔벤스)를 반영할 수 있다. 또한, 자일로오스의 수송은 클로스트리디움 스트라미니솔벤스에 대한 제한일 수 있다. 이러한 잠재적 제한은 써모안에어로박테리움 사카롤리티쿰 및 칼디셀룰로시럽터 크리스티안소니와 같은 자일로오스 이용 생물체로부터 당 수송 유전자를 발현시킴으로써 극복될 수 있다.
- [0156] 클로스트리디움 써모셀룸 및 클로스트리디움 스트라미니솔벤스는 L-아라비노오스를 또한 대사할 수 없는데, 이는 L-아라비노오스 이소머라아제 (도 3에서 5.3.1.4로도 지칭됨) 및 L-리불로오스-5-포스페이트 4-에피머라아제 (도 3에서 5.1.3.4로도 지칭됨)에 대한 유전자가 없음을 반영할 수 있다.
- [0157] 상기 기재된 4개의 유전자, 예를 들어 자일로오스 이소머라아제, 자일롤로키나아제, L-아라비노오스 이소머라아제 및 L-리불로오스-5-포스페이트 4-에피머라아제는 비제한적으로 클로스트리디움 셀룰로리티쿰 (도 4 참조), 써모안에어로박테리움 사카롤리티쿰, 클로스트리디움 스테르코라리움, 칼디셀룰로시럽터 크리스티안소니, 및 클로스트리디움 피토프레멘탄스를 포함하는 수 가지 클로스트리디알 종과 써모안에어로박테리움 사카롤리티쿰 종에 존재하며, 이러한 균주는 그러한 당의 양호한 이용균주이다. 상기 세균 균주가 본원에 기재된 유전자의 공여체로서 사용될 수 있는 것으로 인식될 것이다.
- [0158] 클로스트리디움 피토프레멘탄스는 상기 기재된 2개의 자일로오스 경로 유전자 (자일로오스 이소머라아제 및 자일롤로키나아제)를 발현하지만, 상기 기재된 아라비노오스 경로 유전자 (L-아라비노오스 이소머라아제 및 L-리불로오스-5-포스페이트 4-에피머라아제)를 결여하거나 발현하지 않는다 (도 5 참조).
- [0159] 따라서, 본 발명의 목적은, 예를 들어 바이오매스에서 유래된 펜토오스, 예를 들어 D-자일로오스 또는 L-아라비노오스 대사로부터 에탄올을 생산하기 위해 필요한 하나 이상의 효소에 대한 유전자를 도입시킴으로써, 당 이용 능력을 최적화하도록 상기 기재된 세균 균주의 일부를 변화시키는 데에 있다. 트리오스 포스페이트 이소머라아제 (TPI), GAPDH 및 LDH와 같은 클로스트리디움 써모셀룸 또는 클로스트리디움 스트라미니솔벤스의 고유 프로모터를 포함하는 프로모터가 상기 유전자를 발현시키기 위해 사용될 수 있다. 클로스트리디움 써모셀룸의 고유 프로모터에 상응하는 서열은 (TPI) (SEQ ID NO:16), GAPDH (SEQ ID NO:17) 및 LDH (SEQ ID NO:18)를 포함한다. 유전자가 클로닝된 경우, 발현 전에 코돈 최적화가 수행될 수 있다. 그 후, 클로스트리디움 써모셀룸 또는 클로스트리디움 스트라미니솔벤스를 형질전환시키고 유일한 탄소화물원으로서 D-자일로오스 또는 L-아라비노오스를 함유하는 배지에서 D-자일로오스 및 L-아라비노오스 증식에 대해 선별하기 위해, 예를 들어 고유 프로모터, 자일란분해 유전자 또는 아라비노오스분해 유전자, 및 선별가능 마커를 함유하는 카세트가 사용될 수 있다.
- [0160] 트랜스포손(Transposon)
- [0161] 숙주로 들어온 외래 DNA를 선별하기 위해, 그러한 DNA는 관심있는 생물체에서 안정하게 유지되는 것이 바람직하다. 플라스미드와 관련하여, 이를 일어나게 할 수 있는 2가지 방법이 존재한다. 한 가지 방법은 복제성 플라스미드를 사용하는 것을 통해 이루어진다. 이러한 플라스미드는, 숙주에 의해 인식되어 플라스미드가 딸세포로의 세포 분열 동안 분할되는(partitioned) 안정하고 자율적인 염색체외 엘리먼트로서 복제될 수 있도록 하는, 복제 원점을 지닌다. 두 번째 방법은 플라스미드를 염색체상으로 통합시킴으로써 이루어진다. 이는 주로 상동



재조합에 의해 일어나며, 전체 플라스미드 또는 플라스미드의 부분을 숙주 염색체내로 삽입시킨다. 따라서, 플라스미드 및 선별가능 마커(들)은 염색체의 일체부로서 복제되어 딸세포내로 분리된다. 따라서, 플라스미드 DNA가 선별가능 마커의 사용을 통해 형질전환 사건 동안 세포로 들어가는 지를 확인하기 위해서는 복제성 플라스미드 또는 플라스미드를 염색체상으로 재조합시키는 능력을 사용할 필요가 있다. 이러한 식별자(qualifier)는 항상 충족될 수 있는 것이 아니며, 일련의 유전학적 도구를 지니지 않는 생물체를 취급하는 경우에 특히 그러하다.

[0162] 플라스미드 관련된 마커에 관한 문제를 회피하는 한 가지 방법은 트랜스포손의 사용을 통해 이루어진다. 트랜스포손은 트랜스포사제(transposase)로서 지칭되는 효소적 기구(enzymatic machinery)에 의해 인식되는 모자이크 DNA 서열에 의해 규정되는 이동가능한 DNA 엘리먼트이다. 트랜스포사제의 기능은 트랜스포손 DNA를 숙주 또는 표적 DNA내로 무작위적으로 삽입하는 것이다. 선별가능 마커는 표준 유전공학에 의해 트랜스포손상으로 클로닝될 수 있다. 생성된 DNA 단편은 시험관내 반응으로 트랜스포사제 기구에 커플링될 수 있고, 복합체는 일렉트로포레이션(electroporation)에 의해 표적 세포내로 도입될 수 있다. 염색체상에서의 마커의 안정한 삽입은 단지 트랜스포사제 기구의 기능을 필요로 하고, 상동 재조합 또는 복제성 플라스미드의 필요성을 완화시킨다.

[0163] 트랜스포손의 통합과 관련된 무작위성은 돌연변이유발의 형태로서 작용하는 추가적 이점을 지닌다. 트랜스포손 돌연변이체의 병합물(amalgamation)을 포함하는 라이브러리가 생성될 수 있다. 이러한 라이브러리는 요망되는 표현형을 지닌 돌연변이체를 생산하기 위해 스크린(screen) 또는 선별 과정에서 사용될 수 있다. 예를 들어, CBP 생물체의 트랜스포손 라이브러리는 더 많은 에탄올을 생산하거나 더 적은 락트산 및/또는 더 적은 아세테이트를 생산하는 능력에 대해 스크리닝될 수 있다.

[0164] *고유 셀룰로오스분해 전략*

[0165] 천연 셀룰로오스분해성 미생물이 고유 전략을 통한 CBP 생물체 개발을 위한 출발점이다. 혐기성미생물 및 통성 혐기성미생물이 특히 관심을 끈다. 주요 목적은 산업적 공정의 요건을 만족시키도록 생산물 수율 및 에탄올 역가를 엔지니어링하는 데에 있다. 이러한 목적과 관련된 혼합-산(mixed-acid) 발효의 대사 엔지니어링은 중온성이고 비셀룰로오스분해성인 장내 세균의 경우에 성공적이었다. 적절한 유전자-전달 기술의 최근 개발은 이러한 유형의 작업이 셀룰로오스분해성 세균을 사용하여 수행될 수 있게 한다.

[0166] *재조합 셀룰로오스분해 전략*

[0167] 요망되는 생산물-형성 특성 (예를 들어, 높은 에탄올 수율 및 역가)을 지닌 비셀룰로오스분해 미생물이 재조합 셀룰로오스분해 전략에 의한 CBP 생물체 개발을 위한 출발점이다. 이러한 개발의 주요 목적은 전처리된 리그노셀룰로오스에서의 증식 및 발효를 가능하게 하는 이종성 셀룰라아제 시스템을 엔지니어링하는 데에 있다. 셀룰라아제의 이종성 생산(heterologous production)은 높은 수율로 에탄올을 생산하는 세균 숙주 (대장균, 클레브시엘라 옥시토카(*Klebsiella oxytoca*) 및 자이모모나스 모빌리스의 엔지니어링된 균주)와 효모인 사카로마이세스 세레비시애를 주로 사용하여 추구되어 왔다. 클레브시엘라 옥시토카의 균주에서의 셀룰라아제 발현은 미세 결정 셀룰로오스에 대해 가수분해 수율을 증가시켰지만 첨가된 셀룰라아제없이 증식을 초래하지 않았고, 비정질 셀룰로오스에서의 혐기성 증식을 초래하였다. 수 십개의 당분해 효소가 사카로마이세스 세레비시애에서 기능적으로 발현되었지만, 그러한 발현의 결과로써의 셀룰로오스에서의 혐기성 증식은 명확히 입증된 바 없다.

[0168] 본 발명의 일면들은 고유 셀룰로오스분해 전략을 통해 변화되는 숙주로서 호열성 또는 중온성 미생물을 사용하는 것에 관한 것이다. 생물공학의 공정 응용에서 상기 미생물의 잠재력은 비교적 높은 온도에서 증식하는 능력과 이에 따른 높은 대사율, 물리적으로 그리고 화학적으로 안정한 효소의 생산 및 최종 생산물의 상승된 수율로부터 유래된다. 호열성 세균의 주요 군은 진정세균(eubacteria) 및 고세균(archaeobacteria)을 포함한다. 호열성 진정세균은, 광합성(phototrophic) 세균, 예를 들어 시아노박테리아, 퍼플 박테리아(purple bacteria), 및 그린 박테리아(green bacteria); 그람-양성 세균, 예를 들어 바실루스(*Bacillus*), 클로스트리디움(*Clostridium*), 락트산 세균, 및 액티노마이세스(*Actinomyces*); 및 다른 진정세균, 예를 들어 티오바실루스(*Thiobacillus*), 스피로체트(*Spirochete*), 데суль포토마쿨룸(*Desulfotomaculum*), 그람-음성 호기성세균, 그람-음성 혐기성세균, 및 써모토가(*Thermotoga*)를 포함한다. 메타노젠(Methanogen), 고도 호열성세균(extreme thermophile) (당 분야에서 인지되는 용어), 및 써모플라스마(*Thermoplasma*)가 고세균에 속하는 것으로 간주된다. 특정 구체예에서, 본 발명은 써무스(*Thermus*) 속의 그람-음성 유기영양 호열성세균, 그람-양성 진정세균, 예를 들어 클로스트리디움속, 그리고 또한 구균(rod) 및 간균(coccus) 둘 모두를 포함하는 것들, 진정세균속의 속, 예를 들어 써모시포(*Thermosipho*) 및 써모토가(*Thermotoga*), 고세균의 속, 예를 들어 써모코쿠스(*Thermococcus*), 써모프로테우스(*Thermoproteus*) (막대 모양(rod-shaped)), 써모필룸(*Thermofilum*) (막대 모양), 피로딕티움(*Pyrodictium*), 액



시디아누스(*Acidianus*), 숄폴로부스(*Sulfolobus*), 피로바쿨룸(*Pyrobaculum*), 피로코쿠스(*Pyrococcus*), 써모디스쿠스(*Thermodiscus*), 스태필로써무스(*Staphylothermus*), 데숄푸로코쿠스(*Desulfurococcus*), 아르케오글로부스(*Archaeoglobus*), 및 메타노피루스(*Methanopyrus*)에 관한 것이다. 본 발명에 적합할 수 있는 호열성 또는 중온성 생물체의 일부 예 (세균, 원핵 미생물, 및 진균을 포함함)는 비제한적으로 클로스트리디움 써모숄푸로게네스(*Clostridium thermosulfurogenes*), 클로스트리디움 셀룰로리티쿰(*Clostridium cellulolyticum*), 클로스트리디움 써모셀룸(*Clostridium thermocellum*), 클로스트리디움 써모히드로숄푸리쿰(*Clostridium thermohydrosulfuricum*), 클로스트리디움 써모아세트릭쿰(*Clostridium thermoaceticum*), 클로스트리디움 써모사카롤리티쿰(*Clostridium thermosaccharolyticum*), 클로스트리디움 타르타리보룸(*Clostridium tartarivorum*), 클로스트리디움 써모셀룰라세움(*Clostridium thermocellulaseum*), 클로스트리디움 피토펜멘탄스(*Clostridium phytofermentans*), 클로스트리디움 스트라미노솔벤스(*Clostridium straminosolvens*), 써모안에어로박테리움 써모사카롤리티쿰(*Thermoanaerobacterium thermosaccharolyticum*), 써모안에어로박테리움 사카롤리티쿰(*Thermoanaerobacterium saccharolyticum*), 써모박테로이즈 아세트에틸리쿠스(*Thermobacteroides acetoethylicus*), 써모안에어로비움 브록키(*Thermoanaerobium brockii*), 메타노박테리움 써모오토트로피쿰(*Methanobacterium thermoautotrophicum*), 안에어로셀룸 써모필리움(*Anaerocellum thermophilum*), 피로딕티움 오쿨툼(*Pyrodictium occultum*), 써모프로테우스 뉴트로필루스(*Thermoproteus neutrophilus*), 써모필룸 리브룸(*Thermophilum librum*), 써모트릭스 티오파루스(*Thermothrix thioparus*), 데숄포비브리오 써모필루스(*Desulfovibrio thermophilus*), 써모플라스마 엑시도필룸(*Thermoplasma acidophilum*), 히드로게노모나스 써모필루스(*Hydrogenomonas thermophilus*), 써모마이크로비움 로세움(*Thermomicrobium roseum*), 써무스 플라바스(*Thermus flavas*), 써무스 루베르(*Thermus ruber*), 피로코쿠스 푸리오수스(*Pyrococcus furiosus*), 써무스 아쿠아티쿠스(*Thermus aquaticus*), 써무스 써모필루스(*Thermus thermophilus*), 클로로플렉수스 아우란티아쿠스(*Chloroflexus aurantiacus*), 써모코쿠스 리토랄리스(*Thermococcus litoralis*), 피로딕티움 아비시(*Pyrodictium abyssi*), 바실루스 스테아로써모필루스(*Bacillus stearothermophilus*), 시아니디움 칼다리움(*Cyanidium caldarium*), 마스티고클라두스 라미노수스(*Mastigocladus laminosus*), 클라미도트릭스 칼리디시마(*Chlamydothrix calidissima*), 클라미도트릭스 페니실라타(*Chlamydothrix penicillata*), 티오텍스 카르네아(*Thiothrix carnea*), 포르미디움 테누이시뮴(*Phormidium tenuissimum*), 포르미디움 게이세리콜라(*Phormidium geysericola*), 포르미디움 서브테라네움(*Phormidium subterraneum*), 포르미디움 비야헨시(*Phormidium bijahensi*), 오실라토리아 필리포르미스(*Oscillatoria filiformis*), 시네코코쿠스 리비두스(*Synechococcus lividus*), 클로로플렉수스 아우란티아쿠스(*Chloroflexus aurantiacus*), 피로딕티움 브록키(*Pyrodictium brockii*), 티오바실루스 티오옥시단스(*Thiobacillus thiooxidans*), 숄폴로부스 엑시도칼다리우스(*Sulfolobus acidocaldarius*), 티오바실루스 써모필리카(*Thiobacillus thermophila*), 바실루스 스테아로써모필루스(*Bacillus stearothermophilus*), 세르코술시페르 하마텐시스(*Cercosulcifer hamathensis*), 바흘캄프피아 레이치(*Vahlkampfia reichi*), 시클리디움 시트룰루스(*Cyclidium citrullus*), 닥틸라리아 갠로파바(*Dactylaria gallopava*), 시네코코쿠스 리비두스(*Synechococcus lividus*), 시네코코쿠스 엘롱가투스(*Synechococcus elongatus*), 시네코코쿠스 미네르바(*Synechococcus minervae*), 시네코시스티스 아쿠아틸루스(*Synechocystis aquatilis*), 아파노캡사 써말리스(*Aphanocapsa thermalis*), 오실라토리아 테레브리포르미스(*Oscillatoria terebriformis*), 오실라토리아 앰피비아(*Oscillatoria amphibia*), 오실라토리아 게르미나타(*Oscillatoria germinata*), 오실라토리아 오키니(*Oscillatoria okenii*), 포르미디움 라미노썸(*Phormidium laminosum*), 포르미디움 파르파라시엔스(*Phormidium parparasiens*), 심플로카 써말리스(*Symploca thermalis*), 바실루스 엑시도칼다리아스(*Bacillus acidocaldarias*), 바실루스 코굴란스(*Bacillus coagulans*), 바실루스 써모카테날라투스(*Bacillus thermocatenalatus*), 바실루스 리체니포르미스(*Bacillus licheniformis*), 바실루스 팔밀라스(*Bacillus pamilas*), 바실루스 마세란스(*Bacillus macerans*), 바실루스 써르쿨란스(*Bacillus circulans*), 바실루스 라테로스포루스(*Bacillus laterosporus*), 바실루스 브레비스(*Bacillus brevis*), 바실루스 서브틸리스(*Bacillus subtilis*), 바실루스 스파에리쿠스(*Bacillus sphaericus*), 데숄포토마쿨룸 니그리피칸스(*Desulfotomaculum nigrificans*), 스트렙토코쿠스 써모필루스(*Streptococcus thermophilus*), 락토바실루스 써모필루스(*Lactobacillus thermophilus*), 락토바실루스 불가리쿠스(*Lactobacillus bulgaricus*), 비피도박테리움 써모필룸(*Bifidobacterium thermophilum*), 스트렙토마이세스 프래그멘토스포루스(*Streptomyces fragmentosporus*), 스트렙토마이세스 써모니트리피칸스(*Streptomyces thermonitrificans*), 스트렙토마이세스 써모불가리스(*Streptomyces thermovulgaris*), 슈도노카르디아 써모필리아(*Pseudonocardia thermophila*), 써모액티노마이세스 불가리스(*Thermoactinomyces vulgaris*), 써모액티노마이세스 사카리(*Thermoactinomyces sacchari*), 써모액티노마이세스 칸디다스(*Thermoactinomyces Candidas*), 써모모노스포라 쿠르바타(*Thermomonospora curvata*), 써모모노스포라 비리디스(*Thermomonospora viridis*), 써모모노스포라 시트리나

(*Thermomonospora citrina*), 마이크로비스포라 썬모디아스타티카(*Microbispora thermodiastatica*), 마이크로비스포라 아에라타(*Microbispora aerata*), 마이크로비스포라 비스포라(*Microbispora bispora*), 액티노비피다 디코토미카(*Actinobifida dichotomica*), 액티노비피다 크로모게나(*Actinobifida chromogena*), 마이크로폴리스포라 카에시아(*Micropolyspora caesia*), 마이크로폴리스포라 피에니(*Micropolyspora faeni*), 마이크로폴리스포라 섹티부기다(*Micropolyspora cectivugida*), 마이크로폴리스포라 카브로브루네아(*Micropolyspora cabrobrunea*), 마이크로폴리스포라 썬모비리다(*Micropolyspora thermovirida*), 마이크로폴리스포라 비리디니그라(*Micropolyspora viridinigra*), 메타노박테리움 썬모오토트로피쿰(*Methanobacterium thermoautotrophicum*), 칼디셀룰로시럽터 아세티게누스(*Caldicellulosiruptor acetigenus*), 칼디셀룰로시럽터 사카롤리티쿠스(*Caldicellulosiruptor saccharolyticus*), 칼디셀룰로시럽터 크리스티안소니(*Caldicellulosiruptor kristjanssonii*), 칼디셀룰로시럽터 오웬센시스(*Caldicellulosiruptor owensensis*), 칼디셀룰로시럽터 락토아세티쿠스(*Caldicellulosiruptor lactoaceticus*), 이들의 변이체(variant), 및/또는 이들의 프로제니(progeny)를 포함한다.

[0169] 특정 구체예에서, 본 발명은 페르비도박테리움 곤드와넨세(*Fervidobacterium gondwanense*), 클로스트리디움 썬모락티쿰(*Clostridium thermolacticum*), 무렐라 종(*Moorella sp.*), 및 로도썬무스 마리누스(*Rhodothermus marinus*)로 구성된 군으로부터 선택된 호열성 세균에 관한 것이다.

[0170] 특정 구체예에서, 본 발명은 썬모안에어로박테리움 또는 썬모안에어로박터 속의 호열성 세균에 관한 것이며, 이러한 세균은 비제한적으로 썬모안에어로박테리움 썬모술푸리게네스(*Thermoanaerobacterium thermosulfurigenes*), 썬모안에어로박테리움 아오테아로엔세(*Thermoanaerobacterium aotearoense*), 썬모안에어로박테리움 폴리사카롤리티쿰(*Thermoanaerobacterium polysaccharolyticum*), 썬모안에어로박테리움 제에(*Thermoanaerobacterium zeae*), 썬모안에어로박테리움 자일라노리티쿰(*Thermoanaerobacterium xylanolyticum*), 썬모안에어로박테리움 사카롤리티쿰(*Thermoanaerobacterium saccharolyticum*), 썬모안에어로비움 브록키(*Thermoanaerobium brockii*), 썬모안에어로박테리움 썬모사카롤리티쿰(*Thermoanaerobacterium thermosaccharolyticum*), 썬모안에어로박터 썬모히드로술푸리쿠스(*Thermoanaerobacter thermohydrosulfuricus*), 썬모안에어로박터 에타놀리쿠스(*Thermoanaerobacter ethanolicus*), 썬모안에어로박터 브록키(*Thermo anaerobacter brocki*), 이들의 변이체, 및 이들의 프로제니로 구성된 군으로부터 선택된 종을 포함한다.

[0171] 특정 구체예에서, 본 발명은 게오바실루스, 사카로코쿠스, 패니바실루스, 바실루스, 및 안옥시바실루스 속의 미생물에 관한 것이며, 이러한 미생물은 비제한적으로 게오바실루스 썬모글루코시다시우스(*Geobacillus thermoglucosidasius*), 게오바실루스 스테아로썬모필루스(*Geobacillus stearothermophilus*), 사카로코쿠스 칼독실로실리티쿠스(*Saccharococcus caldxylosilyticus*), 사카로쿠스 썬모필루스(*Saccharoccus thermophilus*), 패니바실루스 캄피나센시스(*Paenibacillus campinasensis*), 바실루스 플라보썬무스(*Bacillus flavothermus*), 안옥시바실루스 캄차트켄시스(*Anoxybacillus kamchatkensis*), 안옥시바실루스 고넨시스(*Anoxybacillus gonensis*), 이들의 변이체, 및 이들의 프로제니로 구성된 군으로부터 선택된 종을 포함한다.

[0172] 특정 구체예에서, 본 발명은 사카로파구스 데그라단스(*Saccharophagus degradans*); 플라보박테리움 존소니아(*Flavobacterium johnsoniae*); 피브로박터 숙시노게네스(*Fibrobacter succinogenes*); 클로스트리디움 훈가테이(*Clostridium hungatei*); 클로스트리디움 피토펙테르멘탄스(*Clostridium phytofermentans*); 클로스트리디움 셀룰로리티쿰(*Clostridium cellulolyticum*); 클로스트리디움 알드리치(*Clostridium aldrichii*); 클로스트리디움 테르미티디디스(*Clostridium termitididis*); 아세티비브리오 셀룰로리티쿠스(*Acetivibrio cellulolyticus*); 아세티비브리오 에타놀기그넨스(*Acetivibrio ethanoligignens*); 아세티비브리오 멀티보란스(*Acetivibrio multivorans*); 박테리로이즈 셀룰로솔벤스(*Bacteroides cellulosolvens*); 및 알칼리박터 사카로포멘탄스(*Alkalibacter saccharofomentans*), 이들의 변이체 및 이들의 프로제니로 구성된 군으로부터 선택된 중온성 세균에 관한 것이다.

[0173] 본 발명의 방법

[0174] 해당과정 동안, 세포는 글루코오스와 같은 단순 당을 피루브산으로 전환시키며 ATP와 NADH를 궁극적으로 생산한다. 산화적 인산화를 위한 기능적 전자 수송 시스템이 없는 경우, 지속적 해당과정 및 ATP 생산을 위한 필수 요건인  $NAD^+$ 를 재생하는 짧은 경로에서 피루브산의 적어도 95%가 소비된다. 이러한  $NAD^+$  재생 시스템의 폐생산물(waste product)은 일반적으로 발효 생산물로 지칭된다.

[0175] 미생물은 다양한 일련의 발효 생산물을 생산하며, 그러한 발효 생산물은 유기산, 예를 들어 락테이트 (락트산의

염 형태), 아세테이트 (아세트산의 염 형태), 숙시네이트, 및 부티레이트, 및 중성 생산물(neutral product), 예를 들어 에탄올, 부탄올, 아세톤, 및 부탄디올을 포함한다. 최종 발효 생산물은 다음과 같은 것들을 포함하는 수 가지 기본적인 특징을 다양한 정도로 공유한다: 그러한 생산물은 이들이 최초로 생산되는 조건하에서 비교적 비독성이지만 축적시에 보다 독성으로 되고; 그리고 그러한 생산물은 피루베이트 보다 더 감소하는데, 이는 해당과정 동안 이들의 직접적 전구체가 최종 전자 수용체로서 작용했기 때문이다. 본 발명의 일면들은 리그노셀룰로오스 바이오매스 기질로부터 에탄올을 생산하는 데에 유용한 신규한 미생물을 제공하기 위해 유전자 녹아웃 기술을 사용하는 것에 관한 것이다. 형질전환된 생물체는 본원에 기재된 유기산으로의 비제한 경로(non-limiting pathway)와 같은 경합 경로를 엔코딩하는 하나 이상의 유전자를 결실 또는 비활성화시킴으로써 제조되거나, 상기와 같이 결실 또는 비활성화시키는 것에 이어 임의로 발효 생산물로서 에탄올을 생산하는 데에 있어서 개선된 성능을 지닌 돌연변이체를 증식을 기초로 선별함으로써 제조된다.

[0176] 특정 구체예에서, 고유한 상태에서 발효 생산물로서 락트산을 생산하는 능력을 부여하는 하나 이상의 유전자를 함유하는 호열성 또는 중온성 미생물은 상기 하나 이상의 유전자의 발현을 감소시키거나 제거하도록 형질전환된다. 발효 생산물로서 락트산을 생산하는 능력을 상기 미생물에게 부여하는 유전자는 락테이트 데히드로게나아제의 발현을 코딩할 수 있다. LDH의 발현을 엔코딩하는 유전자(들) 또는 특정 폴리뉴클레오티드 서열(들)이 결실되거나 억제되면 피루베이트가 락트산으로 전환되게 하는 전체 해당과정 경로에서의 반응 구성(reaction scheme)을 축소시키거나 제거하며, 해당과정의 이러한 초기 단계들로부터의 피루베이트의 결과적인 상대 존재비(relative abundance)는 에탄올의 생산 증가를 가능하게 할 수 있다.

[0177] 특정 구체예에서, 고유한 상태에서 발효 생산물로서 아세트산을 생산하는 능력을 부여하는 하나 이상의 유전자를 함유하는 호열성 또는 중온성 미생물은 상기 하나 이상의 유전자의 발현을 제거하도록 형질전환된다. 발효 생산물로서 아세트산을 생산하는 능력을 상기 미생물에게 부여하는 유전자는 아세테이트 키나아제 및/또는 포스포트랜스아세틸라아제의 발현을 코딩할 수 있다. ACK 및/또는 PTA의 발현을 엔코딩하는 유전자(들) 또는 특정 폴리뉴클레오티드 서열(들)이 결실되거나 억제되면 아세틸 CoA가 아세트산으로 전환되게 하는 전체 해당과정 경로에서의 반응 구성을 축소시키거나 제거하며 (도 1), 해당과정의 이러한 나중 단계들로부터의 아세틸 CoA의 결과적인 상대 존재비는 에탄올의 생산 증가를 가능하게 할 수 있다.

[0178] 특정 구체예에서, 상기 상세히 설명된 유전자 녹아웃 계획들은 개별적으로 또는 동시에 적용될 수 있다. 락테이트의 생산을 위한 메커니즘이 제거되면 (즉, LDH의 발현을 엔코딩하는 유전자 또는 특정 폴리뉴클레오티드 서열이 녹아웃되면) 더 많은 아세틸 CoA를 생성시키며, 이에 따라 아세테이트의 생산을 위한 메커니즘이 또한 제거되는 경우 (즉, ACK 및/또는 PTA의 발현을 엔코딩하는 유전자 또는 특정 폴리뉴클레오티드 서열이 녹아웃되는 경우), 아세틸 CoA의 존재비가 추가로 향상될 것이고, 이는 에탄올의 생산 증가를 초래할 수 있다.

[0179] 특정 구체예에서, 호열성 또는 중온성 미생물은 고유의 또는 내생성 PDC 및/또는 ADH를 지닐 필요가 없다. 특정 구체예에서, PDC 및/또는 ADH를 엔코딩하는 유전자는 본 발명의 유전자 변형된 미생물에서 재조합적으로 발현될 수 있다. 특정 구체예에서, 본 발명의 유전자 녹아웃 기술은 재조합 미생물에 적용될 수 있는데, 이러한 재조합 미생물은 PDC 및/또는 ADH를 코딩하는 이종성 유전자를 포함할 수 있으며, 여기서 상기 이종성 유전자는 발효 생산물로서 에탄올을 생산하는 상기 재조합 미생물 (호열성일 수 있음)의 능력을 증가시키거나 발효 생산물로서 에탄올을 생산하는 능력을 상기 재조합 미생물 (호열성일 수 있음)에게 부여하기에 충분한 수준으로 발현된다.

[0180] 특정 구체예에서, 본 발명의 일면들은 셀룰로오스 함량 또는 헤미셀룰로오스 함량 또는 둘 모두의 함량을 기준으로 하여 이론적 수율의 70% 이상인 농도로 에탄올을 생산하도록 리그노셀룰로오스 기질을 발효시키는 것에 관한 것이다.

[0181] 특정 구체예에서, 본 발명의 일면들은 셀룰로오스 함량 또는 헤미셀룰로오스 함량 또는 둘 모두의 함량을 기준으로 하여 이론적 수율의 80% 이상인 농도로 에탄올을 생산하도록 리그노셀룰로오스 기질을 발효시키는 것에 관한 것이다.

[0182] 특정 구체예에서, 본 발명의 일면들은 셀룰로오스 함량 또는 헤미셀룰로오스 함량 또는 둘 모두의 함량을 기준으로 하여 이론적 수율의 90% 이상인 농도로 에탄올을 생산하도록 리그노셀룰로오스 기질을 발효시키는 것에 관한 것이다.

[0183] 특정 구체예에서, 고유한 상태의 미생물로부터 유기산 생산의 실질적 또는 완전한 제거는 하나 이상의 부위-지정 DNA 상동 재조합 사건을 이용하여 달성될 수 있다.

- [0184] 동시 당화 공동-발효(simultaneous saccharification and co-fermentation, SSCF) 또는 CBP 공정을 호열성 온도에서 작업하는 것은 30 내지 37°C의 통상적인 중온성 발효 온도에 비해 수 가지 중요한 이점을 제공한다. 특히, 셀룰라아제 생산에 대한 전용 공정 단계를 위한 비용이 호열성 SSCF의 경우 상당히 감소되고 (예를 들어, 2배 또는 그 초과), CBP의 경우에는 제거된다. 발효기 냉각과 관련된 비용 그리고 또한 발효 전후의 열교환은 호열성 SSCF 및 CBP 둘 모두의 경우에 감소될 것으로 예상된다. 최종적으로, 호열성 생물축매를 특징으로 하는 공정은 통상적인 중온성 생물축매를 특징으로 하는 공정에 비해 미생물에 덜 오염될 수 있다.
- [0185] 탄소 흐름에 대한 변화에 의해 전자 흐름의 방향을 바꾸는 능력은 광범위한 관련성을 지닌다. 예를 들어, 이러한 방법은 써모안애어로박테리움 사카롤리티쿰 이외의 균주에서 높은 에탄올 수율을 생산하고/하거나 에탄올 이외의 용매, 예를 들어 고급 알코올 (즉, 부탄올)을 생산하기 위해 이용될 수 있다.
- [0186] *안티센스 올리고뉴클레오티드 (asRNA) 전략을 통한 대사 엔지니어링*
- [0187] 효모 및 혐기성 세균과 같은 발효 미생물은 당을 에탄올 및 그 밖의 환원된 유기 최종 생산물로 발효시킨다. 이론적으로, 락테이트 및 아세테이트와 같은 경합 최종-생산물의 형성이 억제될 수 있는 경우 탄소 흐름이 에탄올 생산쪽으로 유도될 수 있다. 본 발명은 본 발명의 CBP 생물체에서 이러한 경합 경로를 제거하도록 설계된 수 가지 유전공학 방법을 제공한다. 이러한 방법들의 대부분은 녹-아웃 구성물 (단일 크로소브 재조합의 경우) 또는 대립유전자-교환(allele-exchange) 구성물 (이중 크로소브 재조합의 경우)을 이용하고, *ack* 및 *ldh*에 대한 유전자좌를 표적화한다. 이러한 도구는 "입증된(tried and true)" 균주 개발 기술을 이용하지만, 진전을 저지할 수 있는 다음과 같은 수 가지 잠재적 문제가 존재한다: (i) 상기 도구는 모든 경우에 CBP 생물체에 대해 알려져 있지 않은 숙주 재조합 효율에 의존하고; (ii) 상기 도구는 한 번에 단지 하나의 경로를 녹아웃시키도록 사용될 수 있고, 이로써 연속적인 유전자 변화는 수 개의 선별가능 마커 또는 리사이클가능(recyclable) 마커를 필요로 하고; (iii) 표적 유전자의 결실은 유해할 수 있거나 다운스트림 유전자 발현에 대해 극성 효과(polar effect)를 지닐 수 있다.
- [0188] 본 발명은 숙주 재조합 효율에 좌우되지 않는 유전공학을 향한 추가의 방법을 제공한다. 이러한 대안적 도구들 중 하나는 안티센스 RNA (asRNA)라 일컬어진다. 안티센스 올리고뉴클레오티드가 시험관내 및 생체내 둘 모두에서 유전자 발현 수준을 억제하기 위해 25년에 걸쳐 사용되어 왔지만, mRNA 구조 예측에서의 최근 진보는 asRNA 분자의 보다 세련된 설계를 촉진시켰다. 이러한 진보는 다수의 집단으로 하여금 세균의 대사 엔지니어링에서 asRNA의 유용성을 입증하게 하였다.
- [0189] asRNA를 사용한 경우 녹-아웃 및 대립유전자-교환 기술에 비해 다음과 같은 수 많은 이점이 존재한다: (i) 다수의 선별가능 마커에 대한 필요성이 완화되는데, 이는 다수의 경로가 단일 asRNA 구성물에 의해 표적화될 수 있기 때문이고; (ii) 표적 mRNA의 약화 수준(attenuation level)이 asRNA 사이의 회합율(association rate)을 증가시키거나 감소시킴으로써 조정될 수 있고; (iii) asRNA 전사체가 조건적 프로모터(conditional promoter)에 의해 유도되는 경우 경로 비활성화가 조건적일 수 있다. 최근에는, 이러한 기술이 그람 양성 중온성세균인 클로스트리디움 아세토부틸리쿰(*Clostridium acetobutylicum*)에서 용매생성(solventogenesis)을 증가시키기 위해 사용되어 왔다 (Tummala et al. (2003)). asRNA가 어떻게 유전자 발현을 약화시키는지와 관련된 정확한 분자 메커니즘은 불명확하지만, 유망한 메커니즘은 asRNA가 표적 mRNA에 하이브리드화된 경우에 촉발된다. 메커니즘은 하기 중 하나 이상을 포함할 수 있다: (i) 리보솜 결합 부위가 리보솜과 적절히 상호작용하는 것을 차단함으로써 mRNA가 단백질로 번역되는 것을 억제하고, (ii) 듀플렉스 RNA를 신속히 분해하는 RNase H와 같은 dsRNA-의존성 RNase를 통해 mRNA의 반감기를 감소시키고, (iii) mRNA의 초기 전사 종결로 인해 전사를 억제한다.
- [0190] *안티센스 서열의 설계*
- [0191] as RNA는 전형적으로 길이가 18개 내지 25개의 뉴클레오티드이다. asRNA 서열을 선택하기 위해 사용될 수 있는 RNA-표적화 핵산의 합리적 설계를 위해 이용가능한 수 개의 계산 도구(computation tool)가 존재한다 (Sfold, Integrated DNA Technologies, STZ Nucleic Acid Design). 예를 들어, 클로스트리디움 써모셀룸 *ack* (아세테이트 키나아제)에 대한 유전자 서열이 합리적 설계 서버(server)에 적용될 수 있고, 수 개의 asRNA 서열이 추려질 수 있다. 요약하면, 설계 파라미터는 예측된 2차 구조를 함유하지 않는 mRNA 표적 서열을 선별해낸다.
- [0192] *전달 벡터의 설계*
- [0193] asRNA 코딩 서열을 표적 생물체에 전달하기 위해 복제성 플라스미드가 사용될 것이다. 비제한적으로 pNW33N, pJIR418, pJIR751, 및 pCTC1와 같은 벡터는 asRNA 코딩 서열을 숙주 세포 내부로 전달하기 위한 asRNA 구성물의 백본(backbone)을 형성할 것이다. 염색체의 (플라스미드 기반) 발현 이외에, asRNA는 asRNA의 안정한 발현



을 획득하기 위해 이중성 유전자좌에서 미생물의 게놈내로 안정하게 삽입될 수 있다. 특정 구체예에서, 관심을 끄는 호열성 또는 중온성 미생물의 균주는 유기산의 생산을 녹아웃시키기 위해 부위 지정 상동 재조합에 의해 엔지니어링될 수 있고, 관심을 끄는 다른 유전자는 asRNA에 의해 부분적으로, 실질적으로 또는 완전히 결실되거나 침묵되거나 비활성화되거나 하향조절될 수 있다.

[0194] 프로모터 선택

[0195] asRNA 전사체의 발현을 보장하기 위해, 주어진 숙주에 대해 적합한 프로모터가 asRNA 코딩 서열에 융합될 것이다. 프로모터-asRNA 카세트(cassette)는 단일 PCR 단계로 구성된다. 프로모터 영역을 증폭시키도록 설계된 센스 및 안티센스 프라이머는, asRNA 서열 (합리적 설계 방법으로부터 추려진 것)이 안티센스 프라이머의 5' 단부에 부착되도록 변형될 것이다. 또한, 최종 PCR 앰플리콘(amplicon)이 제한 효소에 의해 직접 분해될 수 있고 통상적인 클로닝 기술을 통해 벡터 백본내로 삽입될 수 있도록, *EcoRI* 또는 *BamHI*와 같은 제한 부위가 각각의 프라이머의 말단부에 첨가될 것이다.

[0196] 본원에 기재된 바와 같은 펜토오스 당을 대사하는 능력을 지니지 않지만 헥소오스 당을 대사할 수 있는 미생물과 관련하여, 예를 들어 *클로스트리디움 썬모셀럼* 및 *클로스트리디움 스트라미니솔벤스*의 *ack* 및 *ldh* 유전자가 본원에 기재된 방법에 따라 안티센스 RNA를 사용하여 비활성화를 위해 표적화될 수 있는 것으로 인식될 것이다.

[0197] 본원에 기재된 바와 같은 펜토오스 및 헥소오스 당을 대사하는 능력을 부여하는 미생물과 관련하여, 예를 들어 *클로스트리디움 셀룰로리티쿰*, *클로스트리디움 피토펙레멘탄스* 및 *칼디셀룰로시룸터 크리스티안소니*의 *ack* 및 *ldh* 유전자가 본원에 기재된 방법에 따라 안티센스를 사용하여 비활성화를 위해 표적화될 수 있는 것으로 인식될 것이다.

[0198] asRNA 전달 벡터를 발현하는 균주에 대한 항생제 선별 이외에, 그러한 균주는 소듐 플루오로아세테이트(SFA), 브로모아세트산(BAA), 클로로아세트산(CAA), 5-플루오로오로트산(5-FOA) 및 클로로락트산과 같은 수 가지 독성 대사물 중 어느 것을 함유하는 조건적 배지(conditional media)에서 선별될 수 있다. 부분적으로, 실질적으로 또는 완전히 결실되거나 침묵되거나 비활성화되거나 하향조절된 하나 또는 그 초과 유전자를 지닌 균주를 생성시키기 위해, 에탄 메틸 설포네이트(EMS)를 포함하지만 이것에 국한되지 않는 화학적 돌연변이원을 사용하는 것이 안티센스 올리고뉴클레오타이드(asRNA)를 발현시키는 것과 병용하여 이용될 수 있다.

[0199] 예증

[0200] 본 발명은 지금까지 일반적으로 설명되었고, 하기 실시예들을 참조로 하여 보다 용이하게 이해될 것인데, 이러한 실시예들은 단지 본 발명의 특정 일면 및 구체예를 예시하기 위해 포함되며 본 발명을 제한하고자 하는 것이 아니다.

[0201] 실시예 1

[0202] 중온성 및 호열성인 셀룰로오스분해성, 자일란분해성 생물체를 위한 맞춤형 트랜스포손(custom transposon)의 생성

[0203] 본 발명은 셀룰로오스분해성 및/또는 자일란분해성 및/또는 호열성 생물체를 위한 맞춤형 트랜스포손을 생성시키는 방법을 제공한다. 이렇게 하기 위해, 숙주 생물체로부터의 고유 프로모터가 이러한 생물체에서 작용하는 것으로 결정된 선별가능 마커에 융합된다. 이러한 단편은 벡터 pMOD™-2<MCS> (Epicenter®Biotechnologies) 상에 함유된 EZ-Tn5™ 트랜스포손내로 클로닝된다. 예를 들어, *클로스트리디움 썬모셀럼*의 *gapDH* 프로모터가 mLs 약물 마커 뿐만 아니라 *cat* 유전자에 융합된 후, 벡터 pMOD™-2<MCS>내로 서브클로닝된다.

[0204] 시판용 트랜스포손은 셀룰로오스분해성 및/또는 자일란분해성 및/또는 호열성 생물체의 고유 프로모터 및 열안정성 약물 마커를 결합하고 있다. mLs 및 *cat* 마커는 호열성 세균에서 기능하여 왔고, *gapDH* 프로모터는 핵심 해당과정 효소를 조절하며 항시적으로 발현될 수 있다. 상기 약물 마커와 *gapDH* 프로모터가 병용되면 기능적 트랜스포손을 생성시킬 가능성을 크게 향상시킨다. 이러한 방법은 다른 셀룰로오스분해 및/또는 자일란분해성 및/또는 호열성 생물체에 대해 적용될 수 있다.

[0205] 실험 설계

[0206] 도 26은 본원에 참조로 포함된 에피센터®바이오테크놀로지스(Epicenter®Biotechnologies) 사용자 매뉴얼로부터 획득된 도면으로서, pMOD™-2<MCS>의 bp 250-550을 나타낸다. 상부 부분에서, ME로 표지된 검은색 화살촉은 트랜스포손을 규정하는 19 bp 모자이크 단부를 나타낸다. *EcoRI* 및 *HindIII* 부위는 다중-클로닝 부위를 규정하



며, 이는 MCS로 표지된 검은색 박스에 의해 표시된다. 하부 부분에는, MCS와 관련된 제한 효소 및 DNA 서열이 도시되어 있다.

[0207] pMQ87-*gapDH*-cat 및 pMQ87-*gapDH*-mls로부터의 프로모터 융합 단편을 증폭시키기 위해 하기 프라이머가 사용된다: GGCGgaattc CTT GGT CTG ACA ATC GAT GC (SEQ ID NO: 19); GGCGgaattc TATCAGTTATTACCCACTTTTCG (SEQ ID NO:20). 소문자는 엔지니어링된 EcoBI 제한 부위를 나타낸다. 생성된 앰플리콘의 크기는 ~1.9 kb이다. 표준 분자적 방법은 앰플리콘이 *EcoRI*로 분해되어 pMOD™-2<MCS>의 독특한 *EcoRI* 부위내로 클로닝될 수 있게 한다. 제조업자에 의해 설명된 바와 같이 트랜스포손 그리고 후속하여 트랜스포소솜(transpososome)이 생성되어 숙주 생물체내로 도입된다.

[0208] 실시예 2

[0209] **셀룰로오스분해성 및 자일란분해성 균주를 엔지니어링하기 위한 구성물**

[0210] 본 발명은 탄소 흐름을 에탄올쪽이 아닌 다른 방향으로 전환시키는 대사 경로의 핵심 효소를 엔코딩하는 유전자를 돌연변이시킴으로써 CBP를 위해 관심있는 생물체를 유전공학처리하기 위한 조성물 및 방법을 제공한다. 외래 DNA의 거대한 단편을 관심있는 유전자내로 삽입시켜서 이러한 유전자를 부분적으로, 실질적으로 또는 완전히 결실시키거나 침묵시키거나 비활성화시키거나 하향조절하도록, 단일 크로소버 녹아웃 구성물이 설계된다. 염색체로부터 관심있는 유전자를 부분적으로, 실질적으로 또는 완전히 결실시키거나 침묵시키거나 비활성화시키거나 하향조절하도록, 또는 염색체상의 관심있는 유전자를 항생제 내성 카세트에 의해 간접된 유전자의 형태와 같은 그러한 유전자의 돌연변이된 카피로 치환하도록, 이중 크로소버 녹아웃 구성물이 설계된다.

[0211] 단일 크로소버 녹아웃 벡터의 설계는 관심있는 유전자의 내부 단편을 플라스미드 기반 시스템내로 클로닝하는 것을 필요로 한다. 이상적으로는, 이러한 벡터는 숙주 균주에서 발현되지만 숙주 균주에서 복제되지 않는 선별 가능 마커를 함유한다. 따라서, 숙주 균주내로 도입시에, 플라스미드는 복제되지 않는다. 세포가 플라스미드 상에 함유된 마커를 선별해내는 조건적 배지내에 위치하는 경우, 플라스미드를 유지시키는 방법을 찾아낸 세포만이 증식한다. 플라스미드가 자율적 DNA 엘리먼트로서 복제할 수 없기 때문에, 플라스미드가 유지되는 가장 유망한 방법은 재조합을 통해 숙주 염색체상으로 통합되는 것이다. 재조합이 일어나는 가장 유망한 위치는 플라스미드와 숙주 염색체 사이의 상동성 영역에 존재한다.

[0212] 또한, 복제성 플라스미드는 단일 크로소버 간섭을 생성시키기 위해 사용될 수 있다. 녹아웃 벡터를 흡수한 세포는 조건적 배지에서 선별될 수 있고, 그 후 선별없이 계대될 수 있다. 조건적 배지에 의해 제공되는 양성 선별이 없다면, 많은 생물체는 플라스미드를 상실한다. 플라스미드가 숙주 염색체상으로 삽입되는 경우, 그러한 플라스미드는 선별없이 상실되지 않는다. 그 후, 세포는 조건적 배지로 복귀될 수 있고, 염색체 통합을 통해 마커를 보유하는 세포만이 증식한다. 염색체상에 위치한 마커를 함유하는 생물체를 스크리닝하도록 PCR 기반 방법이 고안된다.

[0213] 이중 크로소버 녹아웃 벡터의 설계는 관심있는 유전자를 플랭킹하는 DNA (~ 1 kb)를 플라스미드내로 클로닝하는 것을 적어도 필요로 하고, 일부 경우에는 관심있는 유전자를 클로닝하는 것을 포함할 수 있다. 선별가능 마커는 플랭킹 DNA 사이에 위치할 수 있거나, 관심있는 유전자가 클로닝되는 경우 마커는 그러한 유전자와 관련하여 내부에 위치한다. 이상적으로는 사용되는 플라스미드는 숙주 균주에서 복제될 수 없다. 플라스미드가 숙주내로 도입되고 마커에 대해 조건적인 배지에서 선별시에, 상동 DNA가 염색체상으로 재조합된 세포만이 증식한다. 관심있는 유전자를 선별가능 마커로 치환하기 위해 2회의 재조합 사건이 필요하다.

[0214] 또한, 이중 크로소버 유전자 치환을 생성시키기 위해 복제성 플라스미드가 사용될 수 있다. 녹아웃 벡터를 흡수한 세포는 조건적 배지에서 선별될 수 있고, 그 후 선별없이 계대될 수 있다. 조건적 배지에 의해 제공되는 양성 선별이 없다면, 많은 생물체는 플라스미드를 상실한다. 약물 마커가 숙주 염색체상으로 삽입되는 경우, 그러한 약물 마커는 선별없이 상실되지 않는다. 그 후, 세포는 조건적 배지로 복귀될 수 있고, 염색체 통합을 통해 마커를 보유하는 세포만이 증식한다. 염색체상에 위치한 마커를 함유하는 생물체를 스크리닝하도록 PCR 기반 방법이 고안될 수 있다.

[0215] 항생제 선별 계획 이외에, 소듐 플루오로아세테이트(SFA), 브로모아세트산(BAA), 클로로아세트산(CAA), 5-플루오로오로트산(5-FOA) 및 클로로라트산과 같은 수 가지 독성 대사물 유사체가 상동 재조합 또는 트랜스포손-기반 전략으로부터 발생하는 돌연변이체를 선별하기 위해 사용될 수 있다. 에탄 메틸 설포네이트(EMS)를 포함하지만 이것에 국한되지 않는 화학적 돌연변이원을 사용하는 것이 상동 재조합 또는 트랜스포손-기반 전략을 이용하는 지정 돌연변이유발 계획과 병용하여 이용될 수 있다.

- [0216] *C. 셀룰로리티쿰* 녹아웃 구성물
- [0217] 아세테이트 키나아제 (*클로스트리디움 셀룰로리티쿰* 공개 게놈으로부터의 유전자 131)
- [0218] 단일 크로스오버
- [0219] 클로스트리디움 셀룰로리티쿰의 아세테이트 키나아제 유전자는 길이가 1,110 bp이다. 누클레오티드 91-752에 걸쳐있는 662 bp 내부 단편 (SEQ ID NO:21)을 PCT에 의해 증폭시키고, 다양한 선별가능 마커를 지니는 복제성 벡터 및 자살 벡터(suicide vector)내로 클로닝하였다. 선별가능 마커는 에리트로마이신 및 클로람페니콜 내성을 제공하는 것들을 포함할 수 있다. 이러한 플라스미드는 *ack* 유전자를 붕괴시키기 위해 사용된다. *ack* 유전자의 맵 및 유전자 붕괴를 위해 PCR에 의해 증폭되는 영역은 도 19에 도시되어 있다. 하기 제시된 SEQ ID NO:21의 밑줄그어진 부분은 녹아웃 단편을 플랭킹하는 *EcoBI* 부위인 부위에 상응한다.
- [0220] gaattctgcgcagaatagggtgacaattccttataaagcaatcaagggttcagaagaggctgtattttgaataaagagctaaagaat  
cacaagatgcaatagaggctgtatttctgcactgactgacgataatggcggtataaaaaacatgccgaatatcagcagtgaggaca  
cagaatagtagacggcggtgaaaaattcaacagttctgtatagatgaaacgttatgaatgcagtaagagagtgtagacgttgac  
gcttcataatcgccgaatattataggtatagaggctgccagcagattatgcccaataacattggtagctgtattgataccacttccaca  
gctccatgcctgattatgcatacctttacgcatgccaatgaactttatgaaaagtagcggtataagaaaatatggttccacgggaacatcac  
aaatatgtgcagaaagagcttctgcaatgcttgataagctttgaacgaattaagataattacatgccatctgggaacgggtcaagtattgt  
gctgttaacaagggttaaatcaattgatacttccatgggctttacaccttgagggtgcaatgggtacaagaagcggtacaatagacct  
gaagttgtacgaattc
- [0221] 이러한 부위는 "ack KO 프라이머"의 설계 동안 엔지니어링되었고, 다수의 벡터내로의 단편의 후속 클로닝을 가능하게 한다.
- [0222] 이중 크로스오버
- [0223] 클로스트리디움 셀룰로리티쿰의 *ack* 유전자에 대한 이중 크로스오버 벡터를 구성하기 위해, *ack* 유전자의 각 측면을 플랭킹하는 ~1 kb의 DNA가 클로닝된다. 선별가능 마커는 플랭킹 DNA 사이에 삽입된다. 선별가능 마커는 에리트로마이신 및 클로람페니콜 내성을 제공하는 것들을 포함할 수 있다. *ack* 유전자의 3' 플랭킹 영역은 이용가능한 드래프트 게놈(draft genome)에서 이용가능하지 않다. 이러한 DNA를 획득하기 위해, 클론테크(Clontech)로부터의 게놈 워커(Genome Walker)와 같은 키트가 사용된다.
- [0224] 락테이트 데히드로게나아제 (*클로스트리디움 셀룰로리티쿰* 공개 게놈의 유전자 2262 및 2744):
- [0225] 단일 크로스오버
- [0226] 클로스트리디움 셀룰로리티쿰의 *ldh* 유전자는 길이가 951 bp (유전자 2262의 경우) (SEQ ID NO:22) 및 932 bp (유전자 2744의 경우) (SEQ ID NO:23)이다. 각각의 유전자의 5' 단부 근처의 ~500 bp 내부 단편이 PCR에 의해 증폭되고, 다양한 선별가능 마커를 지니는 복제성 벡터 및 자살 벡터내로 클로닝된다. 선별가능 마커는 에리트로마이신 및 클로람페니콜과 같은 약물 내성을 제공하는 것들을 포함할 수 있다. 이러한 플라스미드는 *ldh* 2262 및 *ldh* 2744 유전자를 붕괴시키기 위해 사용된다. 예로서, *ldh* 2262 유전자의 맵 및 유전자 붕괴를 위해 PCR에 의해 증폭되는 영역은 도 20에 도시되어 있다.
- [0227] 이중 크로스오버
- [0228] 클로스트리디움 셀룰로리티쿰의 *ldh* 유전자(들)에 대한 이중 크로스오버 벡터를 구성하기 위해, *ldh* 유전자(들)의 각 측면을 플랭킹하는 ~1 kb의 DNA가 클로닝된다. 선별가능 마커는 플랭킹 DNA 사이에 삽입된다. 선별가능 마커는 에리트로마이신 및 클로람페니콜과 같은 약물 내성을 제공하는 것들을 포함할 수 있다. 도 21은 클로스트리디움 셀룰로리티쿰 *ldh* (2262) 이중 크로스오버 녹아웃 단편의 예를 제공한다.
- [0229] 하기 제시된 서열 (SEQ ID NO:24)에서, *mLs* 유전자 (선별가능 마커)가 밑줄그어져 있고, 플랭킹 DNA는 나머지 서열이다. 프라이머 설계 동안, 상기 단편이 다수의 복제성 및 비복제성 벡터내로 클로닝될 수 있도록 제한 부위들이 엔지니어링되고 그러한 단편의 5' 및 3' 단부가 엔지니어링된다. 동일한 전략이 *ldh* 2744를 결실시키기 위한 벡터를 생성시키기 위해 사용된다.

gacgcatacaggttgtaacacccatttcccttagcttttgggagatgaataaaacaaacatttccgggtctttaccacacccgccacataaa  
gagctatgccgcatgaaagaaacgatatgttatcatTTTTTctgtaaacgtttattccgaacccggataaagctttaccatatttactgctgcc  
gtccctgcatgtgacacccctataaccactatttcatatcatctctctgttggctgtgtaaatatccatataaccacctaataatatttataa  
acaaatccggtatatactcttggtaataaaaaglacatccgatatagaaatgtacctataaaaaaattattttattgtatgtctttatctgttt  
cattatagtgttctatccattctacggtaaaatcaagtaattccattaaagtaactgacctgtatcttctatctctgtataatccgtattactgattt  
ctcaataaaatcatggttgaacttggggagagaagcttgcgatatcctatgctatgcatgtattcttctcataggttaaaatgaagacagt  
gtaactttttagttccgtaattagccgtacaaatfctacataatttgcgttaataagctgattttctggcctcataaattccgaagcaatcggaaat  
agtttctatgctgtctgtatcttcaattccaagaataaattcgctctcctattctatcatatggacccctcaaatgtaatgtataccaagatta  
tacatactcctagaataataacaaggaataaaatttaataatcgtatatacctacataaatgactaaagctctcaaaaacttctttta  
ttattctatactactaaaaatcaaaaatfctctaaagfcttctcaaaatgttttttgcacaaagtagtatacttttgacccagaatgtttgtta  
taacttcaaaataggggtatatttatagtaatactaaatggaagagtaggatatgtatgaacgagagaaaaataaaaacacagtcaaa  
actttattacttcaaaacataatagataaaaataatgaacataaagataaataacatgataatcttgaatcggtcaggaaggaagg  
gcattttaccctgaattagtcacaggggttaatttgcgaacgtccattgaaatgacacataaattatgcaaaactcagagaaataaaactgtttg  
atcacgataatttccaaagttttaaacagagatatgtcagtttaaaatttccaaaaacaaatcctataaaatattgtataatccttataacata  
agtcacgataataacgcaaaatgttttgcagatagatgagatgagattttaaactgtggaatacgggtttgcataaaagatttataatacaa  
aacgctcattgtcattattttaaagtcagaaagttgatatttctatataaglatggttccaaagaaatatttccctaaacctaagtgaaatgct  
cacttcatgattaaatgaaaaaaatcaagaaatcacacaaagataaaacagagataatttgcgttatgaaatgggttaacaaagaa  
tacaagaaaaatatttcaaaaaatcaatttaacaaatccttaaaacatgcaggaattgacgatttaaacaaatattgactttgaacaaattctatct  
ctttcaatagctataaattttaaataagatcccttatactcggatgcatgcgcaggcaggcatccgaagtagtttctccattatacaagtatct  
cttgtagtactgtcgtctcagcagctgtcttgcctttccgttttccggcacatggagataaagtgatctgtaggcttaatagtgtgtgcat  
gtcaattgccttttgcagagtcactgtccttattttaaaggtttccacaaatgataaaacccgtatcagtcagaaatttactacccgctgatatct  
gtgtctgaacccgtcctataagataggttgcacccaacctgaattccatgaagctgaggtgtctccagcagcttactaaagcatgagat  
attagatgctcactacccgtggtgagcagctgtctgtctatctgcatggcaattccgctcattgtcagagagctaccatttcttataaaag  
aagtttctgtaaacctgtgtgtagggcatcttcaataactgtttactgacttttagcaatcattgcagcaaaatctcaacctttgcgcgcaattgttc  
ctttctcaaaataccagtcacacacggttaatttggatattatgtctccgagacctgaataaataaattcataggtgcattttataacatct  
aaatccactaattattccaaatggcagtcagggcagtcaggaagtagcctgcatttataataaagagcagcctgagctggaaaaacat  
cgtttgaggttgatgtaggtatctgataaaaggaagctgtttaaanaagctatatttggctgcatcaagcaccttctctctactcgcgac  
cactgcatcgttttggaggaatagtaaagcctgagcataagatttcaagcttattgtcatcatagtcgtaagttcaagtagctgcaagag  
atttctgtactttagaatccgaatcttttccaaaaatagtcacgtattcctctccaaaagtagtacaacattactaatttctccttca  
tatgtgc

[0230]

[0231] C. 클로스트리디움 피토펙스멘탄스 녹아웃 구성물

[0232] 아세테이트 키나아제 (클로스트리디움 피토펙스멘탄스 공개 계놈으로부터의 유전자 327)의 경우:

[0233] 단일 크로스오버

[0234] 클로스트리디움 피토펙스멘탄스의 아세테이트 키나아제 유전자는 길이가 1,244 bp이다. 뉴클레오티드 55-626에 걸쳐있는 572 bp 내부 단편이 PCR에 의해 증폭되고, 다양한 선별가능 마커를 지니는 복제성 벡터 및 자살 벡터 내로 클로닝된다. 선별가능 마커는 클로스트리디움 피토펙스멘탄스에 약물 내성을 제공하는 것들을 포함한다. 이러한 플라스미드는 ack 유전자를 붕괴시키기 위해 사용된다. ack 유전자의 맵 및 유전자 붕괴를 위해 PCR에 의해 증폭되는 영역은 도 22에 도시되어 있다. "ack KO 프라이머"의 설계 동안 제한 부위가 엔지니어링되고, 이는 단편이 다수의 벡터내로 후속 클로닝될 수 있게 한다. 상기 기재된 녹아웃 단편의 서열은 SEQ ID NO:25로서 제시된다.

[0235] 이중 크로스오버

[0236] 클로스트리디움 피토펙스멘탄스의 ack 유전자에 대한 이중 크로스오버 녹아웃 벡터를 구성하기 위해, ack 유전자의 각 측면을 플랭킹하는 ~1 kb의 DNA가 클로닝된다. 선별가능 마커는 플랭킹 DNA 사이에 삽입된다. 사용하려는 선별가능 마커는 이러한 군주에 약물 내성을 제공하는 것들을 포함한다. 추정 선별가능 마커로서 mLs 유전자를 지니는 추정 이중 크로스오버 녹아웃 구성물의 예는 도 23에 도시되어 있다.

[0237] 도 23에 도시된 단편에 상응하는 서열 (SEQ ID NO:26)이 하기 제시되어 있다. mLs 유전자 (추정 선별가능 마커)가 밑줄그어져 있고, 서열의 나머지는 플랭킹 DNA에 상응한다. 프라이머 설계 동안, 상기 단편이 다수의 복제성 및 비복제성 벡터내로 클로닝될 수 있도록 제한 부위들이 엔지니어링되고 그러한 단편의 5' 및 3' 단부가 엔지니어링된다.

ctgagtgcaatgtaaaaaaggatgctcaagttcttgaacatccttattatactacaaaatcataaagtaaaattactcagctgtgcaat  
gatctctttttgtgaagatccacaagcttllacaaactctatgagggatcataagtgaccacacttgctgatttaccatggttggagcagtc  
atcttcaggttgcaagcagactatctctctgcttgggaatgttattcttggacaaatagctccattgattacacctctttaaactgttataaaa  
tatctcgatagcagacattctgggtcagttctgtacgggtcacacccgactctctctttaggttagcaccgcagacctgagattcttt  
acagttcttctgcacagaacctcattgggaaccaatcaagactcttcatagataagttattctacgtctaaatcatatccggaaacaaaatt  
tgttcatctaaatcctcggtacgtgttctctgtttcgatacatcaatctctgttagccaogtcatgtcttggatggttctctctcaaacac  
gatcgcaaggaaogcgctaacgttaatttctgtccaccagaattttcgccacctagattagttatctaaagtttaacgggttcttataag  
gtaatagaataaccgacacatttaattcgaatatacaaatcaatcggtgagtgattctttagaccattaggaacattcatgacttcaga  
cattgtatcagcataagtaactctgtctaaaaaaacgcataatgtaagcgcccaaaaattcacactgttagtattataaacgcttaaaatag  
gtttgcaactcctaaactgttaaaaaatgctagaattgtgtaaccatatttctctcattatcggttctcccttataaataattatagctattgaaaaga  
gataaagaattgttcaaacgtaattgtttaaactcgtcaattctgtcatgttttaagggaattgttaaatgtattttgtataattttctgtattcttgtta  
acccattcataacgaaataataactctgtttatcttctgtgataattctgtattttttctatttaactgataagtgagcattacattaggtttacgga  
gaaaattctcttggaccatacttaataatagaataatcaactctgcccattaaaaataatgccaatgagcgtttgtatttaataatctttagca  
aacccgtatttcacgattaaataaactcatcagctatactatacaaaaacaaatttgcgtattatccgtactatgtataaaggtatattaccaaa  
tattttataggattgttttaggaaatttaaaactgcaatatactctgttttaaaacttgggaattatcgtgatacaacaagttatttctgtattttgcat  
aatttatggtctatttcaatgagcattacgaaatcaacactctgtaacttaacagggttaaaatgcccctttccgtgagccgatttcaaaagataatc  
atgttcatltaattcttatttctgattttttatctatatttcttgaagtaataaagtttgaactgtgttttataattttctgttcatgtatttctctataat  
gttctaattcattatcacggggcaacttaataatccgaatatagttctctatctgctcccccagataatgattattatactatttaattctca  
cttaacaaattggagtttccagtttaagaaataaatttaattgaagcaggaatttcgcaatccgcttcatcgtctactgataacctcaacagg  
caatgaagctaaagttaattatttactctgtgctgaacagcagtgattgcaacaacccaacgataatcagagaagaacaccccttgataaa  
cttactctggagctgcaataccctgagtttaattggtccataagctctgcttgcgaagcgtggttaactatataccaattgtaccagcatcaa  
ggctgggaagattatacgtatgctttccagcaatatactacacagcaggtttgaagcacctacacaggaacgattgctgcatcactg  
gaactcgcgtgcatcttatattctgggtataattcattgcaatcttagttgcttcaacactatcaacatctgcatgcttgcgcttcccttgggtga  
atgagaaagcatagctacgataaggttcagagccaactaattgttcaaaactcttgcgtggaacagcgattgctgtaactctctcagcattt  
ggattctgaattaaaccagcatcagagaaaggaagttccatttgcgcccatacaaatagggtactacattacgaagaagcagaa  
ctaactagatttggagcagtttttaaaactcgaagacatggtcttaaggatctgctgtagagtgacaagcaccagatacactaaacatctgca  
tcgcccacttaacacatcacccgtatgtaattgtagtctgtgttaaaagctctttgtctttcaggggtcagcttggcctgtctaagttctaca  
agctgtgtaattgaagc

[0238]

[0239] *락테이트 데히드로게나아제 (클로스트리디움 피토펙르멘탄스 공개 계놈의 유전자 1389 및 2971)의 경우*

[0240] *단일 크로스오버*

[0241] 클로스트리디움 피토펙르멘탄스의 *ldh* 유전자는 길이가 978 bp (유전자 1389의 경우) (SEQ ID NO:27) 및 960 bp (유전자 2971의 경우) (SEQ ID NO:28)이다. 각각의 유전자의 5' 단부 근처의 ~500 bp 내부 단편이 PCR에 의해 증폭되고, 다양한 선별가능 마커를 지니는 복제성 벡터 및 자살 벡터내로 클로닝된다. 사용하려는 선별가능 마커는 약물 내성을 제공하는 것들을 포함한다. 이러한 플라스미드는 *ldh* 1389 및 *ldh* 2971 유전자를 붕괴시키기 위해 사용된다. 예로서, *ldh* 1389 유전자의 맵 및 유전자 붕괴를 위해 PCR에 의해 증폭되는 영역은 도 24에 도시되어 있다.

[0242] *이중 크로스오버*

[0243] 클로스트리디움 피토펙르멘탄스의 *ldh* 유전자(들)에 대한 이중 크로스오버 벡터를 구성하기 위해, *ldh* 유전자(들)의 각 측면을 플랭킹하는 ~1 kb의 DNA가 클로닝된다. 선별가능 마커는 플랭킹 DNA 사이에 삽입된다. 선별가능 마커는 이러한 균주에 약물 내성을 제공하는 것들을 포함한다. 추정 선별가능 마커로서 mLs 유전자를 지니는 추정 이중 크로스오버 녹아웃 구성물의 예는 도 25에 도시되어 있다.

[0244] 도 25에 도시된 단편에 상응하는 서열이 SEQ ID NO:29로서 하기 제시되어 있다. mLs 유전자 (선별가능 마커)가 밑줄그어져 있고, 서열의 나머지 부분은 플랭킹 DNA에 상응한다. 프라이머 설계 동안, 상기 단편이 다수의 복제성 및 비복제성 벡터내로 클로닝될 수 있도록 제한 부위들이 엔지니어링되고 그러한 단편의 5' 및 3' 단부가 엔지니어링된다. 동일한 전략이 *ldh* 2971을 결실시키기 위한 벡터를 생성시키기 위해 사용된다.



tggaatctcactatgacccaatgtggtactaattatcttctatctatggaattagggtttccgcgaatggagatagaggagctgccattgc  
tacttaattgtagaattcttgagagattttagttgattatgtataaggggtgagaaggctactaagatgagacttcttatattttaagagatct  
aaacagtatcttgccttcttggtcgttatgtggtccagctgcttatgagtgagggttaactgggggttggtgattgctgtcagctgcaatcattgg  
gctgtatgggtgttagtttctacagccgcaagcttctaatagtatgacacagcttagccggaatcattgaltggtattggtgtggttgcagc  
attataatagggaattgtatgttgagggaagaagagcatgaggcgagaatgtagcccaataagtaacgtatcagatgatactcggagg  
aattgtgctttgcagtaactctacgtcccaatcgctcctaacttattgaggcgctcaaggaaacagcggatttaattcgtcagatgctattgtt  
tcggcttacccttattctccaagccttctgtatctaataattcataggttttgatggtccattttacgttagcagatgatactacatacaaaacattct  
tgacctatgggtactaaaacttgaggaggtttgctgcaacatagtagtctcatctccactgtatgggttactttacttctaagttagcagatgag  
tggttaagcgctatttaccggtaccgggtgcttaagggaagcttggtattcatgatacaactgcattaaagattcatatgtccagatattttg  
cacagtagcataaattactagagcttattcctataataattcataggttttgatggtccattttacgttagcagatgatactacatacaaaacattct  
atataagtagggtatagtaggaacggaataataaaacacagctcaaaacttattacttcaaaacataatataagataaaataatgaca  
aatataagattaaatgaacatgataatcttgaatcgggtcaggaaagggcatttaccctgaattagtagcagaggtgtaatttctgaac  
tgccattgaaatagacccaataatttcaaaactacagaaaataaaacttctgacacgataatttccaagitttaacaaggatatttgcagt  
ttaaatttctaaaaacccaatcctataaaatttggtaatacttataacataagtagcagataataacgcaaaatttctttgtagatagct  
gatgagatttattatcgtggaatcgggtttgctaaagatttataatacaaaaacgctcattggcatttttaattgagcagaagttgatatttct  
atattaagtagtcttccaagagaaatttctcctaaacctaagtagcctcattcagattaaatagaaaaaatcaagaatcac  
acaaagatataacagaagataaatttctgtagaattggttaacaaagaatacaagaataatttcaaaaaatcaatttaacaattcct  
aaaacatgcaaggaattgacgatttcaacaataatgcttgaacaattcttcttcttcaatagctataaatttataaagaagtaalaggaa  
ataactcgaattattctgcaactctgttctaaaaataaaatgaagaattactatagcaagccaggtttaaattactagctgtctattttgtgca  
tttagtagcagttttgattataaagaataaatttaataactatttgcataaagttatgactatttcaaggttaggttactatacaagtagaata  
aagatacataaaaaataaataatgaacataaattcatgacatgcggaatagaatgaagaatattatgctggttctactataactaaatg  
gatataacaactctattgaaacacttattgggtgtgagtgagagaatttctaagcgccaaagactcactacatgaaattctaagcttca  
cagcggaataatcattatgtatcttattatcataattcaggaggttagtgtaaaataaaaaatagtttctcgttctcattcaggcagtagca  
ttcttaaacaaattgtcgtatggttctggttctgaaacacaaaagcaatttctcacaacttattcgaacaacaaatggtattaaaaattg  
gaggaggttttactatgaaaaaaacggtaacattactgttgggttctgacatgggtggaagcttatttgcagcatgtgttaagaaaaatggat  
caagcgaaacccgcacaaaaagatcctgtgcaacaagcggtgcaaaagaaacctgacaaaacagatccagcgaataaagagcctga  
aaaaacagaccctgttaaaatcaagatttactctgataatgcaaccttaccattaaagaagattggttagtataaaggaaagcgtgagaa  
gagatttaattgttgaattcgaattcgaagtaattccaattgcagattatcaacaaaagtttctttaaactaaatcagggaataacgctccagat  
gcatcttctatcagtagcaacgaggagagaatgcatct

[0245]

[0246]

칼디셀룰로시펄터 크리스티안소니 및 클로스트리디움 스테코라리움 아종 랩토사파르툼

[0247]

본 발명자들이 알고 있는 바로는, 상기 생물체들의 게놈 시퀀싱은 이루어진 바 없으며, 게놈 시퀀싱이 이루어졌다고 하더라도 공중에게 이용가능하지 않았다. 본 발명자들의 실험 결과에 기초해 볼 때, 이러한 생물체들은 셀룰로오스분해성이고 자일란분해성이다. 이러한 생물체들을 유전공학처리하고 탄소 흐름을 에탄올쪽으로 전환시키기 위해 핵심 대사 효소를 엔코딩하는 유전자의 DNA 서열이 이러한 생물체들로부터 필요하다. 이러한 생물체들은 아세테이트 키나아제 및 락테이트 데히드로게나아제와 같은 효소를 포함한다. 이러한 유전자들의 서열을 획득하기 위해, 이러한 생물체들의 게놈이 시퀀싱된다.

[0248]

게놈 서열이 이용가능해진 경우, 엔코딩 유전자와 플랭킹 DNA를 찾아내기 위해 상기 효소들의 보존된 특성이 사용될 수 있다. 이러한 서열들은 단일 및 이중 크로스오버 전략 둘 모두를 이용하는 표적화된 돌연변이유발을 위한 구성물을 설계하기 위해 사용된다. 이러한 전략들은 상기 기재된 것들과 동일하다. 또한, 본 발명자들은 이러한 생물체들에서 어떤 항생제가 선별가능 마커로서 사용될 수 있고 어떤 형질전환용 프로토콜이 가장 잘 기능하는 지를 결정할 것이다.

[0249]

실시예 3

[0250]

### 클로스트리디움 셀룰로리디움의 형질전환

[0251]

세포를 4g/l의 셀로비오스를 지닌 50 mL의 GS 배지중에서 혐기성 조건에서 0.8의 OD로 증식시키고, 34℃에서 인큐베이션하였다. 수거한 후, 상기 세포를 500mM 수크로오스 및 5mM MOPS를 함유하는 세척 완충액을 사용하여 동일한 부피로 3회 세척하고, pH를 7로 조정하였다. 최종 세척 후, 세포 펠릿(pellet)을 동일한 부피의 세척 완충액중에 재현탁시키고, 세포 현탁액의 10ul 분취액을 1mm 전극 간격으로 표준 일렉트로포레이션 큐벳(cuvette)에 넣었다. 1ul 플라스미드 DNA를 첨가하였다. 1:1 내지 10:1의 플라스미드 대 세포의 몰비를 보장하도록 플라스미드 DNA의 농도를 조정하였다. 7kV/cm의 장세기(field strength) (측정치)를 사용하여 샘플을 가로질러 5ms 펄스를 가하였다. 맞춤형 펄스 발생기를 사용하였다. 최초 배양에서 사용된 동일한 배지를 사용하여 샘플을 1000:1로 즉시 희석시키고, 증식이 재개될 때까지 회복시키고, OD의 증가를 통해 결정하였다 (24시간 내지 48시간). 회복된 샘플을 50:1로 희석시키고, 15ug/mL의 에리트로마이신 또는 15ug/mL의 클로람페니콜을 지닌 선별 배지에 넣고, 5일 내지 6일간 증식시켰다. 선별 배지에서 증식을 나타내는 샘플이 실제로 클로스트리디움 셀룰로리디움이며 플라스미드를 지니는 지를 확인하는 시험을 하였다.

[0252]

실시예 4

[0253]

### 셀룰로오스분해성 균주를 엔지니어링하기 위한 구성물

[0254]

셀룰로오스는 바이오매스의 주요 성분들 중 하나이며, 이는 클로스트리디움 썬모셀룸에 의한 발효에 의해 연료 에탄올을 생성시키기 위한 기질로서 잠재적으로 사용될 수 있다. 그러나, 이러한 과정에서, 부산물인 아세테이트와 락테이트를 형성하기 위해 많은 에너지와 탄소원이 사용된다. 락테이트와 아세테이트 생산을 최소화하고



에탄올 형성에 유리한 탄소 흐름 및 에너지를 만들어내기 위해 클로스트리디움 써모셀룸에서 셀룰로오스 이용의 대사 경로를 엔지니어링할 필요가 있다.

[0255] 아세테이트 키나아제는 클로스트리디움 써모셀룸에서 아세테이트를 형성하기 위한 셀룰로오스 이용의 대사 경로에서 중요한 효소이며, 이는 *ack* 유전자에 의해 엔코딩된다. *ack* 유전자의 비활성화는 아세테이트 키나아제를 간접할 수 있고, 이는 아세테이트의 감소 또는 제거를 초래한다.

[0256] 락테이트 데히드로게나아제는 클로스트리디움 써모셀룸에서 락테이트를 형성하기 위한 셀룰로오스 이용의 대사 경로에서 중요한 효소이며, 이는 *ldh* 유전자에 의해 엔코딩된다. *ldh* 유전자의 비활성화는 락테이트 데히드로게나아제를 간접할 수 있고, 이는 락테이트 생성의 감소 또는 제거를 초래한다.

[0257] 플라스미드 *pIKM1*를 기반으로 하는 클로스트리디움 써모셀룸에서의 *ack* 유전자의 비활성화

[0258] *ack* 유전자를 녹아웃시키기 위해, 플라스미드 *pIKM1*의 다중 클로닝 부위(MCS)를 기초로 하여 벡터가 구성되는데, 여기서 클로람페니콜 아세틸트랜스퍼라아제를 엔코딩하는 *cat* 유전자가 3055 bp의 DNA 단편내로 삽입되며, 이러한 DNA 단편은 *ack* 및 *pta* 유전자 (포스포트랜스아세틸라아제를 엔코딩함)를 포함하고, *ack* 유전자의 476 bp 및 *pta* 유전자의 399 bp의 녹아웃을 초래하고, *mLs* 유전자의 양쪽 측면상에 각각 1025 bp 및 1048 bp의 플랭킹 영역을 형성한다 (도 7). *pNW33N*은, 바실루스 코굴란스 및 스태필로코쿠스 아우레우스 (*Staphylococcus aureus*)로부터 분리된 *pBC1* 레플리콘(replicon)을 함유하고, 이는 클로스트리디움 써모셀룸을 포함하는 그람 양성 세균 균주에서 안정하게 복제되는 것으로 예견된다. 플라스미드 *pIKM1*을 기초로 하여 구성된 *ack* 녹아웃 벡터의 서열이 SEQ ID NO:1로서 제시된다.

[0259] 복제성 플라스미드 *pNW33N*을 기반으로 하는 클로스트리디움 써모셀룸에서의 *ack* 유전자의 비활성화

[0260] *ack* 유전자를 녹아웃시키기 위해, 복제성 플라스미드 *pNW33N*의 다중 클로닝 부위(MCS)를 기초로 하여 벡터가 구성되는데, 여기서 마크롤리드(macrolide), 링코사미드(lincosamide) 및 스트렙토그라민 B(streptogramin B) (*MLS<sub>B</sub>*) 내성 유전자 *mLs*가 3345 bp의 DNA 단편내로 삽입되며, 이러한 DNA 단편은 *ack* 유전자, *pta* 유전자 (포스포트랜스아세틸라아제를 엔코딩함) 및 미지의(unknown) 엽스트림 유전자를 포함하고, *ack* 유전자의 855 bp의 녹아웃을 초래하고, *mLs* 유전자의 각각의 측면상에 1195 bp 및 1301 bp의 플랭킹 영역을 형성한다 (도 8). *pNW33N*은, 바실루스 코굴란스 및 스태필로코쿠스 아우레우스로부터 분리된 *pBC1* 레플리콘을 함유하고, 이는 클로스트리디움 써모셀룸을 포함하는 그람 양성 세균 균주에서 안정하게 복제되는 것으로 예견된다. 플라스미드 *pNW33N*을 기초로 하여 구성된 *ack* 녹아웃 벡터의 서열이 SEQ ID NO:2로서 제시된다.

[0261] 플라스미드 *pIKM1*를 기반으로 하는 클로스트리디움 써모셀룸에서의 *ldh* 유전자의 비활성화

[0262] *ldh* 유전자를 녹아웃시키기 위해, 플라스미드 *pIKM1*의 다중 클로닝 부위(MCS)를 기초로 하여 벡터가 구성되는데, 여기서 클로람페니콜 아세틸트랜스퍼라아제를 엔코딩하는 *cat* 유전자가 3188 bp의 DNA 단편내로 삽입되며, 이러한 DNA 단편은 *ldh* 및 *mdh* 유전자 (말레이트 데히드로게나아제를 엔코딩함)를 포함하고, *ldh* 유전자의 1171 bp의 DNA 단편의 녹아웃을 초래하고, *ldh* 및 *mdh* 유전자의 일부를 포함하고, *mLs* 유전자의 양쪽 측면상에 각각 894 bp 및 1123 bp의 플랭킹 영역을 형성한다 (도 9). 플라스미드 *pIKM1*을 기초로 하여 구성된 *ldh* 녹아웃 벡터의 서열이 SEQ ID NO:3으로서 제시된다.

[0263] 플라스미드 *pNW33N*을 기반으로 하는 클로스트리디움 써모셀룸에서의 *ldh* 유전자의 비활성화

[0264] *ldh* 유전자를 녹아웃시키기 위해, 복제성 플라스미드 *pNW33N*의 다중 클로닝 부위(MCS)를 기초로 하여 벡터가 구성되는데, 여기서 마크롤리드, 링코사미드 및 스트렙토그라민 B (*MLS<sub>B</sub>*) 내성 유전자 *mLs*가 2523 bp의 DNA 단편내로 삽입되며, 이러한 DNA 단편은 *ldh* 유전자 및 *mdh* 유전자 (말레이트 데히드로게나아제를 엔코딩함)를 포함하고, *ldh* 유전자의 489 bp의 단편의 녹아웃을 초래하고, *mLs* 유전자의 각각의 측면상에 1034 bp 및 1000 bp의 플랭킹 영역을 형성한다 (도 10). *pNW33N*은, 바실루스 코굴란스 및 스태필로코쿠스 아우레우스로부터 분리된 *pBC1* 레플리콘을 함유하고, 이는 클로스트리디움 써모셀룸을 포함하는 다른 그람 양성 세균 균주에서 안정하게 복제되는 것으로 예견된다. 플라스미드 *pNW33N*을 기초로 하여 구성된 *ldh* 녹아웃 벡터의 서열이 SEQ ID NO:4로서 제시된다.

[0265] 플라스미드 *pUC19*를 기반으로 하는 클로스트리디움 써모셀룸에서의 *ldh* 유전자의 비활성화

[0266] *ldh* 유전자를 녹아웃시키기 위해, *pUC19* 플라스미드의 다중 클로닝 부위(MCS)를 기초로 하여 벡터가 구성되는데, 여기서 클로람페니콜 아세틸트랜스퍼라아제를 엔코딩하는 *cat* 유전자가 717 bp의 *ldh* 유전자 단편

내로 삽입되며, 이러한 DNA 단편은 *cat* 유전자의 각각의 측면상에 245 bp 및 255 bp의 플랜킹 영역을 초래한다 (도 11). pUC19는 pMB1 원점(origin)을 함유하는 대장균 플라스미드 벡터이며, 이는 클로스트리디움 써모셀룸을 포함하는 그람 양성 세균 균주에서 증폭될 수 없다. *mLs* 유전자가 *ldh* 유전자 단편에 의해 플랜킹되어 있는 유사한 벡터가 구성될 수 있다. 플라스미드 pUC19를 기초로 하여 구성된 *ldh* 녹아웃 벡터의 서열이 SEQ ID NO:5로서 제시된다.

[0267] 클로스트리디움 써모셀룸 및 클로스트리디움 스트라미니슐벤스에서의 자일로오스 이소머라아제 및 자일롤로오스 키나아제의 발현 (예언적 실시예(prophetic example))

[0268] 클로스트리디움 써모셀룸에서 자일로오스 이소머라아제 및 자일롤로오스 키나아제를 발현하기 위해, 자일로오스 이소머라아제 및 자일롤로오스 키나아제 유전자를 써모안에어로박테리움 사카롤리티쿰으로부터 클로닝하고, 클로스트리디움 써모셀룸 *gapDH* 프로모터의 조절을 받게 하였다. 이러한 카세트를 pNW33N 백본을 기초로 하는 클로스트리디움 써모셀룸 복제성 플라스미드에 함유시켜서, pMU340 (도 35) SEQ ID NO:74를 생성시킨다. 클로스트리디움 써모셀룸내로 전달되는 경우, 생성된 형질전환이 자일로오스에서 증식하는 능력에 대해 검정될 수 있다. 유사한 구성물이 칼디셀룰로시류터 크리스티안소니 자일로오스 이소머라아제 및 자일롤로오스 키나아제 유전자를 사용하여 생성될 수 있다. 이들 구성물은 클로스트리디움 스트라미니슐벤스에서의 기능성에 대해 또한 시험될 수 있다.

[0269] 클로스트리디움 써모셀룸 및 클로스트리디움 스트라미니슐벤스에서의 피루베이트 데카르복실라아제 및 알코올 데히드로게나아제의 발현 (예언적 실시예)

[0270] 클로스트리디움 써모셀룸에서 피루베이트 데카르복실라아제 및 알코올 데히드로게나아제를 발현하기 위해, 피루베이트 데카르복실라아제 유전자를 공급원인 자이모모나스 모빌리스 및 자이모박터 팔메로부터 클로닝하고, 알코올 데히드로게나아제 유전자를 공급원인 자이모모나스 모빌리스로부터 클로닝한다. 이러한 유전자들 (*pdh* 및 *adh*)은 클로스트리디움 써모셀룸 *pta-ack* 프로모터로부터 오페론(operon)으로서 발현된다. 이러한 카세트를 SEQ ID NOS:75 및 76인 pNW33N 백본을 기초로 하는 클로스트리디움 써모셀룸 복제성 플라스미드 (도 36 및 37)에 함유시킨다. 클로스트리디움 써모셀룸내로 전달되는 경우, 생성된 형질전환은 발현된 효소의 기능성을 측정하기 위해 향상된 에탄올 생산 및/또는 알데히드 생산에 대해 스크리닝될 수 있다. 이들 구성물은 클로스트리디움 스트라미니슐벤스에서의 기능성에 대해 또한 시험된다.

[0271] 실시예 5

[0272] 클로스트리디움 스트라미니슐벤스를 사용하는 아비셀®의 발효

[0273] 클로스트리디움 스트라미니슐벤스를 사용하여 CTFUD 매질을 함유하는 혈청 병에서 1% 아비셀®을 발효시켰다. 생산물 농도 프로파일 및 비는 도 27에 도시되어 있다. 약 2 g/L의 전체 생산물이 3일내에 생성되었는데, 에탄올은 전체 생산물의 약 50%를 구성한다. 도 27은 클로스트리디움 스트라미니슐벤스를 사용하는 경우 1% 아비셀®에 대한 생산물 농도 프로파일을 도시한다. 에탄올 대 아세테이트 비는 E/A로서 표시되고, 에탄올 대 전체 생산물의 비는 E/T로서 표시되어 있다.

[0274] 실시예 6

[0275] 중온성 및 호열성 셀룰로오스분해성, 자일란분해성 생물체에 대한 엔지니어링된 그룹(Group) II 인트론

[0276] 많은 세균 계통에서 발견되는 이동성(mobile) 그룹 II 인트론은 촉매성 RNA이면서 레트로트랜스포서블(retrotransposable) 엘리먼트이다. 이러한 인트론은 레트로트랜스포지션(retrotransposition)으로서 알려진 이동 메커니즘(mobility mechanism)을 이용하는데, 상기 레트로트랜스포지션에서 절제된 인트론 RNA는 DNA 표적 부위내로 직접 역스플라이싱(reverse splice)된 후 인트론-엔코딩된 단백질에 의해 역전사된다. 이동성 락토코쿠스 락티스(*Lactococcus lactis*) L1.LtrB 그룹 II 인트론이 타게트론™(Targetron™) 벡터로서 공지된 유전학 도구로 개발되었고, 이러한 벡터는 시그마 알드리치(Sigma Aldrich)에 의해 시판되고 있다 (Catalog # TA0100). 이러한 제품 및 이의 용도가 인덱스, 엘엘씨.(InGex, LLC.)에 의해 관리되는 미국 특허 번호 5,698,421, 5,804,418, 5,869,634, 6,027,895, 6,001,608, 및 6,306,596 및/또는 다른 계류중인 미국 및 외국 특허 출원의 주제이다.

[0277] 레트로-트랜스포지션을 위해 필요한 모든 서열을 함유하는 타게트론 카세트 (도 28 및 29)는 중온성 또는 호열성 셀룰로오스분해성 생물체에서 복제될 수 있는 벡터내로 서브클로닝될 수 있다. 타게트론 카세트는 lac 프로모터를 임의의 숙주-특이적 또는 종-특이적 항시성 또는 유도성 프로모터로 치환함으로써 변형될 수 있다. 그

그룹 II 인트론이 관심있는 유전자내로 삽입되도록 리타겟팅되어 유전자 녹아웃을 생성시키게 되도록, 고유 인식 서열의 부위-지정 돌연변이유발을 통해 상기 카세트가 추가로 변형될 수 있다. 예를 들어, 그룹 II 인트론은 임의의 중온성 또는 호열성 셀룰로오스분해성 생물체에서 락테이트 데히드로게나아제 또는 아세테이트 키나아제를 녹아웃시키도록 재설계될 수 있다. 표 4는 인트론을 클로스트리디움 셀룰로리티쿰 아세테이트 키나아제 (SEQ ID NO:21)로 리타겟팅하기 위한 프라이머 및 삽입 위치의 예를 도시한다. 표 5는 인트론을 클로스트리디움 셀룰로리티쿰 락테이트 데히드로게나아제 (SEQ ID NO:21)로 리타겟팅하기 위한 프라이머 및 삽입 위치의 예를 도시한다.

[0278] 클로스트리디움 셀룰로리티쿰 ack 유전자 (SEQ ID NO:21)에 삽입되도록 L1.Ltrb 인트론을 리타겟팅하기 위한 벡터의 예는 도 28에 도시되어 있다. pMU367 (클로스트리디움 셀룰로리티쿰 아세테이트 키나아제 KO 벡터)의 벡터 서열은 SEQ ID NO:30이다.

[0279] 클로스트리디움 셀룰로리티쿰 LDH2744 유전자 (SEQ ID NO:23)에 삽입되도록 L1.Ltrb 인트론을 리타겟팅하기 위한 벡터의 예는 도 29에 도시되어 있다. pMU367 (클로스트리디움 셀룰로리티쿰 락테이트 데히드로게나아제 KO 벡터)의 벡터 서열은 SEQ ID NO:31로서 제시된다.

[0280] 표 4

|                                           |                                                               |
|-------------------------------------------|---------------------------------------------------------------|
| 예측된<br>삽입<br>위치 (SEQ<br>ID NO:62)         | ATTTACCTGGCTGGGAATACTGAGACATAT - intron - GTCATTGAGGCCGTA     |
| IBS1 돌연변이유발<br>프라이머 (SEQ ID<br>NO:63)     | AAAAAAGCTTATAATTATCCTTAATTCCTACTACGTGCGCCAGATAGGGTG           |
| EBS1d<br>돌연변이유발<br>프라이머 (SEQ ID<br>NO:64) | CAGATTGTACAAATGTGGTGATAACAGATAAGTCTACTACTGTAACCTTACCTTTCTTTGT |
| EBS2 돌연변이유발<br>프라이머 (SEQ ID<br>NO:65)     | TGAACGCAAGTTTCTAATTTTCGGTTGAAATCCGATAGAGGAAAGTGCT             |

[0281] 표 5

|                                              |                                                              |
|----------------------------------------------|--------------------------------------------------------------|
| 예측된<br>삽입<br>위치<br>(SEQ ID<br>NO:66)         | TTAAATGTTGATAAGGAAGCTCTTTTCAAT - intron - GAAGTTAAGGTAGCA    |
| IBS1<br>돌연변이유발<br>프라이머<br>(SEQ ID<br>NO:67)  | AAAAAAGCTTATAATTATCCTTAGCTCTCTTCAATGTGCGCCAGATAGGGTG         |
| EBS1d<br>돌연변이유발<br>프라이머<br>(SEQ ID<br>NO 68) | CAGATTGTACAAATGTGGTGATAACAGATAAGTCTTCAATGATAACTTACCTTTCTTTGT |
| EBS2<br>돌연변이유발<br>프라이머<br>(SEQ ID<br>NO:69)  | TGAACGCAAGTTTCTAATTTTCGATTAGAGCTCGATAGAGGAAAGTGCT            |

[0283] 실시예 7

[0284] **써모안에어로박터 및 써모안에어로박테리움 균주의 형질전환 (예언적 실시예)**

[0285] 써모안에어로박터 슈도에타놀리쿠스(*Thermoanaerobacter pseudoethanolicus*) 39E, 써모안에어로박테리움 사카롤리티쿰 JW/SL-YS485, 써모안에어로박테리움 사카롤리티쿰 B6A-RI, 및 써모안에어로박터 종 균주 59은 하기 프로토콜에 의해 형질전환된다. 세포를 다음과 같이 변경된 40 mL의 DSMZ M122 배지 [http://www.dsmz.de/microorganisms/media\\_list.php](http://www.dsmz.de/microorganisms/media_list.php) 중에서 55°C에서 광학 밀도가 0.6 내지 0.8이 될 때까지 증식시킨다: 셀룰로오스 대신 5 g/L의 셀로비오스, 1.8 g/L의 K<sub>2</sub>HPO<sub>4</sub>, 글루타티온 비함유, 및 0.5 g/L의 L-시스테인-HCl. 그 후, 세포를 수거하고, 실온에서 40 mL의 0.2 M 셀로비오스를 사용하여 2회 세척한다. 세포를 100 µL의 분취액 중의 0.2 M 셀로비오스에 재현탁시키고, 0.1 내지 1 µg의 플라스미드 DNA를 1 mm 갭-폭(gap-width) 일렉트로포레이션 큐벳 중의 샘플에 첨가한다. 1.8 kV, 25 µF, 200Ω, ~ 3-6 ms의 지수형 펄스(exponential pulse) (Bio-Rad Instruments)를 큐벳에 가하고, 세포를 새로운 M122 중에서 100 내지 200배 희석하고, 55°C에서 12

내지 16시간 동안 인큐베이션한다. 그 후, 회복된 세포를 200  $\mu$ g/mL의 카나마이신과 같은 선별제를 함유하는 새로운 한천-함유 배지를 지닌 페트리-플레이트에서 25 내지 100배 희석한다. 배지가 고화된 경우, 콜로니 형성을 위해 플레이트를 55°C에서 24 내지 72시간 동안 인큐베이션한다. 콜로니는 부위-특이적 재조합의 입증을 위해 PCR에 의해 시험될 수 있다.

[0287]

실시예 8

[0288]

#### 엔지니어링된 써모안에어로박터 및 써모안에어로박테리움 균주의 발효 성능

[0289]

표 6은 엔지니어링된 써모안에어로박터 및 써모안에어로박테리움 균주의 발효 성능을 나타낸다. 배양물을 진탕 시키지 않으며 55°C에서 M122 중에서 24시간 동안 증식시켰다. 하기 약어가 표 6에서 사용된다: 셀로비오스 (CB), 글루코오스 (G), 락트산 (LA), 아세트산 (AA), 및 에탄올 (EtoH). 수치는 리터 당 그램으로 표현된 것이다. YS485 - 써모안에어로박테리움 사카롤리티쿰 JW/SL-YS485, B6A-RI - 써모안에어로박테리움 사카롤리티쿰 B6A-RI, 39E - 써모안에어로박터 슈도에타놀리쿠스 39E.

[0290]

표 6

| 발효 샘플                                         | CB   | G    | LA   | AA   | EtoH |
|-----------------------------------------------|------|------|------|------|------|
| YS485 야생형                                     | 0    | 0    | 0.77 | 1.04 | 1.40 |
| YS485 $\Delta$ L-ldh                          | 0    | 0    | 0    | 0.92 | 1.73 |
| YS485 $\Delta$ pta/ack                        | 2.51 | 0    | 0.75 | 0.06 | 0.62 |
| YS485 $\Delta$ L-ldh, $\Delta$ pta/ack        | 0    | 0    | 0    | 0    | 2.69 |
| B6A-RI 야생형                                    | 0    | 0    | 0    | 1.0  | 1.76 |
| B6A-RI $\Delta$ L-ldh, $\Delta$ pta/ack 균주 #1 | 0    | 0    | 0    | 0    | 2.72 |
| B6A-RI $\Delta$ L-ldh, $\Delta$ pta/ack 균주 #2 | 0.45 | 0    | 0    | 0    | 2.49 |
| 39E 야생형                                       | 0.51 | 0    | 1.51 | 0.15 | 1.87 |
| 배지                                            | 5.10 | 0.25 | 0    | 0    | 0    |

[0291]

[0292]

실시예 9

[0293]

#### 셀룰로오스분해성 및 자일란분해성 균주를 엔지니어링하기 위한 구성물 - 안티센스 RNA 기술 실시예

[0294]

락테이트 데히드로게나아제를 코딩하는 클로스트리디움 써모셀룸 유전자 (Cthe\_1053)를 표적화하는 안티센스 RNA 카세트를 함유하는 복제성 플라스미드 (도 38)을 일렉트로포레이션 및 티아페니콜 선별에 의해 클로스트리디움 써모셀룸 1313에 전달하였다. 이러한 플라스미드에 대해 관찰된 형질전환 효율은 모벡터(parent vector)인 pMU102의 효율과 동일하였다. 상기 플라스미드의 서열은 SEQ ID NO:61로 제시된다. asRNA 카세트는 도 38에 도시되어 있고, 다음과 같이 조직되어 있다: (i) 전체 1827 bp 카세트가 도 38에 도시된 방향으로 pMU102의 다중클로닝 부위내로 클로닝되어 있고, (ii) 고유 프로모터 영역은 카세트의 최초 600 bp내에 함유되어 있고, (iii) ldh 오픈 리딩 프레임의 최초 877 bp은 안티센스 방향으로 고유 프로모터에 융합되어 있고, (iv) 약 300개의 추가 bp가 asRNA ldh 영역의 다운스트림에 포함되어 있다.

[0295]

도 39에 도시된 바와 같이, 6  $\mu$ g/mL의 티아페니콜의 존재하에서 (플라스미드를 유지시키기 위한 것임) M122C 배지에서 정치 배양물(standing culture)을 증식시킴으로써 변환된 최종 생산물 형성에 대해, 생성된 티아페니콜 내성 콜로니를 스크리닝하였다. 9개의 무작위 선택된 티아페니콜-내성 형질전환체의 예비 스크리닝은 4개의 배양물이 야생형에 비해 낮은 락테이트 생산 수준을 나타내었음을 보여주었다. 또한, 둘 모두의 ldh 유전자를 동시에 부분적으로, 실질적으로 또는 완전히 결실시키거나 침묵시키거나 비활성화시키거나 하향조절하기 위해, 둘 모두의 ldh 유전자에 대해 유도된 안티센스 RNA를 함유하는 구성물이 구성될 수 있다.

[0296]

실시예 10

[0297]

SEQ ID NO:44, 45, 및 46은 써모안에어로박테리움 사카롤리티쿰 YS485, 클로스트리디움 써모셀룸 ATCC 27405, 및 클로스트리디움 피토펙테멘탄스로부터의 피루베이트-포르메이트-리아제 (aka 포르메이트 아세틸트랜스퍼라아제, EC. 2.3.1.54, pf1) 유전자이다. Pf1은 피루베이트를 아세틸-CoA 및 포르메이트로 전환시키는 것을 촉매작용한다 (도 34). pf1의 결실은 포르메이트 생산의 제거를 초래하고, 이는 일부 호열성 균주에서 아세트산 수율의 감소를 초래할 수 있으며, 그 결과 에탄올 수율이 증가할 수 있다.

[0298]

도 40 내지 45에 도시된 SEQ ID NO:47 내지 52는 pf1 녹아웃 플라스미드를 도시하는데, 상기 기재된 3개의 생물체 각각에 대해 2개를 도시한다. 각각의 생물체는 pf1 효소를 부분적으로, 실질적으로 또는 완전히 결실시키거나 침묵시키거나 비활성화시키거나 하향조절하도록 설계된 단일 크로스오버 및 이중 크로스오버 플라스미드를

지닌다. 단일 크로스오버 플라스미드는 pfl 유전자의 내부 섹션(internal section)과 상동인 단일 DNA 서열 (400 bp 내지 1000 bp)을 지니도록 설계되고, 이중 크로스오버 플라스미드는 pfl 유전자의 업스트림 (5') 및 다운스트림 (3')에 있는 영역에 상동인 2개의 DNA 서열 (400 내지 1000 bp)을 지니도록 설계된다. 모든 플라스미드는 주어진 생물체에서의 선별을 위해 최상으로 이용가능한 항생제 마커를 사용하도록 설계된다. 플라스미드는 대장균에서 유지될 수 있고, 코돈 디바이시스(Codon Devices) 또는 DNA 2.0과 같은 DNA 합성 계약 회사(contract company)를 통해 구성될 수 있다.

[0299] 참조에 의한 포함

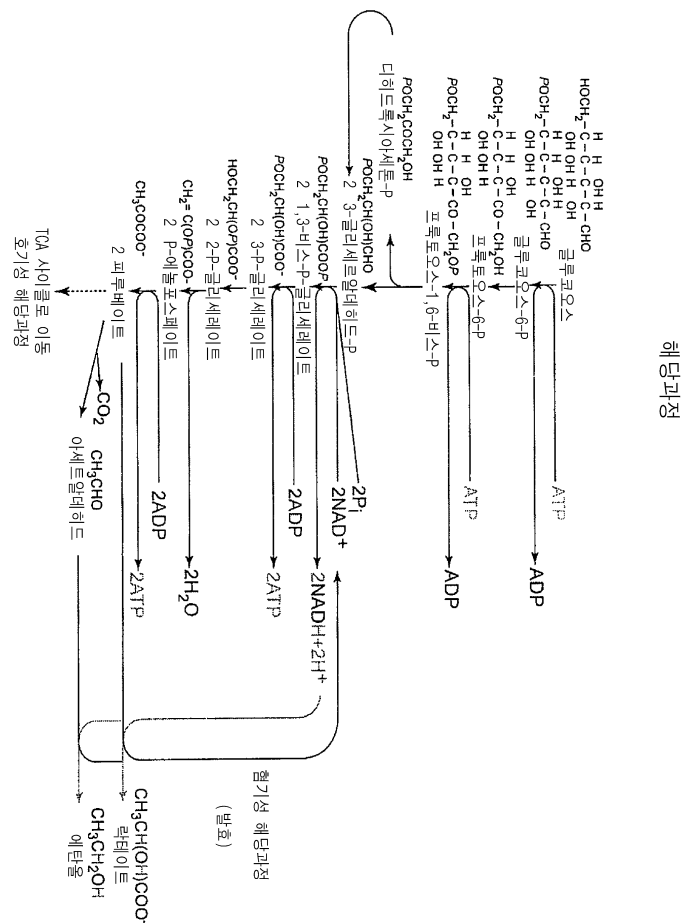
[0300] 본원에 인용된 모든 미국 특허 및 미국 공개 특허 출원은 본원에 참조로 포함된다.

[0301] 균등물

[0302] 당업자는 본원에 기재된 발명의 특정 구체예에 대한 많은 균등물을 인식하거나 단지 정례적 실험을 이용하여 그러한 균등물을 확인할 수 있을 것이다. 이러한 균등물은 하기 청구의 범위에 포함되는 것으로 의도된다.

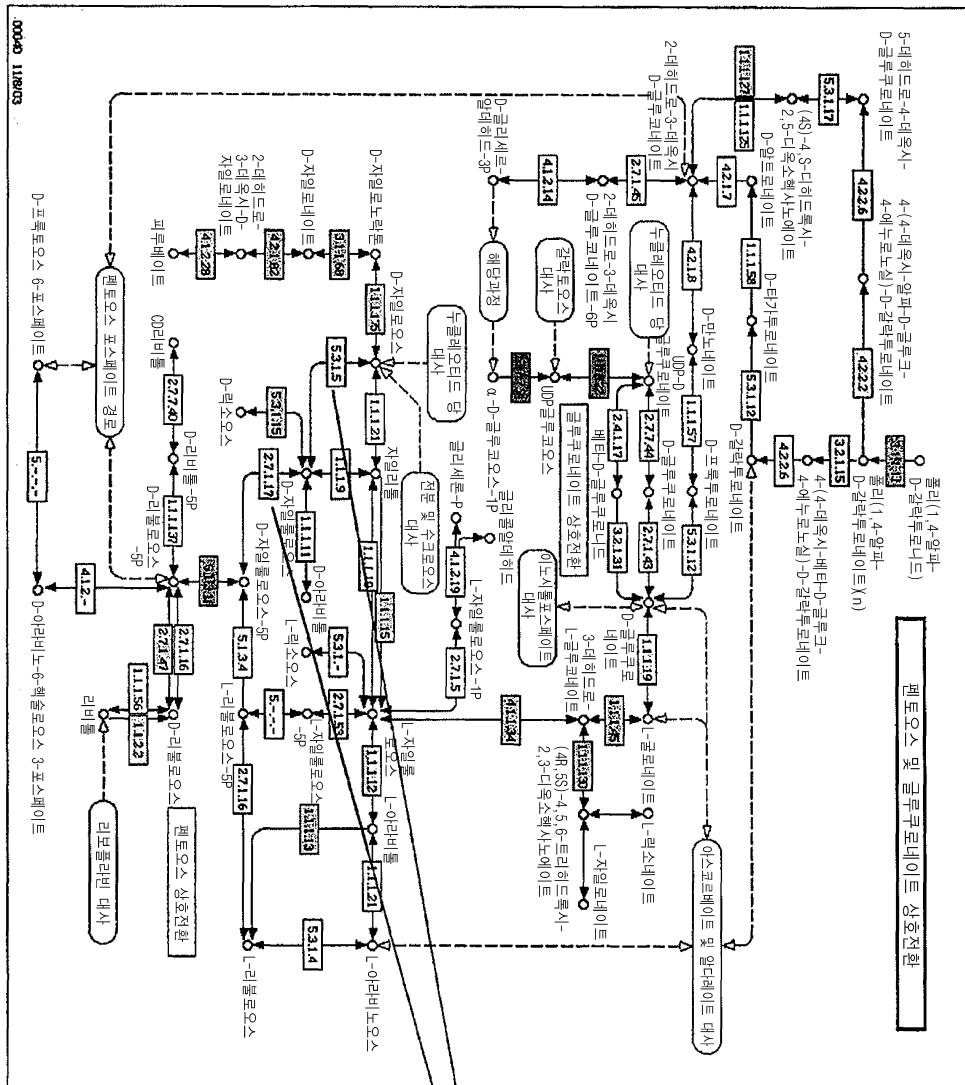
## 도면

도면1

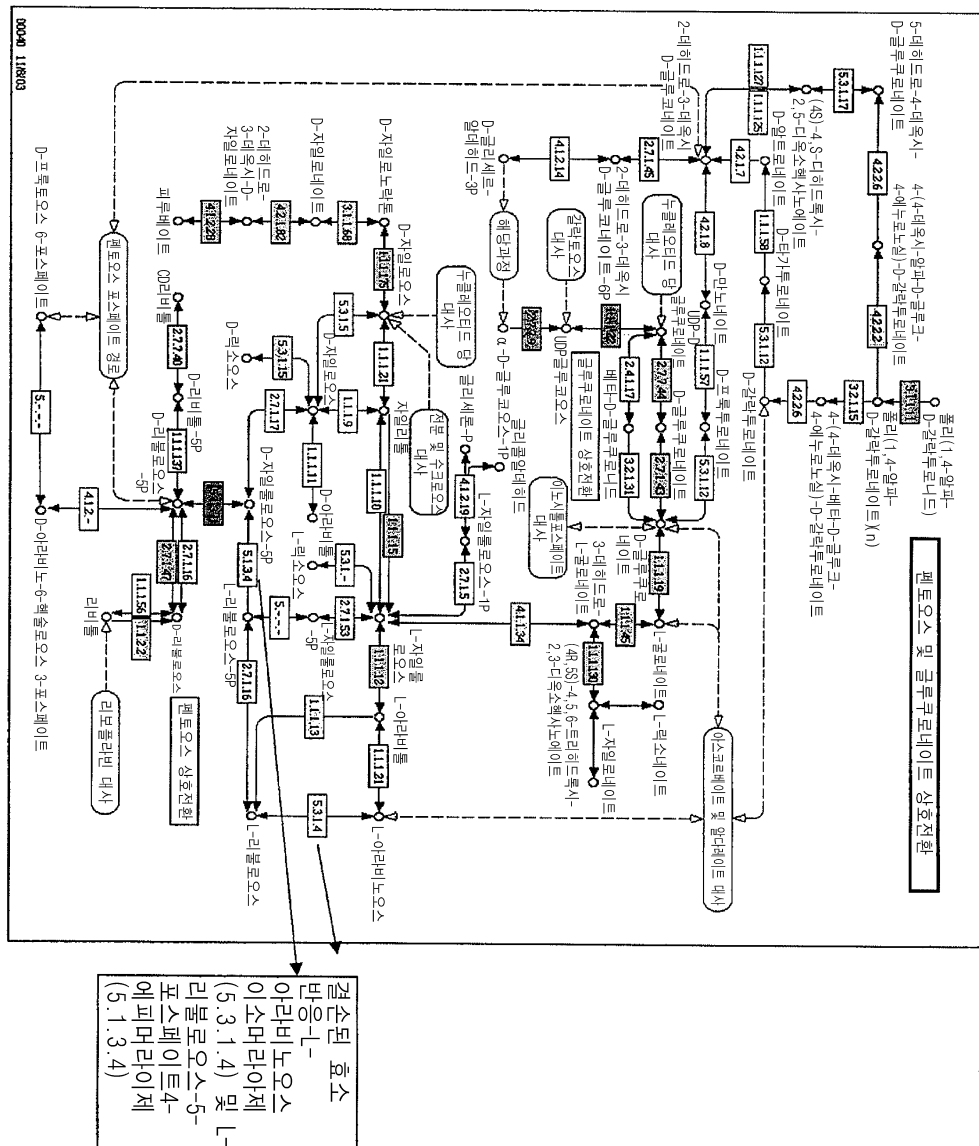




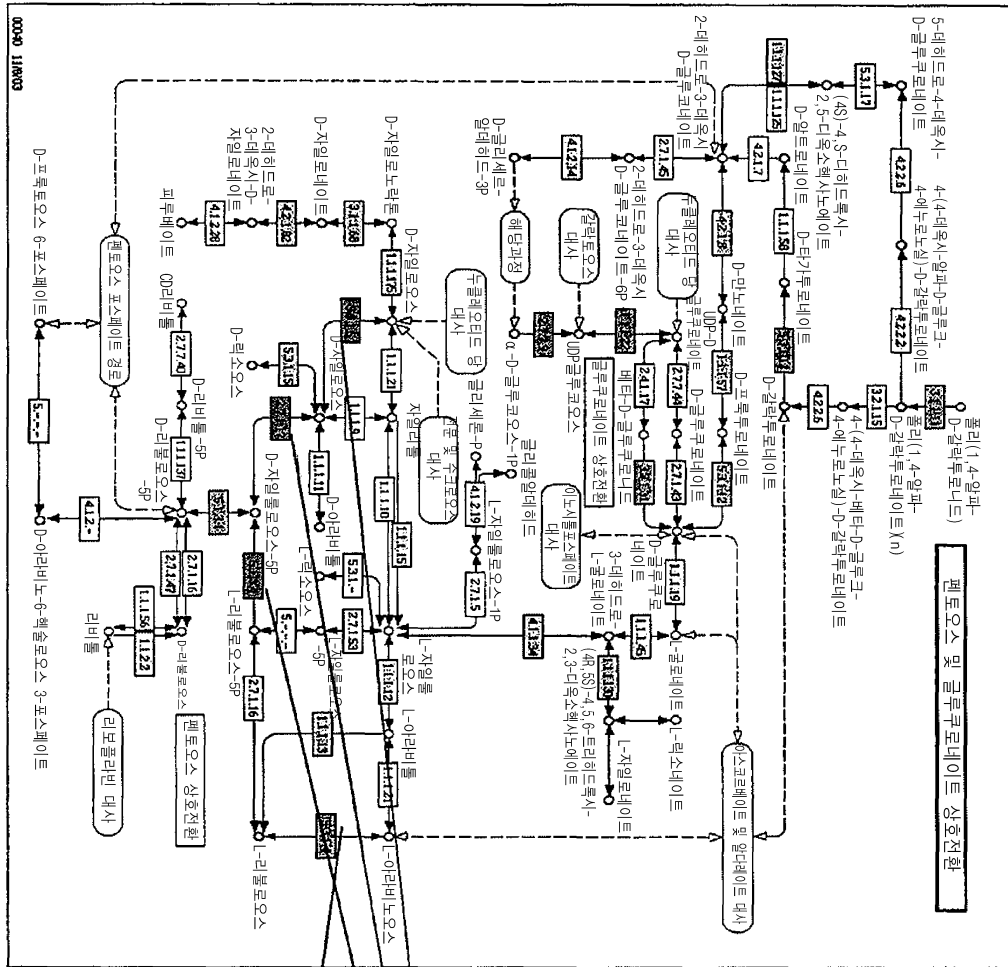
도면2



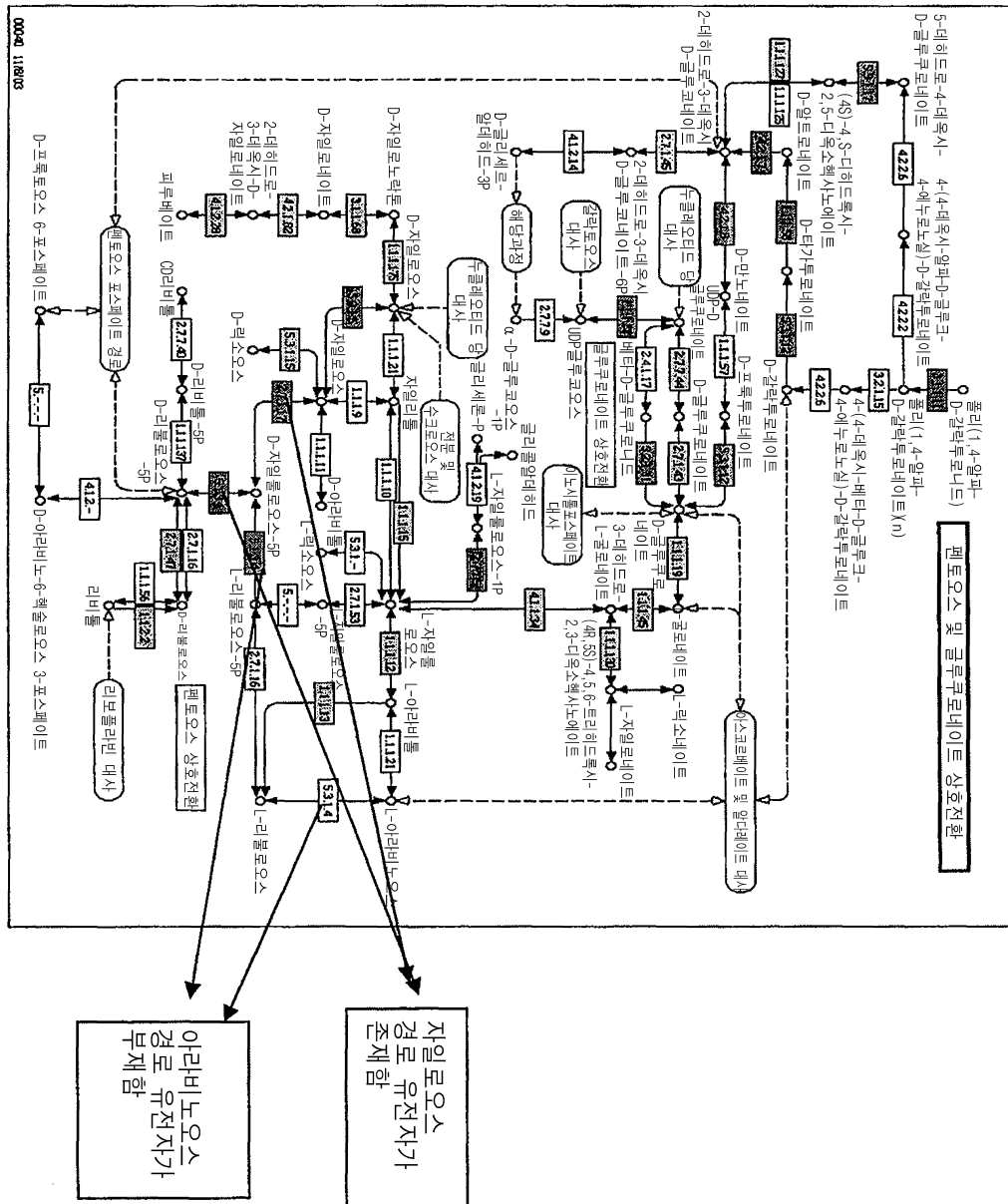
XI  
(5.3.1.5)  
및 XK  
(2.7.1.1  
7-  
결론



도면4



도면5



도면6A

16 Apr 2007

Alignment Results

Alignment: Global DNA alignment against reference molecule  
Parameters: Scoring matrix: Linear (Mismatch 2, OpenGap 4, ExtGap 1)

Reference molecule: Clostridium thermocellum, Region 1 to 1509  
Number of sequences to align: 7  
Total length of aligned sequences with gaps: 1767 bps

| Sequence                   | Start | End  | #Match | NonMatch | %Match |
|----------------------------|-------|------|--------|----------|--------|
| Clostridium thermocellum   | 1     | 1509 |        |          |        |
| Clostridium cellulolyticum | 1     | 1642 | 1372   | 274      | 83     |
| Thermoanaerobacterium sacc | 1     | 1552 | 1286   | 278      | 82     |
| C. stercorarium            | 1     | 1519 | 1374   | 155      | 89     |
| C. stercorarium II         | 1     | 1500 | 1339   | 194      | 87     |
| Caldicellulosiruptor krist | 1     | 1508 | 1250   | 300      | 80     |
| C. phytofermentans         | 1     | 1371 | 1115   | 403      | 73     |

도면6B

16 Apr 2007 Alignment Results

Alignment: Global DNA alignment against reference molecule  
Parameters: Scoring matrix: Linear (Mismatch 2, OpenGap 4, ExtGap 1)

Reference molecule: Clostridium thermocellum, Region 1 to 1509  
Number of sequences to align: 7  
Total length of aligned sequences with gaps: 1767 bps  
Settings: Similarity significance value cutoff: >= 60%

Summary of Percent Matches:

|      |                                       |           |             |     |  |
|------|---------------------------------------|-----------|-------------|-----|--|
| Ref: | Clostridium thermocellum              | 1 to 1509 | ( 1509 bps) | --  |  |
| 2:   | Clostridium cellulolyticum            | 1 to 1642 | ( 1642 bps) | 83% |  |
| 3:   | Thermoanaerobacterium saccharolyticum | 1 to 1552 | ( 1552 bps) | 82% |  |
| 4:   | C. stercorarium                       | 1 to 1519 | ( 1519 bps) | 89% |  |
| 5:   | C. stercorarium II                    | 1 to 1500 | ( 1500 bps) | 87% |  |
| 6:   | Caldicellulosiruptor kristjanssonii   | 1 to 1508 | ( 1508 bps) | 90% |  |
| 7:   | C. phytofermentans                    | 1 to 1371 | ( 1371 bps) | 73% |  |

```

Clostridium      1 tttgatccgggttcaggacggaacgctggcggcgtgacctaacacatgccagtgcagcgg---
Clostridium      1 --tgatccgngacaggnccagagacggnccgggtgcctaacacatgccagtgcagcgg---
Thermoanaero     1 tttgatccgggttcaggacggaacgctggcggcgtgacctaacacatgccagtgcagcggatc
C. stercorar     1 tttgatccgggttcaggacggaacgctggcggcgtgacctaacacatgccagtgcagcgg---
C. stercorar     1 -----cctggctacggaggaacgctggcggcgtgacctaacacatgccagtgcagcgg---
Caldicellulo     1 -----gggtcaggacaggnccagagacggnccgggtgcctaacacatgccagtgcagcgg---
C. phytoferm     1 -----

```

```

Clostridium      58 -gg---gata-----
Clostridium      56 -ga---gttaccttagcncctgagttattcttgganatgatgctgncccgacagcgtcat
Thermoanaero     61 cgg---nact-----
C. stercorar     58 -ggatccgtgt-----
C. stercorar     52 -ggatccgtgt-----
Caldicellulo     49 -ga---gatg-----
C. phytoferm     49 -----

```

```

Clostridium      64 -----taccg--gaa-----g-gttt---a-
Clostridium      111 ccnnnaacaaccttgaat-gaaatatattagttggagttttgcatcacgcg-tttt---a-
Thermoanaero     68 -----caat--taa-----gcgctt---a-
C. stercorar     68 -----taccg--gag-----g-tctttggg-
C. stercorar     62 -----taccg--gag-----g-tctttcggg-
Caldicellulo     55 -----gtagctgaa-----g-gtga--tg
C. phytoferm     55 -----

```

```

Clostridium      77 -scgga-----a-----g-ta-----
Clostridium      164 -tcaagggtgtcaacacataata-----g-tagagagaatgttcagtg
Thermoanaero     82 -cagaa-----aaagagagagaaantgag-ta-----
C. stercorar     84 -scgaa-----g-----t-gg-----
C. stercorar     78 -cagaa-----g-----t-gg-----
Caldicellulo     71 agctggg-----a-----gcta-----
C. phytoferm     71 -----

```

```

Clostridium      86 -----t-ctcc-----t---agcggggacgggtgagtaacccgtgg
Clostridium      205 ctgaaggt-aact-----t---agcggggacgggtgagtaacccgtgg
Thermoanaero     107 -----a-scgcgaagtgagtgccggat---agcggggacgggtgagtaacccgtgg
C. stercorar     93 -----c-atgg-----tgagactggcggacgggcgagtaacccgtga
C. stercorar     87 -----c-atgg-----tgagactggcggacgggcgagtaacccgtga
Caldicellulo     83 -----tcattc-----t---agcggggacgggtgagtaacccgtga
C. phytoferm     1 -----ct-----t---agtggggacgggtgagtaacccgtgg

```



도면6C

Page 2

|              |     |                                                                       |
|--------------|-----|-----------------------------------------------------------------------|
| Clostridium  | 119 | gtatgctacctcaaacagggggatataacaggggaacdtgtgctaatccggctaa-----          |
| Clostridium  | 245 | gcaaacctggctgtttagagggggaataacaggggaacdtgtgctaatccggctaa-----         |
| Thermoanaero | 156 | acaatctacctctgtagttttgggttaaacctcgaaggggtgtgctaatccggctaa-----        |
| C. stercorar | 129 | gcaaacctggcttaggttgggggaataaacaggggaacdtgtgctaatccggctaa-----         |
| C. stercorar | 123 | gcaaacctggcttaggttgggggaataaacaggggaacdtgtgctaatccggctaa-----         |
| Caldicellulo | 117 | gcaaacctacctcagacagggggaataaacaggtcgaaggggtgtgctaatccggctaa-----      |
| C. phytoferm | 31  | gtatgctacctcaaacagggggatataaacgtcgaaggggtgtgctaatccggctaa-----        |
| Clostridium  | 174 | a-----cggg-----gcg-----ggatcgctccattatcaaaaggaga-----a                |
| Clostridium  | 300 | a-----paca-----acgaagaagcaatttcnttgtgtgcaaggagc-----a                 |
| Thermoanaero | 211 | a-----tgtcaagaagt-----gaacacatttttgaagaaaggaga-----a                  |
| C. stercorar | 184 | agaccacagt-----gac-----gcat--gtacaggggt--aaagctg-----a                |
| C. stercorar | 178 | agaccacagt-----gac-----gcatgtcacaggggttaaaagctg-----a                 |
| Caldicellulo | 177 | a-----gggc-----atc-----ggatgggtgctgtgtgtgaagggtagccgnagaggcta         |
| C. phytoferm | 86  | a-----tata-----gcgaaccgcatgattttgctatcaaat-att-----t                  |
| Clostridium  | 207 | atcctggtatgagatggggccgggttcagatagcttagtttggtaggttaacgggtcacccag       |
| Clostridium  | 338 | atcctggtgacagatggggccgggttcagatagcttagtttggtaggttaacgggtcacccag       |
| Thermoanaero | 249 | atcctggtatgagatgggttcgggttcagatagcttagtttggtaggttaacgggtcacccag       |
| C. stercorar | 218 | agggcgctataggatgggttcgggttcagatagcttagtttggtaggttaacgggtcacccag       |
| C. stercorar | 214 | agggcgctataggatgggttcgggttcagatagcttagtttggtaggttaacgggtcacccag       |
| Caldicellulo | 222 | tnccggctgagatgggttcgggttcagatagcttagtttggtaggttaacgggtcacccag         |
| C. phytoferm | 123 | at-aggtatgagatggggccgggttcagatagcttagtttggtaggttaacgggtcacccag        |
| Clostridium  | 267 | gggacgagcgggagggcaaacgggggttggttcgggcgaattgggagtgagagacaggggc         |
| Clostridium  | 398 | gggacgagcgggagggcaaacgggggttggttcgggcgaattgggagtgagagacaggggc         |
| Thermoanaero | 309 | gggacgagcgggagggcaaacgggggttggttcgggcgaattgggagtgagagacaggggc         |
| C. stercorar | 278 | gggacgagcgggagggcaaacgggggttggttcgggcgaattgggagtgagagacaggggc         |
| C. stercorar | 274 | gggacgagcgggagggcaaacgggggttggttcgggcgaattgggagtgagagacaggggc         |
| Caldicellulo | 282 | gggacgagcgggagggcaaacgggggttggttcgggcgaattgggagtgagagacaggggc         |
| C. phytoferm | 182 | gggacgagcgggagggcaaacgggggttggttcgggcgaattgggagtgagagacaggggc         |
| Clostridium  | 327 | agacttctacgggagggagggatgggggaggtttagggagtgaggagaaacctgaagcag          |
| Clostridium  | 458 | agacttctacgggagggagggatgggggaggtttagggagtgaggagaaacctgaagcag          |
| Thermoanaero | 369 | agacttctacgggagggagggatgggggaggtttagggagtgaggagaaacctgaagcag          |
| C. stercorar | 338 | agacttctacgggagggagggatgggggaggtttagggagtgaggagaaacctgaagcag          |
| C. stercorar | 334 | agacttctacgggagggagggatgggggaggtttagggagtgaggagaaacctgaagcag          |
| Caldicellulo | 342 | acacttctacgggagggagggatgggggaggtttagggagtgaggagaaacctgaagcag          |
| C. phytoferm | 242 | nnacgncnsgggagggagggatgggggaggtttagggagtgaggagaaacctgaagcag           |
| Clostridium  | 387 | aacggcgctttagggagggagggatgggggaggtttagggagtgaggagaaacctgaagcag        |
| Clostridium  | 518 | aacggcgctttagggagggagggatgggggaggtttagggagtgaggagaaacctgaagcag        |
| Thermoanaero | 429 | gaagcgctttagggagggagggatgggggaggtttagggagtgaggagaaacctgaagcag         |
| C. stercorar | 398 | gaagcgctttagggagggagggatgggggaggtttagggagtgaggagaaacctgaagcag         |
| C. stercorar | 394 | gaagcgctttagggagggagggatgggggaggtttagggagtgaggagaaacctgaagcag         |
| Caldicellulo | 402 | gaagcgctttagggagggagggatgggggaggtttagggagtgaggagaaacctgaagcag         |
| C. phytoferm | 302 | gacggcgctttagggagggagggatgggggaggtttagggagtgaggagaaacctgaagcag        |
| Clostridium  | 447 | agggagggtagccaaagagcaagcacaaggttaacttaggttgcagagcggcggttaatacg        |
| Clostridium  | 578 | aagggagggtagccaaagagcaagcacaaggttaacttaggttgcagagcggcggttaatacg       |
| Thermoanaero | 486 | agggagggtagccaaagagcaagcacaaggttaacttaggttgcagagcggcggttaatacg        |
| C. stercorar | 456 | -atggagggtagccaaagagcaagcacaaggttaacttaggttgcagagcggcggttaatacg       |
| C. stercorar | 452 | -atggagggtagccaaagagcaagcacaaggttaacttaggttgcagagcggcggttaatacg       |
| Caldicellulo | 461 | gaaggtaggttagccgtttaaaagagcaagcacaaggttaacttaggttgcagagcggcggttaatacg |
| C. phytoferm | 359 | aagggagggtagccaaagagcaagcacaaggttaacttaggttgcagagcggcggttaatacg       |

도면6D

Page 3

|              |     |                                                                        |
|--------------|-----|------------------------------------------------------------------------|
| Clostridium  | 507 | taaggctggcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a       |
| Clostridium  | 638 | taaggctggcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a       |
| Thermoanaero | 544 | taaggctggcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a       |
| C. stercorar | 515 | taaggctggcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a       |
| C. stercorar | 511 | taaggctggcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a       |
| Caldicellulo | 521 | taaggctggcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a       |
| C. phytoferm | 419 | taaggctggcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a       |
| Clostridium  | 564 | agtcagatgctggaatgctcagggccttaaccccggggcctgacacggaatgctgacacggaatgctgga |
| Clostridium  | 696 | agtcagatgctggaatgctcagggccttaaccccggggcctgacacggaatgctgacacggaatgctgga |
| Thermoanaero | 602 | agtcagatgctggaatgctcagggccttaaccccggggcctgacacggaatgctgacacggaatgctgga |
| C. stercorar | 573 | agtcagatgctggaatgctcagggccttaaccccggggcctgacacggaatgctgacacggaatgctgga |
| C. stercorar | 569 | agtcagatgctggaatgctcagggccttaaccccggggcctgacacggaatgctgacacggaatgctgga |
| Caldicellulo | 579 | agtcagatgctggaatgctcagggccttaaccccggggcctgacacggaatgctgacacggaatgctgga |
| C. phytoferm | 477 | agtcagatgctggaatgctcagggccttaaccccggggcctgacacggaatgctgacacggaatgctgga |
| Clostridium  | 624 | gtgctggcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a         |
| Clostridium  | 756 | gtgctggcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a         |
| Thermoanaero | 662 | gtgctggcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a         |
| C. stercorar | 633 | gtgctggcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a         |
| C. stercorar | 629 | gtgctggcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a         |
| Caldicellulo | 639 | gtgctggcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a         |
| C. phytoferm | 537 | gtgctggcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a         |
| Clostridium  | 684 | atccagatgctggaatgctcagggccttaaccccggggcctgacacggaatgctgacacggaatgctgga |
| Clostridium  | 816 | atccagatgctggaatgctcagggccttaaccccggggcctgacacggaatgctgacacggaatgctgga |
| Thermoanaero | 722 | atccagatgctggaatgctcagggccttaaccccggggcctgacacggaatgctgacacggaatgctgga |
| C. stercorar | 693 | atccagatgctggaatgctcagggccttaaccccggggcctgacacggaatgctgacacggaatgctgga |
| C. stercorar | 689 | atccagatgctggaatgctcagggccttaaccccggggcctgacacggaatgctgacacggaatgctgga |
| Caldicellulo | 699 | atccagatgctggaatgctcagggccttaaccccggggcctgacacggaatgctgacacggaatgctgga |
| C. phytoferm | 597 | atccagatgctggaatgctcagggccttaaccccggggcctgacacggaatgctgacacggaatgctgga |
| Clostridium  | 744 | gagcgaacaggaatgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a          |
| Clostridium  | 876 | gagcgaacaggaatgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a          |
| Thermoanaero | 782 | gagcgaacaggaatgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a          |
| C. stercorar | 753 | gagcgaacaggaatgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a          |
| C. stercorar | 749 | gagcgaacaggaatgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a          |
| Caldicellulo | 759 | gagcgaacaggaatgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a          |
| C. phytoferm | 657 | gagcgaacaggaatgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a          |
| Clostridium  | 804 | gagcgaacaggaatgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a          |
| Clostridium  | 936 | gagcgaacaggaatgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a          |
| Thermoanaero | 842 | gagcgaacaggaatgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a          |
| C. stercorar | 813 | gagcgaacaggaatgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a          |
| C. stercorar | 809 | gagcgaacaggaatgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a          |
| Caldicellulo | 818 | gagcgaacaggaatgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a          |
| C. phytoferm | 717 | gagcgaacaggaatgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a          |
| Clostridium  | 863 | ggcgcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a            |
| Clostridium  | 995 | ggcgcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a            |
| Thermoanaero | 897 | ggcgcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a            |
| C. stercorar | 872 | ggcgcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a            |
| C. stercorar | 868 | ggcgcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a            |
| Caldicellulo | 877 | ggcgcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a            |
| C. phytoferm | 777 | ggcgcgaagcctgacacggaatgactgggtgttaagggccgcctaggcggga-tg-c-a            |

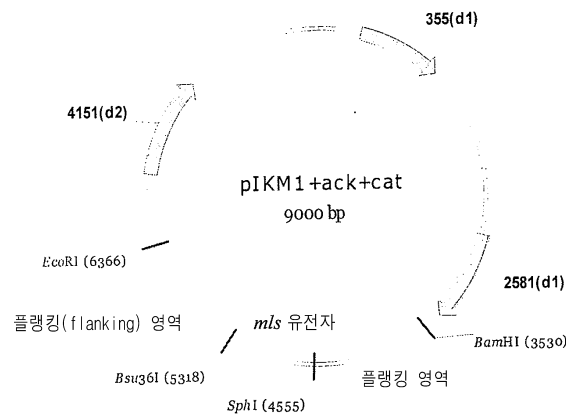
[illegible]

도면6F

Page 5

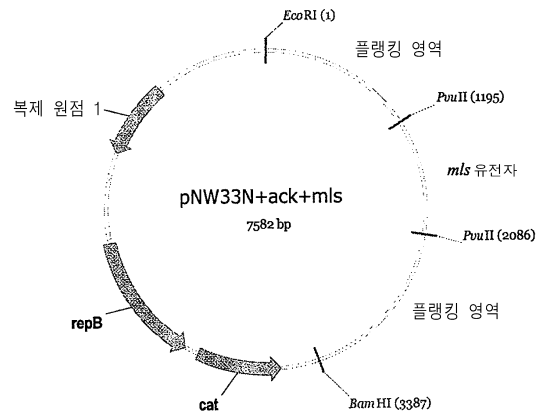
|              |      |                                                                |
|--------------|------|----------------------------------------------------------------|
| Clostridium  | 1314 | agcaatggcaggtcagcaactgccctgaatccctccgggcccctatcacacccggccg     |
| Clostridium  | 1446 | agtaacggttaggtcagtaactgtctcgaataagctcccgggccctgacacacccggccg   |
| Thermoanaero | 1352 | agtaatcggcaggtcagcaactgccctggaataacgtcccgggccctgacacacccggccg  |
| C. stercorar | 1324 | agtaatggcaggtcagcaactgccctgaataagctcccgggccctgacacacccggccg    |
| C. stercorar | 1320 | agtaatggcaggtcagcaactgccctgaataagctcccgggccctgacacacccggccg    |
| Caldicellulo | 1341 | agtaatcggcaggtcagcaactgccctgaataagctcccgggccctgacacacccggccg   |
| C. phytoferm | 1228 | agtaatcgggaacagcaactgtctcgggtgaataacgtcccgggccctgacacacccggccg |
| Clostridium  | 1374 | cccgacatggaatctgcaatgacgaatc-g-taatt-g-accgcaggagggacgg        |
| Clostridium  | 1506 | cccgacatggaatctgcaatgacgaatc-g-taatt-g-accgcaggagggacgt        |
| Thermoanaero | 1412 | cccgacacggaatctttagaatacccggaatc-g-gtgaac-accgcaggagggagac     |
| C. stercorar | 1384 | cccgacacggaatctttagaatacccggaatc-g-gtgaac-accgcaggagggagac     |
| C. stercorar | 1380 | cccgacacggaatctttagaatacccggaatc-g-gtgaac-accgcaggagggagac     |
| Caldicellulo | 1401 | cccgacacggaatcagcaatgacgaatc-g-cagga-g-atctg-----              |
| C. phytoferm | 1288 | cccgacacggaatcagcaatgacgaatc-g-gtgaac-cyaccgtaaggaggagac       |
| Clostridium  | 1430 | tcctgaagctggcaggtcagtaactgggggcaatgcaaac---aggtacccgtatcgga    |
| Clostridium  | 1563 | cccgacacggaatctttagaatacccggaatc-g-gtgaac-accgcaggagggagac     |
| Thermoanaero | 1469 | tcctgaagctggcaggtcagtaactgggggcaatgcaaac---aggtacccgtatcgga    |
| C. stercorar | 1440 | tcctgaagctggcaggtcagtaactgggggcaatgcaaac---aggtacccgtatcgga    |
| C. stercorar | 1437 | tcctgaagctggcaggtcagtaactgggggcaatgcaaac---aggtacccgtatcgga    |
| Caldicellulo | 1445 | tcctgaagctggcaggtcagtaactgggggcaatgcaaac---aggtacccgtatcgga    |
| C. phytoferm | 1345 | tcctgaagctggcaggtcagtaactgggggcaatgcaaac---aggtacccgtatcgga    |
| Clostridium  | 1487 | ggctcggctggatcacctccttt---                                     |
| Clostridium  | 1620 | ggctcggctggatcacctccttt---                                     |
| Thermoanaero | 1526 | ggctcggctggatcacctcctttccct                                    |
| C. stercorar | 1497 | ggctcggctggatcacctccttt---                                     |
| C. stercorar | 1494 | ggctcggctggatcacctccttt---                                     |
| Caldicellulo | 1502 | ggctcggctggatcacctccttt---                                     |
| C. phytoferm |      | ggctcggctggatcacctccttt---                                     |

도면7

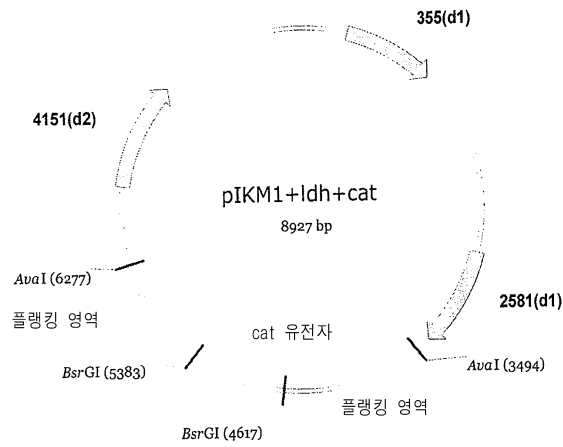




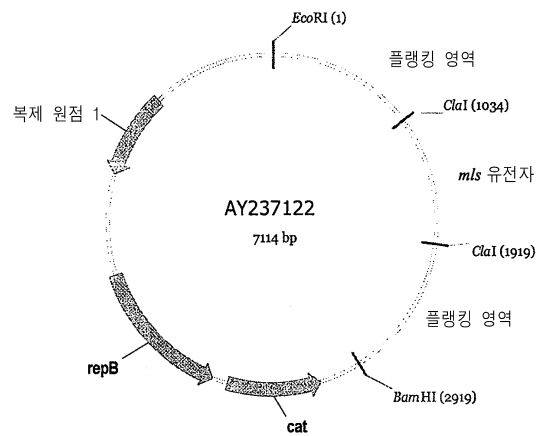
도면8



도면9

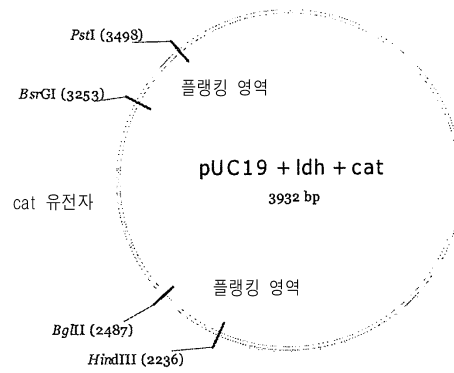


도면10

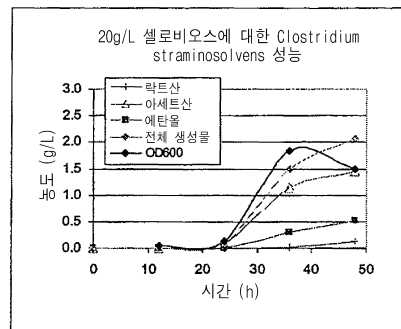




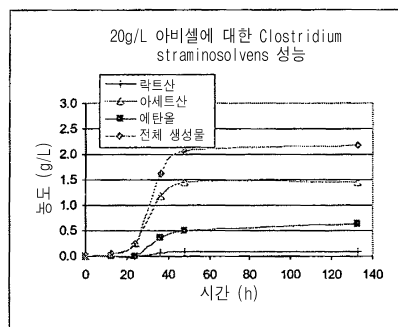
도면11



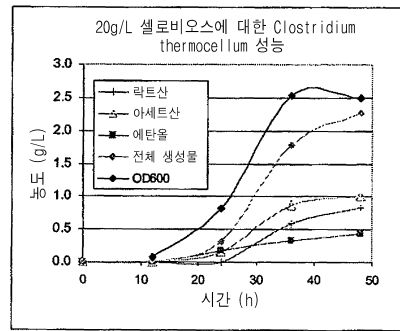
도면12A



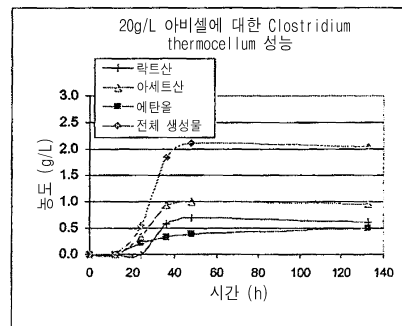
도면12B



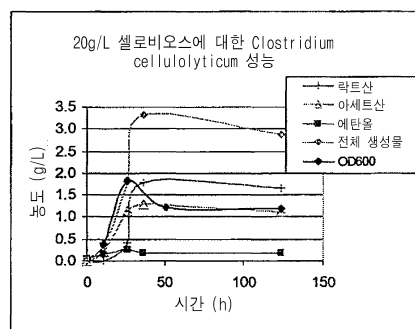
도면13A



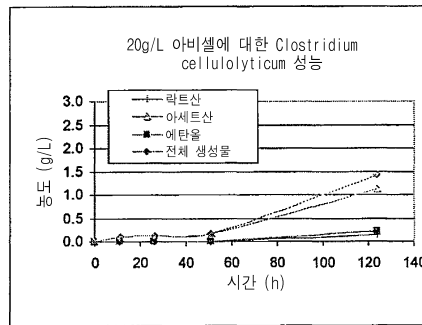
도면13B



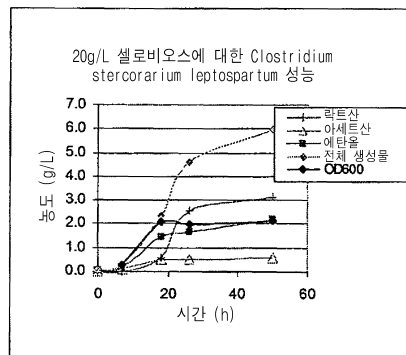
도면14A



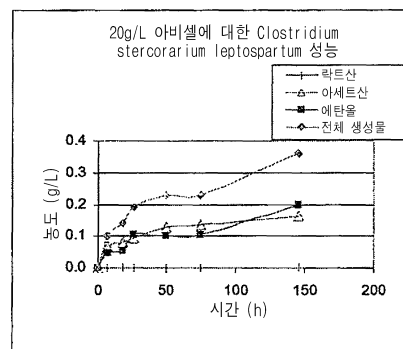
도면14B



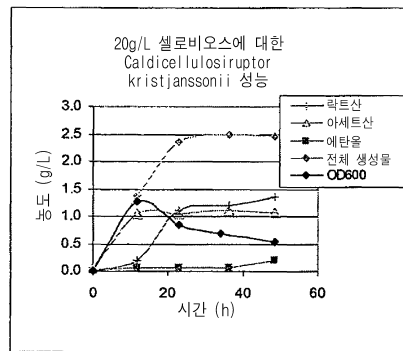
도면15A



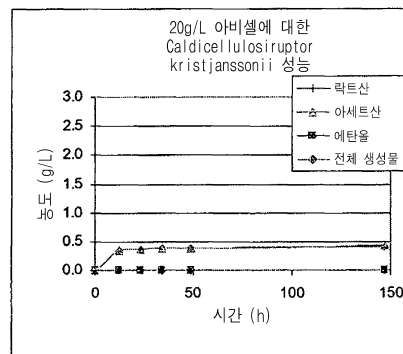
도면15B



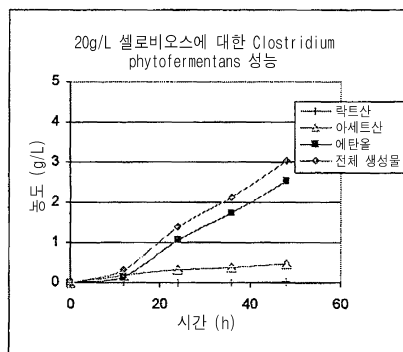
도면16A



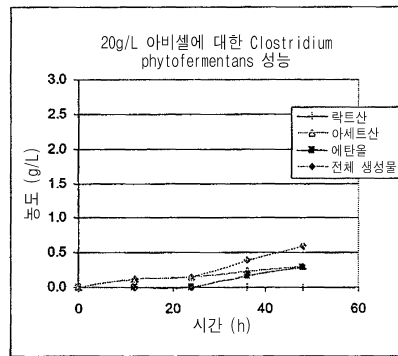
도면16B



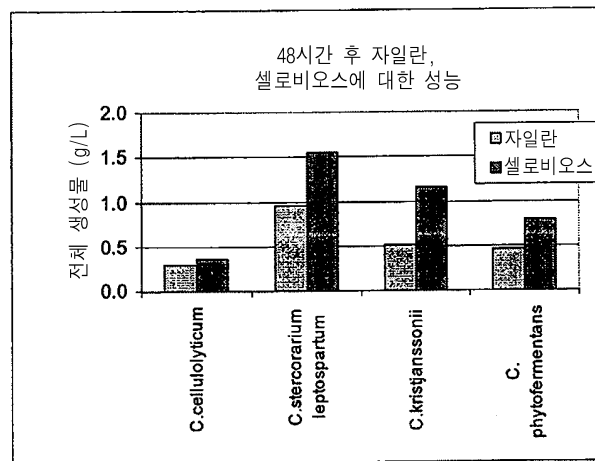
도면17A



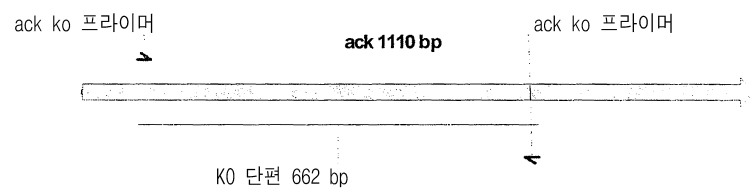
도면17B



도면18

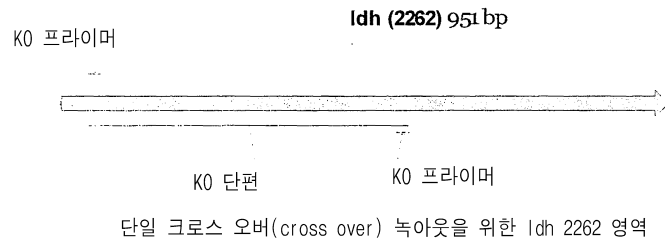


도면19

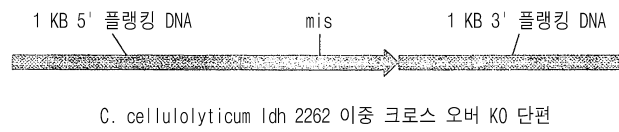




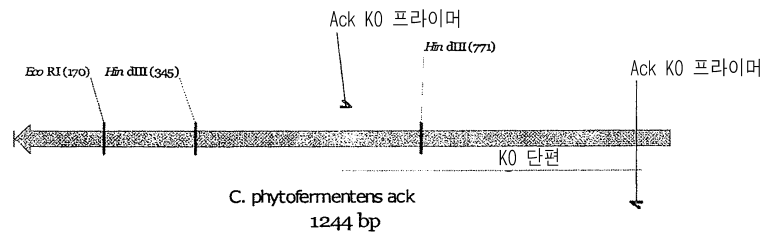
도면20



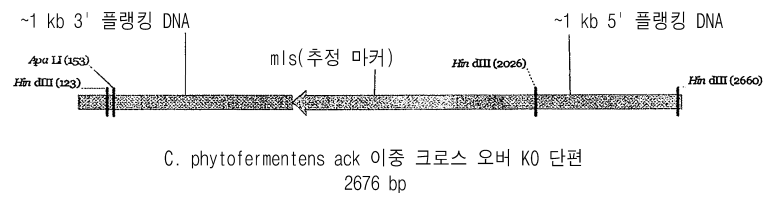
도면21



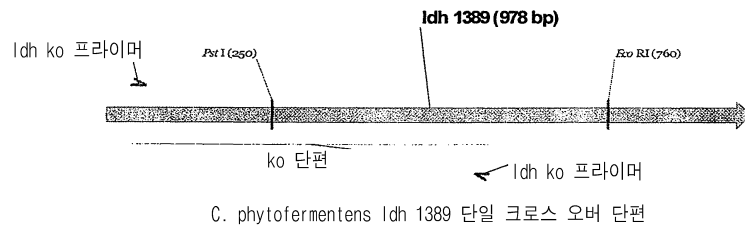
도면22



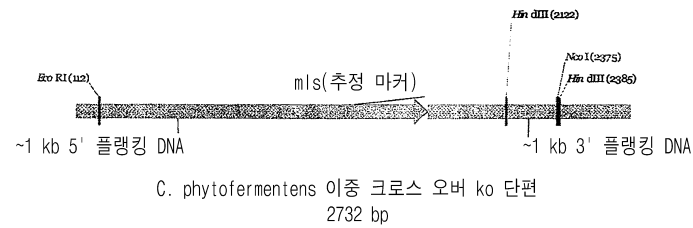
도면23



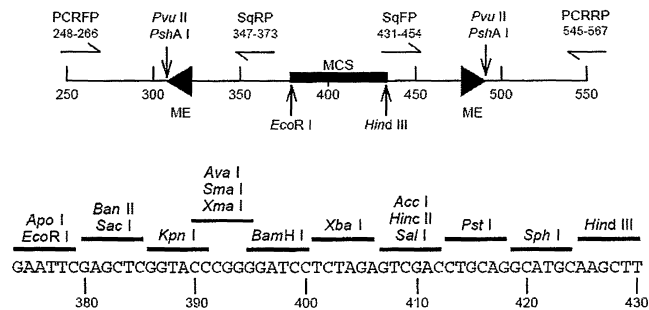
도면24



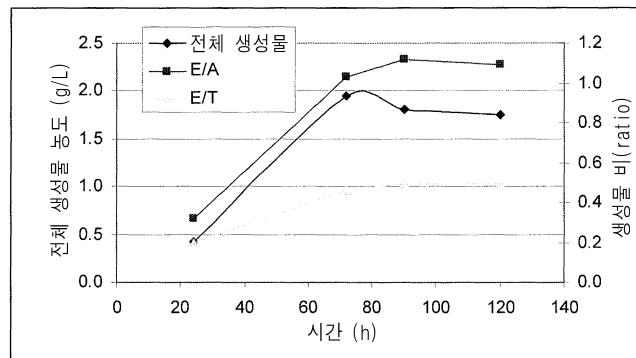
도면25



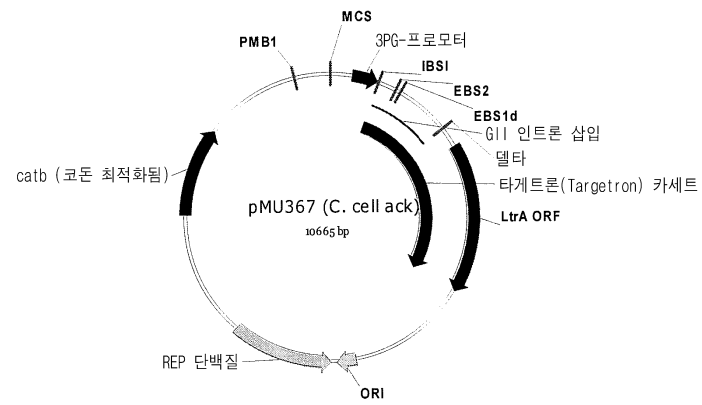
도면26



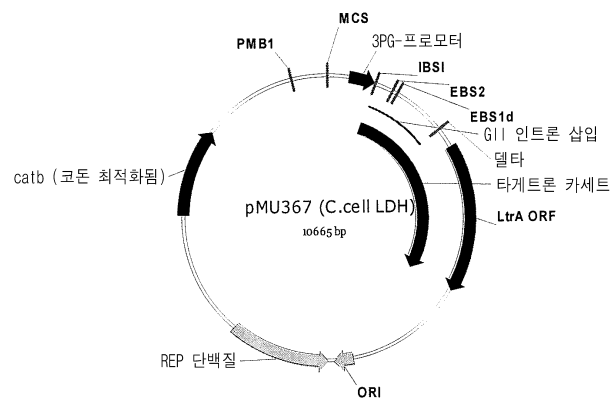
도면27



도면28



도면29



도면30A

Percent identity of 16S sequence to T. saccharolyticum JW/SL-YS485:  
 Thermoanaerobacterium saccharolyticum B6A-RI - 99.0%  
 Thermoanaerobacter sp. strain 59 - 95.7%  
 Thermoanaerobacter pseudoethanolicus 39E - 83.7%

|                                     |       |                                                   |     |
|-------------------------------------|-------|---------------------------------------------------|-----|
|                                     |       | Tethanolicus39E_16S (0.1131)                      |     |
|                                     |       | Thermoanaerobacter_sp_strain_59_16S (0.0197)      |     |
|                                     |       | TsaccharolyticumB6A-RI_16S (0.0043)               |     |
|                                     |       | TsaccharolyticumYS485_16S (0.0054)                |     |
|                                     |       | 1                                                 | 50  |
| Tethanolicus39E_16S                 | (1)   | -----GCTGGCTCAGGACGACGCTGGCGCGCTGCCTAACACATGCAAG  |     |
| Thermoanaerobacter_sp_strain_59_16S | (1)   | -----GCTGGCTCAGGACGACGCTGGCGCGCTGCCTAACACATGCAAG  |     |
| TsaccharolyticumB6A-RI_16S          | (1)   | TTTGATCCTGGCTCAGGACGACGCTGGCGCGCTGCCTAACACATGCAAG |     |
| TsaccharolyticumYS485_16S           | (1)   | TTTGATCCTGGCTCAGGACGACGCTGGCGCGCTGCCTAACACATGCAAG |     |
| Consensus                           | (1)   | TTTGATCCTGGCTCAGGACGACGCTGGCGCGCTGCCTAACACATGCAAG |     |
|                                     |       | 51                                                | 100 |
| Tethanolicus39E_16S                 | (45)  | TCGAGCGATCCGGCACTCAA TAAGCGCTTACAGAAAAGA GAGCGAAA |     |
| Thermoanaerobacter_sp_strain_59_16S | (1)   | -----CTGACCTGCAGTCCTCCG                           |     |
| TsaccharolyticumB6A-RI_16S          | (51)  | TCGAGCGATCCGGCACTCAA TAAGCGCTTACAGAAAAGA GAGCGAAA |     |
| TsaccharolyticumYS485_16S           | (51)  | TCGAGCGATCCGGCACTCAA TAAGCGCTTACAGAAAAGA GAGCGAAA |     |
| Consensus                           | (51)  | TCGAGCGATCCGGCACTCAA TAAGCGCTTACAGAAAAGA GAGCGAAA |     |
|                                     |       | 101                                               | 150 |
| Tethanolicus39E_16S                 | (65)  | -----CTGACCTGCAGTCCTCCG                           |     |
| Thermoanaerobacter_sp_strain_59_16S | (22)  | GCAGTA---GTCCTACAAAT---AGCGCGCGCGCGCTGCCTAAC      |     |
| TsaccharolyticumB6A-RI_16S          | (100) | NTCAGTAACGCACTGATGCGCGATACCGCGCGCGCGCTGCCTAAC     |     |
| TsaccharolyticumYS485_16S           | (100) | ATGAGTAACGCACTGATGCGCGATACCGCGCGCGCGCTGCCTAAC     |     |
| Consensus                           | (101) | TGAGTAACGCTAAGTTGAGTGCCGGATAGCGCGCGCGCGCTGCCTAAC  |     |
|                                     |       | 151                                               | 200 |
| Tethanolicus39E_16S                 | (105) | CGCTGCAATCTACCTGTAGTTGGGATAACACTCGAAGGGGTGCT      |     |
| Thermoanaerobacter_sp_strain_59_16S | (65)  | CGCTGCAATCTACCTGTAGTTGGGATAACACTCGAAGGGGTGCT      |     |
| TsaccharolyticumB6A-RI_16S          | (150) | CGCTGCAATCTACCTGTAGTTGGGATAACACTCGAAGGGGTGCT      |     |
| TsaccharolyticumYS485_16S           | (150) | CGCTGCAATCTACCTGTAGTTGGGATAACACTCGAAGGGGTGCT      |     |
| Consensus                           | (151) | CGCTGCAATCTACCTGTAGTTGGGATAACACTCGAAGGGGTGCT      |     |
|                                     |       | 201                                               | 250 |
| Tethanolicus39E_16S                 | (155) | AATACCGGATAATGTCAGAAGTGGCATCGCTTTTGAAGAAGGAGAG    |     |
| Thermoanaerobacter_sp_strain_59_16S | (115) | AATACCGGATAATGTCAGAAGTGGCATCGCTTTTGAAGAAGGAGAG    |     |
| TsaccharolyticumB6A-RI_16S          | (200) | AATACCGGATAATGTCAGAAGTGGCATCGCTTTTGAAGAAGGAGAG    |     |
| TsaccharolyticumYS485_16S           | (200) | AATACCGGATAATGTCAGAAGTGGCATCGCTTTTGAAGAAGGAGAG    |     |
| Consensus                           | (201) | AATACCGGATAATGTCAGAAGTGGCATCGCTTTTGAAGAAGGAGAG    |     |
|                                     |       | 251                                               | 300 |
| Tethanolicus39E_16S                 | (205) | GCTACCTGATAGGATCGCTGGGCTCCCTCAGCTAGTTGTA          |     |
| Thermoanaerobacter_sp_strain_59_16S | (164) | ---AAT---CCGCTATAGGATCGCTGGGCTCCCTCAGCTAGTTG      |     |
| TsaccharolyticumB6A-RI_16S          | (248) | ---AAT---CCGCTATAGGATCGCTGGGCTCCCTCAGCTAGTTG      |     |
| TsaccharolyticumYS485_16S           | (248) | ---AAT---CCGCTATAGGATCGCTGGGCTCCCTCAGCTAGTTG      |     |
| Consensus                           | (251) | AAT CCGCTATAGGATCGCTGGGCTCCCTCAGCTAGTTG           |     |
|                                     |       | 301                                               | 350 |
| Tethanolicus39E_16S                 | (254) | AAACGCGCTACCAAGCGACGATGGGTAGCCGGCTGAGAGGGTGAACGGC |     |
| Thermoanaerobacter_sp_strain_59_16S | (212) | AAACGCGCTACCAAGCGACGATGGGTAGCCGGCTGAGAGGGTGAACGGC |     |
| TsaccharolyticumB6A-RI_16S          | (295) | AAACGCGCTACCAAGCGACGATGGGTAGCCGGCTGAGAGGGTGAACGGC |     |
| TsaccharolyticumYS485_16S           | (295) | AAACGCGCTACCAAGCGACGATGGGTAGCCGGCTGAGAGGGTGAACGGC |     |
| Consensus                           | (301) | AAACGCGCTACCAAGCGACGATGGGTAGCCGGCTGAGAGGGTGAACGGC |     |
|                                     |       | 351                                               | 400 |
| Tethanolicus39E_16S                 | (304) | CACTGGAAGTGAAGACGGTCCAGACTCCTACGGGAGGACAGTGGG     |     |
| Thermoanaerobacter_sp_strain_59_16S | (262) | CACTGGAAGTGAAGACGGTCCAGACTCCTACGGGAGGACAGTGGG     |     |
| TsaccharolyticumB6A-RI_16S          | (345) | CACTGGAAGTGAAGACGGTCCAGACTCCTACGGGAGGACAGTGGG     |     |
| TsaccharolyticumYS485_16S           | (345) | CACTGGAAGTGAAGACGGTCCAGACTCCTACGGGAGGACAGTGGG     |     |
| Consensus                           | (351) | CACTGGAAGTGAAGACGGTCCAGACTCCTACGGGAGGACAGTGGG     |     |
|                                     |       | 401                                               | 450 |
| Tethanolicus39E_16S                 | (354) | GAATATTGTGCAATGGGGAAACCTTGACACAGCAGCCGCGTGAAGCGA  |     |
| Thermoanaerobacter_sp_strain_59_16S | (312) | GAATATTGTGCAATGGGGAAACCTTGACACAGCAGCCGCGTGAAGCGA  |     |
| TsaccharolyticumB6A-RI_16S          | (395) | GAATATTGTGCAATGGGGAAACCTTGACACAGCAGCCGCGTGAAGCGA  |     |
| TsaccharolyticumYS485_16S           | (395) | GAATATTGTGCAATGGGGAAACCTTGACACAGCAGCCGCGTGAAGCGA  |     |
| Consensus                           | (401) | GAATATTGTGCAATGGGGAAACCTTGACACAGCAGCCGCGTGAAGCGA  |     |
|                                     |       | 451                                               | 500 |
| Tethanolicus39E_16S                 | (404) | GAAGGCGCTCGGTCGTAAGCTCATAGTATGGGAAGATAG           |     |
| Thermoanaerobacter_sp_strain_59_16S | (362) | GAAGGCGCTCGGTCGTAAGCTCATAGTATGGGAAGATAG           |     |
| TsaccharolyticumB6A-RI_16S          | (445) | GAAGGCGCTCGGTCGTAAGCTCATAGTATGGGAAGATAG           |     |
| TsaccharolyticumYS485_16S           | (445) | GAAGGCGCTCGGTCGTAAGCTCATAGTATGGGAAGATAG           |     |
| Consensus                           | (451) | GAAGGCGCTCGGTCGTAAGCTCATAGTATGGGAAGATAG           |     |
|                                     |       | 501                                               | 550 |

도면30B

|                                     |        |                                                    |      |      |
|-------------------------------------|--------|----------------------------------------------------|------|------|
| Tethanolicus39E_16S                 | (453)  | GGTACCATACGAAAGCCCGGCTAACTACGTGCCAGCAGCCGCGTAATA   | 551  | 600  |
| Thermoanaerobacter_sp_strain_59_16S | (412)  | GGTACCATACGAAAGCCCGGCTAACTACGTGCCAGCAGCCGCGTAATA   |      |      |
| TsaccharolyticumB6A-RI_16S          | (492)  | GGTACCATACGAAAGCCCGGCTAACTACGTGCCAGCAGCCGCGTAATA   |      |      |
| TsaccharolyticumYS485_16S           | (492)  | GGTACCATACGAAAGCCCGGCTAACTACGTGCCAGCAGCCGCGTAATA   |      |      |
| Consensus                           | (501)  | GGTACCATACGAAAGCCCGGCTAACTACGTGCCAGCAGCCGCGTAATA   |      |      |
| Tethanolicus39E_16S                 | (503)  | CGTAGGGGCGAGCGTGTGCCGAATTACTGGGCGTAAAGAGCAGCTAGG   | 601  | 650  |
| Thermoanaerobacter_sp_strain_59_16S | (462)  | CGTAGGGGCGAGCGTGTGCCGAATTACTGGGCGTAAAGAGCAGCTAGG   |      |      |
| TsaccharolyticumB6A-RI_16S          | (542)  | CGTAGGGGCGAGCGTGTGCCGAATTACTGGGCGTAAAGAGCAGCTAGG   |      |      |
| TsaccharolyticumYS485_16S           | (542)  | CGTAGGGGCGAGCGTGTGCCGAATTACTGGGCGTAAAGAGCAGCTAGG   |      |      |
| Consensus                           | (551)  | CGTAGGGGCGAGCGTGTGCCGAATTACTGGGCGTAAAGAGCAGCTAGG   |      |      |
| Tethanolicus39E_16S                 | (553)  | CGCTTAATAAGCTCAAGCTTAAGCTTGGGAGGAGGATAT            | 651  | 700  |
| Thermoanaerobacter_sp_strain_59_16S | (512)  | CGCTTAATAAGCTCAAGCTTAAGCTTGGGAGGAGGATAT            |      |      |
| TsaccharolyticumB6A-RI_16S          | (592)  | CGCTTAATAAGCTCAAGCTTAAGCTTGGGAGGAGGATAT            |      |      |
| TsaccharolyticumYS485_16S           | (592)  | CGCTTAATAAGCTCAAGCTTAAGCTTGGGAGGAGGATAT            |      |      |
| Consensus                           | (601)  | CGCTTAATAAGCTCAAGCTTAAGCTTGGGAGGAGGATAT            |      |      |
| Tethanolicus39E_16S                 | (603)  | CTTGAAATGGGCGCAATGGGCGGAGCGGCTTCGAGGCGCGG          | 701  | 750  |
| Thermoanaerobacter_sp_strain_59_16S | (562)  | CTTGAAATGGGCGCAATGGGCGGAGCGGCTTCGAGGCGCGG          |      |      |
| TsaccharolyticumB6A-RI_16S          | (642)  | CTTGAAATGGGCGCAATGGGCGGAGCGGCTTCGAGGCGCGG          |      |      |
| TsaccharolyticumYS485_16S           | (642)  | CTTGAAATGGGCGCAATGGGCGGAGCGGCTTCGAGGCGCGG          |      |      |
| Consensus                           | (651)  | CTTGAAATGGGCGCAATGGGCGGAGCGGCTTCGAGGCGCGG          |      |      |
| Tethanolicus39E_16S                 | (653)  | TACGGGTGAATGCGTAGCTATGGTGGATACCGCTGGCGATGGGA       | 751  | 800  |
| Thermoanaerobacter_sp_strain_59_16S | (612)  | TACGGGTGAATGCGTAGCTATGGTGGATACCGCTGGCGATGGGA       |      |      |
| TsaccharolyticumB6A-RI_16S          | (692)  | TACGGGTGAATGCGTAGCTATGGTGGATACCGCTGGCGATGGGA       |      |      |
| TsaccharolyticumYS485_16S           | (692)  | TACGGGTGAATGCGTAGCTATGGTGGATACCGCTGGCGATGGGA       |      |      |
| Consensus                           | (701)  | TACGGGTGAATGCGTAGCTATGGTGGATACCGCTGGCGATGGGA       |      |      |
| Tethanolicus39E_16S                 | (703)  | ACCTGGGACGTGACCTGCAAGCTTCAGCTGAGGAGGAGGAGGAGG      | 801  | 850  |
| Thermoanaerobacter_sp_strain_59_16S | (662)  | ACCTGGGACGTGACCTGCAAGCTTCAGCTGAGGAGGAGGAGGAGG      |      |      |
| TsaccharolyticumB6A-RI_16S          | (742)  | ACCTGGGACGTGACCTGCAAGCTTCAGCTGAGGAGGAGGAGGAGG      |      |      |
| TsaccharolyticumYS485_16S           | (742)  | ACCTGGGACGTGACCTGCAAGCTTCAGCTGAGGAGGAGGAGGAGG      |      |      |
| Consensus                           | (751)  | ACCTGGGACGTGACCTGCAAGCTTCAGCTGAGGAGGAGGAGGAGG      |      |      |
| Tethanolicus39E_16S                 | (753)  | GATTAGATACCGTGGTAGTCCAGCGGTAAACGATGGATACCTAGTGTGG  | 851  | 900  |
| Thermoanaerobacter_sp_strain_59_16S | (712)  | GATTAGATACCGTGGTAGTCCAGCGGTAAACGATGGATACCTAGTGTGG  |      |      |
| TsaccharolyticumB6A-RI_16S          | (792)  | GATTAGATACCGTGGTAGTCCAGCGGTAAACGATGGATACCTAGTGTGG  |      |      |
| TsaccharolyticumYS485_16S           | (792)  | GATTAGATACCGTGGTAGTCCAGCGGTAAACGATGGATACCTAGTGTGG  |      |      |
| Consensus                           | (801)  | GATTAGATACCGTGGTAGTCCAGCGGTAAACGATGGATACCTAGTGTGG  |      |      |
| Tethanolicus39E_16S                 | (803)  | GAAGCGGAGGCTTGGGCGGCTTGAATCGAGCAATAGCTCAGGAGG      | 901  | 950  |
| Thermoanaerobacter_sp_strain_59_16S | (762)  | GAAGCGGAGGCTTGGGCGGCTTGAATCGAGCAATAGCTCAGGAGG      |      |      |
| TsaccharolyticumB6A-RI_16S          | (842)  | GAAGCGGAGGCTTGGGCGGCTTGAATCGAGCAATAGCTCAGGAGG      |      |      |
| TsaccharolyticumYS485_16S           | (842)  | GAAGCGGAGGCTTGGGCGGCTTGAATCGAGCAATAGCTCAGGAGG      |      |      |
| Consensus                           | (851)  | GAAGCGGAGGCTTGGGCGGCTTGAATCGAGCAATAGCTCAGGAGG      |      |      |
| Tethanolicus39E_16S                 | (853)  | GGGCTTGACATCCACAGAAATCGGGTAGAAATACCCAGGCTCGTGC     | 951  | 1000 |
| Thermoanaerobacter_sp_strain_59_16S | (810)  | GGGCTTGACATCCACAGAAATCGGGTAGAAATACCCAGGCTCGTGC     |      |      |
| TsaccharolyticumB6A-RI_16S          | (890)  | GGGCTTGACATCCACAGAAATCGGGTAGAAATACCCAGGCTCGTGC     |      |      |
| TsaccharolyticumYS485_16S           | (890)  | GGGCTTGACATCCACAGAAATCGGGTAGAAATACCCAGGCTCGTGC     |      |      |
| Consensus                           | (901)  | GGGCTTGACATCCACAGAAATCGGGTAGAAATACCCAGGCTCGTGC     |      |      |
| Tethanolicus39E_16S                 | (903)  | AAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCA | 1001 | 1050 |
| Thermoanaerobacter_sp_strain_59_16S | (860)  | AAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCA |      |      |
| TsaccharolyticumB6A-RI_16S          | (940)  | AAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCA |      |      |
| TsaccharolyticumYS485_16S           | (940)  | AAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCA |      |      |
| Consensus                           | (951)  | AAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCA |      |      |
| Tethanolicus39E_16S                 | (953)  | GGGCTTGACATCCACAGAAATCGGGTAGAAATACCCAGGCTCGTGC     | 1051 | 1100 |
| Thermoanaerobacter_sp_strain_59_16S | (910)  | GGGCTTGACATCCACAGAAATCGGGTAGAAATACCCAGGCTCGTGC     |      |      |
| TsaccharolyticumB6A-RI_16S          | (990)  | GGGCTTGACATCCACAGAAATCGGGTAGAAATACCCAGGCTCGTGC     |      |      |
| TsaccharolyticumYS485_16S           | (990)  | GGGCTTGACATCCACAGAAATCGGGTAGAAATACCCAGGCTCGTGC     |      |      |
| Consensus                           | (1001) | GGGCTTGACATCCACAGAAATCGGGTAGAAATACCCAGGCTCGTGC     |      |      |
| Tethanolicus39E_16S                 | (1003) | AAAGGTGAGGACCTGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG    | 1101 | 1150 |
| Thermoanaerobacter_sp_strain_59_16S | (955)  | AAAGGTGAGGACCTGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG    |      |      |
| TsaccharolyticumB6A-RI_16S          | (1035) | AAAGGTGAGGACCTGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG    |      |      |
| TsaccharolyticumYS485_16S           | (1035) | AAAGGTGAGGACCTGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG    |      |      |
| Consensus                           | (1051) | AAAGGTGAGGACCTGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG    |      |      |
| Tethanolicus39E_16S                 | (1053) | GTGAGATGCTGGGTTAAGTCCGCAACGACCGCAACCGCTCTGCTGCTG   | 1151 | 1200 |
| Thermoanaerobacter_sp_strain_59_16S | (1004) | GTGAGATGCTGGGTTAAGTCCGCAACGACCGCAACCGCTCTGCTGCTG   |      |      |
| TsaccharolyticumB6A-RI_16S          | (1084) | GTGAGATGCTGGGTTAAGTCCGCAACGACCGCAACCGCTCTGCTGCTG   |      |      |
| TsaccharolyticumYS485_16S           | (1084) | GTGAGATGCTGGGTTAAGTCCGCAACGACCGCAACCGCTCTGCTGCTG   |      |      |



도면30C

|                                     |        |                                                    |
|-------------------------------------|--------|----------------------------------------------------|
| Consensus                           | (1101) | GTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTGTTGGTAGTT |
|                                     | 1151   | 1200                                               |
| Tethanolicus39E_16S                 | (1103) | GGCAGCGGGTGCACCTGGGCTGGCTGAGCGGATGGCGGTGGCTACACGG  |
| Thermoanaerobacter_sp_strain_59_16S | (1054) | ACGAGCG--TAAACAGGGGCACTCTACCCAGACGCGCGTGGCTTACACGG |
| TsaccharolyticumB6A-RI_16S          | (1134) | ACGAGCG--TAAACAGGGGCACTCTACCCAGACGCGCGTGGCTTACACGG |
| TsaccharolyticumYS485_16S           | (1134) | ACGAGCG--TAAACAGGGGCACTCTACCCAGACGCGCGTGGCTTACACGG |
| Consensus                           | (1151) | ACGAGCG--TAAAGACGGGACTCTACCGAGACTGCCGTGGAGAACACGG  |
|                                     | 1201   | 1250                                               |
| Tethanolicus39E_16S                 | (1153) | AGCAGCGCTGGGATGGGCTGAAATCGAATGGGCTATGGCGTGGGCT     |
| Thermoanaerobacter_sp_strain_59_16S | (1102) | AGGAGCGCGGCAATGGCTCAAAATCAATCGGCTTATGGCGTGGGCT     |
| TsaccharolyticumB6A-RI_16S          | (1182) | AGGAGCGCGGCAATGGCTCAAAATCAATCGGCTTATGGCGTGGGCT     |
| TsaccharolyticumYS485_16S           | (1182) | AGGAGCGCGGCAATGGCTCAAAATCAATCGGCTTATGGCGTGGGCT     |
| Consensus                           | (1201) | AGGAAGCGGGGATGACGTCAAATCATCATGCCCTTTATGCCCTGGGCTA  |
|                                     | 1251   | 1300                                               |
| Tethanolicus39E_16S                 | (1203) | CACACGTGCTACAAATGGGCTGAAATCGAATGGGCTATGGCGTGGGCT   |
| Thermoanaerobacter_sp_strain_59_16S | (1152) | CACACGTGCTACAAATGGGCTGAAATCGAATGGGCTATGGCGTGGGCT   |
| TsaccharolyticumB6A-RI_16S          | (1232) | CACACGTGCTACAAATGGGCTGAAATCGAATGGGCTATGGCGTGGGCT   |
| TsaccharolyticumYS485_16S           | (1232) | CACACGTGCTACAAATGGGCTGAAATCGAATGGGCTATGGCGTGGGCT   |
| Consensus                           | (1251) | CACACGTGCTACAAATGGGCTGAAATCGAATGGGCTATGGCGTGGGCT   |
|                                     | 1301   | 1350                                               |
| Tethanolicus39E_16S                 | (1253) | GGCAATCCGAGAAACAGGTCCAGTTGCAATTCGAGGCTGCAACCTGGC   |
| Thermoanaerobacter_sp_strain_59_16S | (1202) | GGCAATCCGAGAAACAGGTCCAGTTGCAATTCGAGGCTGCAACCTGGC   |
| TsaccharolyticumB6A-RI_16S          | (1282) | GGCAATCCGAGAAACAGGTCCAGTTGCAATTCGAGGCTGCAACCTGGC   |
| TsaccharolyticumYS485_16S           | (1282) | GGCAATCCGAGAAACAGGTCCAGTTGCAATTCGAGGCTGCAACCTGGC   |
| Consensus                           | (1301) | GCGAATCCGAGAAACAGGTCCAGTTGCAATTCGAGGCTGCAACCTGGC   |
|                                     | 1351   | 1400                                               |
| Tethanolicus39E_16S                 | (1303) | TGCATGAAGACGGAGTTGCTAGTAATCGCGGATCAGCATGCCCGGTGAA  |
| Thermoanaerobacter_sp_strain_59_16S | (1252) | TGCATGAAGACGGAGTTGCTAGTAATCGCGGATCAGCATGCCCGGTGAA  |
| TsaccharolyticumB6A-RI_16S          | (1332) | TGCATGAAGACGGAGTTGCTAGTAATCGCGGATCAGCATGCCCGGTGAA  |
| TsaccharolyticumYS485_16S           | (1332) | TGCATGAAGACGGAGTTGCTAGTAATCGCGGATCAGCATGCCCGGTGAA  |
| Consensus                           | (1351) | TGCATGAAGACGGAGTTGCTAGTAATCGCGGATCAGCATGCCCGGTGAA  |
|                                     | 1401   | 1450                                               |
| Tethanolicus39E_16S                 | (1353) | TACGTT--CCCGGGCTTGTACACACCGCCCGTCACACCGAGAGTTTAC   |
| Thermoanaerobacter_sp_strain_59_16S | (1302) | TACGTT--CCCGGGCTTGTACACACCGCCCGTCACACCGAGAGTTTAC   |
| TsaccharolyticumB6A-RI_16S          | (1382) | TACGTT--CCCGGGCTTGTACACACCGCCCGTCACACCGAGAGTTTAC   |
| TsaccharolyticumYS485_16S           | (1382) | TACGTT--CCCGGGCTTGTACACACCGCCCGTCACACCGAGAGTTTAC   |
| Consensus                           | (1401) | TACGTT--CCCGGGCTTGTACACACCGCCCGTCACACCGAGAGTTTAC   |
|                                     | 1451   | 1500                                               |
| Tethanolicus39E_16S                 | (1402) | AACACCCGAAAGTCACTGACCTAACCGCAAGGAGGAGCTGCCGAAGTGG  |
| Thermoanaerobacter_sp_strain_59_16S | (1351) | AACACCCGAAAGTCACTGACCTAACCGCAAGGAGGAGCTGCCGAAGTGG  |
| TsaccharolyticumB6A-RI_16S          | (1431) | AACACCCGAAAGTCACTGACCTAACCGCAAGGAGGAGCTGCCGAAGTGG  |
| TsaccharolyticumYS485_16S           | (1432) | AACACCCGAAAGTCACTGACCTAACCGCAAGGAGGAGCTGCCGAAGTGG  |
| Consensus                           | (1451) | AACACCCGAAAGTCACTGACCTAACCGCAAGGAGGAGCTGCCGAAGTGG  |
|                                     | 1501   | 1550                                               |
| Tethanolicus39E_16S                 | (1451) | GGCTGGATCACCTCTTC-----                             |
| Thermoanaerobacter_sp_strain_59_16S | (1396) | GGCTGGATCACCTCTTC-----                             |
| TsaccharolyticumB6A-RI_16S          | (1481) | GGCTGGATCACCTCTTC-----                             |
| TsaccharolyticumYS485_16S           | (1482) | GGCTGGATCACCTCTTC-----                             |
| Consensus                           | (1501) | GGCTGGATCACCTCTTC-----                             |
|                                     | 1551   | 1572                                               |
| Tethanolicus39E_16S                 | (1501) | GGCTGGATCACCTCTTC-----                             |
| Thermoanaerobacter_sp_strain_59_16S | (1396) | GGCTGGATCACCTCTTC-----                             |
| TsaccharolyticumB6A-RI_16S          | (1531) | GGCTGGATCACCTCTTC-----                             |
| TsaccharolyticumYS485_16S           | (1532) | GGCTGGATCACCTCTTC-----                             |
| Consensus                           | (1551) | GGCTGGATCACCTCTTC-----                             |

도면31A

|                                      |                                                          |     |
|--------------------------------------|----------------------------------------------------------|-----|
|                                      | 1                                                        | 50  |
| Thermoanaerobacter_sp._strain_59_pta | (1) GTGTATACATATATTTCTCTTTTGTAGTAAGAGGAATGTATAAAAAATAA   |     |
| Tpsuedoethanolicus_pta               | (1) -----                                                |     |
| Tsaccharolyticum_B6A-RI_pta          | (1) GTGTATACATATATTTCTCTTTTGTAGTAAGAGGAATGTATAAAAAATAA   |     |
| Tsaccharolyticum_YS485_pta           | (1) GTGTATACATATATTTCTCTG--TTAGTAAGAGGAATGTATAAAAAATAA   |     |
| Consensus                            | (1) GTGTATACATATATTTCTCTTTTGTAGTAAGAGGAATGTATAAAAAATAA   |     |
|                                      | 51                                                       | 100 |
| Thermoanaerobacter_sp._strain_59_pta | (51) ATATTTTAAAGGAAGGACGATCTTACAGCAATTCTCAACCTG          |     |
| Tpsuedoethanolicus_pta               | (1) -----                                                |     |
| Tsaccharolyticum_B6A-RI_pta          | (51) ATATTTTAAAGGAAGGACGATCTTACAGCAATTCTCAACCTG          |     |
| Tsaccharolyticum_YS485_pta           | (48) ATATTTTAAAGGAAGGACGATCTTACAGCAATTCTCAACCTG          |     |
| Consensus                            | (51) ATATTTTAAAGGAAGGACGATCTTATGAGCATATTCAAACATCATTG     |     |
|                                      | 101                                                      | 150 |
| Thermoanaerobacter_sp._strain_59_pta | (101) AAAAGCTAAAGTGTATAAAAGAAAATGTTCTGCCAGAGGTGCAGAA     |     |
| Tpsuedoethanolicus_pta               | (26) AAAGCTAAAGTGTATAAAAGAAAATGTTCTGCCAGAGGTGCAGAA       |     |
| Tsaccharolyticum_B6A-RI_pta          | (101) AAAAGCTAAAGTGTATAAAAGAAAATGTTCTGCCAGAGGTGCAGAA     |     |
| Tsaccharolyticum_YS485_pta           | (98) AAAAGCTAAAGTGTATAAAAGAAAATGTTCTGCCAGAGGTGCAGAA      |     |
| Consensus                            | (101) AAAAGCTAAAGTGTATAAAAGAAAATGTTCTGCCAGAGGTGCAGAA     |     |
|                                      | 151                                                      | 200 |
| Thermoanaerobacter_sp._strain_59_pta | (151) CCAACAATTAAGCTCTGATATATTAAGCAATTAAGCAATTAAGCAAT    |     |
| Tpsuedoethanolicus_pta               | (76) CCAACAATTAAGCTCTGATATATTAAGCAATTAAGCAATTAAGCAAT     |     |
| Tsaccharolyticum_B6A-RI_pta          | (151) CCAACAATTAAGCTCTGATATATTAAGCAATTAAGCAATTAAGCAAT    |     |
| Tsaccharolyticum_YS485_pta           | (148) CCAACAATTAAGCTCTGATATATTAAGCAATTAAGCAATTAAGCAAT    |     |
| Consensus                            | (151) CCCAGGACATTAAAGCTGCTGAAATAGTTTAAAGAGGAAATGTCAGAA   |     |
|                                      | 201                                                      | 250 |
| Thermoanaerobacter_sp._strain_59_pta | (201) TTGGTGTCTTCTGGAAATGAAGATGAGATAAGAAATGCTGCAAAAGACT  |     |
| Tpsuedoethanolicus_pta               | (126) TTGGTGTCTTCTGGAAATGAAGATGAGATAAGAAATGCTGCAAAAGACT  |     |
| Tsaccharolyticum_B6A-RI_pta          | (201) TTGGTGTCTTCTGGAAATGAAGATGAGATAAGAAATGCTGCAAAAGACT  |     |
| Tsaccharolyticum_YS485_pta           | (198) TTGGTGTCTTCTGGAAATGAAGATGAGATAAGAAATGCTGCAAAAGACT  |     |
| Consensus                            | (201) TTGGTGTCTTCTGGAAATGAAGATGAGATAAGAAATGCTGCAAAAGACT  |     |
|                                      | 251                                                      | 300 |
| Thermoanaerobacter_sp._strain_59_pta | (251) TGGACATATCCAAAGCTGAATCATTGACCTGTAAAGTCTGAATGTTT    |     |
| Tpsuedoethanolicus_pta               | (176) TGGACATATCCAAAGCTGAATCATTGACCTGTAAAGTCTGAATGTTT    |     |
| Tsaccharolyticum_B6A-RI_pta          | (251) TGGACATATCCAAAGCTGAATCATTGACCTGTAAAGTCTGAATGTTT    |     |
| Tsaccharolyticum_YS485_pta           | (248) TGGACATATCCAAAGCTGAATCATTGACCTGTAAAGTCTGAATGTTT    |     |
| Consensus                            | (251) TGGACATATCCAAAGCTGAATCATTGACCTGTAAAGTCTGAATGTTT    |     |
|                                      | 301                                                      | 350 |
| Thermoanaerobacter_sp._strain_59_pta | (301) CATAGGTATGCTAATGATTTTATGAGTAAAGAGAGCAAGGAATCAC     |     |
| Tpsuedoethanolicus_pta               | (226) CATAGGTATGCTAATGATTTTATGAGTAAAGAGAGCAAGGAATCAC     |     |
| Tsaccharolyticum_B6A-RI_pta          | (301) CATAGGTATGCTAATGATTTTATGAGTAAAGAGAGCAAGGAATCAC     |     |
| Tsaccharolyticum_YS485_pta           | (298) CATAGGTATGCTAATGATTTTATGAGTAAAGAGAGCAAGGAATCAC     |     |
| Consensus                            | (301) CATAGGTATGCTAATGATTTTATGAGTAAAGAGAGCAAGGAATCAC     |     |
|                                      | 351                                                      | 400 |
| Thermoanaerobacter_sp._strain_59_pta | (351) ATGGGAAAAGCCAGAGATCAATCGATTAATGATTAATGATTAATGAT    |     |
| Tpsuedoethanolicus_pta               | (276) ATGGGAAAAGCCAGAGATCAATCGATTAATGATTAATGATTAATGAT    |     |
| Tsaccharolyticum_B6A-RI_pta          | (351) ATGGGAAAAGCCAGAGATCAATCGATTAATGATTAATGATTAATGAT    |     |
| Tsaccharolyticum_YS485_pta           | (348) ATGGGAAAAGCCAGAGATCAATCGATTAATGATTAATGATTAATGAT    |     |
| Consensus                            | (351) GTTGGAAAAGCCAGAGAAACAATCAAGGATAATATCTATTGATGTA     |     |
|                                      | 401                                                      | 450 |
| Thermoanaerobacter_sp._strain_59_pta | (401) TCAATGCTTAAAGAGGTTATGCTGATGGATTGATCTGCGCTATTCAT    |     |
| Tpsuedoethanolicus_pta               | (326) TCAATGCTTAAAGAGGTTATGCTGATGGATTGATCTGCGCTATTCAT    |     |
| Tsaccharolyticum_B6A-RI_pta          | (401) TCAATGCTTAAAGAGGTTATGCTGATGGATTGATCTGCGCTATTCAT    |     |
| Tsaccharolyticum_YS485_pta           | (398) TCAATGCTTAAAGAGGTTATGCTGATGGATTGATCTGCGCTATTCAT    |     |
| Consensus                            | (401) TCAATGCTTAAAGAGGTTATGCTGATGGATTGATCTGCGCTATTCAT    |     |
|                                      | 451                                                      | 500 |
| Thermoanaerobacter_sp._strain_59_pta | (451) GCTACTGCAGATTATTAAAGACCTGCATTTCAGATAATTAAAGCGGCTCC |     |
| Tpsuedoethanolicus_pta               | (376) GCTACTGCAGATTATTAAAGACCTGCATTTCAGATAATTAAAGCGGCTCC |     |
| Tsaccharolyticum_B6A-RI_pta          | (451) GCTACTGCAGATTATTAAAGACCTGCATTTCAGATAATTAAAGCGGCTCC |     |
| Tsaccharolyticum_YS485_pta           | (448) GCTACTGCAGATTATTAAAGACCTGCATTTCAGATAATTAAAGCGGCTCC |     |
| Consensus                            | (451) GCTACTGCAGATTATTAAAGACCTGCATTTCAGATAATTAAAGCGGCTCC |     |
|                                      | 501                                                      | 550 |
| Thermoanaerobacter_sp._strain_59_pta | (501) AGGAGCAAAAGATAGTATCAAGCTTTTATATGGAAGTGCCTAATTGTTG  |     |
| Tpsuedoethanolicus_pta               | (426) AGGAGCAAAAGATAGTATCAAGCTTTTATATGGAAGTGCCTAATTGTTG  |     |
| Tsaccharolyticum_B6A-RI_pta          | (501) AGGAGCAAAAGATAGTATCAAGCTTTTATATGGAAGTGCCTAATTGTTG  |     |
| Tsaccharolyticum_YS485_pta           | (498) AGGAGCAAAAGATAGTATCAAGCTTTTATATGGAAGTGCCTAATTGTTG  |     |
| Consensus                            | (501) AGGAGCAAAAGATAGTATCAAGCTTTTATATGGAAGTGCCTAATTGTTG  |     |
|                                      | 551                                                      | 600 |
| Thermoanaerobacter_sp._strain_59_pta | (551) AATATGCTGAAAATGCTGATCTCTGTTGGTGATTTGGTGATTTGGTGAT  |     |
| Tpsuedoethanolicus_pta               | (476) AATATGCTGAAAATGCTGATCTCTGTTGGTGATTTGGTGATTTGGTGAT  |     |
| Tsaccharolyticum_B6A-RI_pta          | (551) AATATGCTGAAAATGCTGATCTCTGTTGGTGATTTGGTGATTTGGTGAT  |     |
| Tsaccharolyticum_YS485_pta           | (548) AATATGCTGAAAATGCTGATCTCTGTTGGTGATTTGGTGATTTGGTGAT  |     |
| Consensus                            | (551) AATATGCTGAAAATGCTGATCTCTGTTGGTGATTTGGTGATTTGGTGAT  |     |
|                                      | 601                                                      | 650 |

도면31B

|                                      |        |                                                     |
|--------------------------------------|--------|-----------------------------------------------------|
| Thermoanaerobacter_sp._strain_59_pta | (601)  | TCGCTAATGCAGAGAACTTGCTTCTATTGCTGACAAATGCTTAATAC     |
| Tpsuedoethanolicus_pta               | (526)  | AATGCAATGACAGCGGATTAAGCAGCAATGCGCATGTGCTGCGCCATAC   |
| Tsaccharolyticum_B6A-RI_pta          | (601)  | TCGCTAATGCAGAGAACTTGCTTCTATTGCTGACAAATGCTTAATAC     |
| Tsaccharolyticum_YS485_pta           | (598)  | TCGCTAATGCAGAGAACTTGCTTCTATTGCTGACAAATGCTTAATAC     |
| Consensus                            | (601)  | TCGCTAATGCAGAGAACTTGCTTCTATTGCTGACAAATGCTTAATAC     |
| Thermoanaerobacter_sp._strain_59_pta | (651)  | TGCAAA-SAAATTGCTGGGCTTTGAACCAAAAGTTGCTATGCTATCATTT  |
| Tpsuedoethanolicus_pta               | (576)  | TGCAAA-ATCCCTCTCGCAATTTGCTTTCGAACTTTGCTGCAATTT      |
| Tsaccharolyticum_B6A-RI_pta          | (651)  | TGCAAA-SAAATTGCTGGGCTTTGAACCAAAAGTTGCTATGCTATCATTT  |
| Tsaccharolyticum_YS485_pta           | (648)  | TGCAAA-SAAATTGCTGGGCTTTGAACCAAAAGTTGCTATGCTATCATTT  |
| Consensus                            | (651)  | TGCAAA-GAATTGTTGGGCTTTGAACCAAAAGTTGCTATGCTATCATTT   |
| Thermoanaerobacter_sp._strain_59_pta | (700)  | TTCGCAAAAGGCTAGTGCATCAGATGAATTAGTAGATAAGTAAAGAAAGC  |
| Tpsuedoethanolicus_pta               | (625)  | TCTATTAAGCAAGTGGCAACATGAATAGTAGATAAGTAAAGAAAGC      |
| Tsaccharolyticum_B6A-RI_pta          | (700)  | TTCGCAAAAGGCTAGTGCATCAGATGAATTAGTAGATAAGTAAAGAAAGC  |
| Tsaccharolyticum_YS485_pta           | (697)  | TCTATTAAGCAAGTGGCAACATGAATAGTAGATAAGTAAAGAAAGC      |
| Consensus                            | (701)  | TCTACAAAGGCTAGTGCATCAGATGAATTAGTAGATAAGTAAAGAAAGC   |
| Thermoanaerobacter_sp._strain_59_pta | (750)  | TTCGCAAAAGGCTAGTGCATCAGATGAATTAGTAGATAAGTAAAGAAAGC  |
| Tpsuedoethanolicus_pta               | (675)  | TCTATTAAGCAAGTGGCAACATGAATAGTAGATAAGTAAAGAAAGC      |
| Tsaccharolyticum_B6A-RI_pta          | (750)  | TTCGCAAAAGGCTAGTGCATCAGATGAATTAGTAGATAAGTAAAGAAAGC  |
| Tsaccharolyticum_YS485_pta           | (747)  | TCTATTAAGCAAGTGGCAACATGAATAGTAGATAAGTAAAGAAAGC      |
| Consensus                            | (751)  | TTCGCAAAAGGCTAGTGCATCAGATGAATTAGTAGATAAGTAAAGAAAGC  |
| Thermoanaerobacter_sp._strain_59_pta | (799)  | TTCGCAAAAGGCTAGTGCATCAGATGAATTAGTAGATAAGTAAAGAAAGC  |
| Tpsuedoethanolicus_pta               | (724)  | TCTATTAAGCAAGTGGCAACATGAATAGTAGATAAGTAAAGAAAGC      |
| Tsaccharolyticum_B6A-RI_pta          | (799)  | TTCGCAAAAGGCTAGTGCATCAGATGAATTAGTAGATAAGTAAAGAAAGC  |
| Tsaccharolyticum_YS485_pta           | (796)  | TTCGCAAAAGGCTAGTGCATCAGATGAATTAGTAGATAAGTAAAGAAAGC  |
| Consensus                            | (801)  | TTGCAATTGGATGCTGCTCTTGTAAAGAAAGTTGCAGAGCTAAAGCGCC   |
| Thermoanaerobacter_sp._strain_59_pta | (849)  | AGGAGCAAAAGTTGGGGATGTCGAATGCTGATATATTCCTGATTATC     |
| Tpsuedoethanolicus_pta               | (774)  | AGGAGCAAAAGTTGGGGATGTCGAATGCTGATATATTCCTGATTATC     |
| Tsaccharolyticum_B6A-RI_pta          | (849)  | AGGAGCAAAAGTTGGGGATGTCGAATGCTGATATATTCCTGATTATC     |
| Tsaccharolyticum_YS485_pta           | (846)  | AGGAGCAAAAGTTGGGGATGTCGAATGCTGATATATTCCTGATTATC     |
| Consensus                            | (851)  | GGGAAGCAAAAGTTGGGGATGTCGAATGCTGATATATTCCTGATTATC    |
| Thermoanaerobacter_sp._strain_59_pta | (899)  | AAGCGGCAAAATAGGATATAGGCTTTGACAGAGATTAGCTT-SAAAT     |
| Tpsuedoethanolicus_pta               | (824)  | AAGCGGCAAAATAGGATATAGGCTTTGACAGAGATTAGCTT-SAAAT     |
| Tsaccharolyticum_B6A-RI_pta          | (899)  | AAGCGGCAAAATAGGATATAGGCTTTGACAGAGATTAGCTT-SAAAT     |
| Tsaccharolyticum_YS485_pta           | (896)  | AAGCGGCAAAATAGGATATAGGCTTTGACAGAGATTAGCTT-SAAAT     |
| Consensus                            | (901)  | AAGCTGGTAATATAGGATATAGGCTTTGACAGAGATTAGCTT-SAAAT    |
| Thermoanaerobacter_sp._strain_59_pta | (947)  | GCATTTGGACCTATAACACAGGAATGGGTGCACCGGTTAATGATTATC    |
| Tpsuedoethanolicus_pta               | (874)  | GCATTTGGACCTATAACACAGGAATGGGTGCACCGGTTAATGATTATC    |
| Tsaccharolyticum_B6A-RI_pta          | (949)  | GCATTTGGACCTATAACACAGGAATGGGTGCACCGGTTAATGATTATC    |
| Tsaccharolyticum_YS485_pta           | (946)  | GCATTTGGACCTATAACACAGGAATGGGTGCACCGGTTAATGATTATC    |
| Consensus                            | (951)  | GCAATTGGACCTATAACACAGGAATGGGTGCACCGGTTAATGATTATC    |
| Thermoanaerobacter_sp._strain_59_pta | (996)  | AAGAGGATGCAGCTATAGAGATTTGTTGACGTAATAGCAACAACAGCTG   |
| Tpsuedoethanolicus_pta               | (924)  | AAGAGGATGCAGCTATAGAGATTTGTTGACGTAATAGCAACAACAGCTG   |
| Tsaccharolyticum_B6A-RI_pta          | (999)  | AAGAGGATGCAGCTATAGAGATTTGTTGACGTAATAGCAACAACAGCTG   |
| Tsaccharolyticum_YS485_pta           | (996)  | AAGAGGATGCAGCTATAGAGATTTGTTGACGTAATAGCAACAACAGCTG   |
| Consensus                            | (1001) | AAGAGGATGCAGCTATAGAGATTTGTTGACGTAATAGCAACAACAGCTG   |
| Thermoanaerobacter_sp._strain_59_pta | (1044) | TAATAGGCTCA-----                                    |
| Tpsuedoethanolicus_pta               | (974)  | TAATAGGCTCA-----                                    |
| Tsaccharolyticum_B6A-RI_pta          | (1049) | TAATAGGCTCA-----                                    |
| Tsaccharolyticum_YS485_pta           | (1046) | TAATAGGCTCA-----                                    |
| Consensus                            | (1051) | TGCAGGCTCAA-----                                    |
| Thermoanaerobacter_sp._strain_59_pta | (1054) | -----                                               |
| Tpsuedoethanolicus_pta               | (1024) | AAATTTAGTCATGAACCTGTGGAAGCTCGTCATTAAAGATATCAATTGTTA |
| Tsaccharolyticum_B6A-RI_pta          | (1060) | -----                                               |
| Tsaccharolyticum_YS485_pta           | (1057) | -----                                               |
| Consensus                            | (1101) | -----                                               |
| Thermoanaerobacter_sp._strain_59_pta | (1054) | -----                                               |
| Tpsuedoethanolicus_pta               | (1074) | GATATGGATAATGGGAAAGTGCTAGCGAAAGGATTGGCGGAAAGGATAGG  |
| Tsaccharolyticum_B6A-RI_pta          | (1060) | -----                                               |
| Tsaccharolyticum_YS485_pta           | (1057) | -----                                               |
| Consensus                            | (1151) | -----                                               |
| Thermoanaerobacter_sp._strain_59_pta | (1054) | -----                                               |
| Tpsuedoethanolicus_pta               | (1124) | TATCAATGATTCTCTTTTAACTCATCAAGTAGAGGGCAAGATAAATAA    |
| Tsaccharolyticum_B6A-RI_pta          | (1060) | -----                                               |
| Tsaccharolyticum_YS485_pta           | (1057) | -----                                               |

도면31C

|                                      |        |                         |       |       |       |
|--------------------------------------|--------|-------------------------|-------|-------|-------|
| Consensus                            | (1201) | -----                   | 1251  | ----- | 1273  |
| Thermoanaerobacter_sp._strain_59_pta | (1054) | -----                   | ----- | ----- | ----- |
| Tpsuedoethanolicus_pta               | (1174) | AAATACAAAAAGATATGAAAAAT | ----- | ----- | ----- |
| Tsaccharolyticum_B6A-RI_pta          | (1060) | -----                   | ----- | ----- | ----- |
| Tsaccharolyticum_YS485_pta           | (1057) | -----                   | ----- | ----- | ----- |
| Consensus                            | (1251) | -----                   | ----- | ----- | ----- |



|                                      |       | 1                                                    | 50       |
|--------------------------------------|-------|------------------------------------------------------|----------|
| Thermoanaerobacter_sp._strain_59_ack | (1)   | -----                                                | -----    |
| Tpsuedoethanolicus_ack               | (1)   | GCTAATGCTATCGGACCAATTTCTCAAGGCTCTTGCAAAAACCTATCAATGA | -----    |
| Tsaccharolyticum_B6A-RI_ack          | (1)   | -----                                                | -----    |
| Tsaccharolyticum_Y5485_ack           | (1)   | -----                                                | -----    |
| Consensus                            | (1)   | -----                                                | -----    |
|                                      |       | 51                                                   | 100      |
| Thermoanaerobacter_sp._strain_59_ack | (1)   | -----                                                | -----    |
| Tpsuedoethanolicus_ack               | (51)  | CTTGTCAAGAGGTTGTAGTGTAGAAGATATTGTTAATGTTATAGCAATAA   | -----    |
| Tsaccharolyticum_B6A-RI_ack          | (1)   | -----                                                | -----    |
| Tsaccharolyticum_Y5485_ack           | (1)   | -----                                                | -----    |
| Consensus                            | (51)  | -----                                                | -----    |
|                                      |       | 101                                                  | 150      |
| Thermoanaerobacter_sp._strain_59_ack | (1)   | -----                                                | -----    |
| Tpsuedoethanolicus_ack               | (101) | CTTGTGTACAAGCTCAAGGGGTGCAAAAATACTTTTGGAGGAGCAGCGAT   | -----    |
| Tsaccharolyticum_B6A-RI_ack          | (1)   | -----                                                | ATGAAATC |
| Tsaccharolyticum_Y5485_ack           | (1)   | -----                                                | ATGAAAT  |
| Consensus                            | (101) | -----                                                | ATGAAAT  |
|                                      |       | 151                                                  | 200      |
| Thermoanaerobacter_sp._strain_59_ack | (1)   | -----                                                | -----    |
| Tpsuedoethanolicus_ack               | (151) | TATGAAATTTTGAATCTGAATCTGTGGAAGCTCTGCTATTAAGAGATACAA  | -----    |
| Tsaccharolyticum_B6A-RI_ack          | (9)   | TATGAAATTTCTGGTTATTAATTGTGGAAGTCTCTTCACTAAAA-TATCAA  | -----    |
| Tsaccharolyticum_Y5485_ack           | (9)   | TATGAAATTAATGGTTATTAATTGTGGAAGTCTCTTGGTAAAA-TATCAA   | -----    |
| Consensus                            | (151) | TATGAAATTTCTGGTTATTAATTGTGGAAGTCTCTTCACTAAAA-TATCAA  | -----    |
|                                      |       | 201                                                  | 250      |
| Thermoanaerobacter_sp._strain_59_ack | (1)   | -----                                                | -----    |
| Tpsuedoethanolicus_ack               | (201) | TTGTGAGTATGGAATATGGGAGAGCTAGCGAAAGGATTTGGCGGAAG      | -----    |
| Tsaccharolyticum_B6A-RI_ack          | (58)  | TGTATGTATGATCAATTSATGGAAATCTGCTGGCAAAAGGCGTTGCGTGAAG | -----    |
| Tsaccharolyticum_Y5485_ack           | (58)  | CGATGATGATCAATTSATGGAAATCTGCTGGCAAAAGGCGTTGCGTGAAG   | -----    |
| Consensus                            | (201) | TTGATTGAATCAA TGATGGAAATGTCTGGCAAAGCGCTTCTGTGAAG     | -----    |
|                                      |       | 251                                                  | 300      |
| Thermoanaerobacter_sp._strain_59_ack | (1)   | -----                                                | -----    |
| Tpsuedoethanolicus_ack               | (251) | GATAGCTTCTCATGATTTCTTTTAACTGATCTATAGTGGCAAAAG        | -----    |
| Tsaccharolyticum_B6A-RI_ack          | (108) | TATCGGCATTAATGATTCCTGCTGTGACGATATGCTCTATGCTG         | -----    |
| Tsaccharolyticum_Y5485_ack           | (108) | TATCGGCATTAATGATTCCTGCTGTGACGATATGCTCTATGCTG         | -----    |
| Consensus                            | (251) | AATCGGCATTAATGATTCCTGCTGTGAC CATAATGCTACACGG         | AGAAA    |
|                                      |       | 301                                                  | 350      |
| Thermoanaerobacter_sp._strain_59_ack | (16)  | AAATCAAGTATTAAGGACATGAAAGCATGCAAAAGCGCAATTAATG       | -----    |
| Tpsuedoethanolicus_ack               | (301) | AAATCAAGTATTAAGGACATGAAAGCATGCAAAAGCGCAATTAATG       | -----    |
| Tsaccharolyticum_B6A-RI_ack          | (155) | AAATCAAGTATTAAGGACATGAAAGCATGCAAAAGCGCAATTAATG       | -----    |
| Tsaccharolyticum_Y5485_ack           | (155) | AAATCAAGTATTAAGGACATGAAAGCATGCAAAAGCGCAATTAATG       | -----    |
| Consensus                            | (301) | AAATCAAGTATTAAGGACATGAAAGCATGCAAAAGCGCAATTAATG       | -----    |
|                                      |       | 351                                                  | 400      |
| Thermoanaerobacter_sp._strain_59_ack | (65)  | GTTTAGATGCTTTGGTAACTGACTGACATGCGCGTATAAAGGATATGTC    | -----    |
| Tpsuedoethanolicus_ack               | (351) | GTTTAGATGCTTTGGTAACTGACTGACATGCGCGTATAAAGGATATGTC    | -----    |
| Tsaccharolyticum_B6A-RI_ack          | (205) | GTTTAGATGCTTTGGTAACTGACTGACATGCGCGTATAAAGGATATGTC    | -----    |
| Tsaccharolyticum_Y5485_ack           | (205) | GTTTAGATGCTTTGGTAACTGACTGACATGCGCGTATAAAGGATATGTC    | -----    |
| Consensus                            | (351) | GTTTAGATGCTTTGGTAACTGACTGACATGCGCGTATAAAGGATATGTC    | -----    |
|                                      |       | 401                                                  | 450      |
| Thermoanaerobacter_sp._strain_59_ack | (114) | ATGATAGATGCTTCTGCGCATGAGCTTCTTCTGCGGAGGAGATGCTTTTA   | -----    |
| Tpsuedoethanolicus_ack               | (401) | ATGATAGATGCTTCTGCGCATGAGCTTCTTCTGCGGAGGAGATGCTTTTA   | -----    |
| Tsaccharolyticum_B6A-RI_ack          | (255) | ATGATAGATGCTTCTGCGCATGAGCTTCTTCTGCGGAGGAGATGCTTTTA   | -----    |
| Tsaccharolyticum_Y5485_ack           | (255) | ATGATAGATGCTTCTGCGCATGAGCTTCTTCTGCGGAGGAGATGCTTTTA   | -----    |
| Consensus                            | (401) | TGAGATAGATGCTTCTGCGCATGAGCTTCTTCTGCGGAGGAGATGCTTTTA  | -----    |
|                                      |       | 451                                                  | 500      |
| Thermoanaerobacter_sp._strain_59_ack | (164) | ATGATAGATGCTTCTGCGCATGAGCTTCTTCTGCGGAGGAGATGCTTTTA   | -----    |
| Tpsuedoethanolicus_ack               | (451) | ATGATAGATGCTTCTGCGCATGAGCTTCTTCTGCGGAGGAGATGCTTTTA   | -----    |
| Tsaccharolyticum_B6A-RI_ack          | (305) | ATGATAGATGCTTCTGCGCATGAGCTTCTTCTGCGGAGGAGATGCTTTTA   | -----    |
| Tsaccharolyticum_Y5485_ack           | (305) | ATGATAGATGCTTCTGCGCATGAGCTTCTTCTGCGGAGGAGATGCTTTTA   | -----    |
| Consensus                            | (451) | CATCATTAGTCTTCTCATAAATGATGAAGTGTATAAAGCGATAACAGATTGT | -----    |
|                                      |       | 501                                                  | 550      |
| Thermoanaerobacter_sp._strain_59_ack | (214) | ATGATAGATGCTTCTGCGCATGAGCTTCTTCTGCGGAGGAGATGCTTTTA   | -----    |
| Tpsuedoethanolicus_ack               | (501) | ATGATAGATGCTTCTGCGCATGAGCTTCTTCTGCGGAGGAGATGCTTTTA   | -----    |
| Tsaccharolyticum_B6A-RI_ack          | (355) | ATGATAGATGCTTCTGCGCATGAGCTTCTTCTGCGGAGGAGATGCTTTTA   | -----    |
| Tsaccharolyticum_Y5485_ack           | (355) | ATGATAGATGCTTCTGCGCATGAGCTTCTTCTGCGGAGGAGATGCTTTTA   | -----    |
| Consensus                            | (501) | ATGATAGATGCTTCTGCGCATGAGCTTCTTCTGCGGAGGAGATGCTTTTA   | -----    |
|                                      |       | 551                                                  | 600      |
| Thermoanaerobacter_sp._strain_59_ack | (264) | ATGATAGATGCTTCTGCGCATGAGCTTCTTCTGCGGAGGAGATGCTTTTA   | -----    |
| Tpsuedoethanolicus_ack               | (551) | ATGATAGATGCTTCTGCGCATGAGCTTCTTCTGCGGAGGAGATGCTTTTA   | -----    |
| Tsaccharolyticum_B6A-RI_ack          | (405) | ATGATAGATGCTTCTGCGCATGAGCTTCTTCTGCGGAGGAGATGCTTTTA   | -----    |
| Tsaccharolyticum_Y5485_ack           | (405) | ATGATAGATGCTTCTGCGCATGAGCTTCTTCTGCGGAGGAGATGCTTTTA   | -----    |
| Consensus                            | (551) | TTGCCAGAAATCATGCGCAACGTTTCAATGTTGGCGGTATTTGATACAG    | -----    |
|                                      |       | 601                                                  | 650      |

도면32B

|                                      |        |                                                      |
|--------------------------------------|--------|------------------------------------------------------|
| Thermoanaerobacter_sp._strain_59_ack | (314)  | CGTTTATGCGATTCCTTTATGCAATATCTTTATCCAAATACCTTAT       |
| Tpsuedoethanolicus_ack               | (601)  | CTTTCATGACACAAATGCCTGATTATGCATATCTTTATCCAAATACCTTAT  |
| Tsaccharolyticum_B6A-RI_ack          | (455)  | CGTTTATGCGATTCCTTTATGCAATATCTTTATCCAAATACCTTAT       |
| Tsaccharolyticum_YS485_ack           | (455)  | CGTTTATGCGATTCCTTTATGCAATATCTTTATCCAAATACCTTAT       |
| Consensus                            | (601)  | CCTTTCATGACACAAATGCCTGATTATGCATATCTTTATCCAAATACCTTAT |
| Thermoanaerobacter_sp._strain_59_ack | (364)  | GAATACACAAAGTACAGGATCAGAAGATATGGATTTCATGGCACATC      |
| Tpsuedoethanolicus_ack               | (651)  | GAATACACAAAGTACAGGATCAGAAGATATGGATTTCATGGCACATC      |
| Tsaccharolyticum_B6A-RI_ack          | (505)  | GAATACACAAAGTACAGGATCAGAAGATATGGATTTCATGGCACATC      |
| Tsaccharolyticum_YS485_ack           | (505)  | GAATACACAAAGTACAGGATCAGAAGATATGGATTTCATGGCACATC      |
| Consensus                            | (651)  | GAATACACACAAAGTACAGGATCAGAAGATATGGATTTCATGGCACATC    |
| Thermoanaerobacter_sp._strain_59_ack | (414)  | GCATAAATATGTTTCAAAATAGGGCTGCAGAGATTTTGAATAAACCTATTG  |
| Tpsuedoethanolicus_ack               | (701)  | GCATAAATATGTTTCAAAATAGGGCTGCAGAGATTTTGAATAAACCTATTG  |
| Tsaccharolyticum_B6A-RI_ack          | (555)  | GCATAAATATGTTTCAAAATAGGGCTGCAGAGATTTTGAATAAACCTATTG  |
| Tsaccharolyticum_YS485_ack           | (555)  | GCATAAATATGTTTCAAAATAGGGCTGCAGAGATTTTGAATAAACCTATTG  |
| Consensus                            | (701)  | GCATAAATATGTTTCAAAATAGGGCTGCAGAGATTTTGAATAAACCTATTG  |
| Thermoanaerobacter_sp._strain_59_ack | (464)  | AAGATTGAAAATCATAACTTGTCTATCTTGGAAATGGCTCCAGATTGGCT   |
| Tpsuedoethanolicus_ack               | (751)  | AAGATTGAAAATCATAACTTGTCTATCTTGGAAATGGCTCCAGATTGGCT   |
| Tsaccharolyticum_B6A-RI_ack          | (605)  | AAGATTGAAAATCATAACTTGTCTATCTTGGAAATGGCTCCAGATTGGCT   |
| Tsaccharolyticum_YS485_ack           | (605)  | AAGATTGAAAATCATAACTTGTCTATCTTGGAAATGGCTCCAGATTGGCT   |
| Consensus                            | (751)  | AAGATTGAAAATCATAACTTGTCTATCTTGGAAATGGCTCCAGATTGGCT   |
| Thermoanaerobacter_sp._strain_59_ack | (514)  | GCTGTCAAATATGGTAAATCATTGACACAGCATGGGATTACACCATT      |
| Tpsuedoethanolicus_ack               | (801)  | GCTGTCAAATATGGTAAATCATTGACACAGCATGGGATTACACCATT      |
| Tsaccharolyticum_B6A-RI_ack          | (655)  | GCTGTCAAATATGGTAAATCATTGACACAGCATGGGATTACACCATT      |
| Tsaccharolyticum_YS485_ack           | (655)  | GCTGTCAAATATGGTAAATCATTGACACAGCATGGGATTACACCATT      |
| Consensus                            | (801)  | GCTGTCAAATATGGTAAATCATTGACACAGCATGGGATTACACCATT      |
| Thermoanaerobacter_sp._strain_59_ack | (564)  | AGAAGGTTTGGCTATGGGTACAGCATCTGGAAGTATAGACCCATCCATT    |
| Tpsuedoethanolicus_ack               | (851)  | AGAAGGTTTGGCTATGGGTACAGCATCTGGAAGTATAGACCCATCCATT    |
| Tsaccharolyticum_B6A-RI_ack          | (705)  | AGAAGGTTTGGCTATGGGTACAGCATCTGGAAGTATAGACCCATCCATT    |
| Tsaccharolyticum_YS485_ack           | (705)  | AGAAGGTTTGGCTATGGGTACAGCATCTGGAAGTATAGACCCATCCATT    |
| Consensus                            | (851)  | AGAAGGTTTGGCTATGGGTACAGCATCTGGAAGTATAGACCCATCCATT    |
| Thermoanaerobacter_sp._strain_59_ack | (614)  | TTTCTTATCTTATGGAAGGATATAGTGTCTGAAGAGGTATGAAT         |
| Tpsuedoethanolicus_ack               | (901)  | TTTCTTATCTTATGGAAGGATATAGTGTCTGAAGAGGTATGAAT         |
| Tsaccharolyticum_B6A-RI_ack          | (755)  | TTTCTTATCTTATGGAAGGATATAGTGTCTGAAGAGGTATGAAT         |
| Tsaccharolyticum_YS485_ack           | (755)  | TTTCTTATCTTATGGAAGGATATAGTGTCTGAAGAGGTATGAAT         |
| Consensus                            | (901)  | TTTCTTATCTTATGGAAGGATATAGTGTCTGAAGAGGTATGAAT         |
| Thermoanaerobacter_sp._strain_59_ack | (664)  | ATATTAAATAAAAAATCTGGTGTTCACGGTATTTACAGGAATAGCAGCGA   |
| Tpsuedoethanolicus_ack               | (951)  | ATATTAAATAAAAAATCTGGTGTTCACGGTATTTACAGGAATAGCAGCGA   |
| Tsaccharolyticum_B6A-RI_ack          | (805)  | ATATTAAATAAAAAATCTGGTGTTCACGGTATTTACAGGAATAGCAGCGA   |
| Tsaccharolyticum_YS485_ack           | (805)  | ATATTAAATAAAAAATCTGGTGTTCACGGTATTTACAGGAATAGCAGCGA   |
| Consensus                            | (951)  | ATATTAAATAAAAAATCTGGTGTTCACGGTATTTACAGGAATAGCAGCGA   |
| Thermoanaerobacter_sp._strain_59_ack | (714)  | TTTGTAGAGCTTTAGAGATGCGCGCTTTAAAAATGGAGATGAAAGAGCTC   |
| Tpsuedoethanolicus_ack               | (1001) | TTTGTAGAGCTTTAGAGATGCGCGCTTTAAAAATGGAGATGAAAGAGCTC   |
| Tsaccharolyticum_B6A-RI_ack          | (855)  | TTTGTAGAGCTTTAGAGATGCGCGCTTTAAAAATGGAGATGAAAGAGCTC   |
| Tsaccharolyticum_YS485_ack           | (855)  | TTTGTAGAGCTTTAGAGATGCGCGCTTTAAAAATGGAGATGAAAGAGCTC   |
| Consensus                            | (1001) | TTTGTAGAGCTTTAGAGATGCGCGCTTTAAAAATGGAGATGAAAGAGCTC   |
| Thermoanaerobacter_sp._strain_59_ack | (764)  | AGTTGGCTTTAAATGTGTTTGCATATCGAGTAAAGAAGAGCATTTGGCGCT  |
| Tpsuedoethanolicus_ack               | (1051) | AGTTGGCTTTAAATGTGTTTGCATATCGAGTAAAGAAGAGCATTTGGCGCT  |
| Tsaccharolyticum_B6A-RI_ack          | (905)  | AGTTGGCTTTAAATGTGTTTGCATATCGAGTAAAGAAGAGCATTTGGCGCT  |
| Tsaccharolyticum_YS485_ack           | (905)  | AGTTGGCTTTAAATGTGTTTGCATATCGAGTAAAGAAGAGCATTTGGCGCT  |
| Consensus                            | (1051) | AGTTGGCTTTAAATGTGTTTGCATATCGAGTAAAGAAGAGCATTTGGCGCT  |
| Thermoanaerobacter_sp._strain_59_ack | (814)  | TATGCAGCAGCTATGGGAGCGTGTGATGTCATTGTATTACAGCAGGTGT    |
| Tpsuedoethanolicus_ack               | (1101) | TATGCAGCAGCTATGGGAGCGTGTGATGTCATTGTATTACAGCAGGTGT    |
| Tsaccharolyticum_B6A-RI_ack          | (955)  | TATGCAGCAGCTATGGGAGCGTGTGATGTCATTGTATTACAGCAGGTGT    |
| Tsaccharolyticum_YS485_ack           | (955)  | TATGCAGCAGCTATGGGAGCGTGTGATGTCATTGTATTACAGCAGGTGT    |
| Consensus                            | (1101) | TATGCAGCAGCTATGGGAGCGTGTGATGTCATTGTATTACAGCAGGTGT    |
| Thermoanaerobacter_sp._strain_59_ack | (864)  | TGGTGAAAATGGTCTGAGATACGAGAATTATACCTTGATGGATTAGAT     |
| Tpsuedoethanolicus_ack               | (1151) | TGGTGAAAATGGTCTGAGATACGAGAATTATACCTTGATGGATTAGAT     |
| Tsaccharolyticum_B6A-RI_ack          | (1005) | TGGTGAAAATGGTCTGAGATACGAGAATTATACCTTGATGGATTAGAT     |
| Tsaccharolyticum_YS485_ack           | (1005) | TGGTGAAAATGGTCTGAGATACGAGAATTATACCTTGATGGATTAGAT     |
| Consensus                            | (1151) | TGGTGAAAATGGTCTGAGATACGAGAATTATACCTTGATGGATTAGAT     |
| Thermoanaerobacter_sp._strain_59_ack | (914)  | CTTACGCGTCTGCTGCTTATGCAATATATATATACGACGATGATGAT      |
| Tpsuedoethanolicus_ack               | (1201) | CTTACGCGTCTGCTGCTTATGCAATATATATATACGACGATGATGAT      |
| Tsaccharolyticum_B6A-RI_ack          | (1055) | CTTACGCGTCTGCTGCTTATGCAATATATATATACGACGATGATGAT      |
| Tsaccharolyticum_YS485_ack           | (1055) | CTTACGCGTCTGCTGCTTATGCAATATATATATACGACGATGATGAT      |

도면32C

|                                      |        |                                                     |
|--------------------------------------|--------|-----------------------------------------------------|
| Consensus                            | (1201) | TTTGTAGGTTTACGCTTGGATAAAGAAAAAATAAAGTCAGAGGAAAAAGAA |
| Thermoanaerobacter_sp._strain_59_ack | (964)  | ACTATTATATCTACCGCTATTCGAAATCTAGCGCATCTGCTGCTGCT     |
| Tpsuedoethanolicus_ack               | (1251) | GAAATTATATCTACCGCTATTCGAAATCTAGCGCATCTGCTGCTGCT     |
| Tsaccharolyticum_B6A-RI_ack          | (1105) | ACTATTATATCTACCGCTATTCGAAATCTAGCGCATCTGCTGCTGCT     |
| Tsaccharolyticum_YS485_ack           | (1105) | ACTATTATATCTACCGCTATTCGAAATCTAGCGCATCTGCTGCTGCT     |
| Consensus                            | (1251) | ACTATTATATCTACCGCTATTCGAAATCTAGCGCATCTGCTGCTGCT     |
| Thermoanaerobacter_sp._strain_59_ack | (1014) | TAAAGCAAGATATCTGCTTGGCTTATGCTATCTGATGATGATGAT       |
| Tpsuedoethanolicus_ack               | (1301) | TAAAGCAAGATATCTGCTTGGCTTATGCTATCTGATGATGATGAT       |
| Tsaccharolyticum_B6A-RI_ack          | (1155) | TAAAGCAAGATATCTGCTTGGCTTATGCTATCTGATGATGATGAT       |
| Tsaccharolyticum_YS485_ack           | (1155) | TAAAGCAAGATATCTGCTTGGCTTATGCTATCTGATGATGATGAT       |
| Consensus                            | (1301) | TAAAGCAAGATATCTGCTTGGCTTATGCTATCTGATGATGATGAT       |
| Thermoanaerobacter_sp._strain_59_ack | (1064) | TAAAGTAG                                            |
| Tpsuedoethanolicus_ack               | (1351) | TAAAGTAG                                            |
| Tsaccharolyticum_B6A-RI_ack          | (1205) | TAAAGTAG                                            |
| Tsaccharolyticum_YS485_ack           | (1205) | TAAAGTAG                                            |
| Consensus                            | (1351) | TAAAGTAG                                            |

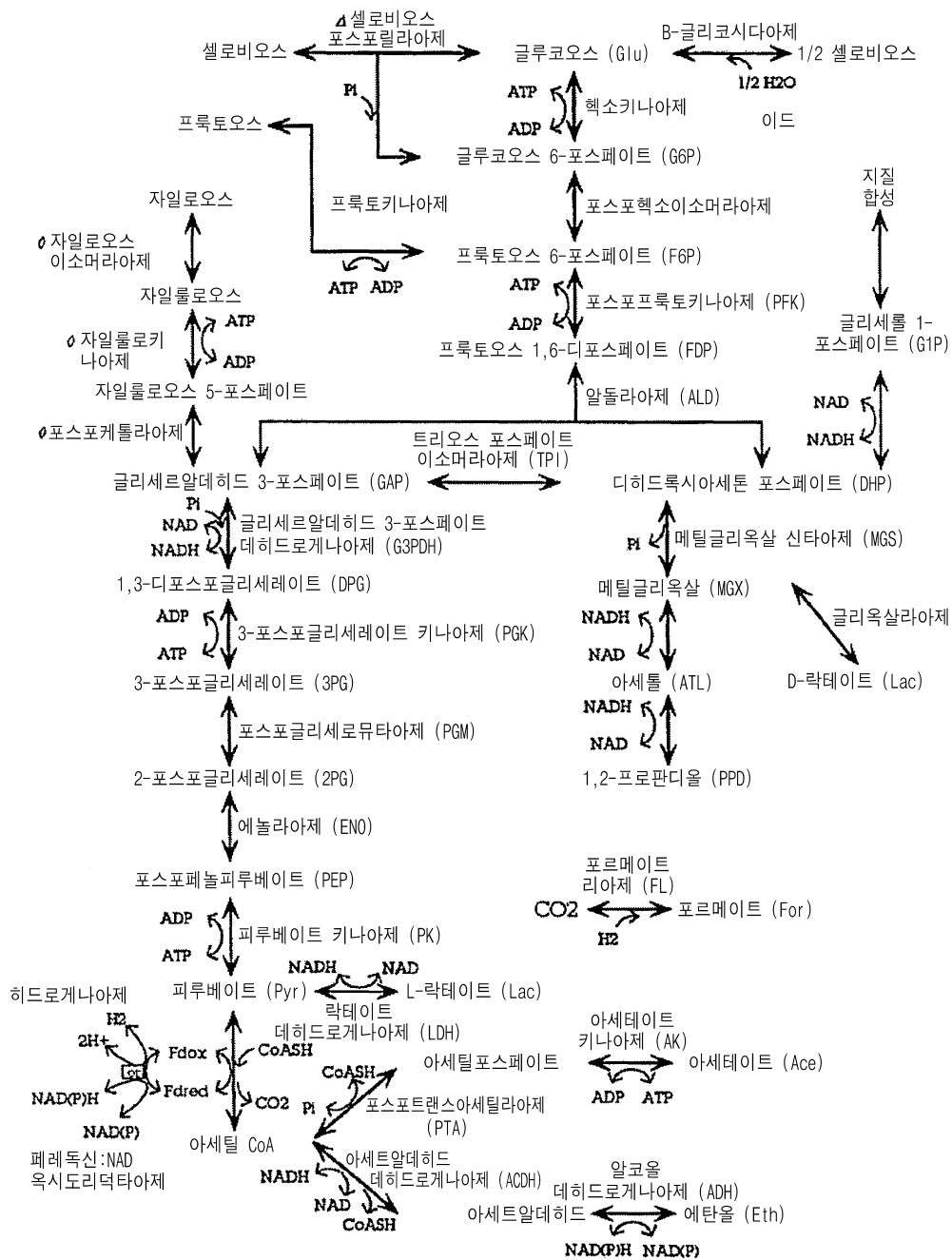


도면33A

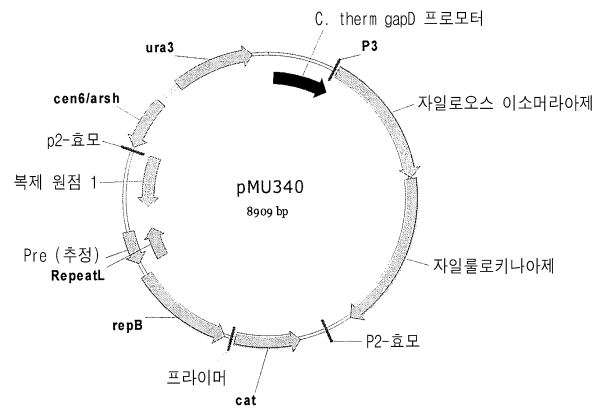
|                                 |       |                                                      |     |
|---------------------------------|-------|------------------------------------------------------|-----|
|                                 |       | 1                                                    | 50  |
| Thermoanaerobacter_sp._strain59 | (1)   | ATCAGTAAAGTGGGCATAATAGGTTGAGCATTGTAGGTGCTACATCTGC    |     |
| Tpseudoeethanolicus_39E         | (1)   | ATGACACAAAATATGTATAATAGGTTCTGGATTGTGCGGTCTACTACTGC   |     |
| Tsaccharolyticum_B6ARI          | (1)   | -----                                                |     |
| Tsaccharolyticum_Y8485          | (1)   | ATCAGCAAGCTAGCAATAATAGGATCTGGTTTGTAGGTGCAACATCGGC    |     |
| Consensus                       | (1)   | ATGAGCAAGTAGC ATAATAGGTTCTGGATTGTAGGTGCTACATCTGC     |     |
|                                 |       | 51                                                   | 100 |
| Thermoanaerobacter_sp._strain59 | (51)  | ATTACATTGGCTCTAAGTGGGACTGTGACAGACATTGTTTAGTAGATT     |     |
| Tpseudoeethanolicus_39E         | (51)  | ATTACACTEGCTTTGAGTGGGATTGCCAAACTATTGTATTATAGATA      |     |
| Tsaccharolyticum_B6ARI          | (1)   | -----                                                |     |
| Tsaccharolyticum_Y8485          | (51)  | ATTACGCTGGCATTAAGTGGGACTGTGACAGATATCGTGCTGGTGGATT    |     |
| Consensus                       | (51)  | ATTACACTGGCTTTAAGTGGGACTGTGACAGATATTGT TTAGTAGATT    |     |
|                                 |       | 101                                                  | 150 |
| Thermoanaerobacter_sp._strain59 | (101) | TAAACAAGSACAAGGGCATGCTTCCAAGTTTCCCTCTA               |     |
| Tpseudoeethanolicus_39E         | (101) | TTAATATAGACAAAGCAGATCTTTTCTTAAAGCCCG                 |     |
| Tsaccharolyticum_B6ARI          | (1)   | -----                                                |     |
| Tsaccharolyticum_Y8485          | (101) | TAAACAAGSACAAGGCTATGCGGATAGGCCTAATCTCA               |     |
| Consensus                       | (101) | TAAACAAGSACAAGGC ATAGGCGATGCACTGGATATAAGCCATGGCATA   |     |
|                                 |       | 151                                                  | 200 |
| Thermoanaerobacter_sp._strain59 | (151) | CCGCTTACAGTTAAATGCTTTCTCCAGTAAAGA                    |     |
| Tpseudoeethanolicus_39E         | (151) | CCGCTTAGTAAATGCTTTCTCCAGTAAAGA                       |     |
| Tsaccharolyticum_B6ARI          | (32)  | CCATTAACAGTTAAATGCTTTCTCCAGTAAAGA                    |     |
| Tsaccharolyticum_Y8485          | (151) | CCGCTTACAGTTAAATGCTTTCTCCAGTAAAGA                    |     |
| Consensus                       | (151) | CCGTTTATACAGCTGTAATGTGTATCAGGTGACTACAAAGATGTTGA      |     |
|                                 |       | 201                                                  | 250 |
| Thermoanaerobacter_sp._strain59 | (201) | GGCGCAGTCTGCGTCTCAATAGTGTGCTTAAAGTAA                 |     |
| Tpseudoeethanolicus_39E         | (201) | AGTTTTCATCAATCAATTAAGTAAAGTAAAGTAA                   |     |
| Tsaccharolyticum_B6ARI          | (82)  | AGCGCGCTCTCAATGCTGCTCAATGCTTAAAGTAA                  |     |
| Tsaccharolyticum_Y8485          | (201) | AGCGCAGTCTGCGTCTCAATAGTGTGCTTAAAGTAA                 |     |
| Consensus                       | (201) | AGCGCAGATGTAATAGTTGTGACAGCAGGTGCTGCTCAAAAGCCGGAG     |     |
|                                 |       | 251                                                  | 300 |
| Thermoanaerobacter_sp._strain59 | (251) | GTCTACCTGCTCTCAATAGTGTGCTTAAAGTAA                    |     |
| Tpseudoeethanolicus_39E         | (251) | TAACCTCAATGCTCTCAATAGTGTGCTTAAAGTAA                  |     |
| Tsaccharolyticum_B6ARI          | (132) | CAAGCTCTGCTCTCAATAGTGTGCTTAAAGTAA                    |     |
| Tsaccharolyticum_Y8485          | (251) | TAACCTCAATGCTCTCAATAGTGTGCTTAAAGTAA                  |     |
| Consensus                       | (251) | AGAC AGGCTTGACCTTGGAAGAAATATACGCTATATTAGTCCATG       |     |
|                                 |       | 301                                                  | 350 |
| Thermoanaerobacter_sp._strain59 | (301) | ATACCTGCTCTCAATAGTGTGCTTAAAGTAA                      |     |
| Tpseudoeethanolicus_39E         | (301) | GGGAAATCAATAGTGTGCTTAAAGTAA                          |     |
| Tsaccharolyticum_B6ARI          | (182) | ATACCTGCTCTCAATAGTGTGCTTAAAGTAA                      |     |
| Tsaccharolyticum_Y8485          | (301) | ATACCTGCTCTCAATAGTGTGCTTAAAGTAA                      |     |
| Consensus                       | (301) | ATACCTGAGCTT TTAAGTACAATGCAAGGCTATATATTGATTGT A      |     |
|                                 |       | 351                                                  | 400 |
| Thermoanaerobacter_sp._strain59 | (350) | CAATCTGCTCTCAATAGTGTGCTTAAAGTAA                      |     |
| Tpseudoeethanolicus_39E         | (350) | CAATCTGCTCTCAATAGTGTGCTTAAAGTAA                      |     |
| Tsaccharolyticum_B6ARI          | (231) | CAATCTGCTCTCAATAGTGTGCTTAAAGTAA                      |     |
| Tsaccharolyticum_Y8485          | (350) | CAATCTGCTCTCAATAGTGTGCTTAAAGTAA                      |     |
| Consensus                       | (351) | CAATCTGTAGATATACTGACGTACGTATACATCAAGATATCTGGACTT     |     |
|                                 |       | 401                                                  | 450 |
| Thermoanaerobacter_sp._strain59 | (400) | CCGCGCGCTCTCAATAGTGTGCTTAAAGTAA                      |     |
| Tpseudoeethanolicus_39E         | (400) | CCATCTCAATAGTGTGCTTAAAGTAA                           |     |
| Tsaccharolyticum_B6ARI          | (281) | CCATCTCAATAGTGTGCTTAAAGTAA                           |     |
| Tsaccharolyticum_Y8485          | (400) | CCATCTCAATAGTGTGCTTAAAGTAA                           |     |
| Consensus                       | (401) | CCATGGGGCAGAGTTTTCGGTTCTGCGACTGTTCTTGACAGTTCAAGGTT   |     |
|                                 |       | 451                                                  | 500 |
| Thermoanaerobacter_sp._strain59 | (450) | TACCAATCTCAATAGTGTGCTTAAAGTAA                        |     |
| Tpseudoeethanolicus_39E         | (450) | CAGATCTCAATAGTGTGCTTAAAGTAA                          |     |
| Tsaccharolyticum_B6ARI          | (331) | TACCAATCTCAATAGTGTGCTTAAAGTAA                        |     |
| Tsaccharolyticum_Y8485          | (450) | TACCAATCTCAATAGTGTGCTTAAAGTAA                        |     |
| Consensus                       | (451) | TAGGTATCTTTTAAAGCAAGCAATTGCAATATAGAT CCGAGAAATGTCCAC |     |
|                                 |       | 501                                                  | 550 |
| Thermoanaerobacter_sp._strain59 | (499) | CGAGCAATCTCAATAGTGTGCTTAAAGTAA                       |     |
| Tpseudoeethanolicus_39E         | (499) | CGATCTCAATAGTGTGCTTAAAGTAA                           |     |
| Tsaccharolyticum_B6ARI          | (381) | CGAGCAATCTCAATAGTGTGCTTAAAGTAA                       |     |
| Tsaccharolyticum_Y8485          | (499) | CGAGCAATCTCAATAGTGTGCTTAAAGTAA                       |     |
| Consensus                       | (501) | GGAAGGATAATTGGCGAGCATGGTATACAGAGTTTGCAGCATGGAGCAT    |     |
|                                 |       | 551                                                  | 600 |
| Thermoanaerobacter_sp._strain59 | (549) | ATCAATCTCAATAGTGTGCTTAAAGTAA                         |     |
| Tpseudoeethanolicus_39E         | (549) | TGCAATCTCAATAGTGTGCTTAAAGTAA                         |     |
| Tsaccharolyticum_B6ARI          | (431) | ATCAATCTCAATAGTGTGCTTAAAGTAA                         |     |
| Tsaccharolyticum_Y8485          | (549) | ATCAATCTCAATAGTGTGCTTAAAGTAA                         |     |
| Consensus                       | (551) | AACAACATATCAGGAATATCATTTAATGAGTACTGCAGTTTATGCGGAC    |     |
|                                 |       | 601                                                  | 650 |

도면33B

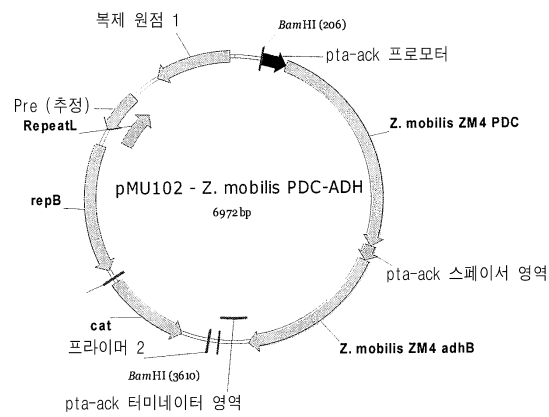
|                                 |       |                                                               |
|---------------------------------|-------|---------------------------------------------------------------|
| Thermoanaerobacter_sp._strain59 | (599) | <del>GACTTTCGTAATTCGAAATTTCCGPAAGCAAGCGAAGATGAGCTGTCGAT</del> |
| Tpseudoeothanolicus_39E         | (599) | <del>AAGCAATTCGAAAGATTTTAGGAGCGAGATTTTATGATTTCTCTAGA</del>    |
| Tsaccharolyticum_B6ARI          | (481) | <del>GAATCTCCAGCCGAAATTTTCAGAAAGCGAGTAGAGCAGACGCTGAAAT</del>  |
| Tsaccharolyticum_Y8485          | (599) | <del>GGCTCTCCAGCCGAAATTTTCAGAAAGCGAGTAGAGCAGACGCTGAAAT</del>  |
| Consensus                       | (601) | GCGTCTGTAAACAAATTTTCAGAAAGGAGTAGAAGATGAAGTTGTAAT              |
|                                 |       | 651 700                                                       |
| Thermoanaerobacter_sp._strain59 | (649) | <del>GGCGTTTACGAAATTTTCTTAAAGCGTGCCAGGCTTTCGCTGCGC</del>      |
| Tpseudoeothanolicus_39E         | (649) | <del>GGCGCTTACGCAATATGTAAGAAAGCGTGCCAGATTTTATCGCTGCG</del>    |
| Tsaccharolyticum_B6ARI          | (531) | <del>GGCGCTTACGCAATATGTAAGAAAGCGTGCTACGTTCTATCGCTGCGC</del>   |
| Tsaccharolyticum_Y8485          | (649) | <del>GGCGCTTACGCAATATGTAAGAAAGCGTGCTACGTTCTATCGCTGCGC</del>   |
| Consensus                       | (651) | GCTGCTTATAAGATAATAGACAAAAAGGTGCTACATATTATGCTGTGCG             |
|                                 |       | 701 750                                                       |
| Thermoanaerobacter_sp._strain59 | (699) | <del>TCATACGACTAAGAAGATTATTSAGTCTATCATAGCGATCAAAATTCAT</del>  |
| Tpseudoeothanolicus_39E         | (699) | <del>TCATACGACTAAGAAGATTCTAGAGCTATTTCTCAGATCAAAATTCAT</del>   |
| Tsaccharolyticum_B6ARI          | (581) | <del>AGTTGCGACTAAGAAGATTCTGCGAGTGCATCTTANAGATCAAAATTCAT</del> |
| Tsaccharolyticum_Y8485          | (699) | <del>AGTTGCGACTAAGAAGATTCTGCGAGTGCATCTTANAGATCAAAATTCAT</del> |
| Consensus                       | (701) | TGTTGCAGTAAGAAGATTGTGGAGTGTATCTTAAGAGATGAAATTCAT              |
|                                 |       | 751 800                                                       |
| Thermoanaerobacter_sp._strain59 | (749) | <del>TTCTTTGAGTTTCATCTCCATTAAATGGTGAATACGGTGTACAGATGTA</del>  |
| Tpseudoeothanolicus_39E         | (749) | <del>TTCTGAGTCTCTCATCTCCGCTTACCCGCCAATTTGGCTTCTCATCTCG</del>  |
| Tsaccharolyticum_B6ARI          | (631) | <del>TTCTGAGTCTCTCATCTCCGCTTACCCGCCAATTTGGCTTCTCATCTCG</del>  |
| Tsaccharolyticum_Y8485          | (749) | <del>TTCTGAGTCTCTCATCTCCATTAAATGGTGAATACGGTGTACAGATGTA</del>  |
| Consensus                       | (751) | TTCTACAGTTCATCTCCATTAAATGGCAATACGGTGTACAGATGTA                |
|                                 |       | 801 850                                                       |
| Thermoanaerobacter_sp._strain59 | (799) | <del>TCCTTAAGCTTGGCATCATTGTTGCGCAAAATGGTGTGCGAGGCTGCT</del>   |
| Tpseudoeothanolicus_39E         | (799) | <del>CTTTGCGGCTTCTCTGCGTGTTCGACGAATGGAACTGATATTAAT</del>      |
| Tsaccharolyticum_B6ARI          | (632) | <del>TCCTTAAGCTTGGCATCATTGTTGCGCAAAATGGTGTGCGAGGCTGCT</del>   |
| Tsaccharolyticum_Y8485          | (799) | <del>TCCTTAAGCTTGGCATCATTGTTGCGCAAAATGGTGTGCGAGGCTGCT</del>   |
| Consensus                       | (801) | TCTTTAAGCTTGGCATCATTGTTGCGCAAAATGGTGTGCAAGGATTCT              |
|                                 |       | 851 900                                                       |
| Thermoanaerobacter_sp._strain59 | (849) | <del>GGATTTCGCTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT</del>  |
| Tpseudoeothanolicus_39E         | (849) | <del>TGATTAATGACCTTCACAGCGAGGATTTCTGCTTTTAGAGATCAGGCG</del>   |
| Tsaccharolyticum_B6ARI          | (632) | <del>GGATTTCGCTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT</del>  |
| Tsaccharolyticum_Y8485          | (849) | <del>GGATTTCGCTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT</del>  |
| Consensus                       | (851) | GGA TTGCTTTTCTGTA GAAGAAGTTGAGAAGTTTAGACATTAGACAA             |
|                                 |       | 901 938                                                       |
| Thermoanaerobacter_sp._strain59 | (899) | <del>GCCTTATGGCTGATGTTATAAAACAGTTGACATG</del>                 |
| Tpseudoeothanolicus_39E         | (899) | <del>AACTTATCAAAAGTCTAATAAAGAGCTTATATTTAA</del>               |
| Tsaccharolyticum_B6ARI          | (632) | <del>GCCTTATGGCTGATGTTATAAAACAGTTGACATG</del>                 |
| Tsaccharolyticum_Y8485          | (899) | <del>GCCTTATGGCTGATGTTATAAAACAGTTGACATG</del>                 |
| Consensus                       | (901) | G GTTATGGCAGATGT ATAAAACAGTT GATATA                           |



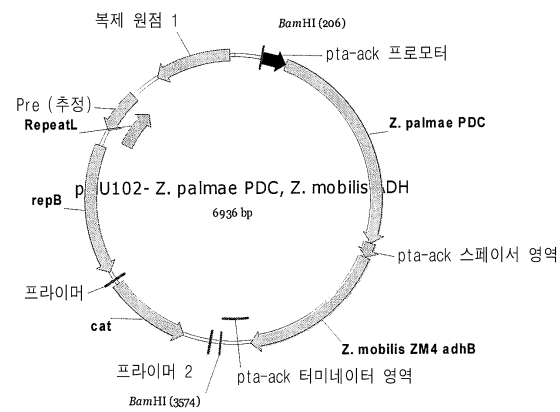
도면35



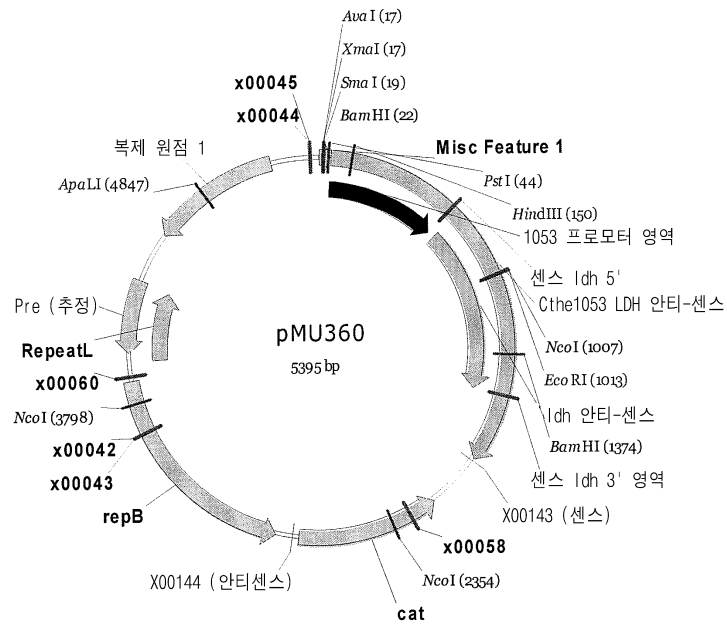
도면36



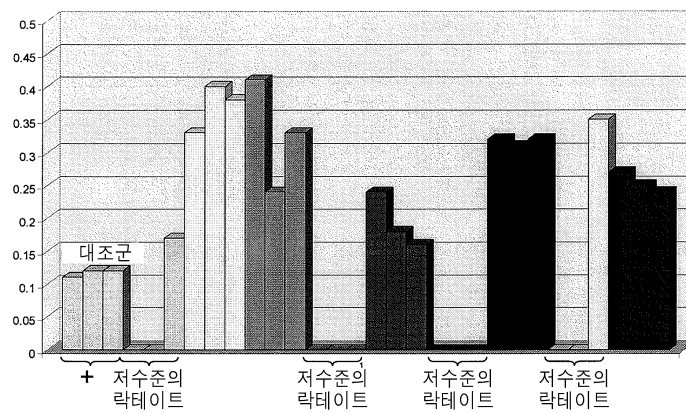
도면37



도면38

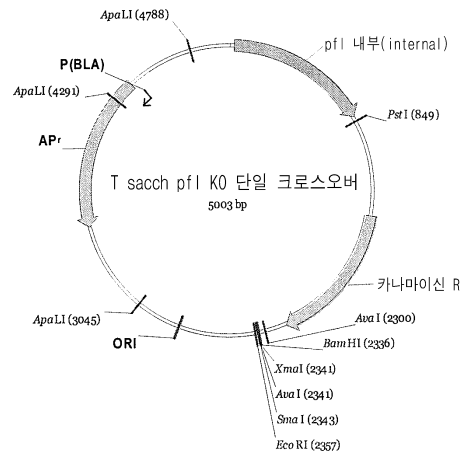


도면39

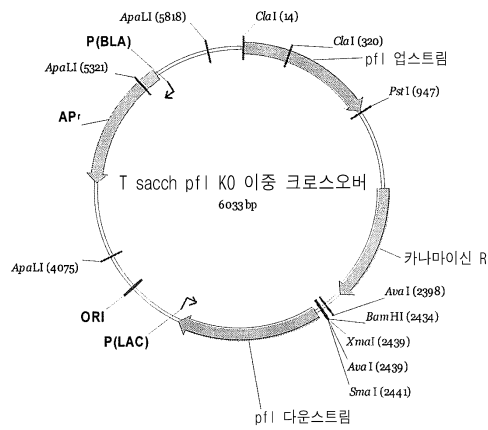




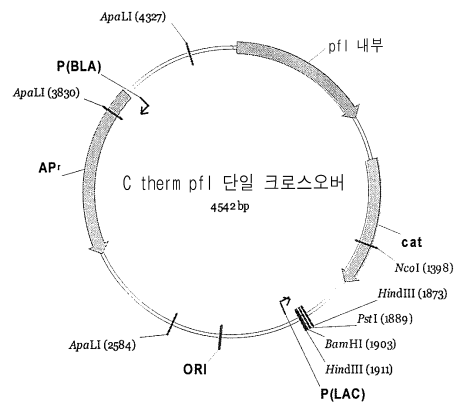
도면40



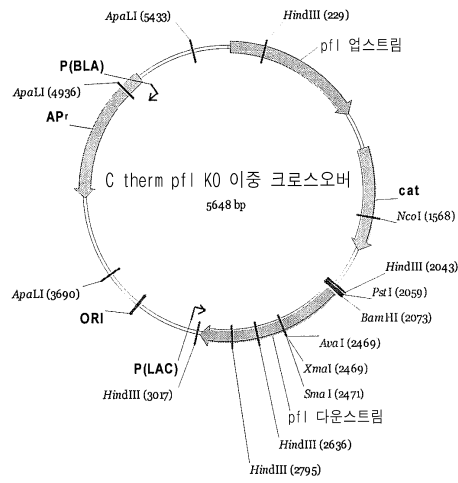
도면41



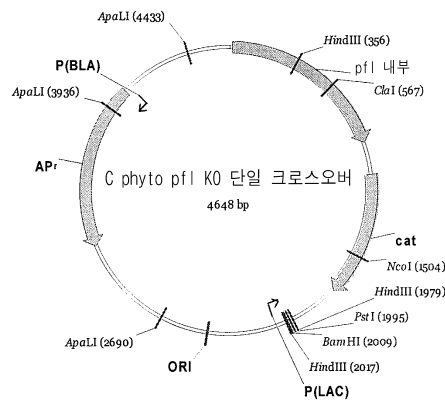
도면42



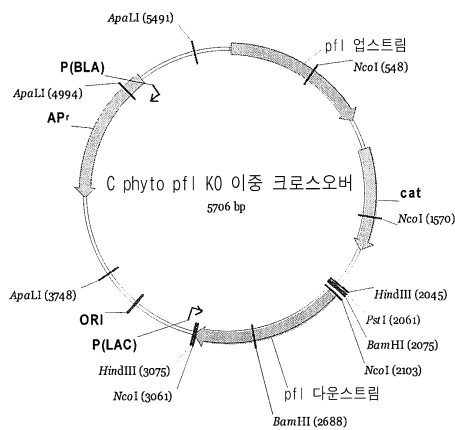
도면43



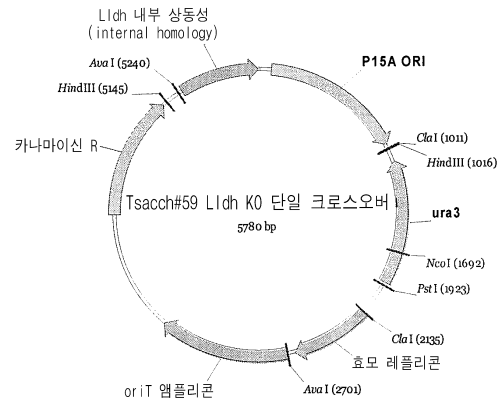
도면44



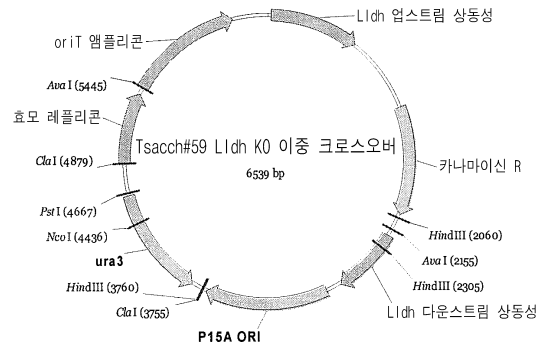
도면45



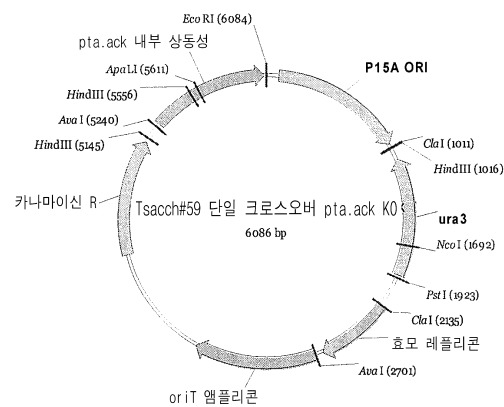
도면46



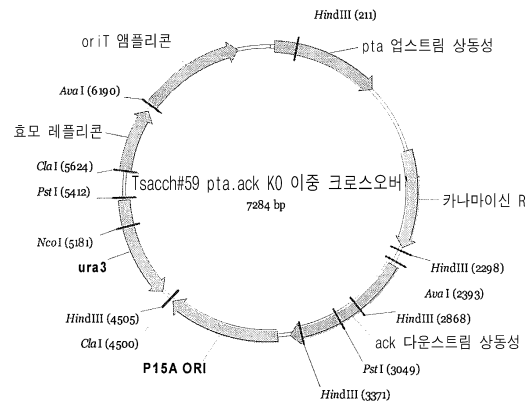
도면47



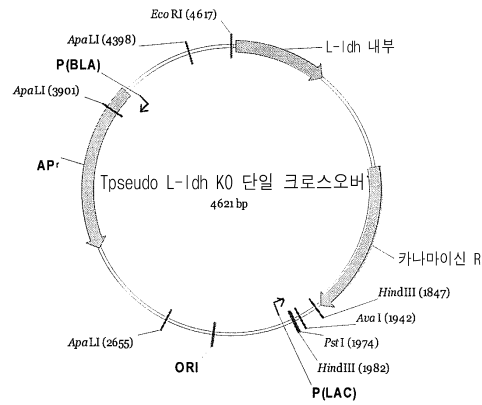
도면48



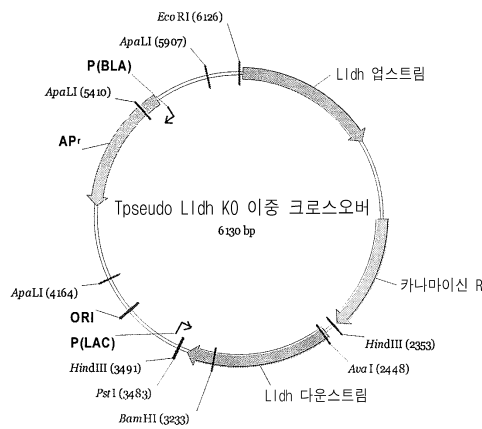
도면49



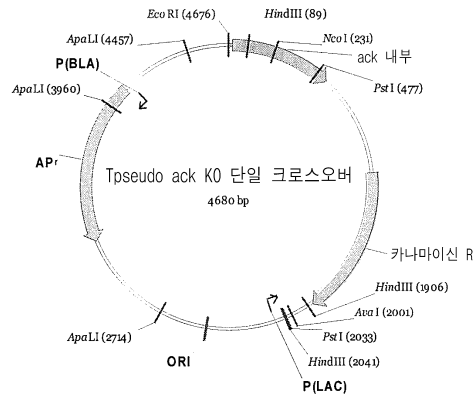
도면50



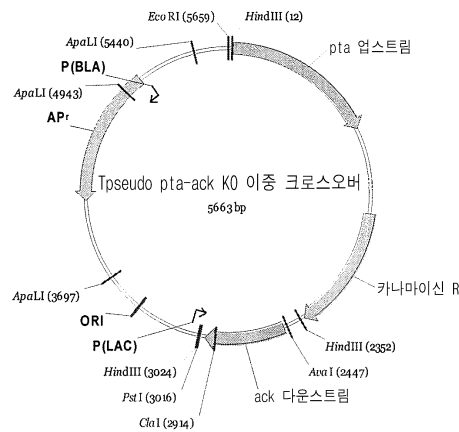
도면51



도면52



도면53



## 서열 목록

### SEQUENCE LISTING

<110> MASCOMA CORPORATION

<120> GENE KNOCKOUT MESOPHILIC AND THERMOPHILIC ORGANISMS, AND METHODS OF USE THEREOF

<130> MCX-011.25

<140> PCT/US08/63237

<141> 2008-05-09

<150> 60/916,978

<151> 2007-05-09



<160> 92

<170> PatentIn version 3.3

<210> 1

<211> 9000

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 1

```

agctttggct aacacacacg ccattccaac caatagtttt ctcggcataa agccatgctc      60

tgacgcttaa atgcactaat gccttaaaaa aacattaaag tctaacacac tagacttatt      120

tacttcgtaa ttaagtcgtt aaaccgtgtg ctctacgacc aaaagtataa aacctttaag      180

aaccttcttt tttcttgtaa aaaaagaaac tagataaatc tctcatatct tttattcaat      240

aatcgcatca gattgcagta taaatttaac gatcactcat catgttcata tttatcagag      300

ctcgtgctat aattatacta attttataag gaggaaaaaa taaagagggt tataatgaac      360

gagaaaaata taaaacacag tcaaaacttt attacttcaa aacataatat agataaaata      420

atgacaaata taagattaaa tgaacatgat aatatctttg aaatcggctc aggaaaaggg      480

cattttacc ttgaattagt acagagggtg aatttcgtaa ctgccattga aatagaccat      540

aaattatgca aaactacaga aaataaactt gttgatcacg ataatttcca agttttaaac      600

aaggatatat tgcagtttaa atttcctaaa aaccaatcct ataaaatatt tggtaatata      660

ccttataaca taagtacgga tataatacgc aaaattgttt ttgatagtat agctgatgag      720

atttatttaa tcgtggaata cgggtttgct aaaagattat taaatacaaa acgctcattg      780

```

gcattatittt taatggcaga agttgatatt tctatatataa gtatgggtcc aagagaatat 840

tttcatccta aacctaaagt gaatagctca cttatcagat taaatagaaa aaaatcaaga 900

atatcacaca aagataaaca gaagtataat tatttcgtta tgaaatgggt taacaaagaa 960

tacaagaaaa tatttacaaa aaatcaattt aacaattcct taaaacatgc aggaattgac 1020

gatttaaaca atattagctt tgaacaattc ttatctcttt tcaatagcta taaattattt 1080

aataagtaag ttaagggatg cataaactgc atccctaac ttgtttttcg tgtacctatt 1140

ttttgtgaat cgattatgtc ttttgcgcat tcacttcttt tctatataaa tatgagcgaa 1200

gcgaataagc gtcggaaaag cagcaaaaag tttccttttt gctgttggag catgggggtt 1260

caggggggtgc agtatctgac gtcaatgccg agcgaaagcg agccgaagg tagcatttac 1320

gttagataac cccctgatat gtcctgacgc ttatataga aaagaagatt caactaggta 1380

aaatcttaat ataggttgag atgataaggt ttataaggaa tttgtttgtt ctaatttttc 1440

actcatittt ttctaatttc ttttaacaaa tgttcttttt tttttagaac agttatgata 1500

tagttagaat agtttaaaat aaggagttag aaaaagatga aagaaagata tggaacagtc 1560

tataaaggct ctgagaggct catagacgaa gaaagtggag aagtcataga ggtagacaag 1620

ttataccgta acaaacgtc tggtaacttc gtaaaggcat atatagtca attaataagt 1680

atgttagata tgattggcgg aaaaaaactt aaaatcgta actatctct agataatgtc 1740

cacttaagta acaatacaat gatagctaca acaagagaaa tagcaaaagc tacaggaaca 1800

agtcacaaaa cagtaataac aacacttaaa atcttagaag aaggaaatat tataaaaaga 1860

aaaactggag tattaatgtt aaacctgaa ctactaatga gaggcgacga ccaaaaacaa 1920

aaatacctct tactcgaatt tgggaacttt gagcaagagg caaatgaaat agattgacct 1980

cccaataaca ccacgtagtt attgggaggt caatctatga aatgcgatta agcttggctg 2040  
 caggtcgata aaccagcga accatttgag gtgataggta agattatacc gaggtatgaa 2100  
 aacgagaatt ggacctttac agaattactic tatgaagcgc catatttaaa aagctaccaa 2160  
 gacgaagagg atgaagagga tgaggaggca gattgccttg aatatattga caatactgat 2220  
 aagataatat atcttttata tagaagatat cgccgtatgt aaggatttca gggggcaagg 2280  
 cataggcagc gcgcttatca atatatctat agaatgggca aagcataaaa acttgcattg 2340  
 actaatgctt gaaaccagg acaataacct tatagcttgt aaattctatc ataattgttg 2400  
 ttccaaaac ggctccgtcg atactatgtt atacccaac ttccaaaaca actttgaaaa 2460  
 agctgttttc tggatattaa ggttttagaa tgcaaggaac agtgaattgg agttcgtctt 2520  
 gttataatta gcttcttggg gtatctttaa atactgtaga aaagaggaag gaaataataa 2580  
 atggctaaaa tgagaatata accggaattg aaaaaactga tcgaaaaata ccgctgcgta 2640  
 aaagatacgg aaggaatgac tcctgctaag gtatataagc tggggggaga aaatgaaaac 2700  
 ctatatttaa aaatgacgga cagccggtat aaagggacca cctatgatgt ggaacgggaa 2760  
 aaggacatga tgctatggct ggaaggaaag ctgcctgttc caaaggctct gcactttgaa 2820  
 cggcatgatg gctggagcaa tctgctcatg agtgaggccg atggcgtcct ttgctcggaa 2880  
 gagtatgaag atgaacaaag ccctgaaaag attatcgagc tgtatgcgga gtgcatcagg 2940  
 ctcttttact ccacgacat atcggtattg ccctatacga atagcttaga cagccgctta 3000  
 gccgaattgg attacttact gaataacgat ctggccgatg tggattgcga aaactgggaa 3060  
 gaagacactc catttaaga tccgcgcgag ctgtatgatt ttttaagac ggaaaagccc 3120

gaagaggaac ttgtcttttc ccacggcgac ctgggagaca gcaacatctt tgtgaaagat 3180

ggcaaagtaa gtggctttat tgatcttggg agaagcggca gggcggacaa gtggtatgac 3240

attgccttct gcgtccggtc gatcagggag gatatcgggg aagaacagta tgtcgagcta 3300

tttttgact tactggggat caagcctgat tgggagaaaa taaaatatta tattttactg 3360

gatgaattgt ttagtacct agatttagat gtctaaaaag ctttttagac atctaacttt 3420

ttctgaagta catccgaac tgtccatact ctgatgtttt atatcttttc taaaagttcg 3480

ctagataggg gtcccagcg cctacgagga atttgatcg actctagagg atccctcagc 3540

gaagctccac tatgtttcaa aatgtcagat atatcaattt tcatcaaagt cacctcttaa 3600

aaccgacaag gactattata ctaactaata accctcatgt caagaattat atgacagatt 3660

ggcttaaata acaaaaaata ttttgtttag ttaaattcgg aatttcttct taatattatt 3720

aacatattcc acatattaat acaagaaaaa acccgcaaa aaaataaaaa aattttataa 3780

gcccgtttcc taaaaaaca ggcttgtaaa attataacgc atcttttata agttttttac 3840

aagtcttaaa gtctcccttg caatctcaag ctctctattt gtcgggataa ccaaagtctt 3900

tactttcgca tcgggagcac tgatatccgc ttccttgcct ttactttcat ttttatccaa 3960

atctatttta attccgaaaa agtcataatc cttcaaaact tctcttctta tataagcatt 4020

gttttcgccg atacctgcag tgaataccac cgcatacaag ccgttcagca ctgcaatata 4080

ttttccaata tatttcctaa caccatagca gaaaatatcc aatgccagct gcgccctgtc 4140

atctcccttt tctgcggcat cctgaacatc tctgaaatca ctgcttacac ctgaaattcc 4200

aagcacacct gatttcttgt taaggaaatt gtttatatcg ttaatattca ttttttctt 4260

ttcatcaaa taagttataa ccgcaggggtc aacattgccg cttctgtgtac ccatgcacaa 4320

cccctgcaga ggagtaaadc ccattgaggt gtcaacggat tttccgcctt ttaccgcaca 4380  
 aatacttgct ccgtttccaa gatggcaggt tatcagcttc aggetctcaa taggtttgcc 4440  
 cagcatctga gccgccctgt gggccacata tttgtgggaa gttccgtgga atccgtatct 4500  
 tctcaattta tactttctcat atatctcata agggagggca taaatatatg catgctagtt 4560  
 caacaaacgg gattgacttt taaaaagga ttgattctaa tgaagaaagc agacaagtaa 4620  
 gcctcctaaa ttcactttag ataaaaattt aggaggcata tcaaatgaac tttataaaaa 4680  
 ttgattttag caattggaag agaaaagaga tatttaataca ttatttgaac caacaaacga 4740  
 ctttttagtat aaccacagaa attgatatta gtgttttata ccgaaacata aaacaagaag 4800  
 gatataaatt ttaccctgca tttattttct tagtgacaag ggtgataaac tcaaatacag 4860  
 cttttagaac tggttacaat agcgacggag agttaggtta ttgggataag ttagagccac 4920  
 tttatacaat ttttgatggt gtatctaaaa cattctctgg tatttggact cctgtaaaga 4980  
 atgacttcaa agagttttat gatttatacc tttctgatgt agagaaatat aatgggttcgg 5040  
 ggaaattgtt tcccaaaaca cctatacctg aaaatgcttt ttctctttct attattccat 5100  
 ggacttcatt tactgggttt aacttaata tcaataataa tagtaattac cttctacca 5160  
 ttattacagc aggaaaattc attaataaag gtaattcaat atatttaccg ctatctttac 5220  
 aggtacatca ttctgtttgt gatggttatc atgcaggatt gtttatgaac tctattcagg 5280  
 aattgtcaga taggcctaatt gactggcttt tataacctga ggttttgctc caaccagcat 5340  
 ctcaaaagat ttggatgcag atattgcaat ttcagaaagc tggctctgcat ccggttttc 5400  
 caccaagccg caatcgcat atacaaaggt tccgttatga ccatattcac agttgggtac 5460



aaccataaca aaaaaggatg atacagatgt tgcceccggg gccgtcttta atatctgcaa 5520

agccggtctc aaagtatttg cagtgggaatt gacagcaccc gccaccatac catccgcttc 5580

accttttttt accatcataa ctccataata aagagggtct ttgatcggtt cctttgcggc 5640

ttctatagtc atacccttcg attttctaag ctcatcacgt gtatttgcat aatcctccaa 5700

tttttcggaa ttaaggaat cctctatcat cactccttca agatcaatat ccccccgcag 5760

actcttaate tccttttcat tgcctatcag tacaaccttt gcaattccct ttttcattat 5820

catggatgcg gctttaataa ccttcagatc cgtactttcc ggcaaaaacta tggtttttac 5880

gtctgatttc gccctttcaa ttatttggtc caaaaaactc ataaattctt ctcccttcat 5940

aatcccaaaa ctgttatcat aaaaactgta ttgtaatac ttataactat atattatcac 6000

caggtaataa tacctactca ctataaacag ctattttact gggttccaag caactctaat 6060

tatatacaaa atgttttttg tataacaacac cctccttate ttttttcgg ctttagccat 6120

aaataacggc aagtaactcc aaaatacagg atatttcatg ctttagaaa ctttttatta 6180

gtcttcttaa ttattcagat ttgtggcaa ttaaactttg cagctcctcc aaatagtgtg 6240

ccagctcttc ttctttaaga ttgctgagat atgacaatct gtaattttta gcctttttgg 6300

ccatctctag cgactctcc gtcattccca aatctttcaa aacacagcta tagttatagt 6360

atgcgaattc actggccgtc gttttacaac gtcgtgactg ggaaaaccct ggcgtttacc 6420

aacttaateg ccttgcagca catccccctt tcgccagctg gcgtaatagc gaagaggccc 6480

gcaccgatcg cctttcccaa cagttgcgca gcctgaatgg cgaatggcgc ctgatgcggt 6540

attttctect tacgcatctg tgcggtatgt cacaccgat atggtgcact ctcagtacaa 6600

tctgctctga tgccgatag ttaagccagc cccgacaccc gccaacaccc gctgacgcgc 6660

cctgacgggc ttgtctgctc ccggcatccg cttacagaca agctgtgacc gtctccggga 6720

gtgcatgtg tcagaggttt tcaccgtcat caccgaaacg cgcgagacga aagggcctcg 6780

tgatacgctt atttttatag gttaatgtca tgataataat ggtttcttag acgtcaggtg 6840

gcacttttcg gggaaatgtg cgcggaaccc ctatttgttt atttttctaa atacattcaa 6900

atatgtatcc gctcatgaga caataacctt gataaatgct tcaataatat taaaaagga 6960

agagtatgag tattcaacat ttccgtgtcg cccttattcc cttttttgcg gcattttgcc 7020

ttctgtttt tgctcaccca gaaacgctgg tgaaagtaaa agatgctgaa gatcagttgg 7080

gtgcacgagt gggttacatc gaactggatc tcaacagcgg taagatcctt gagagttttc 7140

gccccgaaga acgttttcca atgatgagca cttttaagt tctgctatgt ggcgcggtat 7200

tatcccgat tgacgccggg caagagcaac tcggtcgccg catacactat tctcagaatg 7260

acttggttga gtactcacca gtcacagaaa agcatcttac ggatggcatg acagtaagag 7320

aattatgcag tgctgccata accatgagtg ataacactgc ggccaactta ctctgacaa 7380

cgatcggagg accgaaggag ctaaccgctt ttttgacaa catgggggat catgtaactc 7440

gccttgatcg ttgggaaccg gagctgaatg aagccatacc aaacgacgag cgtgacacca 7500

cgatgcctgt agcaatggca acaacgttgc gaaaactatt aactggcgaa ctacttactc 7560

tagcttcccc gcaacaatta atagactgga tggaggcgga taaagttgca ggaccacttc 7620

tgcgctcggc cttccggct ggctggttta ttgctgataa atctggagcc ggtgagcgtg 7680

ggtctcgcg tatcattgca gcaactgggc cagatggtaa gccctccgt atcgtagtta 7740

tctacacgac ggggagtcag gcaactatgg atgaacgaaa tagacagatc gctgagatag 7800

gtgcctcact gattaagcat tggtaactgt cagaccaagt ttactcatat atactttaga 7860

ttgatttaaa acttcatttt taatttaaaa ggatctaggt gaagatcctt ttgataatc 7920

tcattgaccaa aatcccttaa cgtgagtttt cgttccactg agcgtcagac cccgtagaaa 7980

agatcaaagg atcttcttga gatccttttt tctgcgcgt aatctgctgc ttgcaaaaa 8040

aaaaaccacc gctaccagcg gtggtttgtt tgccgatca agagctacca actctttttc 8100

cgaaggtaac tggcttcagc agagcgcaga taccaaatac tgtccttcta gtgtagccgt 8160

agttaggcca ccacttcaag aactctgtag caccgcctac atacctcgt ctgctaattc 8220

tgttaccagt ggtcgtgcc agtgccgata agtcgtgtct taccgggttg gactcaagac 8280

gatagttacc ggataaggcg cagcggtcgg gctgaacggg gggttcgtgc acacagccca 8340

gcttggagcg aacgacctac accgaactga gatacctaca gcgtgagcta tgagaaagcg 8400

ccacgttcc cgaagggaga aaggcggaca ggtatccggt aagcggcagg gtcggaacag 8460

gagagcgcac gagggagctt ccagggggaa acgcttgga tctttatagt cctgtcgggt 8520

ttcgccacct ctgacttgag cgtcgatttt tgtgatgctc gtcagggggg cggagcctat 8580

ggaaaaacgc cagcaacgcg gcctttttac ggttcctggc cttttgctgg ccttttgctc 8640

acatgttctt tctgcgtta tcccctgatt ctgtggataa ccgtattacc gcctttgagt 8700

gagctgatac cgctcgccgc agccgaacga ccgagcgcag cgagtcagtg agcgaggaag 8760

cggaagagcg cccaatacgc aaaccgcctc tccccgcgcg ttggccgatt cattaatgca 8820

gctggcacga caggtttccc gactggaaaag cgggcagtga gcgcaacgca attaatgtga 8880

gttagctcac tcattaggca cccaggtt tacactttat gcttcggct cgtatgttgt 8940

gtggaattgt gagcggataa caatttcaca caggaaacag ctatgacct gattacgcca 9000

<210> 2  
 <211> 7582  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: Synthetic  
 polynucleotide

<400> 2  
 aattctgttc tggctttaga ccatttacgc ttgggtttg ccatgcatca cacctcctcg 60  
  
 tcaccataca acaaaccttg tgtataaata aaatcttgta cacatggaag gtactcgatt 120  
  
 ttgtttttac ttaaaacaaa gctacactaa ttgtctgtaa aaaagttttt gaaaacttcc 180  
  
 agccgcggat ctatatcttc atttacgcat ttgcactcat cgttgttaag atccgtcca 240  
  
 cacttcgggc aatatecctt gcaggcctcg tcacaaacct gctttgccgg aagattcagt 300  
  
 atgatattgt ctatcattac cttttccagc tcgaggaact taccatggta cgtataatat 360  
  
 tcctcgtcgg ttttgttact gccctcttct acaaagtfff cctttacatc aatatgcatc 420  
  
 tttgattcaa tatccttgag gcaccttgag catTTTgccc tgtaatccgc ccagagttca 480  
  
 ccgtcaagtt ttataatccc tccggcattt accaaagtgc ccttaaaagt taccggtttc 540  
  
 gcaaagtcaa aatcctcagc tataaaatca ttaattttaa ttgactcact aaagtccagt 600  
  
 ctcagcgaag ctccactatg tttcaaaatg tcagatatat caattttcat caaagtcacc 660  
  
 tcttaaaacc gacaaggact attatactaa ctaataaccc tcatgtcaag aattatatga 720  
  
 cagattggct taaataacaa aaataatfff gtttagttaa attcggaatt tcttcttaat 780  
  
 attattaaca tattccacat attaatacaa gaaaaaaccc ggcaaaaaaa taaaaaaatt 840

ttataagccc gtttcctaaa aaaacaggct tgtaaaatta taacgcatct ttataagtt 900

ttttacaagt cttaaagtc cccttgcaat ctcaagctcc tcatttgcg ggataaccaa 960

agtcctttact ttgcgcatcg gagcactgat atccgcttct ttgcctttca ctcatTTTT 1020

atccaaatct attttaattc cgaaaaagtc catatccttc aaaacttctc ttcttatata 1080

agcattgttt tcgccgatac ctgcagtga taccaccgca tcaacgccgt tcagcactgc 1140

aatatatTTT ccaatatatt tcctaacacc atagcagaaa atatccaatg ccagctgctg 1200

cagtaatcgc atcagattgc agtataaatt taacgatcac tcatcatgtt catatttate 1260

agagctcgtg ctataattat actaatttta taaggaggaa aaaataaaga gggttataat 1320

gaacgagaaa aatataaaac acagtcaaaa ctttattact tcaaaacata atatagataa 1380

aataatgaca aatataagat taaatgaaca tgataatate ttgaaatcg gctcaggaaa 1440

agggcatTTT acccttgaat tagtacagag gtgtaatttc gtaactgcc ttgaaataga 1500

ccataaatta tgcaaaacta cagaaaataa acttggtgat cagcataatt tccaagtttt 1560

aaacaaggat atattgcagt ttaaatttcc taaaaaccaa tcctataaaa tatttggtaa 1620

tataccttat aacataagta cggatataat acgcaaaatt gtttttgata gtatagctga 1680

tgagatttat ttaatcgtgg aatacgggtt tgctaaaaga ttattaaata caaacgctc 1740

attggcatta tttttaatgg cagaagttga tatttctata ttaagtatgg ttccaagaga 1800

atattttcat cctaaaccta aagtgaatag ctacttate agattaaata gaaaaaatc 1860

aagaatatca cacaaagata aacagaagta taattatttc gttatgaaat gggttaacaa 1920

agaatacaag aaaatatTTT caaaaaatca atttaacaat tccttaaac atgcaggaat 1980

tgacgattta aacaatatTA gctttgaaca attcttatct cttttcaata gctataaatt 2040

atttaataag taagttaagg gatgcataaa ctgcatccct tacagctgat acttttagtga 2100

tgagcttccg gtatttaataa ccaaaatatt catttcaaaa actcactccc gtcttgtttt 2160

ttttaatttt cctattccta aacttcgata aacagatgtt tttattaaac gctgcgcaac 2220

accttcttca atgtccggtt ttaacagaat ttatgccttg acatattgag cctgaaccgc 2280

agtaattgcc gcaaccccgga ctatatcttc ggcaactgcag cctcgtgaca gatcatttac 2340

cggctcttgc aaaccttgtg ttatcgggcc gtaagcttca gcttttgcca atctctgtgt 2400

aagcttgtat gcaatatttc cggcatcaag atccgggaaa ataagaacat tggcctttcc 2460

tgcaacactg cttccctttg ccttcgattt tgccacttcc ggaacaatgg cggcatccac 2520

ctgaagtctt cctgcaattg caaggtgggg agctttttcc tttgcaagct gtgttgcttt 2580

gattaccttt tcggtcagct cacttttggc actgccgtaa gaagaataag aaagcattgc 2640

cacctgaggt ttgctccaa ccagcatctc aaaagatttg gatgcagata ttgcaatttc 2700

agaaagctgg tctgcatccg gattttccac caagccgcaa tcggcatata caaaggttcc 2760

gttatgacca tattcacagt tgggtiacaac cataacaaaa aaggatgata cgagttttgt 2820

ccccggggcc gtctttaata tctgcaaagc cggcttcaaa gtatttgagc tggaattgac 2880

agcaccgcc accataccat ccgttcacc tttttttacc atcataactc cataataaag 2940

agggtctttg atcgtttccc ttgcggcttc tatagtcata ccttcgatt ttctaagctc 3000

atacagtgtt tttgcataat cctccaattt ttcggaattt aaggaatcct ctatcatcac 3060

tccttcaaga tcaatatccc ccgccagact cttaatctcc ttttcattgc ctatcagtac 3120

aacctttgca attccctttt tcattatcat ggatgcggct ttaataaccc tcagatccgt 3180



actttccggc aaaactatgg tttttacgtc tgatttcgcc ctttcaatta ttgtttccaa 3240

aaaactcata aattcttctc ctttcataat cccaaaactg ttatcataaa aactgtattt 3300

gtaatactta taactatata ttatcaccag gtaataatac ctactcacta taaacagcta 3360

ttttactggg ttccaagcaa ctctaggatc ctctagagtc gacctgcagg catgcaagct 3420

tggcgtaatc atggtcatag ctgtttcctg tgtgaaattg ttatccgctc acaattccac 3480

acaacatacg agccggaagc ataaagtga aagcctgggg tgcctaataa gtgagctaac 3540

tcacattaat tgcgttgccg tcaactgccc ctttcagtc gggaaacctg tcgtgccagc 3600

ccttcaaact tcccaaagcg gagccctagt gacattagaa aaccgactgt aaaaagtaca 3660

gtcggcatta tctcatatta taaaagccag tcattaggcc tatctgacaa ttctgaata 3720

gagttcataa acaatcctgc atgataacca tcacaaacag aatgatgtac ctgtaaagat 3780

agcggtaaat atattgaatt acctttatta atgaattttc ctgctgtaat aatgggtaga 3840

aggtaattac tattattatt gatatttaag ttaaaccag taaatgaagt ccatggaata 3900

atagaaagag aaaaagcatt ttcaggataa ggtgttttgg gaaacaattt cccgaacca 3960

ttatatattc ctacatcaga aaggataaaa tcataaaact cttgaagtc attctttaca 4020

ggagtccaaa taccagagaa tgttttagat acaccatcaa aaattgtata aagtggctct 4080

aacttatccc aataacctaa ctctccgtcg ctattgtaac cagtctaaa agctgtattt 4140

gagtttatca ccctgtcac taagaaaata aatgcagggt aaaatttata tccttcttgt 4200

tttatgtttc ggtataaaac actaatatca atttctgtgg ttatactaaa agtcgtttgt 4260

tggttcaaat aatgattaaa tatctctttt ctcttccaat tgtctaaac aattttatta 4320

aagttcattt gatatgcctc ctaaatTTTT atctaaagtg aatttaggag gcttacttgt 4380

ctgctttctt cattagaatc aatccttttt taaaagtcaa tcccgtttgt tgaactactc 4440

tttaataaaa taatttttcc gtccccaatt ccacattgca ataataaaaa atccatcttc 4500

atcggttttt tcgtcatcat ctgtatgaat caaatcgctt tcttctgtgt catcaaggtt 4560

taatttttta tgtattttct ttaacaaacc accataggag attaaccttt tacgggtgtaa 4620

accttctcc aaatcagaca aacgtttcaa attcttttct tcatcatcgg tcataaaatc 4680

cgatcctttt acaggatatt ttgcagtctt gtcaattgcc gattgtatat ccgatttata 4740

tttatttttc ggtcgaatca ttgaaacttt tacatttggg tcatagtcta atttcattgc 4800

ctttttccaa aattgaatcc attgtttttg attcacgtag ttttctgtat tcttaaaata 4860

agttggttcc acacatacca atacatgcat gtgctgatta taagaattat ctttattatt 4920

tattgtcact tccgttgca cgcataaaacc aacaagattt ttattaattt ttttatattg 4980

catcattcgg cgaatcctt gagccataatc tgacaaactc ttatttaatt cttcgccatc 5040

ataaacattt ttaactgtta atgtgagaaa caaccaacga actgttggtt ttgttttaat 5100

aacttcagca acaacctttt gtgactgaat gccatgttcc attgctctcc tccagttgca 5160

cattggacaa agcctggatt tacaaaacca cactcgatac aactttcttt cgctgtttc 5220

acgattttgt ttatactcta atatttcagc acaatctttt actctttcag cttttttaa 5280

ttcaagaata tgcagaagtt caaagtaatc aacattagcg attttctttt ctctccatgg 5340

tctcactttt ccactttttg tctgttccac taaaaccctt gatttttcat ctgaataaat 5400

gtactatta ggacacataa tattaaaaga aacccccatc tatttagtta ttgttttgt 5460

cattataac tttaacagat ggggtttttc tgtgcaacca attttaaggg ttttcaatac 5520

tttaaaacac atacatacca acacttcaac gcacctttca gcaactaaaa taaaaatgac 5580

gttatttcta tatgtatcaa gaatagaaag aactcgtttt tcgctacgct caaaacgcaa 5640

aaaaagcact cattcgagtg ctttttctta tcgctccaaa tcatgcgatt ttttctctt 5700

tgcttttctt tgctcacgaa gttctcgatc acgctgcaaa acatcttgaa gcgaaaaagt 5760

attcttcttt tcttccgacg gctcatgctg acgcacgaaa agccctctag gcgcatagga 5820

acaactccta aatgcatgtg aggggttttc tcgtccatgt gaacagtcgc atacgcaata 5880

ttttgtttcc catactgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta 5940

ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggtgcggc 6000

gagcggtatc agctcactca aaggcggtaa tacggttacc cacagaatca ggggataacg 6060

caggaaagaa catgtgagca aaagccagc aaaaggccag gaaccgtaaa aaggccgcgt 6120

tgctggcggtt ttcccatagg ctccgcccc ctgacgagca tcacaaaaat cgacgtcaa 6180

gtcagagggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct 6240

ccctcgtagc ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc 6300

cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg 6360

tcgttcgctc caagctgggc tgtgtgcacg aacccccgt tcagcccgac cgctgcgcct 6420

tatccggtaa ctatcgtctt gagtccaacc cggtagaca cgacttatcg ccaactggcag 6480

cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga 6540

agtgggggcc taactacggc tacactagaa gaacagtatt tggtatctgc gctctgctga 6600

agccagttac cttcgaaaa agagtggta gctcttgatc cggcaaaca accaccgtg 6660

gtagcggtgg ttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag 6720

aagatccttt gatcttttct acgggggtctg acgctcagtg gaacgaaaac tcacgttaag 6780

ggatttttgg caigagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat 6840

gaagttttta atcaatctaa agtatataig agtaaacttg gtcigacagt taccaatgct 6900

taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcattcata gttgcctgac 6960

tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa 7020

tgataccgcg agaccacgc tcaccggctc cagatttata agcaataaac cagccagccc 7080

gatatgggaa acaaaatatt gcgtatgca ctgttcacat ggacgagaaa acccctcaca 7140

tgcathtagg agttgttctt atgcgcctag agggcttttc gtgcgtcagc atgagcgatc 7200

ggaagaaaag aagaatactt tttcgttca agatgttttg cagcgtgac gagaacttcg 7260

tgagcaaaga aaagcaaaga ggaaaaatc gcatgatttg gagcgataag aaaaagcact 7320

cgaatgagtg ctttttttgc gttttgagcg tagcgaaaaa cgagtctttt ctattcttga 7380

tacatataga aataacgtca tttttatatt agttgctgaa aggtgcgttg aagtgttgg 7440

atgtatgtga ttcaataatt tcttttactc gctcggtata gtcgatcggt tcatcattca 7500

ccaaatcata attttcatgt gaccgttctt tatcaatata gggattcggt ttactttccc 7560

gttctctctg attgtgaaat tg 7582

<210> 3

<211> 8927

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 3  
agctttggct aacacacacg ccattccaac caatagtttt ctcggcataa agccatgctc 60  
  
tgacgcttaa atgcactaat gccttaaaaa aacattaaag tctaacacac tagacttatt 120  
  
tacttcgtaa ttaagtcgtt aaaccgtgtg ctctacgacc aaaagtataa aacctttaag 180  
  
aactttcttt tttcttgtaa aaaaagaaac tagataaatc tctcatatct tttattcaat 240  
  
aatcgcatca gattgcagta taaatttaac gatcactcat catgttcata tttatcagag 300  
  
ctcgtgctat aattatacta attttataag gaggaaaaaa taaagagggt tataatgaac 360  
  
gagaaaaata taaaacacag tcaaaacttt attacttcaa aacataatat agataaaata 420  
  
atgacaaata taagattaaa tgaacatgat aatatctttg aaatcggctc aggaaaaggg 480  
  
cattttacc c ttgaattagt acagagggtg aatttcgtaa ctgccattga aatagaccat 540  
  
aaattatgca aaactacaga aaataaactt gttgatcacg ataatttcca agttttaaac 600  
  
aaggatatat tgcagtttaa atttcctaaa aaccaatcct ataaaatatt tggtaatata 660  
  
ccttataaca taagtacgga tataatacgc aaaattgttt ttgatatgat agctgatgag 720  
  
at ttattttaa tctgtgaata cgggtttgct aaaagattat taaatacaaa acgtcattg 780  
  
gcattat ttt taatggcaga agttgatatt tctatat taa gtatggttcc aagagaatat 840  
  
tttcatccta aacctaaagt gaatagctca cttatcagat taaatagaaa aaaatcaaga 900  
  
atatcacaca aagataaaca gaagtataat tatttcgtta tgaaatgggt taacaaagaa 960  
  
tacaagaaaa tatttcaaaa aaatcaattt aacaattcct taaaacatgc aggaattgac 1020  
  
gatttaaca atattagctt tgaacaattc ttatctcttt tcaatagcta taaattattt 1080  
  
aataagtaag ttaagggatg cataaactgc atcccttaac ttgtttttcg tgtacctatt 1140

ttttgtgaat cgattatgtc ttttcgcat tcacttcttt tctatataaa tatgagcgaa 1200  
 gcgaataagc gtcggaaaag cagcaaaaag tttccttttt gctgttggag catgggggtt 1260  
 caggggggtgc agtatctgac gtcaatgccg agcgaaagcg agccgaaggg tagcatttac 1320  
 gttagataac cccctgatat gctccgacgc ttatataga aaagaagatt caactaggtta 1380  
 aaatcttaat ataggttgag atgataaggt ttataaggaa tttgtttgtt ctaatttttc 1440  
 actcattttg ttctaatttc ttttaacaaa tgttcttttt tttitagaac agttatgata 1500  
 tagttagaat agtttaaaat aaggagtgag aaaaagatga aagaaagata tggaacagtc 1560  
 tataaaggct ctgagaggt catagacgaa gaaagtggag aagtcataga ggtagacaag 1620  
 ttataccgta acaaacgtc tggttaacttc gtaaaggcat atatagtca attaataagt 1680  
 atgttagata tgattggcgg aaaaaactt aaaatcgta actatctct agataatgtc 1740  
 cacttaagta acaatacaat gatagctaca acaagagaaa tagcaaaagc tacaggaaca 1800  
 agtctacaaa cagtaataac aacacttaaa atcttagaag aaggaaatat tataaaaaga 1860  
 aaaactggag tattaatgtt aaacctgaa ctactaatga gaggcgacga ccaaaaacaa 1920  
 aaatacctct tactcgaatt tgggaacttt gagcaagagg caaatgaaat agattgacct 1980  
 cccaataaca ccacgtagtt attgggaggt caatctatga aatgcgatta agcttggctg 2040  
 caggtcgata aaccagcga accatttgag gtgataggta agattatacc gaggtatgaa 2100  
 aacgagaatt ggacctttac agaattactc tatgaagcgc catatttaaa aagctaccaa 2160  
 gagcaagagg atgaagagga tgaggaggca gattgccttg aatatattga caatactgat 2220  
 aagataatat atcttttata tagaagatat cgccgtatgt aaggatttca gggggcaagg 2280  
 cataggcagc gcgcttatca atatatctat agaatgggca aagcataaaa acttgcatgg 2340



actaatgctt gaaacccagg acaataacct tatagcttgt aaattctatc ataattgtgg 2400

tttcaaaatc ggctccgtcg atactatggt atacgccaac tttcaaaaca actttgaaaa 2460

agctgttttc tggatattaa ggtttttagaa tgcaaggaac agtgaattgg agttcgtctt 2520

gttataatta gcttcttggg gtatctttta atactgtaga aaagaggaag gaaataataa 2580

atggctaaaa tgagaatatc accggaattg aaaaaactga tcgaaaaata ccgctgcgta 2640

aaagatacgg aaggaatgtc tcctgctaag gtatataagc tggtagggaga aaatgaaaac 2700

ctatatttta aaatgacgga cagccggtat aaaggaccca cctatgatgt ggaacgggaa 2760

aaagacatga tgctatggct ggaaggaaag ctgcctgttc caaaggtcct gcactttgaa 2820

cggcatgatg gctggagcaa tctgtcatg agtgaggccg atggcgtcct ttgctcgaa 2880

gagtatgaag atgaacaaag ccctgaaaag attatcgagc tgtatgcgga gtgcatcagg 2940

ctctttcact ccacgacat atcggattgt ccctatacga atagcttaga cagccgctta 3000

gccgaattgg attacttact gaataacgat ctggccgatg tggattgcga aaactgggaa 3060

gaagacactc catttaaaga tccgcgcgag ctgtatgatt ttttaaagac ggaaaagccc 3120

gaagaggaac ttgtcttttc ccacggcgac ctgggagaca gcaacatctt tgtgaaagat 3180

ggcaaagtaa gtggctttat tgatcttggg agaagcggca gggcggacaa gtggtatgac 3240

attgccttct gcgtccggtc gatcaggag gatatcgggg aagaacagta tgtcgagcta 3300

ttttttgact tactggggat caagcctgat tgggagaaaa taaaatatta tattttactg 3360

gatgaattgt tttagtacct agatttagat gtctaaaaag ctttttagac atctaatttt 3420

ttctgaagta catccgcaac tgtccatact ctgatgtttt atatcttttc taaaagttag 3480

ctagataggg gtccccgagag ccccatactc atgagcagtc ttgttacagc tatgccgga 3540

gcacctgaac cgtttacaac aacttctata tctctgattt tcttgttgac aagctttaat 3600

gcattgatca ttgctgcaac agtaacaacg gctgtaccgt gctggtcac atggaatatt 3660

ggaatgtcac attcctcttt gagtcttctt tctatttcaa agcatctcgg agcggatata 3720

tcttcgaggt ttataccgcc aaagcttccg gagatgagct tgattgtctt tacaatttca 3780

tctacgtctt ttgatttgat acagagcgga aatgcgtcca catcaccaa cttcttgaag 3840

agtacgcatt taccttccat aacaggcatt ccggcttcag gtcctatgtc tccgagccct 3900

aaaaccgccg taccgtcggg aataaccgt accaggttcc aacgtcttgt atattcataa 3960

gaaagattaa catctttctg aattgcaaga catggttctg caacacccgg tgtataagca 4020

agcgacaact cttccttggg tgaacaggt accttgtgta taacctaat ttacccttc 4080

cactcacgt gaagccttag tgattctttt ctgtaacca tttgattcta cctccaaatt 4140

atattattaa atatctgca tattaatgca caattataaa ttcttaactt cgttcaatac 4200

ttttttaacc tgctccgtg agaactttaa agcttcttct tcttcaggag tcagattaaa 4260

ttggagaact tctgaacac cttcggaatt tacgatggat ggaaggctta ttgcaacatc 4320

ttctattcca tacatgccgt ttataacggg tctacgggt cttattgtat tctgattctt 4380

aaggagtgtt tcaactattg tgttgattga aactgcaata ccatagtatg ttgcaccttt 4440

gttcttgata atggttgca cgcagtttt aacatcttca gcgatttttt tcttgtcttc 4500

ttcttgaaa ttgcatttcg gatcatgat atattcgttg atatttttac cggcgatatg 4560

tgtgcagctc cacaacggaa gctgtgaatc accgtgttcg cctattatgt agccgtgtac 4620

atagttcaac aaacgggatt gactttttaa aaaggattga ttctaataa gaaagcagac 4680

aagtaagcct cctaaattca ctttagataa aaatttagga ggcatatcaa atgaacttta 4740

ataaaattga tttagacaat tggaagagaa aagagatatt taatcattat ttgaaccaac 4800

aaacgacttt tagtataacc acagaaattg atattagtgt tttataccga aacataaaac 4860

aagaaggata taaattttac cctgcattta ttttcttagt gacaagggtg ataaactcaa 4920

atacagcttt tagaactggt tacaatagcg acggagagtt aggttattgg gataagttag 4980

agccacttta tacaattttt gatggtgtat ctaaaacatt ctctggtatt tggactcctg 5040

taaagaatga cttcaaagag ttttatgatt tatacctttc tgatgtagag aaatataatg 5100

gttcggggaa attgtttccc aaaacaccta tacctgaaaa tgctttttct ctttctatta 5160

ttccatggac ttcatttact gggtttaact taaatatcaa taataatagt aattaccttc 5220

taccattat tacagcagga aaattcatta ataaaggtaa ttcaatatat ttaccgctat 5280

ctttacaggt acatcattct gtttgtgatg gttatcatgc aggattgttt atgaactcta 5340

ttcaggaatt gtcagatagg cctaatagact ggcttttata atgtacattt attggttaaca 5400

ttgtcttttg gggttttctt tcttatatcc gtctttgccg ccgcgggttc ggaaaaattt 5460

gaaatattgc ttgtttccct tcttttgttg gtacttttga taccttatat tgccattat 5520

tacaaactgg agaacggagt tcagaggctt tatgagcttt ataacaaaat tgatgaaaaa 5580

tgtgtaagga aaaacaagac cgcctgagtt ctcaccaga cggtcggtat tggcagtttc 5640

actttcgta gtcgatgttt ttcatgccgg caaagaaatt attttcttgc aagaaccttt 5700

ttcagtttg caaatcttgg aagaccatct tcgataggag gtcttgattc tcctgaatt 5760

aacggaagag catacttaat aaagtcttct gtaaggcctg ctccgtcagg tttaatccat 5820

tccaacggaa ctttcttctc agtatttgca acttcactga gggttcagaag ctgatattg 5880

cacttgtatt caggaccttc cgctctttca aaagcaacca ttagtctgt tttcccttca 5940

acggcatatt gtacggctgc ctgtcctgca agataagctt catttacgtc ggtaagagaa 6000

gctacgtgag ctgcgcatct ttggagaagg ctgaattcaa tgccgcgaac ctttgcgccg 6060

gtcttctctt taacaatgtt agccagtgtt gaagcaagac cgcccaactg tgcattgtcca 6120

aaggagtctt ttgttttcgc aaggctgaa ccgtattcgg aaatatattt tccgtttttg 6180

tctttgatac cttcagatac ggctacaata acctttccgt tttctttgta gattcttgc 6240

acatcttcaa caaatttgc tatgtcaaag gaaagctcgg gtaccgagct cgaattcact 6300

ggccgtcgtt ttacaacgtc gtgactggga aaaccctggc gttaccaac ttaatcgct 6360

tgcagacat ccccttttcg ccagctggcg taatagcgaa gaggccgca ccgatcgccc 6420

ttccaacag ttgcgcagcc tgaatggcga atggcgctg atgcggtatt ttctcttac 6480

gcatctgtgc ggtatttcac accgcataatg gtgcactctc agtacaatct gctctgatgc 6540

cgcatagtta agccagcccc gacacccgcc aacaccgct gacgcgcct gacgggcttg 6600

tctgtccccg gcatccgctt acagacaagc tgtgaccgtc tccgggagct gcatgtgtca 6660

gaggttttca ccgtcatcac cgaaacgcgc gagacgaaag ggctcgtga tacgcctatt 6720

tttatagggt aatgtcatga taataatggt ttcttagacg tcaggaggca cttttcgggg 6780

aaatgtgcgc ggaacccta ttgttttatt ttctaaata cattcaaata tgtatccgt 6840

catgagacaa taacctgat aaatgcttca ataatatga aaaaggaaga gtatgagtat 6900

tcaacatttc cgtgtcgecc ttattccctt ttttgcggca ttttgccttc ctgtttttgc 6960

tcaccagaa acgttggtga aagtaaaaga tctgaagat cagttgggtg cacgagtggg 7020

ttacatcgaa ctggatctca acagcggtaa gatccttgag agttttcgcc ccgaagaacg 7080

ttttccaatg atgagcactt ttaaagttct gctatgtggc gcggtattat cccgtattga 7140

cgccgggcaa gagcaactcg gtcgccgcat acactattct cagaatgact tggttgagta 7200

ctcaccagtc acagaaaagc atcttaccga tggcatgaca gtaagagaat tatgcagtgc 7260

tgccataacc atgagtgata aactgcggc caacttactt ctgacaacga tcggaggacc 7320

gaaggagcta accgcttttt tgcacaacat gggggatcat gtaactcgcc ttgatcgttg 7380

ggaaccggag ctgaatgaag ccatacaaaa cgacgagcgt gacaccacga tgcctgtagc 7440

aatggcaaca acgttgcga aactattaac tggcgaacta cttactctag cttcccggca 7500

acaattaata gactggaagg aggcggataa agttgcagga ccacttctgc gctcggccct 7560

tccggtggc tggtttatig ctgataaatc tggagccggt gagcgtgggt ctgcgggtat 7620

cattgcagca ctggggccag atggtaagcc ctcccgtatc gtagttatct acacgacggg 7680

gagtcaggca aciatggatg aacgaaatag acagatcgct gagatagggt cctcactgat 7740

taagcattgg taactgtcag accaagttaa ctcatatata ctttagattg atttaaaact 7800

tcatttttaa tttaaaagga tctaggtgaa gatccttttt gataatctca tgacaaaat 7860

cccttaacgt gagttttcgt tccactgagc gtcagacccc gtagaaaaga tcaaaggatc 7920

ttcttgagat ctttttttc tgcgcgtaat ctgctgcttg caaacaaaaa aaccaccgt 7980

accagcgggt gtttgtttgc cggatcaaga gctaccaact cttttccga aggttaactgg 8040

cttcagcaga gcgcagatac caaatactgt ctttctagt tagccgtagt taggccacca 8100

cttcaagaac tctgtagcac cgctacata cctcgctctg ctaatctgt taccagtggc 8160

tgctgccagt ggcgataagt cgtgtcttac cgggttggac tcaagacgat agttaccgga 8220

taaggcgag cggtcgggct gaacggggggg ttctgtcaca cagcccagct tggagcgaac 8280

gacctacacc gaactgagat acctacagcg tgagctatga gaaagcgcca cgcttcccga 8340

agggagaaag gcggacaggt atccggtaag cggcagggtc ggaacaggag agcgcacgag 8400

ggagcttcca gggggaaacg cctggtatct ttatagtcct gtcgggttcc gccacctctg 8460

acttgagcgt cgatttttgt gatgtctgtc aggggggcgg agcctatgga aaaacgccag 8520

caacgcggcc tttttacggt tcttggcctt ttgttgcct tttgtcaca tgtttttcc 8580

tgctttatcc cctgattctg tggataaccg tattaccgcc tttgagttag ctgataccgc 8640

tcgccgcagc cgaacgaccg agcgcagcga gtcagttagc gaggaagcgg aagagcgccc 8700

aatacgcaaa ccgcctctcc ccgcgcgttg gccgattcat taatgcagct ggcacgacag 8760

gtttcccgac tggaaagcgg gcagttagcg caacgcaatt aatgtgagtt agctcactca 8820

ttaggcaccc caggctttac actttatgct tccggctcgt atgttgtgtg gaattgtgag 8880

cgataacaa tttcacacag gaaacagcta tgaccatgat tacgcca 8927

<210> 4

<211> 7114

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 4

aattcagccc catactcatg agcagtcctt ttacagctat gccggcagca cctgaaccgt 60

ttacaacaac ttctatatcc tcgattttct tgttgacaag ctttaatgca ttgatcattg 120

ctgcaacagt aacaacggct gtaccgtgct ggatcatcatg gaataattgga atgtcacatt 180

cctctttgag tcttctttct atttcaaagc atctcggagc ggatataatc tcgaggttta 240

taccgcaaaa gcttccggag atgagcttga ttgtctttac aatttcattc acgtcttttg 300

atttgataca gagcggaaat gcgtccacat caccaaactt cttgaagagt acgcatttac 360

cttcataac aggcatctcg gcttcaggtc ctatgtctcc gagccctaaa accgccgtac 420

cgtcggtaat aaccgtacc aggttccaac gtcttgataa ttcataagaa agattaacat 480

ctttctgaat tgcaagacat ggttctgcaa caccgggtgt ataagcaagc gacaactctt 540

ccttggttga aacaggtacc ttgtgtataa cctcaatttt acccttcac tcaccgtgaa 600

gccttagtga ttctttctg taatccattt gattctacct ccaaattata ttattaaata 660

tctgcgatat taatgcacaa ttataaatc ttaacttctg tcaatacttt tttaacctgc 720

tccgtgaga atcttaaagc ttcttcttct tcaggagtca gattaaattg gagaacttcc 780

tgaacacctt cggaatttac gatggatgga aggccttattg caacatcttc tattccatac 840

atgccgttta taacggttcc tacggttctt attgtattct gattcttaag gagtgtttca 900

actattgtgt tgattgaaac tgcaatacca tagtatgttg cacctttgtt cttgataatg 960

gttgaccccg cagttttaac atcttcagcg attttttctt tgtcttcttc tgtgaaattg 1020

catttcggat catcgattaa tcgcatcaga ttgcagtata aatttaacga tcactcatca 1080

tgttcatatt tatcagagct cgtgtctataa ttataactat ttataagga ggaaaaata 1140

aagagggtta taatgaacga gaaaaatata aaacacagtc aaaactttat tacttcaaaa 1200

cataatatag ataaaataat gacaaatata agattaaatg aacatgataa tatctttgaa 1260

atcggtcag gaaaaggga ttttaccctt gaattagtag agagggtgaa tttcgtaact 1320



gccattgaaa tagaccataa attatgcaaa actacagaaa ataaacttgt tgatcacgat 1380

aatttccaag ttttaacaa ggatatattg cagtttaaat ttcctaaaaa ccaatcctat 1440

aaaatatttg gtaatatacc ttataacata agtacggata taatagcaa aattgttttt 1500

gatagtatag ctgatgagat ttatttaatc gtggaatagc ggtttgctaa aagattatta 1560

aatacaaac gtcattggc attattttta atggcagaag ttgatatttc tatattaagt 1620

atggttccaa gagaatattt tcctcctaaa cctaaagtga atagctcact taccagatta 1680

aatagaaaaa aatcaagaat atcacacaaa gataaacaga agtataatta tttcgttatg 1740

aaatgggtta acaaagaata caagaaaata ttacaaaaa atcaatttaa caattcctta 1800

aaacatgcag gaattgacga ttttaacaat attagctttg aacaattctt atctcttttc 1860

aatagctata aattatttaa taagtaagtt aagggatgca taaactgcat cccttaatcg 1920

atgagaacaa gticatttgc ggtttgccgc aaagccattg tgaaggctgc agacgcacct 1980

acaaaaccag caccaatgat tgcaactttt gacctacttt ttaccatttc cataccattc 2040

ctttcaatta ccagtatat ttaacggtta gticgtttat aaatttgaga ttaattcttt 2100

aaattttaac tggaaccgc gticacaggt attatcatta atttcagtat atgtgtttaa 2160

taaaaattag tgaattttg caactgcaag catttaaat tgtaaacgat aaataaatcc 2220

aggcaacaaa tttccccat ttttaaatagc ccagttaaac acattgataa cattttaaca 2280

ttattttata tctgcgtcca taactgaaaa agggaaatcc attactttat gaaatcaaat 2340

tttgaagtta tcaagaaatt atgacgattt tctccgtggc atgcaagatt tcgcgatatt 2400

tcattcgttt atattaatTT tttatgaaaa ctgcggtttg ggctgacaat tgcgatggaa 2460

gtttcaatta gactttttgt caaatattat gtataataat attatctata ataattgtatg 2520

aaaaaattgt cctaagatgg aagacggggg tggtttcata tatggttaaa tatctgaaga 2580

ggcaggaaga gttagtagaa gaagctctct caaaggataa ctgttctgac tgggaaagt 2640

tgagaaatta tcataagtc caaatgaat ttttcagca tgagagactt gtacattat 2700

tggtaacatt gtcttttggg tttttcttc ttatatccgt tcttgccgcc gcggtttcgg 2760

aaaaattga aatattgctt gtttccttc tttgttggg actttgata ctttatattg 2820

cccattatta caaactggag aacggagttc agaggcttta tgagctttat aacaaaattg 2880

atgaaaaatg tgtaaggaaa aacaagaccg cctgagtga tcctctagag tcgacctgca 2940

ggcatgcaag cttggcgtaa tcatggtcac agctgtttcc tgtgtgaaat tgttatccgc 3000

tcacaattcc acacaacata cgagccggaa gcataaagt taaagcctgg ggtgcctaat 3060

gagttagcta actcacatta attgcgttgc gctcactgcc cgctttccag tcgggaaacc 3120

tgctgtgcca gcccttcaaa ctcccaaag gcgagcccta gtgacattag aaaaccgact 3180

gtaaaaagta cagtcggcat tatctcatat tataaaagcc agtcattagg cctatctgac 3240

aattcctgaa tagagttcat aaacaatcct gcatgataac catcacaac agaattgatg 3300

acctgtaaag atagcggtaa atatattgaa ttacctttat taatgaattt tctgtctgta 3360

ataatgggta gaaggtaatt actattatta ttgatattta agttaaaccc agtaaatgaa 3420

gtccatggaa taatagaaag agaaaaagca ttttcaggta taggtgtttt gggaaacaat 3480

ttccccgaac cattatattt ctctacatca gaaaggtata aatcataaaa ctctttgaag 3540

tcattcttta caggagtcca aataccagag aatgttttag atacaccatc aaaaattgta 3600

taaagtggt ctaacttacc ccaataacct aactctccgt cgctattgta accagttcta 3660

aaagctgtat ttgagtttat cacccttgtc actaagaaaa taaatgcagg gtaaaattta 3720  
 tatecttctt gttttatgtt tcggtataaa acactaatat caatttctgt ggttatacta 3780  
 aaagtcgttt gtiggttcaa ataatgattia aatatctctt ttccttcca attgtctaaa 3840  
 tcaattttat taaagttcat ttgatatgcc tcctaaattt ttaactaaag tgaatttagg 3900  
 aggccttactt gtctgctttc ttcattagaa tcaatccttt tttaaaagtc aatcccgttt 3960  
 gtigaactac tctttaataa aataattttt cgttcccaa ttccacattg caataataga 4020  
 aaatccatct tcatcggttt ttcgtcatc atctgtatga atcaaatcgc cttcttctgt 4080  
 gtcacaaagg ttttaatttt tatgtatttc ttttaacaaa ccaccatagg agattaacct 4140  
 tttacggtgt aaaccttctt ccaaatcaga caaacgttcc aaattctttt ctccatcatc 4200  
 ggtcataaaa tccgtatcct ttacaggata ttttgtagtt tcgtcaattg ccgattgtat 4260  
 atccgattta tatttatttt tcggtcgaat catttgaact tttacatttg gatcatagtc 4320  
 taatttcatt gcctttttcc aaaattgaat ccattgtttt tgattcacgt agttttctgt 4380  
 attcttaaaa taagttaggt ccacacatac caatacatgc atgtgctgat tataagaatt 4440  
 atctttatta ttattgttca ctccggttgc acgcataaaa ccaacaagat ttttattaat 4500  
 tttttatat tgcattcttc ggcgaatcc ttgagccata tctgacaaac tcttatttaa 4560  
 ttcttcgcca tcataaacat ttttaactgt taatgtgaga aacaaccaac gaactgttgg 4620  
 cttttgttta ataacttcag caacaacctt ttgtgactga atgcatgtt tcattgctct 4680  
 cctccagttg cacattggac aaagcctgga tttacaaaac cactctgat acaactttct 4740  
 ttgcctgtt tcagattttt gtttatactc taatatttca gcacaatctt ttactcttcc 4800

agccttttta aattcaagaa tatgcagaag ttcaaagtaa tcaacattag cgattttctt 4860  
 ttctctccat ggctcactt ttccactttt tgccttgcc actaaaaccc ttgatttttc 4920  
 atctgaataa atgttactat taggcacat aatattaaaa gaaaccccca tctatttagt 4980  
 tatttgtttg gtcacttata actttaacag atggggtttt tctgtgcaac caattttaag 5040  
 ggttttcaat actttaaac acatacatac caacacttca acgcaccttt cagcaactaa 5100  
 aataaaaatg acgttatttc tatatgtatc aagaatagaa agaactcgtt ttctgctacg 5160  
 ctcaaaacgc aaaaaaagca ctcatcagag tgctttttct tctgctcca aatcatgcga 5220  
 tttttctctc ttgcttttc ttgctcagc aagtctcga tcacgtgca aaacatcttg 5280  
 aagcgaaaaa gtattcttct tttcttcga tcgctcatgc tgacgcacga aaagccctct 5340  
 aggcgcatag gaacaactcc taaatgatg tgagggttt tctcgtccat gtgaacagtc 5400  
 gcatacgcaa tattttgttt ccatactgc attaatgaat cggccaacgc gcggggagag 5460  
 gcggtttgcg tattgggcgc tcttcgctt cctcgtcac tgactcgtg cgctcggtcg 5520  
 ttcggtgcg gcgagcggtc tcagctcact caaaggcggc aatacggtta tccacagaat 5580  
 caggggataa cgcaggaaag aacatgtgag caaaaggcca gaaaaggcc aggaaccgta 5640  
 aaaaggccgc gtgctggcg tttttcata ggctccgcc cctgacgag catcacaaaa 5700  
 atcgacgtc aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc 5760  
 ccctggaag ctccctcgtg cgctctctg ttccgacct gccgcttacc ggatactgt 5820  
 ccgcctttct cccttcggga agcgtggcg tttctcatag ctacgctgt aggtatctca 5880  
 gtccgtgtga ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc gttcagcccg 5940  
 accgtgcgc cttatccgtt aactatcgtc ttgagtcaa cccgtaaga cagacttat 6000

cgccactggc agcagccact ggtaacagga ttagcagagc gagtatgta ggcggtgcta 6060

cagagttctt gaagtgggtg cctaactacg gctacactag aagaacagta tttggatatc 6120

gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac 6180

aaaccaccgc tggtagcggg ggtttttttg ttgcaagca gcagattacg cgcagaaaaa 6240

aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa 6300

actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt 6360

taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact tggcttgaca 6420

gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcaccca 6480

tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggttta ccatctggcc 6540

ccagtgtgc aatgataccg cgagaccac gtcacccggc tccagattta tcagcaataa 6600

accagccagc ccgatatggg aaacaaaata ttgcgtatgc gactgttcac atggacgaga 6660

aaaccctca catgcattta ggagtgttc ctatgcgcct agagggtttt tcgtgcgtca 6720

gcatgagcga tcggaagaaa agaagaatac ttttcgctt caagatgttt tgcagcgtga 6780

tcgagaactt cgtgagcaaa gaaaagcaaa gaggaaaaaa tcgcatgatt tggagcgata 6840

agaaaaagca ctggaatgag tgcttttttt gcgttttgag cgtagcgaaa aacgagttct 6900

ttctattctt gatacatata gaaataacgt catttttatt ttagttgctg aaaggtgcgt 6960

tgaagtgttg gtatgtatgt gattcaataa tttcttttac tcgctcgta tagtcgatcg 7020

gttcacatt caccaaatca taattttcat gtgaccgttc tttatcaata tcgggattcg 7080

ttttactttc ccgttctctc tgattgtgaa attg 7114

<210> 5  
 <211> 3932  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: Synthetic  
 polynucleotide

<400> 5  
 gacgaaaggg cctcgtgata cgcctatfff tataggftaa tgtcatgata ataattgfff 60  
 cttagacgtc aggtggcact ttfcggggaa atgtgcgcgg aaccctatt tgtttatfff 120  
 tctaaataca ttcaaatatg tatccgtca tgagacaata accctgataa atgcttcaat 180  
 aatattgaaa aaggaagagt atgagtattc aacatttcg tgtcgccctt attccctfff 240  
 ttgcggcatt ttgccttctt gtttttgctc acccagaaac gctggtgaaa gtaaaagatg 300  
 ctgaagatca gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga 360  
 tccttgagag ttttcgcccc gaagaacgtt ttccaatgat gagcactfff aaagtctgct 420  
 tatgtggcgc ggtattatcc cgtattgacg ccgggcaaga gcaactcggc gcccgcatat 480  
 actattctca gaatgacttg gttgagtact caccagtcac agaaaagcat ctacggatg 540  
 gcatgacagt aagagaatta tgcagtgtg ccataacat gatgataac actgcggcca 600  
 acttacttct gacaacgacg ggaggaccga aggagctaac cgcttttttg cacaacatgg 660  
 gggatcatgt aactcgctt gatcgttggg aaccggagct gaatgaagcc ataccaaagc 720  
 acgagcgtga caccacgatg cctgtagcaa tggcaacaac gttgcgcaaa ctattaactg 780  
 gcgaactact tactctagct tcccggaac aattaataga ctggatggag gcggataaag 840  
 ttgcaggacc acttctgcgc tcggcccttc cggctggctg gtttattgct gataaatctg 900

gagccggtga gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct 960

cccgtatcgt agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac 1020

agatcgctga gatagggtgcc tcactgattia agcattggta actgtcagac caagtttact 1080

cataatact ttagattgat ttaaaacttc atttttaatt taaaaggatc taggtgaaga 1140

tcctttttga taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt 1200

cagaccccg agaaaagatc aaaggatctt ctgagatcc ttttttctg cgcgtaatct 1260

gctgcttgca aacaaaaaaaa ccaccgtac cagcgggtgt ttgtttgccg gatcaagagc 1320

taccaactct tttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc 1380

ttctagtga gccgtagtta ggccaccact tcaagaactc tgtagcacgc cctacatacc 1440

tcgctctgct aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg 1500

ggttggactc aagacgatag ttaccggata aggcgcagcg gtcgggctga acggggggtt 1560

cgtgcacaca gccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg 1620

agctatgaga aagcgccacg ctccccgaag ggagaaaggc ggacaggtat ccggttaagcg 1680

gcagggtcgg aacaggagag cgcacagagg agcttccagg gggaaacgcc tggatatctt 1740

atagtcctgt cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtcag 1800

gggggaggag cctatggaaa aacgccagca acgcggcctt ttacggttc ctggcctttt 1860

gctggccttt tgctcacatg ttctttctg cgttatcccc tgattctgtg gataaccgta 1920

ttaccgcctt tgagttagct gataccgtc gccgcagccg aacgaccgag cgcagcgagt 1980

cagtgagcga ggaagcgga gagcgcccaa tacgcaaacc gcctctcccc gcgcgttggc 2040



cgattcatta atgcagctgg cagcacaggt ttcccgactg gaaagcgggc agtgagcgca 2100

acgcaattaa tgtgagttag ctactcatt aggcacccca ggctttacac tttatgcttc 2160

cggctcgtat gtgtgtgga attgtgagcg gataacaatt tcacacagga aacagctatg 2220

accatgatta cgccaagctt gtagttggig caggctttgt aggttccacc acagcttata 2280

cattgatgct cagcggactt atatctgaaa ttgtactgat agacataaat gcaaaaaaag 2340

ccgacggaga agtcatggac ttaaatacag gcatgccttt tgtaaggccc gttgaaattt 2400

atcgtggtga ctacaaagac tgtgccgat ccgacatagt aatcattacc gccggtgcca 2460

accaaaaaga aggcgaaacg agaatagatc ttagttcaac aaacgggatt gactttttaa 2520

aaaggattga ttctaataa gaaagcagac aagtaagcct cctaaattca ctttagataa 2580

aaatttagga ggcatatcaa atgaacttta ataaaattga tttagacaat tggaagagaa 2640

aagagatatt taatcattat ttgaaccaac aaacgacttt tagtataacc acagaaattg 2700

atattagtgt ttataaccga aacataaaac aagaaggata taaattttac cctgcattha 2760

ttttcttagt gacaagggtg ataaactcaa atacagcttt tagaactggt tacaatagcg 2820

acggagagtt aggttattgg gataagttag agccacttta tacaattttt gatgggtgat 2880

ctaaaacatt ctctggtatt tggactcctg taaagaatga cttcaaagag ttttatgatt 2940

tatacctttc tgatgtagag aaatataatg gttcggggaa attgtttccc aaaacaccta 3000

tacctgaaaa tgctttttct ctttctatta ttccatggac ttcattttact gggtttaact 3060

taaatatcaa taataatagt aattaccttc taccattat tacagcagga aaattcatta 3120

ataaaggtaa ttcaatatat ttaccgctat ctttacaggt acatcattct gtttgtgatg 3180

gttatcatgc aggattgttt atgaactcta ttcaggaatt gtcagatagg cctaatact 3240

ggcttttata atgtacatgc ttatatattt ggcgaaacacg gtgacaccga agttgcggcc 3300

tggagtccttg caaatattgc gggaaattccc atggatcgct actgtgacga atgccaatcag 3360

tgcgaggagc agatttcccg gaataaaaata tatgaaagtg ttaaaaatgc agcttatgaa 3420

atcatcagga acaaaggatgc aacctattat gccgtagccc ttgccgtaag aagaatcggt 3480

gaagccattg tactgcaggt cgactctaga ggatccccgg gtaccgagct cgaattcact 3540

ggccgtcgtt ttacaacgtc gtgactggga aaacctggc gttaccaac ttaatcgctt 3600

tgcagacat ccccttttcg ccagctggcg taatagcgaa gagggccgca ccgatcgccc 3660

ttccaacag ttgcgcagcc tgaatggcga atggcgctg atgcggtatt ttctccttac 3720

gcatctgtgc ggtatttcac accgcatatg gtgactctc agtacaatct gctctgatgc 3780

cgcatagtta agccagcccc gacacccgcc aacacccgct gacgcgcct gacgggcttg 3840

tctgtcccg gcatccgctt acagacaagc tgtgaccgtc tccgggagct gcatgtgtca 3900

gaggttttca ccgtcatcac cgaaacgcgc ga 3932

<210> 6  
 <211> 1320  
 <212> DNA  
 <213> Thermoanaerobacterium saccharolyticum

<400> 6  
 atgaataaat attttgagaa cgtatctaaa ataaaatatg aaggaccaa atcaaataat 60

ccttattcct ttaaatttta caatccagag gaagtaatcg atggcaagac gatggaggag 120

catctccgt tttctatagc ttattggcac acttttactg ctgatggaac agatcaattt 180

ggcaaggcta ctatgcaaag accatggaac cactacacag atcctatgga tatagcgaaa 240

cgaaggtag aagcagcatt tgagtttttt gataagataa atgcaccttt ctctctcttc 300

catgataggg atattgcccc tgaaggagat actcttagag agacaaacaa aaacttagat 360

acaatagttg ctatgataaa ggattactta aagaccagca agacaaaagt tttgtggggt 420

accgcaaatc ttttctccaa tccgagattt gtacatgggtg catcaacatc ctgcaatgct 480

gacgtttttg catattctgc agcgcaagtc aaaaaagccc ttgagattac taaggagctt 540

ggccgcgaaa actacgtatt ttgggggtgga agagaagggt acgagacgct tctcaatata 600

gatatggagt tagagcttga taactttgca agatttttgc acatggctgt tgactatgca 660

aaggaaatcg gctttgaagg tcagttcttg attgagccga agccaaagga gcctacaaaa 720

catcaatacg actttgacgt ggcaaatgta ttggcattct tgagaaaata cgaccttgac 780

aaatatttca aagtaaatat cgaagcaaac catgcgacat tggcattcca cgacttccaa 840

catgagctaa gatagccag aataaacggt gtattaggat caattgacgc aaatacaggc 900

gacatgcttt tgggatggga tacggaccag ttccctacag atatacgcat gacaacgctt 960

gctatgtatg aagtcataaa gatgggtgga ttgacaaaag gtggccttaa ctttgatgca 1020

aaagtaagac gtgcttcatt tgagccagaa gatcttttct taggtcacat agcaggaatg 1080

gatgcttttg caaaaggctt taaagttgct tacaagcttg tgaaagatgg cgtatttgac 1140

aagttcatcg aagaaagata cgcaagctac aaagaaggca ttggcgctga tattgtaagc 1200

ggtaaagctg acttcaagag ccttgaaaag tatgcattag agcacagcca gattgtaaac 1260

aaatcaggca gacaagagct attagaatca atcctaaatc agtattttgtt tgcagaataa 1320

<210> 7

<211> 1566

<212> DNA

<213> Thermoanaerobacterium saccharolyticum

<400> 7

|                                                                     |      |
|---------------------------------------------------------------------|------|
| atgaggcg cttcatgctt cattaaagct gccctcaaca aaaatcatgg aggtaaatgt     | 60   |
| atgtattttt tagggataga tttagggaca tcatcagtta agataatact gatgaatgaa   | 120  |
| agcggcaatg tggatatcaag cgtttcaaaa gaatatcctg tgtactatcc agagccaggc  | 180  |
| tgggctgagc aaaatccaga agattgggtgg aatggcaciaa gggatggaat aagagagatt | 240  |
| attgcgaaaa gcggcgtaaa tggcgaatgaa ataaagggtg ttggcttaag cgggcagatg  | 300  |
| catggactgg tgcttttaga caaagacaat aacgttttaa cgccagccat actttgggtg   | 360  |
| gaccagagga cacaggaaga atgcgactac atcacagaga aaataggaaa agaagcctt    | 420  |
| ttgaagtaca cagggaataa agcattgaca ggttttactg caccaaagat attatgggta   | 480  |
| aagaagcacc ttaaagacgt atatgaaaga atcgctcata tccitttgcc aaaagattat   | 540  |
| ataaggttta aattgacagg tgagtacgct acagaagttt cagatgcac aggtacactt    | 600  |
| cttttcgatg tggaaaatag aagatgggtca aaggaaatga tagacatatt tgaaataccg  | 660  |
| gaaaaagccc ttctaagtg ctacgaatca acagatgtca cagggtatgt caccaaagag    | 720  |
| gcagcagatt tgacagggtc tcatgaaggg actattgtcg taggcggtgg tggtagcaaa   | 780  |
| gccagcgcg ctgtaggcac tggcacggtg aaaagcggca tagtgtccat cgcattagga    | 840  |
| acttcaggcg tcgtatttgc atcacaggac aagtacgcag cagatgatga gcttaggctt   | 900  |
| cactcattct gccatgcaaa cggcaaatgg catgtgatgg gtgtcatgct ttcggctgca   | 960  |
| tcatgtctta aatgggtgggt agatgatgta aataattaca agaccgatgt tatgacattt  | 1020 |
| gatggactct tagaagaagc agagaagggtg aagccaggca gtgatggatt gatattcttg  | 1080 |

ccatacctga tgggtgaaag gaccccttac agcgatcctt atgcgagagg cagctttgta 1140

ggtttaacaa ttacacacaa tagaagccac atgacaagat ctatattaga aggcgtcgca 1200

tttggactta gggattcgct ggagcttata aaggctttaa atatacctgt aaatgaagcc 1260

agggttaagt gtggtgggtgc taaaagcagg ctttggaggc aaatacttgc cgatgtattc 1320

aatgtaagga tagacatgat aaatgctaca gaaggacctt catttgggtgc agcaataatg 1380

gcgtctgtgg gatatggcct ttacaaaaat gtagatgatg catgcaatag ttttaataaaa 1440

gttacagaca gcgtatatcc aatcaaagaa aacgtcgaaa agtacaacaa actgtatcca 1500

atctacgtga gcttgtattc aaggcttaaa ggcgcccttg aagaaattgg gaagttggat 1560

ttgtaa 1566

<210> 8  
 <211> 1407  
 <212> DNA  
 <213> Thermoanaerobacterium saccharolyticum

<400> 8  
 atgattattg tgtacaaaga tgaaaagccc aggataggtt ttttgggtat tatgcaggag 60

ttatacgatg atatgttgcc tggattact gaaaggcaag agatgtatgc acaacaggtt 120

ataggtagat taggtgatgt tgctgatttt tatttcccag gtgctgcaaa aaacagaaat 180

gatatagaaa ggatagttaa ggaattcaac gataaagatc ttgacggaat aatgatcgtg 240

atgctgacat acggaccagc cacaaatctt gtgaatgctt taagaaacaa taggcttccg 300

attatgctgg cgaatataca gccagaaagc actgtgacag acgattggga tatgggggac 360

ttgacctaca accaaggtgt tcatgggtgca caggatactt caaatattat tctgagaatg 420

ggcataactt gtctgttat aacagaagat tggcattctg atgaatttaa agattttgtg 480

aatgattggg caaaaactgt aaagacagta aaagctttga ggaatatgaa gatagcacia 540

tttgaagaa tgcattggtat gtatgacata atgggtgatg atgcagcttt tacaagaaaa 600

ttggggccgc aaataaacca ggagtacatt ggccaagttt ttagatatat ggaagaagct 660

acaaatgaag aaattgacaa agtgatagag gaaaacaaga agaactttta tatagatcct 720

aaattaagtg atgagagcca cagatatgct gcaaggcttc aaataggatt taagaaattg 780

cttgaggaga aagggtactc tggctttagt gctcactttg atgtgtttta aggcatgga 840

agatttaagc agatacacat gatggcagca tcaaaactga tggcagaagg atatggctat 900

gcggcagagg gcatgtagt tacggcaagc ctggtggcag caggatcatgt ttgataggc 960

aatgcacact ttaccgatg gtatgcgatg gattttaaga gagattcaat ttgatgagc 1020

cacatgggag agggcaattg gaagatagcc agaaaagata gacctataaa attatcgat 1080

cgagagcttg gcataggaat gcttgataat cctccaacag tgggttttat ggctcaacca 1140

ggcattgcga cattggcatc attagtgtct ttagaaggcg aaaaatatag acttgttgtt 1200

tcaaaggag aaattttaga tacagaagaa gcgaaaaata tagagatgcc gtatttccat 1260

tttagacctg aaaacggagt tagggcttgt ctaaattggct ggcttaaaaa tgggtgtaca 1320

catcatcagt gcttgacatt aggtgatgct actaaaagat ggaagctttt atgcgaatta 1380

ttagatatcg agtatgttga agtgtaa 1407

<210> 9

<211> 642

<212> DNA

<213> Thermoanaerobacterium saccharolyticum

<400> 9

atgttagaga acctaaaaca acgtgtatat aaaatgaaca tgatgcttcc taaaaacaat 60

ttagtcacaa tgacaagcgg caatgtcagc ggaagagatc ctgagacaaa tctttagtc 120

ataaagccca gcggagtttt gtacgatgaa atgacgccag atgatatggt agtcgtggat 180

ttggatggca atgtggttga gggtaagcta aaacctctg tcgatactgc tacacatctt 240

tacgtctaca ggcatagaaa tgatgtaaac ggcatgttcc atacacactc accgtatgct 300

acaagttttg ccgcacttgg ccggtcaatt ccggtctatc ttacagctat tgcagacgag 360

tttggatgcg caattcctgt agggccttat gccaaaattg gcggggaaga gataggaaaa 420

gccatcgtag attatatagg tgagagtcct gcaatactta tgaaaaatca cggcgttttt 480

accattggca attcacctga agcagcctta aaagctgctg ttatggtaga agatacagct 540

aagacggtgc acttatcact gcttttaggc acacctgatg taataccaga tgaagaagta 600

aaaagagccc atgaaagata tcttacaaaa tacggtcaat ga 642

<210> 10  
 <211> 1320  
 <212> DNA  
 <213> Clostridium cellulolyticum

<400> 10

atgtcagaag tatttagcgg tatttcaaac attaaatttg aaggaagcgg gtcagataat 60

ccattagctt ttaagtacta tgaccctaag gcagttatcg gcggaaagac aatggaagaa 120

catctgagat tcgcagttgc ctactggcat acttttgcag caccaggtgc tgacatgttc 180

gggtgcaggat catatgtaag accttggaaat acaatgtccg atcctctgga aattgcaaaa 240

tacaaagttg aagcaaactt tgaattcatt gaaaagctgg gagcaccttt cttegttttc 300

catgacaggg atattgtcc tgaaggcgac aactcgtcg aaacaataa aaaccttgat 360

acaatagttt cagtaattaa agatagaatg aaatccagtc cggtaaagt attatgggga 420

actacaaatg ctttcgga aaagattt atgcatggg catcaacttc gccaaacgct 480

gacatatttg cgtatgcagc agctcaggtt aaaaaggcaa tggaaatcac aaaggaatta 540

ggcggagaaa actatgtatt ctgggggtgt agagaaggtt atgaaactct cttgaataca 600

gacatgaagc tggaaactga taatttagca agattcttga agatggctgt tgactatgct 660

aaggaaatcg gttttgacgg acaattccta atcgaaccaa agccaaaaga accaactaag 720

caccaatatg attttgatac agctacagtt atcggcttcc tgaagacata tggattagac 780

ccatacttca agatgaatat cgaagctaac catgctacat tagcaggaca cacattccaa 840

catgagcttg ctatgtgcag aatcaacgac atgcttggaa gtattgatgc taaccaaggt 900

gatgtaatgc tcggatggga tacagaccaa ttcccaacga acctatatga tgcaacacta 960

gcaatggtgg aagtattaaa ggccggcgga ttgaaaaagg gaggtttgaa cttcgactca 1020

aaagttagaa gaggatcatt cgaaccatca gacttgttct atggacatat tgcaggtatg 1080

gatacttttg caaagggtct tatcatagca aataagatcg ttgaggacgg taagtttgat 1140

gcatttggtg ctgacagata ctcaagctac acaaatggta tcggaaaaga tattgttgaa 1200

ggaaaagtgt gctttaagga attggagcaa tatgactta ctgcaaagat tcagaacaag 1260

tctggacgtc aggaaatgct ggaagctttg ttaaaccagt atatcctga aacaaaataa 1320

<210> 11

<211> 1608

<212> DNA

<213> Clostridium cellulolyticum

<400> 11



|                                                                    |      |
|--------------------------------------------------------------------|------|
| atgaagcatg aactaaatga cgggagaaat gctattctaa atggaaagac agcaattggg  | 60   |
| attgaactcg gatcaactag aataaaaacg gtattgatag gtgcagacaa tgcacctatc  | 120  |
| gcatccggta gtcatgactg ggaaaacagc tatatcaata atatttggac ttacagcttg  | 180  |
| gaagatatct ggaaaggcgt acagagcagc taticagaaa tggttaaaga tgtagggac   | 240  |
| aaattcggag taagtctaaa gacaaccgga gcaatagggt ttagcggaat gatgcacggt  | 300  |
| tatatggttt ttgataagga aggtaatctt ctgactcatt tcagaacatg gcgtaacact  | 360  |
| ataactgcac aggcttccga ggaactaacc aagttgttta attatcctat tcctcaaagg  | 420  |
| tggagcattg cccatcttta ccaagccata ctgaacaatg aagagcatgt atccaatatc  | 480  |
| gattttatga ctacattggc cggatttata cactggaagt tgacaggaga aaaagtctt   | 540  |
| ggtgtcggag aggcatcagg tgttttccca atagatttag atactaagga ttttaattca  | 600  |
| agtatgatta atcagtttaa tgaggtacc accaatcgaa attttcatg gaagcttcaa    | 660  |
| aatattcttc caaaagtttt ggtttcgggt actgaagcag gtaggctgac agaagaaggt  | 720  |
| gcaaagcttc ttgatgttac cggggagctt caggcgggta ttcctttttg tccccctgag  | 780  |
| ggagatgcgg gaaccggtat gggtgcaact aacagcgttg ctgtccgtac aggcaatgtg  | 840  |
| tctgccggga ctctgtttt tgctatggtt gttctcgaaa aggaattatc caaagtgtat   | 900  |
| tcggaaattg acctggtgac tacacctgac gcaaactctg tggctatggt tcattcaa    | 960  |
| aattgtacat cggactatga cgcatggaig ggtatatattg ctgaggcagt taagaccttg | 1020 |
| ggctttgacg tgaaaaaacc acagctatat gataccctgc tgggagccgc acttcaaggt  | 1080 |
| gacctgatt gcggagggtt gcttgcgtac ggttatattt caggtagca tattacccat    | 1140 |
| tttgaagaag gtcgccgat ggttgttcgt tcatcaaca gcaaattcaa cctggccaac    | 1200 |

tttatcaggg tcaatttggt tacatctctt ggagccttga agaccggttt ggatattctt 1260

tttcaaaagg aagctgttaa agtggacggt attaccggac acggcgggtt cttaagacg 1320

aaggaagtag gacagaagat tatggcggct gcctttaatg tccctgtatc tgttatgaag 1380

actgcgggtg aaggcgggtc atgggggtatt gcctacttg ctctgtatat gattaatagg 1440

gaaagctcac agtccttga ggattttctt aaacaaatg tgtttgggga aagccaaggt 1500

gagactgtac agccagattc gaaggatgtt gacggtttca acgagtttat gaaaaggtag 1560

acaaagggac tgggtattga aagggtcgcg ataaacttct tgaactga 1608

<210> 12  
 <211> 1380  
 <212> DNA  
 <213> Clostridium cellulolyticum

<400> 12  
 atgataacca aacaaaaacc aagaatcgga tttttgggcc taatgcaggg attgtatgac 60

gaatcacagc cggaactgcc gaaaatgcag gaggcatttg ccagagaagt ggttgaacaa 120

ttaaaagatg tggcagatat tgattttccc ggtccagcaa aagaaagaga agatatagaa 180

agatatgtaa aatatttcaa tgataaagag tacgatggaa taatgatagt aaatctgttg 240

tacagtccgg gaaatcggtt aatacaggct atgaagaata ataatctgcc aatattgctg 300

gctaatatcc aaccacttcc cgatgttaca tcaaactggg attggatttt gtgcacaact 360

aatcagggaa ttcatggaat acaggataca agtaatgttc tcatgcgttg tggattataa 420

ccggctatta taacagatga ttggaaggct gaatccttta aagcctactt tgaagattgg 480

gcattggctg ccaacacgca taacagacta aaaaagacaa aggttgcgat tttcgccgt 540

atgcacaata tgggtgacat acttggatgat gatgcggcat tgtgcagaaa atttgggtga 600

gaggcaaacc atgtaacaat cggtcgggtt tattacaaca tggaaggatt gtcagataaa 660

gaagtagatg cccagattga ggaagataaa aagaatttta aaattgatcc taatcttcct 720

gaagaaagtc atcggtatgc tgcacgtatg caattagcct ttgaaaaatt ccttaatgat 780

aacggttatg aaggtttttc acagtcttct aacatatata aggaagacgg cagggtcaaa 840

caaataccga tattggcagg ctccagtctc ctgcaaaag gttatggta ttcggcggaa 900

ggtgatacaa atgtacttct catgactgtg atcggtcaca tgatgatagg ggatcctcat 960

tttactgaga tgtactccct ggactttggt aaggattcag caatgctaag ccatatggga 1020

gaaggcaact ggaaggttgc aaggaaggat cgcggagtga cactgattga caggcctctt 1080

gatattggtg gtcttggtaa tcctccgaca ccaaagtta acgtagaacc aggaacagct 1140

acccttgttt cctcgttgc agtagaagga gaaaaatacc aactaattgt atcaaagggt 1200

actatccttg atactgagga ctgccagat gttcctatga accatgcttt tttcagaccg 1260

gattccggca tcaaaaaggc tatggacgaa tggtttagcta atggtggtac acatcacgaa 1320

gtactattcc tgggtgattt tagaagacgt tttgaattat tatgtaaatt ctgacataa 1380

<210> 13  
 <211> 690  
 <212> DNA  
 <213> Clostridium cellulolyticum

<400> 13  
 atgttggaa aactaaaaca agcgggtgtg gaagccaatc tagagctgcc tgaaaaagga 60

cttgtaacat atacatgggg aaatgtaagc ggtatcgaca gagaaagcag acttattgca 120

attaaacca gtggtgttga gtataatgtt atgacagctg atgatattgt attaatacgac 180

cttacaggta aagtgggtga aggaaaattg aagccgtctt ctgatgcacc aacacatgta 240

gtctctgata atgcatttcc tgatatagga ggtgtaacac acacccattc caggtgggca 300

actgcttttg cacaggctgg tatggggatt cctgcttacg ggactactca tgcggattac 360

ttttatgggtg aaatcccatg tactcgggaa atgacaaagg atgagattga gtccgattat 420

gaagcaaata ccggaacggt gataatagag acttttaaag atttaaatcc taactatatc 480

cctgccgtac ttgtaaaaaa tcatgcacct ttacatggg gaaaaagtgc agcggaatcg 540

gttcataatt ctgttgtttt agaagaagta gctatgatgg ctattcagtg cagacaactg 600

aacccaaatg taactcccat gccgcagggtg ctgctagaca agcattttat gaggaagcac 660

ggcccgaag cttattacgg acaaaaataa 690

<210> 14  
 <211> 1317  
 <212> DNA  
 <213> Clostridium phytofermentans

<400> 14  
 atgaaaaatt actttccaaa tgttcagaa gtaaaatag aaggcccaa ttcaacgaat 60

ccatttgctt ttaaataatta tgacgcaa ataaagttag cgggtaaaac aatgaaagag 120

cactgtcgtt ttgcattatc ttggtggcat actctttgtg caggtggtgc tgatccattc 180

ggtgtaacaa ctatggatag aacctacgga aatcacag atccaatgga acttgctaag 240

gcaaaagttg acgtgggtt cgaattaatg actaaattag gaattgaatt ctctgtttc 300

catgacgcag atattgctcc agaaggtag acttttgaag agtcaaagaa gaatctttt 360

gaaatcggtg attacatcaa agagaagatg gatcagactg gtatcaagtt attatggggt 420

actgctaata actttagtca tccaagattt atgcatgggtg cttccacatc ttgcaacgca 480

gacgtatttg catatgctgc tgctaagatt aagaatgcat tagatgcaac aattaaatta 540

ggcggtaaag gttatgtatt ctgggggtgggt cgtgaagggtt atgaaacact tcttaataca 600

gatttaggac ttgagcttga taatatggct agacttatga agatggctgt agagtatggc 660

cgtgcaaatg gttttgatgg cgacttctat attgagccaa agccaaagga accaaccaag 720

catcaatatg attttgatac agcaaccgta cttgctttcc ttcgcaaata tggcttagaa 780

aaagatttca agatgaacat tgaagcaaac catgctactc ttgcaggtca tacctttgaa 840

catgaacttg caatggctag agttaatggt gcatttgggt ctgtagatgc aaaccagggt 900

gatccaaacc ttggatggga tacggatcaa ttcccaactg atgttcatag tgcaactctt 960

gcaatgcttg aagtacttaa ggctgggtgga ttcactaacg gcggacttaa ctttgatgca 1020

aaggtaaagac gtggttcctt cgaatttgat gatattgcat acggttatat tgcaggaatg 1080

gatacttttg cacttgggtt aattaaggct gctgagatta tgcacgatgg tagaatcgca 1140

aaattttag atgatcgta tgcaagctat aaaacaggaa ttggtaaagc aattgtggat 1200

ggaactacat ctcttgaaga attagagcag tatgttttaa cacatagtga accagtaatg 1260

cagagtggtc gtcaggaagt tcttgaaaca atcgtaaata atattttatt tagataa 1317

<210> 15

<211> 1599

<212> DNA

<213> Clostridium phytofermentans

<400> 15

atgggcatgg agcattttta agatgcgatt cttacgggta aaacaacact tggaattgag 60

cttgggtcca ctagaataaa agctgtttta gtaaatgaag aaaacgaacc aattgcgtca 120

ggaagccatg attgggaaaa tcaatatatt gataatgtat ggacttacaa tctggatgat 180

atctggaggg gcgttcagaa tagttatgga caaatgacaa gtgatgttaa gaataagtac 240

ggagtagaac ttacaacaat tggagccatt ggttttagt gaaatgatgca tggctatatg 300

gcttttgacg aaagtggaga gttacttgta ccatttcgta cctggagaaa tacaataaca 360

ggaccagcat ccgagcagtt gaccaatgta tttcagtatc aaattccaca acgttggagt 420

attgcccac tatatcaagc tatcttaaatt gggaatccc acgtgaaaaa tattagattc 480

ctgacaacat tggcaggata tattcactgg aagctaacag gagaaaaagt attaggagtc 540

ggagaagcat ctggaatgtt tccaatcgat ataaatacga aagattttaa taaatcaatg 600

ttagctcagt ttaatgaact gggtgcttcg aatgactatt catggaaaat agaagatatt 660

ctaccgaaag tactagtgc aggagagtct gctggagtat taaccgaaga aggagtaaaa 720

cttcttgatg tttcaggtaa attaaaagca ggaattcctc tttgtccgcc ggaaggagat 780

gctggaactg gtatggtagc aaccaacagc gtagcaaaga gaactggtaa tgtatctgct 840

ggtacttctg tatttgcaat ggctgtattt gaaaaagagc tttcaaaagt ttacgaagaa 900

attgaccttg tgacgactcc aagcggagat cttgtggcta tggtgactg caataactgt 960

acttctgatt tgaatgcgtg gggttctatc tttaaagaat ttgcttcggc aatgggcatg 1020

gaagctgata tgtcaaagat attctcaacg ctatacaata aggcgttaga aggcaatgca 1080

gagtgtggag gcttactcgc atacaattat tttccggtg aacatataac acactttgaa 1140

gaagccgcc cattgtttgt aagaactcca gagagtaagt ttaaccttgc gaatttcatg 1200

agagttcatc tattcacagc acttgggtgt ttaaagatag gtcttgatat cctattaaaa 1260

caagaatcag tacaattgga tgagatTTTT ggtcatggtg gattatttaa gacgaaagat 1320

gtcggacaaa aaattatggc tggtgcaatc aatgttcctg tttctgtgat ggagactgcc 1380

ggagaaggcg gagcatgggg aatcgcaatc ttggcttctt atatgatttc taaggaagaa 1440

ggtcagtcct tagatgagta tctttctaaa catgtattcc aaggaaagac aggtagcaag 1500

atgcagccgg atccaaggga tgtagaaggt tttgaacagt ttatcaaacg atatattgac 1560

ggacttgaaa ttgagcgtaa agcagtgagg atattataa 1599

<210> 16  
 <211> 240  
 <212> DNA  
 <213> Clostridium thermocellum

<400> 16  
 actttgccaa cgttacaagg gaagttgcaa gagcggttgc cgagtcgga gcaatttcaa 60

taatcggagg cggagattct gccgcagcta tagaacagct tggttttgcc gataagatta 120

cccacatttc aaccggaggc ggcgctctt tggagtttct tgaaggaaaa gtattgccgg 180

gaattgatgt attaatggat aaataaggag agaagaggtc atgagtagaa aagttattgc 240

<210> 17  
 <211> 420  
 <212> DNA  
 <213> Clostridium thermocellum

<400> 17  
 aattactgta tctctctggc attgccaggt tttataaag attaaaatta ttgactagaa 60

ataaaaaaat tgtccataat attaatggac aaaaaaaca agaattacat caaaggaaga 120

taaaaatact ttgttaaaaa attaattatt ttttatctaa actattgaaa atgaaaataa 180

aataatataa aatgaatcat agtgcaagag atacttgcca gaggatgaat attttactgc 240

attcatgctt tatggcagct aatagaggca ttaaattaaa ttttaattta caataggagg 300

cgatattaat ggcagtaaaa attggtatca acggttttgg acgtatcggc cgtcttgtgt 360

tcagggccag tcicaacaac ccgaacgttg aggttgtagg tataaacgac ccatttattg 420

<210> 18

<211> 300

<212> DNA

<213> Clostridium thermocellum

<400> 18

ctgccccatt aaaagctcgg ttccaaccgc taatatctcc gcattcatat tgaaagaccc 60

cttaaattta aactttttgt aacttattat atcaattagt gttataaaat aaaagggaaa 120

aagaattaaa atcaaaggtt tcaagagcag ccgtatcacc cgtaaaagtt tcagccgatt 180

caaccttttt acacataaaa ctttcaaaaa ttgatgactt acaattatca agtaggatat 240

aatattacta atgctaaaca gttattgata aaggaggaag gaatatgaac aataacaaag 300

<210> 19

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 19

ggcgggaattc ctiggtctga caatcgatgc 30

<210> 20

<211> 33

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic



primer

<400> 20

ggcgggaattc tatcagttat taccacttt tcg 33

<210> 21

<211> 674

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 21

gaattctgcg acagaatagg gattgacaat tcctttataa agcaatcaag gggttcagaa 60

gaggctgtta ttttgaataa agagctaaag aatcacaaag atgcaataga ggctgttatt 120

tctgcactga ctgacgataa tatgggcgtt ataaaaaaca tgtccgaaat atcagcagtg 180

ggacacagaa tagtacacgg cggtgaaaaa ttcaacagtt ctgtagttat agatgaaaac 240

gttatgaatg cagtaagaga gtgtatagac gttgcaccgc ttcataatcc gccgaatatt 300

ataggtatag aggcttgcca gcagattatg cccaatatac ctatggtagc tgtatttgat 360

accactttcc acagctccat gcctgattat gcataccttt acgattgcc atatgaactt 420

tatgaaaagt acggtataag aaaatatggt ttccacggaa catcacacaa atatgttgca 480

gaaagagctt ctgcaatgct tgataagtct ttgaacgaat taaagataat tacatgccat 540

cttggaacg gticaaglat ttgtgctgtt aacaagggt aatcaattga tacttccatg 600

ggctttacac ctttgcaggg acttgcaatg ggtacaagaa gcggtacaat agaccctgaa 660

gttgttacga attc 674

<210> 22

<211> 951  
 <212> DNA  
 <213> Clostridium thermocellum

<400> 22  
 atgaaaaata aatctataaa taaaatagta attgtaggta cgggttttgt cggttcaaca 60  
  
 actgcctata cttaaatggt cagcggacta gtttccgaga ttgtacttat tgaccgtaac 120  
  
 acaagcaaag ccgaaggaga ggcaatggat atgaatcacg gtatgccctt tgtaagacct 180  
  
 gtcagaatat acaaaggiga ttatcctgat tgcaaagggt ctgatattgt tgtaataaca 240  
  
 ggtggagcaa accagaagcc cggtgaaacc agaattgacc ttgtaaataa aaatactgaa 300  
  
 gtttttaaag acattgttgg aaatatcatt aaatacaata cagactgtat tttacttgtt 360  
  
 gttacaaacc cggttgatat cttaacctat gtaacataca aattatccgg atttcccaaa 420  
  
 aacagagtta taggctccgg aacagttctt gatactgcac gtttcaaata tatgcttggt 480  
  
 gaacacatgg gaggtagacc aagaaacgtt catgcttata taatcgggtga acatggagat 540  
  
 acagaggtag ctacatggag tctggcatcc atagccggga taccgatgga tgcttattgc 600  
  
 aaggaatgta aatcctgtga tgctgaaaac tttagagtg aaacttttga caaagtaaaa 660  
  
 aatgcagctt atgaaattat tgatagaaaa aatgcaacct actacgccgt tgctcttgca 720  
  
 gtaagaagaa ttgtagaggc tatcgttcgt aatgaaaact ccatattgac ggtatcaagc 780  
  
 ctattcgaag gagaatacgg cctcaatgac atatgtctca gtattcccag ccaggtaaat 840  
  
 tcggagggtg tttcaaggat tttgaatatt cctctgagca gtgaggaaac aggtttactt 900  
  
 aataaatctg cccaggcctt gaaacaggtt atcagtgggc tgaatttata a 951

<210> 23  
 <211> 933  
 <212> DNA

<213> Clostridium thermocellum

<400> 23

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| atgggtttta aagttgcgat cataggagca ggattttgtg gagcatcagc tgcgtatgcg  | 60  |
| atgtctataa acaacttggg ttctgaattg gtattaattg atgtaaataa agagaaggct  | 120 |
| tatggatgaag cacttgatat cagccatggc ttatcattct caggaaatat gacagtttat | 180 |
| tccggcgact attctgatgt taaggattgt gatgttatag ttgtaactgc aggggcagca  | 240 |
| agaaaaccgg gagaaactcg tttggacctt gctaaaaaga atactatgat catgaagagc  | 300 |
| atagttactg atataatgaa gtactacaat aagggtgtta ttgtaagtgt atcaaactct  | 360 |
| gttgatgtat tggcatatat gacacaaaag tggtcaggat tgcctgcaa taaagttata   | 420 |
| ggatcaggaa cagtctctga cagtgaaga ctgagaactc atatcagtca ggcattggat   | 480 |
| gtagacattg ctaacgttca cggttatatt gttggtgaac atggtgattc tcagttgcc   | 540 |
| ttatggagtg caacacatat agcaggagta caatttgacg actatgtaa agctactggc   | 600 |
| ttaaatgttg ataaggaagc tcttttcaat gaagttaagg tagcaggtgc aactattatt  | 660 |
| aagaacaagg gagcaactta ctacgtata gctctttcaa ttaacagaat agttgaatca   | 720 |
| atcctgaagg acttcaatac tattatgcct gttggtacag ttcttgacgg acagtacga   | 780 |
| ttaaaggatg ttttattaaa cgttcctacg atagttggcg gaaacggagc tgaaaaagtt  | 840 |
| cttgaagtga acattacaga tgcagaatta caacttttga agcattcagc tgaacaggtt  | 900 |
| agggcagtta ttaacgaagt taaagacata taa                               | 933 |

<210> 24

<211> 2830

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 24

|                                                                    |      |
|--------------------------------------------------------------------|------|
| gacgcataca ggttgaaca cccatttccc ttagcttttc gggagatgaa taaaacaaac   | 60   |
| tttccgggtc cttaccaca ccgcccacat aaagagctat gccgcatgaa agaaacgata   | 120  |
| tgttatcatt tttttcgtaa actgttatit ccgaaccggg ataaagcttt accatattat  | 180  |
| taactgtgc cgtccctgca tgtgtacacc ctataaccac tattttcata tacatcctcc   | 240  |
| tttgtttgct tgtaaataata tcccatatat accacctaaa tatattttat aaacaaattc | 300  |
| ggtatatcat tcttttggtg aataaaaagt acatccgata ttagaatgta cctaaaaaaa  | 360  |
| attattattt tattgtatat gctttatctg ttttcattat atggtttgc atccattcta   | 420  |
| cggtaaaatc aagtaattcc attaagtact gatcctgac cttgtctatc ctgctataat   | 480  |
| ccgtattact gatttttctca ataaaatcat ggtgttcaac tttgtgggag agaagcttgc | 540  |
| gatatcctat gctatgcatg tattcttctt cataggtaaa atgaaagaca gtgtaattct  | 600  |
| ttagttccgt aattagccgt acaatttcat catatttgc tgtaataagc tgatttttcg   | 660  |
| tggcctcata aatttccgaa gcaatctgga atagtttctt atgctgttcg tcgattttct  | 720  |
| caattccaag aataaattcg tctctccatt ctatcatatg gacctccta aattgtaatg   | 780  |
| tataccaaga ttatacatc ttcttagaat ataaacaata caaggataaa attttaatat   | 840  |
| cgtataccta cataaatgac taacttaaag ctctctaaaa cttctttttt attatttcta  | 900  |
| tactactaaa atcaaaaata ttctctaaag tatttctaca aatgttgttt ttgcaacaaa  | 960  |
| gtagtatact ttgacacca gaatgttttg ttataactta caaattaggg gtatatttat   | 1020 |

agtaaatact aaatggaaga gtaggatatt gattatgaac gagaaaaata taaaacacag 1080

tcaaaacttt attacttcaa aacataatat agataaaata atgacaaata taagattaaa 1140

tgaacatgat aatatctttg aaatcggctc aggaaaaggg cattttacc ttgaattagt 1200

acagaggtgt aatttcgtaa ctgccattga aatagacat aaattatgca aaactacaga 1260

aaataaactt gttgatcacg ataatttcca agttttaaac aaggatatat tgcagittaa 1320

atttcctaaa aaccaatcct ataaaatatt tggtaatata cttataaca taagtacgga 1380

tataatacgc aaaattgttt ttgatagtat agctgatgag atttatttaa tcgtggaata 1440

cgggtttgct aaaagattat taaatacaaa acgtcattg gcattatfff taatggcaga 1500

agttgatatt tctatattaa gtatggttcc aagagaatat tttcatccta aacctaaggt 1560

gaatagctca cttatcagat taaatagaaa aaaatcaaga atatcacaca aagataaaca 1620

gaagtataat tatttcgtta tgaaatgggt taacaaagaa tacaagaaaa tatttacaaa 1680

aatcaatff aacaattcct taaaacatgc aggaattgac gatttaacaa atattagctt 1740

tgaacaattc ttatctcttt tcaatagctt taaattatff aataagatcc cttttacttc 1800

ggatgcatgc cgcaggcagg catccgaagt agtttctcca ttatacaagt attctcttga 1860

gtacgtcgtc gcttctcagc agctgctttg ctttttcctt gttttcggc acatggagat 1920

aagtgtatct gttaggctta atagtgtgtg ccatgtcaat tgccttttcg aagtcacttg 1980

ctttcatfff taaggtttcc acaaaattga taaaaccctg atcagtcaga aattttacta 2040

cccgctgata tctgtgttct tgaaccctgc tcataagata ggttgcaatc ccaacctgaa 2100

ttcatgaag ctgaggtgtc tccagcagct tatctaaagc atgagatatt agatgtcac 2160

taccgctggc tggagcactg ctgtctgcta tctgcatggc aattccgctc attgtcagag 2220

agtcctacat ttccctttaa aagaagtttt ctgtaacctg tgtgtagggc atccttaca 2280

tactgtttac tgacttttta gcaatcattg cagcaaaatc gtcaaccttt gccgcattgt 2340

tcctttcttc aaaataccag tcatacacag ccgtaatttt ggatattatg tctccgagac 2400

ctgaataaat aaatttcata ggtgcatttt ttaatacatc taaatccact aatattccaa 2460

atggcatcga ggcattgtac gaagtacgcc tgccatttat aatcaaagag cagcctgagc 2520

tggaaaaacc atcgtttgag gttgatgtag gtatactgat aaaaggaagc ttgtttaaaa 2580

aagctatata tttggctgca tcaagcacct ttcctcctcc tactccgacc actgcatcgg 2640

ttttggaggg aatagtaaaa gccttgagca taagattttc aagctttatg tcatcatagt 2700

cgtaagtttc aagtactgca agagattttc ttgactttat ggaatccaga atcttttcac 2760

caaataagtc acgtattccc tctccaaaaa gtactacaac attactaatt cctgcccttt 2820

caatatgtgc 2830

<210> 25

<211> 572

<212> DNA

<213> Clostridium thermocellum

<400> 25

ccaaggtgac aaacgataac ttttgagtta ttacatcta agccagcaag cgtggttgct 60

cttttagaaa catagctgtg acttggtccg tggaaacat atcttcttac cttatattta 120

tcatagtact caaatggaat accataaagg taagcttctt ttggcattgt ctgatggaat 180

gcagtatcaa aaacagctac cattggtaca tttggcataa ttgatattaca agcgttgata 240

ccaataaggt ttgctgggtt gtgtaaaggt gcaagatcat tacactcttc aattgcattt 300

aagacttcat cattgattac tacggaatga gcaaatttct caccacatg tactactcta 360

tgtccaacag cgttgatttc atctaaggac ttaatcacac cataatttct attcataaga 420

gcagcgatta cttttttaat agcaacctca tggtttggaa gtgcatcctc aagaactacc 480

ttctcaccgt cagctgactt gtgagtaaga cggccatcaa taccgattct ttcacaaaga 540

cctactgcta atgcttgctc tgtcacagag tc 572

<210> 26

<211> 2676

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 26

ctgagtgcac tgtaaaaaag gatgcctcaa gtattcttga aacatcctta tattatacta 60

caaaatcata aagtaaatta ctgagctgia gcaatgatct cttttttgtt gtaagatcca 120

caagctttac aaactctatg aggcatcata agtgcaccac acttgctgca tttcactaag 180

tttggagcag tcattcttcca gtttgacaga cgactatctc ttctagcttt ggaatgttta 240

ttctttggac aaatagctcc cattgattac acctccttaa acttggttaa aatatctcgg 300

atagcagaca ttcttgggtc tagttctgta cggtcacacc cgcactctcc ttcatttagg 360

ttagcaccgc agaccttgca gattccttta cagtcttctt tgcacagaac cttcattggg 420

aaaccaatca agacttcttc atagataagt ttatctacgt ctaaatacata tccggaacaa 480

aaatttggtt catctaaatc ctcggtacgc tgttctcttg ttttcgatac atcaatctct 540

gtagccacgt cgatgtcttg ttggatgggt tcttccttca aacaacgac gcaaggaacg 600  
 gctaacgcta atttcgtttt tgcttcacc agaatttttc ggccacctag attagttaat 660  
 ctaagtttaa ccggttcttt ataggttaata gaataaccga caccatttaa ttcgaatata 720  
 tcaaattcaa tcggtgcagt gtattctttg agaccattag gaacattcat gacttcagac 780  
 atttgtatca gcataagtaa ctctgtctta aaaaaacgca taatgtaagc gcccaaaaat 840  
 tcacactgtt agtattataa acgcttaaaa taggtttgtc aactcctaac tgttaaaaat 900  
 gtcagaattg tgtaaccata ttttctcttc attatcggtc ttccttatt aaataattta 960  
 tagctattga aaagagataa gaattgttca aagctaatat tgtttaaatc gtcaattcct 1020  
 gcatgtttta aggaattgtt aaattgattt ttgttaaata ttttcttgta ttctttgtta 1080  
 acccatttca taacgaaata attatacttc tgtttatctt tgtgtgatat tcttgatttt 1140  
 tttctattta atctgataag tgagctattc actttagggt taggatgaaa atattctctt 1200  
 ggaaccatac ttaatataga aatatcaact tctgccatta aaaataatgc caatgagcgt 1260  
 ttgtatttta ataatctttt agcaaaccg tattccacga ttaaataaat ctcacagct 1320  
 atactatcaa aaacaatttt gcgtattata tccgtactta tgttataagg tatattacca 1380  
 aatattttat aggattgggt tttaggaaat ttaaactgca atatatcctt gtttaaaact 1440  
 tggaaattat cgtgatcaac aagtttatit tctgtagttt tgcataattt atggctctatt 1500  
 tcaatggcag ttacgaaatt acacctctgt actaattcaa gggtaaaatg cccttttctt 1560  
 gagccgattt caaagatatt atcatgttca tttaacttta tatttgtcat tattttatct 1620  
 atattatgtt ttgaagtaat aaagttttga ctgtgtttta tatttttctc gttcattgta 1680  
 tttctcctta taatgttctt aaattcattt atcacggggc aacttaatat atccgaaata 1740



tagttcttct atatcgttcc cccagtataa tgattattat actatittaat cttcaactta 1800

acaattggag tticcagtta agaaataata atttaatgcc aaagcggata ttcgcaatcc 1860

gcttacgcta ctigctcata acctcaacag gcaatgaagc taagttaatt atttactctg 1920

tgccgaaca gcagtgattg caacaacacc aacgatatca tcagaagaac aacctcttga 1980

taaatacattt actggagctg caataccctg agttaatggt ccataagctt ctgcctttgc 2040

aagacgctgt gttacttat atccaatgtt accagcatca aggtctggga agattaatac 2100

gttagctttt ccagcaatat cactaccagg agcttttgaa gcacctacac taggaacgat 2160

tgctgcatct aactggaact cgccgtcgat cttatatctt gggtataatt catttgcaat 2220

cttagttgct tctacaacct tatcaacatc tgcattgttt gcgcttcctt ttgttgaatg 2280

agaaagcata gctacgatag gttcagagcc aactaattgt tcaaaactct tcgctgtgga 2340

accagcgatt gctgctaact cttcagcatt tggattctga tttaaaccag catcagagaa 2400

aaggaaagtt ccatttgcgc ccataacaca attaggtact accattacga agaaagcaga 2460

aactaactta gtatttggag cagtttttaa aatctgaaga catggtctta aggtatctgc 2520

tgtagagtga caagcaccag atactaaacc atctgcatcg cccatcttaa ccatcattac 2580

accgtatgta atgtagtctg ttgttaaaag ctcttttgct ttttcagggg tcatgccttt 2640

tgccgtgcta agttctacaa gcttgtaaat gtaagc 2676

<210> 27

<211> 978

<212> DNA

<213> Clostridium thermocellum

<400> 27

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| atggcgatta caataaacg aagtaaagtt attgtgtgg gtgcaggttt agttgttact   | 60  |
| tcaacggcgt ttagtctaataacgaaagt gttgtgatg aggttatgtt gatagatatc    | 120 |
| aatcgtgcta aggcgcattg ggaagtaagt gatttgtgtc atagtatcga gtatttaaat | 180 |
| cgaaatgttt tggtaacgga aggagattat acagactgta aggacgctga tattgttgta | 240 |
| ataactgcag ggcttcgcc aaaaccagga cagtgcggc ttgatactct tgggttatcc   | 300 |
| gcagatattg tgagcacgat tgtggaacct gtcataaga gtgggttcaa tggaaatattc | 360 |
| ttagtcgtga cgaatccgtt ggattcgatt gctcaatatg tttatcaatt atcggggctt | 420 |
| ccaaagcaac aagtcttgg aactggaaca gcgattgact ctgcaagatt aaaacacttt  | 480 |
| attggagata tttacatgt agatcctaga agcatacagg cttatacgat gggagagcat  | 540 |
| ggagattctc aaatgtgtcc ttggtcgtt gttacggtt gcggtaaaaa tattatggac   | 600 |
| atcgtacggg ataacaaaga gtattccgat attgacttta atgaaatctt atataaggtt | 660 |
| accagggtag gttttgatat tttatcagt aagggtacta cttgttatgg aatagcgtca  | 720 |
| gcagctgtgg ggattataaa agcaattctt tatgatgaga attccatcct tccggtctct | 780 |
| accttattgg agggggaata tggtagttt gatgtatatg cagggtacc atgcattcta   | 840 |
| aatcgtttcg gcgtgaagga ttagtgga gtaaatatga cagaagtaga gttaaatcaa   | 900 |
| ttccgagcct ctgttcacgt tgtgaggga gctattgaaa acttaaaaga cagagataaa  | 960 |
| aaggcattat tttataa                                                | 978 |

<210> 28

<211> 960

<212> DNA

<213> Clostridium thermocellum

<400> 28

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| ttatgatagc gtcaatgcat actgaaaatt ttctttcatt gttctacaag aagcatcgaa  | 60  |
|                                                                    |     |
| ttttcccttt tcttcagggtg tcaaatttag ctcaatgatt tcttctacac catgaattcc | 120 |
|                                                                    |     |
| aagtaccgta ggaacagatg catagacatc atgctggcca tactcaccat ttaagagagt  | 180 |
|                                                                    |     |
| agatactggt aataccttct tctcatctga gaaaatggct cgtgtaacct cagctagtga  | 240 |
|                                                                    |     |
| tgcaccaata ccaaattccg ttgagccttt tccagttagg atatgccatc cccctgctct  | 300 |
|                                                                    |     |
| agcttcatca gaaagcttag aaagatcaat ctgcccatat ttttcagggtt tttccttgat | 360 |
|                                                                    |     |
| tagttccaaa attggttttc cagctataga taccgttgac catgcaacca tctggctttc  | 420 |
|                                                                    |     |
| tccgtgttct ccaagaacaa atccatagat tgatttttga tcaatttcaa cagcatctgc  | 480 |
|                                                                    |     |
| aattgctctt ctaagtctgg cagagtctag taccgtactt gttgaaataa ttttattgga  | 540 |
|                                                                    |     |
| tgagtactga agtaaagct gtaaataatg tgttattaca tctgctggat ttgaaatgct   | 600 |
|                                                                    |     |
| aacaatcata ccatcaaac ctgaattttt gatatgcaa gctacctctt taataattag    | 660 |
|                                                                    |     |
| agcagtattc gtaagggtac tcattcttgt ttcaccctta tttttatctg gattggttcc  | 720 |
|                                                                    |     |
| tactgcaatc accatgagat ctgcatcagc tgcatacta taatcaccg attttacctt    | 780 |
|                                                                    |     |
| aactctgtgt ggtaggtata ctgtagcatc gtagatatcc agtcttctg ctttcgcttt   | 840 |
|                                                                    |     |
| ttctctatca atatcaataa agataatttc ttctgcaagc cctgctctg ccagtgcata   | 900 |
|                                                                    |     |
| tccagcatga gatcctacgt gacctgctcc gataataatg acttttcttg gttttgcat   | 960 |

<210> 29

<211> 2732

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 29

|                                                                    |      |
|--------------------------------------------------------------------|------|
| tggaatctca ctatgcacca atgtggtact aaattatatac tttatctatg gaaaattagg | 60   |
| ttttccgcga atggagatag agggagctgc cattgctact ttaatttgta gaattcttga  | 120  |
| gagtatttta gttgttattt atatgtataa gggtgagaag gtacttaaga tgagactttc  | 180  |
| ttatatTTTT aagagatcta aacagtattt tcgctctttg gctcgttata gtgcgccagt  | 240  |
| gcttatgagt gaggttaact gggggcttgg gattgctgtt cagtctgcaa tcattgggcg  | 300  |
| tatgggtgtt agttttctta cagccgccag cttcattaat gtagtacaac agttagccgg  | 360  |
| aatcattctg attggtattg gtgtgggttc gagcattata ataggaatt tgattggtga   | 420  |
| gggaaaagag catgaggcga gaatgctagc caataagtta atacgtatca gtatgatact  | 480  |
| cggaggaatt gttgcttttg cagtaatctt actacgtcca atcgctccta actttattga  | 540  |
| ggcgtctaag gaaacagcgg atttaattcg tcagatgcta tttgtttcgg cttacctctt  | 600  |
| attcttccaa gccttatctg tattaactat ggccggaata ttacgtggtg caggggatac  | 660  |
| cctttactgt gcaacctttg atgttttgac cttatgggta ctaaaacttg gaggaggttt  | 720  |
| gcttgcaacc atagtacttc atcttcacc tgatgggtt tactttatct taagtagcga    | 780  |
| tgagtgtgtt aaagcgctat ttacggtacc gcgggtctta aaggacgtt ggattcatga   | 840  |
| tacaacactg cattaagatt tcatatgtcc agatatTTTT gcacagtagc ataattacta  | 900  |
| gagcttattc ctataatatt cataggTTTT gatgtccat tttacgttac gatagcatat   | 960  |
| attacatcaa aaccaattct atataagatg aggttatagt atgaacgaga aaaatataaa  | 1020 |
| acacagtcaa aactttatta cttcaaaaca taatatagat aaaataatga caaatataag  | 1080 |

attaaatgaa catgataata tctttgaaat cggctcagga aaagggcatt ttacccttga 1140

attaglacag aggtgtaatt tcgtaactgc cattgaaata gaccataaat tatgcaaaac 1200

tacagaaaat aaacttgttg atcacgataa ttccaagtt ttaacaagg atatattgca 1260

gtttaaatct cctaaaaacc aatcctataa aatatttggg aatatacctt ataacataag 1320

tacggatata atacgcaaaa ttgtttttga tagtatagct gatgagattt attaatcgt 1380

ggaatacggg ttgctaaaa gattattaaa taaaaacgc tcattggcat tatttttaat 1440

ggcagaagtt gatatttcta tattaagtat ggtccaaga gaatattttc atcctaaacc 1500

taaagtgaat agctcactta tcagattaaa tagaaaaaaa tcaagaatat cacacaaaga 1560

taaacagaag tataattatt tcgttatgaa atgggttaac aaagaatata agaaaatatt 1620

tacaaaaaat caatttaaca attccttaaa acatgcagga attgacgatt taaacaatat 1680

tagctttgaa caattcttat ctcttttcaa tagctataaa ttatttaata agaagtaata 1740

ggaaataata ctgaattat tctgcaatct gtcttaaaaa ataaaattaa gaaattacta 1800

tagcaagcca ggtaaaaatt actagcttgc tatttttgtg catttagtac agttttgatt 1860

attaaagaat aaatttaata actattttgc aataagttat tgactatttc acaagttagt 1920

gttactatac aagtatgaaa taaagataca taaaaaata aataatatga aacataaatt 1980

catgacatgc ggaatagaat gaaagaatat tatgtcgggt cctaatacta aatggatata 2040

acaatctatt gaaacactta tggggtgtaa gtgtggagag aatttctaaa gcgccaaaag 2100

actctacata tgaatttcta aagcttcaca cgggaataat ctaatttatg tatcttatta 2160

tcataattca ggaaggtagt gtgaaaatat aaaaattagt ttctctgttt cattcaggca 2220

gtagcatctt ttaacaaaat ttgctatgca ttgggtgtta tctgaaaaac aaaaagcaat 2280

tttctcacia cttatttctg aacaacaatg gtattaaaaa tttggaggag gatitttacta 2340

tgaaaaaac ggtaacatta ctgttggttc tgaccatggt ggtaagctta ttgcagcat 2400

gtggtaagaa aaatggatca agcgaaccg gcacaaaaga tcctgtggca acaagcgggtg 2460

caaaagaacc tgacaaacaa gatccaggca ataaagagcc tgaaaaacaa gaccctgtta 2520

aatcaagat ttattactct gataatgcaa ccttaccatt taaagaagat tggtagtta 2580

taaaggaagc tgagaagaga tttaatgttg atttcgattt cgaagtaatt ccaattgcag 2640

attatcaaac aaaagtttct ttaacattaa atacaggaaa taacgtcca gatgtcatcc 2700

tttatcagtc aacgcaggga gagaatgcat ct 2732

<210> 30  
 <211> 10665  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: Synthetic  
 polynucleotide

<220>  
 <221> modified\_base  
 <222> (1522)..(1522)  
 <223> a, c, g or t

<400> 30  
 gaattcgagc tcggtaccg gggatcctct agagtcgacc tgcaggcatg caagcttggc 60

actggccgtc gttttacaac gtcgtgactg ggaaaaccct ggcggtaccc aacttaatcg 120

ccttgacga catccccctt tcgccagctg gcgtaatagc gaagaggccc gcaccgcgta 180

gaggatcgag atctcgatcc cgcgaaatta atacgactca ctataggga attgtgagcg 240

gataacaatt cccctctagg ctcataactt cacgtcctg tataatattt tatttatita 300

aaaatgagtc aaaatttagg aaatgattgc aatgtata atatccaaat ttccattcaa 360

ataaccaaag taattttacc tctttttatg agctatttca atactttgtt agtaaattaa 420

cataatgag ctgtatatgg tttaatgaaa aaagttattt tgaagggata ttgtaaaaaa 480

cataatatat tatatggata aattttacat ttgacttadc atatgttaat atatgtaata 540

tgaatagcta atctaagcag gctactgcct agaaaaaagc ttataattat ccttaatttc 600

ctactacgtg cgcccagata ggggtttaag tcaagtagtt taaggtaacta ctctgtaaga 660

taacacagaa aacagccaac ctaaccgaaa agcgaaagct gatacgggaa cagagcacgg 720

ttggaaagcg atgagttacc taaagacaat cgggtacgac tgagtcgcaa tgtaaatcag 780

atataaggta taagtttgtt ttactgaacg caagtttcta atttcggttg aaatccgata 840

gaggaaagtg tctgaaacct ctagtacaaa gaaaggttaag ttacagtagt agacttatct 900

gttatcacca catttgtaca atctgtagga gaacctatgg gaacgaaacg aaagcgatgc 960

cgagaatctg aatttacaa gacttaacac taactgggga taccctaaac aagaatgcct 1020

aatagaaagg aggaaaaagg ctatagcact agagcttgaa aatcttgcaa gggtagcgag 1080

tactcgtagt agtctgagaa gggtaacgcc ctttacatgg caaaggggta cagttattgt 1140

gtactaaaat taaaaattga ttagggagga aaacctcaaa atgaaaccaa caatggcaat 1200

tttagaaaga atcagtaaaa attcacaaga aaatatagac gaagttttta caagacttta 1260

tcgttatctt ttacgtccag atatttatta cgtggcgacg cgttgggaaa tggcaatgat 1320

agcgaaacaa cgtaaaactc ttgttgtatg ctttcattgt catcgtcacg tgattcataa 1380

acacaagtga atgtcgacag tgaattttta cgaacgaaca ataacagagc cgtatactcc 1440

gagaggggta cgtacgggttc ccgaagaggg tggtgcaaac cagtcacagt aatgtgaaca 1500

aggcggtacc tccctacttc ancatatcat tttctgcagc cccctagaaa taattttgtt 1560

taactttaag aaggagatat acatatatgg ctagatcgtc cattccgaca gcatcgccag 1620

tcactatggc gtgctgctag cgctatatgc gttgatgcaa tttctatgca ctgtagtag 1680

tctgagaagg gtaacgccct ttacatggca aagggtaca gttatttgt actaaaatta 1740

aaaattgatt agggaggaaa acctcaaaat gaaaccaaca atggcaattt tagaaagaat 1800

cagtaaaaat tcacaagaaa atatagacga agtttttaca agactttatc gttatctttt 1860

acgtccagat atttattacg tggcgtatca aaatttatat tccaataaag gagcttccac 1920

aaaaggaata ttagatgata cagcggatgg ctttagtgaa gaaaaataa aaaagattat 1980

tcaatcttta aaagacggaa ctactatcc tcaacctgta cgaagaatgt atattgcaaa 2040

aaagaattct aaaaagatga gaccttagg aattccaact ttacagata aattgatcca 2100

agaagctgtg agaataatc ttgaatctat ctatgaaccg gtattcgaag atgtgtctca 2160

cggttttaga ccacaacgaa gctgtcacac agctttgaaa acaatcaaaa gagagtttgg 2220

cggcgcaaga tggtttggg agggagatat aaaaggctgc ttcgataata tagaccacgt 2280

tacactcatt ggactcatca atcttaaaat caaagatatg aaaatgagcc aattgattta 2340

taaatttcta aaagcaggtt atctggaaaa ctggcagtat cacaaaactt acagcggaac 2400

acctcaaggt ggaattctat ctctctttt ggccaacatc tatcttcag aattggataa 2460

gtttgtttta caactcaaaa tgaagtttga ccgagaaagt ccagaaagaa taacacctga 2520

atatcgggag ctccacaatg agataaaaag aatttctcac cgtctcaaga agttggaggg 2580



tgaagaaaaa gctaaagttc ttttagaata tcaagaaaaa cgtaaaagat tacccacact 2640

cccctgtacc tcacagacaa ataaagtatt gaaatacgtc cggtatgcgg acgacttcac 2700

tatctctgtt aaaggaagca aagaggactg tcaatggata aaagaacaat taaaactttt 2760

tattcataac aagctaaaaa tggaattgag tgaagaaaaa acactcatca cacatagcag 2820

tcaaccgcgt cgttttctgg gatatgatat acgagtaagg agatctggaa cgataaaacg 2880

atctggtaaa gtcaaaaaga gaacactcaa tgggagtgtga gaactcctta ttcctcttca 2940

agacaaaatt cgtcaattta tttttgacaa gaaaatagct atccaaaaga aagatagctc 3000

atggtttcca gttcacagga aatatcttat tcgttcaaca gacttagaaa tcatcacaat 3060

ttataattct gaactccgcg ggatttgtaa ttactacggt ctagcaagta attttaacca 3120

gtcgaattat tttgcttacc ttatggaata cagctgtcta aaaacgatag cctccaaaca 3180

taaggaaca ctttcaaaaa ccatttccat gtttaaagat ggaagtgggt cgtgggggat 3240

cccgatgag ataaagcaag gtaagcagcg ccgttatatt gcaaatttta gtgaatgtaa 3300

atccccttat caatttacgg atgagataag tcaagtcctt gtattgtatg gctatgcccc 3360

gaatactctt gaaaacaggt taaaagctaa atgttgtgaa ttatgtggga cgtctgatga 3420

aaatacttcc tatgaaattc accatgtcaa taaggtcaaa aatcttaaag gcaaagaaaa 3480

atgggaaatg gcaatgatag cgaaacaacg taaaactctt gttgtatgct ttcattgtca 3540

tcgtcacgtg attcataaac acaagtgaga tatctcgagc acccgttctc ggagcactgt 3600

ccgaccgctt tgccgccgcg ccagtcctgc tcgcttcgct acttgagacc actatcgact 3660

acgcgatcat ggcgaccaca cccgtcctgt ggatcgccaa gtcgcccgat ggtagtgtgg 3720

ggtctcccca tgcgagagta gggaactgcc aggcatacaa taaaacgaaa ggctcagtcg 3780

aaagactggg cctttcgttt tatctgttgt ttgtcgtga acgtctctct gagtaggaca 3840  
 aatccgccgg gagcggattt gaacgttgcg aagcaacggc ccggagggtg gcgggcagga 3900  
 cgcccccat aaactgccag gcatcaaatt aagcagaagg ccatcctgac ggatggcctt 3960  
 ttgtcgtttc taaaaactct tctgtctgtc atatctacaa gccatcccg ctttccaac 4020  
 agttgcgcag ccigaatggc gaatggcgcc tgatgcggtt ttttctcctt acgcatctgt 4080  
 gcggtatttc acaccgcata tgggtgcactc tcagtacaat ctgctctgat gccgcatagt 4140  
 taagccagcc ccgacaccgg ccaacaccgg ctgacgcgcc ctgacgggct tgtctgctcc 4200  
 cgcatccgc ttacagacaa gctgtgaccg tctccgggag ctgcatgtgt cagaggtttt 4260  
 caccgtcatc accgaaacgc gcgagacgaa agggcctcgt gatacgcta tttttatagg 4320  
 ttaatgtcat gataataatg gtttcttaga cgtcaggtgg cacttttcgg gaaaatgtgc 4380  
 gcggaacccc tatttgttta tttttctaaa tacattcaaa tatgtatccg ctcatgagac 4440  
 aataaccctg ataatgctt caataatatt gaaaaaggaa gagtatgagt attcaacatt 4500  
 tccgtgtcgc cttattccc ttttttgcgg cattttgcct tctgttttt gctcaccag 4560  
 aaacgttgtt gaaagtaaaa gatgtgaag atcagttggg tgcacgagtg ggttacatcg 4620  
 aactggatct caacagcggc aagatccttg agagttttcg cccgaagaa cgttttcaa 4680  
 tgatgagcac ttttaaatta aaaatgaagt tttaaaactt catttttaat ttaaattaaa 4740  
 aatgaagttt tatcaaaaaa atttccaata atccactct aagccacaaa cagccctat 4800  
 aaaatccgc tttaatccca ctttgagaca catgtaatat tactttacgc cctagtatag 4860  
 tgataatfff ttacattcaa tgccacgcaa aaaaataaag gggcactata ataaaagttc 4920

cttcggaact aactaaagta aaaaattatc ttacaaacct ccccaaaaaa aagaacaggt 4980

acaaagtacc ctataatata agcgtaaaaa aatgagggtta aaaataaaaa aataaaaaaa 5040

taaaaaaata aaaaaataaa aaaaataaaa aaataaaaaa ataaaaaaat aaaaaataa 5100

aaaaataaaa aaataaaaaa ataaaaaaat ataaaaataa aaaaatataa aaataaaaaa 5160

atataaaaaat aaaaaaatat aaaaataaaa aaataaaaaa atataaaaaat aaaaaataa 5220

aaaaatataa aaatatTTTT tatttaaagt ttgaaaaaa tttttttata ttatataatc 5280

tttgaagaaa agaataaaa aaatgagcct ttataaaagc ccattttttt tcatatacgt 5340

aatatgacgt tctaagtgtt ttattggtac ttctaacatt agagtaattt ctttttttt 5400

aaagcctttt tctttaaggg cttttttttt ttttcttaac acatttaatt cctctttttt 5460

tgttgctttt cctttagctt ttaattgctc ttgataattt tttttacctc taatatattc 5520

tcttctctta tattcctttt tagaaattat tattgtcata tatttttgtt cttcttctgt 5580

aatttctaac aactctataa gagtttcatt cttatactta tattgcttat ttttatctaa 5640

ataacatctt tcagcacttc tagttgctct tataacttct ctttcactta aatgttgtct 5700

aaacatacta ttaagtctta aaacatcatt taatgccttc tcaatgtctt ctgtaaagct 5760

acaaagataa tatctatata aaaataatat aagctctctg tgcctttta aatcatattc 5820

tcttagttca caaagtttta ttatgtcttg tattcttcca taatataaac ttctttctct 5880

ataaatataa tttattttgc ttggcttacc ctttttcctt tcatatggtt ttaattcagg 5940

taaaaatcca ttttgtattt ctcttaagtc ataaatatat tcgtactcat ctaatatatt 6000

gactactgtt tttgatttag agtttatact tcctggaact cttaatattc tggttgcac 6060

taaggettgt ctatctgctc caaagtattt taattgatta tataaatatt ctggaaccgc 6120

tttccataat ggtaatgctt tactaggtac tgcatttatt atccatatta aatacattec 6180

tcttcacta tctattacat agtttggtat aggaatactt tgattaaaat aattcttttc 6240

taagtcatt aatactggg ctttagtttt gccagtttta taataatcca agtctataaa 6300

cagtgtattt aactctttta tattttctaa tcgcctacac ggcttataaa aggtatttag 6360

agttatatag atatattcat cactcataac taaatctttt aattcagcgt atttatagtg 6420

ccattggcta tatccttttt tatctataac gctcctgggt atccaccctt tacttctact 6480

atgaatatta tctatatagt tctttttatt cagctttaat gcgtttctca cttattcacc 6540

tccccttctg taaaactaag aaaattatat catattttca ataattatta actattctta 6600

aactcttaat aaaaaataga gtaagtcgcc aatlgaaact taatctattt tttatgtttt 6660

aatttattat ttttattaaa atatattaaa ctaaattaaa tgattctttt taatttttta 6720

ctatttcatt ccataatata ttactataat tatttcaaaa taatatttct tcatttgtaa 6780

tatttagatg atttactaat tttagttttt atatattaaa taattaatgt ataatttata 6840

taaaaaatca aaggagctta taaattatga ttatttccaa agatactaaa gatttaattt 6900

tttcaatttt aacaatactt ttgtgaatat tatgtttaaa ttttaattgta tttttttcat 6960

ataataaagc cgttgaagta aaccaatcca ttttccttat gatgttatta ttaaatttaa 7020

gttttataat aatatcttta ttatatattat tgtttttaaa aaaactagtg aaatttcggg 7080

ctttattaaa cttattttta ggaattttat tttcattttc atctttacag gatttgatta 7140

tatctttaaa tatgttttat caaatattat ctttttctaa atttatatat atttttatta 7200

tatttattat tatatatatt ttatttttaa gtttctttct aacagctatt aaaaagaaac 7260

ttaaaaaataa aaacacgtac tctaaaccaa taaataaaac tatttttatt attgctgcct 7320  
 tgattggaat agtttttagt aaaattaatt tcaatatcc acaatattat attataagct 7380  
 agctttgcat tgiacttttc aatcgcttca cgaatgcggt tatctccgaa agataaagtc 7440  
 ttttcatctt ccttgatgaa gataagattt tctccgtctc cgccggcaga attgaagcgg 7500  
 ggtactacgg tatcgtctgc gtcatttcc gttgtctgat agatgatagt cataggctca 7560  
 ttttcttccg tttcggtaaa ggggataggt tcgcccttg agagcagggc ggcgatggaa 7620  
 agcattaact tgcttttccc atgccccgga tctccctgca atagcgtaac ttgccaac 7680  
 ggaatatacg gataccacag ccactttact tctttcggct cgatttact tgccttgatg 7740  
 atttcaagag gtacgtgaa attcatttcg ttttcattha gtttcatttt ttcttgttct 7800  
 ctttttctct gaaaataataa aaaccacaga ttgatactaa aaccttggtt gtgttgcttt 7860  
 tcggggctta aatcaaggaa aaatccttgt ttaagcctt tcaaaaagaa acacaagtc 7920  
 tttgtactaa cctgtggta tgtataaaat tgtagatttt agggtaacaa aaaacaccgt 7980  
 atttctacga tgtttttgc taaatacttg tttttagtta cagacaaacc tgaagttgaa 8040  
 ttcatattta ttaaattaag cgtatatact attgaaaatg tttttgaaat attataaaat 8100  
 taactttggt ttaggaaaag taaccagttc tttgtcgat aagcattaat ttgcttgact 8160  
 aattaataaa aaacttagga ggtaacacta atggtattcg agaaaattga caagaacagt 8220  
 tggaacagaa aagaatactt tgatcactat tttgctagtg taccttgac atacagtatg 8280  
 actgtaaagg ttgatataac acagattaaa gagaaggga tgaaattgta ccctgcaatg 8340  
 ctttattaca tagcaatgat agtaaacaga catagtgaat ttaggaccgc tatcaatcag 8400  
 gatggatgaac ttggaattta tgatgaaatg attccatcat atactatatt ccataatgac 8460

accgagacat tctcaagtct ttggactgaa tgcaagtcag attttaagtc atttcttgca 8520

gattatgaat ctgatactca aagatacggg aataatcacc gtatggaagg aaaacctaata 8580

gcacctgaga atatatttcaa tgtttccatg ataccttggg caacatttga cggatttaata 8640

ctgaatctgc aaaaaggcta cgattactta atccctatct ttacaatggg caagtattat 8700

aaggaagaca ataaaatcat ccttccccctt gcaatccagg tacatcatgc agtatgtgat 8760

ggatttcata ttgtcgttt tgtaaataaa ctgcaagaat taataaatc ctaactcgag 8820

ggcagtagcg cggtaggtccc acctgacccc atgccgaact cagaagtgaac acgataaaac 8880

gaaaggctca gtcgaaagac tgggcctttc gttttatctg ttgtttgtcg gtgaacgctc 8940

tcctgagtag gacaaatccg ccgggagcgg atttgaacgt tgcgaagcaa cggcccgag 9000

ggtggcgggc aggacgcccc ccataaactg ccaggcatca aattaagcag aaggccatcc 9060

tgacggatgg cttttttat tgtaaattcc ggtaacctt gtagcttagt gggaatttgt 9120

acctttatc gatacaaat cccgtaggc gctagggaca ctttttact cgttaaaaag 9180

ttttgagaat attttatatt ttgttcatg taatcactcc ttcttaatta caaattttta 9240

gcatctaatt taacttcaat tcctattata caaaatttta agatactgca ctatcaacac 9300

actcttaagt ttgcttctaa gtcttatttc cataacttct ttacgtttc cgggtacaat 9360

tcgtaatcat gtcatactg tttcctgtgt gaaattctta tccgtcaca attccacaca 9420

acatacgagc cggaagcata aagtgtaaag cctgggggtgc ctaatgagtg agctaactca 9480

cattaattgc gttgcgctca ctgcccgtt tccagtcggg aaacctgtcg tgccagaaaa 9540

cttcattttt aatttaaaag gatctagggtg aagatccttt ttgataatct catgacaaaa 9600

atcccttaac gtgagttttc gttccactga gctcagacc ccgtagaaaa gatcaaagga 9660

tcttcttgag atcccttttt tctgcgcgta atctgctgct tgcaaacaaa aaaaccaccg 9720

ctaccagcgg tggtttggtt gccggatcaa gagctaccaa ctctttttcc gaaggtaact 9780

ggcttcagca gagcgcagat accaaatact gtccttctag tgtagccgta gttaggccac 9840

cattcaaga actctgtagc accgcctaca tacctcgtc tgctaatect gttaccagtg 9900

gctgctgcca gtggcgataa gtcgtgtctt accgggttgg actcaagacg atagttaccg 9960

gataaggcgc agcggtcggg ctgaacgggg ggttcgtgca cacagcccag cttggagcga 10020

acgacctaca ccgaactgag atacctacag cgtgagctat gagaaagcgc cacgcttccc 10080

gaagggagaa agcgggacag gtatccggtg agcggcaggg tcggaacagg agagcgcacg 10140

agggagcttc cagggggaaa cgcttggtat ctttatagtc ctgtcgggtt tcgccacctc 10200

tgacttgagc gtcgattttt gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc 10260

agcaacgcgg cctttttacg gttcctggcc ttttgctggc cttttgctca catgttcttt 10320

cttgcgttat cccctgattc tgttgataac cgtattaccg cctttgagtg agctgatacc 10380

gctcgcgcga gccgaacgac cgagcgcagc gagtcagtga gcgaggaagc ggaagagcgc 10440

ccaatacgca aaccgcctct ccccgcgctt tggccgattc attaatagcag ctggcacgac 10500

aggtttcccg actggaaaagc gggcagttag cgcaacgcaa ttaatgtgag ttagctcact 10560

cattaggcac ccaggtctt acactttatg ctccggctc gtatgttggtg tggaattgtg 10620

agcggataac aatttcacac aggaacacgc tatgacatg attac 10665

<210> 31

<211> 10665

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<220>

<221> modified\_base

<222> (1522)..(1522)

<223> a, c, g or t

<400> 31

gaattcgagc tcggtaccgc gggatcctct agagtcgacc tgcaggcatg caagcttggc 60

actggccgctc gttttacaac gtcgtgactg ggaaaaccct ggcgttacc aacttaatcg 120

ccttgacgca catccccctt tcgccagctg gcgtaatagc gaagaggccc gcaccgcgta 180

gaggatcgag atctcgatcc cgcgaaatta atacgactca ctatagggga attgtgagcg 240

gataacaatt cccctctagg ctcataactt cacgctcctg tatatatttt tatttattta 300

aaaatgagtc aaaatttagg aaatgattgc aatatgtata atatccaaat ttccattcaa 360

ataaccaaag taattttacc tctttttatg agctatttca atactttgtt agtaaattaa 420

catatatgag ctgtatatgg tttaatgaaa aaagttatit tgaaggata ttgtaaaaaa 480

cataatatat tatatggata aattttacat ttgacttata atagttaat atatgtaata 540

tgaatagcta atctaagcag gctactgcct agaaaaaagc ttataattat ccttagctct 600

cttcaatgtg cgcccagata ggggtgtaag tcaagtagtt taaggtaata ctctgtaaga 660

taacacagaa aacagccaac ctaaccgaaa agcgaaagct gatacgggaa cagagcacgg 720

ttggaaagcg atgagttacc taaagacaat cgggtacgac tgagtcgcaa tgtaaatcag 780

atataaggta taagttgtgt ttactgaacg caagtttcta atttcgatta gagctcgata 840



gaggaaagtg tctgaaacct ctagtacaaa gaaaggttaag ttatcattga agacttatct 900

gttatcacca catttgtaca atctgttaga gaacctatgg gaacgaaacg aaagcgatgc 960

cgagaatctg aatttaccaa gacttaacac taactgggga taccctaaac aagaatgcct 1020

aatagaaagg aggaaaaagg ctatagcact agagcttgaa aatcttgcaa gggtacggag 1080

tactcgtagt agtctgagaa gggtaacgcc cttacatgg caaaggggta cagttattgt 1140

glactaaaat taaaaattga ttagggagga aaacctcaaa atgaaaccaa caatggcaat 1200

tttagaaaga atcagtaaaa attcacaaga aaatatagac gaagttttta caagacttta 1260

tcgttatctt ttacgtccag atatttatta cgtggcgacg cgttgggaaa tggcaatgat 1320

agcgaaacaa cgtaaaactc ttgttgtatg ctttcattgt catcgtcacg tgattcataa 1380

acacaagtga atgtcgacag tgaattttta cgaacgaaca ataacagagc cgtatactcc 1440

gagaggggta cgtacgggtc ccgaagaggg tggtgcaaac cagtcacagt aatgtgaaca 1500

aggcggtacc tccctacttc ancatatcat tttctgcage cccctagaaa taattttgtt 1560

taactttaag aaggagatat acatatatgg ctagatcgtc cattccgaca gcatcgccag 1620

tcactatggc gtgctgctag cgctatatgc gttgatgcaa tttctatgca ctcgtagtag 1680

tctgagaagg gtaacgcctt ttacatggca aaggggtaca gttatttgtt actaaaatta 1740

aaaattgatt agggaggaaa acctcaaat gaaaccaaca atggcaattt tagaaagaat 1800

cagtaaaaat tcacaagaaa atatagacga agtttttaca agactttatc gttatctttt 1860

acgtccagat atttattacg tggcgatatc aaatttatat tccaataaag gagcttccac 1920

aaaaggaata ttagatgata cagcggatgg ctttagtgaa gaaaaataa aaaagattat 1980

tcaatcttta aaagacggaa cttactatcc tcaacctgta cgaagaatgt atattgcaaa 2040

aaagaattct aaaaagatga gacctttagg aattccaact ttcacagata aattgatcca 2100  
 agaagctgtg agaataattc ttgaatctat ctatgaaccg gtattcgaag atgtgtctca 2160  
 cggttttaga ccicaacgaa gctgtcacac agctttgaaa acaatcaaaa gagagtttgg 2220  
 cggcgcaaga tggtttgtgg agggagatat aaaaggctgc ttcgataata tagaccacgt 2280  
 tacactcatt ggactcatca atcttaaaat caaagatatg aaaatgagcc aattgattta 2340  
 taaatttcta aaagcaggtt atctggaaaa ctggcagtat cacaaaactt acagcggaac 2400  
 acctcaaggt ggaattctat ctctctttt ggccaacatc tatcttcacg aattggataa 2460  
 gtttgtttta caactcaaaa tgaagtttga ccgagaaaagt ccagaaagaa taacacctga 2520  
 atatcgggag ctccacaatg agataaaaag aatttctcac cgtctcaaga agttggaggg 2580  
 tgaagaaaaa gctaaagttc ttttagaata tcaagaaaaa cgtaaaagat tacccacact 2640  
 cccctgtacc tcacagacaa ataaagtatt gaaatacgtc cggtatgcgg acgacttcat 2700  
 tatctctgtt aaaggaagca aagaggactg tcaatggata aaagaacaat taaaactttt 2760  
 tattcataac aagctaaaaa tggaattgag tgaagaaaaa acactcatca cacatagcag 2820  
 tcaaccgcgt cgttttctgg gatatgatat acgagtaagg agatctggaa cgataaacg 2880  
 atctggtaaa gtcaaaaaga gaacactcaa tgggagtgtga gaactcctta ttctcttca 2940  
 agacaaaatt cgtcaattta tttttgacaa gaaaatagct atccaaaaga aagatagctc 3000  
 atggtttcca gttcacagga aatatcttat tcgttcaaca gacttagaaa tcatacaat 3060  
 ttataattct gaactccgcg ggatttgtaa ttactacggt ctagcaagta attttaacca 3120  
 gctcaattat tttgcttacc ttatggaata cagctgtcta aaaacgatag cctccaaaca 3180

taaggaaca ctttcaaaaa ccatttccat gtttaaagat ggaagtgggt cgtgggggat 3240

cccgtatgag ataaagcaag gtaagcagcg ccgttatitt gcaaatttta gtgaatgtaa 3300

atccccittat caatttacgg atgagataag tcaagctcct gtattgtatg gctatgcccc 3360

gaatactctt gaaaacaggt taaaagctaa atgttgtgaa ttaatgtggga cgtctgatga 3420

aaatacttcc tatgaaattc accatgtcaa taaggtcaaa aatcttaaag gcaaagaaaa 3480

atgggaaatg gcaatgatag cgaaacaacg taaaactctt gttgtatgct ttcattgtca 3540

tcgtcacgtg attcataaac acaagtgaga tatctcgagc acccgttctc ggagcactgt 3600

ccgaccgctt tggccgccgc ccagtcctgc tcgcttcgct acttgagacc actatcgact 3660

acgcgatcat ggcgaccaca cccgtcctgt ggatcgccaa gctcgccgat ggtagtgtgg 3720

ggtctcccca tgcgagagta gggaactgcc aggcatacaa taaaacgaaa ggctcagtcg 3780

aaagactggg ctttcgttt tatctgttgt ttgtcggtga acgctctcct gagtaggaca 3840

aatccgccgg gagcggattt gaacgttgcg aagcaacggc ccggaggggtg gcgggcagga 3900

cgcccccat aaactgccag gcatcaaatt aagcagaagg ccatcctgac ggatggcctt 3960

tttgcgtttc taaaactct tctgtctgc atatctacaa gccatccgc ctttccaac 4020

agttgcgcag ccigaatggc gaatggcgcc tgatcggtta ttttctcctt acgcatctgt 4080

gcggtatttc acaccgata tgggtgactc tcagtacaat ctgctctgat gccgcatagt 4140

taagccagcc ccgacaccgc ccaacaccgc ctgacgcgcc ctgacgggct tgtctgctcc 4200

cggcatccgc ttacagacaa gctgtgaccg tctccgggag ctgcatgtgt cagaggtttt 4260

caccgtcatc accgaaacgc gcgagacgaa agggcctcgt gatacgcccta tttttatagg 4320

ttaatgtcat gataataatg gtttcttaga cgtcaggtgg cacttttcgg ggaaatgtgc 4380

gcggaacccc tatttgttta tttttctaaa tacattcaaa tatgtatccg ctcatgagac 4440

aaataaccctg ataaatgctt caataatatt gaaaaaggaa gagtatgagt attcaacatt 4500

tccgtgtcgc ccttatccccc ttttttgcgg cattttgcct tccgtttttt gctcacccag 4560

aaacgctggt gaaagtaaaa gatgctgaag atcagttggg tgcacgagtg ggttacatcg 4620

aactggatct caacagcggg aagatccttg agagttttcg cccgaagaa cgttttccaa 4680

tgatgagcac ttttaaatta aaaaatgaagt tttaaaactt catttttaat ttaaattaaa 4740

aatgaagttt tatcaaaaaa atttccaata atcccactct aagccacaaa cagccctat 4800

aaaatcccgc tttaatccca ctttgagaca catgtaatat tactttacgc cctagtatag 4860

tgataatfff ttacattcaa tgccacgcaa aaaaataaag gggcactata ataaaagttc 4920

cttcggaact aactaaagta aaaaattatc tttacaacct ccccaaaaaa aagaacaggt 4980

acaaagtacc ctataataca agcgtaaaaa aatgagggtg aaaataaaaa aataaaaaaa 5040

taaaaaaata aaaaaataaa aaaaataaaa aaataaaaaa ataaaaaaat aaaaaataa 5100

aaaaataaaa aaataaaaaa ataaaaaaat ataaaaataa aaaaatataa aaataaaaaa 5160

atataaaaaat aaaaaaatat aaaaataaaa aaataaaaaa atataaaaaat aaaaaataa 5220

aaaaatataa aaatatffff tatttaaagt ttgaaaaaaa tttttttata ttatataatc 5280

tttgaagaaa agaataataa aaatgagcct ttataaaagc ccattttttt tcatatacgt 5340

aatatgacgt tctaagtfff ttattggtac ttctaacatt agagtaattt ctttattttt 5400

aaagcctfff tcttaagggt cttttatttt ttttcttaat acatttaatt cctctttttt 5460

tgttgctfff cctttagctt ttaattgctc ttgataatff tttttacctc taatatfftc 5520

tcttctctta tattcctttt tagaaattat tatgtcata tatttttgtt ctctctctgt 5580

aatttcta atactataa gagtttcatt ctatactta tatgtctat tttatctaa 5640

ataacatctt tcagcacttc tagttgctct tataacttct ctctcactta aatgttgtct 5700

aaacatacta ttaagtctta aaacatcatt taatgccttc tcaatgtctt ctgtaaagct 5760

acaaagataa tatctatata aaaataatat aagctctctg tgcctttta aatcatattc 5820

tcttagttca caaagtttta ttatgtcttg tattcttcca taatataaac ttctttctct 5880

ataaatataa tttattttgc ttggctctacc ctttttcctt tcatatgggt ttaattcagg 5940

taaaaatcca ttttgtatct ctcttaagtc ataaatatat tegtactcat ctaatatatt 6000

gactactgtt tttgatttag agtttatact tcctggaact cttaatatcc tggttgcatc 6060

taaggttgt ctatctgctc caaagtattt taattgatta tataaatatt ctgaaccgc 6120

ttcccataat ggtaatgctt tactaggtac tgcatttatt atccatatta aatacatcc 6180

tcttccacta tctattacat agtttggtat aggaatactt tgattaaaat aattcttttc 6240

taagtccatt aatacctggt ctttagtttt gccagtttta taataatcca agtctataaa 6300

cagtgtatct aactctttta tattttctaa tcgcctacac ggcttataaa aggtatttag 6360

agttatatag atattttcat cactcataac taaatctttt aattcagcgt atttatagt 6420

ccattggcta tatccttttt tatctataac gctcctgggt atccaccctt tacttctact 6480

atgaatatta tctatatagt tctttttatt cagctttaat gcgtttctca ctattcacc 6540

tcccctctg taaaactaag aaaattatat catattttca ataattatta actattctta 6600

aactctta ataaaaataga gtaagtcctc aatgaaact taatctatct tttatgtttt 6660

aatttattat ttttattaaa atattttaaa ctaaaataaa tgattctttt taatttttta 6720

ctatttcatt ccataatata ttactataat tatttacaaa taatatttct tcatttgtaa 6780

tatttagatg atttactaat tttagttttt atatattaaa taattaatgt ataatttata 6840

taaaaaatca aaggagctta taaattatga ttatttccaa agatactaaa gatttaattt 6900

tttcaatttt aacaatactt tttgtaatat tatgtttaaa ttttaattgta ttttttcat 6960

ataataaagc cgttgaagta aaccaatcca ttttccttat gatgttatta ttaaatttaa 7020

gttttataat aatatcttta ttatatttat tgtttttaaa aaaactagtg aaatttccgg 7080

ctttattaaa cttattttta ggaattttat tttcattttc atctttacag gatttgatta 7140

tatctttaaa tatgttttat caaatattat ctttttctaa atttatatat atttttatta 7200

tatttattat tatatatatt ttatttttaa gtttctttct aacagctatt aaaaagaaac 7260

ttaaaaataa aaacacgtac tctaaaccaa taaataaaac tatttttatt attgctgcct 7320

tgattggaat agtttttagt aaaattaatt tcaatatcc acaatattat attataagct 7380

agctttgcat tgiacttttc aatcgcttca cgaatgcggg tatctccgaa agataaagtc 7440

ttttcatctt ccttgatgaa gataagattt tctccgtctc cgccggcaga attgaagcgg 7500

ggtactacgg tatcgtctgc gtcattttcc gttgtctgat agatgatagt cataggctca 7560

ttttcttccg tttcggtaaa ggggataggt tcgccctttg agagcagggc ggcgatggaa 7620

agcattaaact tgcttttccc atcgcccgga tctccctgca atagcgtaac ttgccaac 7680

ggaatatacg gataccacag ccactttact tctttcggt cgatttcact tgccttgatg 7740

atttcaagag gtacgtgaa attcatttcg ttttcattha gtttcatttt ttcttgttct 7800

ccttttctct gaaaatataa aaaccacaga ttgatactaa aaccttggtt gtgttgcttt 7860

tcggggctta aatcaaggaa aaatccttgt ttttagcctt tcaaaaagaa acacaaggtc 7920

tttgtactaa cctgtgggta tgtataaaat tgtagatttt agggttaacaa aaaacaccgt 7980

atttctacga tgtttttgct taaatacttg tttttagtta cagacaaacc tgaagttgaa 8040

ttcatattta ttaaattag cgtatatact attgaaaatg tttttgaaat attataaaat 8100

taactttggt ttaggaaaaa taaccagttc ttttgtcgat aagcattaat ttgcttgact 8160

aattaataaa aaacttagga ggtaacacta atggtatctg agaaaattga caagaacagt 8220

tggaacagaa aagaatactt tgatcactat tttgctagtg taccttgacac atacagtatg 8280

actgtaaagg ttgatataac acagattaaa gagaaggga tgaattgta ccctgcaatg 8340

ctttattaca tagcaatgat agtaaacaga catagtgaat ttaggaccgc tatcaatcag 8400

gatggatgaac ttggaattta tgatgaaatg attccatcat atactatatt ccataatgac 8460

accgagacat tctcaagtct ttggactgaa tgcaagtcag attttaagtc atttcttgca 8520

gattatgaat ctgatactca aagatacggg aataatcacc gtatggaagg aaaacctaata 8580

gcacctgaga atattttcaa tgtttccatg ataccttggg caacatttga cggatttaata 8640

ctgaatctgc aaaaaggcta cgattactta atccctatct ttacaatggg caagtattat 8700

aaggaagaca ataaatcat ccttccccct gcaatccagg tacatcatgc agtatgtgat 8760

ggatttcata ttgtcggtt tgtaaatgaa ctgcaagaat taataaatc ctaactcgag 8820

ggcagtagcg cgggtgggtccc acctgacccc atgccgaact cagaagtga acgataaaac 8880

gaaaggctca gtcgaaagac tgggcctttc gttttatctg ttgtttgtcg gtgaacgctc 8940

tcctgagtag gacaaatccg ccgggagcgg atttgaacgt tgcgaagcaa cggccccggag 9000

ggtggcgggc aggacgcccc ccataaactg ccaggcatca aattaagcag aaggccatcc 9060

tgacggatgg cttttttat tgtaaattcc ggtaaccctt gtagcttagt gggaatttgt 9120  
 accccttacc gatacaaat ccccgtaggc gctagggaca ctttttact cgttaaaaag 9180  
 ttttgagaat attttatatt ttgttcatg taatcactcc ttcttaatta caaatittta 9240  
 gcatctaatt taacttcaat tcctattata caaaatttta agatactgca ctatcaacac 9300  
 actcttaagt ttgcttctaa gtcttatttc cataacttct ttacgtttc cgggtacaat 9360  
 tcgtaatcat gtcatactg tttcctgtgt gaaattctta tccgtcaca attccacaca 9420  
 acatacgagc cggaagcata aagtgtaaag cctgggggtgc ctaatgagtg agctaactca 9480  
 cattaattgc gttgcgtca ctgcccgtt tccagtcggg aaacctgtcg tgccagaaaa 9540  
 ctcatTTTT aatttaaaag gatctagggtg aagatccttt ttgataatct catgaccaa 9600  
 atccctaac gtgagtttc gtccactga gcgtcagacc ccgtagaaaa gatcaaagga 9660  
 tcttcttag atccttttt tctgcgcgt atctgtgtg tgcaaaaaa aaaaccaccg 9720  
 ctaccagcgg tggtttgtt gccggtacaa gagctacaa ctctttttcc gaaggtaact 9780  
 ggcttcagca gagcgcagat accaaatact gtccttctag ttagaccgta gtaggccac 9840  
 cacttaaga actctgtac accgcctaca tacctcgtc tgctaactct gttaccagt 9900  
 gctgtgcca gtggcgataa gtcgtgtct accgggttg actcaagac atagttaccg 9960  
 gataaggcgc agcggtcggg ctgaacgggg ggttcgtgca cacagcccag ctggagcga 10020  
 acgacctaca ccgaactgag atacctacag cgtgagctat gagaaagcgc cacgttccc 10080  
 gaaggagaa agcggacag gtatccgta agcggcagg tcggaacagg agagcgacg 10140  
 agggagcttc caggggaaa cgctggtat ctttatagtc ctgtcgggt tcgccacctc 10200



tgacttgagc gtcgattttt gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc 10260

agcaacgcgg cctttttacg gtctctggcc ttttgcctgc cttttgtca catgttcttt 10320

cctgcgttat cccctgattc tgttgataac cgtattaccg cctttgagtg agctgatacc 10380

gctcgccgca gccgaacgac cgagcgcagc gagtcagtga gcgaggaagc ggaagagcgc 10440

ccaatacgca aaccgcctct ccccgcgctt tggccgattc attaatagcag ctggcacgac 10500

aggtttcccg actggaaagc gggcagtgag cgcaacgcaa ttaatgtgag ttagctcact 10560

cattaggcac ccaggtctt acactttatg cttccgctc gtatgttgtg tggaattgtg 10620

agcgataac aatttcacac aggaacagc tatgacatg attac 10665

<210> 32  
 <211> 1056  
 <212> DNA  
 <213> Thermoanaerobacterium saccharolyticum

<400> 32  
 gtgtatacaa tatatttctt cttagtaaga ggaatgtata aaaataaata ttttaaagga 60

agggacgac ttatgagcat tattcaaac atcattgaaa aagctaaaag cgataaaaag 120

aaaattgttc tgccagaagg tgcagaacc aggacattaa aagctgctga aatagtitta 180

aaagaaggga ttgcagattt agtgcttctt ggaaatgaag atgagataag aaatgctgca 240

aaagacttgg acatatccaa agctgaaatc attgacctg taaagtctga aatgtttgat 300

aggtatgcta atgatttcta tgagttaagg aagaacaaag gaatcacgtt ggaaaaagcc 360

agagaaacaa tcaaggataa tatctatttt ggatgtatga tggttaaaga aggttatgct 420

gatggattgg tatctggcgc tattcatgct actgcagatt tattaagacc tgcatttcag 480

ataattaaaa cggtccagg agcaaagata gtatcaagct tttttataat ggaagtcct 540

aattgtgaat atggtgaaaa tgggtgtattc ttgtttgctg attgtgcggt caacccatcg 600  
 cctaattgcag aagaacttgc ttctattgcc gtacaatctg ctaatactgc aaagaatttg 660  
 ttgggctttg aaccaaagt tgccatgcta tcattttcta caaaaggtag tgcacacat 720  
 gaattagtag ataaagtaag aaaagcgaca gagatagcaa aagaattgat gccagatgtt 780  
 gctatcgacg gtgaattgca attggatgct gctcttgta aagaattgc agagctaaaa 840  
 gcgccgggaa gcaaagttgc gggatgtgca aatgtgctta tattccctga ttacaagct 900  
 ggtaatatag gatataagct tgtacagagg ttagctaagg caaatgcaat tggacctata 960  
 acacaaggaa tgggtgcacc ggттаатgat ttatcaagag gatgcagcta tagagatatt 1020  
 gttgacgtaa tagcaacaac agctgtgcag gctcaa 1056

<210> 33  
 <211> 1209  
 <212> DNA  
 <213> Thermoanaerobacterium saccharolyticum

<400> 33  
 atgaaaatta tgaaaatact ggттаттаат tgcggaagtt cttcgctaaa atatcaactg 60  
 attgaatcaa ctgatggaaa tgtgttggca aaaggccttg ctgaaagaat cggcataaat 120  
 gattccatgt tgacacataa tgctaacgga gaaaaaatca agataaaaaa agacatgaaa 180  
 gatcacaag acgcaataaa attggtttta gatgctttgg taaacagtga ctacggcgtt 240  
 ataaaagata tgtctgagat agatgctgta ggacatagag ttgttcacgg aggagaatct 300  
 ttacatcat cagtтссат ааатgatgaa gtgttaaaag cgataacaga ttgcatagaa 360  
 ttagctccac tgcacaatcc tgctaataata gaaggaatta aagcttgcca gcaaatcatg 420

ccaaacgttc caatggtggc ggtatttgat acagcctttc atcagacaat gcctgattat 480

gcataicttt atccaatacc ttatgaatac tacacaaagt acaggattag aagatatgga 540

tttcatggca catcgcataa atatgtttca aatagggtcg cagagatttt gaataaacct 600

attgaagatt tgaaaatcat aacttgtcat cttggaaatg gctccagcat tgctgctgtc 660

aaatatggta aatcaattga cacaagcatg ggatttacac cattagaagg tttggctatg 720

ggtacacgat ctggaagcat agacccatcc atcatttcgt atcttatgga aaaagaaaat 780

ataagcgctg aagaagtagt aaatatatta aataaaaaat ctggtgttta cggtatattca 840

ggaataagca gcgatttttag agacttagaa gatgccgcct ttaaaaatgg agatgaaaga 900

gctcagttgg ctttaaatgt gtttgcatac cgagtaaaga agacgattgg cgcttatgca 960

gcagctatgg gaggcgtcga tgtcattgta ttacagcag gtgttggtga aaatgttcct 1020

gagatacgag aatttatact tgatggatta gagtttttag ggttcagctt ggataaagaa 1080

aaaaataaag tcagaggaag agaaactatt atatctacgc cgaattcaaa agttagcgtg 1140

atggttgtgc ctactaatga agaatacatg attgctaaag atactgaaaa gattgtaaag 1200

agtataaaa 1209

<210> 34  
 <211> 1358  
 <212> DNA  
 <213> Thermoanaerobacter pseudoethanolicus

<400> 34  
 gctaagtcta tcggaccaat ttctcaaggt cttgcaaac ctatcaatga cttgtcaaga 60

ggtttagtagt tagaagatat tgttaatgtt atagcaataa cttgtgtaca agctcaaggg 120

gtgcaaaaat aactttgagg aggcagcgat tatgaaaatt ttagtcatga actgtggaag 180

ctcgtcatta aaagtatcaa ttgtagata tggataatgg gaaagtgcta gcgaaaggat 240

tggcggaag gataggtatc aatgatttct ttttaactca tcaagtagag ggcaaagata 300

aaataaaaat acaaaaagat atgaaaaatc ataaagaagc tatacaaatt gtttagagg 360

cttagtaga taaagaaatc ggaatattaa aagatatgaa agaaatagat gcagtaggac 420

atagagtgtg gcacggggga gagtttttta ctgattccgt attgattgac gatgaggtaa 480

tcaaaaaatt agaagcatgt attgaccttg cacctttgca caatcctgct aatattgagg 540

gaataaaagc ttgtcggcag ataatgccag ggggtccaat ggtagcagtt ttgatacgg 600

ctttccatca aacaatgcca gattatgcgt atatttatcc cattccttat gaatactacg 660

aaaaatatag aataagaaga tatggattcc atgggacttc tcataaatat gtatctttaa 720

gagctgctga aatattaaag aggcctattg aagagttaaa aattattact tgccatttag 780

ggaatgggtc tagtattgct gcggttaaag gcggttaagtc gatagataca agtatgggat 840

ttactccatt agaagggctg gctatgggia caaggtccgg aaatgttgat ccttcaatta 900

taactttctt aatggaaaaa gaaggattga ctgcagaaca gggtatagat atacttaata 960

agaaatcagg tgtatacggg atttcaggaa taagtaatga ctttagagat atagaaatg 1020

cagcttttaa agaagggcat aaaagggcta tgttggcatt aaaagtttct gcttataggg 1080

tgaaaaagac aataggttct tatacagctg ctatgggtgg ggttgatgta attgtgttta 1140

ctgctggagt tggagaaaat ggaccagaaa tgagagagtt tattttagag gatctagagt 1200

ttttaggctt taaactggac aaagagaaga ataagtaag aggaaaagag gaaattatat 1260

ctacagaaga ttcaaaagt ttatgtatgg ttattcctac aaatgaagaa tatatgattg 1320

ctaaagatac tgaaaaattg gtaaaaggtt taaagtag 1358

<210> 35  
 <211> 1196  
 <212> DNA  
 <213> Thermoanaerobacter pseudoethanolicus

<400> 35  
 atggcagtaa tggatagtat catacaaaag gctaaagcta ataaaaaaag gatttgtgctt 60  
 cctgagggaa gtgaagctcg aactttaaaa gctgctgaaa aggttattaa agaaggtatt 120  
 gctgatgtag ttttattagg gaaggaagaa gaaataaaag aaaaagcaaa gggattggat 180  
 atctcgaaag cagaaattat agacctgaa aagtcgcctc ttttcaaaa atatgctgaa 240  
 gaatattata atttgagaaa aaccaaagga gttacagaag aacaggcata tcaaattatg 300  
 aaagacccta tgtactatgg gtgcatgatg gtcaaattag acgatgttga tggatatgta 360  
 tctggggcga ttcacgtac tgctgatggt ttcagaccgg cttttcaaat tgtaaaaact 420  
 gctgcagggtg tcaaagtagt atccagcgcc tttataatgg aagtiaccta tigtacttat 480  
 ggaagcgatg gatlatttat ttttctgat tgtgcaataa atcctaattc taatgaagag 540  
 gaattagcag caattgccat tgcttctgcc catactgcaa aagtccttgc tggaaattgag 600  
 cctagaattg ctatgctgtc attttctact aaaggaagtg caaacatga attagtagat 660  
 aagtgaaaa atgcgactaa aatcgcaaaa gaattggcgc ctgatttgct aattgatggt 720  
 gagcttcaat tagatgctgc gattgtcaaa gaagtaggag agttaaggc tccaggaagt 780  
 cctgtagcgg ggaatgcaaa tgtgcttatt ttcccagatt tgcaagcggg aaacattgga 840  
 tataagctag tgcaaagact tgctaaagct aatgctatcg gaccaatttc tcaaggtctt 900  
 gcaaaaccta tcaatgactt gtcaagaggt tgtagttag aagatattgt taatgttata 960

gcaataactt gtgtacaagc tcaaggggtg caaaaataac tttgaggagg cagcgattat 1020

gaaaatttta gtcatgaact gtggaagctc gtcattaaaa gtatcaattg ttagatatgg 1080

ataatgggaa agtgctagcg aaaggattgg cggaaaggat aggtatcaat gattctcttt 1140

taactcatca agtagagggc aaagataaaa taaaaataca aaaagatatg aaaaat 1196

<210> 36

<211> 1053

<212> DNA

<213> Thermoanaerobacter sp.

<400> 36

gtgtatacaa tatatttctt ctttttagta agaggaatgt ataaaaataa atattttaaa 60

ggaagggacg atcttatgag cattattcaa aacatcattg aaaaagctaa aagtgataaa 120

aagaaaattg ttctgccgga aggtgcagaa ccagaaacat taaaagctgc tgaaatagtt 180

ttaaaagaag gaattgcaga tttggtgctt cttggaatg aagatgagat aagaaatgct 240

gcaaaagact tggacatatc taaagctgaa atcattgatc ctgtaaaatc tgaaatgttt 300

gataggtatg ctaatgattt ttatgagtta aggaagagca aaggaatcac gttggaaaaa 360

gccagagaaa caatcaagga taatatctat tttggatgta tgatggttaa agaaggttat 420

gctgatggat tggatatctg cgctattcat gctactgcag atttattaag acctgcattt 480

cagataatta aaacggctcc aggagcaaag atagtatcaa gcttttttat aatggaagtg 540

cctaattgtg aatatggiga aaatgggtga ttcttgtttg ctgattgcgc ggtcaaccca 600

tcgcctaagc cagaagaact tgcttctatt gctgtacaat ctgctaatac tgcaaagaat 660

ttgttgggct ttgaaccaa agttgctatg ctatcathtt ccacaaaagg tagtgcatca 720

catgaattag tagataaagt aagaaaagcg acagaaatag caaaagaatt gatgccagat 780  
gttgctatcg acggtgaatt gcaattggat gctgctcttg tcaaagaagt tgcagagcta 840  
aaagcgccag gaagcaaagt tgcgggatgt gcaaatgtgc ttatattccc tgatttaciaa 900  
gctggtaata taggatataa gcttgtacag agattagcta gcaaatgcaa ttggacctat 960  
aacacaggaa tgggtgcacc ggttaatgat ttatcaagag gatgcagcta tagagatatt 1020  
gttgacgtaa tagcacacag ctgtacaggc tca 1053

<210> 37  
<211> 1068  
<212> DNA  
<213> Thermoanaerobacter sp.

<400> 37  
atgctaacgg agaaaaatca agataaaaaa agacatgaaa gatcacaaag acgcaataaa 60  
attgttttag atgctttggt aagcagtgac tacggcgtaa taaaggatat gtctgagata 120  
gatgctgtag gacatagagt tggtcacgga ggagaatctt ttacatcatc agtttcata 180  
aatgatgatg tgtaaaagc gataacagat tgcatagaat tagctccact gcacaatcct 240  
gccaatatag aaggaattaa agcttgccag caaatcatgc caaacgttcc aatggtggcg 300  
gtatttgata cagcctttca tcagacaatg cctgattatg catacttta tccaatacct 360  
tatgaatact acacaaagta caggatcaga agatatggat ttcatggcac atgcataaa 420  
tatgtttcaa atagggtgc agagatttta aataaaccta ttgaagattt gaaaatcata 480  
acttgctatc ttggaaatgg ctccagcatt gctgctgtca aatatggtaa atcaattgac 540  
acaagcatgg gatttacacc attagaaggt ttggctatgg gtacacgatc tggaagcata 600  
gacccatcca ttatttcgta tcttatggaa aaagaaaata taagcgctga agaagtagta 660

aatatattaa ataaaaaatc tgggtgtttac ggtatttcag gaataagcag cgattttaga 720  
gacttagaag atgccgcctt taaaaatgga gatgaaagag ctacagttggc tttaaatgtg 780  
tttgcataac gagtaaagaa gatgattggc gcttatgcag cagctatggg aggcgtcgat 840  
gccattgtat ttacagcagg tgttggtgaa aatggctctg agatacgaga atttatactt 900  
gatggattag agttcttagg gttcagcttg gataaagaaa aaaataaagt cagaggaaaa 960  
gaaactatta tatctacgcc gaattcaaaa gttagcgtga tggttgtgcc cactaatgaa 1020  
gaatacatga ttgctaaaga tactgaaaag attgtaaaga gtataaaa 1068

<210> 38  
<211> 1059  
<212> DNA  
<213> Thermoanaerobacterium saccharolyticum

<400> 38  
gtgtatacaa tatatttctt ctttttagta agaggaatgt ataaaaataa atattttaaa 60  
ggaagggatg atcttatgag cattattcag aacatcattg aaaaagctaa aagcgataaa 120  
aagaaaattg tttgccaga aggtgcagaa cccaggacat taaaagctgc tgaaatagtt 180  
ttaaagaag gaattgcaga tttggtgctt cttggaaatg aagatgagat aagaaatgca 240  
gcaaaagact tggacataac caaagctgaa ataattgacc ctgtaaaatc tgaaatgttt 300  
gataggtatg ctaatgatit ttacgaatta agaaagagca agggaatcac attggaaaaa 360  
gccagagaaa caatcaagga taatatctat ttggatgta tgatggttaa agaaggttat 420  
gctgatggat tagtatctgg cgctattcat gctactgcag atttattaag acctgcattt 480  
cagataatta aaacagctcc aggagcaaag atagtatcaa gcttttttat aatggaagtg 540



cctaattgtg aatatgggtga aaatggcgta ttcttgtttg ctgattgtgc ggtcaatcca 600

tcacctaatag cagaagaact tgcttctatt gctgtacaat ctgctaatac tgcaaagaat 660

ttgttgggtt ttgaacaaa agttgccatg ctatcathtt ccacaaaagg tagtgcatca 720

catgaattag tagacaaggt aagaaaagcg acagagatag caaaggatht gatgccagat 780

gttgctatcg atggtgaatt gcaactggat gctgctattg ttaaagaagt tgcagagcta 840

aaagcaccgg gaagcaaagt tgcgggatgt gcaaatgtgc ttatattccc tgacttacia 900

gctggttaata taggatafaa gcttgtacag agattagcta aggcaaatgc aattggaccg 960

ataacgaag gaatgggtgc accagttaat gatttatcaa gaggatgcag ctataaagat 1020

attgttgacg taatagcgac aacagctgtg caggctcaa 1059

<210> 39  
 <211> 1209  
 <212> DNA  
 <213> Thermoanaerobacterium saccharolyticum

<400> 39  
 atgaaaacta tgaaattct gggtattaat tgtggaagtt cttactaaa atatcaattg 60

attgaatcaa ttgatgaaa tgtgtggca aaaggccttg ctgaaagaat cggcataaat 120

gattccctgt tgacgataa tgctaacgga gaaaaaatca agataaaaaa agacatgaaa 180

gatcaciaag acgcaataaa attggtttta gatgcttttg taagtagcga ctacggcgtt 240

ataaaggata tgtctgagat agatgctgta ggacatagag ttgttcatgg aggagagtct 300

tttacctat cagttcttat aaatgatgaa gtgttaaagg caataacaga ttgtatagaa 360

ttagctccac tgcataatcc tgctaataata gaaggaatta aagcttgcca gcaaatcatg 420

ccaaacgttc caatgggtgc ggtatttgat acagcctttc atcaaacat gcctgattat 480

gcataatcttt atccaatacc ttatgagtag tacacaaagt acaggatcag aagatatgga 540

tttcatggca cgtcgcataa atatgtttca agtagggctg cagagatttt gaataaacct 600

attgaagatt tgaaaatcat aacttgtagt cttggaaatg gctccagtat tgctgccgtc 660

aaatatggta aatcaattga cacaagcatg ggattttacac cattagaagg tttggctatg 720

ggtacacgat ctggaagtag agacccatcc atcatttctt atcttatgga aaaagaaaat 780

ataagtgtg aagaggtagt aaatatatta aataaaaaat ctggtgttta cggtatttcg 840

ggaataagca gcgattttag agatttagaa gatgctgcct ttaaaaatgg agatgaaaga 900

gctcagttgg ccttaaatgt gtttgcatac cgagtaaaga agacgattgg agcttatgca 960

gcagctatgg gaggcgttga tgctattgta ttacggcag gtgttggtga aaatgggcct 1020

gagataagag aatttatact tgatggattg gatttcttag ggttcagctt ggataaagaa 1080

aaaaataaag tcagaggaaa ggaaactatt atatctacgc caaattcaaa aattagcgtg 1140

atggtttgtc cgactaatga agaataatg attgctaaag atactgaaaa gattgtaaag 1200

agtataaaa 1209

<210> 40  
 <211> 933  
 <212> DNA  
 <213> Thermoanaerobacterium saccharolyticum

<400> 40  
 atgagcaagg tagcaataat aggatctggt tttgtaggtg caacatcggc atttacgctg 60

gcattaagtg ggactgtgac agatatcgtg ctggtggatt taaacaagga caaggtata 120

ggcgaatgac tggacataag ccatggcata cgcctaatac agcctgtaaa tgtgtatgca 180

ggtgactaca aagatgtgaa aggcgcagat gtaatagttg tgacagcagg tgctgctcaa 240

aagccgggag agacacggct tgaccttgta aagaaaaata cagccatatt taagtccatg 300

atactgagc ttttaaagta caatgacaag gccatatatt tgattgtgac aaatcccgtg 360

gataactga cgtacgttac atacaagatt tctggacttc catggggcag agtttttgg 420

tctggcaccg ttcttgacag ctcaagggtt agataccttt taagcaagca ctgcaatata 480

gatccgagaa atgtccacgg aaggataatc ggcgagcatg gtgacacaga gtttgacga 540

tggagcataa caaacatatac gggatatatca tttaatgagt actgcagcat atgcggacgc 600

gtctgcaaca caaatctcag aaaggaagta gaagaagaag tcgtaaatgc tgcttacaag 660

ataatagaca aaaaagggtc tacatactat gctgtggcag ttgcagtaag aaggattgtg 720

gagtgcattc taagagatga aaattccatc ctacacgtat catctccatt aaatggacag 780

tacggcgtga aagatgtttc attaaagctt ccatctatcg taggcaggaa tggcgttgcc 840

aggattttgg acttgccctt atctgacgaa gaagtggaga agtttaggca ttcagcaagt 900

gtcatggcag atgtcataaa acaattagat ata 933

<210> 41  
 <211> 933  
 <212> DNA  
 <213> Thermoanaerobacter sp.

<400> 41

atgagtaaag tggccataat aggttcagga tttgtaggtg ctacatctgc atttacattg 60

gctctaagtg ggactgtgac agacattgtt ttagtagatt taaacaagga caaggcgata 120

ggcgatgcac tggatattag ccacgttata ccgcttatac agcctgtaaa tgtgtatgct 180

ggcgactaca aggatatacg gggcgagat gtagtagttg taacagcagg tgcggctcaa 240

aagccaggag agtctaggct ggaccttgta aaaaagaata catctatatt caagtccatg 300

atacctgaac ttttaaaata caatgataaa gctatatacc tgattgtaac aaatcctgtt 360

gatatattaa cgtatgttac atacaaaata gcgaaacttc cgtgggggcg tgtattcggg 420

tcaggctactg tccttgacag ttcccgaattt aggtatcttt taagtaaaca ttgcaatatt 480

gatcctagaa atgtacatgg aaggataatt ggagaacacg gcgatacaga atttgccgcg 540

tggagcataa caaatatttc aggaatatca tttaatgagt actgcaattt gtgcggacga 600

gtttgtaata caaatctcag aaaggaagtg gaagatgaag ttgtcaatgc ggcttacaaa 660

attattgata aaaagggtgc cacgtattac gctgtggctg tagcagtaag aagaatagtt 720

gagtgtatca taagggaaga aaattcaatt ctacagttt catctccatt aaatggtcaa 780

tacgggtgaa gagatgtatc ttttaagcttg ccatcaattg tgggcaaaaa tgggtgttga 840

agggttctgg atttgccttt ggctgatgac gaagttgaga agtttaaaca ttcggcaagc 900

gttatggctg atgttataaa acagttggac ata 933

<210> 42  
 <211> 936  
 <212> DNA  
 <213> Thermoanaerobacter pseudoethanolicus

<400> 42  
 atgaacaaaa tatctataat aggttctgga tttgtcgggtg ctactactgc atacacactg 60

gctttgagtg ggattgccaa aactattgta ttaatagata ttaataaaga caaagcagaa 120

ggcgaatgctc ttgatataag ccacggcgta ccgtttatta gtccagttga attgtacgcg 180

ggagattata gtgatgtttc aggttctgac ataataatca ttacagcggg agcagcacao 240

aaaccgggag aaaccagact tgacttagtg aagagaaata cgatgatttt taaagacata 300

gtggcaaac ttattaaagt aaatgacaca gcaatatacc ttatagttac aaatccagta 360

gatatcttta catacgttac ctataaaata tctggcttgc catacgaag agtattgggg 420

tctggcacag ttctcgacag tgcgagattc agatatcttt taagcaaaca ttgtaacata 480

gatccgagga atatacacgg atataaatt ggggagcatg gcgattctga gcttgcagct 540

tggagcatta cgaacatagc aggcatacca attgataatt actgcaattt atgtggaaaa 600

gcatgtgaaa aagattttag agaggagatt ttttaaatg ttgtaagagc tgcctatacg 660

ataatagaaa aaaagggtgc gacatattat gcggttgctc tcgcagtaag aagaatcgta 720

gaagctatth tcagagatga aaattccatt ttgactgtgt catctccgt aaccggccaa 780

tatggtgtta caaatgtggc tttagacctt cctccgttg ttggacgaaa tggaaatcgta 840

aatatacttg aattaccact ttcacaggaa gaaattgctg cttttagaag atcagccgaa 900

gttatcaaaa gtgtaataca agagcttgat atataa 936

<210> 43  
 <211> 631  
 <212> DNA  
 <213> Thermoanaerobacterium saccharolyticum

<400> 43

aggcgatgca ctggacataa gccatggcat accattaata cagcctgtaa atgtgtatgc 60

aggtgactac aaagatgttg aaggcgcgga cgtaatagtt gtgacagcag gggctgctca 120

aaagccaggt gagacgagge ttgacctgtg gaagaaaaat acagctatat ttaagtcctat 180

gatacttgag cttttaaagt acaatgacaa ggtatataat ttgattgtca caaatcctgt 240

agacatactg acgtacgtta catacaagat atctggactt ccatggggca gagttttcgg 300

ttctggcact gttcttgaca gttcaaggtt taggtacctt ttaagcagc actgcaatat 360  
 agattccgag aaatgtccac ggaaggataa tcggcgagca tggtagacaca gagtttgag 420  
 catggagcat aacaaacata tctggaatat catttaatga gtactgcagc atatgcgggc 480  
 gcatctgcaa cacaaatttc agaaaggaag tagaagaaga agtcgtaaat gctgcttata 540  
 agataataga caaaaaaggt gctacatact atgctgtcgc agttgcagta agaaggattg 600  
 tggagtgcac cttaagagat gaaaattcca t 631

<210> 44  
 <211> 2229  
 <212> DNA  
 <213> Thermoanaerobacterium saccharolyticum

<400> 44  
 atgatcaatg aatggcgagg gtttcaggag ggcaaatggc aaaagactat tgacgttcaa 60  
 gatattatcc agaaaaatta cacattatac gaaggcgatg atagtttttt agaaggcct 120  
 acagaaaaga ctattaagct ttggaacaaa gttcttgagc taatgaagga agaactgaaa 180  
 aaaggtgtgt tagatatgga taaaaaact gtatcgtcta taacatcca tgatgcgggg 240  
 tatatagaca aagatcttga ggaaatagtt ggattgcaga cagacaaacc tcttaaaga 300  
 gctataatgc cttacggtgg cataagaatg gtcaaaaaag cttgcgaagc ttatggatat 360  
 aaagtggacc caaaagtaga agagatatat acgaagtaca gaaagacca caatgatggt 420  
 gtatttgatg catatactcc agaaataaga gcagcaagac atgccggcat aataacaggt 480  
 ctccagatg catatggcag aggaagaatc ataggtgatt acagaagagt tgctctttat 540  
 ggaattgata gactcatcga agaaaaggaa aaagaaaaac ttgagcttga ttacgatgaa 600

tttagatgaag caactattcg cttgagagaa gaattgacag aacagataaa agcattaaac 660  
 gaaatgaaag agatggcttt aaagtacggt tatgacatat caaagcctgc aaaaaatgca 720  
 aaagaagctg tgcagtgac ttactttgcc ttccttgctg ctataaagga acaaaatggt 780  
 gccgctatgt cgctgggcag agtatctact ttttagata tatacattga aagagatctt 840  
 aaagaaggaa cattgacaga gaaacaagca caagagttaa tggatcactt tgtcatgaag 900  
 cttagaatgg tgaggttctt aaggactcct gattacaatg aactatttag tggcgatcct 960  
 gtttgggtga ctgaatcaat tggcgggtga ggcgtagacg gaagacctct tgtcactaaa 1020  
 aattcattca ggatattaaa tactttatat aacttaggtc ctgcacctga gccaaacttg 1080  
 acggttttat ggtccaaaaa ctttcctgaa aactttaaaa gattctgtgc caaggtatca 1140  
 atagatacaa gtctattca atatgaaaat gacgacttaa tgaggccaat atacaatgac 1200  
 gactatagca tcgcctgctg tgtgtcagct atgaagacgg gagaacagat gcaatttttt 1260  
 ggagcaaggg caaatctgc gaaggcgcta ctgtatgcta taaacggcgg tatcgatgaa 1320  
 aggtataaaa cgcaagtggc accaaaattt aatcctataa cgtctgagta tttagactac 1380  
 gatgaggtaa tggcagcata tgacaatatg ttagagtggc ttgcaaaagt gtatgttaaa 1440  
 gctatgaata taatacacta catgcacgat aaatacgctt atgaaagatc ctttatggct 1500  
 ttgcatgata gagacatcgt aaggacgatg gcttttggaa tcgcaggtct ttctgttgcg 1560  
 gcagattcgt taagcgccat aaagtatgct aaagtaaaag ccataagaga tgaaaatggc 1620  
 atagcaatag attatgaagt ggaaggagat ttcctaagt ttggcaatga tgatgacagg 1680  
 gttgactcaa tagcattga cattgtagaa agattcatga ataagcttaa aaagcacaag 1740  
 acttacagaa actctatacc aacactgtct gttttgacaa taacgtcaaa tgtggtgtac 1800

ggcaaaaaga cgggtgctac acctgacgga agaaaagcgg gagaaccttt tgcgccaggc 1860

gcaaattccga tgcacggcag agatacaaaa ggtgccatag catcaatgaa ttcagtatca 1920

aaaatacctt atgacagttc attggatggt atatcataca catttacgat tgtaccaaat 1980

gcgcttggca aggatgacga agataaaatt aataatcttg taggactatt agatggatat 2040

gcatttaatg cggggcacca cataaacatc aatgttttaa acagagatat gttgcttgat 2100

gctatggagc atcctgaaaa atatccgcag cttactataa gggtttcagg gtatgctgtc 2160

aatttcaata aattaacgag agagcaacag ttggaggtta tatccgcac ttttcacgaa 2220

tctatgtag 2229

<210> 45  
 <211> 2229  
 <212> DNA  
 <213> Clostridium thermocellum

<400> 45  
 atggatgcat ggcgcggtt taataaaggc aactggtgcc aggaaattga cgttcgtgat 60

tttataatta gaaattatac tccttatgaa ggcatgaaa gctttcttgt aggacctacg 120

gatagaacgc ggaaactttg ggagaagggt tccgaactgt taaagaaaga acgggagAAC 180

ggcgggggtat tggatgttga taccataca atttcaacga ttacgtctca taaacctgga 240

tatatagata aagaacttga agttattgtc gggcttcaga cggatgagcc tttaaaaaga 300

gccataatgc cgtttggcgg tatacgtatg gtgattaagg gagccgaagc ttatggccac 360

agtgtggacc ctcaggttgt tgaaatattc acaaagtaca gaaagactca taaccaggga 420

gtttatgatg tatatactcc cgaaatgaga aaagccaaaa aagccgggat tattacagga 480



cttcccgacg catacggcag aggaagaata attggcgatt acagaagggt tgcactttat 540

ggcgttgaca ggctgattgc tgaaaaagag aaagaaatgg caagtcctga aagagattac 600

attgactatg agactgttcg agacagagaa gaaataagcg agcagattaa atctttaaaa 660

caacttaag aaatggcttt aagtiacggt ttgacatat cttgtcctgc aaaggatgcc 720

agagaagcct ttcaatggtt gtattttgca tatcttgacg cagtcaagga acagaacggc 780

gcggcaatga gtattggaag aatttcgact ttccttgaca tatacattga aagggatctc 840

aaagaaggaa aatcacgga ggagtggct caggaaactgg ttgaccagct ggttataaag 900

ctgagaattg tgagattttt gagaactcct gagtatgaaa agctcttcag cggagacccc 960

acttgggtaa ccgaaagtat cggaggtatg gcgctggatg gaagaacgct ggttacaaaa 1020

tcttcgttca ggtttttgca cactcttttc aacctgggac atgcaccgga gcccaacctt 1080

acagtacttt ggtccgtcaa tcttcccgaa ggctttaaaa agtactgtgc aaaggatatca 1140

attcattcaa gctccatcca gtatgaaagc gacgacataa tgaggaaaca ctggggagac 1200

gattatggaa tagcatgctg tgtttctgct atgagaattg gaaaacagat gcagtcttc 1260

ggtgcaagat gcaatcttcg aaaagctctt cttacgcta ttaacggcgg aaaggatgaa 1320

atgacgggag aacagattgc tccgatgttt gcaccggtgg aaaccgaata ccttgattac 1380

gaggacgtaa tgaagaggtt tgacatggtg cttgactggg tggcaaggct ttatatgaac 1440

acctcaata taattcacta catgcatgac aaatatgcct atgaggcgct gcagatggca 1500

ttgcatgaca aagacgtgtt caggacgatg gcatgcggaa tagccggttt gtctgtggtg 1560

gcagactccc ttagcgcgat aaaatatgca aaggttaaac cgatacgaac tgaaaacaac 1620

ctcgttgttg actacgaagt tgagggtgat taccctaaat tcggaataa cgacgaacgt 1680

gttgatgaaa ttgcagtga agtagtaaaa atgttcatga acaagcttag aaagcaaagg 1740

gcttacagaa gtgccactcc gaccttttcc atacttacca taacttcaaa cgtgggtatat 1800

ggaaagaaaa ccggaaacac tcctgacggc agaaaagctg gagaaccttt ggcgccggga 1860

gcaaatccga tgcattgaag ggatataaac ggagcattgg ctgtactgaa cagtattgcg 1920

aagcttcctt atgaatatgc ccaggacggc atttcatata ctttctccat aattccaaaa 1980

gctctgggaa gagacgagga aaccagaata aacaatctta aatcaatgct tgacggatat 2040

ttcaagcagg gcggccacca cataaatgta aatgtgtttg aaaaagagac actgttagat 2100

gccatggaac atccgaaaa atatccacaa cttaccataa gagtgtccgg gtatgcagtg 2160

aactttataa agcttacacg ggagcaacag ctggatgtta ttaacagaac gattcacgga 2220

aagatttaa 2229

<210> 46  
 <211> 2061  
 <212> DNA  
 <213> Clostridium phytofermentans

<400> 46  
 atgatgactt cagttaatgaa acaggaatgg gaagggttta aacaaggttag atggatcact 60

tcagtaaatg ttcgagactt catacagaac aattacacaa tgtatgatgg tgatgaatcc 120

tttttagcag gtccaaccga agccaccaat aaactatggg cccaggttat ggagctttca 180

aagcaggaaa gtgagaaagg tggagtcctt gatatggaca ccaagatagt atctactatt 240

gtttctcagc gtccctggta tttagataaa gatattgaaa caattgttgg ttttcagacc 300

gataagccat ttaagagatc actacaggtc tttggtggta ttcgtatggc acagagtgct 360

|                                                                     |      |
|---------------------------------------------------------------------|------|
| tgccatgaat atggatatga ggtagacgaa gaggtagcac gtatttttac agactaccgc   | 420  |
| aagacacata atcaagggtg atttgatgca tacactgacg aaatgaagct cgctagaaaa   | 480  |
| tcagcaatca ttactgggtt gcctgatgct tatggtagag gtagaattat tggcgattac   | 540  |
| cgtcgagtgg cactttacgg tactgattta cttattgaag acaagaaaga acaacttaca   | 600  |
| acttccttaa agagaatgac tagtgataat attcgcttaa gagaagaatt agcagaacaa   | 660  |
| attcgtgcat taaaagaatt agcgaagctt ggtgaaatct atggttacga tattacgaag   | 720  |
| ccagcaataa atgcaaagga agcaattcag tggttttact ttggatatct tgcagcggtg   | 780  |
| aaagagcaaa acggtgctgc aatgagctta ggccgtactt ctacattcct tgatatttat   | 840  |
| atccagagag atttagataa tgggtttatc acagaaaaag aagcacaaga gtatatcgat   | 900  |
| cattttatta tgaacttcg tctagtgaag ttgcaagaa ctccagaata caatgcctta     | 960  |
| ttctccggtg accctacttg ggtaacagaa agtatcgctg ggattggtag agatggacgc   | 1020 |
| catatggtaa caaagacatc ctccggttac cttcatacgt tagacaacct tggaaactgct  | 1080 |
| ccagaaccaa acatgacagt tctatgggtc actagattac caagattatt taaagagtag   | 1140 |
| tgtgctaaga tgicaattaa gtcacacctt attcaatagc aaaatgatga tatcatgcgt   | 1200 |
| ccaactcatg gtgatgatta tgcaattgct tgttgtgtat cctctatgaa aattggtaaa   | 1260 |
| gagatgcagt tctttggagc acgtgcaaat cttgctaagt gtcttcttta cgcaatcaat   | 1320 |
| ggtgggtgtag atgaagttct taaaattcag gttgggtccaa agtaccgtcc agttgagggt | 1380 |
| gaatacctta attatgagga cgtaatgtcg aaatacaaag atatgatgga gtggctagca   | 1440 |
| gaactttatg tgaatacttt aaatgtaatc cactacatgc atgataaata tagctatgaa   | 1500 |
| agaattcaaa tggcacttca tgatcgtgaa gtaaaacgtt actttgcaac tggatttgcg   | 1560 |

ggctctttctg ttgtagcgga ctctttaagt gcaattaagt atgctaaggt aaaagtaatt 1620

cgtgatgaga atggcgttgt aaccgattac gaaattgaag gtgattatcc aaagtacggc 1680

aacaatgatg atcgtgtaga cgatatcgct gtacagttag tgcatgactt tatgaacatg 1740

attcgcaagc atcatactta tcgtgatgga tacccaacga tgtcaatctt aacgataact 1800

tctaattgtag tttatggaaa gaagacaggt aatactccag acggacgtaa gaagggtgaa 1860

ccattagcac caggtgctaa cccaatgcat cgtcgtgata ctcatggtgc agcagcgctc 1920

ctagcatcgg tagcaaagct tccattccgt gatgcgcagg atggtatttc taatacgttc 1980

tctattgtac caggagcatt aggtagaagt gatgtgttat ttgctggaga cttagattta 2040

gacgatatgt ctgagaacta a 2061

<210> 47

<211> 5003

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 47

agagctataa tgccttacgg tggcataaga atggtcaaaa aagcttgca agcttatgga 60

tataaagtgg acccaaaagt agaagagata ttacgaagt acagaaagac ccacaatgat 120

ggtgtatttg atgcatatac tccagaaata agagcagcaa gacatgccg cataataaca 180

ggctctccag atgcatatgg cagaggaaga atcataggtg attacagaag agttgctctt 240

tatggaattg atagactcat cgaagaaaag gaaaaagaaa aacttgagct tgattacgat 300

gaatttgatg aagcaactat tcgcttgaga gaagaattga cagaacagat aaaagcatta 360

aacgaaatga aagagatggc tttaaagtac ggttatgaca tatcaaagcc tgcaaaaaat 420

gcaaaagaag ctgtgcagtg gacttacttt gccttccttg ctgctataaa ggaacaaaat 480

ggtgccgcta tgtcgtggg cagagtatct acttttttag atatatacat tgaaagagat 540

cttaaagaag gaacattgac agagaaacaa gcacaagagt taatggatca ctttgtcatg 600

aagcttagaa tggtagagtt cttaggact cctgattaca atgaactatt tagtggcgat 660

cctgtttggg tgactgaatc aattggcggg gtaggcgtag acggaagacc tcttgtcact 720

aaaaattcat tcaggatatt aaatacttta tataacttag gtcctgcacc tgagccaaac 780

ttgacggttt tatggtccaa aaaccttctt gaaggtcaat ctatgaaatg cgattaagct 840

tggctgcagg tcgataaacc cagcgaacca tttgaggtga taggtaagat tataccgagg 900

tatgaaaacg agaattggac ctttacagaa ttactctatg aagcgccata tttaaaaagc 960

taccaagacg aagaggatga agaggatgag gaggcagatt gccttgaata tattgacaat 1020

actgataaga taatatactt tttatataga agatatcgcc gtatgtaagg atttcagggg 1080

gcaaggcata ggcagcgcgc ttatcaatat atctatagaa tgggcaaagc ataaaaactt 1140

gcatggacta atgcttgaat cccaggacaa taaccttata gcttgtaaat tctatcataa 1200

ttgtggtttc aaaatcggtt ccgtcgatac tatgttatac gccaaacttc aaaacaactt 1260

tgaaaaagct gttttctggt atttaagggt ttagaatgca aggaacagtg aattggagtt 1320

cgtcttgtaa taattagctt cttgggggtat ctttaaatac tgtagaaaag aggaaggaaa 1380

taataaatgg ctaaatgag aatatcaccg gaattgaaaa aactgatcga aaaataccgc 1440

tgcgtaaaag atacggaagg aatgtctcct gctaaggtat ataagctggt gggagaaaat 1500

gaaaacctat atttaaaat gacggacagc cggataaag ggaccaccta tgatgtggaa 1560

cgggaaaagg acatgatgct atggctggaa ggaaagctgc ctgttccaaa ggtcctgcac 1620

tttgaacggc atgatggctg gagcaatctg ctcagtgtg aggccgatgg cgtcctttgc 1680

tcggaagagt atgaagatga acaaagccct gaaaagatta tcgagctgta tgcggagtgc 1740

atcaggctct ttactccat cgacatatcg gattgtccct atacgaatag cttagacagc 1800

cgcttagccg aattggatta cttactgaat aacgatctgg ccgatgtgga ttgcgaaaac 1860

tgggaagaag acactccatt taaagatccg cgcgagctgt atgatttttt aaagacggaa 1920

aagcccgaag aggaacttgt cttttccac ggcgacctgg gagacagcaa catctttgtg 1980

aaagatggca aagtaagtg ctttattgat ctggggagaa gcggcagggc ggacaagtgg 2040

tatgacattg ctttctgctg ccggtcgatc agggaggata tcggggaaga acagtatgtc 2100

gagctatttt ttgacttact ggggatcaag cctgattggg agaaaataaa atattatatt 2160

ttactggatg aattgtttta gtacctagat ttagatgtct aaaaagcttt ttagacatct 2220

aatcttttct gaagtacatc cgcaactgtc catactctga tgttttatat cttttctaaa 2280

agttcgctag ataggggtcc cgagcgccia cgaggaattt gtatcgactc tagaggatcc 2340

ccgggtaccg agtcgaatt cactggccgc aagcttggcg taatcatggt catagctgtt 2400

tcctgtgtga aattgttata cgctcacaat tccacacaac atacgagccg gaagcataaa 2460

gtgtaaagcc tggggtgcct aatgagttag ctaactcaca ttaattgcgt tgcgtcact 2520

gcccgtttc cagtcgggaa acctgtctg ccagctgcat taatgaatcg gccaacgcgc 2580

ggggagaggc ggtttgcgta ttgggcgtc ttccgcttcc tcgtcactg actcgtgcg 2640

ctcggtcgtt cggtgcgcg gagcggatc agtcactca aaggcggtaa tacggttattc 2700

cacagaatca ggggataacg caggaagaa catgtgagca aaaggccagc aaaaggccag 2760

gaaccgtaaa aaggccgcgt tgctggcggtt ttccatagg ctccgcccc ctgacgagca 2820

tcacaaaaat cgacgtcaa gtcagagggtg gcgaaacccg acaggactat aaagatacca 2880

ggcgtttccc cciggaagct ccctcgtgcg ctctcctgtt ccgacctgc cgcttaccgg 2940

atacctgtcc gcctttctcc ctccgggaag cgtggcgctt tctcatagct cagctgtag 3000

gtatctcagt tcggtgtagg tcgttcgtc caagctgggc tgtgtgcacg aacccccgt 3060

tcagcccgac cgtcgcct tatccggtaa ctatcgtctt gattccaacc cgtaagaca 3120

cgacttatcg ccaactggcag cagccactgg taacaggatt agcagagcga ggtatgtagg 3180

cggtgctaca gatttctga agtggtggcc taactacggc taactagaa ggacagtatt 3240

tggtatctgc gctctgctga agccagttac ctccgaaaa agagtggta gctcttgatc 3300

cggcaaaaa accaccgtg gtagcgggtg ttttttgtt tgcaagcagc agattacgcg 3360

cagaaaaaaa ggatctcaag aagatcttt gatctttct acggggtctg acgtcagtg 3420

gaacgaaaac tcacgttaag ggattttgt catgagatta tcaaaaagga tcttcaccta 3480

gatcctttta aattaaaaat gaagtttta atcaatctaa agtatatatg agtaacttg 3540

gtctgacagt taccaatgct taatcagtga ggcacctatc tcagcgatct gtctatttcg 3600

ttcatccata gttgcctgac tccccgtcgt gtagataact acgatacggg agggcttacc 3660

atctggcccc agtctgcaa tgataccgcg agaccacgc tcaccggctc cagatttattc 3720

agcaataaac cagccagccg gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc 3780

ctccatccag tctattaatt gttgccggga agctagagta agtagttcgc cagttaatag 3840

tttgcgcaac gttgttgcca ttgctacagg catcgtgggtg tcacgctcgt cgtttggtat 3900

ggcttcattc agctccggtt cccaacgata aaggcgagtt acatgatccc ccatgttgtg 3960

caaaaaagcg gtiagctcct tcggctctcc gatcgttgctc agaagtaagt tggccgcagt 4020

gttatcactc atggttatgg cagcactgca taattctctt actgtcatgc catccgtaag 4080

atgcttttct gtgactgggtg agtactcaac caagtcattc tgagaatagt gtatgcggcg 4140

accgagttgc tcttgcccgg cgtcaatacg ggataatacc gcgccacata gcagaacttt 4200

aaaagtgtc atcattggaa aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct 4260

gttgagatcc agttcgatgt aaccactcgt tgcaccaaac tgatcttcag catcttttac 4320

tttaccagc gtttctgggt gagcaaaaac aggaaggcaa aatgccgcaa aaaaggaat 4380

aaggcgaca cggaaatgtt gaatactcat actcttctt tttcaatatt attgaagcat 4440

ttatcagggt tattgtctca tgagcggata catatttgaa tgtatttaga aaaataaaca 4500

aataggggtt ccgcgcacat tccccgaaa agtgcacct gacgtctaag aaaccattat 4560

tatcatgaca ttaacctata aaaataggcg taccacgagg ccttttcgtc tcgcgcgttt 4620

cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca cagcttgtct 4680

gtaagcggat gccgggagca gacaagcccg tcaggcgcg tcagcgggtg ttggcgggtg 4740

tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc accatatgcg 4800

gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgcc attcgccatt 4860

caggctgcgc aactgttggg aaggcgatc ggtgcgggcc tcttcgtat tacgccagt 4920

ggcgaaaggg ggatgtgtg caaggcgatt aagtgggta acgccagggt tttccagtc 4980



acgacgttgt aaaacgacgg cca 5003

<210> 48  
 <211> 6033  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 48  
 cccattgtgc aatcgatcaa ttgaaatagc tattcaattg atttaagcaa ttttatttct 60  
 tcttcattta attctcgcca ttccccctct tttaaatctt catctaattt taattggcct 120  
 attgaaagcc tttttaagta gacgactttt gagcctatcg cttcaaacat tctttttatt 180  
 tgatgatatt taccttctct gattgagaca tatactttcg atgtactgcc tgaagatatt 240  
 atttcagtt ttgccggcat agtcttgtaa ccatcgctta atagtatgcc atctgaaaac 300  
 aaagatacgt catcttcacg gataaaaccc aaaacttctg catagtattt ttgaaaacg 360  
 tgtttttttg gcgataagag cttatgagat agttcacctg catttgtaat taaaagtaat 420  
 ccctctgtgt ctttatcaag ccttctgtct gggaaaacct ttctgcctt tatatggtgt 480  
 ggcaaaagat ctacaacggt tttttctgat ggatcatatg ttgcacagat tacacctttc 540  
 ggttttattca tcattatata tatgtattct ttgtacgata ttttttact tctaaacgtg 600  
 attatatctt tatcaggttg tactgcaaaa cgggggtcgt caatcgtcac attatttatt 660  
 gcaacaaggc cttcttttat aaaattttta atttcttttc ttgtgccata acccatattt 720  
 gataaaagct tatctattct catttttgac atcttaaatt cctcctaaac aatatgactg 780  
 tgcttcttag taaattatat cccaaaaata taaaatttgt agcaaaaatg tgatatatat 840

catatttttt gcgttttcct gatgatacaa ttaagatgat gtttcaagat aataaatTTT 900

tctgaagtgt atacagtata ttgactacaa agaacaaaat actgcaggTc gataaaccca 960

gcgaaccatt tgaggTgata ggtaagatta taccgaggta tgaaaacgag aattggacct 1020

ttacagaatt acTctatgaa gcgccatatt taaaaagcta ccaagacgaa gaggatgaag 1080

aggatgagga ggcagattgc cttgaatata ttgacaatac tgataagata atatatcttt 1140

tatatagaag atatcgccgt atgtaaggat ttCagggggc aaggcatagg cagcgcgctt 1200

atcaatatat ctatagaatg ggcaaagcat aaaaactTgc atggactaat gcttgaaacc 1260

caggacaata accttatagc ttgtaaattc tatcataatt gtggtttcaa aatcggtcc 1320

gtcgatacta tgttatacgc caactttcaa aacaactttg aaaaagctgt tttctggtat 1380

ttaaggtttt agaatgcaag gaacagtga tTggagtTcg tcttgttata attagcttct 1440

tggggTatct ttaaatactg tagaaaagag gaaggaaata ataaatggct aaaatgagaa 1500

tatcaccgga attgaaaaaa ctgatcgaaa aataccgctg cgtaaaagat acggaaggaa 1560

tgtctcctgc taaggtatat aagctggTgg gagaaaatga aaacctatat ttaaaaatga 1620

cggacagccg gtataaaggg accacctatg atgtggaacg ggaaaaggac atgatgctat 1680

ggctggaagg aaagctgcct gttccaaagg tcctgcactt tgaacggcat gatggctgga 1740

gcaatctgct catgagtgag gccgatggcg tcctttgctc ggaagagtat gaagatgaac 1800

aaagccctga aaagattatc gagctgtatg cggagtgcac caggctcttt cactccatcg 1860

acatatcgga ttgtccctat acgaatagct tagacagccg cttagccgaa ttggattact 1920

tactgaataa cgatctggcc gatgtggatt gcgaaaactg ggaagaagac actccattta 1980

aagatccgcg cgagctgtat gatTTTTTaa agacggaaaa gcccgaaagag gaacttgtct 2040

ttccccacgg cgacctggga gacagcaaca tctttgtgaa agatggcaaa gtaagtggct 2100  
 ttattgatct tgggagaagc ggcagggcgg acaagtggta tgacattgcc tctgcgtcc 2160  
 ggtcgatcag ggaggatatac ggggaagaac agtatgtcga gctatTTTT gacttactgg 2220  
 ggatcaagcc tgattgggag aaaataaaat attatatTTT actggatgaa ttgttttagt 2280  
 acctagattt agatgtctaa aaagctTTTT agacatctaa tcttttctga agtacatccg 2340  
 caactgtcca tactctgatg ttttatatct tttctaaaag ttcgctagat aggggtcccg 2400  
 aggcctacg aggaatttgt atcgactcta gaggatcccc gggtaccgaa aaggtgattg 2460  
 tcatggttat ggggaagata cattcaatag agacatgtgg tactgtagat gggcctggca 2520  
 taaggtagct agtctttatg caaggttgc ctttaagggt cgcttattgc cataaccctg 2580  
 acacatggaa ttataacggt ggtaaagaag tatcaacaga tgagatatTT aacgatgcaa 2640  
 aaagatatat accgtacatg aaatcatcag gcggcgccgt gacgtgaca ggtggagagc 2700  
 ctacattaca gcctgaattt tgcgaagatc tatttaaaaa gcttaaagcg tctggcatac 2760  
 aactgcatt agacacatcg ggatatgtga atatagataa agtaaaagaa cttgtaaaac 2820  
 aactgatct tttttgtctt gatataaagc acattgatga tgaaagccat aaaaagctta 2880  
 caggagtgtc gaatagaaag actttggagt ttgcaagata cctttccgat gaaggcaaga 2940  
 aaatgtggat aaggcatgtg atagtacctg gaataacgga tgatatggaa gagataagga 3000  
 aattggctga ttttgtctca tcattgaaaa atgtagatag agttgagata cttccgtatc 3060  
 ataaaatggg tgtgtataaa tatgaggcac ttgggatacc atatagattg aagggaataa 3120  
 atcctcctga cacatcaaaa attaaagaga taaaagaaga gtttaggaaa agagatataa 3180

aagtgtcta aaagcctcat gattcgtatc atggggcttt tcctttgaat taatttgata 3240

aagggtgtaa aattatcatg tgatgatgtg attttgagg taatcgcatg aatttaaata 3300

agataaatag aaacacgtac tacatagata atcctacgaa tattggcggtt tatgcctata 3360

aaaataaaaa ttgtctatta gtagatactg gtataaacgc aagcttggcg taatcatggt 3420

catagctgtt tcctgtgtga aattgttata cgctcacaat tccacacaac atacgagccg 3480

gaagcataaa gtgtaaagcc tggggtgcct aatgagttag ctaactcaca ttaattgcgt 3540

tgcgctcact gcccgtttc cagtcgggaa acctgtcgtg ccagctgcat taatgaatcg 3600

gccaacgcgc ggggagaggc ggtttgcgta ttgggcgctc ttccgcttcc tcgctcactg 3660

actcgtcgcg ctcggtcgtt cggctgcggc gagcggatc agctcactca aaggcggtaa 3720

tacggttata cacagaatca ggggataacg caggaaagaa catgtgagca aaaggccagc 3780

aaaaggccag gaaccgtaaa aaggccgcgt tgcctggcgtt tttccatagg ctccgcccc 3840

ctgacgagca tcacaaaaat cgacgtcaa gtcagagggtg gcgaaaccg acaggactat 3900

aaagatacca ggcgtttccc cctggaagct cctcgtcgcg ctctcctgtt ccgacctgc 3960

cgcttaccgg atacctgtcc gcctttctcc ctccgggaag cgtggcgctt tctcatagct 4020

cacgctgtag gtatctcagt tcggtgtagg tcgttcgctc caagctgggc tgtgtgcacg 4080

aacccccgt tcagcccgcg cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc 4140

cggtaagaca cgacttatcg ccaactggcag cagccactgg taacaggatt agcagagcga 4200

ggtatgtagg cgggtctaca gatttcttga agtggtggcc taactacggc tacactagaa 4260

ggacagtatt tggatctgc gctctgctga agccagtta cttcggaata agagtggta 4320

gctcttgatc cgcaaaaca accaccgtg gtacgggtgg ttttttgtt tgcaagcagc 4380

agattacgcg cagaaaaaaa ggatctcaag aagatccttt gatcttttct acggggtctg 4440

acgctcagtg gaacgaaaac tcacgttaag ggatttttgg catgagatta tcaaaaagga 4500

tcttcaccta gatcctttta aattaaaaat gaagtttta atcaatctaa agtatatatg 4560

agtaaacctg gtctgacagt taccaatgct taatcagtga ggcacctatc tcagcgatct 4620

gtctatttcg ttcatccata gttgcctgac tccccgtcgt gtagataact acgatacggg 4680

agggttacc atctggcccc agtgcctgca tgataccgcg agaccacgc tcaccggctc 4740

cagatttacc agcaataaac cagccagccg gaaggccga gcgcagaagt ggtcctgcaa 4800

ctttatccgc ctccatccag tctattaatt gttgccggga agctagagta agtagttcgc 4860

cagttaatag ttgcgcaac gttgttgcca ttgctacagg catcgtggtg tcacgctcgt 4920

cgtttggat ggttcattc agtccgggt cccaacgac aaggcgagtt acatgatccc 4980

ccatgttgtg caaaaaagcg gttagctcct tcggtcctcc gatcgttgc agaagtaagt 5040

tggccgcagt gttatcactc atggttatgg cagcactgca taattctctt actgtcatgc 5100

catccgtaag atgcttttct gtgactggtg agtactcaac caagtcattc tgagaatagt 5160

glatgcggcg accgagttgc tcttgcccgg cgtcaatag ggataatacc gcgccacata 5220

gcagaacttt aaaagtgtc atcattggaa aacgttcttc ggggcgaaaa ctctcaagga 5280

tcttaccgct gttgagatcc agttcgatgt aaccactcg tgcaccaac tgatcttcag 5340

catcttttac tttcaccagc gtttctgggt gagcaaaaac aggaaggcaa aatgccgcaa 5400

aaaagggaat aaggcgaca cggaaatgt gaatactcat actcttctt tttcaatatt 5460

attgaagcat ttatcagggt tattgtctca tgagcgata catattgaa tgtatttaga 5520

aaaataaaca aataggggtt ccgcgcacat ttccccgaaa agtgccacct gacgtctaag 5580  
 aaaccattat tatcatgaca ttaacctata aaaataggcg tatcacgagg ccctttcgtc 5640  
 tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca 5700  
 cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 5760  
 ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 5820  
 accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgcc 5880  
 attcgccatt caggctgcgc aactgttggg aaggcgatc ggtgcgggcc tcttcgtat 5940  
 tacgccagct ggcgaaaggg ggatgtgctg caaggcgatt aagtgggta acgccagggt 6000  
 ttcccgatc acgacgttgt aaaacgacgg cca 6033

<210> 49

<211> 4542

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 49

gtttatgatg tatatactcc cgaaatgaga aaagccaaaa aagccgggat tattacagga 60  
 ctccccgacg catacggcag aggaagaata attggcgatt acagaagggt tgcactttat 120  
 ggcgttgaca ggctgattgc tgaaaaagag aaagaaatgg caagtcttga aagagattac 180  
 attgactatg agactgttcg agacagagaa gaaataagcg agcagattaa atctttaaaa 240  
 caacttaaag aaatggcttt aagttacggt ttgacatat cttgtcctgc aaaggatgcc 300  
 agagaagcct ttcaatggtt gtattttgca tatcttgacg cagtcaagga acagaacggc 360

gcggcaatga gtattggaag aatttcgact ttccttgaca tatacattga aagggatctc 420

aaagaaggaa aactcacgga ggagttggct caggaactgg ttgaccagct ggttataaag 480

ctgagaattg tgagattttt gagaactcct gagtatgaaa agctcttcag cggagacccc 540

acttgggtaa ccgaaaglat cggaggtatg gcgctggatg gaagaacgct ggttacaaaa 600

tcttcgttca ggtttttgca cactcttttc aacctgggac atgcaccgga gcccaacctt 660

acagtacttt ggtccgtcaa tcttcccga ggctttaaaa agtactgtgc aaaggtatca 720

attcattcaa gctccatcca gtatgaaagc gacgacataa tgaggaaaca ctggggagac 780

gattatggaa tagcagatgg attttctatt attgcaatgt ggaattggga acggaaaaat 840

tattttatta aagagtagtt caacaaacgg gatlgacttt taaaaaagga ttgattctaa 900

tgaagaaagc agacaagtaa gcctcctaaa ttcacttttag ataaaaattt aggaggcata 960

tcaaatgaac ttaataaaaa ttgatttaga caattggaag agaaaagaga tatttaataca 1020

ttatttgaac caacaaacga cttttagtagt aaccacagaa attgatatta gtgttttata 1080

ccgaaacata aaacaagaag gatataaatt ttaccctgca tttatcttct tagtgacaag 1140

ggtgataaac tcaaatagc cttttagaac tggttacaat agcgacggag agttaggtta 1200

ttgggataag ttagagccac tttatacaat ttttgatggt gtatctaaaa cattctctgg 1260

tatttggact cctgtaaaga atgacttcaa agagttttat gatttatacc tttctgatgt 1320

agagaaatat aatgggttcgg ggaaattgtt tcccaaaaca cctatactg aaaatgcttt 1380

ttctctttct attattccat ggacttcatt tactgggttt aacttaaata tcaataataa 1440

tagtaattac ctctaccca ttattacagc aggaaaattc attaataaag gtaattcaat 1500

atatttaccg ctatctttac aggtacatca ttctgtttgt gatggttata atgcaggatt 1560

gtttatgaac tctattcagg aattgtcaga taggcctaata gactggcttt tataatatga 1620

gataatgccg actgtacttt ttacagtcgg ttttctaatag tctactagggc tcgcctttgg 1680

gaagtttgaa gggctggcac gacaggtttc ccgactggaa agcgggcagt gagcgcaacg 1740

caattaatgt gagttagctc actcattagg caccacaggc tttacacttt atgcttccgg 1800

ctcgtatgtt gtgtggaatt gtgagcggat aacaatttca cacaggaaac agctatgacc 1860

atgattacgc caagcttgca tgcctgcagg tcgactctag aggatccgca agcttggcgt 1920

aatcatggtc atagctgttt cctgtgtgaa attgttatcc gtcacaatt ccacacaaca 1980

tacgagccgg aagcataaag tgtaaagcct ggggtgccta atgagtgagc taactcacat 2040

taattgcgtt gcgtcactg cccgctttcc agtcgggaaa cctgtcgtgc cagctgcatt 2100

aatgaatcgg ccaacgcgcg gggagaggcg gtttgcgtat tgggcgtct tccgcttcc 2160

cgctcactga ctgctgcgc tcggtcgttc ggctgcggcg agcggtatca gctcactcaa 2220

aggcggtaat acggttatcc acagaatcag gggataacgc aggaaagaac atgtgagcaa 2280

aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt gctggcgctt ttccataggc 2340

tccgcccccc tgacgagcat cacaaaaac gacgtcaag tcagaggtgg cgaaaccga 2400

caggactata aagataccag gcgtttcccc ctggaagctc cctcgtgcgc tctcctgttc 2460

cgacctgcc gcttaccgga tacctgtccg cttttctccc ttcgggaagc gtggcgcttt 2520

ctcatagctc acgtgtagg tatctcagtt cgggttaggt cgttcgtcc aagctgggct 2580

gtgtgcacga accccccgtt cagcccgacc gctgcgcctt atccggtaac tategtcttg 2640

agtccaaccc ggtaagacac gacttatcgc cactggcagc agccactggt aacaggatta 2700



gcagagcgag gtaigttagc ggtgctacag agttcttgaa gtggtggcct aactacggct 2760

acactagaag gacagtattt ggtatctgcg ctctgctgaa gccagttacc ttcggaaaaa 2820

gagttggtag ctcttgatcc ggcaaaaaa ccaccgtgg tagcgggtgt tttttgttt 2880

gcaagcagca gattacgcgc agaaaaaag gatctcaaga agatccttg atcttttcta 2940

cggggtctga cgctcagttg aacgaaaact cacgttaagg gattttgttc atgagattat 3000

caaaaaggat cttcacctag atccttttaa attaaaaatg aagttttaa tcaatctaaa 3060

gtatatatga gtaaacttgg tctgacagtt accaatgctt aatcagtgag gcacctatct 3120

cagcgatctg tctatttcgt tcatccatag ttgctgact ccccgctgtg tagataacta 3180

cgatacggga gggtttacca tctggcccca gtgctgcaat gataccgcga gaccacgct 3240

caccgctcc agatttatca gcaataaacc agccagccgg aagggccgag cgcagaagtg 3300

gtcttgaac ttatccgcc tccatccagt ctattaattg ttgccgggaa gctagagtaa 3360

gtagttgcc agttaatagt ttgcgcaacg ttgttgccat tgctacaggc atcgtggtgt 3420

cacgctctgc gtttggatg gcttcattca gctccgttc ccaacgatca aggcgagtta 3480

catgatcccc catgttgtgc aaaaaagcgg ttagctcctt cggtcctccg atcgttgtca 3540

gaagtaagtt ggccgcagtg ttatcactca tggttatggc agcactgcat aattctctta 3600

ctgtcatgcc atccgtaaga tgcttttctg tgactgggtga gtactcaacc aagtcattct 3660

gagaatagtg tatgcggcga ccgagttgct cttgccgggc gtcaatacgg gataataccg 3720

cgccacatag cagaacttta aaagtgtca tcattggaaa acgttcttcg gggcgaaaac 3780

tctcaaggat cttaccgtg ttgagatcca gttegatga acccactcg gcaccaact 3840

gatcttcagc atcttttact ttaccagcg tttctgggtg agcaaaaaca ggaaggcaaa 3900

atgccgcaaa aaaggaata agggcgacac ggaaatgttg aatactcata ctcttccttt 3960

ttcaatatta ttgaagcatt tatcagggtt attgtctcat gagcggatac atatttgaat 4020

gtatttagaa aaataaaca ataggggttc cgcgcacatt tccccgaaaa gtgccacctg 4080

acgtctaaga aaccattatt atcatgacat taacctataa aaataggcgt atcacgaggc 4140

cctttcgtct cgcgcgtttc ggtgatgacg gtgaaaacct ctgacacatg cagctcccgg 4200

agacggtcac agcttgtctg taagcggatg ccgggagcag acaagcccgt cagggcgcggt 4260

cagcgggtgt tggcgggtgt cgggctggc ttaactatgc ggcatcagag cagattgtac 4320

tgagagtgca ccatatgcgg tgtgaaatac cgcacagatg cgtaaggaga aaataccgca 4380

tcaggcgcca ttgccattc aggtcgcga actgttggga agggcgatcg gtgcgggcct 4440

cttcctatt acgccagctg gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa 4500

cgccagggtt ttccagtcg cgacgttgta aaacgacggc ca 4542

<210> 50

<211> 5648

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 50

cccgcataa tggaagtaaa gagcatgaaa gtgggaatgc tggcttacac cgataatggcg 60

gaaattgtgt acaagggcaa tccgaactac aagtttgcgg ccggagagga caagccgggg 120

gttgcaccaa gacctttgaa atttgacgat tccataaaaa aagacataga agagttacgg 180

agcaaggtgg atatitaaat tgtttcactt cactggggag tggaggaaag ctttgaagtt 240

ctgcctgaac agagggaatt tgcccacagt cttatagata acggagtgga tgtaaatattg 300

ggacaccatc cccaccagtt ccaaggtata gaaatctaca agggcaaacc tgttttctac 360

agtctgggta attttatttt tgatcagaac gatcccgaag accaggagtc ctttattgtg 420

acacttgatt acaaaggcag cagactgaca ggaatagagg ctgtaccgtg gagaacaatc 480

ggaaaaatac aggtagtctc tcaaaaagga gatgaagcaa aacctatttt ggaaagagag 540

aaaaatttat gtaataggct tgatacaaac tgcattataa aagatgaca attatatttt 600

gaaattggaa aataatgata atataattaa gtggacgta ttttgacaaa ataaatcat 660

aaagtgttg cattgtcga gatttgtgat atcattggat agtaaattat attttaggtt 720

aaaaatggaa aaatagtttt ttatttaaac tttattttta aactttattt aaaatatcaa 780

aataattgcc ttgtattttt acttattgta caatatattt gtacaatata ttaaggaaaa 840

aaatactttt gtagcgactt aaaagtcaat tgaatggacc aataaaggac cttttcaaat 900

ttgtcaaggt attttaggac aatttttttt attttggata ttgttcttgt ttattgggta 960

aataagatgg attttctatt attgcaatgt ggaattggga acggaaaaat tattttatta 1020

aagagtagtt caacaaacgg gattgacttt taaaaaagga ttgattctaa tgaagaaagc 1080

agacaagtaa gcctcctaaa ttcacttttag ataaaaattt aggaggcata tcaaatgaac 1140

tttaataaaa ttgattttag caattggaag agaaaagaga tatttaataca ttatttgaac 1200

caacaaacga cttttagtat aaccacagaa attgatatta gtgttttata ccgaaacata 1260

aaacaagaag gatataaatt ttacctgca tttattttct tagtgacaag ggtgataaac 1320

tcaaatacag cttttagaac tggttacaat agcgacggag agttaggtta ttgggataag 1380

ttagagccac tttatacaat ttttgatggt gtatctaaaa cattctctgg tatitggact 1440  
 cctgtaaaga atgacttcaa agagttttat gatttatacc tttctgatgt agagaaatat 1500  
 aatggttcgg ggaaattgtt tcccaaaaca cctatacctg aaaatgcttt ttctctttct 1560  
 attattccat ggacttcatt tactgggttt aacttaaata tcaataataa tagtaattac 1620  
 ctctaccca ttattacagc aggaaaattc attaataaag gtaattcaat atatttaccg 1680  
 ctatctttac aggtacatca ttctgtttgt gatggttata atgcaggatt gtttatgaac 1740  
 tctattcagg aattgtcaga taggcctaata gactggcttt tataatatga gataatgccg 1800  
 actgtacttt ttacagtcgg ttttctaatag tcactagggc tcgccttttg gaagtttgaa 1860  
 gggtcggcac gacaggtttc ccgactggaa agcgggcagt gagcgcaacg caattaatgt 1920  
 gagttagctc actcattagg caccacaggc ttacacttt atgcttccgg ctctgatgtt 1980  
 gtgtggaatt gtgagcggat aacaatttca cacaggaaac agctatgacc atgattacgc 2040  
 caagcttgca tgcctgcagg tcgactctag aggatcccat taaaggcag gataactca 2100  
 ttgaaatctt ttgggacact ggacggaccg ggtataagat ttgtggtttt catgcagggc 2160  
 tgtcccttgc gtgtatata ttgccacaac agggatacct gggatgttaa tgcggggagt 2220  
 gagtacactc cccggcaagt aattgatgaa atgatgaaat acatagacta tataaaggtc 2280  
 tccggaggcg gaataactgt taccggcggg gagcctgttc tccaggccga ttttgtggcc 2340  
 gaggtgttca gacttgcaaa agagcaggga gtgcatacgg cgctggatac caatggattt 2400  
 gctgacatag agaaggttga aaggcttata aaatacacgc atcttgtatt gctggatata 2460  
 aagcatgccg gggaggataa acataagata attaccggtg tgtccaacga aaaaatcaag 2520

cgttttgcgc tgatatcttc ggaccagga gtgcctatct ggataagata tgtccttgtc 2580

cccggatata ccgacgatga agatgacctt aaaatggcgg ctgatttcat aaaaaagctt 2640

aaaacggtgg aaaaaatcga agttcttcct taccacaaca tgggagcata caaatgggaa 2700

aaacttggtc agaaatacat gcttgaagga gtaaaggggc cgagtgcgca agaggtggaa 2760

aaagcaaaga ggattctgtc aggcaaataa taaaagcttt tttcttttat tatttgcctt 2820

ttctattac caatttgcct tgccttaagt taggtttggt ttgatgagt ttttaaatgt 2880

ttcttttata ttatctttt atatgaacag tgttgtaaac ttccaatcc agtttgcga 2940

atattgattt aaaaatcttt gccgtatact gggcgtcagt taatgcccg tgaagattt 3000

cgctctattc aacgcaagct tggcgtatc atggtcatag ctgtttcctg tgtgaaattg 3060

ttatccgctc acaattccac acaacatacg agccggaagc ataaagtga aagcctgggg 3120

tgcctaata gtagctaac tcacattaat tgcgttgcgc tcaactgccg ctttccagtc 3180

gggaaacctg tcgtgccagc tgcattaatg aatcgccaa cgcgcgggga gaggcggtt 3240

gcgtattggg cgctcttccg ctctctcgct cactgactcg ctgcgctcgg tcgttcggct 3300

gcggcgagcg gtatcagctc actcaaaggc ggtaatcgg ttatccacag aatcagggga 3360

taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc 3420

cgctttgctg gcgtttttcc ataggtccg cccccctgac gagcatcaca aaaatcgacg 3480

ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt tccccctgg 3540

aagtccttc gtgcgtctc ctgttccgac cctgccgctt accggatacc tgtccgctt 3600

tctcccttcg ggaagcgtgg cgctttctca tagctcacgc ttaggtatc tcagttcgg 3660

gtaggtcgtt cgtccaagc tgggtgtgt gcacgaacc cccgttcagc ccgaccgtg 3720

cgcttatcc ggtaactatc gtcttgagtc caacccggta agacacgact tatcgccact 3780

ggcagcagcc actggtaaca ggattagcag agcgaggtat gtaggcggtg ctacagagtt 3840

cttgaagtgg tggcctaact acggctacac tagaaggaca gtatttggtat tctgcgctct 3900

gctgaagcca gtiaccttcg gaaaaagagt tggtagctct tgatccggca aacaaaccac 3960

cgctggtagc ggigggtttt ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc 4020

tcaagaagat cctttgatct tttctacggg gtctgacgct cagtggaaacg aaaactcacg 4080

ttaagggatt ttggtcatga gattatcaaa aaggatcttc acctagatcc ttttaaatta 4140

aaaatgaagt tttaaatcaa tctaaagtat atatgagtaa acttggctctg acagttacca 4200

atgcttaatc agtgaggcac ctatctcagc gatctgtcta tttcgttcat ccatagttagc 4260

ctgactcccc gtcgtgtaga taactacgat acgggagggc ttaccatctg gccccagtgc 4320

tgcaatgata ccgcgagacc cacgtcacc ggctccagat ttatcagcaa taaaccagcc 4380

agccggaagg gccgagcgca gaatgggtcc tgcaacttta tccgcctcca tccagictat 4440

taattgttgc cggaagcta gagtaagtag ttcgccagtt aatagtttgc gcaacgttgt 4500

tgccattgct acaggcatcg tgggtgcacg ctgctgcttt ggtatggctt cattcagctc 4560

cggttcccaa cgatcaaggc gagttacatg atccccatg ttgtgcaaaa aagcggttag 4620

ctccttcggt cctccgatcg ttgtcagaag taagttggcc gcagtgttat cactcatggt 4680

tatggcagca ctgcataatt ctcttactgt catgccatcc gtaagatgct tttctgtgac 4740

tggtagtac tcaaccaagt cattctgaga atagtgtatg cggcgaccga gttgctcttg 4800

cccggcgta atacgggata ataccgccc acatagcaga actttaaaag tgctcatcat 4860

tggaaaacgt tcttcggggc gaaaactctc aaggatctta ccgctgttga gatccagttc 4920  
 gatgtaaccc actcgtgcac ccaactgata ttcagcatct tttactttca ccagcgtttc 4980  
 tgggtgagca aaaacaggaa ggcaaaatgc cgcaaaaaag ggaataaggg cgacacggaa 5040  
 atgttgaata ctcatactct tcctttttca atattattga agcattttatc agggttattg 5100  
 tctcatgagc ggatacatat ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg 5160  
 cacatttccc cgaaaagtc cactgacgt ctaagaaacc attattatca tgacattaac 5220  
 ctataaaaat aggcgtatca cgaggccctt tcgtctcgcg cgtttcggtg atgacggtga 5280  
 aaacctctga cacatgcagc tcccggagac ggtcacagct tgtctgtaag cggatgccgg 5340  
 gagcagacaa gcccgtcagg gcgcgtcagc ggggtgttggc ggggtgctggg gctggcttaa 5400  
 ctatgcggca tcagagcaga ttgtactgag agtgcacat atgcggtgtg aaataccgca 5460  
 cagatgcgta aggagaaaaat accgcatcag gcgccattcg ccattcaggc tgcgcaactg 5520  
 ttgggaaggg cgatcgggtgc gggcctcttc gctattacgc cagctggcga aagggggatg 5580  
 tgctgcaagg cgattaagtt gggtaacgcc aggggtttcc cagtcacgac gttgtaaaac 5640  
 gacggcca 5648

<210> 51

<211> 4648

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 51

tttgggtgga ttcgtatggc acagagtgc tgcctgaat atggatatga ggtagacgaa 60

gaggtagcac gtatittttac agactaccgc aagacacata atcaaggtgt atttgatgca 120

tacactgacg aaatgaagct cgctagaaaa tcagcaatca ttactgggtt gcctgatgct 180

tatggtagag gtagaattat tggcgattac cgtcgagtgg cacittacgg tactgattta 240

cttattgaag acaagaaaga acaacttaca acttccttaa agagaatgac tagtgataat 300

attcgcttaa gagaagaatt agcagaacaa attcgtgcat taaaagaatt agcgaagctt 360

ggtgaaatct atggttacga tattacgaag ccagcaataa atgcaaagga agcaattcag 420

tggctttact ttggatatct tgcagcggta aaagagcaaa acggtgctgc aatgagctta 480

ggccgtactt ctacattcct tgatatttat atccagagag atttagataa tgggtgttacc 540

acagaaaaag aagcacaaga gtatatcgat catitttatta tgaaacttcg tctagtgaag 600

tttgcaagaa ctccagaata caatgcctta ttctccgggtg accctacttg ggtaacagaa 660

agtatcgctg ggattggtag agatggacgc catatggtaa caaagacatc ctccggttac 720

cttcatacgt tagacaacct tggaaactgct ccagaaccaa acatgacagt tctatggtea 780

actagattac caagattatt taaagagtac tgtgctaaga tgtcaattaa gtcacacctt 840

attcaatacg aaaatgatga tatcatgcgt ccaactcatg gtgatgatta tgcaattgct 900

agatggattt tctattattg caatgtggaa ttgggaacgg aaaaattatt ttattaaaga 960

glagttcaac aaacgggatt gactttttaa aaaggattga ttctaataa gaaagcagac 1020

aagtaagcct cctaaattca ctttagataa aaatttagga ggcatatcaa atgaacttta 1080

ataaaattga tttagacaat tggaagagaa aagagatatt taatcattat ttgaaccaac 1140

aaacgacttt tagtataacc acagaaattg atattagtgt tttataccga aacataaaac 1200

aagaaggata taaattttac cctgcattta ttttcttagt gacaagggtg ataaactcaa 1260



atacagcttt tagaactggg tacaatagcg acggagagtt aggttattgg gataagttag 1320  
 agccacttta tacaattttt gatggtgtat ctaaaacatt ctctgggtatt tggactcctg 1380  
 taaagaatga cticaaagag ttttatgatt tatacctttc tgatgtagag aaatataatg 1440  
 gttcggggaa attgtttccc aaaacaccia tacctgaaaa tgctttttct ctttctatta 1500  
 ttccatggac ttcatctact gggtttaact taaatatcaa taataatagt aattaccttc 1560  
 taccattat tacagcagga aaattcatta ataaaggtaa ttcaatatat ttaccgtat 1620  
 ctttacaggt acatcattct gtttgtgatg gttatcatgc aggattgttt atgaactcta 1680  
 ttcaggaatt gtcagatagg cctaatact ggcctttata atatgagata atgccgactg 1740  
 tactttttac agtcggtttt ctaatgtcac tagggctcgc ctttggaag ttgaagggc 1800  
 tggcacgaca ggtttccga ctggaagcg ggcagtgagc gcaacgcaat taatgtgagt 1860  
 tagctcactc attaggcacc ccaggcttta cactttatgc ttccggctcg tatgttgtgt 1920  
 ggaattgtga gcgataaca atttcacaca ggaacagct atgacatga ttacccaag 1980  
 ctgcatgcc tgcaggtcga ctctagagga tccgcaagct tggcgtaatc atggcatag 2040  
 ctgtttcctg tggaaattg ttatccgctc acaattccac acaacatacg agccggaagc 2100  
 ataaagtga aagcctgggg tgcctaatga gtgagctaac tcacattaat tgcgttgccg 2160  
 tcatgcccc cttccagtc gggaaacctg tcgtgccagc tgcattaatg aatcgccaa 2220  
 cgcgcgggga gaggcggttt gcgtattggg cgctcttcg cttcctcgt cactgactcg 2280  
 ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc actcaaagc ggtaatcgg 2340  
 ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag 2400

gccaggaacc gtaaaaaggc cgcgttctg gcgtttttcc ataggctccg cccccctgac 2460

gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga 2520

taccaggcgt tttcccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt 2580

accggatacc tgtccgcctt tctcccttcg ggaagcgtgg cgctttctca tagctcacgc 2640

tgtaggtatc tcagttcggg gtaggtcgtt cgtccaagc tgggctgtgt gcacgaaccc 2700

cccgttcagc ccgaccgtg cgccttatcc ggtaactatc gtcttgagtc caaccggta 2760

agacacgact tatcgccact ggcagcagcc actggtaaca ggattagcag agcgaggat 2820

gtaggcgggtg ctacagagtt cttgaagtgg tggcctaact acggctacac tagaaggaca 2880

gtatttggtg tctgcgctct gctgaagcca gttaccttcg gaaaaagagt tggtagctct 2940

tgatccggca aacaaaccac cgctggtagc ggtgggtttt ttgtttgcaa gcagcagatt 3000

acgcgcagaa aaaaaggatc tcaagaagat cctttgatct tttctacggg gtctgacgt 3060

cagtggaacg aaaactcacg ttaagggatt ttggtcatga gattatcaa aaggatcttc 3120

acctagatcc ttttaaatta aaaatgaagt tttaaatcaa tctaaagtat atatgagtaa 3180

acttggctctg acagttacca atgcttaatc agtgaggcac ctatctcagc gatctgtcta 3240

tttcgttcat ccatagtgc ctgactcccc gtcgtgtaga taactacgat acgggagggc 3300

ttaccatctg gccccagtgc tgcaatgata ccgcgagacc cacgtcacc ggctccagat 3360

ttatcagcaa taaaccagcc agccggaagg gccgagcgca gaagtgggtcc tgcaacttta 3420

tccgcctcca tccagtctat taattgttgc cgggaagcta gagtaagtag ttcgccagtt 3480

aatagtttgc gcaacgttgt tgccattgct acaggcatcg tgggtgcacg ctctctgttt 3540

ggtatggctt cattcagctc cggttcccaa cgatcaaggc gagttacatg atcccccatg 3600

ttgtgcaaaa aagcggttag ctcttcggt cctccgatcg ttgtcagaag taagttggcc 3660  
 gcagtgttat cactcatggt tatggcagca ctgcataatt ctcttactgt catgccatcc 3720  
 gtaagatgct ttictgtgac tggtagtac tcaaccaagt cattctgaga atagtgtatg 3780  
 cggcgaccga gtigtcttgg cccggcgtca atacgggata ataccgcgcc acatagcaga 3840  
 actttaaaag tgctcatcat tggaaaacgt tcttcggggc gaaaactctc aaggatctta 3900  
 ccgctgttga gatccagttc gatgtaacc actcgtgcac ccaactgac ttcagcatct 3960  
 ttactttca ccagcgtttc tgggtgagca aaaacaggaa ggcaaaatgc cgcaaaaaag 4020  
 ggaataaggg cgacacggaa atgttgaata ctcatactct tcctttttca atattattga 4080  
 agcatttattc agggttattg tctcatgagc ggatacatat ttgaatgtat ttagaaaaat 4140  
 aaacaaatag gggttcgcg cacatttccc cgaagagtc cacctgacgt ctaagaaacc 4200  
 attattatca tgacattaac ctataaaaat aggcgtatca cgaggccctt tcgtctcgcg 4260  
 cgtttcgggtg atgacgggtga aaacctctga cacatgcagc tcccggagac ggtcacagct 4320  
 tgtctgtaag cggatgccgg gagcagacaa gcccgtcagg gcgcgtcagc ggggtgttggc 4380  
 ggggtgtcggg gctggcttaa ctatgcggca tcagagcaga ttgtactgag agtgcacat 4440  
 atgcgggtgtg aaataccgca cagatgcgta aggagaaaat accgcatcag gcgccattcg 4500  
 ccattcaggc tgcgcaactg ttgggaaggg cgatcgggtgc gggcctcttc gctattacgc 4560  
 cagctggcga aagggggatg tgctgcaagg cgattaagtt gggtaacgcc agggttttcc 4620  
 cagtcacgac gttgtaaaac gacggcca 4648

<210> 52

<211> 5706

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 52

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| tcaccgacga gcagccatga ttataaacac cggagtccg gagcttagta gcgaagcatc   | 60  |
| taaaagcggg aagccttata ttacggcgg aacgggaaat ggaccagtct ttattgaacg   | 120 |
| taccgtgat gtaagaaaag cggtagagga tatcattgca agccgcacct ttgattacgg   | 180 |
| aatcgtgtct gcggcagaac aatatatggt agtagacagt cttattgcag ctgaagtaaa  | 240 |
| agctgagatg ttaagaaacg gtgcctactt catgaacgag gaagaggaga aaaagctaata | 300 |
| agacctccta aaccttacga gtggaaagc agatacagaa attatgggaa gaccagccga   | 360 |
| agaacttgcc aaacgagcag gatttatggt acctaatacc acgactgtgc tggtttccga  | 420 |
| acagaaatat atttccgaca ggaacccatt tgcaaaagag cttctttgtc ctgtattggc  | 480 |
| ttactacatc gaaaatgact ggatgcatgc ttgtgagaag tgcatgagtc ttttagtaaa  | 540 |
| cgaaagccat ggacataccc tgggtattca ttccagggat gaagaagtaa taggccagtt  | 600 |
| cgctttaaag aaaccagtag gcagagtact tgtaaatacc cccgctaccc tgggtagtat  | 660 |
| gggtgcaacc acaaacttgt ttccggctat gaccctagga agcattacag caggcgccgg  | 720 |
| aatcacagcg gacaatgttt ctccatgaa ttcatatac attcgtaaag taggatatgg    | 780 |
| agttcgggga gtacaagaat ttcttggttc ggttgagaaa acctcaagcg gatacgcgaa  | 840 |
| agtcctgaa acaatcagga acaatgcct tgaacaaac aaggtcaatg cctttgaaac     | 900 |
| aagcaaagc atggaagatg ctagagatct ttgaaacag atttacaag ccttgtccaa     | 960 |

agaactagat ggattttcta ttattgcaat gtggaattgg gaacggaaaa attattttat 1020

taaagagtag ttcaacaac gggattgact tttaaaaaag gattgattct aatgaagaaa 1080

gcagacaagt aagcctccta aattcacttt agataaaaaat ttaggaggca tatcaaatga 1140

actttaataa aattgattta gacaattgga agagaaaaga gatatttaaat cattatttga 1200

accaacaaac gacttttagt ataaccacag aaattgatat tagtgtttta taccgaaaca 1260

taaaacaaga aggatataaa ttttacctg catttatfff cttagtgaca agggtgataa 1320

actcaatac agctttttaga actggttaca atagcgacgg agagttaggt tattgggata 1380

agttagagcc actttataca atttttgatg gtgtatctaa aacattctct ggtatttga 1440

ctctgtaaa gaatgacttc aaagagtttt atgatttata cttttctgat gtagagaaat 1500

ataatggttc ggggaaattg tttcccaaaa cacctatacc tgaaaatgct ttttctcttt 1560

ctattattcc atggacttca tttactgggt ttaacttaaa tatcaataat aatagtaatt 1620

accttctacc cattattaca gcaggaaaat tcattaataa aggttaattca atatatattac 1680

cgctatcttt acaggtacat cattctgttt gtgatggtta tcatgcagga ttgtttatga 1740

actctattca ggaattgtca gataggccia atgactggct tttataatat gagataatgc 1800

cgactgtact ttttacagtc ggttttctaa tgcactagg gctcgcttt gggaagtttg 1860

aagggtggc acgacaggtt tcccactgg aaagcgggca gtgagcgcaa cgcaattaat 1920

gtgagttagc tcactcatta ggcaccccag gctttacact ttatgcttcc ggctcgtatg 1980

ttgtgtggaa ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac 2040

gccaagcttg catgcctgca ggtcgactct agaggatcca acggagagta atcaaaatgg 2100

accatggaaa atcaggagag attgaacgta aggcgtttat atttaacgtg cagaagtaca 2160

acatgtatga cgggcggga atcagaacct tggatattctt taaaggctgt cctcttcggt 2220  
 gtaaattggtg ctccaatccg gaaggctctgg aacgaaaatt tcaggtaatg tataagcaaa 2280  
 gttttgtac aaactgcggg gcgtgcgctg atgtgtgccc cgtaggaatc cacgtgatgt 2340  
 cgaacggaac acatgaaatt gttcgggaaa aggaatgcat cggctgcatg aagtgtaaaa 2400  
 acatctgccc aaagtcggcg cttaccattg caggagaggt aaagaccatt tcagaactgc 2460  
 ttaagattgt ggaagaggac gctgcttttt atgatatgtc cggaggtggc gtgacccttg 2520  
 ggggtggtga agtaaccgca caaccagaag cgcccttaaa tcttttgatg gcttgtaaac 2580  
 aggagggaat caacacagca attgaaactt gcggttattc gaatacagag aacattttaa 2640  
 aaattgcgga atatgtggat cttttcctgt ttgatatcaa acatatggat ccagtacgtc 2700  
 acaacgagtt aacagggtg aacaatgaac agattcttac taaccttgag gaactgcttc 2760  
 accgccgcta taacgtaaaa gtccgtatgc caatgttaaa aggaattaat gacagcaggg 2820  
 aagaaattga tgcggttacc aagtttttaa tgccataccg tactgataag aactttaagg 2880  
 gaattgactt acttccatc cataagctcg gagttaataa atacaatcag cttgataagg 2940  
 tatatccgat tgacggcgat cctagcttaa gtgctgagga tttagaccga attgaaggtt 3000  
 ggatgaaaga atacgatattt ccggttaacg tggtaaaaca ctaagaaagg ggaaggacgc 3060  
 catggaagaa ggcaagcttg gcgtaatcat ggcatagct gtttcctgtg tgaaattgtt 3120  
 atccgtcac aattccacac aacatacag cgggaagcat aaagtgtaaa gcctggggtg 3180  
 cctaattgagt gagctaactc acattaattg cgttcgctc actgcccgtt ttccagtcgg 3240  
 gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc 3300

gtattgggcg ctcttcgct tcctcgctca ctgactcgct gcgctcggtc gttcggtgc 3360

ggcgagcggg atcagctcac tcaaaggcgg taatacgggt atccacagaa tcaggggata 3420

acgcaggaag gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg 3480

cgttgctggc gtttttccat aggtccgccc cccctgacga gcatcacaaa aatcgacgt 3540

caagtacagag gtggcgaaac ccgacaggac tataaagata ccaggcggtt cccctggaa 3600

gctccctcgt gcgtctcct gttccgacc tgccgcttac cggatacctg tccgccttc 3660

tcccttcggg aagcgtggcg ctttctcata gctcacgtg taggtatctc agttcgggtg 3720

aggctgttcg ctccaagctg ggctgtgtgc acgaaccccc cggtcagccc gaccgtgcg 3780

ccttatccgg taactatcgt cttgagtcca acccgtaag acacgactta tcgccactgg 3840

cagcagccac tggtaacagg attagcagag cgaggatgt aggcgggtgt acagattct 3900

tgaagtgggt gcctaactac ggctacacta gaaggacagt atttggtatc tgcgctctgc 3960

tgaagccagt taccttcgga aaaagagttg gtagctcttg atccggcaaa caaaccacg 4020

ctggtagcgg tggttttttt gtttcaagc agcagattac gcgcagaaaa aaaggatctc 4080

aagaagatcc ttgatcttt tctacggggt ctgacgtca gtggaacgaa aactcacgtt 4140

aagggtttt ggtcatgaga ttatcaaaaa ggatcttcac ctagatcctt ttaaattaaa 4200

aatgaagttt taaatcaatc taaagtatat atgagtaaac ttggtctgac agttaccaat 4260

gcttaatcag tgaggcacct atctcagcga tctgtctatt tcgttcaccc atagttgcct 4320

gactccccgt cgtgtagata actacgatac gggagggtt accatctggc cccagtgtg 4380

caatgatacc gcgagaccca cgctcaccgg ctccagattt atcagcaata aaccagccag 4440

ccggaagggc cgagcgaga agtggtcctg caactttatc cgctccatc cagtctatta 4500

atgttgccg ggaagctaga gtaagtagtt cgccagttaa tagtttgcgc aacgttgttg 4560

ccattgctac aggcatcgtg gtgtcacgct cgtcgtttgg tatggcttca ttcagctccg 4620

gttcccaacg atcaaggcga gttacatgat ccccatgtt gtgcaaaaaa gcggttagct 4680

ccttcggtcc tccgatcgtt gtcagaagta agttggccgc agtggtatca ctcattggtta 4740

tggcagcact gcataattct cttactgtca tgccatccgt aagatgcttt tctgtgactg 4800

gtgagtactc aaccaagtca ttctgagaat agtgtatgcg gcgaccgagt tgctcttgcc 4860

cggcgtcaat acgggataat accgcgccac atagcagaac tttaaaagtg ctcattcattg 4920

gaaaacgttc ttccggggcga aaactctcaa ggatcttacc gctgttgaga tccagttcga 4980

tgtaaccac tcgtgcacc aactgatctt cagcatcttt tactttcacc agcgtttctg 5040

ggtgagcaaa aacaggaagg caaaatgccg caaaaaaggg aataagggcg acacggaaat 5100

gttgaatact catactcttc ctttttcaat attattgaag cattatcag ggttattgtc 5160

tcatgagcgg atacatat tt gaatgtattt agaaaaataa acaaataggg gtcccgcgca 5220

catttccccg aaaagtgcc cctgacgtct aagaaacat tattatcatg acattaacct 5280

ataaaaatag gcgtatcacg aggccctttc gtctcgcgcg tttcgggtgat gacggtgaaa 5340

acctctgaca catgcagctc ccggagacgg tcacagcttg tctgtaagcg gatgccggga 5400

gcagacaagc ccgtcagggc gcgtcagcgg gtgttgccgg gtgtcggggc tggcttaact 5460

atgcggcatc agagcagatt gtactgagag tgcaccatat gcggtgtgaa ataccgcaca 5520

gatgcgtaag gagaaaaatc cgcacaggc gccattcgcc attcaggctg cgcaactgtt 5580

gggaagggcg atcgggtcgg gcctcttcgc tattaccca gctggcgaaa gggggatgtg 5640



ctgcaaggcg attaagttgg gtaacgccag ggttttccca gtcacgacgt tgtaaaacga 5700

cggcca 5706

<210> 53  
 <211> 5780  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: Synthetic  
 polynucleotide

<400> 53  
 tctatcagct gtccctcctg ttcagctact gacggggtgg tgcgtaacgg caaaagcacc 60

gccggacatc agcgctagcg gagtgtatac tggcttacta tgttggcact gatgagggtg 120

tcagtgaagt gcttcatgtg gcaggagaaa aaaggctgca ccggtgcgtc agcagaatat 180

gtgatacagg atatattccg ctctctcgtc cactgactcg ctacgctcgg tcgttcgact 240

gcggcgagcg gaaatggctt acgaacgggg cgagatttc ctggaagatg ccaggaagat 300

acttaacagg gaagtgaagc ggccgcggca aagccgtttt tccataggct ccgccccct 360

gacaagcatc acgaaatctg acgctcaaat cagtgggtggc gaaacccgac aggactataa 420

agataccagg cgtttcccc tggcggtctc ctctgtcgtc ctctgttcc tgcctttcgg 480

tttaccggtg tcattccgct gttatggcgg cgtttgtctc attccacgcc tgacactcag 540

ttccgggtag gcagttcgtc ccaagctgga ctgtatgcac gaaccccccg ttcagtcgga 600

ccgtcgcgcc ttatccggtg actatcgtct tgagtccaac ccggaagac atgcaaaagc 660

accactggca gcagccactg gtaattgatt tagaggagtt agtcttgaag tcatgcgccg 720

gttaaggcta aactgaaagg acaagttttg gtgactgcgc tctccaagc cagttacctc 780

ggttcaaaga gttggtagct cagagaacct tcgaaaaacc gccctgcaag gcggtttttt 840

cgttttcaga gcaagagatt acgcgcagac caaacgatac tcaagaagat catcttatta 900

atcagataaa atattttctag atttcagtgc aatttatctc ttcaaagtga gcacctgaag 960

tcagcccat acgataaag ttgtaattct catgtttgac agcttatcat cgataagctt 1020

taatgcggta gtttatcaca gttaaattgc taacgcagtc aggcacctat acatgcattt 1080

acttataata cagtttttta gttttgctgg ccgcactctc tcaaatatgc ttcccagcct 1140

gcttttctgt aacgttcacc ctctacctta gcataccctc cctttgcaa tagtcctctt 1200

ccaacaataa taatgtcaga tctgttagag accacatcat ccacggttct atactgttga 1260

cccaatgcgt ctcccttgtc atctaaacc acaccgggtg tcataatcaa ccaatcgtaa 1320

ccttcactc ttccacccat gtctctttga gcaataaagc cgataacaaa atctttgtcg 1380

ctcttcgcaa tgtaacagt acccttagta tattctccag tagataggga gcccttgcatt 1440

gacaattctg ctaacatcaa aaggcctcta gggtcctttg ttacttcttc tgccgcctgc 1500

ttcaaaccgc taacaatacc tgggcccacc acaccgttg cattcgtaat gctgcccatt 1560

tctgctattc tgiatacacc cgcagagtac tgcaatttga ctgtattacc aatgtcagca 1620

aattttctgt cttcgaagag taaaaaattg tacttggcgg ataatgcctt tagcggctta 1680

actgtgccct ccatggaaaa atcagtcaag atatccacat gtgttttttag taaacaaatt 1740

ttgggacctt atgtttcaac taactccagt aattccttgg tggtagaac atccaatgaa 1800

gcacacaagt ttgtttgctt ttctgtcatg atattaaata gcttggcagc aacaggacta 1860

ggatgagtag cagcacgttc cttatatgta gctttcgaca tgatttatct tcgtttcctg 1920

caggtttttg ttctgtgcag ttgggttaag aatactgggc aatttcattt ttcttcaaca 1980

ctacatatgc gtatatatac caatctaagt ctgtgctcct tccttcgttc ttccttctgt 2040

tccgagatta ccgaatcaaa aaaatttcaa agaaaccgaa atcaaaaaaa agaataaaaa 2100

aaaaatgatg aattgaattg aaaagctagc ttatcgatgg gtccttttca tcacgtgcta 2160

taaaaataat tataatttaa attttttaat ataaatatat aaattaaaaa tagaaagtaa 2220

aaaaagaaat taaagaaaaa atagtttttg ttttccgaag atgtaaaaga ctctaggggg 2280

atcgccaaca aatactacct tttatcttgc tcttctgct ctcaggtatt aatgccgaat 2340

tgtttcatct tgtctgtgta gaagaccaca cagaaaaatc ctgtgatttt acattttact 2400

tatcgtaaata cgaatgtata tctattttaa ctgcttttct tgtctaataa atatatatgt 2460

aaagtacgct ttttgttgaa atttttttaa cttttgttta ttttttttc ttcattccgt 2520

aactcttcta ctttctttat ttactttcta aaatccaaat acaaaacata aaaataaata 2580

aacacagagt aaattcccaa attattccat cattaaaaga tacgaggcgc gtgtaagtta 2640

caggcaagcg atctctaaga aaccattatt atcatgacat taacctataa aaaagcctc 2700

tcgagctaga gtcgatcttc gccagcaggg cgaggatcgt ggcatcacgc aaccgcgccg 2760

tgcgcgggtc gtcggtgagc cagagtttca gcaggccgcc caggcggccc aggtcgccat 2820

tgatgcgggc cagctcgcg acgtgctcat agtccacgac gccgtgatt ttgtagccct 2880

ggccgacggc cagcaggtag gccgacaggc tcatgccggc cgccgccgc ttttctcaa 2940

tcgtctttcg ttcgtctgga aggcagtaca ccttgatagg tgggctgccc ttcctggttg 3000

gcttggtttc atcagccatc cgtttgccct catctgttac gccggcggtg gccggccagc 3060

ctcgcagagc aggattcccg ttgagcaccg ccaggtgcga ataagggaca gtgaagaagg 3120

aacacccgct cgcggttggg cctacttcac ctatcctgcc cggtgacgc cgttggatac 3180  
 accaaggaag gtctacacga accctttggc aaaatcctgt atatcgtgcg aaaaaggatg 3240  
 gatataccga aaaaatcgct ataatgaccc cgaagcaggg ttatgcagcg gaaaagcgct 3300  
 gcttccttgc tgttttggg aatatctacc gactggaaac aggcaaatgc aggaaattac 3360  
 tgaactgagg ggacaggcga gagacgatgc caaagagcta caccgacgag ctggccgagt 3420  
 gggttgaatc ccgcgcggcc aagaagcgcc ggctgatga ggctgcggtt gcgttcctgg 3480  
 cggtaggggc ggatgtcgat atgcgtaagg agaaaatacc gcatcaggcg catatttgaa 3540  
 tgtatttaga aaaataaaca aaaagagttt gtagaaacgc aaaaaggcca tccgtcagga 3600  
 tggccttctg ctttaattga tgcctggcag tttatggcgg gcgtcctgcc cgccaccctc 3660  
 cgggcggtt cttcgcaacg ttcaaatccg ctcccggcgg atttgtccta ctcaggagag 3720  
 cgttcacga caaacaacag ataaaacgaa aggccagtc tttcgactga gcctttcgtt 3780  
 ttatttgatg cctggaaacc cagcgaacca tttgaggtga taggtaagat tataccgagg 3840  
 tatgaaaacg agaattggac ctttacagaa ttactctatg aagcgccata tttaaaaagc 3900  
 taccaagacg aagaggatga agaggatgag gaggcagatt gccttgaata tattgacaat 3960  
 actgataaga taatatatct tttatataga agatatgcc gtatgtaagg atttcagggg 4020  
 gcaaggcata ggcagcgcgc ttatcaatat atctatagaa tgggcaaagc ataaaaactt 4080  
 gcatggacta atgcttgaaa ccagggacaa taaccttata gcttgtaaat tctatcataa 4140  
 ttgtggtttc aaaatcggtc ccgtcgatac tatgttatac gccaaacttc aaaacaactt 4200  
 tgaaaaagct gttttctggt atttaagggt ttagaatgca aggaacagtg aattggagtt 4260  
 cgtcttgta taattagctt cttggggtat ctttaaatac tgtagaaaag aggaaggaaa 4320

taataaatgg ctaaatgag aatatcaccg gaattgaaaa aactgatcga aaaataccgc 4380

tgcgtaaaag atacggaagg aatgtctcct gctaaggat ataagctggt gggagaaaaat 4440

gaaaacctat atttaaaaat gacggacagc cgtataaag ggaccaccta tgatgtggaa 4500

cgggaaaagg acatgatgct atggctggaa ggaaagctgc ctgttccaaa ggtcctgcac 4560

tttgaacggc atgatggctg gagcaatctg ctcagtgtg aggccgatgg cgtcctttgc 4620

tcggaagagt atgaagaatg acaaagccct gaaaagatta tcgagctgta tgcggagtgc 4680

atcaggtctt ttactccat cgacatatcg gattgtccct atacgaatag cttagacagc 4740

cgcttagccg aattggatta cttactgaat aacgatctgg ccgatgtgga ttgcgaaaac 4800

tgggaagaag acactccatt taaagatccg cgcgagctgt atgatTTTT aaagacggaa 4860

aagccgaag aggaacttgt cttttccac ggcgacctgg gagacagcaa catctttgtg 4920

aaagatggca aagtaagtgg ctttattgat cttgggagaa gcggcagggc ggacaagtgg 4980

tatgacattg ccttctgcgt ccggtcgatc agggaggata tcggggaaga acagtatgtc 5040

gagctatTTT ttgacttact ggggatcaag cctgattggg agaaaataaa atattatatt 5100

ttactggatg aattgtttta gtacctagat ttagatgtct aaaaagcttt ttagacatct 5160

aatcttttct gaagtacatc cgcaactgtc catactctga tgttttatat cttttctaaa 5220

agttcgctag ataggggtcc cgagcgcta cgaggaattt gtatcgtaca aggatatcga 5280

gggcgcagat gtagtagttg taacagcagg tgcggctcaa aagccaggag agtctaggct 5340

ggaccttgta aaaaagaata catctatatt caagtccatg atacctgaac ttttaaaata 5400

caatgataaa gctatatacc tgattgtaac aaatcctgtt gatataattaa cgtatgttac 5460

atacaaaata gcgaaacttc cgtgggggcg tgtattcggg tcaggtactg tccttgacag 5520

ttcccgattt aggtatcttt taagtaaaca ttgcaatatt gatcctagaa atgtacatgg 5580

aaggataatt ggagaacacg gcgatacaga atttgcggcg tggagcataa caaatatttc 5640

aggaatatca tttaatgagt actgcaattt gtgcggacga gtttgaata caaatccag 5700

aaaggaagtg gaagatgaag ttgtcaatgc ggcttacaaa attattgata aaaagggtgc 5760

cacgtattac gctgtggctg 5780

<210> 54  
 <211> 6539  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: Synthetic  
 polynucleotide

<220>  
 <221> modified\_base  
 <222> (92)..(92)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (151)..(151)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (153)..(153)  
 <223> a, c, g or t

<400> 54  
 gcacatattg atagagaaca tggctggata attactataa ctccaagaaa acgaatagta 60

aaagaatgga ggcgaaattaa tgagtaatgt cncaatacaa ttaatagaaa tttgtcgga 120

|                                                                    |      |
|--------------------------------------------------------------------|------|
| atatgtaaat aataacttaa acataaatga ntntatcgaa gatttccaag tgctttatga  | 180  |
| acaaaagcaa gatttattaa cagatgaaga aatgagtttg tttgatgata tttatatggc  | 240  |
| ttgtgaatat taigaacagg atgaaaatat aagaaatgaa tatcacttgt atattggaga  | 300  |
| aatgaatta agacaaaaag tgcaaaaact tgtaaaaaag ttagcagcat aataaacgcg   | 360  |
| taaggcatga tagctaaagc ggtattttta tgcaattaaa aggaaaaatg atatctgata  | 420  |
| aaccgcggaa aagtatttta gaaaacaact ataaagataa tatttcaaag caagaaggat  | 480  |
| aaaataagat taaactatta gacactttta ttagaaaatg ttataatatt attaagagaa  | 540  |
| aatttatatt atttaggagg taattttatg agtaaatggg ccataatagg ttcaggattt  | 600  |
| gtaggtgcta catctgcatt tacattggct ctaagtggga ctgtgacaga cattgtttta  | 660  |
| gtagatttaa acaaggacaa ggcgataggc gatgcactgg atattagcca aaaccagcg   | 720  |
| aaccatttga ggtgataggt aagattatac cgaggatatga aaacgagaat tggaccttta | 780  |
| cagaattact ctatgaagcg ccatatttaa aaagctacca agacgaagag gatgaagagg  | 840  |
| atgaggaggc agattgcctt gaatatattg acaatactga taagataata tatctttat   | 900  |
| atagaagata tcgccgtatg taaggatttc agggggcaag gcataggcag cgcgcttadc  | 960  |
| aatatatcta tagaatgggc aaagcataaa aacttgcatg gactaatgct tgaaaccag   | 1020 |
| gacaataacc ttatagcttg taaattctat cataattgtg gtttcaaat cggtccgctc   | 1080 |
| gatactatgt tatacgccaa ctttcaaac aactttgaaa aagctgtttt ctggtattta   | 1140 |
| aggttttaga atgcaaggaa cagtgaattg gagttcgtct tgttataatt agcttcttgg  | 1200 |
| ggtatcttta aatactgtag aaaagaggaa ggaaataata aatggctaaa atgagaatat  | 1260 |
| caccggaatt gaaaaaactg atcgaaaaat accgtgcgtg aaaagatacg gaaggaaatg  | 1320 |

ctcctgctaa ggtatataag ctgggtgggag aaaatgaaaa cctatatatta aaaatgacgg 1380

acagccggta taaagggacc acctatgatg tggaacggga aaaggacatg atgctatggc 1440

tggaaggaaa gctgcctgtt ccaaaggtcc tgcactttga acggcatgat ggctggagca 1500

atctgctcat gagtgaggcc gatggcgtcc ttgctcgga agagtatgaa gatgaacaaa 1560

gccctgaaaa gattatcgag ctgtatgcgg agtgcacag gctctttcac tccatcgaca 1620

tatcgattg tccctatacg aatagcttag acagccgctt agccgaattg gattacttac 1680

tgaataacga tctggccgat gtggattgcg aaaactggga agaagacact ccatttaaag 1740

atccgcgcga gctgtatgat tttttaaga cggaaaagcc cgaagaggaa ctgtctttt 1800

cccacggcga cctgggagac agcaacatct ttgtgaaaga tggcaaagta agtggcttta 1860

ttgatcttgg gagaagcggc agggcggaca agtggatga cattgccttc tgcgtccggt 1920

cgatcaggga ggatatcggg gaagaacagt atgtcgagct attttttgac ttactgggga 1980

tcaagcctga ttgggagaaa ataaaatatt atattttact ggatgaattg ttttagtacc 2040

tagatttaga tgictaaaaa gctttttaga catctaactt tttctgaagt acatccgcaa 2100

ctgtccatac tctgatgttt tatactttt ctaaaagttc gctagatagg ggtcccgagc 2160

gcctacgagg aatttgatc gacgtattac gctgtggctg tagcagtaag aagaatagtt 2220

gagtgtatca taagggatga aaattcaatt cttacagttt catctccatt aaatggtcaa 2280

tacgggtgtaa gagatgtatc ttttaagcttg ccatcaattg tgggcaaaaa tgggtgttga 2340

agggttctgg atttgcttt ggctgatgac gaagttgaga agtttaaca ttcggcaagc 2400

gttatggctg atgttataaa acagttggac atataaaaata aatcattgta taaggtttat 2460



aagacggctt ttatcatgta tggtaaagc cgctttttta tgaatataaa aatacaaagt 2520

ggaaaaatcta aataaagggtg atgcaatatg cagaatatga gtcctcaaga aattatatcg 2580

agtgccttta tgaaggcaaa aaaatctgag aatattatac atgctaaggc tatagattat 2640

gggaaaaata tatcagataa ccagatgcaa gcgatatgta agcaaataga gataacggct 2700

ttaaaccatg tggacaaaat agtgacagct gagaagacga tgcactatc agctgtccct 2760

cctgttcagc tactgacggg gtggcgcgta acggcaaaag caccgccgga catcagcgct 2820

agcggagtgt atactggctt actatgttgg cactgatgag ggtgtcagt aagtgttca 2880

tgtggcagga gaaaaaaggc tgcaccgtg cgtcagcaga atatgtgata caggatatat 2940

tccgttctc cgctcactga ctgcctacgc tcggcgttc gactgcggcg agcggaaatg 3000

gcttacgaac ggggcggaga ttctctggaa gatgccagga agatactta cagggaagtg 3060

agagggccgc ggcaaagccg tttttccata ggctccgcc cctgacaag catcacgaaa 3120

tctgacgctc aaatcagtgg tggcgaaacc cgacaggact ataaagatac caggcgtttc 3180

ccctggcgg ctccctcgtg cgctctcctg ttctgcctt tcggtttacc ggtgtcattc 3240

cgctgttatg gccgcgttg tctcattcca cgcctgacac tcagttccgg gtaggcagtt 3300

cgctccaagc tggactglat gcacgaacc cccgttcagt ccgaccgtg cgccttatcc 3360

ggtaactatc gtcttgagtc caaccggaa agacatgcaa aagcaccact ggcagcagcc 3420

actggtaatt gatttagagg agttagtctt gaagtcatgc gccggttaag gctaaactga 3480

aaggacaagt tttggtgact gcgtcctcc aagccagtta cctcggttca aagagttggt 3540

agctcagaga accttcgaaa aaccgccctg caaggcggtt tttcgtttt cagagcaaga 3600

gattacgcgc agacaaaaac gatctcaaga agatcatctt attaatacaga taaaatattt 3660

ctagatttca gtgaattta tctcttcaaa ttagcacct gaagtcagcc ccatacgata 3720

taagttgtaa ttctcatgtt tgacagctta tcatcgataa gctttaatgc ggtagtttat 3780

cacagttaaa ttgctaacgc agtcaggcac ctatacatgc atttacttat aatacagttt 3840

tttagttttg ctggccgcat cttctcaaat atgcttccca gccctgtttt ctgtaacgtt 3900

cacctctac cttagcatcc cttccctttg caaatagtcc tcttccaaca ataataatgt 3960

cagatcctgt agagaccaca tcatccacgg ttctatactg ttgaccaat gcgtctccct 4020

tgatcatctaa acccacaccg ggtgtcataa tcaaccaatc gtaaccttca tctcttccac 4080

ccatgtctct ttgagcaata aagccgataa caaaatcttt gtcgtcttcc gcaatgtcaa 4140

cagtaccctt agtatattct ccagtagata gggagccctt gcatgacaat tctgctaaca 4200

tcaaaaggcc tctaggttcc ttgttactt cttctgccgc ctgcttcaaa ccgctaacaa 4260

tacctgggcc caccacaccg tgtgcattcg taatgtctgc ccattctgct attctgtata 4320

cacccgcaga gtactgcaat ttgactgtat taccaatgtc agcaaatttt ctgtcttcga 4380

agagtaaaaa attgtacttg gcggataatg cctttagcgg cttactgtg ccctccatgg 4440

aaaaatcagt caagatatcc acatgtgttt ttagtaaca aattttggga cctaagtctt 4500

caactaactc cagtaattcc ttggtggtac gaacatccaa tgaagcacac aagtttgttt 4560

gcttttcgtg catgatatta aatagcttgg cagcaacagg actaggatga gtagcagcac 4620

gttccttata ttagcttttc gacatgattt atcttcgttt cctgcaggtt ttgttctgt 4680

gcagttgggt taagaatact gggcaatttc atgtttcttc aacctacat atgcgtatat 4740

ataccaatct aagtctgtgc tccttccttc gttcttcctt ctgttcggag attaccgaat 4800

caaaaaaatt tcaaagaaac cgaaatcaaa aaaaagaata aaaaaaaat gatgaattga 4860

attgaaaagc tagcttatcg atgggtcctt ttcacacgt gctataaaaa taattataat 4920

ttaaattttt taatataaat atataaatta aaaatagaaa gtaaaaaaag aaattaaaga 4980

aaaaatagtt ttigtitttc gaagatgtaa aagactctag ggggatcgcc aacaaatact 5040

accttttacc ttgctcttcc tgctctcagg tattaatgcc gaattgttcc atcttgtctg 5100

tgtagaagac cacacacgaa aatcctgtga ttttacattt tacttatcgt taatcgaatg 5160

tatatctatt taatctgctt ttcttgtcta ataaatatat atgtaaagta cgctttttgt 5220

tgaaattttt taaacctttg tttatttttt ttcttctatt ccgtaactct tctaccttct 5280

ttatttactt tctaaaatcc aaatacaaaa cataaaaaata aataaacaca gagtaaattc 5340

ccaaattatt ccatcattaa aagatacgag gcgcgtgtaa gttacaggca agcgatctct 5400

aagaaacat tattatcatg acattaacct ataaaaaagg cctctcgagc tagagtcgat 5460

cttcgccagc agggcgagga tcgtggcatc accgaaccgc gccgtgcgcg ggtcgtcggg 5520

gagccagagt ttcagcaggc cgcccaggcg gcccaggctg ccattgatgc gggccagctc 5580

gcggacgtgc tcatagtcca cgacccccgt gatattgtag ccttggccga cggccagcag 5640

gtaggccgac aggtcatgc cggccgccgc cgccttttcc tcaatcgctc ttcgttcgtc 5700

tggaaggcag tacaccttga taggtgggct gcccttctg gttggcttgg tttcatcagc 5760

catccgcttg cctcatctg ttacgccggc gtagccggc cagcctcgca gagcaggatt 5820

cccgttgagc accgccaggt gcgaataagg gacagtgaag aaggaacacc cgctcgcggg 5880

tgggcctact tcacctatcc tgcccggctg acgccgttgg atacaccaag gaaagtctac 5940

acgaaccctt tggcaaaatc ctgtatatcg tgcgaaaaag gatggatata ccgaaaaaat 6000

cgctataatg accccgaagc agggttatgc agcggaaaag cgctgcttcc ctgctgtttt 6060

gtggaatatc taccgactgg aaacaggcaa atgcaggaaa ttactgaact gaggggacag 6120

gcgagagacg atgccaaga gctacaccga cgagctggcc gagtgggttg aatccgcgc 6180

ggccaagaag cgccggcgtg atgaggctgc ggttgcgttc ctggcggatga gggcggatgt 6240

cgatatgcgt aaggagaaaa taccgcatca ggcgcatatt tgaatgtatt tagaaaaata 6300

aacaaaaaga gttttagaa acgcaaaaag gccatccgtc aggatggcct tctgcttaat 6360

ttgatgcctg gcagtttatg gcgggcgtcc tgcccgcac cctccgggcc gttgcttcgc 6420

aacgttcaaa tccgtcccg gcggatttgt cctactcagg agagcgttca ccgacaaaca 6480

acagataaaa cgaaagcccc agtctttcga ctgagccttt cgttttattt gatgcctgg 6539

<210> 55

<211> 6086

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 55

tctatcagct gtccctcctg ttcagctact gacgggtgg tgcgtaacgg caaaagcacc 60

gccggacatc agcgctagcg gagtgtatac tgcttacta tgttggcact gatgagggtg 120

tcagtgaagt gcttcatgtg gcaggagaaa aaaggctgca ccggtgcgtc agcagaatat 180

gtgatacagg atatattccg ctctctcgt cactgactcg ctacgtcgg tcgttcgact 240

gcggcgagcg gaaatggctt acgaacgggg cggagatttc ctggaagatg ccaggaagat 300

acttaacagg gaagtgagag ggccgaggca aagccgtttt tccatagget ccgccccct 360

gacaagcatc acgaaatctg acgctcaaat cagtgggtgc gaaacccgac aggactataa 420

agataccagg cgtttcccc tggcggtcc ctcgtgcgt ctctgttcc tgcctttcg 480

tttacgggtg tcattccgct gttatggccg cgtttgtctc attccacgcc tgacactcag 540

ttccgggtag gcagttcgct ccaagctgga ctgtatgcac gaacccccg ttcagtccga 600

ccgtgcgcc ttatccggtg actatcgtct tgagtccaac ccggaagac atgcaaaagc 660

accactggca gcagccactg gtaattgatt tagaggagtt agtcttgaag tcatgcgccg 720

gttaaggcta aactgaaagg acaagttttg gtgactgcgc tcctccaagc cagttacctc 780

ggttcaaaga gttgtagct cagagaacct tcgaaaaacc gccctgcaag gcggtttttt 840

cgttttcaga gcaagagatt acgcgcagac caaacgac tcaagaagat catcttatta 900

atcagataaa atatttctag atttcagtgc aatttatctc ttcaaatgta gcacctgaag 960

tcagcccat acgataataa ttgtaattct catgtttgac agcttatcat cgataagctt 1020

taatgcggta gtttatcaca gttaaattgc taacgcagtc aggcacctat acatgcattt 1080

acttataata cagtttttta gttttgctgg ccgcatcttc tcaaatatgc ttcccagcct 1140

gcttttctgt aacgttcacc ctctacctta gcatcccttc cctttgcaa tagtcctctt 1200

ccaacaataa taatgtcaga tcctgtagag accacatcat ccacggttct atactgttga 1260

cccaatgcgt ctcccttctc atctaacc acaccgggtg tcataatcaa ccaatcgtaa 1320

ccttcactctc ttccacccat gtctctttga gcaataaagc cgataacaaa atctttgtcg 1380

ctcttcgcaa tgtcaacagt acccttagta tattctccag tagataggga gcccttgc 1440

gacaattctg ctaacatcaa aaggcctcta ggttcccttg ttaacttctc tgccgcctgc 1500

ttcaaaccgc taacaatacc tgggccacc acaccgtgtg cattcgtaat gtctgccat 1560

tctgtatttc tgiatacacc cgcagagtag tgcaatttga ctgtattacc aatgtcagca 1620

aattttctgt cticgaagag taaaaaattg tacttggcgg ataatgcctt tagcggctta 1680

actgtgccct ccatggaata atcagtcaag atatccacat gtgttttttag taaacaaatt 1740

ttgggaccta atgtttcaac taactccagt aattccttgg tggtagaac atccaatgaa 1800

gcacacaagt ttgtttgctt ttcgtgcag atattaaata gcttggcagc aacaggacta 1860

ggatgagtag cagcacgttc cttatatgta gctttcgaca tgatttatct tcgtttcctg 1920

caggtttttg ttcgtgcag ttgggttaag aatactgggc aatttcattg ttcttcaaca 1980

ctacatatgc gtatatatac caatctaagt ctgtgtcct tcttcgttc ttccttctgt 2040

tccgagatta ccgaatcaaa aaaatttcaa agaaaccgaa atcaaaaaaa agaataaaaa 2100

aaaaatgatg aattgaattg aaaagctagc ttatcgatgg gtccttttca tcacgtgcta 2160

taaaaataat tataatttaa attttttaat ataaatatat aaattaaaaa tagaaagtaa 2220

aaaaagaaat taaagaaaaa atagtttttg tttccgaag atgtaaaaga ctctaggggg 2280

atcgccaaca aatactacct tttatcttgc tcttcctgct ctccaggtatt aatgccgaat 2340

tgtttcatct tgctgtgtga gaagaccaca cagaaaaac ctgtgatttt acattttact 2400

tatcgtaaat cgaatgtata tctatttaat ctgttttct tgtctaataa atatatatgt 2460

aaagtacgct tttgttgaa attttttaaa cctttgttta ttttttttc ttcattccgt 2520

aactcttcta cttcttttat ttactttcta aaatccaaat acaaaacata aaaataaata 2580

aacacagagt aaattcccaa attattccat cattaaaaga tacgaggcgc gtgtaagtta 2640

caggcaagcg atctctaaga aaccattatt atcatgacat taacctataa aaaagcctc 2700

tcgagctaga gtcgatcttc gccagcaggg cgaggatcgt ggcatcacg aaccgcgccg 2760

tcgcggggtc gtcggtgagc cagagtttca gcaggccgcc caggcggccc aggtcgccat 2820

tgatcggggc cagctcgcgg acgtgctcat agtccacgac gccctgatt ttgtagccct 2880

ggccgacggc cagcaggtag gccgacaggc tcatgccggc cgccgccgcc ttttctcaa 2940

tcgctcttcg ttcgtctgga aggcagtaca ccttgatagg tgggctgcc ttcctggttg 3000

gcttggtttc atcagccatc cgcttgccct catctgttac gccggcggtg gccggccagc 3060

ctcgcagagc aggattcccg ttgagcaccg ccaggtgcga ataagggaca gtgaagaagg 3120

aacacccgct cgcggttggg cctacttcac ctatcctgcc cggtgacgc cgttgatac 3180

accaaggaaa gtctacacga accctttggc aaaatcctgt atatcgtgcg aaaaaggatg 3240

gatataccga aaaaatcgct ataatgacc cgaagcaggg ttatgcagcg gaaaagcgct 3300

gcttccctgc tgtttttgg aatatctacc gactggaaac aggcaaatgc aggaaattac 3360

tgaactgagg ggacaggcga gagacgatgc caaagagcta caccgacgag ctggccgagt 3420

gggttgaatc ccgcgcggcc aagaagcgcc ggctgatga ggctgcggtt gcgttcctgg 3480

cggtgagggc ggatgtcgat atgcgtaagg agaaaatacc gcatcaggcg catatttgaa 3540

tgtatttaga aaaataaaca aaaagagttt gtagaaacgc aaaaaggcca tccgtcagga 3600

tggccttctg ctttaattga tgcctggcag tttatggcgg gcgtcctgcc cgccaccctc 3660

cgggccgttg cttcgcaacg ttcaaatccg ctcccggcgg atttgctcta ctcaggagag 3720

cgttcaccga caaacaacag ataaaacgaa aggcccagtc tttcgactga gcctttcgtt 3780

ttatttgatg cctggaaacc cagcgaacca tttgaggtga taggtaagat tataccgagg 3840

tatgaaaacg agaattggac ctttacagaa ttactctatg aagcgccata tttaaaaagc 3900

taccaagacg aagaggatga agaggatgag gaggcagatt gccttgaata tattgacaat 3960

actgataaga taatataatct tttatataga agatatcgcc gtatgtaagg atttcagggg 4020

gcaaggcata ggcagcgcgc ttatcaatat atctatagaa tgggcaaagc ataaaaactt 4080

gcatggacta atgcttgaag cccaggacaa taaccttata gcttgtaaat tctatcataa 4140

ttgtggtttc aaaatcggtc ccgtcgatac tatgttatac gccaaacttc aaaacaactt 4200

tgaaaaagct gttttctggt atttaaggtt ttagaatgca aggaacagtg aattggagtt 4260

cgctcttgta taattagctt cttgggggtat ctttaaatac tgtagaaaag aggaaggaaa 4320

taataaatgg ctaaaatgag aatatcaccg gaattgaaaa aactgatcga aaaataccgc 4380

tgcgtaaaag atacggaagg aatgtctcct gctaaggtat ataagctggt gggagaaaat 4440

gaaaacctat atttaaaaat gacggacagc cgggtataaag ggaccaccta tgatgtggaa 4500

cgggaaaagg acatgatgct atggctggaa ggaaagctgc ctgttccaaa ggtcctgcac 4560

tttgaacggc atgatggctg gagcaatctg ctcatgagtg aggccgatgg cgtcctttgc 4620

tcggaagagt atgaagatga acaaagccct gaaaagatta tcgagctgta tgcggagtgc 4680

atcaggctct ttactccat cgacatatcg gattgtccct atacgaatag cttagacagc 4740

cgcttagccg aattggatta cttactgaat aacgatctgg ccgatgtgga ttgcgaaaac 4800

tgggaagaag acactccatt taaagatccg cgcgagctgt atgatttttt aaagacggaa 4860

aagcccgaag aggaacttgt cttttccac ggcgacctgg gagacagcaa catctttgtg 4920

aaagatggca aagtaagtgg ctttatlgat cttgggagaa gcggcagggc ggacaagtgg 4980



tatgacattg ccttctgcgt cgggtcgatc agggaggata tcggggaaga acagtatgtc 5040

gagctatatt ttgacttact ggggatcaag cctgattggg agaaaataaa atattatatt 5100

ttactggatg aattgtttta gtacctagat ttagatgtct aaaaagcttt ttagacatct 5160

aatcttttct gaagtacatc cgcaactgtc catactctga tgttttatat cttttctaaa 5220

agttcgttag ataggggtcc cgagcgccia cgaggaattt gtatcggctg tacaatctgc 5280

taatactgca aagaatttgt tgggctttga accaaaagtt gctatgctat cattttccac 5340

aaaaggtagt gcatcacatg aattagtaga taaagtaaga aaagcgacag aaatagcaaa 5400

agaattgatg ccagatgttg ctatcgacgg tgaattgcaa ttggatgctg ctcttgtcaa 5460

agaagtigca gagctaaaag cgccaggaag caaagttgcg ggatgtgcaa atgtgcttat 5520

attccctgat ttacaagctg gtaatatagg atataagctt gtacagagat tagctagcaa 5580

atgcaattgg acctataaca caggaatggg tgcaccggtt aatgatattat caagaggatg 5640

cagctataga gatattgttg acgtaatagc acacagctgt acaggctcat aaatgtaaag 5700

tatggaggat gaaattatga aaatactggt atatgcgaag tctcactaaa tatcactgat 5760

gatcatgatg aaatgtgctg gcaaaggcct tgctgagaga atcggcataa atgatccctg 5820

ttgacacata tgctaacgga gaaaaatcaa gataaaaaaa gacatgaaag atcacaaga 5880

cgcaataaaa ttgttttaga tgcttttgta agcagtgact acggcgttat aaaggatatg 5940

tctgagatag atgctgtagg acatagagtt gttcacggag gagaatcttt tacatcatca 6000

gttctcataa atgatgatgt gttaaaagcg ataacagatt gcatagaatt agctccactg 6060

cacaatcctg ccaatataga aggaat 6086

<211> 7284  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: Synthetic  
 polynucleotide

<400> 56  
 taacagggcc atgaccttct gcaaagcact tttttaaaat ttcaagagaa tattgtggca 60  
  
 tatatttttc taatgcagaa tagtttttat catttataac ctttctaaga tatgtagcac 120  
  
 ttgcaaaaga ttctactgtt tccagagaat tatatcctgg accaaatctt tttaatgtca 180  
  
 aaggcttaat attgctgttt aacttgatta agctttttat gtattcaatt ccaagtatgt 240  
  
 tatttggatt tccgataata ttgttaatgg tccttccgaa atattttgat aatgctgctt 300  
  
 cacgtgcctt tgcatatgtg atgccgcttt ttaaataagg tttcaaagag ctttataatc 360  
  
 ttctggttcg tctaaaagaa attttgatat ttggcataat tcgtcaatag agccatgttc 420  
  
 gcttccaaac gatatgtaat caatgacatt taatgaatcc agcagcttta ctgctccata 480  
  
 tgcgaaatth tctgcggtag aaactgcata tatagtgggc aattctatta ctaagtcaat 540  
  
 acctgatagc aatgcagctt ctgtccttga ccacttgtca ataataagac gtattcctcg 600  
  
 ctgaacaaag ttaccactca taattgcaat gacaaaatct gcacctgttg tttcaattga 660  
  
 tttttttata tggatatatg gtccgttatg gagagggtta tattctacaa ttactcccaa 720  
  
 tatactcatt attaaaaacc tttctaaaaa attattaatt gtacttatta ttttataaaa 780  
  
 aatatgttaa aatgtaaaat gtgtatacaa tatatttctt ctttttagta agaggaaatgt 840  
  
 ataaaaataa atatttttaa ggaaggacg atcttatgag cattattcaa aacatcattg 900  
  
 aaaaagctaa aagtgataaa aagaaaattg ttttgccgga aggtgcagaa acccagcgaa 960

ccatttgagg tgataggtaa gattataccg aggtatgaaa acgagaattg gacctttaca 1020

gaattactct atgaagcgcc atatttaaaa agctaccaag acgaagagga tgaagaggat 1080

gaggaggcag attgccttga atatatggac aatactgata agataatata tcttttatat 1140

agaagatata gccgtatgta aggatttcag ggggcaaggc ataggcagcg cgcttatcaa 1200

tatatctata gaatgggcaa agcataaaaa ctgcatgga ctaatgcttg aaaccagga 1260

caataacctt atagcttgta aattctatca taattgtggt ttcaaaatcg gctccgtcga 1320

tactatgtta tacgccaact ttcaaaacaa ctttgaaaaa gctgttttct ggtatttaag 1380

gttttagaat gcaaggaaca gtgaattgga gttcgtcttg ttataattag ctctttgggg 1440

tatctttaaa tactgtagaa aagaggaagg aaataataaa tggctaaaat gagaatatca 1500

ccggaattga aaaaactgat cgaaaaatac cgtgcgtaa aagatacga aggaatgtct 1560

cctgctaagg tatataagct ggtgggagaa aatgaaaacc tatatttaaa aatgacggac 1620

agccggtata aagggaccac ctatgatgtg gaacgggaaa aggacatgat gctatggctg 1680

gaaggaaagc tgctgttcc aaaggtcctg cactttgaac ggcatgatgg ctggagcaat 1740

ctgctcatga gtgaggccga tggcgctctt tgctcggaag agtatgaaga tgaacaaagc 1800

cctgaaaaga ttatcgagct gtatgcggag tgcacaggc tctttcactc catcgacata 1860

tcggattgtc cctatacgaa tagcttagac agccgcttag ccgaattgga ttacttactg 1920

aataacgata tggccgatgt ggattgcgaa aactgggaag aagacactcc atttaaagat 1980

ccgcgcgagc tgatgatatt tttaaagacg gaaaagcccg aagaggaact tgtcttttcc 2040

cacggcgacc tgggagacag caacatcttt gtgaaagatg gcaaagtaag tggctttatt 2100

gatcttggga gaagcggcag ggcggacaag tggatatgaca ttgccttctg cgtccggtcg 2160

atcagggagg atacgggga agaacagtat gtcgagctat tttttgactt actggggatc 2220

aagcctgatt gggagaaaaat aaaatattat attttactgg atgaattgtt ttagtaccta 2280

gatttagatg tctaaaaagc tttttagaca tctaattctt tctgaagtac atccgcaact 2340

gtccatactc tgatgtttta tatcttttct aaaagtctgc tagatagggg tcccagcgc 2400

ctacgaggaa tttgtatcga aagttagcgt gatggttggtg ccactaatg aagaatacat 2460

gattgctaaa gatactgaaa agattgtaa gagtataaaa tagcattctt gacaaatgtt 2520

tacccatta gtataattaa ttttggcaat tatattgggg tgagaaaatg aaaattgatt 2580

tatcaaaatt aaggacata ggggccgcag catcgaagtc aactacgtag aaaatctgag 2640

tgttcttgag gcaaatagca atagatactg agttataaag cctattagcg taactggaag 2700

cataacatac gatagtgaag gaatagtttt aaaacttttg gcacgcgggg ctattaaagt 2760

aacatcgat aggtgccttg acgaatttga gtatgagttc gtaataccta ttgacgaaat 2820

agtaaacgag tctgatgatg aattttcagg tgaagtggaa gatgaaaagc ttgatttgac 2880

gaaaattgtg atigaaaatg tggaactttc tcttccgatg aagttcattt gctcgaatga 2940

ttgcaaggtt ctatgttcta ctgacggtta aaatcttaat catgaaaaat gcgattgcca 3000

aataaaagaa attgatccac gcctttcagt tttgaataaa ttactgcaga agatgtagga 3060

ggtgtataat atgccagttc caaagcgtag aacatctaag gcaagaagag ataaaagaag 3120

gcatagccat agtttagctg tacctgctta tgttttgtgc ccacaatgtc atgaacaaaa 3180

attgccccac agagtttgtt taagctgtgg ttattacgac ggtaaagagg tattgaaagt 3240

ggaagaaaag taatggagtt ttctctatta cttttctttt ttatttcttg acttttatgt 3300

atggcgtaat ttataattat gagtaagtca taaaaacaac ctatatTTGG agctgataat 3360

gtggccacga agcttagtaa aagagataga ttaaaaaagt taaaaattga aatcgaaaaa 3420

tatccatttt acactgatga tgagttagct gatttgtttt cggttagcgt tcagacgata 3480

aggctggatt ctatcagctg tccctcctgt tcagctactg acgggggtgt gcgtaacggc 3540

aaaagcaccg ccggacatca gcgctagcgg agtgtatact ggcttactat gttggcactg 3600

atgagggtgt cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtgcgta 3660

gcagaatatg tgatacagga tatattccgc ttctcgcgc actgactcgc tacgctcgg 3720

cgttcgactg cggcgagcgg aaatggctta cgaacggggc ggagatttcc tggaagatgc 3780

caggaagata cttaacaggg aagtgagagg gccgcggcaa agccgttttt ccataggctc 3840

cgccccctg acaagcatca cgaaatctga cgtcaaatc agtggtggcg aaaccgcaca 3900

ggactataaa gataccaggc gtttccccct ggcggtcccc tcgtgcgcgc tctgttcc 3960

gcctttcggg ttaccgggtg cattccgctg ttatggccgc gtttgtctca ttccacgcct 4020

gacactcagt tccgggtagg cagttcgctc caagctggac tgtatgcacg aacccccgt 4080

tcagtccgac cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc cggaaagaca 4140

tgcaaaagca cactggcag cagccactgg taattgattt agaggagtta gtcttgaagt 4200

catgcgccgg ttaaggctaa actgaaagga caagttttgg tgactgcgct cctccaagcc 4260

agttacctcg gttcaaagag ttggtagctc agagaacctt cgaaaaaccg ccctgcaagg 4320

cggttttttc gttttcagag caagagatta cgcgcagacc aaaacgatct caagaagatc 4380

atcttattaa tcagataaaa ttttctaga ttccagtga atttatctct tcaaatgtag 4440

cacctgaagt cagcccccata cgatataagt tgaattctc atgtttgaca gcttatcatc 4500

gataagcttt aatgcggttag tttatcacag ttaaattgct aacgcagtca ggcacctata 4560

catgcattta cttataatac agtttttttag ttttgctggc cgcatcttct caaatatgct 4620

tcccagcctg cttttctgta acgttcaccc tctaccttag catcccttcc ctttgcaaat 4680

agtcctcttc caacaataat aatgtcagat cctgtagaga ccacatcatc cacggttcta 4740

tactgttgac ccaatgcgtc tcccttgta tctaaacca caccgggtgt cataatcaac 4800

caatcgtaac cttcatctct tccacccatg tctctttgag caataaagcc gataacaaaa 4860

tctttgtcgc tcttcgcaat gtcaacagta cccttagtat attctccagt agataggag 4920

cccttgcatg acaattctgc taacatcaaa aggctcttag gttcctttgt tacttcttct 4980

gccgcctgct tcaaaccgtt aacaatacct gggcccacca caccgtgtgc attcgtaatg 5040

tctgcccatt ctgtattctt gtatacacc gcagagtact gcaatttgac tgtattacca 5100

atgtcagcaa attttctgtc ttcgaagagt aaaaaattgt acttggcgga taatgccttt 5160

agcggcttaa ctgtgccctc catggaaaaa tcagtcaaga tatccacatg tgtttttagt 5220

aaacaaattt tgggacctaa tgcttcaact aactccagta attccttggt ggtacgaaca 5280

tccaatgaag cacacaagtt tgtttgcttt tcgtgcatga tattaaatag cttggcagca 5340

acaggactag gatgagtagc agcacgttcc ttatatgtag ctttcgacat gatttatctt 5400

cgtttcctgc aggtttttgt tctgtgcagt tgggttaaga atactgggca atttcatgtt 5460

tcttcaacac tacatatgcg tatatatacc aatctaagtc tgtgtctctt ccttcgttct 5520

tccttctgtt cggagattac cgaatcaaaa aaatttcaaa gaaaccgaaa tcaaaaaaaaa 5580

gaataaaaaa aaaatgatga attgaattga aaagctagct tatcgatggg tccttttcat 5640

cacgtgctat aaaataatt ataattttaa ttttttaata taaatatata aattaaaaat 5700

agaaagtaaa aaaagaaatt aaagaaaaaa tagtttttgt tttccgaaga tgtaaaagac 5760

tctaggggga tcgccaacaa atactacctt ttatcttgct cttcctgctc tcaggtatta 5820

atgccgaatt gtttcatctt gtctgtgtag aagaccacac acgaaaatcc tgtgatttta 5880

cattttactt atcgttaatc gaatgtatat ctatttaatc tgcttttctt gtctaataaa 5940

tatatatgta aagtacgctt tttgttgaaa ttttttaaac ctttgtttat tttttttct 6000

tcattccgta actctttctac cttctttatt tactttctaa aatccaaata caaaacataa 6060

aaataaataa acacagagta aattcccaaa ttattccatc attaaaagat acgaggcgcg 6120

tgtaagttac aggcaagcga tctctaagaa accattatta tcatgacatt aacctataaa 6180

aaaggcctct cgagctagag tcgatcttcg ccagcagggc gaggatcgtg gcatcaccga 6240

accgcgccgt gcgcgggtcg tcggtgagcc agagtttcag caggccgccc aggcggccca 6300

ggtcgccatt gatcggggcc agctcgcgga cgtgctcata gtccacgacg cccgtgattt 6360

tgtagccctg gccgacggcc agcaggtagg ccgacaggct catgccggcc gccgccgcct 6420

tttctcaat cgctcttcgt tcgtctggaa ggcagtacac cttgataggt gggctgcctt 6480

tcctggtttg cttggtttca tcagccatcc gcttgccctc atctgttacg ccggcggtag 6540

ccggccagcc tcgcagagca ggattcccgt tgagcaccgc caggtagcgaa taaggacag 6600

tgaagaagga acacccgctc gcgggtgggc ctacttcacc tatcctgccc ggctgacgcc 6660

gttgataca ccaaggaaag tctacacgaa ccttttgga aaatcctgta tatcgtgcga 6720

aaaaggatgg atataccgaa aaaatcgcta taatgacccc gaagcagggt tatgcagcgg 6780

aaaagcgctg cttccctgct gttttgtgga atatctaccg actggaaaca ggcaaatgca 6840  
 ggaaattact gaactgaggg gacaggcgag agacgatgcc aaagagctac accgacgagc 6900  
 tggccgagtg ggttgaatcc cgcgcgccca agaagcgccg gcgtgatgag gctgcggttg 6960  
 cgttcctggc ggtgagggcg gatgtcgata tgcgtaagga gaaaataccg catcaggcgc 7020  
 atatttgaat gtatttagaa aaataaaca aaagagtttg tagaaacgca aaaaggccat 7080  
 ccgtcaggat ggccttctgc ttaatttgat gcctggcagt ttatggcggg cgtcctgccc 7140  
 gccaccctcc gggccgttgc ttgcaacgt tcaaatccgc tcccggcgga ttgtcctac 7200  
 tcaggagagc gttcacgac aaacaacaga taaaacgaaa ggcccagtct ttcgactgag 7260  
 cctttcgttt tatttgatgc ctgg 7284

<210> 57

<211> 4621

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 57

cagcacaaaa accgggagaa accagacttg acttagtgaa gagaaatacg atgattttta 60  
 aagacatagt ggcaaaactt attaaagtaa atgacacagc aatatacctt atagttacaa 120  
 atccagtaga tattcttaca tacgttacct ataaaatatac tggcttgcca tacggaagag 180  
 tattggggtc tggcacagtt ctcgacagtg cgagattcag atatctttta agcaaacatt 240  
 gtaacataga tccgaggaat atacacggat atataattgg ggagcatggc gattctgagc 300  
 ttgcagcttg gagcattacg aacatagcag gcataccaat tgataattac tgcaatttat 360



gtggaaaagc atgtgaaaaa gatttttagag aggagatttt taataatgtt gtaagagctg 420

cctatacgat aatagaaaaa aagggtgcga catattatgc ggttgctctc gcagtaagaa 480

gaatcgtaga agctattaaa ccacgcgaac catttgaggt gataggttaag attataccga 540

ggtatgaaaa cgagaattgg acctttacag aattactcta tgaagcgcca tatttaaaaa 600

gctaccaaga cgaagaggat gaagaggatg aggaggcaga ttgccttgaa tatattgaca 660

atactgataa gataatatat cttttatata gaagatatcg ccgtatgtaa ggatttcagg 720

gggcaaggca taggcagcgc gcttatcaat atatctatag aatgggcaaa gcataaaaac 780

ttgcatggac taatgcttga aaccaggac aataacctta tagcttgtaa attctatcat 840

aattgtggtt tcaaaatcgg ctccgtcgat actatgttat acgccaactt tcaaaacaac 900

tttgaaaaag ctgttttctg gtatttaagg ttttagaatg caaggaacag tgaattggag 960

ttcgtcttgt tataattagc ttcttggggt atctttaaat actgtagaaa agaggaagga 1020

aataataaat ggctaaaatg agaatatcac cggaattgaa aaaactgatc gaaaaatacc 1080

gctgcgtaaa agatacggaa ggaatgtctc ctgctaaggt atataagctg gtgggagaaa 1140

atgaaaacct atatttaaaa atgacggaca gccggtataa agggaccacc tatgatgtgg 1200

aacgggaaaa ggacatgatg ctatggctgg aaggaaagct gcctgttcca aaggtcctgc 1260

actttgaacg gcatgatggc tggagcaatc tgctcatgag tgaggccgat ggcgtccttt 1320

gctcggaaga gtatgaagat gaacaaagcc ctgaaaagat tatcgagctg tatgcggagt 1380

gcatcaggct ctttactcc atcgacatat cggattgtcc ctatacgaat agcttagaca 1440

gccgcttagc cgaattggat tacttactga ataacgatct ggccgatgtg gattgcgaaa 1500

actgggaaga agacactcca tttaaagatc cgcgcgagct gtatgatgtt ttaaagacgg 1560

aaaagccga agaggaactt gtcttttccc acggcgacct gggagacagc aacatctttg 1620  
 tgaaagatgg caaagtaagt ggctttattg atcttgggag aagcggcagg gcggacaagt 1680  
 ggtatgacat tgccttctgc gtccggtcga tcaggaggga tatcggggaa gaacagtatg 1740  
 tcgagctatt ttttgactta ctggggatca agcctgattg ggagaaaata aaatattata 1800  
 ttttactgga tgaattgttt tagtacctag atttagatgt ctaaaaagct ttttagacat 1860  
 ctaatctttt ctgaagtaca tccgcaactg tccatactct gatgttttat atcttttcta 1920  
 aaagttcgct agataggggt cccgagcgcc tacgaggaat ttgtatcgct gcaggcatgc 1980  
 aagcttggcg taatcatggt catagctgtt tcctgtgtga aattgtttatc cgctcacaat 2040  
 tccacacaac atacgagccg gaagcataaa gtgtaaagcc tggggtgcct aatgagttag 2100  
 ctaactcaca ttaattgctg tgcgtcact gcccgcttcc cagtcgggaa acctgtcgtg 2160  
 ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta ttgggcgctc 2220  
 ttccgcttcc tcgtcactg actcgtcgcg ctccgtcgtt cggctgcggc gagcggtatc 2280  
 agctcactca aaggcggtaa tacggttatt cacagaatca ggggataacg caggaaagaa 2340  
 catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt tgctggcgtt 2400  
 ttccatagg ctccgcccc ctgacgagca tcacaaaaat cgacgtcaa gtcagaggtg 2460  
 gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct ccctcgtgcg 2520  
 ctctcctgtt ccgacctgc cgttaccgg atacctgtcc gcctttctcc ctccgggaag 2580  
 cgtggcgctt tctcatagct cacgtgtag gtatctcagt tcggtgtagg tcgttcgctc 2640  
 caagctgggc tgtgtgcacg aacccccgt tcagcccgac cgctgcgcct tatccggtaa 2700

ctatcgtctt gagtccaacc cggtaaagaca cgacttatcg ccactggcag cagccactgg 2760

taacaggatt agcagagcga ggtatgtagg cggtagctaca gatttcttga agtggtaggc 2820

taactacggc tacactagaa ggacagtatt tggtagctgc gctctgctga agccagttac 2880

cttcggaaaa agagttagta gctcttgatc cggcaaaaa accaccgctg gtagcgttgg 2940

ttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag aagatccttt 3000

gatcttttct acggggctcg acgctcagtg gaacgaaaac tcacgttaag ggatttttgt 3060

catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat gaagtttta 3120

atcaatctaa agtatatatg agtaaacttg gcttgacagt taccaatgct taatcagtga 3180

ggcacctatc tcagcgatct gtctatttcg ttcattcata gttgcctgac tccccgtcgt 3240

gtagataact acgatacggg agggcttacc atctggcccc agtctgcaa tgataccgcg 3300

agaccacgc tcaccggctc cagatttacc agcaataaac cagccagccg gaagggccga 3360

gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt gttgccggga 3420

agctagagta agtagttcgc cagttaatag ttgacgcaac gttgttgcca ttgctacagg 3480

catcgtggtg tcacgctcgt cgtttgggat ggcttcattc agtccggtt cccaacgac 3540

aaggcgagtt acatgatccc ccatgttgtag caaaaaagcg gtagctcct tcggtcctcc 3600

gatcgtgtc agaagtaagt tggccgcagt gttatcactc atggttatgg cagcactgca 3660

taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg agtactcaac 3720

caagtcattc tgagaatagt gtatcggcgc accgagttgc tcttccccgg cgtcaatacg 3780

ggataatacc gcgccacata gcagaacttt aaaagtgtc atcatggaa aacgtttctc 3840

ggggcgaaaa ctctcaagga tcttaccgt gttgagatcc agttcgatgt aaccactcg 3900

tgcaccaac tgatcttcag catcttttac tttcaccage gtttctgggt gagcaaaaac 3960

aggaaggcaa aatgccgcaa aaaagggaaat aagggcgaca cggaaatgtt gaatactcat 4020

actcttcctt tticaatatt attgaagcat ttatcagggt tattgtctca tgagcggata 4080

catatttgaa tgiattttaga aaaataaaca aataggggtt ccgcgcacat tccccgaaa 4140

agtgccacct gacgtctaag aaaccattat tatcatgaca ttaacctata aaaataggcg 4200

tatcacgagg ccttttcgtc tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat 4260

gcagctcccg gagacgggtca cagcttgtct gtaagcggat gccgggagca gacaagcccc 4320

tcagggcgcg tcagcgggtg ttggcgggtg tcggggctgg cttaactatg cggcatcaga 4380

gcagattgta ctgagagtgc accatatgcg gtgtgaaata ccgcacagat gcgtaaggag 4440

aaaataccgc atcaggcgcc attcgccatt caggctgcgc aactgttggg aagggcgatc 4500

ggtgcgggcc tcttcgctat tacgccagct ggcgaaaggg ggatgtgctg caaggcgatt 4560

aagttgggta acgccagggt tttccagtc acgacgttgt aaaacgacgg ccagtgaatt 4620

c 4621

<210> 58

<211> 6130

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 58

aggtttatcg ccgccttgta cagttttaaa ttgcatagaa atcatgagac ccattgcaat 60

|                                                                   |      |
|-------------------------------------------------------------------|------|
| aaatacaagt atcaaagata aagtttgcaa tgcctttcct ttcaatttct ccacatcctt | 120  |
| tctctatata aaaagacatc ttcgctttgc ttttggtttc agcttatatg cacttttata | 180  |
| aataactatg atactctata aatactataa catggaaaat gttaaaattt attaagaaat | 240  |
| tattaagttt ttattacaaa aaagttacaa aacctctgac atttttcata tcagaggttg | 300  |
| tcatttttta ttttatttc tatagaattt tttagtaca atatttcttc taattcttta   | 360  |
| ttgtatttat ctattttcaa catggtactt ctatataggc gtatatcttc ttcgtttttt | 420  |
| tgtatacatt ttttaaggga gttttttaca gtttcaaaaa gcgtatcata agtaatgtaa | 480  |
| ttatgcattt caaggtcgga gattggaact gcgattaatt cctcccttc tattttatag  | 540  |
| tgataaaaaa tgttgcggg ttttaaaact atttttcat ctatgctttt atcgtaaatt   | 600  |
| attatacctt ctctatttc aaaagttttt cgcctataac ttctaagttt tatattttcg  | 660  |
| actcctatgt atttttttat tgccaaaagt atgtttttta tttctgaaat ggattttaca | 720  |
| agcacttctt gcattttttt atttgccatc tctttatctt tttcactttt tatgagtctt | 780  |
| tccatcaaag actttatttc atgtatatct cccataaaat atcacctctt tcttaatatt | 840  |
| ccacagagga atcattttaa acgttgaata ttttaaatTA ttagagaaaa aatagacttg | 900  |
| actatttttt gaaatttgat agactattat taatagaaaa ttaatatga aaaggagaag  | 960  |
| atattatgaa caaaatatct ataataggtt ctggatttgt cggaaccca gcgaaccatt  | 1020 |
| tgaggtgata ggtaagatta taccgaggta tgaaaacgag aattggacct ttacagaatt | 1080 |
| actctatgaa gcgccatatt taaaaagcta ccaagacgaa gaggatgaag aggatgagga | 1140 |
| ggcagattgc cttgaatata ttgacaatac tgataagata atatatcttt tatatagaag | 1200 |
| atatcgccgt atgtaaggat ttacaggggc aaggcatagg cagcgcgctt atcaatatat | 1260 |

ctatagaatg ggcaaagcat aaaaacttgc atggactaat gcttgaaacc caggacaata 1320

accttatagc ttgtaaattc tatcataatt gtggtttcaa aatcggtcc gtcgatacta 1380

tgttatacgc caactttcaa aacaactttg aaaaagctgt tttctggtat ttaaggtttt 1440

agaatgcaag gaacagtga tggagttcg tcttggtata attagcttct tggggtatct 1500

ttaaatactg tagaaaagag gaaggaaata ataaatggct aaaatgagaa taccacgga 1560

attgaaaaaa ctgatcga aataccgctg cgtaaaagat acggaaggaa tgtctcctgc 1620

taaggtatat aagctggtgg gagaaaatga aaacctatat ttaaaaatga cggacagccg 1680

gtataaagg accacctatg atgtggaacg gaaaaaggac atgatgctat ggctggaagg 1740

aaagctgcct gttccaaagg tctgcactt tgaacggcat gatggctgga gcaatctgct 1800

catgagttag gccgatggcg tcctttgctc ggaagagtat gaagatgaac aaagccctga 1860

aaagattatc gagctgtatg cggagtgc atcaggctctt cactccatcg acatctcgga 1920

ttgtccctat acgaatagct tagacagccg cttagccgaa ttggattact tactgaataa 1980

cgatctggcc gatgtggatt gcgaaaactg ggaagaagac actccattta aagatccgcg 2040

cgagctgtat gatTTTTTaa agacggaaaa gcccgagag gaacttgtct tttccacgg 2100

cgacctggga gacagcaaca tctttgtgaa agatggcaaa gtaagtggct ttattgatct 2160

tgggagaagc ggcagggcgg acaagtggta tgacattgcc ttctgcgtcc ggtcgatcag 2220

ggaggatatc ggggaagaac agtatgtcga gctatttttt gacttactgg ggatcaagcc 2280

tgattgggag aaaataaaat attatatttt actggatgaa ttgttttagt acctagattt 2340

agatgtctaa aaagcttttt agacatctaa tctttctga agtacatccg caactgtcca 2400

tactctgatg ttttatactt tttctaaaag ttcgctagat aggggtcccg agcgcctacg 2460

aggaatttgt atcgaagatc agccgaagtt atcaaaagtg taatacaaga gcttgatata 2520

taagagggga aaccctcttt tttgtatat aaaaagtcac agcgtgaaaa tataataatt 2580

aaaaaatga ttttttaggg tgtgatagtc gtgcagaaaa taactcagca ggagattatt 2640

ttaagtcct ttgttgaagc acaaaattta gaaaagatac tgttggataa agtaagagaa 2700

tatgggaaag aatcagtaga taatcaaata aaagcattgt taaagcaaat tgaaataatg 2760

ataaaaaatc ataaagaaga cataaaaaag gcacaaaaga ctatgcatat taattccctt 2820

gtcaaaaaaa atatgtctca agagccttta gacatgcttc aagatttatt aaaaaattta 2880

gttaatatcc aagcctttta taatgaaact gtgtgaata ttactaatcc ttacgttaga 2940

cagttgttta ctcaaatgag ggatgatgtt atgagattta tttctattct tcaaatggag 3000

atgaaaagtc tggaatcgaa accttctatt ccaaataaca cagttttaaa tacaccggag 3060

atgagttaat atgaaagtg ctattattgg tgctgggtgt tcagggtgg ctgcggcaat 3120

tacttttcaa aggtatggca ttacaccaga tatttttgaa aaaaagtga aaataggtga 3180

attatttaac catgttgagg ggttattaaa agtgataaat aggcctataa aggatccgct 3240

tcatcatctt aaaaatgttt atggaataga agttaaacca attaacacca ttgacaaaat 3300

agtaatgaag gggccaactg taacagcttc tgttactggg agtaatcttg ggtatatgat 3360

tttaagagga caggacgcaa actctcttga aaatcaattg tataataagt tagaaatacc 3420

agttaatttc aatatagaag ctgattataa gaagttaaaa aataattacg attatgtctg 3480

caggcatgca agcttggcgt aatcatggtc atagctgttt cctgtgtgaa attgttatcc 3540

gtcacaatt ccacacaaca tacgagccgg aagcataaag tgtaaagcct ggggtgccta 3600

atgagtgagc taactacat taattgcgtt gcgtcactg cccgtttcc agtcgggaaa 3660

cctgtcgtgc cagctgcatt aatgaatcgg ccaacgcgcg gggagaggcg gtttgcgtat 3720

tgggcgtctt tccgtttcct cgctcactga ctgcgtgcgc tcggtcgttc ggctgcggcg 3780

agcggtatca gtcactcaa aggcggtaat acggttatcc acagaatcag gggataacgc 3840

aggaaagaac atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt 3900

gctggcgttt ttccataggc tccgcccccc tgacgagcat cacaaaaatc gacgtcaag 3960

tcagaggtgg cgaaaccga caggactata aagataccag gcgtttcccc ctggaagtc 4020

cctcgtgcgc tctcctgttc cgacctgcc gcttaccgga tacctgtccg cttttctccc 4080

ttcgggaagc gtggcgttt ctcatagtc acgtgtagg tatctcagtt cgggttaggt 4140

cgttcgtcc aagctgggt gtgtgcacga acccccgtt cagcccgacc gtgcgcctt 4200

atccgtaac tatcgtcttg agtccaacc ggtaagacac gacttatcgc cactggcagc 4260

agccactggt aacaggatta gcagagcgag gtatgtaggc ggtgctacag agttcttgaa 4320

gtggtggcct aactacggct acactagaag gacagtattt ggtatctgcg ctctgtgaa 4380

gccagttacc ttcggaaaaa gaggtagtag ctcttgatcc ggcaaaaaa ccaccgtgg 4440

tagcggtagt tttttgttt gcaagcagca gattacgcgc agaaaaaaag gatctcaaga 4500

agatcctttg atcttttcta cggggtctga cgctcagtgg aacgaaaact cacgttaagg 4560

gattttggtc atgagattat caaaaaggat cttcacctag atccttttaa attaaaaatg 4620

aagttttaaa tcaatctaaa gtatatatga gtaaaactgg tctgacagtt accaatgctt 4680

aatcagttag gcacctatct cagcgatctg tctatttcgt tcatccatag ttgcctgact 4740



ccccgctgtg tagataacta cgatacggga gggcttacca tctggcccca gtgctgcaat 4800  
 gataccgcga gaccacgct caccggctcc agatttatca gcaataaacc agccagccgg 4860  
 aagggccgag cgcagaagtg gtccigcaac ttatccgcc tccatccagt ctattaattg 4920  
 ttgccgggaa gctagagtaa gtagttcgcc agttaatagt ttgcgcaacg ttgttgccat 4980  
 tgctacaggc atcgtgggtg cagctcgtc gtttggtatg gcttcattca gctccggttc 5040  
 ccaacgatca aggcgagtta catgatcccc catgttgtgc aaaaaagcgg ttagtcctt 5100  
 cggctctcgc atcgttgta gaagtaagt ggccgcagtg ttatcactca tggttatggc 5160  
 agcactgcat aattctctta ctgtcatgcc atccgtaaga tgcttttctg tgactggtga 5220  
 gtactcaacc aagtcattct gagaatagt tatgcggcga ccgagttgct ctgcccggc 5280  
 gtcaatacgg gataataccg cgccacatag cagaacttta aaagtgtca tcattggaaa 5340  
 acgttcttcg gggcgaaaac tctcaaggat cttaccgctg ttgagatcca gttcgatga 5400  
 acccactcgt gcacccaact gatcttcagc atcttttact ttcaccagcg tttctgggtg 5460  
 agcaaaaaca ggaaggcaaa atgccgcaaa aaagggaata agggcgacac ggaaatgttg 5520  
 aatactcata ctcttccttt ttcaatattia ttgaagcatt tatcagggtt attgtctcat 5580  
 gagcggatac atatitgaat gtattitagaa aaataaaca ataggggttc cgcgcacatt 5640  
 tccccgaaaa gtgccacctg acgtctaaga aaccattatt atcatgacat taacctataa 5700  
 aaataggcgt atcacgaggc ctttcgtct cgcgcgtttc ggtgatgacg gtgaaaacct 5760  
 ctgacacatg cagctcccg agacggtcac agcttgctg taagcggatg ccgggagcag 5820  
 acaagcccgt caggcgcggt cagcgggtgt tggcgggtgt cggggctggc ttaactatgc 5880  
 ggcatcagag cagattgtac tgagagtga ccatatgcgg tgtgaaatac cgcacagatg 5940

cgtaaggaga aaataaccga tcaggcgcca ttccgcatc aggtgcgca actgttgga 6000

aggcgatcg gtgcgggcct cttcgctatt acgccagctg gcgaaaggg gatgtgctgc 6060

aaggcgatta agttgggtaa cgccagggtt ttcccagtc cgacgttgta aaacgacggc 6120

cagtgaattc 6130

<210> 59

<211> 4680

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 59

tattgattga cgatgaggtg atcaaaaaat tagaagcatg tattgacctt gcacctttgc 60

acaatcctgc taatattgag ggaataaaaag cttgtcggca gataatgcc ggggtgccaa 120

tggtagcagt tttgatacg gctttccatc aaacaatgcc agattatgcg tatatttacc 180

ccattcctta tgaatactac gaaaaataa gaataagaag atatggattc catgggactt 240

ctcataaata tgtatcttta agagctgctg aaatattaaa gaggcctatt gaagagttaa 300

aaattattac ttgccattta gggaatgggt ctagtattgc tgcggttaaa ggcggttaagt 360

cgatagatac aagtatggga ttactccat tagaagggtt ggctatgggt acaaggtccg 420

gaaatgttga tccttcaatt ataactttct taatggaaaa agaaggattg actgcagaac 480

aggttataga tatacttaat aagaaatcag gtgtatacgg aatttcagga ataagtaatg 540

actttagaga tatagaaaac ccagcgaacc atttgagggt ataggtaaga ttataccgag 600

gtatgaaaac gagaattgga cctttacaga attactctat gaagcgccat atttaaaaag 660  
 ctaccaagac gaagaggatg aagaggatga ggaggcagat tgccttgaat atattgacaa 720  
 tactgataag ataatatatc ttttatatag aagatatcgc cgtatgtaag gatttcaggg 780  
 ggcaaggcat aggcagcgcg cttatcaata tatctataga atgggcaaag cataaaaact 840  
 tgcattggact aatgcttgaa acccaggaca ataacctat agcttgtaaa ttctatcata 900  
 attgtggttt caaaatcggc tccgtcgata ctatgttata cgccaacttt caaaacaact 960  
 ttgaaaaagc tgttttctgg tatttaaggt tttagaatgc aaggaacagt gaattggagt 1020  
 tcgtcttgtt ataattagct tcttggggta tctttaaata ctgtagaaaa gaggaaggaa 1080  
 ataataaatg gctaaaatga gaatatcacc ggaattgaaa aaactgatcg aaaaataccg 1140  
 ctgcgtaaaa gatacggaaag gaatgtctcc tgctaaggta tataagctgg tgggagaaaa 1200  
 tgaaaaccta tatttaaaaa tgacggacag ccggtataaa gggaccacct atgatgtgga 1260  
 acgggaaaag gacatgatgc tatggctgga aggaaagctg cctgttccaa aggtcctgca 1320  
 ctttgaacgg catgatggct ggagcaatct gctcatgagt gaggccgatg gcgtcctttg 1380  
 ctcggaagag tatgaagatg aacaaagccc tgaaaagatt atcgagctgt atgcggagtg 1440  
 catcaggctc ttacttcca tcgacatatc ggattgtccc tatacgaata gcttagacag 1500  
 ccgcttagcc gaattggatt acttactgaa taacgatctg gccgatgtgg attgcgaaaa 1560  
 ctgggaagaa gacactccat ttaaagatcc gcgcgagctg tatgattttt taaagacgga 1620  
 aaagcccgaa gaggaacttg tcttttccca cggcgacctg ggagacagca acatctttgt 1680  
 gaaagatggc aaagtaagtg gctttattga tcttgggaga agcggcaggg cggacaagtg 1740  
 gtatgacatt gccttctgcg tccggtcgat caggaggat atcggggaag aacagtatgt 1800

cgagctatTT tttgacttac tggggatcaa gcctgattgg gagaaaataa aatattatat 1860

tttactggat gaattgtttt agtacctaga tttagatgtc taaaaagctt tttagacatc 1920

taatcttttc tgaagtacat ccgcaactgt ccatactctg atgttttata tcttttctaa 1980

aagttcgcta gataggggtc ccgagcgcct acgaggaatt tgtatcgctg caggcatgca 2040

agcttggcgt aatcatggtc atagctgttt cctgtgtgaa attgttatcc gtcacaatt 2100

ccacacaaca tacgagccgg aagcataaag tgtaaagcct ggggtgccta atgagtgagc 2160

taactcacat taattgcgtt gcgctcactg cccgctttcc agtcgggaaa cctgtcgtgc 2220

cagctgcatt aatgaatcgg ccaacgcgcg gggagaggcg gtttgcgtat tgggcgctct 2280

tccgttctct cgctcactga ctgcctgcgc tcggtcgttc ggctgcggcg agcggtatca 2340

gtcactcaa aggcggtaat acggttatcc acagaatcag gggataacgc aggaaagaac 2400

atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa aggcgcgctt gctggcgttt 2460

ttccataggc tccgcccc tgacgagcat cacaaaaatc gacgctcaag tcagaggtgg 2520

cgaaaccgga caggactata aagataccag gcgtttcccc ctggaagctc cctcgtgcgc 2580

tctcctgttc cgaccctgcc gcttaccgga tacctgtccg cttttctccc ttcgggaagc 2640

gtggcgcttt ctcatagctc acgtgtagg tatctcagtt cggtgtaggt cgttcgctcc 2700

aagctgggct gtgtgcacga accccccgtt cagcccgacc gctgcgcctt atccggtaac 2760

tatcgtcttg agtccaacct ggtaagacac gacttatcgc cactggcagc agccactggt 2820

aacaggatta gcagagcgag gtatgtaggc ggtgctacag agttcttgaa gtggtggcct 2880

aactacggct aactagaag gacagtatTT ggtatctgcg ctctgctgaa gccagttacc 2940

ttcggaaaaa gaggtaggtag ctcttgatcc ggcaaaaaa ccaccgtgg tagcggtggt 3000

ttttttgttt gcaagcagca gattacgcgc agaaaaaaag gatctcaaga agatcctttg 3060

atcttttcta cggggtctga cgctcagtg aacgaaaact cacgttaagg gattttggtc 3120

atgagattat caaaaaggat cttcacctag atccttttaa attaaaaatg aagttttaaa 3180

tcaatctaaa gtatatatga gtaaacctgg tctgacagtt accaatgctt aatcagtgag 3240

gcacctatct cagcgatctg tctatttcgt tcatccatag ttgcctgact ccccgctgtg 3300

tagataacta cgatacggga gggcttacca tctggcccca gtgctgcaat gataccgcga 3360

gaccacgct caccggtcc agatttatca gcaataaacc agccagccgg aagggccgag 3420

cgcagaagtg gtccctgaac ttatccgcc tccatccagt ctattaattg ttgccgggaa 3480

gctagagtaa gtagttgcc agttaatagt ttgcgcaacg ttgttgccat tgctacaggc 3540

atcgtgggtg cagctcgtc gtttggtatg gcttcattca gctccggttc ccaacgatca 3600

aggcgagtta catgatcccc catgttgtgc aaaaaagcgg ttagctcctt cggctcctcg 3660

atcgttgta gaagtaagtt ggccgcagtg ttatcactca tggttatggc agcactgcat 3720

aattctctta ctgtcatgcc atccgtaaga tgcttttctg tgactggtga gtactcaacc 3780

aagtcattct gagaatagtg tatgcggcga ccgagttgct cttgcccggc gtcaatacgg 3840

gataataccg cgccacatag cagaacttta aaagtgtca tcattggaaa acgtttctcg 3900

gggcgaaaac tctcaaggat cttaccgtg ttgagatcca gttcgatga acccactcgt 3960

gcaccaact gatcttcagc atcttttact ttcaccagcg tttctgggtg agcaaaaaca 4020

ggaaggcaaa atgccgcaaa aaagggaata agggcgacac ggaaatgttg aatactcata 4080

ctcttccttt ttcaatatia ttgaagcatt taccaggtt attgtctcat gagcggatac 4140

atatttgaat gtatttagaa aaataaacia ataggggttc cgcgcacatt tccccgaaaa 4200

gtgccacctg acgtctaaga aaccattatt atcatgacat taacctataa aaataggcgt 4260

atcacgaggc cttttcgtct cgcgcgtttc ggtgatgacg gtgaaaacct ctgacacatg 4320

cagctcccg agacggtcac agcttgtctg taagcggatg cggggagcag acaagcccgt 4380

cagggcgctg cagcgggtgt tggcgggtgt cggggctggc ttaactatgc ggcatcagag 4440

cagattgtac tgagagtga ccatatgcgg tgtgaaatac cgcacagatg cgtaaggaga 4500

aaataccgca tcaggcgcca ttgccattc aggtgctgca actgttggga agggcgatcg 4560

gtcggggcct cttcgtatt acgccagctg gcgaaagggg gatgtgctgc aaggcgatta 4620

agttgggtaa cgccagggtt tcccgatca cgacgttgta aaacgacggc cagtgaattc 4680

<210> 60

<211> 5663

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 60

gttttataag aagctttgta aattttattt tctaagcctt cagagacatc aaatatgtga 60

ttgagaggaa tgtgatttct caacagataa attaaaaatat tagaaaattc ctctaaagat 120

acgggacctt gtttttttc aaaagaactt tgcaaaattt cttttgtaaa gtccgggata 180

tatttttcta agccttctag tcccttttca aaaatatgct ttcttataga agaggcactg 240

gcaaattctc cttttagctc taaagagggtg tacaagagc cttttctctt tatagtaaaa 300

ggtgtaatag aactacctat tttctttaaa gatttaaggt attctattgc caatatgtta 360  
 ttggatgttt gtaaaatctt ttctatttca ttattattta taactttttg taatgctaata 420  
 tcccgtagctt ttgcaaaggt tatgccgctt tttaaatatt cttttaatgc ttttctataa 480  
 taaattggct ctctaaaag tatttcagca atttttgtga gttcgtttaa atcgcttttt 540  
 tcactcccaa aggaaaaaca atctactatt tttaaagagt ctaatagttt caccgctcca 600  
 taagcgaaat tttcagctgt agaggttagca taaactactg gtaactcgat taccaaatct 660  
 ataccggctt ttaatgcat ttgagttcgt ttccatttgc ctacaattgc tggttctcct 720  
 ctttgacga agtttccact cattactgct atagtataat cgcatttggg taattctttt 780  
 gaagtttgca gatggttaaag gtggccattg tgaaaaggat tatattcgac aataattcct 840  
 aaaattccca tacaacttct taccctttca aaaaattttt taagatatata ttattatttt 900  
 acataaaata tgataaaatg taaaaggagc atcggtgtata caatattata gtgataaaat 960  
 taaaaaagga agggagattt taaatggcag taatggatag taaaaccag cgaaccattt 1020  
 gaggtgatag gtaagattat accgaggtat gaaaacgaga attggacctt tacagaatta 1080  
 ctctatgaag cgccatatat aaaaagctac caagacgaag aggatgaaga ggatgaggag 1140  
 gcagattgcc ttgaatatat tgacaatact gataagataa tatacttttt atatagaaga 1200  
 tatcgccgta tgtaaggatt tcagggggca aggcattaggc agcgcgctta tcaatatatc 1260  
 tatagaatgg gcaaagcata aaaacttgca tggactaatg cttgaaaccc aggacaataa 1320  
 ccttatagct tgtaaattct atcataattg tggtttcaaa atcggtccg tcgatactat 1380  
 gttatacgcc aactttcaaa acaactttga aaaagctgtt ttctgggtatt taaggtttta 1440  
 gaatgcaagg aacagtgaat tggagttcgt cttgttataa ttagcttctt ggggtatctt 1500

taaatactgt agaaaagagg aaggaaataa taaatggcta aaatgagaat atcacccgaa 1560  
 ttgaaaaaac tgatcgaata ataccgctgc gtaaaagata cggaaggaat gtctcctgct 1620  
 aaggtatata agctgggtggg agaaaatgaa aacctatatt taaaaatgac ggacagccgg 1680  
 tataaaggga ccacctatga tgtggaacgg gaaaaggaca tgatgctatg gctggaagga 1740  
 aagctgcctg ttccaaaggt cctgcacttt gaacggcatg atggctggag caatctgctc 1800  
 atgagtggag ccgatggcgt cctttgctcg gaagagtatg aagatgaaca aagccctgaa 1860  
 aagattatcg agctgtatgc ggagtgcac aggtctcttc actccatcga catatcgat 1920  
 tgtccctata cgaatagctt agacagccgc ttagccgaat tggattactt actgaataac 1980  
 gatctggccg atgtggattg cgaaaactgg gaagaagaca ctccatttaa agatccgcgc 2040  
 gagctgtatg atttttttaa gacgaaaaag cccgaagagg aacttgtctt tttccacggc 2100  
 gacctgggag acagcaacat ctttgtgaaa gatggcaaag taagtggctt tattgatctt 2160  
 gggagaagcg gcagggcgga caatgggtat gacattgcct tctgcgtccg gtcgatcagg 2220  
 gaggatatcg gggaagaaca gtatgtcgag ctatTTTTTg acttactggg gatcaagcct 2280  
 gattgggaga aaataaaata ttatatTTTt ctggatgaat tgttttagta cctagattta 2340  
 gatgtctaaa aagctTTTTt gacatctaatt cttttctgaa gtacatccgc aactgtccat 2400  
 actctgatgt tttatatctt ttctaaaagt tcgctagata ggggtcccg ggcctacga 2460  
 ggaatttgta tcgtgacttt agagatatag aaaatgcagc ttttaaagaa gggcataaaa 2520  
 gggctatgtt ggcattaaaa gttttcgctt atagggtgaa aaagacaata ggttcttata 2580  
 cagctgctat ggggtgggtt gatgtaattg tgtttactgc tggagtggga gaaaatggac 2640



cagaaatgag agagttttatt ttagaggatc tagagttttt aggcctttaaa ctggacaaaag 2700

agaagaataa ggtaagagga aaagaggaaa ttatatctac agaagattca aaagttaaag 2760

ttatggttat tcctacaaat gaagaatata tgattgctaa agatactgaa aaattggtaa 2820

aaggtttaaa gtagataatc ttgacaacgg gttgtggggg tagtataata ggtgatgtca 2880

attattttaa gggtgagaa gaaaaatgaa aatcgatcta ttaaaatca aaggacagct 2940

tggccgcagc ataaatatag actatgtaga ggacatagag aacattgaat ttaaaggga 3000

agaatacaaa ctgcaggcat gcaagcttgg cgtaatcatg gtcatactg tttcctgtgt 3060

gaaattgtta tccgtcaca attccacaca acatacgagc cggaagcata aagtgtaaag 3120

cctgggggtgc ctaatgagt agctaactca cattaatgc gttgcgtca ctgcccgctt 3180

tccagtcggg aaacctgtcg tgccagctgc attaatgaat cgccaacgc gcggggagag 3240

gcggtttgcg tattgggcgc tcttcgctt cctcgctcac tgactcgctg cgctcggtcg 3300

ttcggctgcg gcgagcggtc tcagctcact caaaggcggg aatacgggt tccacagaat 3360

caggggataa cgcaggaaag aacatgtgag caaaaggcca gaaaaggcc aggaaccgta 3420

aaaaggccgc gttgctggcg tttttccata ggctccgcc cctgacgag catcacaaaa 3480

atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc 3540

cccctggaag ctccctctg cgctctctg ttccgacct gccgcttacc ggatacctgt 3600

ccgcctttct ccttcggga agcgtggcgc tttctcatag ctcacgtgt aggtatctca 3660

gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc gttcagcccc 3720

accgctgcgc cttatccgtt aactatctc ttgagtccaa cccgtaaga cagacttat 3780

cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta 3840

cagagtcttt gaagtggtag cctaactacg gctacactag aaggacagta ttggtatct 3900

gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac 3960

aaaccaccgc tggtagcggg ggtttttttg ttgcaagca gcagattacg cgcagaaaaa 4020

aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa 4080

actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt 4140

taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact tggcttgaca 4200

gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctatct cgttcaccca 4260

tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggtta ccatctggcc 4320

ccagtgtgc aatgataccg cgagaccac gtcaccggc tccagattta tcagcaataa 4380

accagccagc cggaagggcc gagcgagaa gtggtcctgc aactttatcc gcctccatcc 4440

agtcatttaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca 4500

acgtttgtgc cattgttaca ggcatcgtgg tgcacgctc gtcgtttggt atggcttcat 4560

tcagctccgg ttccaacga tcaaggcgag ttacatgac cccatgttg tgcaaaaag 4620

cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac 4680

tcattggtat ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt 4740

ctgtgactgg tgagtactca accaagtcac tctgagaata gtgtatgcgg cgaccgagtt 4800

gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc 4860

tcattcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat 4920

ccagttcgat gtaaccact cgtgcacca actgatcttc agcatctttt actttacca 4980

gcgtttcttg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataagggcga 5040

cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg 5100

gttattgtct catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg 5160

ttccgcgcac atttccccga aaagtgccac ctgacgtcta agaaaccatt attatcatga 5220

cattaaccta taaaaatagg cgtatcacga ggccctttcg tctcgcgcgt ttcggtgatg 5280

acggtgaaaa cctctgacac atgcagctcc cggagacggt cacagcttgt ctgtaagcgg 5340

atgccgggag cagacaagcc cgtcagggcg cgtcagcggg tgttggcggg tgcggggct 5400

ggcttaacta tgcggcatca gacgagattg tactgagagt gcaccatatg cgggttgaaa 5460

taccgcacag atgcgtaagg agaaaaatacc gcatcaggcg ccattcgcca ttcagctgc 5520

gcaactgttg ggaaggcgga tcggtgcggg cctcttcgct attacgccag ctggcgaaag 5580

gggatgtgc tgcaaggcga ttaagtggg taacgccagg gttttccag tcacgacgtt 5640

gtaaaacgac ggccagtga ttc 5663

<210> 61

<211> 5395

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 61

aattcgagct cggtaccgg ggatcctcia gagtcgacct gcaggcatgc aagggaagat 60

atgcctgctt gacattatg tccgtcattt ttcggtttat cctggtaaaa aaagttttaa 120

tcctctcaag gctttcttcg tgtaaaacaa gctttttccc caaaacttcc gcaacagtct 180

cctttgtaag gtcatcctgc gtggggccga gtcctccggt cataataaca aggtcgcacc 240  
 ttccaaagc tgcaagaaga cttttttca gccgaacgga attgtccccc accacactgt 300  
 gataatacac attcacacca atgtcattga gccttttga tatatactgg gcattggtat 360  
 ttgctatctg ccccatiaaa agctcgggtc caaccgctaa tatctccgca ttcattatga 420  
 aagaccctt aaatttaaac tttttgtaac ttattatc aattagtgt ataaaataaa 480  
 agggaaaaag aattaaaatc aaagtttca agagcagccg taccacccgt aaaagtttca 540  
 gccgattcaa cttttttaca cataaaactt tcaaaaattg atgacttaca attatcaagt 600  
 aggatataat attactaatg ctaaacagtt attgataaag gaggaaggaa tatgaacaat 660  
 aacaaagtaa ttaaaaaagt aaccgggact ttctgagaag ctgaatttct tcatcgttga 720  
 aaggcacgtt caatatttcc tcaataccgt ttacaccac gattgtcgga acacttaagc 780  
 atacatcgct aagtccttac tgctcttcca aaaggcttga aacgtaagg atggagtttt 840  
 catttcttac aatggcttca acgattcttc ttacggcaag ggctacggca taataggttg 900  
 cacctttgtt ccgatgatt tcataagctg ctttttaac actttcatat attttattcc 960  
 gggaaatctg ctctcgcac tgatggcatt cgtcacagta gcgatccatg ggaattcccg 1020  
 caatatttgc aagactccag gccgcaactt cgggtgtcacc gtgttcgcca ataataaag 1080  
 catgtacatt tcgtgcatcc acttttcat gtctgcttaa aagataacgg aacctggctg 1140  
 tgtccaaaac cgttcggaa cctattactt tgtttttcgg gaatccgat agtttgtaag 1200  
 ttacataggt taaaatatcc accgatttg tgactaccag aagaatacaa tcgttggtgt 1260  
 actttacaat ttcatattatg atatttttga atacttccgt gtttctttta acaagatcta 1320  
 ttctcgttgc gccttctttt tggttggcac cggcggtaat gattactatg tcggatccgg 1380

cacagtcctt gtagtcacca cgataaatit caacgggcct tacaaaagc atgccgtgat 1440

ttaagtcctat gacttctccg tcggctttit ttgcatttat gtctatcagt acaatttcag 1500

atataagtcg gctgagcatc aatgtataag ctgtgggtga acctacaaag cctgcaccaa 1560

ctacgaaaca ctctaaaaga aataataaaa aactagata tatgaaagtt ctctttttct 1620

tttatgaaaa ggagaacttt cattattgat aaatatataa actagtatat aattttaata 1680

taaaacctat ttacataat ggaaattatc tatcggggga ggaaatatga acaattcagt 1740

ggaaatttta aataaatcg tgtcaaatat tgaaaaagtc attgttgga aaaagaaagc 1800

tatcgagttg atattaatat cacttatttg cgatggacat gttttgattg aagatgtccc 1860

cgggtgcgga aaaaccagta cttggcgtaa tcatggtcac agctgtttcc tgtgtgaaat 1920

tgttatccgc tcacaattcc acacaacata cgagccggaa gcataaagtg taaagcctgg 1980

gggtgcctaat gagtgagcta actcacatta attgcgttgc gctcactgcc cgctttccag 2040

tcgggaaacc tgcgtgcca gcccttcaaa ctcccaaag gcgagcccta gtgacattag 2100

aaaaccgact gtaaaaagta cagtcggcat tatctcatat tataaaagcc agtcattagg 2160

cctatctgac aattcctgaa tagagttcat aaacaatcct gcatgataac catcacaac 2220

agaatgatgt acctgtaaag atagcggtaa atatattgaa ttacctttat taatgaattt 2280

tcctgtctga ataatgggta gaaggtaatt actattatta ttgatattta agttaaaccc 2340

agtaaatgaa gtccatggaa taatagaaag agaaaaagca ttttcaggta taggtgtttt 2400

gggaaacaat ttccccgaac cattatattt ctctacatca gaaaggata aatcataaaa 2460

ctctttgaag tcattcttta caggagtcga aataccagag aatgttttag atacaccatc 2520

aaaaattgta taaagtggct ctaacttatt ccaataacct aactctccgt cgctattgta 2580  
 accagttcta aaagctgtat ttgagtttat cacccttgtc actaagaaaa taaatgcagg 2640  
 gtaaaattta tatccttctt gttttatggt tcggtataaa acactaatat caatttctgt 2700  
 ggttatacta aaagtcgttt gttggttcaa ataatgatta aatatctctt ttctcttcca 2760  
 attgtctaaa tcaattttat taaagttcat ttgatatgcc tcctaaattt ttatctaaag 2820  
 tgaatttagg aggcttactt gtctgctttc ttcatlagaa tcaatccttt tttaaaagtc 2880  
 aatcccgttt gttgaactac tctttaataa aataattttt ccgttcccaa ttccacattg 2940  
 caataataga aaatccatct tcatcggtt ttctgtcatt atctgtatga atcaaatcgc 3000  
 ctctctctgt gtcacaaagg ttttaatttt tatgtatttc ttttaacaaa ccaccatagg 3060  
 agattaacct tttaggtgt aaaccttctt ccaaatcaga caaacgtttc aaattctttt 3120  
 ctcatcattc ggcatataaa tccgtatcct ttacaggata ttttgagtt tcgtcaattg 3180  
 ccgattgtat atccgattta tatttatttt tcggtcgaat catttgaact ttacatttg 3240  
 gatcatagtc taatttcatt gcctttttcc aaaattgaat ccattgtttt tgattcacgt 3300  
 agttttctgt attcttaaaa taagtgggtt ccacacatac caatacatgc atgtgctgat 3360  
 tataagaatt atctttatta tttattgtca ctccgttgc acgcataaaa ccaacaagat 3420  
 ttttattaat tttttatat tgcattctt gcgaaatcc ttgagccata tctgacaaac 3480  
 tcttatttaa ttcttcgcca tcataaacat ttttaactgt taatgtgaga aacaaccaac 3540  
 gaactgttgg cttttgttta ataacttcag caacaacctt ttgtgactga atgcatgtt 3600  
 tcattgctct cctccagtgt cacattggac aaagcctgga ttacaaaaac cacactgat 3660  
 acaactttct ttgcctgtt tcacgatttt gtttatactc taatatttca gcacaatctt 3720

ttactctttc agccttttta aattcaagaa tatgcagaag ttcaaagtaa tcaacattag 3780

cgattttctt ttctctccat ggtctcacit ttccactttt tgtcttgccc actaaaaccc 3840

ttgatttttc atctgaataa atgctactat taggacacat aatattaaaa gaaaccccca 3900

tctatttagt tatttgtttg gtcacttata actttaacag atgggggttt tctgtgcaac 3960

caattttaag ggitttcaat actttaaac acatacatc caacacttca acgcaccttt 4020

cagcaactaa aataaaaatg acgttatttc tatatgtatc aagaatagaa agaactcgtt 4080

tttcgtacg ctcaaaacgc aaaaaagca ctcatcagag tgctttttct tatcgtccca 4140

aatcatgcga ttttttcctc ttgtcttttc ttgtctcag aagtctcga tcacgtgca 4200

aaacatcttg aagcgaaaaa gtattcttct tttcttcga tcgtcatgc tgacgcacga 4260

aaagccctct agcgcatag gaacaactcc taaatgcatg tgagggggtt tctcgtccat 4320

gtgaacagtc gcatacga tattttgttt ccatactgc attaatgaat cggccaacgc 4380

gcggggagag gcggtttgcg tattgggcgc tcttcgctt cctcgtcac tgactcgtg 4440

cgctcggtcg ttcggtcgcg gcgagcggta tcagctcact caaaggcggc aatacgttta 4500

tccacagaat caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaggcc 4560

aggaaccgta aaaaggccgc gttgtggcg tttttccata ggctccgcc cctgacgag 4620

catcacaaaa atcgacgtc aagtcagagg tggcgaaacc cgacaggact ataaagatac 4680

caggcgtttc cccctggaag ctccctcgtg cgtctctctg ttccgacct gccgcttacc 4740

ggatacctgt ccgcctttct ccttcggga agcgtggcg tttctcatag ctacacgtgt 4800

aggtatctca gttcgggtga ggtcgttcgc tccaagctgg gctgtgtgca cgaacccccc 4860

gttcagcccg accgctgcgc cttatccggt aactatcgtc ttgagtcgaa cccggtgaga 4920

cagcacttat cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta 4980

ggcggtgcta cagagttctt gaagtgggtg cctaactacg gctacactag aagaacagta 5040

tttggtatct gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga 5100

tccggcaaac aaaccaccgc tggtagcggg ggtttttttg ttgcaagca gcagattacg 5160

cgcagaaaaa aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag 5220

tggaacgaaa actcacgtta agggattttg gtcattgagat tatcaaaaag gatcttcacc 5280

tagatccttt taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact 5340

tggcttgaca gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtc 5395

<210> 62  
 <211> 45  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: Synthetic  
 primer

<400> 62  
 atttacctgg ctgggaatac tgagacatat gtcattgagg ccgta 45

<210> 63  
 <211> 53  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: Synthetic  
 primer

<400> 63  
 aaaaaagctt ataattatcc ttaatttcct actacgtgcg cccagatagg gtg 53



<210> 64  
 <211> 60  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: Synthetic  
 primer

<400> 64  
 cagattgtac aaatgtggtg ataacagata agtctactac tgtaacttac ctttctttgt 60

<210> 65  
 <211> 49  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: Synthetic  
 primer

<400> 65  
 tgaacgcaag ttictaatTT cggttgaaat ccgatatagagg aaagtgtct 49

<210> 66  
 <211> 45  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: Synthetic  
 primer

<400> 66  
 ttaaattgttg ataaggaagc tcttttcaat gaagttaagg tagca 45

<210> 67  
 <211> 53  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: Synthetic  
 primer

<400> 67  
aaaaaagctt ataattatcc ttagctctct tcaatgtcg cccagatagg gtg 53

<210> 68  
<211> 60  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
primer

<400> 68  
cagattgtac aaatgtggtg ataacagata agtcttcaat gataacttac ctttctttgt 60

<210> 69  
<211> 49  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
primer

<400> 69  
tgaacgaag ttctaatctt cgattagagc tcgatagagg aaagtgtct 49

<210> 70  
<211> 1509  
<212> DNA  
<213> Caldiscellulosiruptor kristjanssonii

<400> 70  
atgtatitta ttggaattga cggttgaaca tctggaacaa agacaatcct gattgactca 60

aaaggtgaaga ttctggcttc tgcaaccttt gaatatcctc tttatcagcc tcagattggc 120

tgggctgagc aaaatccga agactggtgg gatgcaagcg taaaaggaat taaagctgtg 180

cttgaaaagt caaaagtaga cccaaggaa gttaaggctg tgggacttac cgggcagatg 240

cacgggcttg tgctgcttga caaaaactac aacgttataa gaccatcaat catctggtgt 300

gaccagagaa cggcaaaaga atgtgatgaa ataacagaaa aggttggcaa ggaaaagctt 360

gtggagatta cagcaaaccc tgcactgaca ggttttacag cgtccaagat tctgtgggtg 420

agaaacaacg agccccaaaa ctatgagaag gtctacaaaa ttttgcttcc caaagactat 480

ataaggttta aacttacagg cgagtttgca acagatgtgt cggacgcctc gggtatgcag 540

cttttggaca ttaaaaacag gtgctgggtct gatgaggtac ttgaaaagct tgagatagac 600

aaagggttc ttggaaaagt ctatgagtcg ccagaggtaa cgggaaaagt tagcgggcaa 660

gcaagcgaac ttacaggctt ttgtgaaggt acgcttgttg ttgcaggtgg aggagaccag 720

gcagcaggtg cagttggaat tggcatagta aagacgggtg tgatttcac tacaattggt 780

tcgtctggcg ttgtttttgc ccatcttgac gagtttaaga ttgaccaca gggaagggtt 840

cacacatitt gtcatgcagt gccgggaaaa tggcatgtga tgggtgtaac acaagtgcc 900

ggactttctc tcaagtgggt tagagacaac ttgacacaca tcgaaaaggc tgcgtttgag 960

tttattgaca aagaccata cattttgatg gaccaggagg cagaacttgc aaaccaggc 1020

gcagacggac ttgttttctt gccatatttg atgggggaaa gaacgcccatt ttggaccca 1080

tacgccaag gaatattctt tggataaaca gcaaagcata cacgaagaga gttcattaga 1140

gctgtcatgg aaggtgttgt attttcactt aaaaactgtc ttgatatttt gtatgagatg 1200

ggcatcgagg tgaaggaggt aagagtttca ggcggtggtg caaagagcaa gctctggaga 1260

cagatgcagg cagacatatt tgagatggat gtatggacac tgaattccaa agaaggacct 1320

gcgtttggtg cagctatcct ggcagcagtt ggtgcaggag aatatcagaa ggttgaagaa 1380

gccttgata ctatgattca aaaggtagat aactgcagcc caaatgaaaa actatttgaa 1440

atatatagaa aaacttataa actttacaac agtatatc caagagttaa ggacttattc 1500

aacatgtaa 1509

<210> 71

<211> 1317

<212> DNA

<213> Caldiscellulosiruptor kristjanssonii

<400> 71

atgaaatact tcaaagacat tccagaagta aaatatgaag gaccacagtc agacaatcca 60

tttgctttca agtattacaa tcttgacgag gttattgatg gcaagccttt aaaagaccac 120

cttcgttttg caattgctta ctggcacacg ttcgtgcaa ccggtacgca cccttttgga 180

caacctacaa ttaatcgtec atgggacagg ttctcaaacc caatggacaa tgcaaaagca 240

agagtgaag ctgcatttga attttttgaa aagctaaatg ttccattttt ctgcttcac 300

gacagagaca tcgcacctga aggagaaaaa ttaagagaaa caaacaagaa ttgggatgag 360

atagtctcta tgataaaaga atatttaaag acaagcaaaa caagagtttt gtggggaaca 420

gcaaacctat tttcacatcc gcgatatgtt catggtgctg caacatcctg caatgccgat 480

gtttttgcgt atgcagcagc gcaggtgaaa aaggcgtag aggttacaaa agagcttggc 540

ggcgaaaact atgtgttctg gggcggaaga gaaggttatg agacactttt gaacaccgac 600

atggagcttg agcttgacaa cttggcaaga tttttgcaca tggcagttga ctatgcaaaa 660

gagatagggg ttgacggta gtttttgatt gaaccaaagc caaaagaacc aactaagcat 720

cagtacgatt ttgatgccgc tcatgtttat ggatttttga aaaaatatga ccttgacaag 780

tacttcaagc tcaacataga ggtaaaccat gcgactttgg caggacatga tttccacat 840

gagttgagat ttgcacgaat aaacaacatg cttggctcaa ttgatgctaa catgggcgac 900

ttgcttttgg gctgggatac agaccagttc ccaacagatg taagacttac cacgcttgct 960

atgtatgagg ttattaaagc tggtaggcttt gacaaaggcg ggctcaactt tgacgcaaag 1020

gtaagaagag gticttttga gcttgaagac ttggtcattg gtcacattgc agggatggac 1080

gcttttgcaa aaggatttaa gatagcatat aagcttgtca aagacggtgt atttgataag 1140

tttattgaag aaaggtatag aagctacaaa gaaggaattg gagctaagat tgtaagtgg 1200

caggcagatt ttaagacgtt agaagaatat gctttgaatc tttcaaagat agaaaacaaa 1260

tctggcaagc aagagcttct tgagatgatt ttgaacaaat atatgttcag tgaataa 1317

<210> 72

<211> 1536

<212> DNA

<213> Clostridium straminisolvens

<400> 72

ttgtcatatt tactgggagt agatataggt acatcaggca cgaaaactgt tttatatgat 60

gaactgggaa ataccgtagc aagcagcctt gaggaatatc cattgtacca gcccacatatt 120

gggtgggcag agcaggaacc ggaagactgg tggagggcaa catgcctatc tatcaaacat 180

gttatttcca aaagaggaat tgatgcttcc tctattaagg gaatcggact ttcaggacag 240

atgcacgggg ctgttcttct ggacaaagac ggcaaagtgc taagaaaagc aattatatgg 300

tgtgaccaga gaagttttgc cgagtgcgag cagattactt caattatagg gaaggaaagg 360

ctcgttgaga taactgcca cctgcactg acgggattta cagcatcaaa ggttatgtgg 420

gttaaaaata atgaacctga aatttttgag aagatttata agatacttct ccctaaagac 480

tatataagat ataaattaac gggagaattt gctacagagg tatctgatgc cagtggaatg 540

cagtttatgg atataccggg gagaaaatgg agcgacgaag tcataagtaa actcggactt 600

gataaaagca tgctgggaga actctatgag tctcaggaag ttagcgggaa agtgaataag 660

tatgctgctt cattaaccgg acttaaggaa ggaactcctg tcgtgggtgg agcaggagac 720

caggcagcag gagctgtcgg taatggaatt gtgagaccg gggtgggttc atccactata 780

ggaacttcag gagtagtatt tgcattctct gaaaaggtta ctattgatcc aaagggtaga 840

gttcatactt tttgtcatgc ggtaccaaatt acctggcaca ttatgggggt tacacaaggg 900

gccgggctgt ctcttaagtg gttccgtgac aatttctgta tagaagaaaa gagaactgca 960

gagctaataa aaatagaccc gtacataatt atggataaag aagctgaaaa agtggctccg 1020

ggctgtaacg gtttaactta tttaccttat ctgatgggag aaagaacgcc acatcttgac 1080

cctaatagcca aggggtgtctt tttcggatta acagcaaagc atgaaaaaca ggatatctta 1140

aggtcgatta tggaaggtgt tgtatatagc cttagagatt gccttgaaat tattgaggaa 1200

atgggtgtta acgtttctga agtaagagct tccggtggag gcggtaaaag tgaattgtgg 1260

agaaaaatgc aggcggatat attcggcact gatattacaa ccgtaaagtc aagtgaggga 1320

ccggcacttg gggtagcact tcttgccgga gtaggaacgg gtgtgtacaa caacattaat 1380

gaagcatgtg aagcagtaat aaaagaaaat acccggcagg cttcggaccc ggagctatat 1440

gtaaaaatac cgaagtttta tgatatattat aaacgtctgt ataactcttt gaaaaaggaa 1500

tttgcagacc tttcggctat gctgcaaagt ttatag 1536

<210> 73

<211> 1320

<212> DNA

<213> Clostridium straminisolvens

<400> 73

|                                                                    |      |
|--------------------------------------------------------------------|------|
| atggcagagt attttaaaaa tgtaccgaaa atcaaatatg aaggaaagga ttcggacaat  | 60   |
| ccttttagcgt ttaagtacta taatcccgat gaggtcattg gcggtaaaac aatgaaagag | 120  |
| catctaaggt ttgctgttgc atattggcat acatatcagg gaacgggtgc agaccattt   | 180  |
| gggccgggta ctgctgtaag accgtgggat gacatatcgg acccaatgga tcttgcaaag  | 240  |
| gccaaagtgg ccgcaaattt cgagctgtgt gaaaaattgg gagtaccatt ttctctcttc  | 300  |
| catgacagag atattgcgcc tgaagcttca actttaagag agaccaataa aagacttgat  | 360  |
| gagattgttg cactgataaa ggactatgat aaaacaagta gtgtaaaact actctggggt  | 420  |
| acaacaaatg cttttagcca cccaagggtt gtccatgggtg catctacttc tccgaatgca | 480  |
| gatgtatttg catatgcagc agctcagggt aaaaaggcta tggaaattac cctggaactt  | 540  |
| ggcggtcaga actatgtgtt ctggggtgga agagaaggct atgaaacctt acttaatact  | 600  |
| gatatgaaat tggagcttga caatatggca aggttccttg gaatggcagt tgactatgca  | 660  |
| aaagagattg gttttaaaagg gcagctcttg attgaaccta agccaaaaga accgacaaag | 720  |
| caccagtatg actttgatac agctacagtt atcggtttct taaggactta tggctcttgag | 780  |
| aattacttca aaatgaatat tgaagcaaat cacgtacac ttgcagctca tactttccag   | 840  |
| catgaactta gggtttcaag aattaacggt gtgctaggaa gtatcgatgc aaaccagggt  | 900  |
| gatcttcttt taggatggga cactgaccaa ttcccgacaa atatctacga tactaccctt  | 960  |
| gctatgtatg aagtaattaa ggcaggcgga ttacaacgg gaggtctgaa ttttgattct   | 1020 |
| aaagtcagaa gaggatcatt tgagcctgtg gacctgttct atgcacatat tgcaggtatg  | 1080 |
| gacgcttttg caaaaggatt taaaatagca tataaatgg tttccgacgg taagtttgac   | 1140 |
| aaatttattg atgaaagata tgaaagctat aagagcggta ttggaaaaga tattgtagat  | 1200 |

ggaaaaagtag ggtttaaaga gcttgaaaaa tatgcttttag agcttgatgg tatcaagaat 1260

gtgtcgggaa gacaggaagt tctcgaagct atgttaaaca aatatattct tgaggactag 1320

<210> 74

<211> 8909

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 74

catgagacaa taaccctgat aaatgcttca ataatatga aaaaggaaga gtatccagta 60

ttctgacatg ggtgtatcaa taaccatgc gtttccgtat tgtatcgga tggtttcgga 120

cagggcgggtg ggaatagaca tggaaaagat tttttgccc gaggatgcat tgataaagta 180

tttcttttcc gaaagagagg aaaagattct aaagagtttt ggaaatactg atgaatattg 240

tgtgcagagt acaattctat ggacaagaaa agaggctttg tcaaaacttt ttcgtctggg 300

aatgaggatg gattttaaaa agctggatac ttggaggac gaggtggttt ttcaggaaac 360

aaacagggcg cgtctgtttt cttttatatg caataattac tgtatctctc tggcattgcc 420

aggttttaat aaagattaaa attattgact agaaataaaa aaattgtcca taatattaat 480

ggacaaaaaa acaaagaatt acatcaaagg aagataaaaa tactttgtta aaaaattaat 540

tattttttat ctaaactatt gaaaatgaaa ataaaataat ataaaatgaa tcatagtgca 600

agagatactt gccagaggat gaatatatta ctgcattcat gctttatggc agctaataga 660

ggcattaaat taaattttaa tttaaatag gaggcgatat taatgaataa atattttgag 720



aacgtatcta aaataaaata tgaaggacca aaatcaaaca atccttattc ttttaaattt 780

tacaatccag aagaagtaat cgatggcaag acgatggagg agcatctacg cttttctata 840

gcttactggc acacttttac tgctgatgga acagatcaat ttggcaaagc taccatgcaa 900

agaccatgga accactacac agatcctatg gacatagcaa aggcaagggt agaagcagca 960

tttgagtttt ttgataagat aaatgcacct ttcttctgct tccatgacag ggatatgca 1020

cctgaaggag acactcttag agagacaaac aaaaacttag atacaatagt tgccatgata 1080

aaggattact tgaagaccag caagacgaaa gttttgtggg gcaccgcaa tcttttctcc 1140

aatccgagat ttgtacatgg tgcataaca tcttgcaatg ctgatgtttt cgcataattct 1200

gcagctcaag ttaaaaaagc tcttgagatt actaaggagc ttggcggcga aaactacgta 1260

ttctggggtg gcagagaagg atatgaaaca cttctcaata cagacatgga gtttgagctt 1320

gacaactttg caagattttt gcacatggct gttgactacg cgaaggaaat cggctttgaa 1380

ggccagtctt tgattgagcc gaagccaaag gagcctacga aacaccaata cgactttgac 1440

gtggcaaattg tattggcatt cttgagaaaa tacggccttg acaaatattt caaagtgaat 1500

atcgaggcaa accatgcgac attggcattc cagcacttcc aacatgagct aagatacgcc 1560

agaataaacg gtgtattagg atcaattgac gcaaatacag gcgatatgct tttaggatgg 1620

gatacagacc agttccctac agatatacgc atgacaacgc ttgctatgta tgaagtcata 1680

aagatgggtg gatttgacaa aggcggcctt aacttcgatg caaaagtaag acgtgcttca 1740

tttgaaccag aagatctttt cttaggtcat atagccggaa tggatgcctt tgcaaaaggc 1800

ttcaaagttg cttacaaact tgtgaaagat ggcgatatgt acaagtcat cgaggaaaga 1860

tagcaagct acaaagacgg cattggcgct gacattgtaa gcgggaaagc tgacttcaag 1920

agccttgaaa agtacgcatt agagcacagc cagattgtca acaaatcagg caggcaagag 1980  
 ctgttagaat caatcctaaa tcagtatttg ttgcagaat aatgaaacat gagggcggct 2040  
 tcatgcttca ttaaagctgc cctcaacaaa aatcatggag gtaaatgtat gtatttttta 2100  
 gggatagatt tagggacatc atcagttaag ataatactga tgaatgaaag cggcaatgtg 2160  
 glatcaagcg ttcaaaaaga atatcctgtg tactatccag agccaggctg ggctgagcaa 2220  
 aatccagaag attggtggaa tggcacaagg gatggaataa gagagattat tgcgaaaagc 2280  
 ggcgtaaatg gcgatgaaat aaagggtgtt ggcttaagcg ggcagatgca tggactggtg 2340  
 cttttagaca aagacaataa cgttttaacg ccagccatac tttggtgtga ccagaggaca 2400  
 caggaagaat gcgactacat cacagagaaa ataggaaaag aaggcctttt gaagtacaca 2460  
 gggaataaag cattgacagg ttttactgca ccaaagatat tatgggtaaa gaagcacctt 2520  
 aaagacgtat atgaaagaat cgctcatac cttttgcaa aagattatat aaggtttaaa 2580  
 ttgacagggtg agtacgctac agaagtttca gatgcatcag gtacacttct tttcgatgtg 2640  
 gaaaatagaa gatggtcaaa ggaaatgata gacatatattg aaataccgga aaaagccctt 2700  
 cctaagtgtc acgaatcaac agatgtcaca gggatatgtca ccaaagaggc agcagatttg 2760  
 acagggttc atgaaggac tattgtcgta ggcggtggtg gtgaccaagc cagcggcgct 2820  
 gtaggcactg gcacggtgaa aagcggcata gtgtccatcg cattaggaac ttcaggcgctc 2880  
 gtatttgcac cacaggacaa gtacgcagca gatgatgagc ttaggcttca ctcattctgc 2940  
 catgcaaacg gcaaatggca tgtgatgggt gtcatgcttt cggctgcac atgtcttaaa 3000  
 tgggtgggtag atgatgtaaa taattacaag accgatgtta tgacatttga tggactctta 3060

gaagaagcag agaaggtgaa gccaggcagt gatggattga tattcttgcc atacctgatg 3120

ggtgaaagga ccccttacag cgatccttat gcgagaggca gctttgtagg tttacaatt 3180

acacacaata gaagccacat gacaagatct atattagaag gcgtcgatt tggacttagg 3240

gattcgctgg agcttataaa ggctttaaat atacctgtaa atgaagccag ggtaagtgg 3300

ggtggtgcta aaagcaggct ttggaggcaa atacttgccg atgtattcaa tgtaaggata 3360

gacatgataa atgctacaga aggaccttca ttggtgcag caataatggc gtctgtggga 3420

tatggccttt acaaaaatgt agatgatgca tgcaatagtt taataaaagt tacagacagc 3480

gtatatccaa tcaaagaaaa cgtcgaaaag tacaacaaac tgtatccaat ctacgtgagc 3540

ttgtattcaa ggcttaaaagg cgctttgaa gaaattggga agttggattt gtaaaataaa 3600

ttcatttga aataaattta tgacagtaca agggacattg attaacaaag cticaggta 3660

ataatagtaa agttaatatt tgctatgaaa tgaagcata ataactgtt ccttgtactt 3720

tgctttatca tgttattta agatactaata taataaaagt caatttagcc aataataaaa 3780

tcctatatat agtaaatatt tacaataaaa tcactacaaa ataaaaaact ttatttaatc 3840

tcttaaaaat atctacataa gggggtgtta gatgaaaag gccgtaatca tggatcatagc 3900

tgtttctgt gtgaaattgt tatccgtca caattccaca caacatacga gccggaagca 3960

taaagtgtaa agcctggggt gcctaagtag tgagctaact cacattaatt gcgttgcgt 4020

cactgcccgc ttccagtcg ggaaacctgt cgtgccagcc ctcaaaactt cccaaaggcg 4080

agccctagt acattagaaa accgactgta aaaagtacag tcggcattat ctcatattat 4140

aaaagccagt cattaggcct atctgacaat tcctgaatag agttcataaa caatcctgca 4200

tgataacat cacaacaga atgatgtacc tgtaagata gcggtaaata tattgaatta 4260

cctttattaa tgaattttcc tgctgtaata atgggtagaa ggtaattact attattattg 4320  
 atatttaagt taaacccagt aaatgaagtc catggaataa tagaaagaga aaaagcattt 4380  
 tcaggtatag gtgttttggg aaacaatttc ccggaacat tatattttctc tacatcagaa 4440  
 aggtataaat cataaaactc tttgaagtca ttttttacag gagtccaaat accagagaat 4500  
 gttttagata caccatcaaa aattgtataa agtggctcta acttatccca ataacctaac 4560  
 tctccgtcgc tattgtaacc agttctaaaa gctgtatttg agtttatcac ccttgtcact 4620  
 aagaaaataa atgcagggtg aaatttatat ccttcttggt ttatgtttcg gtataaaaca 4680  
 ctaatatcaa tttctgtggt tatactaaaa gtcgtttggt ggttcaaata atgattaaat 4740  
 atctcttttc tcttccaatt gtctaaatca atttatttaa agttcatttg atatgcctcc 4800  
 taaattttta tctaaagtga atttaggagg cttacttgtc tgctttcttc attagaatca 4860  
 atcctttttt aaaagtcaat cccgtttggt gaactactct ttaataaaat aatttttccg 4920  
 ttcccaattc cacattgcaa taatagaaaa tccatcttca tcggcttttt cgtcatcatc 4980  
 tgtatgaatc aaatcgcttt cttctgtgtc atcaaggttt aattttttat gtatttcttt 5040  
 taacaaacca ccataggaga ttaacctttt acggtgtaaa ccttcctcca aatcagacaa 5100  
 acgtttcaaa ttcttttctt catcatcggt cataaaatcc gtatccttta caggatattt 5160  
 tgcagtttcg tcaattgccg attgtatatc cgatttatat ttatttttcg gtcgaatcat 5220  
 ttgaactttt acatttggat catagtctaa tttcattgcc tttttccaaa attgaatcca 5280  
 ttgtttttga ttcacgtagt tttctgtatt cttaaaataa gttggttcca cacataccaa 5340  
 tacatgatg tgctgattat aagaattatc tttattattt attgtcactt ccgttgcacg 5400

cataaaacca acaagatttt tattaatttt tttatatgc atcattcggc gaaatccttg 5460

agccatatct gacaaactct tatttaattc ttcgcatca taaacatttt taactgttaa 5520

tgtgagaaac aaccaacgaa ctgttggcctt ttgtttaata acttcagcaa caaccttttg 5580

tgactgaatg ccatgtttca ttgctctcct ccagttgcac attggacaaa gcctggattt 5640

acaaaaccac actcgataca actttctttc gcctgtttca cgattttgtt tatactctaa 5700

tatttcagca caatctttta ctctttcagc ctttttaaata tcaagaatat gcagaagttc 5760

aaagtaatca acattagcga ttttcttttc tctccatggc ctcacttttc cactttttgt 5820

cttgtccact aaaacccttg atttttcatc tgaataaatg ctactattag gacacataat 5880

attaaaagaa acccccatct atttagttat ttgtttggc acttataact ttaacagatg 5940

gggtttttct gtgcaaccaa ttttaagggt ttcaataact ttaaaacaca tacataccaa 6000

cattcaacg cacttttcag caactaaaat aaaaatgacg ttatttctat atgtatcaag 6060

aatagaaaga actcgttttt cgctacgctc aaaacgcaa aaaagcactc attcgagtgc 6120

ttttcttat cgctccaaat catgcgattt tttctctttt gcttttcttt gctcacgaag 6180

ttctcgaatc cgctgcaaaa catcttgaag cgaaaaagta ttcttctttt ctccgatcg 6240

ctcatgctga cgcacgaaaa gccctctagg cgcataggaa caactcctaa atgcatgtga 6300

ggggttttct cgtccatgtg aacagtcgca tacgcaatat ttgtttccc atactgcatt 6360

aatgaatcgg ccaacgcgcg gggagaggcg gtttgcgtat tgggcgctct tccgcttctt 6420

cgctcactga ctgctgcgc tcggtcgttc ggctgcggcg agcggtatca gctcactcaa 6480

aggcggtaat acggttatcc acagaatcag gggataacgc aggaagaac atgtgagcaa 6540

aaggccagca aaaggccagg aaccgtaaaa aggcgcggtt gctggcggtt ttccataggc 6600

tccgcccc tgacgagcat cacaaaaatc gacgtcaag tcagaggtgg cgaaacccga 6660

caggactata aagataccag gcgtttcccc ctggaagctc cctcgtgcgc tctcctgttc 6720

cgaccctgcc gcttaccgga tacctgtccg cctttctccc ttcgggaagc gtggcgcttt 6780

ctcatagctc acgctgtagg tatctcagtt cgggtgtagg cggtcgtcc aagctgggct 6840

gltgtcacga accccccgtt cagcccgacc gctgcgcctt atccggtaac tatcgtcttg 6900

agtccaacc ggtaagacac gacttatcgc cactggcagc agccactggt aacaggatta 6960

gcagagcgag gtatgtaggc ggtgtctacag agttcttgaa gtggtggcct aactacggct 7020

acactagaag aacagtattt ggtatctgcg cctcgtctgaa gccagttacc ttcggaaaaa 7080

gagttgtag ctcttgatcc ggcaaaaaa ccaccgtgg tagcgggtgt tttttgttt 7140

gcaagcagca gattacgcgc agaaaaaaag gatctcaaga agatcctttg atcttttcta 7200

cggggatcgc ttgcctgtaa cttacacgcg cctcgtatct tttaatgatg gaataatttg 7260

ggaatttact ctgtgtttat ttatttttat gttttgtatt tggatttttag aaagtaaata 7320

aagaaggtag aagagttacg gaatgaagaa aaaaaataa acaaaggttt aaaaaatttc 7380

aacaaaaagc gtactttaca tataatttta ttagacaaga aaagcagatt aaatagatat 7440

acattcgatt aacgataagt aaaatgtaa atcacaggat tttcgtgtgt ggtcttctac 7500

acagacaaga tgaacaatt cggcattaat acctgagagc aggaagagca agataaaagg 7560

tagtatttgt tggcgatccc cctagagtct ttacatctt cggaaaacaa aaactatttt 7620

ttctttaatt tcttttttta ctttctatct ttaatttata tatttatatt aaaaaattta 7680

aattataatt atttttatag cacgtgatga aaaggacca tcgataagct agcttttcaa 7740

ttcaattcat catttttttt ttattctttt ttgtgatttc ggtttctttg aaattttttt 7800

gattcggtaa tctccgaaca gaaggaagaa cgaaggaagg agcacagact tagattggta 7860

tatatagca taigtatgt tgaagaaaca tgaattgcc cagtattctt aacccaactg 7920

cacagaacaa aaacctgcag gaaacgaaga taaatcatgt cgaaagctac atataaggaa 7980

cgtgctgcta ctcatcctag tctgtttgct gccaaagctat ttaatatcat gcacgaaaag 8040

caaacaaact tgtgtgcttc attggatgtt cgtaccacca aggaattact ggagttagtt 8100

gaagcattag gtcccaaat ttgtttacta aaaacacatg tggatatctt gactgatttt 8160

tccatggagg gcacagttaa gccgctaaag gcattatccg ccaagtacaa ttttttactc 8220

ttcgaagaca gaaaatttgc tgacattggt aatacagtca aattgcagta ctctgcgggt 8280

gtatacagaa tagcagaatg ggcagacatt acgaatgcac acggtgtggt gggcccaggt 8340

attgttagcg gtttgaagca ggcggcagaa gaagtaacaa aggaacctag aggccttttg 8400

atgttagcag aattgtcatg caagggtcc ctatctactg gagaatatac taagggtact 8460

gttgacattg cgaagagcga caaagatttt gttatcggct ttattgctca aagagacatg 8520

ggtggaagag atgaaggta cgattggttg attatgacac ccggtgtggg ttagatgac 8580

aaggagacg cattgggtca acagtataga accgtggatg atgtggtctc tacaggatct 8640

gacattatta ttgttgaag aggactattt gcaaaggga gggatgctaa ggtagagggt 8700

gaacgttaca gaaaagcagg ctgggaagca tatttgagaa gatgcggcca gcaaaactaa 8760

aaaactgtat tataagtaaa tgcattgata ctaaaactcac aaattagagc ttcaatttaa 8820

ttatatcagt tattaccac ttttcgggga aatgtgcgcg gaaccctat ttgtttattt 8880

ttctaaatac attcaaatat gtatccgt 8909

<210> 75

<211> 6972

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 75

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| gacgctcagt ggaacgaaaa ctcacgttaa gggattttgg tcatgagatt atcaaaaagg  | 60  |
|                                                                    |     |
| atcttcacct agatcctttt aaattaaaaa tgaagtttta aatcaatcta aagtatatat  | 120 |
|                                                                    |     |
| gagtaaaactt ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc | 180 |
|                                                                    |     |
| tgtcaattcg agctcggtac ccggggatcc ttaagaagac taataaaaag ttctctaaaag | 240 |
|                                                                    |     |
| catgaaatat cctgtatttt ggagttactt gccgttattt atggctaaag ccgaaaaaaa  | 300 |
|                                                                    |     |
| gataaggagg gtgttgtata caaaaaacat ttgttatata attagagttg ctggaaccc   | 360 |
|                                                                    |     |
| agtaaaatag ctgtttatag tgagiaggia ttattacctg atgagttata ctgtcggtag  | 420 |
|                                                                    |     |
| ctatttagcg gagcggcttg tccagattgg tctcaagcat cacttcgcag tcgcgggcga  | 480 |
|                                                                    |     |
| ctacaacctc gtccttcttg acaacctgct ttgaacaaa aacatggagc aggtttattg   | 540 |
|                                                                    |     |
| ctgtaacgaa ctgaactgcg gtttcagtcg agaaggttat gtcgtgcc aaggcgcagc    | 600 |
|                                                                    |     |
| agcagccgtc gttacctaca gcgtcggtag gctttccgca ttgatgcta tcggtggcgc   | 660 |
|                                                                    |     |
| ctatgcagaa aaccttccgg ttatcctgat ctccggtgct ccgaacaaca atgatcacgc  | 720 |
|                                                                    |     |
| tgtgtgtcac gtgttgcatc acgctcttgg caaaaccgac taccactatc agttggaaat  | 780 |
|                                                                    |     |
| ggccaagaac atcacggccg ccgctgaagc gatttacacc ccggaagaag ctccggctaa  | 840 |



aatcgatcac gtgattaaaa ctgctcttcg tgagaagaag ccggtttatc tcgaaatcgc 900

ttgcaacatt gcttccatgc cctgcgcgcg tcttgaccg gcaagcgcat tgttcaatga 960

cgaagccagc gacgaagctt ctttgaatgc agcggttgaa gaaaccctga aattcatcgc 1020

caaccgcgac aaagttagcg tctctgctcg cagcaagctg cgcgcagctg gtgctgaaga 1080

agctgctgtc aaatttgctg atgctctcgg tggcgcagtt gctaccatgg ctgctgcaaa 1140

aagcttcttc ccagaagaaa acccgcatc catcggcacc tcatggggtg aagtcagcta 1200

tccgggcgtt gaaaagacga tgaaagaagc cgaatgcggtt atcgtctcgg ctctgtctt 1260

caacgactac tccaccactg gttggacgga tattcctgat cctaagaaac tggttctcgc 1320

tgaaccgcgt tctgtcgtcg ttaacggcat tcgcttcccc agcgtccatc tgaaagacta 1380

tctgaccctg ttggctcaga aagtttccaa gaaaaccggt gcattggact tcttcaaatc 1440

cctcaatgca ggtgaactga agaaagccgc tccggctgat ccgagtgtc cgttggtcaa 1500

cgcagaaatc gcccgctcagg tcgaagctct tctgaccccg aacacgacgg ttattgctga 1560

aaccggtgac tcttggttca atgctcagcg catgaagctc ccgaacggtg ctgcggttga 1620

atatgaaatg cagtggggtc acattggttg gtccgttctt gccgccttcg gttatgccgt 1680

cgggtgtccg gaacgtcga acatcctcat ggttggtgat ggttccttcc agctgacggc 1740

tcaggaagtc gctcagatgg ttgcctgaa actgccggtt atcatcttct tgatcaataa 1800

ctatggttac accatcgaag ttatgatcca tgatggtccg tacaacaaca tcaagaactg 1860

ggattatgcc ggtctgatgg aagtgttcaa cggtaacggt gggtatgaca gcggtgctgg 1920

taaaggcctg aaggctaaaa ccggtggcga actggcagaa gctatcaagg ttgctctggc 1980

aaacaccgac ggcccaacc tgatcgaatg cttcatcggt cgtgaagact gactgaaga 2040

atgggtcaaa tggggtaaagc gcgttgctgc cgccaacagc cgtaagcctg ttaacaagct 2100

cctctagatt ctgttaaaac cggacattga agaagggtgtt gcgcagcgtt taataaaaac 2160

atctgtttat cgaagtttag gaataggaaa attaaaaaaa acaagacggg agtgagtttt 2220

tgaaatggct tcttcaactt tttatattcc tttcgtcaac gaaatgggcg aaggttcgct 2280

tgaaaaagca atcaaggatc ttaacggcag cggctttaaa aatgcgctga tcgtttctga 2340

tgctttcatg aacaaatccg gtgttgtgaa gcaggttgct gacctgttga aagcacaggg 2400

tattaattct gctgtttatg atggcgttat gccgaaccgc actgttaccg cagttctgga 2460

aggccttaag atcctgaagg ataacaattc agacttcgtc atctccctcg gtggtggttc 2520

tccccatgac tgcgccaaag ccacgcctct ggtcgcaacc aatggtggtg aagtcaaaga 2580

ctacgaaggt atcgacaaat ctaagaaacc tgcctgcct ttgatgtcaa tcaacacgac 2640

ggctggtacg gcttctgaaa tgacgcgttt ctgcatcatc actgatgaag tccgtcacgt 2700

taagatggcc atgttgacc gtcacgttac cccgatgggt tccgtcaacg atcctctgtt 2760

gatggttggt atgccccaaag gcctgaccgc cgccaccggt atggatgctc tgaccacgc 2820

attgaagct taitcttcaa cggcagctac tccgatcacc gatgcttgcg ctttgaaagc 2880

agcttccatg atcgctaaga atctgaagac cgcttgcgac aacggttaagg atatgccggc 2940

tcgtgaagct atggcttatg cccaattcct cgctggtatg gccttcaaca acgcttcgct 3000

tggttatgtc catgctatgg ctcaccagtt gggcggttac tacaacctgc cgcattggtg 3060

ctgcaacgct gttctgcttc cgcattgtct ggcttataac gcctctgtcg ttgctggtcg 3120

tctgaaagac gttggtgttg ctatgggtct cgatatcgcc aatctcggtg ataaagaagg 3180

cgcagaagcc accattcagg ctgttcgca tctggctgct tccattggta ttccagcaaa 3240  
 cctgaccgag ctgggtgcta agaaagaaga tgtgccgctt cttgctgacc acgctctgaa 3300  
 agatgcttgt gcctgacca acccgctca gggatgacag aaagaagttg aagaactctt 3360  
 cctgagcgtt ttctaaaaga tgcgttataa ttttacaagc ctgttttttt aggaaacggg 3420  
 cttataaaat ttttttattt ttttgccggg ttttttcttg tattaatatg tggaatatgt 3480  
 taataatatt aagaagaaat tccgaattta actaaacaaa attatttttg ttatttaagc 3540  
 caatctgtca tataattctt gacatgaggg ttattagtta gtataatagt ccttgctcgtt 3600  
 ttttaagaggg atcctctaga gtcgacctgc aggcatgcaa gcttggcgta atcatggta 3660  
 tagctgtttc ctgtgtgaaa ttgttatccg ctccacaattc cacacaacat acgagccgga 3720  
 agcataaagt gtaaagcctg gggtcctaa tgagttagct aactcacatt aattgcgttg 3780  
 cgctcactgc ccgttttcca gtcgggaaac ctgtcgtgcc agcccttcaa actteccaaa 3840  
 ggcgagccct agtgacatta gaaaaccgac tgtaaaaagt acagtcggca ttatctcata 3900  
 ttataaaagc cagtcattag gcctatctga caattcctga atagagttca taaacaatcc 3960  
 tgcataataa ccatcacaaa cagaatgatg tacctgtaaa gatagcggta aatatattga 4020  
 attaccttta ttaatgaatt ttctgctgt aataatgggt agaaggtaat tactattatt 4080  
 attgatattt aagttaaacc cagtaaatga agtccatgga ataatagaaa gagaaaaagc 4140  
 attttcaggt ataggtgttt tgggaacaaa tttccccgaa ccattatatt tctctacatc 4200  
 agaaaggtat aaatcataaa actctttgaa gtcattcttt acaggagtcc aaataccaga 4260  
 gaatgtttta gatacccat caaaaattgt ataaagtggc tctaacttat cccaataacc 4320  
 taactctccg tcgtatttgt aaccagttct aaaagctgta tttgagttta tcaccttgt 4380

cactaagaaa ataatgcag ggtaaaattt atatccttct tgttttatgt ttcggtataa 4440

aacactaata tcaatttctg tggttatact aaaagtcgtt tgttggttca aataatgatt 4500

aaatatctct ttctcttcc aattgtctaa atcaatttta ttaaagttca ttgatatgc 4560

ctcctaaatt ttatctaaa gtgaatttag gaggccttact tgtctgcttt ctccattaga 4620

atcaatcctt ttttaaaagt caatcccggt tgttgaacta ctctttaata aaataatttt 4680

tccgttccca attccacatt gcaataatag aaaatccatc ttcacggct ttttcgtcat 4740

catctgtatg aatcaaatcg ccttcttctg tgcacatcaag gtttaatttt ttatgtattt 4800

cttttaacaa accaccatag gagattaacc ttttacggtg taaaccttcc tccaaatcag 4860

acaaacgttt caaattcttt tcttcatcat cggtcataaa atccgtatcc ttacaggat 4920

attttgcatg ttcgtcaatt gccgattgta tatccgattt atatttattt ttcggtcgaa 4980

tcatttgaac tttaacattt ggatcatagt ctaatttcat tgcctttttc caaaattgaa 5040

tccattgttt ttgattcacg tagttttctg tattcttaaa ataagttggt tccacacata 5100

ccaatacatg catgtgctga ttataagaat tatctttatt atttattgtc acttccgttg 5160

cacgcataaa accaacaaga tttttattaa tttttttata ttgcatcatt cggcgaaatc 5220

cttgagccat atctgacaaa ctcttattta attcttcgcc atcataaaca tttttaactg 5280

ttaatgtgag aaacaaccaa cgaactgttg gcttttgttt aataacttca gcaacaacct 5340

tttgtgactg aatgccatgt ttcatgtctc tcctccagtt gcacattgga caaagcctgg 5400

atttacaaaa ccacactga tacaactttc ttgcgctgt ttacagattt tgtttatact 5460

ctaataattc agcacaatct ttactcttt cagccttttt aaattcaaga atatgcagaa 5520

gttcaaagta atcaacattt gcgattttct tttctctcca tggctcact tttccacttt 5580

ttgtcttgc cactaaaacc cttgatTTTT caictgaata aatgctacta ttaggacaca 5640

taatatataa agaaaccccc atctatttag ttatttgttt ggtcacttat aactttaaca 5700

gatgggggtt ttctgtgcaa ccaattttta gggttttcaa tactttaaaa cacatacata 5760

ccaacacttc aacgcacctt tcagcaacta aaataaaaat gacgttattt ctatatgtat 5820

caagaataga aagaactcgt ttttcgtac gctcaaaacg caaaaaaagc actcattcga 5880

gtgctttttc ttatcgctcc aaatcatgcg attttttctt ctttgctttt ctttgctcac 5940

gaagtctcgc atcacgctgc aaaacatctt gaagcgaaaa agtattcttc ttttcttcgc 6000

atcgctcatg ctgacgcacg aaaagccctc taggcgcata ggaacaactc ctaaatgcat 6060

gtgaggggtt ttctcgtcca tgtgaacagt cgcatacgca atattttgtt tcccatactg 6120

cattaatgaa tcggccaacg cgcggggaga ggcgggtttgc gtattgggcg ctcttcgcgt 6180

tcctcgtca ctgactcgt gcgctcggtc gttcggctgc ggcgagcggt atcagctcac 6240

tcaaaggcgg taatacggtt atccacagaa tcaggggata acgcaggaaa gaacatgtga 6300

gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat 6360

aggctccgcc cccctgacga gcatcacaaa aatcgacgt caagtcagag gtggcgaaac 6420

ccgacaggac tataaagata ccaggcgttt cccctggaa gtcctcctgt gcgtctcct 6480

gttccgacct tgccgcttac cggatactg tccgcctttc tcccttcggg aagcgtggcg 6540

ctttctcata gctcacgtg taggtatctc agttcgggtg aggtcgttcg ctccaagctg 6600

ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg cttatccgg taactatcgt 6660

cttgagtcca acccggttaag acacgactta tcgccactgg cagcagccac tggtaacagg 6720

attagcagag cgaggtatgt aggcgggtgct acagagttct tgaagtgggtg gcctaactac 6780

ggctacacta gaagaacagt atttggtaic tgcgctctgc tgaagccagt taccttcgga 6840

aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggttttttt 6900

gtttgcaagc agcagattac ggcagaaaa aaaggatctc aagaagatcc tttgatcttt 6960

tctacggggt ct 6972

<210> 76

<211> 6936

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 76

gacgctcagt ggaacgaaaa ctcacgttaa gggatttttg tcatgagatt atcaaaaagg 60

atcttcacct agatcctttt aaattaaaaa tgaagtttta aatcaatcta aagtatatat 120

gagtaaactt ggcttgacag ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc 180

tgtcaattcg agctcggtag ccggggatcc ttaagaagac taataaaaag tttctaaaag 240

catgaaatat cctgtatttt ggagttactt gccgttattt atggctaaag ccgaaaaaaa 300

gataaggagg gtgttgtata caaaaaacat tttgtatata attagagttg cttggaaccc 360

agtaaaatag ctgtttatag tgagtaggta ttattacctg atgtataccg ttggtatgta 420

cttggcagaa cgcctagccc agatcggcct gaaacaccac tttgccgtgg ccggtgacta 480

caacctgggtg ttgcttgatc agctcctgct gaacaaagac atggagcagg tctactgctg 540

|                                                                    |      |
|--------------------------------------------------------------------|------|
| taacgaactt aactgcggct ttagcgccga aggttacgt cgtgcacgtg gtgccgccgc   | 600  |
| tgccatcgtc acgttcagcg taggtgctat ctctgcaatg aacgccatcg gtggcgcccta | 660  |
| tgcagaaaac ctgccggica tcctgatctc tggctcaccg aacaccaatg actacggcac  | 720  |
| aggccacatc ctgcaccaca ccattgggtac tactgactat aactatcagc tggaaatggt | 780  |
| aaaacacgtt acctgcgcac gtgaaagcat cgtttctgcc gaagaagcac cggcaaaaat  | 840  |
| cgaccacgtc atccgtacgg ctctacgtga acgcaaaccg gcttatctgg aaatcgcgtg  | 900  |
| caacgtcgtt ggcgtgaat gtgttcgtcc gggcccgatc aatagcctgc tgcgtgaact   | 960  |
| cgaagttgac cagaccagtg tcaactgccg ttagatgcc gccgtagaat ggctgcagga   | 1020 |
| ccgccagaac gtcgtcatgc tggtcggtag caaactgcgt gccgctgccg ctgaaaaaca  | 1080 |
| ggctgttgcc ctacgggacc gcctgggctg cgtgttcacg atcatggctg ccgaaaaagg  | 1140 |
| ctttttcccg gaagatcatc cgaacttccg cggcctgtac tgggtgaag tcagctccga   | 1200 |
| aggtgcacag gaactgggtg aaaacgccga tgccatcctg tgtctggcac cggatattca  | 1260 |
| cgactatgct accgttggct ggaactcctg gccgaaaggc gacaatgtca tggatcatga  | 1320 |
| caccgaccgc gtcactttcg caggacagtc cttcgaaggt ctgtcattga gcacattcgc  | 1380 |
| cgcagcactg gctgagaaag caccitctcg cccggcaacg actcaaggca ctcaagcacc  | 1440 |
| ggtactgggt attgaggccg cagagcccaa tgcaccgtg accaatgacg aatgacgcg    | 1500 |
| tcagatccag tcgtgatca ctccgacac tactctgaca gcagaaacag gtgactcttg    | 1560 |
| gttcaacgtt tctcgcgtc cgattcctgg cgggtcctgt gtcgaactgg aatgcaatg    | 1620 |
| gggtcatatc ggttgggtccg taccttctgc attcggtaac gccgttgggt ctccggagcg | 1680 |
| tcgccacatc atgatgggtg gtgatggctc ttccagctg actgctcaag aagttgctca   | 1740 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| gatgatccgc tatgaaatcc cggatcatcat ctctctgata aacaaccgcg gttacgtcat | 1800 |
| cgaaatcgct atccatgacg gcccttataa ctacatcaaa aactggaact acgtggcct   | 1860 |
| gatcgacgtc ttcaatgacg aagatgggtca tggcctgggt ctgaaagctt ctactggtgc | 1920 |
| agaactagaa ggcgctatca agaaagcact cgacaatcgt cgcgggtccga cgctgatcga | 1980 |
| atgtaacatc gctcaggacg actgcactga aaccctgatt gcttggggta aacgtgtagc  | 2040 |
| agctaccaac tctcgaaac cacaagcgta aattctgtta aaaccggaca ttgaagaagg   | 2100 |
| tgttgcgtag cgtttaataa aaacatctgt ttatcgaagt ttaggaatag gaaaattaaa  | 2160 |
| aaaaacaaga cgggagttag tttttgaaat ggcttcttca actttttata ttcctttcgt  | 2220 |
| caacgaaatg ggcgaaggtt cgcttgaaaa agcaatcaag gatcttaacg gcagcggctt  | 2280 |
| taaaaatgag ctgatcgctt ctgatgcttt catgaacaaa tccggtgttg tgaagcaggt  | 2340 |
| tgctgacctg ttgaaagcac agggattataa ttctgctgtt tatgatggcg ttatgccgaa | 2400 |
| cccgaactgt accgcagttc tggaaggcct taagatcctg aaggataaca attcagactt  | 2460 |
| cgatcatctc ctccggtggtg gttctcccca tgactgcgcc aaagccatcg ctctggtcgc | 2520 |
| aaccaatggt ggtagagtca aagactacga aggtatcgac aaatctaaga aacctgccct  | 2580 |
| gcctttgatg tcaatcaaca cgacggctgg tacggcttct gaaatgacgc gtttctgcat  | 2640 |
| catcactgat gaagtccgtc acgttaagat ggccattgtt gaccgtcacg ttaccccgat  | 2700 |
| ggtttccgtc aacgatcctc tgttgatggt tggatatcca aaaggcctga ccgccgccac  | 2760 |
| cggtatggat gctctgacct acgcatttga agcttatctt tcaacggcag ctactccgat  | 2820 |
| caccgatgct tgcgctttga aagcagcttc catgatcgct aagaatctga agaccgcttg  | 2880 |



cgacaacggt aaggatatgc cggctcgtga agctatggct tatgcccaat tcctcgtcgg 2940

tatggccttc aacaacgctt cgcttggtta tgcctatgct atggctcacc agttgggcgg 3000

ttactacaac ctgccgcatg gtgtctgcaa cgtgttctg cttccgcatg ttctggctta 3060

taacgcctct gtcgttgctg gtcgtctgaa agacgttggg gttgctatgg gtctcgatat 3120

cgccaatctc ggtgataaag aaggcgcaga agccaccatt caggctgttc gcgatctggc 3180

tgcttccatt ggtattccag caaacctgac cgagctgggt gctaagaaag aagatgtgcc 3240

gcttcttgct gaccacgctc tgaaagatgc ttgtgctctg accaaccgcg gtcagggtga 3300

tcagaaagaa gttgaagaac tcttctgag cgctttctaa aagatgcgtt ataattttac 3360

aagcctgttt ttttaggaaa cgggcttata aaattttttt atttttttgc cgggtttttt 3420

cttgtattaa tatgtggaat atgttaataa tattaagaag aaattccgaa tttaactaaa 3480

caaaattatt ttgttatit aagccaatct gtcataata tcttgacatg agggttatta 3540

gttagtataa tagtcttgt cggttttaag agggatcctc tagagtcgac ctgcaggcat 3600

gcaagcttgg cgtaatcatg gtcatactg tttcctgtgt gaaattgtta tccgctcaca 3660

attccacaca acatacgagc cggaagcata aagtgtaaag cctgggggtgc ctaatgagtg 3720

agctaactca cattaattgc gttgcgtca ctgccgctt tccagtcggg aaacctgtcg 3780

tgccagccct tcaaacttcc caaaggcgag ccctagtgac attagaaaac cgactgtaaa 3840

aagtacagtc ggcattatct catattataa aagccagtca ttaggcctat ctgacaattc 3900

ctgaatagag ttcataaaca atcctgcatg ataaccatca caaacagaat gatgtacctg 3960

taaagatagc ggtaaatata ttgaattacc ttattaatg aattttcctg ctgtaataat 4020

gggtagaagg taattactat tattatgat atttaagtta aaccagtaa atgaagtcca 4080

tggataata gaaagagaaa aagcattttc aggtataggt gttttgggaa acaatttccc 4140  
 cgaaccatta tatttctcta catcagaaaag gtataaatca taaaactctt tgaagtcatt 4200  
 ctttacagga gtccaaatac cagagaatgt tttagatata ccatcaaaaa ttgtataaag 4260  
 tggctctaac ttatcccaat aacctaacct tccgtcgcta ttgtaaccag ttctaaaagc 4320  
 tgtatttgag tttatcacc ttgtcactaa gaaaataaat gcagggtaaa atttatatcc 4380  
 ttcttgtttt atgtttcggg ataaaacact aatatcaatt tctgtgggta tactaaaagt 4440  
 cgtttgttgg ttcaaataat gattaaatat ctcttttctc ttccaattgt ctaaatacat 4500  
 tttattaaag ttcatattgat atgcctccta aatttttata taaagtgaat ttaggaggct 4560  
 tacttgtctg ctttcttcat tagaatcaat ccttttttaa aagtcaatcc cgtttgttga 4620  
 actactcttt aataaaaataa tttttccgtt cccaattcca cattgcaata atagaaaatc 4680  
 catcttcate ggctttttcg tcatcatctg tatgaatcaa atgccttct tctgtgtcat 4740  
 caagggttaa tttttatgt atttctttta acaaaccacc ataggagatt aaccttttac 4800  
 ggtgtaaacc ttcttccaaa tcagacaaac gtttcaaatt ctttcttca tcatcggtca 4860  
 taaaatccgt atcctttaca ggatatattg cagtttcgtc aattgccgat tgtatatccg 4920  
 atttatattt atttttcggg cgaatcattt gaacttttac atttggatca tagtctaatt 4980  
 tcattgcctt ttccaaaaat tgaatccatt gtttttgatt cacgtagttt tctgtattct 5040  
 taaaataagt tggttccaca cataccaata catgcatgtg ctgattataa gaattatctt 5100  
 tattatttat tgtcacttcc gttgcacgca taaaaccaac aagattttta ttaatttttt 5160  
 tatattgcat cattcggcga aatccttgag ccatatctga caaactctta ttttaattctt 5220

cgccatcata aacatitttta actgttaatg tgagaaacaa ccaacgaact gttggctttt 5280

gtttaataac ttcagcaaca accttttgtg actgaatgcc atgtttcatt gctctcctcc 5340

agttgcacat tggacaaagc ctggatttac aaaaccacac tcgatacaac ttcttttcgc 5400

ctgtttcacg attttgttta tactctaata tttcagcaca atcttttact ctttcagcct 5460

ttttaaatc aagaatatgc agaagttcaa agtaatcaac attagcgatt ttcttttctc 5520

tccatggctc cacttttcca ctttttgtct tgtccactaa aacccttgat ttttcactcg 5580

aataaatgct actattagga cacataatat taaaagaaac ccccatctat ttagttattt 5640

gtttggtcac ttataacttt aacagatggg gtttttctgt gcaaccaatt ttaagggttt 5700

tcaatacttt aaaacacata cataccaaca ctccaacgca ctttcagca actaaaataa 5760

aatgacgtt atttctatat gtatcaagaa tagaaagaac tcgtttttcg ctacgtcaa 5820

aacgcaaaaa aagcactcat tcgagtgtt ttctttatcg ctccaaatca tgcgattttt 5880

tcctctttgc ttttctttgc tcacgaagt ttctgatcag ctgcaaaaca tcttgaagcg 5940

aaaaagtatt ctctttttct tccgatcgct catgtgacg cagcaaaagc cctctaggcg 6000

cataggaaca actcctaaat gcatgtgagg ggttttctcg tccatgtgaa cagtcgcata 6060

cgcaatattt tgtttcccat actgcattaa tgaatcggcc aacgcgcggg gagaggcggg 6120

ttgcgtattg ggcgctcttc cgcttctctg ctactgact cgctgcgctc ggtcgttcgg 6180

ctgcggcgag cggtatcagc tcaactcaaag gcggttaatac ggttatccac agaatacagg 6240

gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag 6300

gccgcgttgc tggcggtttt ccataggctc cgccccctg acgagcatca caaaaatcga 6360

cgctcaagtc agaggtggcg aaacccgaca ggactataaa gataccaggc gtttccccct 6420

ggaagctccc tcgtgcctc tcctgttccg accctgccgc ttaccggata cctgtccgcc 6480

tttctccctt cgggaagcgt ggcgttttct catagctcac gctgtaggta tctcagttcg 6540

gtgtaggctg ttcgtccaa gctgggctgt gtgcacgaac ccccggttca gcccgaccgc 6600

tgcgccttat ccgtaacta tcgtcttgag tccaaccgg taagacacga ctatcgcca 6660

ctggcagcag ccactggtaa caggattagc agagcgaggt atgtaggcgg tgctacagag 6720

ttcttgaagt ggtggcctaa ctacggctac actagaagaa cagtatttgg tatctgcgt 6780

ctgctgaagc cagttacctt cggaaaaaga gttggtagct cttgatccgg caaacaacc 6840

accgctggta gcggtggtt tttgtttgc aagcagcaga ttacgcgcag aaaaaagga 6900

tctcaagaag atcctttgat cttttctacg gggctct 6936

<210> 77  
 <211> 1509  
 <212> DNA  
 <213> Clostridium thermocellum

<400> 77

tttgatcctg gctcaggacg aacgctggcg gcgtgcctaa cacatgcaag tcgagcgggg 60

atatacgga ggittaccgg aagtatatcc tagcggcgga cgggtgagta acgctgggt 120

aacctacctc atacaggggg ataacacagg gaaacctgtg ctaataccgc ataacggggc 180

ggcatcgtcc tgttatcaaa ggagaaatcc ggtatgagat gggcccgcgt ccgattagct 240

agttggtgag gtaacggctc accaaggcga cgatcggtag ccgaactgag aggttggtcg 300

gccacattgg gactgagaca cggcccagac tcctacggga ggcagcagtg gggaatattg 360

cgcaatgggg gaaacctga cgcagcaacg ccgctgaag gaagaaggcc ttcgggttgt 420

aaacttcctt gattggggac gaaggaagtg acggtaccca aagaacaagc cacggctaac 480  
 tacgtgccag cagccgcggt aatacgtagg tggcgagcgt tgtccggaat tactgggtgt 540  
 aaagggcgcg taggcgggat gcaagtcaga tgtgaaattc cggggcctta ccccggggct 600  
 gcatctgaaa ctgtatctct tgagtgtctg agaggaaagc ggaattccta gtgtagcgg 660  
 gaaatgcgta gatattagga ggaacaccag tggcgaaggc ggctttcttg acagtaactg 720  
 acgctgaggc gcgaaagcgt ggggagcaaa caggattaga taccctggtg gtccacgccg 780  
 taaacgatgg atactaggtg taggaggtat cgacccttc tgtgccggag ttaacacaat 840  
 aagtatccca cctggggagt acggccgcaa ggttgaaact caaaggaatt gacggggggc 900  
 cgcaacaaga gtggagtatg tggtttaatt cgaagcaacg cgaagaacct taccagggt 960  
 tgacatccct ctgacagctc tagagatagg gcttccttcg gggcagagga gacagtggt 1020  
 gcatggttgt cgtcagctcg tgtctgaga tgttgggtta agtcccga cagcgcaac 1080  
 ccttctcgtt agttgccagc acgttaaggt gggcactcta gcgagactgc cggcgacaag 1140  
 tcggaggaag gtggggacga cgtcaaatca tcatgccctt tatgtccttg gctacacacg 1200  
 tactacaatg gctgctacaa agggaagcga taccgcgagg tggagcaaat ccccaaaagc 1260  
 agtcccagtt cggattgcag gctgaaactc gcctgcatga agtcggaatt gctagtaatg 1320  
 gcaggtcagc atactgccgt gaatacgttc cgggccttg tacacaccgc ccgtcacacc 1380  
 atgagagtct gcaacacccg aagtcatagt ctaaccgcaa ggaggcgct gccgaaggtg 1440  
 gggcagatga ttgggtgaa gtcgtaacaa ggtagccgta tcggaaggtg cggtggatc 1500  
 acctccttt 1509

<211> 1642  
 <212> DNA  
 <213> Clostridium cellulolyticum

<220>  
 <221> modified\_base  
 <222> (9)..(9)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (16)..(16)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (26)..(26)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (70)..(70)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (86)..(86)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (96)..(96)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (113)..(115)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (320)..(320)  
 <223> a, c, g or t

<220>  
 <221> modified\_base

<222> (443)..(443)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (1109)..(1109)

<223> a, c, g or t

<400> 78

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| tgatcctgng acaggncgag cgctgncggc gtcctaaca catgcgagtc gagcggagtt   | 60  |
|                                                                    |     |
| acctttagcn ctgagtattc ttgganatga tgcigncccg acagcgtcat ccnnaacaa   | 120 |
|                                                                    |     |
| ccttaatgaa atatattagt ggagttttgc atcacgcgtt ttatcaaagt gtcaacacat  | 180 |
|                                                                    |     |
| aatagtagaa gagaatgttc agtgctgaag gtaacttagc ggcggacggg tgagtaacgc  | 240 |
|                                                                    |     |
| gtgggcaacc tgcctgttac agggggataa cacagggaaa cttgtgctaa taccgataa   | 300 |
|                                                                    |     |
| cacaacgaag aagcatttcn ttgttgtcaa aggagcaatc cggtagacaga tgggcccgcg | 360 |
|                                                                    |     |
| tccaattagc tagtttgtga tgtaacggat caccaaggcg acgatttgta gccgaactga  | 420 |
|                                                                    |     |
| gaggttgatc ggccacattg ggnctgagac acggcccaga ctctacggg aggcagcagt   | 480 |
|                                                                    |     |
| ggggaatatt gcacaatggg ggaaacctg atgcagcaac gccgcgtgaa ggatgaaggt   | 540 |
|                                                                    |     |
| tttcggattg taaacttctt tagtcaggga cgaagaaaat gacggtacct gaagaataag  | 600 |
|                                                                    |     |
| ccacggctaa ctacgtgcca gcagccgagg taatacgtag gtggcaagcg ttgtccggaa  | 660 |
|                                                                    |     |
| ttactgggtg taaaggcggt gtagcgaggga atgtaagtca gatgtgaaat cccagggtt  | 720 |
|                                                                    |     |
| aacctggag ctgcatctga aactatgttt cttgagtgcc ggagaggaaa gcggaattcc   | 780 |
|                                                                    |     |
| tagttagcgc gtgaaatgcg tagatattag gaggaacacc agtggcgaag gcggctttct  | 840 |
|                                                                    |     |
| ggacggtaac tgacgtgag gcgcgaaagc gtggggagca aacaggatta gataacctgg   | 900 |
|                                                                    |     |
| tagtccacgc tgtaaagcat ggatactagg ttaggaggt atcgaccct tctgtgccgg    | 960 |

agttaacaca ataagtatcc cacctgggga gtacggccgc aaggttgaaa ctcaaaggaa 1020  
 ttgacggggg cccgcacaag cagtggagta tgtggtttaa ttcgaagcaa cgcaagaac 1080  
 cttaccaagg cttagacatat agcgggaatnc ggcagagatg tcgtagtcct tcgggactgc 1140  
 tatacacagg tggtagcatgg ttgtcgtcag ctctgtgtcgt gagatgttgg gttaagtccc 1200  
 gcaacgagcg caaccctgt tgctagttga taacattaag atgatcactc tagcgagact 1260  
 gccggtgaca aatcggagga aggtggggac gacgtcaaat catcatgccc cttatgtctt 1320  
 gggctacaca cgtactacaa tggctataac agagggaagc taagctgcaa agtggagcaa 1380  
 atccccaaaa atagtccag ttcagatggt gggctgcaac ccgccacat gaagtcggaa 1440  
 ttgctagtaa tggtaggtca gtatactgtc gtgaatacgt tcccgggcct tgtacacacc 1500  
 gcccgtcaca ccatgagagt ctgcaacacc cgaagtcgat agtctaaccg caaggaggac 1560  
 gtcgccgaag gtggggccga tgattggtgt gaagtcgtaa caagtagcc gtatcggaag 1620  
 gtgcggctgg atcacctcct tt 1642

<210> 79  
 <211> 1552  
 <212> DNA  
 <213> Thermoanaerobacterium saccharolyticum

<220>  
 <221> modified\_base  
 <222> (64)..(64)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (100)..(100)  
 <223> a, c, g or t



<220>

<221> modified\_base

<222> (343)..(343)

<223> a, c, g or t

<400> 79

|                                                                   |      |
|-------------------------------------------------------------------|------|
| tttgatcctg gctcaggacg aacgctggcg gcgtgcctaa cacatgcaag tcgagcgatc | 60   |
| cggnactcaa ttaagcgctt acagaaaaag agagagaaan tgagtaaagc caaagttgag | 120  |
| tgccgatag cgccggacgg gtgagtaacg cgtggacaat ctaccctgta gtttgggata  | 180  |
| acacctcgaa aggggtgcta ataccggata atgtcaagaa gtggcatcac tttttgaaga | 240  |
| aaggagaaat ccgctatagg atgagtcgc gtccattag ctagtggcg gggtaaaagc    | 300  |
| ccaccaaggc gacgatgggt agccggcctg agagggtgaa cgnccacact ggaactgaga | 360  |
| cacggtccag actcctacgg gaggcagcag tggggaatat tgttcaatgg gggaaaccct | 420  |
| gacacagcga cgccgcgtga gcgaagaagg cttcgggtc gtaaagctca atagtatggg  | 480  |
| aagatagtga cggtaccata cgaaagcccc ggctaactac gtgccagcag ccgcgtaat  | 540  |
| acgtaggggg cgagcgttgt ccggaattac tgggcgtaaa gagcacgtag gcggctgtaa | 600  |
| aagtcagatg taaaaacct gggctcaacc gaggggtgtgc atctgaaact aaacagcttg | 660  |
| agtcaaggag aggagagcgg aattcctgtg ttagcgtga aatgcgtaga gatcaggaag  | 720  |
| aataccagtg gcgaaagcgg ctctctggac ttgaactgac gctgaggtgc gaaagcgtgg | 780  |
| ggagcaaaca ggattagata ccctggtagt ccacgccgta aacgatggat actaggtgtg | 840  |
| ggtaagcat catccgtgcc ggagttaacg caataagtat cccgcctggg gagtacggcc  | 900  |
| gcaagttga aactcaaagg aattgacggg ggccgcaca agcagcggag catgtggttt   | 960  |
| aattcgaagc aacgcgaaga acctaccag ggcttgacat ccacagaatc aggtagaaat  | 1020 |

accagagtgc ctcgaaagag gagctgtgag acaggtgggtg catggtttgc gtcagctcgt 1080

gtcgtgagat gttgggttaa gtcccgaac gagcgcaacc cctgttggta gttaccagcg 1140

taaagacggg gactctaccg agactgccgt ggagaacacg gaggaaggcg gggatgacgt 1200

caaatcatca tgccctttat gccctgggct acacacgtgc tacaatggcc tgaacagagg 1260

gcagcgaagg agcgatccgg agcgaatccc agaaaacagg tcccagttca gattgcaggc 1320

tgcaacccgc ctgcatgaag acggagtgtc tagtaatcgc ggatcagcat gccgcggtga 1380

atacgttccc gggccttgta cacaccgcc gtcacaccac gagagtttac aacaccgaa 1440

gtcagtgacc taaccgaag ggaggagctg ccgaagggtg ggtaaatgat tggggtgaag 1500

tcgtaacaag gtagccgtat cggaagggtc ggctggatca cctcctttcc ct 1552

<210> 80

<211> 1519

<212> DNA

<213> Clostridium stercorarium

<400> 80

tttgatcctg gctcaggacg aacgtggcg gcgtgcctaa cacatgcaag tcgaacggga 60

tccgtgttac ggaggtcttt ggaccgaagt ggcatggtga gagtggcgga cgggcgagta 120

acgcgtgagc aacctgcctt atgctggggg ataacaccgg gaaaccggtg ctaataccgc 180

ataagaccac agtgacgcat gtacagtgtt aaagctgagg cggcatagga tgggctcgcg 240

gtccattagc tagttggtag ggtaacggcc taccaaggcg acgatcggtg gccggactga 300

gaggttggcc ggccgcattg ggactgagac acggcccaga ctctacggg aggcagcagt 360

ggggaatatt gcgcaatggg ggaaacctg acgcagcgac gccgcgtgga ggaagaaggc 420

ctttgggttg taaactcctt tgatcgggga cgaagatgac ggtaccgaa gaacaagcca 480

cggtacta cgtgccagca gccgcggtaa tacgtagggt gcgagcgttg tccggaatta 540  
 ctgggtgtaa agggcgtgta ggccgggtgc caagtcaggt gtgaaatacc ggggcctaac 600  
 ctccgggtg catctgaaac tgggtctctt gaggccgga gaggaagcg gaattcccag 660  
 ttagcgggtg aaatgcgtag atattgggag gaacaccagt gccgaaggcg gctttctgga 720  
 cggtactga cgctgaggcg cgaaagcgtg gggagcaaac aggattagat accctggtag 780  
 tccacgtgt aaacgatgga tactagggtg aggaggtatc gacccttct gtgccgtagt 840  
 taacacaata agtatccac ctggggagta cgcccgcaag gctgaaactc aaaggaattg 900  
 acgggggccc gcacaagcag tggagcatgt ggtttaattc gaagcaacgc gaagaacctt 960  
 accagggtt gacatcccc tgacgatgt agagatacat cttctccga aggagcagg 1020  
 gagacaggtg gtcatggtg cagctcagct cgtgtcgtga gatgttgggt taagtccgc 1080  
 aacgagcgca acccttctg ttagttgcca gcagtaagat gggcactcta acgagactgc 1140  
 cgcgagaag tcggaggaag gtgggatga cgtcaaatca tcatgccct tatgtcctgg 1200  
 gctacacacg tgctacaatg gcgactacag aggaagcaa atccggcagg aggagcaaat 1260  
 cccgaaaggt cgtcccagtt cggattgcag gctcgaactc gcctgcatga agccggaatt 1320  
 gctagtaatg gcaggtcagc atactgccgt gaatacgttc ccgggccttg tacacaccgc 1380  
 ccgtcacacc atgagagctg gcaacaccg aagccgtagc ctaaccgaga gggggcgcc 1440  
 gtcgaaggtg gggcaggtga ttgggtgaa gtcgtaacaa ggtagccgta tcggaaggtg 1500  
 cggtggatc acctcttt 1519

<210> 81

<211> 1500

<212> DNA

<213> Clostridium stercorarium II

<400> 81

|                                                                   |      |
|-------------------------------------------------------------------|------|
| cctggctcag gacgaacgct ggcgcgctgc ctaacacatg caagtcgaac gggatccgtg | 60   |
| ttacggaggt cticggaccg aagtggcatg gtgagagtgg cggacgggcg agtaacgcgt | 120  |
| gagcaacctg ccctatgctg ggggataaca ccgggaaacc ggtgctaata cgcataaga  | 180  |
| ccacagtgc gcatgtcaca gtggtaaaag ctgaggcggc ataggatggg ctgcgctccg  | 240  |
| attagctagt tggtagggta acggcctacc aaggcgacga tcggtagccg gactgagagg | 300  |
| ttggccggcc gcattgggac tgagacacgg ccagactcc tacgggaggc agcagtgggg  | 360  |
| aatattgcgc aatgggggaa accctgacgc agcgacgccg cgtggaggaa gaaggccttt | 420  |
| gggttgtaaa ctctttgat cggggacgaa gatgacggta cccgaagaac aagccacggc  | 480  |
| taactacgtg ccagcagccg cggtaatagc taggtggcga gcgttgctcg gaattactgg | 540  |
| gtgtaaaggc cggttaggcg ggggtccaag tcaggtgtga aataccgggg cttaacctcg | 600  |
| ggggtgcac tgaaactggt gctcttgagt gccggagagg aaagcgaat tcccagtgtg   | 660  |
| gcggtgaaat gcgtagatat tgggaggaac accagtggcg aaggcggctt tctggacggt | 720  |
| aactgacgct gaggcgcgaa agcgtgggga gcaaacagga ttagataccc tggtagtcca | 780  |
| cgctgtaaac gatggatact aggtgtagac cccttctgtg ccgtagttaa cacaataagt | 840  |
| atcccacctg gggagtacga ggtatcgggc cgcaaggctg aaactcaaag gaattgacgg | 900  |
| gggcccgcac aagcagtgga gcatgtggtt taattcgaag caacgcgaag aaccttacca | 960  |
| gggcttgaca tccccctgac ggatgtagag atacatcttc tccgcaagga gcaggggaga | 1020 |
| caggtgggtc atggtgtgct tcagctcgtg tcgtgagatg ttgggttaag tcccgaacg  | 1080 |

agcgcaaccc ttgtcgtagg ttgccagcag taagatgggc actctaacga gactgccggc 1140

gagaagtcgg aggaaggtgg ggatgacgtc aaatcatcat gcccttatg tcctgggcta 1200

cacacgtgct acaatggcga ctacagaggg aagcaaatcc gcgaggagga gcaaattccg 1260

aaaggtcgtc ccagttcgga ttgcaggctg caactcgctt gcatgaagcc ggaattgcta 1320

glaatggcag gtcagcatat tgccgtgaat acgttcccgg gccttgtaca caccgcccgt 1380

cacacatga gagctggcaa caccgaagc cggtagccta accgagaggg gggcgccgtc 1440

gaaggtgggg caccgaagc cggtagccta accgagaggg gggcgccgtc gaaggtgggg 1500

<210> 82  
 <211> 1508  
 <212> DNA  
 <213> Caldiscellulosiruptor kristjanssonii

<220>  
 <221> modified\_base  
 <222> (213)..(213)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (223)..(223)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (405)..(405)  
 <223> a, c, g or t

<400> 82  
 ggctcaggac gaacgtggc ggcgtgccta acgcatgcaa gtcgagcgga gatggtagct 60

gaaggtgatg agctggaagc tatcatctta gcggcggacg ggtgagtaac acgtgagcaa 120

cctacccctca gcacggggat aacagctcga aagggtctgt aatacccgat gggaccacgg 180

catcgcatgg tgctgtggtg aaagggtagc cgnagaggct atnccggctg gggatgggct 240

cgcggcccat cagctagtig gtggggtaac ggcctaccaa ggcgacgacg ggtagccggc 300

ctgagagggt gtacggccac agtgggactg agacacggcc cacactccta cgggaggcag 360

cagcggggaa tcttgcgcaa tgggcgaaag cctgacgcag cgacnccgcg tgagggaaga 420

agcccttcgg ggtgtaaacc tctttggacg gggagaagtg gaagatagta cccgtttaa 480

aagccacggc taactacgtg ccagcagccg cgtaatacgt taggtggcga gcgttgtccg 540

gaattactgg gcgtaaaggg tgcgtaggcg gcctggtgta ttgagcgtga aatttttggg 600

ctcaacccaa aaggagcgct caagactgcc gggcttgagt gcgggagagg acggcggaat 660

tcccgtgta gcgtgaaat gcgtagatat cgggaggaac accagtggcg aaggcggccg 720

tctggaccgt aactgacgtg gaggcacgaa agcgtgggga gcaaacagga ttagataccc 780

tggtagtcca cgctgtaaac gatggatgct aggtgtgggg gagaagaact ctccgtgcc 840

gtagttaaca caataagcat cccgcctggg gactacggtc gcaaggttga aactcaaagg 900

aattgacggg ggccgcaca agcgttgag catgtggttt aattcgaagc aacgcgaaga 960

accttaccag ggcttgacat gccgggaacc ctgccgaaag gcgggggtgc ctgcttgta 1020

agagcaggag cccggacaca ggtggtgcat ggtgtcgtc agctcgtgc gtgagatgtt 1080

gggttaagtc ccgaacgag cgcaaccct gcccttagtt gccagcggtt ttagccgggc 1140

actctaaggg gactgccgcc gatgaggcgg aggaaggtgg gcatgacgtc aatcatcat 1200

gccccttatg ccttgggcta cacacgtgct acaatgggtg ctacagaggg cggcgaaggc 1260

gcgagccgga gcgaatccca aaaaagcacc cccagttcgg attgcaggt gcaactcgcc 1320

tgcatagaat cggaatcgct agtaatcgcg gatcagcatg ccgcggtgaa tacgttcceg 1380

ggccttgtag acaccgcccg tcacacatg agagtcagca acacctgaag acacaggata 1440

tctgtgttga aggtggggct gatgattggg gtgaagtcgt aacaaggtag ccgtacggga 1500

acgtgcgg 1508

<210> 83

<211> 1370

<212> DNA

<213> Clostridium phytofermentans

<220>

<221> modified\_base

<222> (240)..(243)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (247)..(247)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (250)..(250)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (293)..(293)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (413)..(413)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (421)..(427)

<223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (492)..(492)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (510)..(511)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (530)..(530)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (587)..(587)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (616)..(617)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (646)..(648)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (674)..(674)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (687)..(687)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (722)..(723)  
 <223> a, c, g or t



<220>  
 <221> modified\_base  
 <222> (771)..(771)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (812)..(813)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (848)..(849)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (905)..(906)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (916)..(916)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (923)..(923)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (1024)..(1025)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (1076)..(1077)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (1092)..(1093)  
 <223> a, c, g or t

<220>  
 <221> modified\_base

<222> (1149)..(1150)

<223> a, c, g or t

<400> 83

|                                                                    |      |
|--------------------------------------------------------------------|------|
| cttagtggcg gacgggtgag taacgcgtgg gtaacctgcc tcatacaggg ggataacagt  | 60   |
| cggaacgat tgctaaaacc gcataatata gcgaaaccgc atgattttgc tatcaaatat   | 120  |
| ttataggtat gagatgggcc cgcgtctgat tagctagttag gtggggtaat ggcctaccaa | 180  |
| ggcgacgac agtagccggc ttgagagagt gaccggccac attgggactg agacacggcn   | 240  |
| nnnactnctn cgggaggcag cagtggggaa tattggacaa tgggggaaac ccngatccag  | 300  |
| cgacgccg cg tgagtgaaga agtatttcgg tatgtaaagc tctatcagca gggaagataa | 360  |
| tgacagtacc tgactaagaa gccccggcta actacgtgcc agcagccg cg gtnatacgta | 420  |
| nnnnnnnagc gttatccgga ttactgggt gtaaaggag cgtaggtggt aggtcaagtc    | 480  |
| agatgtgaaa gnccagggt caacctgg nctgcatttg aaactggctn actgagtga      | 540  |
| ggagaggtaa gtggaattcc tagttagcg gtgaaatcg tagatantag gaggaacacc    | 600  |
| agtggcgaag gcggcnact ggactgtaac tgacactgag gctcgnnngc gtggggagca   | 660  |
| aacaggatta gainccctgg tagtcnccg cgtaaagat gaatactagc tgttcggggt    | 720  |
| cnnacagggc ttcggtggcg cagtaacgc aataagtatt ccacctggg ngtagcttcg    | 780  |
| caagaatgaa actcaaagga attgacgggg annccacaa gcggtggagc atgtggttta   | 840  |
| attcgaanna acgcaagaa ccttaccag tcttgacatc cctctgacaa ccgagtaacg    | 900  |
| tcggnnttct tcggncaga ggngacaggt ggtgcatggt tgtcgtcagc tcgtgtcgtg   | 960  |
| agatgttggg ttaagtcccg caacgagcg aacctatc tttagtagcc agcagttcgg     | 1020 |
| ctgnncactc tagagagact gccagggata acctggagga aggcggggat gacgttnaat  | 1080 |

catcatgccc cnnatgattt gggctacaca cgtgctacaa tggtgactac aaagagaagc 1140

aagcctgcnn gggggagcaa atctcaaaaa ggtcatccca gttcggattg tactctgcaa 1200

ctcgagtaca tgaagctgga atcgctagta atcgcgatc agaatgtcgc ggtgaatacg 1260

ttcccggtc ttgtacacac cgyycgtcac tccatgggag taggtaacgc ccgaagtcag 1320

tgaccyaacc gtaaggaggg agctgccgaa ggcgggatct ataactgggg 1370

<210> 84

<211> 57

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 84

gaattcgagc tcgtaccgc gggatcctct agagtcgacc tgcagggcatg caagctt 57

<210> 85

<211> 1515

<212> DNA

<213> Thermoanaerobacter pseudoethanolicus

<220>

<221> modified\_base

<222> (69)..(69)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (71)..(71)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (1427)..(1427)

<223> a, c, g or t

<400> 85

|                                                                   |      |
|-------------------------------------------------------------------|------|
| cctggctcag gacgaacgt ggcggcgtgc ctaacacatg caagtcgagc ggtccggcag  | 60   |
| ccaacttang ncgggagccg gatagcggcg gacgggtgag taacgcgtgg gcaacctacc | 120  |
| cttaagaccg ggataacacc tcgaaagggg tgctaatact ggataagctc cttgtagggc | 180  |
| atggtatgag gaggaaggta gcgggactac cgcttaagga tgggcccgcg tcccatcagc | 240  |
| tagttggtag ggtaacggcc taccaaggcg acgacgggta gccggcctga gagggtggtc | 300  |
| ggccacactg ggactgagac acggcccaga ctctacggg aggcagcagt ggggaatctt  | 360  |
| gcgcaatggg cgaaagcctg acgcagcgac gccgcgtgag cgaggaaggc cttcgggtcg | 420  |
| taaagctcga tagtgtggga agaagggatg acggtaccac acgaaagccc cggctaacta | 480  |
| cgtgccagca gcctcggtaa gacgtagggg gcgagcgttg tccggaatta ctgggcgtaa | 540  |
| agggcgcgta ggcgccgctt caagtcaggt gtaaaatacc cgggctcaac ccggggatag | 600  |
| catttgaaac tgggcggcta gagggcagga gaggggagtg gaattcccgg tgtagcggtg | 660  |
| aatgcgtag atatcgggag gaataccagt gccgaaggcg actctctgga ctgaccctga  | 720  |
| cgctgaggcg cgaaagcgtg gggagcaaac aggattagat accctggtag tccacgccgt | 780  |
| aaacgatggg tactaggtgt gggatgcgga agcattccgt gccgtagtta acgcaataag | 840  |
| tacccgcct ggggagtacg gccgcaaggt tgaaactcaa aggaattgac gggggccgc   | 900  |
| acaagcggtg gagcatgttg tttaattcga agcaacgca agaaccttac cagggttga   | 960  |
| catgcaggta gtagcgagcc gaaaggtag cgaccttacc ttaaaggtag ggagcctgca  | 1020 |
| caggtgggtc atggttgtcg tcagctcgtg tcgtgagatg ttgggttaag tcccgaacg  | 1080 |

agcgcaaccc ctgcctctag ttgccagcgg gtgaagccgg gcacgctaga gggactgccg 1140

tggacaacac ggaggaaggt ggggatgacg tcaaatcatt atgccctata tgccctgggc 1200

cacacacgtg ctacaatggc cgggtacagag ggaagcgaag ccgcgaggtg gagcgaaacc 1260

caaaaagccg gtccaagttc ggattgcagg ctgcaactcg cctgcatgaa gtcggaatcg 1320

ctagtaatcg cggatcagca tgccgcggtg aatacgttcc cgggccttgt acacaccgcc 1380

gtcacacca cgagagtctg caacaccga agccgtgacc caaccgnaag gagggagccg 1440

tcgaaggtgg ggcagatgat tgggtgaag tcgtaacaag gtagccgtat cggaaggtgc 1500

ggctggatca cctcc 1515

<210> 86

<211> 1395

<212> DNA

<213> Thermoanaerobacter sp.

<400> 86

ctacacatgc agtcgagcga agggagtact acggtacgaa cttagcggcg gacgggtgag 60

taacgcgtgg acaatctacc ctgtagaccg ggataacacc tcgaaagggg tgctaatacc 120

ggataatgtc gagaagcggc atcgcttttt gaagaaagga gagaatccgc tataggagga 180

gtccgcgtcc cattagctag ttggcgaggg taaaagccca ccaaggcgac gatgggtagc 240

cggcctgaga gggatgaacgg ccacactgga actgagacac ggtccagact cctacgggag 300

gcagcagtgg ggaatatgtt gcaatggggg aaaccctgac acagcgacgc cgcgtgagtg 360

aagaaggcct tcgggtcgta aagctcaata gtatgggaag aaagaaatga cggtaccata 420

cgaaagcccc ggctaactac gtgccagcag ccgcggtaat acgtaggggg cgagcgttgt 480

ccggaattac tgggcgtaaa gagcacgtag gcggtataa aagtcagatg tgaaaaacct 540

gggctcaacc gaggtatgc atctgaaact aaatagcttg agtcaaggag aggagagcgg 600  
 aattcctggt gtagcggtag aatgcgtaga gatcaggaag aataccagtg gcgaaagcgg 660  
 ctctctggac ttgaactgac gctgaggtag gaaagcgtgg ggagcaaaca ggattagata 720  
 ccctggtagt ccacgccgta aacgatggat actaggtgtg ggtagatat aatccgtgcc 780  
 ggagttaacg caataagtat cccgcctggg gtagcggcc gcaaggttga aactcaaagg 840  
 aattgacggg ggcccgcaca agcagcggag catgtggttt aattcgaagc aacgcgaaga 900  
 accttaccag ggcttgacat ccacagaatc gagtagaat acttgagtgc ctgtaagag 960  
 gagctgtgag acaggtggtg catggtgtc gtcagctcgt gtcgtgagat gttgggttaa 1020  
 gtcccgcaac gagcgcaacc cctgttggtg gttaccagcg taaagacggg gactctaccg 1080  
 agactgccgt ggataacacg gaggaaggcg gggatgacgt caaatcatca tgccctttat 1140  
 gccctgggct acacacgtgc tacaatggcc tgaacagagg gcagcgaagg agcgatccgg 1200  
 agcgaatccc agaaaacagg tcccagttca gattgcaggc tgcaacccgc ctgcatgaag 1260  
 acggagttag tagtaatcgc ggatcagcat gccgcggtga atacgttccc gggccttgta 1320  
 cacaccgccc gtcacaccac gagagtttac aacaccgaa gtcagtgacc taaccgaag 1380  
 ggaggagctg ccgaa 1395

<210> 87  
 <211> 1552  
 <212> DNA  
 <213> Thermoanaerobacterium saccharolyticum

<220>  
 <221> modified\_base  
 <222> (64)..(64)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (100)..(100)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (343)..(343)

<223> a, c, g or t

<400> 87

tttgatcctg gctcaggacg aacgctggcg gcgtgcctaa cacatgcaag tcgagcgatc 60

cggnactcaa ttaagcgtt acagaaaaag agagagaaan tgagtaaagc caaagttgag 120

tgccgtagat cgccggacgg gtgagtaacg cgtggacaat ctaccctgta gtttgggata 180

acacctcgaa aggggtgcta ataccggata atgtcaagaa gtggcatcac tttttgaaga 240

aaggagaaat ccgctatagg atgagtcgc gtccattag ctagtggcg gggtaaaagc 300

ccaccaaggc gacgatgggt agccggcctg agagggtgaa cgncacact ggaactgaga 360

cacgtccag actcctacgg gaggcagcag tggggaatat tgttcaatgg gggaaacct 420

gacacagcga cgccgcgtga gcgaagaagg cttcgggtc gtaaagctca atagtatggg 480

aagatagtga cggtaccata cgaaagcccc ggctaactac gtgccagcag ccgcgtaat 540

acgtaggggg cgagcgttgt ccggaattac tgggcgtaaa gagcacgtag gcggctgtaa 600

aagtcagatg tgaaaaacct gggctcaacc gaggtgtgc atctgaaact aaacagcttg 660

agtcaaggag aggagagcgg aattccttgt gtagcgtga aatgcgtaga gatcaggaag 720

aataccagtg gcgaaagcgg ctctctggac ttgaactgac gctgaggtgc gaaagcgtgg 780

ggagcaaaca ggattagata ccctggtagt ccacgccgta aacgatggat actaggtgtg 840

ggtgaagcat catccgtgcc ggagttaacg caataagtat cccgcctggg gagtacggcc 900  
 gcaaggttga aactcaaagg aattgacggg ggccccgaca agcagcggag catgtggttt 960  
 aattcgaagc aacgcgaaga accttaccag ggcttgacat ccacagaatc aggtagaaat 1020  
 accagagtgc ctcgaaagag gagctgtgag acaggtgggtg catggttgtc gtcagctcgt 1080  
 gtcgtgagat gttgggttaa gtcccgaac gagcgcaacc cctgttggtg gttaccagcg 1140  
 taaagacggg gactctaccg agactgccgt ggagaacacg gaggaaggcg gggatgacgt 1200  
 caaatcatca tgccctttat gccctgggct acacacgtgc tacaatggcc tgaacagagg 1260  
 gcagcgaagg agcgatccgg agcgaatccc agaaaacagg tcccagttca gattgcaggc 1320  
 tgcaacccgc ctgcatgaag acggagttgc tagtaatcgc ggatcagcat gccgcggtga 1380  
 atacgttccc ggcccttgta cacaccgcc gtcacaccac gagagtttac aacacccgaa 1440  
 gtcagtgacc taaccgcaag ggaggagctg ccgaaggtgg ggtaaagat tggggtgaag 1500  
 tcgtaacaag gtagccgtat cggaaggtgc ggctggatca cctcctttcc ct 1552

<210> 88

<211> 1553

<212> DNA

<213> Thermoanaerobacterium saccharolyticum

<400> 88

ttgatcctg gtcaggacg aacgctggcg gctgcctaa cacatgcaag tcgagcgatc 60  
 cggcactcaa ctaagcgctt acagaaaaag agagagaaaa tgagtaaagc caaagttgag 120  
 tgccgtagat cggcggacgg gtgagtaacg cgtggacaat ctacctgta gtttgggata 180  
 acacctgaa aggggtgcta ataccgata atgtcaagaa gtggcatcac tttttgaaga 240  
 aaggagaaat ccgctatagg atgagtccgc gtccattag ctagttggcg gggtaaaagc 300



ccaccaaggc gacgatgggt agccggcctg agagggtgaa cgccacact ggaactgaga 360

cacggtccag acicctacgg gaggcagcag tggggaatat tgtgcaatgg gggaaaccct 420

gacacagcga cgccgcgtga gcgaagaagg ccttcgggtc gtaaagctca atagtatggg 480

aagatagtga cggtaccata cgaaagcccc gggctactac gtgccagcag ccgcggtaat 540

acgtaggggg cgagcgttgt ccggaattac tgggcgtaaa gagcacgtag gcggctgtaa 600

aagtcagatg taaaaaacct gggctcaacc gaggtgtgc atctgaaact aaacagcttg 660

agtcaaggag aggagagcgg aattcctgtg ttagcgtga aatgcgtaga gatcaggaag 720

aataccagtg gcgaaagcgg ctctctggac ttgaactgac gctgaggtgc gaaagcgtgg 780

ggagcaaaca ggattagata ccctggtagt ccacgccgta aacgatggat actaggtgtg 840

ggtgaagcat catccgtgcc ggagttaacg caataagtat cccgcctggg gagtacggcc 900

gcaaggttga aactcaaagg aattgacggg ggcccgcaaca agcagcggag catgtggttt 960

aattcgaagc aacgcgaaga acctiaccag ggcttgacat ccacagaatc tggtagaaat 1020

accggagtgc ctgaaagag gagctgtgag acaggtggtg catggttgtc gtcagctcgt 1080

gtcgtgagat gtgggttaa gtcccgaac gagcgcaacc cctgttgta gttaccagcg 1140

taaagacggg gactctaccg agactgccgt ggagaacacg gaggaaggcg gggatgacgt 1200

caaatcatca tgccctttat gccctgggct acacacgtgc tacaatggcc tgaacagagg 1260

gcagcgaagg agcgatccgg agcgaatccc agaaaacagg tcccagttca gattgcaggc 1320

tgcaaccgc ctgcatgaag acggagttgc tagtaatcgc ggatcagcat gccgcggtga 1380

atacgtttcc cgggccttgt acacaccgcc cgtcacacca cgagagttta caacaccga 1440

agtcagtgc ctaaccgaaa ggaaggagct gccgaagggtg gggtaaataa ttggggtgaa 1500

gtcgtaacaa ggtagccgta tcggaagggtg cggctggatc acctcctttc taa 1553

<210> 89

<211> 1569

<212> DNA

<213> Unknown

<220>

<223> Description of Unknown Organism: Consensus  
Sequence

<220>

<221> modified\_base

<222> (71)..(71)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (92)..(92)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (101)..(101)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (250)..(251)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (255)..(255)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (296)..(296)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (494)..(494)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (496)..(496)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (852)..(852)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (864)..(864)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (1016)..(1017)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (1048)..(1050)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (1067)..(1067)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (1158)..(1159)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (1407)..(1407)

<223> a, c, g or t

<400> 89

tttgatcctg gctcaggacg aacgctggcg gcgtgcctaa cacatgcaag tcgagcgatc

60

cggcactcaa ntaagcgctt acagaaaaag angagcgaaa ntgagtaaac gctaagtiga 120

gtgccggata gcggcggacg ggtgagtaac gcgtggacaa tctacctgt agtttgggat 180

aacacctga aaggggtgct aataccggat aatgtcaaga agtggcatcg cttttigaag 240

aaaggagagn naatnccgct ataggatgag tccgcgtccc attagctagt tggcnggggt 300

aaaagccac caaggcgacg atgggtagcc ggctgagag ggtgaacggc cacactggaa 360

ctgagacacg gtccagactc ctacgggagg cagcagtggg gaatatgtg caatggggga 420

aacctgaca cagcgacgcc gcgtgagcga agaaggcctt cgggtcgtaa agctcaatag 480

tatgggaaga tagnantgac ggtaccatac gaaagccccg gctaactacg tgccagcagc 540

cgcgtaata ctagggggc gagcgttgc cggaattact gggcgtaaag agcacgtagg 600

cggtgtaaa agtcagatgt gaaaaacctg ggctcaaccg aggtgtgca tctgaaacta 660

aacagcttga gtcaaggaga ggagagcgga attcctgggtg tagcgggtgaa atgcgtagag 720

atcaggaaga ataccagtgg cgaaagcggc tctctggact tgaactgacg ctgaggtgcg 780

aaagcgtggg gagcaaacag gattagatac cctggtagtc cacgccgtaa acgatggata 840

ctaggtgtgg gnigaggcat catnccgtgc cggagttaac gcaataagta tcccgcctgg 900

ggagtacggc cgcaaggttg aaactcaaag gaattgacgg gggcccgac aagcagcgga 960

gcatgtggtt taattcgaag caacgcgaag aaccttacca gggcttgaca tccacnaga 1020

atcgggtaga aataccagag tgcctcgnnn aaagaggagc tgtgagnaca ggtggtgcat 1080

ggttgcgtc agtcgtgtc gtgagatgtt gggttaagtc ccgcaacgag cgcaaccct 1140

gttggtagt accagcgnnt aaagacgggg actctaccga gactgccgtg gagaacacgg 1200

aggaaggcgg ggtgacgtc aaatcatcat gccctttatg ccctgggcta cacacgtgct 1260

acaatggcct gaacagaggg cagcgaagga gcatccgga gcgaatcca gaaaacaggt 1320

cccagttcag attgcaggct gcaacccgcc tgcataaga cggagttgct agtaatcgcg 1380

gatcagcatg ccgcggtgaa tacgttcccc gggccttgta cacaccgccc gtcacaccac 1440

gagagtttac aacacccgaa gtcagtgacc taaccgcaag ggaggagctg ccgaaggtgg 1500

ggtaaatgat tggggtgaag tcgtaacaag gtagccgtat cggaaggtgc ggctggatca 1560

cctcctttc 1569

<210> 90  
 <211> 1061  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Description of Unknown Organism: Consensus  
 Sequence

<220>  
 <221> modified\_base  
 <222> (657)..(657)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (789)..(789)  
 <223> a, c, g or t

<400> 90  
 gtgtatacaa tatatttctt ctttttagta agaggaaatgt ataaaaataa atattttaaa 60

ggaaggagc atcttatgag cattattcaa aacatcattg aaaaagctaa aagtgataaa 120

aagaaaattg ttctgccaga aggtgcagaa cccaggacat taaaagctgc tgaaatagtt 180

ttaaaagaag gaattgcaga ttgtgtgctt cttggaaatg aagatgagat aagaaatgct 240

gcaaaagact tggacataac caaagctgaa atcattgacc ctgtaaagtc tgaaatgttt 300

gataggtatg ctaatgattt ttatgagtta aggaagagca aaggaatcac gttggaaaaa 360

gccagagaaa caatcaagga taatatctat ttggatgta tgatggttaa agaaggttat 420

gctgatggat tggatatcgg cgctattcat gctactgcag atttattaag acctgcattt 480

cagataatta aaacggctcc aggagcaaag atagtatcaa gcttttttat aatggaagtg 540

cctaattgtg aatatggatg aaatggatga ttcttgtttg ctgattgtgc ggtcaatcca 600

tcgcctaata cagaagaact tgcttctatt gctgtacaat ctgctaatac tgcaaangaa 660

tttgttgggc ttgaaccaa aagttgctat gctatcattt tctacaaaag gtagtgcac 720

acatgaatta gtagataagg taagaaaagc gacagagata gcaaaagaat tgatgccaga 780

tgttgctant cgatggatga ttgcaattgg atgctgctct tgttaaagaa gttgcagagc 840

taaaagcgcc gggaagcaaa gttgcgggat gtgcaaatgt gcttatattc cctgatttac 900

aagctggtta tataggatat aagcttgtac agagattagc taaggcaaat gcaattggac 960

ctataacaca aggaatgggt gcaccgggtt atgatttatt aagaggatgc agctatagag 1020

atattgttga cgtaatagca acaacagctg tgcaggctca a 1061

<210> 91  
 <211> 1213  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Description of Unknown Organism: Consensus  
 Sequence

<220>  
 <221> modified\_base  
 <222> (52)..(52)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (72)..(72)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (136)..(136)

<223> a, c, g or t

<220>

<221> modified\_base

<222> (151)..(153)

<223> a, c, g or t

<400> 91

atgaaaatta tgaaaattct gggttattaat tgtggaagtt cttcactaaa antatcaatt 60

gattgaatca antgatggaa atgtgctggc aaaaggcctt gctgaaagaa tcggcataaa 120

tgattccctg ttgacncata atgctaacgg nnnagaaaaa atcaagataa aaaaagacat 180

gaaagatcac aaagacgcaa taaaatttgt tttagatgct ttggttaagta gtgactacgg 240

cgttataaag gatatgtctg agatagatgc tgtaggacat agagttgttc acggaggaga 300

gtcttttaca tcatacagttc tcataaatga tgaagtgtta aaagcgataa cagattgtat 360

agaattagct ccaactgcaca atcctgctaa tatagaagga attaaagctt gccagcaaat 420

catgccaaac gtccaatgg tggcggtatt tgatacagcc tttcatcaga caatgcctga 480

ttatgcatat ctttatccaa taccttatga atactacaca aagtacagga tcagaagata 540

tggatttcat ggcacatcgc ataaatatgt ttcaaataagg gctgcagaga ttttgaataa 600

acctattgaa gatttgaaaa tcataacttg tcatcttgga aatggctcca gtattgctgc 660

tgtaaataat ggtaaatcaa ttgacacaag catgggattt acaccattag aaggtttggc 720

tatgggtaca cgatctggaa gtatagacc atccattatt tcttatctta tggaaaaaga 780

aaatataagt gctgaagagg tagtaaatat attaaataaa aaatctgggtg tttacggtat 840

ttcaggaata agcagcgatt ttagagattt agaagatgcc gcctttaaaa atggagatga 900

aagagctcag ttggctttaa atgtgtttgc atatcgagta aagaagacga ttggcgctta 960

tgcagcagct atgggaggcg ttgatgtcat tgtatttaca gcaggtgttg gtgaaaatgg 1020

tcctgagata cgagaattta tacttgatgg attagagttt ttagggttca gcttggataa 1080

agaaaaaaat aaagtcagag gaaaagaaac tattatatct acgccgaatt caaaagttag 1140

cgatgatggtt gtgcctacta atgaagaata tatgattgct aaagatactg aaaagattgt 1200

aaagagtata aaa 1213

<210> 92  
 <211> 935  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Description of Unknown Organism: Consensus  
 Sequence

<220>  
 <221> modified\_base  
 <222> (15)..(15)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (90)..(90)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (117)..(117)  
 <223> a, c, g or t



<220>  
 <221> modified\_base  
 <222> (255)..(255)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (313)..(313)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (349)..(349)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (485)..(485)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (755)..(755)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (761)..(761)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (779)..(779)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (791)..(791)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (793)..(793)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (800)..(800)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (818)..(818)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (824)..(824)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (836)..(836)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (854)..(854)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (863)..(863)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (869)..(869)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (902)..(902)  
 <223> a, c, g or t

<220>  
 <221> modified\_base  
 <222> (917)..(917)  
 <223> a, c, g or t

<220>  
 <221> modified\_base

<222> (929)..(929)

<223> a, c, g or t

<400> 92

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| atgagcaaag tagcnataat aggttctgga ttgttaggtg ctacatctgc atttacactg | 60  |
|                                                                   |     |
| gctttaagtg ggactgtgac agatattgtg ttagtagatt taaacaagga caaggcnata | 120 |
|                                                                   |     |
| ggcgaatcac tggatataag ccatggcata ccgtttatac agcctgtaaa tgtgtatgca | 180 |
|                                                                   |     |
| ggtgactaca aagatgttga aggcgcagat gtaatagttg tgacagcagg tgctgctcaa | 240 |
|                                                                   |     |
| aagccgggag agacnaggct tgaccttgtg aagaaaaata cagctatatt taagtccatg | 300 |
|                                                                   |     |
| atacctgagc ttnttaaagt acaatgacaa ggctatatat ttgattgtga caaatcctgt | 360 |
|                                                                   |     |
| agataactg acgtacgtta catacaagat atctggactt ccatggggca gagttttcgg  | 420 |
|                                                                   |     |
| ttctggcact gttcttgaca gttcaagggt taggtatctt ttaagcaagc attgcaatat | 480 |
|                                                                   |     |
| agatnccgag aaatgtccac ggaaggataa ttggcgagca tggtgataca gagtttgcag | 540 |
|                                                                   |     |
| catggagcat aacaaacata tcaggaatat catttaatga gtactgcagt ttatgcggac | 600 |
|                                                                   |     |
| gcgtctgtaa cacaaatttc agaaaggaag tagaagatga agttgtaaat gctgcttata | 660 |
|                                                                   |     |
| agataataga caaaaagggt gctacataat atgctgtggc tgttgcagta agaaggattg | 720 |
|                                                                   |     |
| tggagtgtat cttaagagat gaaaattcca ttctnacagt ntcactcca ttaaattgnc  | 780 |
|                                                                   |     |
| aatacgggtg nanagatgtg tctttaagct tgccatcnat tgtnggcaga aatggngttg | 840 |
|                                                                   |     |
| caaggattct gganttgctt ttntctgang aagaagtga gaagtttaga cattcagcaa  | 900 |
|                                                                   |     |
| gngttatggc agatgtgnata aaacagttna atata                           | 935 |