



등록특허 10-2542130



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2023년06월13일

(11) 등록번호 10-2542130

(24) 등록일자 2023년06월07일

(51) 국제특허분류(Int. Cl.)

A01H 6/82 (2018.01) A01H 1/06 (2006.01)
A01H 5/10 (2018.01) C12N 15/113 (2010.01)
C12N 15/82 (2006.01) C12N 9/02 (2006.01)
C12N 9/22 (2006.01)

(52) CPC특허분류

A01H 6/825 (2022.05)
A01H 1/06 (2021.01)

(21) 출원번호 10-2022-0009633

(22) 출원일자 2022년01월24일

심사청구일자 2022년01월24일

(56) 선행기술조사문헌

KR1020190066718 A*

WO2019163601 A1*

NCBI Reference Sequence, Solanum lycopersicum
7-dehydrocholesterol reductase (DWF5-2),
mRNA, ACCESSION NM_001366114.1 (2018.10.18.)*

NCBI Reference Sequence, Solanum lycopersicum
7-dehydrocholesterol reductase
(LOC101248602), mRNA, ACCESSION
NM_001366115.1 (2018.10.18.)*

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

(주)지플러스생명과학

충청북도 청주시 흥덕구 오송읍 의료단지길 123,
2동

(72) 발명자

최선미

서울특별시 강서구 마곡중앙로 36, 1502동 1604
호(마곡동, 마곡엠밸리15단지)

김진화

서울특별시 강남구 광평로34길 55, 113동 1104호
(수서동, 강남대시앙포레)

(뒷면에 계속)

(74) 대리인

류종우, 이내영

전체 청구항 수 : 총 11 항

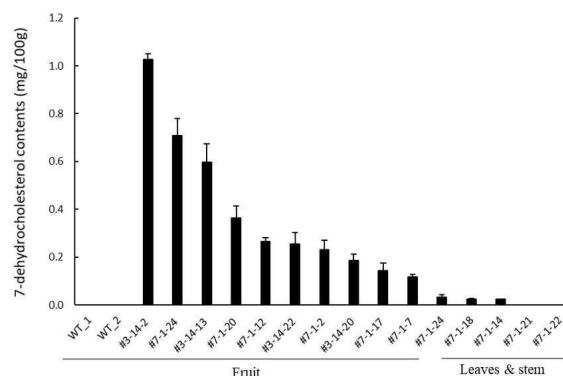
심사관 : 이근혜

(54) 발명의 명칭 7-디하이드로콜레스테롤이 고농도로 함유된 토마토 및 이의 제조 방법

(57) 요약

본 발명은 유전자 편집 기술을 이용하여 제조한 고농도의 비타민 D 전구체를 함유한 토마토 및 그 제조방법에 관한 것이다. 구체적으로, 토마토의 DWF5 유전자가 결실된 경우 고농도의 7-디하이드로콜레스테롤(7-dehydrocholesterol)이 축적될 뿐 아니라, 종자 생장이 가능함을 확인하였다. 따라서, 이와 같이 생산된 토마토는 고농도의 비타민 D3 전구체를 함유할 수 있으며, 상업적으로 활용도가 높다.

대표도 - 도28



(52) CPC특허분류

A01H 5/10 (2022.05)
C12N 15/1137 (2013.01)
C12N 15/8205 (2013.01)
C12N 15/8235 (2013.01)
C12N 9/001 (2013.01)
C12N 9/22 (2013.01)
C12N 2310/20 (2017.05)

(72) 발명자

김정모

서울특별시 강서구 화곡로64길 12, 202호(등촌동
만석빌라)

유민경

경기도 용인시 기흥구 사은로126번길 33, 신창 미
션힐 아파트 205-1903호

전윤아

대전광역시 중구 산성로 108-23 (문화주공3단지아
파트)306동 201호

명세서

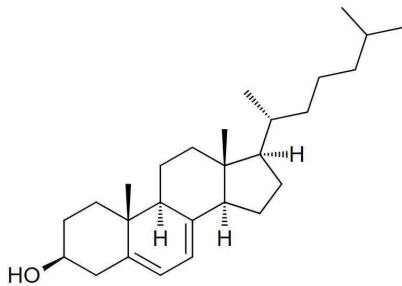
청구범위

청구항 1

하기 화학식 1의 7-디하이드로콜레스테롤(7-dehydrocholesterol)이 과실에 고농도로 함유된 형질전환된 토마토로서,

상기 토마토는 서열번호 1의 염기서열을 포함하는 DWF5-1 유전자의 기능이 상실된 것인 형질전환된 토마토:

[화학식 1]



청구항 2

제1항에 있어서,

상기 7-디하이드로콜레스테롤의 농도는 토마토 중량 100 g당 0.01 mg 내지 1.5 mg인 것인, 형질전환된 토마토.

청구항 3

삭제

청구항 4

제1항에 있어서,

상기 유전자의 기능 상실은 DWF5-1 유전자의 엑손 6, 엑손 7, 엑손 9, 엑손 10, 엑손 11 및 이들의 조합으로 이루어진 군에서 선택되는 어느 하나의 도메인의 변이에 의한 것으로,

상기 변이는 치환(conversion), 결실(deletion), 역위(inversion) 또는 삽입(insertion)에 의해 야생형과 상이한 것인, 형질전환된 토마토.

청구항 5

삭제

청구항 6

삭제

청구항 7

제4항에 있어서,

상기 엑손 6 도메인은 서열번호 9의 염기서열을 포함하는 것이며;

상기 엑손 7 도메인은 서열번호 10의 염기서열을 포함하는 것이며;

상기 엑손 9 도메인은 서열번호 11의 염기서열을 포함하는 것이며;

상기 엑손 10 도메인은 서열번호 12의 염기서열을 포함하는 것이며;

상기 엑손 11 도메인은 서열번호 13의 염기서열을 포함하는 것인, 형질전환된 토마토.

청구항 8

삭제

청구항 9

삭제

청구항 10

삭제

청구항 11

삭제

청구항 12

삭제

청구항 13

삭제

청구항 14

삭제

청구항 15

삭제

청구항 16

삭제

청구항 17

삭제

청구항 18

삭제

청구항 19

제1항에 있어서,

상기 형질전환된 토마토는 동형접합체(homozygote)인 것인, 형질전환된 토마토.

청구항 20

제1항에 있어서,

상기 형질전환된 토마토는 종자를 형성하는 것을 특징으로 하는, 형질전환된 토마토.

청구항 21

서열번호 1의 염기서열을 포함하는 DWF5-1 유전자의 엑손 6 도메인의 염기서열에 상보적으로 결합하는 적어도 하나의 sgRNA(single guide RNA) 및 CRISPR(clustered regularly interspaced palindromic repeats) 연관 단백질질을 암호화하는 염기서열을 포함하는 벡터.

청구항 22

삭제

청구항 23

제21항에 있어서,

상기 sgRNA는 서열번호 20, 서열번호 21 및 이들의 조합으로 이루어진 군으로부터 선택되는 어느 하나의 염기서열을 포함하는 것인, 벡터.

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

제21항에 있어서,

상기 CRISPR 연관 단백질은 Cas9, Cpf1, c2c1, C2c2, Cas13, c2c3, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5e(CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a, Cas8a1, Cas8a2, Cas8b, Cas8c, Csn1, Csx12, Cas10, Cas10d, Cas10, Cas10d, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1(CasA), Cse2(CasB), Cse3(CasE), Cse4(CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 및 Cul966으로 구성된 군으로부터 선택되는 어느 하나인 것인, 벡터.

청구항 31

제30항에 있어서,

상기 CRISPR 연관 단백질은 Cas9인 것인, 벡터.

청구항 32

제21항, 제23항, 제30항 및 제31항 중 어느 한 항의 벡터를 아그로박테리움을 이용하여 토마토에 도입하는 단계를 포함하는 과실에 7-디하이드로콜레스테롤이 고농도로 함유된 형질전환된 토마토를 제조하는 방법.

발명의 설명

기술 분야

본 발명은 비타민 D 전구체인 7-디하이드로콜레스테롤을 고농도로 함유하는 토마토 및 이의 제조 방법에 관한

[0001]

것이다.

배경 기술

- [0002] 비타민 D는 생명 유지에 필수적인 영양소로서, 비타민 D가 충분하지 못하면 구루병, 골연화증, 골다공증 뿐만 아니라 유방암, 대장암, 전립선암 등의 악성 종양, 고혈압을 비롯한 심혈관질환, 당뇨병, 다발성 경화증, 건선, 류머티스성 관절염, 결핵 등도 증가한다고 보고되고 있다. 비타민 D의 경우 식품을 통해 섭취할 수 있지만 적당한 햇빛을 통하여 체내에서도 합성이 되는 비타민이다. 따라서, 햇빛에 충분히 노출되지 못한 경우에는 비타민 D가 많이 함유된 식품이나 비타민 D 영양 보충제를 섭취하여야 한다.
- [0003] 특히 현대인은 하루 중 대부분을 실내에서 보내고 실외 활동을 할 때도 옷과 모자 등으로 일광을 차단하기 때문에 비타민 D의 결핍에 주의하여야 한다. 햇빛에 충분히 노출되지 못한 경우에는 식품을 통해 비타민 D를 충분히 섭취하여야 하지만, 실제로 비타민 D 함량이 높은 식품은 그리 많지 않다는 문제점이 있다. 비타민 D가 풍부하다고 알려진 식품이라 하더라도 100 g당 함유된 비타민 D의 양은, 슬라이스 치즈 0.29 μg /약 7장(100 g), 참치 2.9 μg /100 g, 달걀 1.36 μg /2개(100 g), 건포도 4.4 μg /약 1컵(100 g)으로 이와 같은 식품으로 매일 권장량을 다 채우기는 불가능하다.
- [0004] 또한, 비타민 D 영양 보충제로서 합성 비타민 D는 그 부작용 및 유해 가능성이 보고되고 있으며, 천연 비타민 D는 값이 매우 비싸고 구하기 어렵다는 문제점이 있다. 한편, 대한민국 공개공보 제10-2017-0138657호에는 DWF5(delta 5, 7-sterol deta 7 reductase, DWARF5), CPD(constitutive photomorphogenic DWAF, DWF3) 및 SMT1(sterol methyltransferase 1) 유전자를 결실 시킬 경우 비타민 D의 전구체인 7-디하이드로캠페스테롤(7-dehydrocampesterol)이 농축된 상추를 제조할 수 있다는 점을 개시하고 있다. 그러나, 전구비타민 D3로서 7-디하이드로콜레스테롤(7-dehydrocholesterol)이 농축된 토마토 제조에 대해서는 연구된 바가 없다.

선행기술문헌

특허문헌

- [0005] (특허문헌 0001) KR 10-2017-0138657 A1

발명의 내용

해결하려는 과제

- [0006] 이에, 본 발명자는 CRISPR 기술을 통해 DWF5, CPD 및 SMT1 유전자를 결실시켜 실험한 결과, DWF5 유전자가 결실 되었을 시 7-디하이드로콜레스테롤을 고농도로 함유한 토마토가 생산되는 것을 확인하여 본 발명을 완성하였다. 구체적으로, 본 발명은 프로비타민 D를 생산하기 위한 전구체인 7-디하이드로콜레스테롤을 고농도로 함유한 토마토를 개발하기 위해 최적화된 DWF5, CPD 및 SMT1 결실 부위를 확인하고, 이를 포함하는 비타민 D3 전구체가 농축된 토마토를 개발하는 것이다.

과제의 해결 수단

- [0007] 이를 달성하기 위하여, 본 발명의 일 측면은, 하기 화학식 1의 7-디하이드로콜레스테롤(7-dehydrocholesterol)이 고농도로 함유된 형질전환된 토마토를 제공한다.
- [0008] 본 발명의 다른 측면은, DWF5 유전자의 염기서열에 상보적으로 결합하는 하나 또는 하나 이상의 sgRNA(single guide RNA) 및 CRISPR(clustered regularly interspaced palindromic repeats) 관련 단백질을 암호화하는 염기서열을 포함하는 벡터를 제공한다.
- [0009] 본 발명의 또 다른 측면은, 상기 벡터를 아그로박테리움을 이용하여 토마토에 도입하는 단계를 포함하는 7-디하이드로콜레스테롤이 고농도로 함유된 형질전환된 토마토를 제조하는 방법을 제공한다.

발명의 효과

- [0010] 본 발명에 의해서 제조된 형질전환된 토마토가 동형접합체로서 DWF5 유전자의 변이를 포함하는 경우, 토마토의 과실 부위에 7-디하이드로콜레스테롤이 다량 축적되었다. 또한 상기 동형접합체(T2)는 종자를 생산할 수 있어,

원하는 토마토 형질을 보존할 수 있었다. 따라서, 본 발명의 형질전환 토마토는 상업적으로 활용도가 높다.

도면의 간단한 설명

[0011]

- 도 1은 DWF5 sgRNA에 의해 타겟팅되는 DWF5의 위치를 나타낸 도면이다.
- 도 2는 CPD sgRNA에 의해 타겟팅되는 CPD의 위치를 나타낸 도면이다.
- 도 3은 SMT1 sgRNA에 의해 타겟팅되는 SMT1의 위치를 나타낸 도면이다.
- 도 4는 RGEN을 이용하여 DWF5 sgRNA의 유전자 편집 효율을 확인한 결과를 나타낸 도면이다.
- 도 5는 RGEN을 이용하여 CPD sgRNA의 CPD 유전자 편집 효율을 확인한 결과를 나타낸 도면이다.
- 도 6은 RGEN을 이용하여 SMT1 sgRNA의 SMT1 유전자 편집 효율을 확인한 결과를 나타낸 도면이다.
- 도 7은 DWF5 sgRNA가 적재된 Cas9-sgRNA 벡터의 모식도를 나타낸 도면이다.
- 도 8은 DWF5 sgRNA 및 CPD sgRNA가 적재된 Cas9-sgRNA 벡터의 모식도를 나타낸 도면이다.
- 도 9는 DWF5 sgRNA, CPD sgRNA 및 SMT1 sgRNA가 적재된 Cas9-sgRNA 벡터의 모식도를 나타낸 도면이다.
- 도 10은 아그로박테리움을 이용한 형질전환 토마토를 제작하는 과정을 나타낸 흐름도이다.
- 도 11은 유전자 편집을 통해 제작된 토마토의 발아 과정을 나타낸 것이다.
- 도 12는 DWF5 유전자가 편집된 토마토 식물체를 관찰한 도면이다.
- 도 13은 DWF5 및 CPD 유전자가 편집된 토마토 식물체를 관찰한 도면이다.
- 도 14는 DWF5, CPD 및 SMT1 유전자가 편집된 토마토 식물체를 관찰한 도면이다.
- 도 15는 DWF5, CPD 또는 SMT1 유전자 편집 여부를 T7E1 시험(T7E1 assay)을 통해 확인한 결과를 나타낸 도면이다.
- 도 16은 DWF5 유전자 결실을 생어 염기분석을 통해 확인한 결과를 나타낸 도면이다.
- 도 17은 DWF5, CPD 및 SMT1 유전자 결실을 생어 염기분석을 통해 확인한 결과를 나타낸 도면이다.
- 도 18은 DWF5 유전자가 결실된 T1세대 형질전환 토마토 식물체에서 DWF5 유전자 표현형을 차세대 염기분석(next generation sequencing)을 통해 확인한 결과의 나타낸 도면이다.
- 도 19는 DWF5 유전자가 결실된 T1세대 동형접합체 형질전환 토마토 식물체의 유전자 편집 효율 및 수확된 종자량(위), 유전자 서열(중간) 및 식물체의 표현형(아래)을 나타낸 도면이다.
- 도 20은 DWF5 유전자가 결실된 동형접합체(T1)의 T2세대 형질전환 토마토 식물체에 함유된 7-디하이드로콜레스테롤을 측정하기 위한 시료 준비 과정을 나타낸 흐름도이다.
- 도 21은 DWF5 유전자가 결실된 동형접합체(T1)의 T2세대 형질전환 토마토 식물체에 함유된 7-디하이드로콜레스테롤을 LC/MS로 측정하기 위한 측정 조건을 나타낸 도면이다.
- 도 22는 야생형 토마토(WT) 식물체에 함유된 7-디하이드로콜레스테롤을 LC/MS로 측정한 결과를 나타낸 도면이다.
- 도 23은 표준시약으로 시판되고 있는 7-디하이드로콜레스테롤을 LC/MS로 측정한 결과를 나타낸 도면이다.
- 도 24는 DWF5 유전자가 결실된 동형접합체(T1)의 T2세대 형질전환 토마토 식물체에 함유된 7-디하이드로콜레스테롤을 LC/MS로 측정한 결과를 나타낸 도면이다.
- 도 25는 DWF5 유전자가 결실된 동형접합체(T1)의 T2세대 형질전환 토마토 식물체에 함유된 7-디하이드로콜레스테롤을 LC/MS로 측정한 결과를 나타낸 도면이다.
- 도 26은 DWF5 유전자가 결실된 동형접합체(T1)의 T2세대 형질전환 토마토 식물체에 함유된 7-디하이드로콜레스테롤을 LC/MS로 측정한 결과를 나타낸 도면이다.
- 도 27은 DWF5 유전자가 결실된 동형접합체(T1)의 T2세대 형질전환 토마토 식물체에 함유된 7-디하이드로콜레스테롤을 LC/MS로 측정한 결과를 나타낸 도면이다.

테를을 LC/MS로 측정한 결과를 나타낸 도면이다.

도 28은 DWF5 유전자가 결실된 동형 접합체(T1)의 T2세대 형질전환 토마토 식물체의 과실에 함유된 7-디하이드로콜레스테를을 LC/MS로 측정한 결과를 나타낸 그래프이다.

도 29는 DWF5 유전자가 결실된 동형접합체(T1)의 T2세대 형질전환 토마토 식물체에서 하이그로마이신(hygromycin) 및 포스피노트리신(phosphinothricin, PPT) 저항성 유전자의 발현을 RT-PCR로 확인한 결과를 나타낸 도면이다.

도 30은 항생제 저항성 유전자가 발현되지 않는 T2세대 형질전환 토마토 식물체(D100_3-14-29 및 D100_7-1-15)에서 DWF5의 엑손 6 도메인의 염기서열을 야생형(WT) 토마토 식물체와 비교 분석한 결과를 나타낸 도면이다.

도 31은 항생제 저항성 유전자가 발현되지 않는 T2세대 형질전환 토마토 식물체(D100_3-14-29 및 D100_7-1-15)에서 DWF5의 엑손 6 도메인의 아미노산 서열을 야생형(WT) 토마토 식물체와 비교 분석한 결과를 나타낸 도면이다.

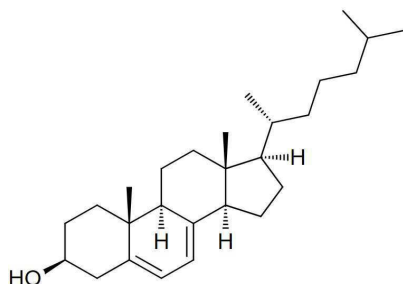
도 32는 DWF5 유전자가 결실된 동형 접합체(T2) 형질전환 토마토 식물체(T2)의 표현형을 나타낸 도면이다.

도 33은 토마토 식물체의 뿌리, 줄기, 잎 및 과실에서 DWF5-1 또는 DWF5-2 유전자의 발현량을 q-PCR을 통해 확인한 결과를 나타낸 그래프이다. ****p<0.0001

발명을 실시하기 위한 구체적인 내용

[0012] 본 발명의 일 측면은 하기 화학식 1의 7-디하이드로콜레스테를(7-dehydrocholesterol)이 고농도로 함유된 형질전환된 토마토를 제공한다.

[0013] [화학식 1]



[0014]

[0015] 본 명세서에서 사용된 용어, "7-디하이드로콜레스테를(7-dehydrocholesterol)"은 비타민 D3로 전환되는 프로비타민 D3이다. 상기 형질전환된 토마토는 중량 100 g 당 0.01 mg 내지 1.5 mg의 7-디하이드로콜레스테를을 포함할 수 있다. 구체적으로, 형질전환된 토마토는 중량 100 g 당 0.10 mg, 0.20 mg, 0.30 mg, 0.40 mg, 0.50 mg, 0.60 mg, 0.70 mg, 0.80 mg, 0.90 mg, 1.00 mg, 1.10 mg, 1.20 mg, 1.30 mg, 1.40 mg 또는 1.50 mg의 7-디하이드로콜레스테를을 포함할 수 있다.

[0016] 본 명세서에서 사용된 용어, "토마토"는 학명이 Solanum lycopersicum으로서 가지목 가지과의 식물 또는 그의 과실(열매)을 일컫는다. 라틴아메리카가 원산지인 한해살이풀로, 키는 1 내지 3 m에 노란색 꽃이 핀다. 열매는 리코펜에 의해 붉은색을 띠며 식용 목적으로 이용된다. 본 명세서에서 토마토는 식물체일 수 있고, 과실(열매)일 수 있다.

[0017] 상기 형질전환된 토마토는 DWF5 유전자의 변이를 포함하는 것일 수 있다.

[0018] 본 명세서에서 사용된 용어, "변이"는 치환(conversion), 결실(deletion), 역위(inversion) 또는 삽입(insertion)에 의해 야생형과 상이한 것을 의미하는 것으로, 그 결과 목표 유전자의 기능이 상실된다. 상기 기능의 상실은 목표 유전자가 변형이 없는 경우에 통상적으로 수행하는 생물학적 기능 또는 역할의 전부 또는 일부가 없어진 것을 의미한다. 예를 들어, 목적 유전자에 의해 발현되는 단백질이 조기에 종결되거나, 단백질로서의 정상 기능을 상실한 것일 수 있다. 구체적으로, 상기 유전자 편집은 표적 부위에 종료 코돈을 생성시키거나, 야생형과 다른 아미노산을 코딩하는 코돈을 생성시킴으로써 목표 유전자를 녹아웃(knock-out) 시키는 것일 수 있다. 또한, 단백질을 생성하지 않는 비코딩 DNA 서열에 변이를 도입하거나 목표 유전자의 기능을 활성화시키는 등 다양한 형태일 수 있으나, 이에 제한되는 것은 아니다. 본 발명에서 변이는 아미노산 서열의 변이일 수 있으

며, 아미노산을 암호화하는 염기서열의 변이일 수 있다. 바람직하게는, 염기서열의 변이일 수 있다.

- [0019] 본 명세서에서 사용된 용어, "DWF5"는 7-디하이드로콜레스테롤 리덕타제(7-dehydrocholesterol reductase)를 의미한다. 상기 효소는 델타 5,7-스테롤 델타 7 리덕타제(delta 5,7-sterol delta 7 reductase)라 불리우기도 한다. 상기 효소는 5-디하이드로에피스테롤(5-dehydroepisterol)을 환원시켜 24-메틸렌콜레스테롤(24-methylencholesterol)로 전환시키는 활성을 가진다.
- [0020] 이때, 상기 DWF5는 서열번호 2 또는 서열번호 4의 아미노산 서열을 포함할 수 있다. 또한, 이를 코딩하는 핵산은 서열번호 1 또는 서열번호 3의 염기서열을 포함할 수 있다. 바람직하게는, 상기 DWF5는 본 명세서에서 DWF5-1로 표기되는 서열번호 2의 아미노산을 가질 수 있다. 또한, 상기 DWF5-1는 서열번호 1의 염기서열에 의해 암호화될 수 있다.
- [0021] 본 발명에서 서열번호 1의 염기서열을 포함하는 DWF5(DWF5-1) 유전자는 토마토에서 발현되며, 특히 뿌리 및 성숙된 열매에서 주로 발현된다(도 33). 서열번호 2의 염기서열을 포함하는 DWF5(DWF5-2)는 DWF5-1의 스플라이싱 변이체(splicing variant)로 주로 미성숙 열매에서 발현된다(Tomato eFP Browser 2.0 참조). 본 발명에서 DWF5-1 및/또는 DWF5-2 유전자 특이적으로 결손을 유도하였을 시, 상기 DWF5-1 유전자만을 결손시킨 변이 식물체는 프로비타민 D3가 농축된 토마토 열매를 생산할 수 있었으며, 생장 및 종자 생성 형태도 야생형 토마토와 유사하였다(도 19, 도 32 및 표 12). 반면, 상기 DWF5-1 및 DWF5-2 유전자를 동시에 결손시킨 유전자 변이 식물체의 경우 정상적으로 생장하지 못하였으며, 종자 생성이 불가능하였다. 또한, 상술한 바와 같이 과실에서 DWF5-2 유전자의 발현량이 DWF5-1에 비하여 현저히 낮기 때문에 DWF5-2 유전자만을 결손시킨 변이 식물체는 정상적인 성장을 보이지만 과실에 농축된 프로비타민 D3 함량은 DWF5-1 유전자 변이 식물체에 비하여 현저히 낮을 것으로 예상되었다. 따라서, 본 발명에서는 서열번호 1을 포함하는 DWF5(DWF5-1) 유전자 변이를 통한 프로비타민 D3가 고농도로 함유된 토마토를 생산하였다. 이하, 본 명세서에서 서열번호 1을 포함하는 DWF5 유전자는 DWF5-1 또는 DWF5 유전자로 혼용될 수 있다.
- [0022] 구체적으로 상기 DWF5 유전자의 변이는 엑손 6(서열번호 9), 엑손 7(서열번호 10), 엑손 9(서열번호 11), 엑손 10(서열번호 12), 엑손 11(서열번호 13) 및 이들의 조합으로 이루어진 군에서 선택되는 어느 하나의 도메인의 변이를 포함할 수 있다. 본 발명의 일 구체예에서 상기 DWF5 유전자의 변이는 엑손 6 도메인의 변이를 포함할 수 있다. 구체적으로, DWF5의 엑손 6 도메인을 암호화하는 서열번호 9의 염기서열에 치환, 결실, 역위 또는 삽입으로 변이가 생긴 것일 수 있다. 더욱 구체적으로, DWF5 유전자의 엑손 6 도메인을 암호화하는 서열번호 9의 염기서열이 치환, 결실, 역위 또는 삽입되어 DWF5의 활성이 낮아질 수 있으며 종료 코딩이 생성될 수 있다.
- [0023] 본 발명의 형질전환된 토마토는 CPD 유전자의 변이를 추가로 포함할 수 있다.
- [0024] 본 명세서에서 사용된 용어, "CPD(constitutive photomorphogenic DWAF)"는 시토크롬 P450 모노옥시게나제인 CP90A 계열의 효소로 DWF3로도 알려져 있다. 브라시놀리드 생합성(brassinolide biosynthesis) 과정 중, C6 산화 경로 작용하여 카테스테론(cathasteron)을 테스테론(teasterone)으로 전환시키며, 6-데옥소카스테론(6-deoxocathasterone)을 6-데옥소테스테론(6-deoxoteasterone)으로 전환시키는 역할을 한다. 본 발명에서 상기 효소는 7-디하이드로캄페스테롤(7-dehydrocampesterol)을 7-디하이드로캄페스타놀(7-dehydrocampestanol)로 환원시키는 역할을 한다.
- [0025] 이때, 상기 CPD는 서열번호 6의 아미노산 서열을 포함할 수 있다. 또한, 이를 코딩하는 핵산은 서열번호 5의 염기서열을 포함할 수 있다.
- [0026] 구체적으로 상기 CPD 유전자의 변이는 엑손 7(서열번호 14), 엑손 8(서열번호 15) 및 이들의 조합으로 이루어진 군에서 선택되는 어느 하나의 도메인의 변이를 포함할 수 있다. 본 발명의 일 구체예에서 상기 CPD 유전자의 변이는 엑손 8 도메인의 변이를 포함할 수 있다. 구체적으로, CPD 유전자의 변이는 엑손 8 도메인을 암호화하는 서열번호 15의 염기서열에 치환, 결실, 역위 또는 삽입으로 변이가 생긴 것일 수 있다. 더욱 구체적으로, CDP 유전자의 엑손 8 도메인을 암호화하는 서열번호 15의 염기서열이 치환, 결실, 역위 또는 삽입되어 CPD의 활성이 낮아질 수 있으며 종료 코딩이 생성될 수 있다.
- [0027] 본 발명의 DWF5 및 CPD 유전자의 변이를 포함하는 형질전환된 토마토는 SMT1 유전자의 변이를 추가로 포함할 수 있다.
- [0028] 본 명세서에서 사용된 용어, "SMT1(sterol methyltransferase 1)"은 스테로이드 골가지에 메틸기를 첨가시키는 효소이다. 구체적으로, 상기 SMT1은 5-디하이드로에피스테롤(5-Dehydroepisterol)을 7-디하이드로캄페스테롤(7-

Dehydrocampesterol)로 전환시키는 역할을 한다.

- [0029] 이때, 상기 SMT1은 서열번호 8의 아미노산 서열을 포함할 수 있다. 또한, 이를 코딩하는 핵산은 서열번호 7의 염기서열을 포함할 수 있다.
- [0030] 구체적으로 상기 SMT1 유전자의 변이는 엑손 7(서열번호 16), 엑손 9(서열번호 17), 엑손 11(서열번호 18), 엑손 12(서열번호 19) 및 이들의 조합으로 이루어진 군에서 선택되는 어느 하나의 도메인의 변이를 포함할 수 있다. 본 발명의 일 구체예에서 상기 SMT1 유전자의 변이는 엑손 7 및 엑손 9 도메인의 변이를 포함할 수 있다. 구체적으로, SMT1 유전자의 변이는 엑손 7 및 엑손 9 도메인을 암호화하는 서열번호 16 및 서열번호 17의 염기서열에 치환, 결실, 역위 또는 삽입으로 변이가 생긴 것일 수 있다. 더욱 구체적으로, SMT1 유전자의 엑손 7 및 엑손 9 도메인을 암호화하는 서열번호 16 및 서열번호 17의 염기서열이 치환, 결실, 역위 또는 삽입되어 SMT1의 활성이 낮아질 수 있으며, 종료 코딩이 생성될 수 있다.
- [0031] 상기 형질전환된 토마토는 동형접합체(homozygote)인 것일 수 있다. 즉, DWF5를 암호화하는 유전자 두 개에 변이를 포함할 수 있다. 구체적으로, DWF5 유전자의 엑손 6, 엑손 7, 엑손 9, 엑손 10, 엑손 11 및 이들의 조합으로 이루어진 군에서 선택되는 어느 하나의 도메인을 암호화하는 유전자 두 개에 변이를 포함할 수 있다. 본 발명의 일 구체예로 상기 형질전환된 토마토는 DWF5 유전자의 엑손 6 도메인을 암호화하는 유전자 두 개에 변이를 포함할 수 있다.
- [0032] 상기 형질전환된 토마토는 DWF5, CPD 및 SMT1를 암호화하는 유전자 두개에 변이를 포함할 수 있다. 이때, CPD 유전자의 엑손 7, 엑손 8 및 이들의 조합으로 이루어진 군에서 선택되는 어느 하나의 도메인을 암호화하는 유전자 두 개에 변이를 포함할 수 있다. 또한, SMT1 유전자의 엑손 7, 엑손 9, 엑손 11, 엑손 12 및 이들의 조합으로 이루어진 군에서 선택되는 어느 하나의 도메인을 암호화하는 유전자 두 개에 변이를 포함할 수 있다. 여기서 DWF5 유전자의 변이는 상술한 바와 동일하다.
- [0033] 본 발명의 일 구체예로 상기 형질전환된 토마토는 DWF5의 엑손 6 도메인, CPD의 엑손 7 도메인, 및 SMT1의 엑손 7 및 엑손 9 도메인을 암호화하는 각각의 유전자 두 개에 변이를 포함할 수 있다.
- [0034] 상기 형질전환된 토마토가 동형접합체로서 DWF5 유전자의 변이를 포함하는 경우, 상기 토마토는 종자를 형성할 수 있다. 상기 형질전환된 토마토가 동형접합체로 DWF5, CPD 및 SMT1 유전자의 변이를 포함하는 경우, 상기 토마토는 열매가 맺히지 않을 수 있으며, 종자가 형성되지 않을 수 있다.
- [0035] 본 발명의 다른 측면은, DWF5 유전자의 염기서열에 상보적으로 결합하는 하나 또는 하나 이상의 sgRNA 및 CRISPR 연관 단백질을 암호화하는 염기서열을 포함하는 벡터를 제공한다. 여기서, DWF5는 상술한 바와 동일하다.
- [0036] 본 명세서에서 사용된 용어, "가이드 RNA(guide RNA, gRNA)"는 유전체 편집을 통해 세포 내에서 표적 핵산을 인식하여 표적 핵산을 절단, 삽입 또는 연결시키는 폴리뉴클레오티드를 의미한다. 상기 gRNA는 표적 핵산 중 표적 서열에 상보적인 서열을 포함할 수 있다. 상기 gRNA는 상기 표적 핵산에서 PAM의 5' 방향 또는 3' 방향으로 연속적인 2 내지 24 뉴클레오티드(이하, 'nt'라 함)의 염기서열과 상보적인 폴리뉴클레오티드일 수 있다. 상기 gRNA의 길이는 17 nt, 18 nt, 19 nt, 20 nt, 21 nt, 22 nt, 23 nt 또는 24 nt일 수 있다.
- [0037] 상기 gRNA는 단일 가이드 RNA(single guide RNA, sgRNA)일 수 있다. 상기 sgRNA는 표적 핵산 서열에 특이적인 CRISPR RNA(crRNA) 및/또는 CRISPR 연관 단백질과 복합체를 형성하는 transactivating crRNA(tracrRNA)를 포함할 수 있다. 상기 sgRNA는 목표 유전자(표적 부위) 내의 표적 핵산 서열과 상보적인 염기서열(표적화 서열)을 가지는 부분(이를 spacer region, target DNA recognition sequence, base pairing region 등으로도 명명함) 및 CRISPR 연관 단백질 결합을 위한 헤어핀(hairpin) 구조를 포함할 수 있다. 보다 구체적으로, 목표 유전자 내의 표적 핵산 서열과 상보적인 염기서열(표적화 서열)을 포함하는 부분, CRISPR 연관 단백질 결합을 위한 hairpin 구조 및 종결(terminator) 서열을 포함할 수 있다. 상기 기술된 구조는 5'에서 3' 순으로 순차적으로 존재하는 것일 수 있으나, 이에 제한되는 것은 아니다. 상기 sgRNA가 crRNA 및 tracrRNA의 주요 부위 및 목표 유전자에 대한 상보적인 염기서열을 포함하는 경우라면 어떠한 형태의 sgRNA도 사용될 수 있다.
- [0038] 본 발명에서 상기 sgRNA는 DWF5 유전자의 엑손 6, 엑손 7, 엑손 9, 엑손 10 및 엑손 11로 이루어진 군에서 선택되는 어느 하나의 도메인의 염기서열에 상보적으로 결합하는 것일 수 있다. 상기 DWF5 유전자의 엑손 6의 도메인에 결합하는 sgRNA는 서열번호 20 또는 서열번호 21의 염기서열을 포함할 수 있다. 상기 엑손 7 도메인에 결합하는 sgRNA는 서열번호 22의 염기서열을 포함할 수 있다. 상기 엑손 9 도메인에 결합하는 sgRNA는 서열번호 23 또는 서열번호 24의 염기서열 포함할 수 있다. 상기 엑손 10 도메인에 결합하는 sgRNA는 서열번호 25의 염기

서열을 포함할 수 있다. 상기 엑손 11 도메인에 결합하는 sgRNA는 서열번호 26의 염기서열을 포함할 수 있다. 본 발명의 일 구체예에서 sgRNA는 엑손 6 도메인의 염기서열에 상보적으로 결합하는 서열번호 20 및 서열번호 21의 염기서열을 포함할 수 있다.

[0039] 본 발명에서 상기 벡터는 CPD 유전자의 염기서열에 상보적으로 결합하는 sgRNA를 추가로 포함할 수 있다. 상기 CPD 유전자에 상보적으로 결합하는 sgRNA는 CPD 유전자의 엑손 7 또는 엑손 8 도메인의 염기서열에 상보적으로 결합하는 것일 수 있다. 상기 CPD 유전자의 엑손 7의 도메인에 결합하는 sgRNA는 서열번호 27의 염기서열을 포함할 수 있다. 상기 CPD 유전자의 엑손 8 도메인에 결합하는 sgRNA는 서열번호 28, 서열번호 29, 서열번호 30 또는 서열번호 31의 염기서열을 포함할 수 있다. 본 발명의 일 구체예에서 sgRNA는 엑손 8 도메인의 염기서열에 상보적으로 결합하는 서열번호 28 및 서열번호 29의 염기서열을 포함할 수 있다.

[0040] 본 발명에서 상기 DWF5 및 CPD 유전자의 염기서열에 상보적으로 결합하는 sgRNA를 포함하는 벡터는 SMT1 유전자의 염기서열에 상보적으로 결합하는 sgRNA를 추가로 포함할 수 있다. 이때, SMT1 유전자의 염기서열에 상보적으로 결합하는 sgRNA는 SMT1 유전자의 엑손 7, 엑손 9, 엑손 11 및 엑손 12로 이루어진 군에서 선택되는 하나 또는 하나 이상의 도메인의 염기서열에 상보적으로 결합하는 것일 수 있다. 상기 SMT1 유전자의 엑손 7의 도메인에 결합하는 sgRNA는 서열번호 32의 염기서열을 포함할 수 있다. 상기 SMT1 유전자의 엑손 9 도메인에 결합하는 sgRNA는 서열번호 33의 염기서열을 포함할 수 있다. 상기 SMT1 유전자의 엑손 11의 도메인에 결합하는 sgRNA는 서열번호 34의 염기서열을 포함할 수 있다. 상기 SMT1 유전자의 엑손 12 도메인에 결합하는 sgRNA는 서열번호 35의 염기서열을 포함할 수 있다. 본 발명의 일 구체예에서 sgRNA는 엑손 7 및 엑손 9 도메인의 염기서열에 상보적으로 결합하는 서열번호 32 및 서열번호 33의 염기서열을 포함할 수 있다. 이때, DWF5 및 CPD 유전자의 염기서열에 상보적으로 결합하는 sgRNA는 상술한 바와 동일하다.

[0041] 본 명세서에서 사용된 용어, "CRISPR(clustered regularly interspaced palindromic repeats) 연관 단백질"은 CRISPR 연관 단백질은 DNA 또는 RNA와 같은 핵산이 이중 가닥 혹은 단일 가닥을 가질 경우(dsDNA/RNA 및 ssDNA/RNA) 이를 인식하여 절단할 수 있는 효소이다. 구체적으로, 이들은 sgRNA와 결합된 이중가닥 혹은 단일가닥 핵산을 인식하여 이를 절단할 수 있다. 상기 CRISPR 연관 단백질은 CRISPR 연관 단백질 및 그의 기능을 보유한 변이체를 포함한다.

[0042] 상기 CRISPR 연관 단백질은 스트렙토코커스 피오게네스(*Streptococcus pyogenes*), 스트렙토코커스 서모필루스(*Streptococcus thermophilus*), 스트렙토코커스속(*Streptococcus* sp.), 스태필로코커스 오레우스(*Staphylococcus aureus*), 노카르디옵시스 다손빌레이(*Nocardia opsis dassonvillei*), 스트렙토마이세스 프리스티나에 스피랄리스(*Streptomyces pristinae spiralis*), 스트렙토마이세스 비리도크로모게네스(*Streptomyces viridochromogenes*), 스트렙토스포란지움 로세움(*Streptosporangium roseum*), 알리사이클로바실러스 아시도칼다리우스(*Alicyclobacillus acidocaldarius*), 바실러스 슈도마이코이데스(*Bacillus pseudomycoides*), 바실러스 셀레니티레두센스(*Bacillus selenitireducens*), 엑시구오박테리움 시비리쿰(*Exiguobacterium sibiricum*), 락토바실러스 델부루에키(*Lactobacillus delbrueckii*), 락토바실러스 살리바리우스(*Lactobacillus salivarius*), 마이크로실라 마리나(*Microscilla marina*), 부르크홀데리아레스 박테리움(*Burkholderiales bacterium*), 폴라로모나스 나프탈레니보란스(*Polaromonas naphthalenivorans*), 폴라로모나스속(*Polaromonas* sp.), 크로코스파에라 와트소니(*Crocospaera watsonii*), 시아노테스속(*Cyanothece* sp.), 마이크로시스티스 에루기노사(*Microcystis aeruginosa*), 슈도모나스 에루기노사(*Pseudomonas aeruginosa*), 시네코코커스속(*Synechococcus* sp.), 아세토할로비움 아라바티쿰(*Acetohalobium arabaticum*), 암모니펙스 데젠시(*Ammonifex degensii*), 칼디셀룰로시럽토르 벅시(*Caldicelulosiruptor beccii*), 칸디다투스 데술포루디스(*Candidatus Desulforudis*), 클로스트리듐 보툴리눔(*Clostridium botulinum*), 클로스트리듐 디피실레(*Clostridium difficile*), 피네골디아 마그나(*Fingoldia magna*), 나트라나에로비우스 서모필루스(*Natranaerobius thermophilus*), 펠로토마쿨룸 서모프로피오니쿰(*Pelotomaculum thermopropionicum*), 아시디티오바실러스 칼두스(*Acidithiobacillus caldus*), 아시디티오바실러스 페로옥시단스(*Acidithiobacillus ferrooxidans*), 알로크로마티움 비노숨(*Allochromatium vinosum*), 마린노박터속(*Marinobacter* sp.), 니트로소코커스 할로필루스(*Nitrosococcus halophilus*), 니트로소코커스 와트소니(*Nitrosococcus watsonii*), 슈도알테로모나스 할로플란크티스(*Pseudoalteromonas haloplanktis*), 테도노박터 라세미퍼(*Ktedonobacter racemifer*), 메타노할로비움 에베스티가툼(*Methanohalobium evestigatum*), 아나바에나 바리아빌리스(*Anabaena variabilis*), 노둘라리아 스푼미게나(*Nodularia spumigena*), 노스톡속(*Nostoc* sp.), 아르트로스피라 막시마(*Arthrospira maxima*), 아르트로스피라 플라텐시스(*Arthrospira platensis*), 아르트로스피라속(*Arthrospira* sp.), 링비아속(*Lyngbya* sp.), 마이크로콜레우스 크토노플라스테스(*Microcoleus chthonoplastes*), 오실라토리아속(*Oscillatoria* sp.), 페트로토가 모빌리스(*Petrotoga mobilis*), 서모시포 아

프리카누스(*Thermosiphon africanus*), 아카리오클로리스 마리나(*Acaryochloris marina*), 렙토티리키아 샤히(*Leptotrichia shahii*), 프레보텔라(*Prevotella*) 또는 프란시셀라 노비시다(*Francisella novicida*)의 세균으로부터 유래된 것일 수 있다.

[0043] 본 발명에서 상기 CRISPR 연관 단백질은 Cas9, Cpf1, c2c1, C2c2, Cas13, c2c3, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5e(CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a, Cas8a1, Cas8a2, Cas8b, Cas8c, Csn1, Csx12, Cas10, Cas10d, Cas10, Cas10d, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1(CasA), Cse2(CasB), Cse3(CasE), Cse4(CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 및 Cul966으로 구성된 군으로부터 선택되는 어느 하나일 수 있다. 구체적으로 상기 CRISPR 연관 단백질은 Cas9일 수 있다.

[0044] 상기 Cas9 단백질은 가이드 sgRNA와 복합체를 형성하여 리보핵산 단백질(RNP)형태로 작용할 수 있다. 여기서 "리보핵단백질(ribonucleoprotein, RNP)"은 RNA와 단백질의 복합체로, 본 발명에서 Cas9-sgRNA 복합체일 수 있다.

[0045] 본 명세서에서 사용된 용어, "백터"는 숙주 세포에 도입되어 숙주 세포 유전체 내로 재조합 및 삽입될 수 있다. 또는 상기 백터는 에피솜으로서 자발적으로 복제될 수 있는 폴리뉴클레오티드 서열을 포함하는 핵산 수단으로 이해된다. 상기 백터는 선형 핵산, 플라스미드, 파지미드, 코스미드, RNA 백터, 바이러스 백터, 미니-염색체 및 이의 유사체들을 포함한다. 바이러스 백터의 예로는 레트로바이러스, 아데노바이러스 및 아데노-관련 바이러스를 포함하나 이에 제한되지 않는다.

[0046] 본 발명의 일 구체예에서 상기 백터는 서열번호 36 내지 서열번호 38의 어느 하나의 염기서열을 포함할 수 있다. 또한, 본 발명에서 상기 백터는 서열번호 36 내지 서열번호 38 중 어느 하나에 대해 약 70%, 약 75%, 약 80%, 약 85%, 약 90%, 약 91%, 약 92%, 약 93%, 약 94%, 약 95%, 약 96%, 약 97%, 약 98%, 약 99% 또는 약 100% 서열 상동성을 갖는다. 또한, 상기 백터는 표적 유기체 또는 세포에서의 발현을 위해 부분적으로 또는 완전히 코돈 최적화될 수 있다.

[0047] 상기 백터는 숙주세포에서 상기 폴리뉴클레오티드가 발현될 수 있도록 적절한 프로모터에 작동 가능하도록 연결된 것일 수 있다. 본 발명의 백터가 식물 세포에 적용되는 경우, 본 발명의 프로모터는 식물체의 유전자 도입을 위해 이용되는 프로모터를 포함할 수 있다. 예를 들어, SP6 프로모터, T7 프로모터, T3 프로모터, PM 프로모터, 옥수수의 유비퀴틴 프로모터(Ubi), 칼리플라워 모자이크 바이러스(CaMV) 35S 프로모터, 노팔린 신타아제(nos) 프로모터, 피그위트 모자이크 바이러스 35S 프로모터, 수가크레인 바실리폼 바이러스 프로모터, 콤멜리나 엘로우 모틀 바이러스 프로모터, 리볼로오스-1,5-비스-포스페이트 카르복실라아제 스몰 서브유닛(ssRUBISCO)의 광유도성 프로모터, 벼 사이토솔 트리오포스페이트 이소머라아제(TPI) 프로모터, 아라비도시스스의 아데닌 포스포리보실트랜스퍼라아제(APRT) 프로모터 및 옥토파인 신타아제 프로모터를 포함할 수 있으나, 이에 제한되지 않는다.

[0048] 구체적으로, 상기 백터는 플라스미드 DNA, 파아지 DNA 등이 될 수 있고, 상업적으로 개발된 플라스미드(pUC18, pBAD, pIDTSAMRT-AMP 등), 대장균 유래 플라스미드(pYG601BR322, pBR325, pUC118, pUC119 등), 바실러스 서브틸리스 유래 플라스미드(pUB110, pTP5 등), 효모-유래 플라스미드(YEp13, YEp24, YCp50 등), 파아지 DNA(Charon4A, Charon21A, EMBL3, EMBL4, λgt10, λgt11, λZAP 등), 동물 바이러스 백터(레트로바이러스(retrovirus), 아데노바이러스(adenovirus), 백신시아 바이러스(vaccinia virus) 등), 곤충 바이러스 백터(배칼로바이러스(baculovirus) 등) 등이 될 수 있다. 상기 백터는 숙주 세포에 따라서 단백질의 발현량과 수식 등이 다르게 나타나므로, 목적에 가장 적합한 숙주세포를 선택하여 사용함이 바람직하다.

[0049] 본 발명에서 상기 백터는 선택성 마커를 포함할 수 있다. 상기 선택성 마커는 통상적으로 화학적인 방법으로 선택될 수 있는 특성을 갖는 핵산 서열로, 형질전환된 세포를 비형질전환 세포로부터 구별할 수 있는 모든 유전자가 이에 해당된다. 예를 들어, 글리포세이트(glyphosate), 글루포시네이트암모늄(glufosinate ammonium) 또는 포스포노트리신(phosphinothricin)과 같은 제초제 저항성 유전자, 카나마이신(kanamycin), G418, 블레오마이신(bleomycin), 하이그로마이신(hygromycin), 클로람페니콜(chloramphenicol)과 같은 항생제 내성 유전자가 있으나, 이에 한정되는 것은 아니다. 본 발명의 일 구체예로 상기 선택성 마커는 하이그로마이신 및/또는 포스포노트리신일 수 있다.

[0050] 본 발명의 또 다른 측면은, 상기 sgRNA를 암호화하는 유전자 및 CRISPR 연관 단백질을 암호화하는 유전자를 포함하는 백터를 아그로박테리움을 이용하여 토마토에 도입하는 단계를 포함하는 7-디하이드로콜레스테롤이 고농

도로 함유된 형질전환된 토마토를 제조하는 방법을 제공한다.

[0051] 상기 형질전환 식물은 당업계에 공지된 식물의 형질전환 방법을 통해 제작될 수 있다. 당업자는 숙주로 선택한 식물의 특성을 고려하여 특정 식물에 적절한 공지의 형질전환 방법을 선택하여 실시할 수 있다. 구체적으로, 식물의 형질전환 방법으로 아그로박테리움을 이용한 형질전환 방법을 사용할 수 있다.

[0052] 상기 "아그로박테리움을 이용한 형질전환 방법"은 식물의 뿌리와 줄기에 종양을 일으키는 토양의 그람 음성 세균인 아그로박테리아를 이용하여 식물 세포에 외부 유전자를 전달하는 방법이다. 이는 아그로박테리움 투메파시엔스(*Agrobacterium tumefaciens*), 아그로박테리움 리조게네스(*Agrobacterium rhizogenes*) 등의 아그로박테리아에서 발견되는 종양 유발 플라스미드(Ti plasmid)의 T-DNA(transfer DNA)가 식물의 유전체(genome)에 삽입되는 현상을 이용한 방법이다.

[0053] 본 발명의 일 구체예에서 DWF5 유전자의 엑손 6 도메인에 변이를 포함하는 경우, 동형접합체를 가지는 토마토(T1)는 종자를 생성할 수 있다(도 19). 또한, DWF5 유전자의 엑손 6 도메인에 변이를 가진 동형접합체 T2세대 토마토도 정상적으로 성장하여 종자를 생성할 수 있으며(도 32), 7-디하이드로콜레스테롤이 고농도로 축적되었다(표 13 및 도 22 내지 도 28). 하지만, 형질전환된 토마토가 DWF5 유전자의 엑손 6 및 CPD 유전자의 엑손 8 도메인에 변이를 포함하는 경우, 상기 토마토는 온전한 식물체로 성장하지 못하였다. 또한, 형질전환된 토마토가 DWF5 유전자의 엑손 6, CPD 유전자의 엑손 8 및 SMT1 유전자의 엑손 7 및 엑손 9 도메인에 변이를 가진 동형접합체인 경우, 열매가 맺히지 않았다.

[0055] 이하 본 발명을 실시예를 통하여 보다 상세하게 설명한다. 그러나, 이들 실시예는 본 발명을 예시적으로 설명하기 위한 것으로 본 발명의 범위가 이들 실시예에 한정되는 것은 아니다.

[0056] 실시예 1. DWF5, CPD 및 SMT1 유전자 편집을 위한 Cas9-sgRNA 벡터의 제작

[0057] DWF5(도 1), CPD(도 2) 및 SMT1(도 3) 유전자 편집을 위한 sgRNA의 디자인 및 선정은 Cas-Designer site(<http://www.rgenome.net/cas-designdesigner/>)를 통하여 진행하였다. 타겟 유전자의 이름 또는 서열, 동물 또는 식물의 종을 선택한 후, 사용할 Cas9을 선택하여 유전자내 필요한 엑손(exon) 위치, sgRNA의 GC 함량, mismatch가 존재하는 off-target(비표적)등의 결과를 얻어 sgRNA를 디자인하였다. Cas9은 PAM이 NGG-3'이며 식물은 토마토(*S. lycopersicum*)를 선정하였다(도 1 내지 도 3). 또한, 유전자 편집 효율을 확인하기 위하여 sgRNA를 스크리닝하였다(표 2 및 도 1 내지 도 3).

[0058] 구체적으로, sgRNA의 표적화 적절성은 RGEN(RNA-guided endonuclease) 분석을 이용하여 진행하였다. DNA template는 상기의 증폭된 PCR 생성물을 PCR Purification 키트를 이용하여 정제한 후(PCR product), DNA template의 양을 측정하여 표 1과 같은 조합으로 RGEN을 수행하였다. 37℃에서 1시간 배양한 다음, 6x DNA-Purple dye를 3 µL씩 넣어 열린 후 2.5% Gel에 로딩하였다. DNA template에 잘리지 않은 밴드의 정도를 분석하여 유전자 편집의 효율을 분석하였다.

[0059] [표 1]

RGEN 최종 농도 조건	
10X 3.1 NEB buffer	2 µL
sgRNA(120 nM)	5 µL
spyCas9(100 nM)	5 µL
DNA template(PCR product)	200 ng
DW up to 20 µL, 37℃ 1h	

[0060]

[0061] 그 결과, DWF5 유전자를 표적으로 제작된 sgRNA의 경우, D5-4 및 D5-6을 제외한 모든 sgRNA가 DWF5 유전자를 표적할 수 있음을 확인하였다(도 4). CPD sgRNA는 C1을 제외한 모든 sgRNA가 CPD 유전자를 표적화하였으며(도 5), SMT1 sgRNA는 제작된 모든 sgRNA가 CPD 유전자를 표적화 하였다(도 6).

[0062] [표 2]

Target Gene (Cas9)	sgRNA (primer name)	DNA cleavage assay	sgRNA A (sub name)	sgRNA sequence (PAM site 3')	유전자 내 위치 (exon)
SolycDWF5 (서열번호 1)	SolycDWF5-1	O	D5-1	TGGAGCTTTATCCTCGT ATCGGG(서열번호 20)	6 (서열번호 9)
	SolycDWF5-2	O	D5-2	GCAGATTTGGTATGATG TCTTGG(서열번호 21)	6 (서열번호 9)
	SolycDWF5-3	O	D5-3	TGTAAATACGATATTGA TGTTGG(서열번호 22)	7 (서열번호 10)
	SolycDWF5-4	△	D5-4	GACCAGCCACTAGTATG TAAAGG(서열번호 23)	9 (서열번호 11)
	SolycDWF5-5	O	D5-5	TGCATTTTCCATTTGTT CTCCGG(서열번호 24)	9 (서열번호 11)
	SolycDWF5-6	X	D5-6	CCTCTGGTGAGAAAAAG ACCAGC(서열번호 25)	10 (서열번호 12)
	SolycDWF5-7	O	D5-7	AGAAAGATGACGTAGAA GTAGGG(서열번호 26)	11 (서열번호 13)
SolycCPD (서열번호 5)	SolycCPD gRNA1	△	C1	AACGAGTCACCAAGTGG TGAAGG(서열번호 27)	7 (서열번호 14)
	SolycCPD gRNA2	O	C2	CGCTTCAGCATCCTTGT CGTTGG(서열번호 28)	8 (서열번호 15)
	SolycCPD gRNA3	O	C3	TGAGCAGAAAGAAGAAG AGAAGG(서열번호 29)	8 (서열번호 15)
	SolycCPD gRNA4	O	C4	TGAGATTGTAGAATGCA GAGTGG(서열번호 30)	8 (서열번호 15)
	SolycCPD gRNA5	O	C5	GTAGAATGCAGAGTGGA GGGAGG(서열번호 31)	8 (서열번호 15)
SolycSMT1 (서열번호 7)	SolycSMT1 gRNA1	O	S1	GTTGAACCGCAAAGTGG GATTGG(서열번호 32)	7 (서열번호 16)
	SolycSMT1 gRNA2	O	S2	GATGCTATAGAGAGATC TATAGG(서열번호 33)	9 (서열번호 17)
	SolycSMT1 gRNA4	O	S4	GTTTCGGCACTTGAATA TGTGGG(서열번호 34)	11 (서열번호 18)
	SolycSMT1 gRNA5	O	S5	GCCGCAGAAAGTCTTGT TGGTGG(서열번호 35)	12 (서열번호 19)

[0063]

[0064]

상기 sgRNA 스크리닝 결과를 바탕으로 최종적으로 DWF5, CPD 또는 SMT1를 표적으로 제작된 sgRNA를 포함하는 Cas9-sgRNA 벡터를 제작하였다.

[0065]

카나마이신과 하이그로마이신 저항성 유전자를 가진 pCambia1300 플라스미드를 기반으로 하여 pNGPJ0014 벡터 (서열번호 39)를 제작하였다.

[0066]

pNGPJ0014 벡터는 카나마이신 및 하이그로마이신 항생제에 의해 선별되도록 항생제 카세트를 가지며, Cas9과 합성된 polycistronic tRNA-gRNA 카세트를 갖는다. 상기 Cas9을 애기장대 Ubiquitin 10(Ubi 10) 프로모터에 의해 발현되도록 제작하였으며, 상기 tRNA-gRNA는 애기장대 Ubiquitin 6 프로모터에 의해 발현되도록 제작하였다.

[0067]

Cas9 단백질을 핵으로 전달하기 위해 Cas9 오픈리딩프레임의 N-말단에 SV40(PKKRKRV) 핵이행신호 서열과 C-말단에 Bipartite(KEPAATKKAGQAKKKK) 핵이행신호 서열을 추가하였다. 이때, Cas9-sgRNA 벡터는 1 내지 3 종류의 유전자를 표적화 할 수 있게 제작하였다. 또한, 하나의 표적당 2 종류의 sgRNA를 적재할 수 있도록 제작하였다. DWF5 유전자를 표적으로 하는 D5-1 sgRNA(서열번호 20) 및 D5-2 sgRNA(서열번호 21)의 염기서열을 포함하는 Cas-sgRNA 벡터를 "D100"(서열번호 36)이라고 명명하였다(도 7).

[0068]

CPD 및 DWF5 유전자를 표적화하는 Cas9-sgRNA 벡터를 "D120"(서열번호 37)이라고 명명하였다. 이때, CPD sgRNA는 C2(서열번호 28) 및 C3(서열번호 29)의 염기서열을 포함하며, DWF5 sgRNA는 D5-1(서열번호 20) 및 D5-2(서열번호 21)의 염기서열을 포함한다(도 8).

[0069]

SMT1, CPD 및 DWF5 유전자를 표적화 하는 Cas9-sgRNA 벡터를 "D121"(서열번호 38)이라고 명명하였다. 이때, SMT1 sgRNA는 S1(서열번호 32) 및 S2(서열번호 33)의 염기서열을 포함하며, CPD sgRNA는 C2(서열번호 28) 및

C3(서열번호 29)의 염기서열을 포함하고, DWF5 sgRNA는 D5-1(서열번호 20) 및 D5-2(서열번호 21)의 염기서열을 포함한다(도 9).

[0070] 실시예 2. DWF5, CPD 및 SMT1 유전자 편집을 통한 토마토의 형질 전환

[0071] 토마토의 형질 전환은 도 10의 모식도에 기재된 바와 같이 진행하였다. 각각의 단계별 사용된 배지 및 호르몬은 표 3 내지 표 8에 기재된 바와 동일하게 제조하여 사용하였다.

[표 3]

TMS(V)배지	MS SALT	4.3 g/L	4 종류의 비타민은 100배 농축액(stock solution)으로 제조한 후 첨가하였다. ; 100배(100X) 비타민 농축액은 10 mL/L씩 사용하였다.
	Myo-inositol	100 mg/L	
	Nicotinic acid	0.5 mg/L	
	Pyridoxine HCl	0.5 mg/L	
	Thiamin HCl	0.1 mg/L	
	Sucrose	30 g/L	
	Plant Agar	0.6~0.7%/ L	
	배지 조제 후 pH를 1 N NaOH을 이용하여 5.7로 교정한 후, autoclave로 121℃, 15분 동안 멸균하여, 90x15 mm petridish에 약 30 mL씩 분주하였다. 비타민 농축액은 상기 배지를 멸균한 후, 배지 온도가 55~60℃ 사이일 때 첨가하였다.		

[표 4]

TPC배지 (Tomato Pre-Culture Medium)	MS SALT	4.3 g/L	4 종류의 비타민은 100배 농축액(stock solution)으로 제조한 후 첨가하였다. ; 100배(100X) 비타민 농축액은 10 mL/L씩 사용하였다.
	Myo-inositol	100 mg/L	
	Nicotinic acid	0.5 mg/L	
	Pyridoxine HCl	0.5 mg/L	
	Thiamin HCl	0.1 mg/L	
	Sucrose	30 g/L	
	Gelrite	0.25%/L	
	BA	1 mg/L	
	NAA	0.1 mg/L	
	배지 조제 후 pH를 1 N NaOH을 이용하여 5.7로 교정한 후, autoclave로 121℃, 15분 동안 멸균하여, 90x15 mm petridish에 약 30 mL씩 분주하였다. 비타민 농축액은 상기 배지를 멸균한 후, 배지 온도가 55~60℃ 사이일 때 첨가하였다.		

[표 5]

TCC 배지 (T o m a t o Co-Culture Medium)	MS SALT	4.3 g/L	4 종류의 비타민은 100배 농축액(stock solution)으로 제조한 후 첨가하였다. ; 100배(100X) 비타민 농축액은 10 mL/L씩 사용하였다.
	Myo-inositol	100 mg/L	
	Nicotinic acid	0.5 mg/L	
	Pyridoxine HCl	0.5 mg/L	
	Thiamin HCl	0.1 mg/L	
	Sucrose	30 g/L	
	Gelrite	0.25%/L	
	BA	1 mg/L	
	NAA	0.1 mg/L	
	Acetosyringone(AS)	1 0 0 mM/L	
배지 조제 후 pH를 1 N NaOH을 이용하여 5.2로 교정한 후, autoclave로 121℃, 15분 동안 멸균하여, 90x15 mm petridish에 약 30 mL씩 분주하였다. 비타민 농축액, BA, NA 및 AS는 상기 배지를 멸균한 후, 배지 온도가 55~60℃ 사이일 때 첨가하였다.			

[표 6]

균 희석 배지	MS SALT	4.3 g/L	pH를 1 N NaOH 및 1 N HCl로 5.7로 교정한 후, autoclave로 121℃, 15분 동안 멸균하여 냉장 보관하였다.
	Sucrose	30 g/L	

[0078] [표 7]

TSI 배지 (Tomato Shoot Induction Medium)	MS SALT	4.3 g/L	4 종류의 비타민은 100배 농축액(stock solution)으로 제조한 후 첨가하였다. : 100배(100X) 비타민 농축액은 10 mL/L씩 사용하였다.
	Myo-inositol	100 mg/L	
	Nicotinic acid	0.5 mg/L	
	Pyridoxine HCl	0.5 mg/L	
	Thiamin HCl	0.1 mg/L	
	Sucrose	30 g/L	
	Gelrite	0.25%	
	IAA	0.2 mg/L	
	Zeatin	2 mg/L	
	PPT	5 mg/L	
		carbenicillin	5 mg/L
배지 조제 후 pH를 1 N NaOH 및 1 N HCl로 5.7로 교정한 후, autoclave로 121℃, 15분 동안 멸균하여, 90x15 mm petridish에 약 30 mL 분주하였다. 비타민 농축액, BA, NA 및 AS는 상기 배지를 멸균한 후, 배지 온도가 55~60℃ 사이일 때 첨가하였다.			

[0079]

[0080] [표 8]

Rooting 배지	1/2 MS SALT(비타민 포함)	2.2 g/L
	Sucrose	30 g/L
	Plant Agar	0.6~0.7%/L
	IAA	0.2 mg/L
	Carbenicillin	300 mg/L
배지 조제 후 pH를 1 N NaOH 및 1 N HCl로 5.7로 교정한 후, autoclave로 121℃, 15분 멸균하여, 마젠타 박스에 분주하였다. IAA는 멸균 후, 배지 온도가 55~60℃ 사이에 첨가하였다.		

[0081]

[0082] [표 9]

호르몬 및 비타민 농축액 조제		
	농도	제품명 또는 제조방법
BA	1 mg/mL	Calsson(BSL01-100mL) 1 mg/mL solution을 구매하여 사용.
NAA	1 mg/mL	Calsson(NSL01-100mL) 1 mg/mL solution을 구매하여 사용.
Zeatin	2 mg/mL	phytottechlab(Z860) 1 mg/mL solution 구매하여 사용.
IAA(Indole acetic acid)	1 mg/mL	Duchefa, I0901 시약 50 mL을 D.W 50 mL에 녹여 필터한 후 분주하여 냉동 보관함.
Acetosyringone(AS)	1 0 0 mM/mL	Acetosyringone 0.5887 g을 DMSO 20 mL에 녹이고 10 mL D.W를 추가하여 총 30 mL로 제조함. 제조된 용액은 1 mL씩 분주하여 냉동 보관함(얼지 않음).
Hygromycin	1 0 mg/mL	바이오테크(CAT.No.BM463H) 10 mg/mL solution을 구매하여 사용.
carbenicillin	3 0 0 mg/mL	Duchefa(C0109) 시약 1.5 g을 D.W 50 mL에 녹여 필터한 후 분주하여 냉동 보관함.
Modified Vitamin solution	100x	100배 농축액로 제조하여 배지 조제시 사용함. Myo-inositol(MB cell, MB-I4715) 10 g/L; Nicotinic Acid(Showa, 1414-0130) 50 mg/L; Pyridoxine HCl(Sigma, P-8666) 50 mg/L; 및 Thiamine HCl(Duchefa, T0614) 10 mg/L;을 1 L의 D.W에 녹여 autoclave로 121℃, 15분 동안 멸균한 후 냉장 보관함.

[0083]

[0084] 구체적으로, 토마토 종자를 70% 에탄올로 5 초간 소독하고, 2% 락스로 20분간 소독하였다. 소독한 종자는 4~5회 멸균수로 씻은 후, 멸균된 여과지에 올려 수분을 제거하였다. 소독한 토마토 종자를 종자 파종 배지 TMS(표 3)에 옮겨 25℃에서 3일 동안 암배양한 후, 명배양을 배양하였다.

- [0085] 과종 후 떡잎(cotyledone)이 1~1.5 cm 자랐을 때(9~12일) 토마토의 하배축(hypocotyl) 위쪽을 0.7~1.0 cm로 잘라 TPC 배지 plate(표 4)에 약 18 내지 20개(4, 5, 5 및 4개 또는 5, 5, 5 및 5개씩 총 18~20개) 치상하고 명배양(16L/8D)으로 1일 배양하였다.
- [0086] 아그로박테리아(Agrobacterium, GV3101 strain, DSMZ, CAT.No.DSM12364)는 50 mL 균배양 배지(YEP)에 100 mM AS(acetosyringone) 및 50 mg/L rifampicin, 50 mg/L kanamycin(Km)을 넣고, OD600 값이 0.8~1.2 정도 되도록 28℃에서 1~2일 동안 암배양하였다.
- [0087] 배양된 균을 4℃, 6,000 rpm으로 15분 동안 원심분리하여 상층액을 제거하고, pellet은 동량의 Tomato Co-Culture(TCC) 배지(표 5)에 재현탁하고 OD값이 0.5 가 되도록 희석하였다.
- [0088] 50 mL 튜브에 상기 방법으로 준비된 40 mL 균 배양액(OD600 값이 0.5) 및 명배양에서 전처리 한 식물체를 넣고 (과종 200립 기준), 25℃ 상온에서 20~30분간 배양하였다. 공동배양된 절편을 멸균된 여과지에 옮겨 건조시켜 수분을 가능한 제거한 후, TCC 배지 plate(표 5)에 약 18 내지 20개(4, 5, 5 및 4개 또는 5, 5, 5 및 5개씩 총 18~20개) 치상하여 2일 동안 25℃에서 명배양(16 L/8D)하였다. 공동배양 2 일 후 절편체를 신초 선발 배지(TSI, 표 7)로 옮긴 후, 식물 배양기 내에서 25℃로 명배양(16 L/8D) 하였다. 이때, 아그로박테리아가 자라 있는 경우, 공동배양 액체 배지에 300 mg/L carbenicillin을 넣어 세척 후, 필터페이퍼에 건조하여 다시 배지에 치상하였다.
- [0089] 치상 후 3~4주 사이에 잘려진 단면에서 캘러스가 분화되고, 분화된 캘러스에서 신초가 유도되어 1 cm 이상 자랐을 때, 뿌리를 유도시키기 위하여 Rooting 배지(표 9)로 옮겼다. 이때, 캘러스는 가능한 제거하고 신초 부분을 이식하여 발근을 유도하였다. 뿌리 발근이 충분히 이루어졌을 때, 수돗물로 개체를 세척하여 개체에 남아있는 아그로박테리아를 제거하고 흙으로 옮겨 배양실 환경을 유지하면서 외부 환경(습도변화 및 비멸균상태)에 적응시키면서 배양하였다.
- [0090] 상기 캘러스에서 식물체 분화 단계에 따른 개체 변화는 도 11에 기재하였다.
- [0091] 그 결과, 도 12 내지 도 14에 나타난 바와 같이, D100(도 12) 또는 D121(도 14)로 형질전환 시킨 토마토 신초의 경우 정상적으로 식물체로 성장하였으나, D120으로 형질전환 시킨 토마토 절편은 정상적인 성장하지 않았다(도 13).
- [0092] **실시예 3. DWF5, CPD 및 SMT1 유전자 편집으로 생성된 토마토 식물체의 염기서열 분석**
- [0093] 실시예 2에서 수득된 토마토 식물체의 형질전환 여부를 확인하기 위하여 상기 D100 또는 D121로 형질전환시킨 토마토 개체에 대해 T7E1 시험(도 15) 및 생어 염기서열 분석(sanger sequencing)(도 16 및 도 17)을 수행하였다.
- [0094] 구체적으로, T7E1 시험은 먼저 타겟 위치를 PCR 통해 증폭한 후 PCR Purification 키트를 이용하여 증폭된 PCR 결과물을 정제하였다. 상기 정제된 PCR 결과물을 DNA template로 사용하여 표 10과 같은 조합과 표 11의 조건으로 Heteroduplex formation을 수행하였다. 그리고, T7 endonuclease I(T7E1)를 1 μ L 첨가하여 37℃에서 30분간 반응시켰다. 마지막으로 Proteinase K 1 μ L를 처리하고 37℃에서 5분간 반응시켜 T7E1 효소 활성을 비활성화 하였다. PCR Product에 Fragment 진하기 정도로 유전자 편집의 효율을 평가하였다.

[0095] [표 10]

PCR Product (250 ng)	5 μ L
10x NEB Buffer 2	2 μ L
Nuclease-free water	Up to 14 μ L

[0096]

[0097] [표 11]

온도	Ramp rate
95	4.0/sec
95-85	-2.0/sec
85-25	-0.1/sec
25-4	4.0/sec

[0098]

[0099] 생어 염기서열 분석은 하기와 같이 수행하였다. sgRNA 표적 영역을 20 μ L 반응에서 Q5 High-Fidelity DNA Polymerase(NewEngland Biolabs)를 사용하여 게놈 DNA 추출물로부터 증폭하였다. 이후, PCR 생성물을 All in one Cloning Kit(Biofact, South Korea)를 사용하여 TA 벡터에 클로닝하고, 클로닝된 15 내지 20개의 클론을 각각의 샘플에 대해 개별적으로 시퀀싱하였다.

[0100] 그 결과, D100으로 형질전환되어 생산된 DWF5 유전자 결실 토마토 동형접합체는 14개체로 확인되었다(도 15 및 도 16). 또한, D121로 형질전환 되어 DWF5, CPD 및 SMT1 유전자 결실 토마토 동형접합체는 1개체로 확인되었다(도 15 및 도 17). D121 형질전환 식물체의 경우, 유전자 결실로 인한 식물 호르몬 불균형이 초래되어 열매가 맺히지 않는 현상이 관찰되었다.

[0101] **실시예 4. DWF5 유전자 편집으로 생성된 토마토의 종자 획득 및 유전자 표현형 분석**

[0102] 유전자 가위를 통해 DWF5 유전자 결실 토마토(T0 세대) 중 7개체에서 확보된 종자(DWF5 homozygote)를 파종하여 다음 세대(T1)의 DWF5 유전자 결실 토마토 수득 여부를 차세대 염기분석(next generation sequencing, NGS)을 통해 확인하였다.

[0103] 그 결과, T0 7 개체에서 수득된 종자(T1) 23개체 중 12개체가 동형접합체(homozygote)임을 확인할 수 있었다(도 18). 또한, 상기 12개체 중 유전자 편집 효율이(indel 효율)이 100%인 6개체를 선별하고 종자(T2) 발생 여부 및 수득된 종자량을 확인하였다. 그 결과, 총 6개체 중 5개체에서 종자가 생산되었다(도 19 및 표 12). 또한, 야생형과 비교하여도 DWF5 유전자 결실 토마토의 과실, 잎 및 가지 줄기의 표현형에서 큰 차이를 관찰할 수 없었다. 이때, 야생형 토마토(Wild type, WT)는 대조군이다.

[0104] [표 12]

Label	Total reads	sgRNA 1			sgRNA 2			T1 Zygote	T2 종자량(g)	
		Insertions	Deletions	Indel freq. (%)	Insertions	Deletions	Indel freq. (%)			
D100-1-4	54192	53220	51	100	0	52983	100	Mono-allelic	0.37	1.06
D100-7-1	89757	89183	54	100	0	88705	100	Mono-allelic	0.35	
D100-12-8	7532	7377	13	100	0	7267	100	Mono-allelic	0.34	
D100-3-11	54964	0	54465	100	0	53656	100	Mono-allelic	0.13	0.45
D100-3-14	44027	0	42936	100	0	42304	100	Mono-allelic	0.32	
D100-4-8	83297	0	82686	100	0	81960	100	Mono-allelic	—	

[0105]

[0106] **실시예 5. T2 세대의 표현형 분석**

[0107] **실시예 5.1. 비타민 D 전구체 분석**

[0108] 실시예 4에서 수득한 종자(T2)를 파종하여 T2세대 토마토에 함유되어 있는 프로비타민 D3의 함량을 LC/MS로 측정하였다. 이때, 시료로 토마토 야생형(WT), #3-14 라인(D100-3-14)의 T2 세대(4개체) 및 #7-1 라인(D100-7-1)의 T2 세대(11개체)를 사용하였고, 표준 시약으로 동물 프로비타민 D3 전구체인 7-디하이드로콜레스테롤(7-dehydrocholesterol)(provitamin D3, sigma, Cat30800, Lot.BCBS6021)을 사용하였다. 또한, 대조군으로 야생형(WT) 토마토를 시료로 사용하였다.

[0109] 구체적으로, 각 시료의 일정량을 칭량하여 30 mL의 에탄올 및 1 mL의 10% 에탄올성 피로갈로를 첨가하였다. 그 뒤, 3 mL의 90% 수산화칼륨(KOH)을 첨가하여 80℃ 수욕조에 10분마다 vortexing하면서 30분간 반응시켰다. 반응이 끝난 뒤, 실온에서 냉각시킨 후 30 mL의 증류수 및 n-헥산(n-hexane)을 각각 첨가하고 1시간 진탕하여 혼합하였다. 그리고 상층액을 분액여두에 옮겨 물층에 30 mL의 n-헥산을 첨가하여 vortexing하여 혼합하였다. 다시 상층액을 분액여두에 옮겨 헥산층에 100 mL의 증류수를 첨가하여 수세하였다. 상기 수세 과정은 2회 반복하였다. 그 뒤, 헥산층을 무수황산나트륨으로 탈수한 뒤 농축하고, 5 mL 메탄올(MeOH)에 녹여 폴리테트라플루오로에틸렌(Polytetrafluoroethylene, PTFE) 필터로 여과하여 시험 용액을 준비하였다(도 20).

[0110] 상기 과정으로 준비된 시험 용액에 함유된 7-디하이드로콜레스테롤을 도 21의 조건으로 LC/MS를 통해 측정하였다.

[0111] 그 결과, WT 토마토의 과실부위에서는 7-디하이드로콜레스테롤이 검출되지 않은 반면, T2세대의 #3-14 라인에서 4개체, #7-1 라인에서 11개체(총 15개체)에서 7-디하이드로콜레스테롤이 검출되었다(표 13 및 도 22 내지 도 28).

[0112] [표 13]

No	시료명	시료상태	최종부피(L)	표준품순도 (%)	결과값 (mg/100 g)
1	WT 토마토	성숙 과실	0.005	95.3	불검출
2	WT 토마토	성숙 과실	0.005	95.3	불검출
3	#3-14-2 토마토	성숙 과실	0.005	95.3	1.03
4	#3-14-13 토마토	성숙 과실	0.005	95.3	0.60
5	#3-14-20 토마토	성숙 과실	0.005	95.3	0.19
6	#3-14-22 토마토	성숙 과실	0.005	95.3	0.26
7	#7-1-2 토마토	성숙 과실	0.005	95.3	0.23
8	#7-1-7 토마토	성숙 과실	0.005	95.3	0.12
9	#7-1-12 토마토	미성숙 과실 포함	0.005	95.3	0.27
10	#7-1-14 토마토	익줄기	0.005	95.3	0.02
11	#7-1-17 토마토	성숙 과실	0.005	95.3	0.14
12	#7-1-18 토마토	익줄기	0.005	95.3	0.03
13	#7-1-20 토마토	미성숙 과실	0.005	95.3	0.36
14	#7-1-21 토마토	익줄기	0.005	95.3	불검출
15	#7-1-22 토마토	익줄기	0.005	95.3	불검출
16	#7-1-24 토마토	미성숙 과실	0.005	95.3	0.71
17	#7-1-27 토마토	익줄기	0.005	95.3	0.03

[0113]

[0114] 실시예 5.2. T2 세대의 항생제 free 개체 선별

[0115] 실시예 4에서 수득한 T2세대의 #3-14 라인(D100-3-14) 및 #7-1 라인(D100-7-1) 중 항생제 내성이 없는 개체를 선별하기 위하여, 상기 개체의 mRNA 내의 항생제 내성 유전자의 발현을 RT-PCR을 통해 확인하였다.

[0116] 그 결과, #3-14 라인(D100-3-14)의 5개체 및 #7-1 라인(D100-7-1)의 11개체가 항생제 내성이 없는 개체임을 확인할 수 있었다(도 29).

[0117] 상기 항생제 내성이 없는 개체로 선별된 개체 중, #3-14-29 및 #7-1-15 개체에 대하여 DWF5 유전자 편집 여부를 확인하였다.

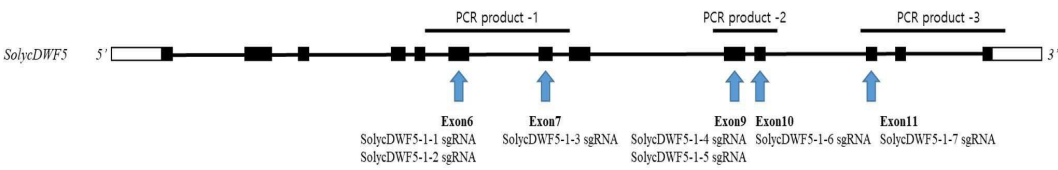
[0118] 그 결과, 도 30 및 도 31에 나타난 바와 같이 상기 개체의 염기서열은 야생형(WT) 토마토의 염기서열과 상이하며(도 30), 아미노산 서열도 엑손 6에 해당되는 서열이 결실되어 있는 것을 확인할 수 있었다(도 31).

[0119] 실시예 5.3. T2 세대의 표현형 및 종자 생성 확인

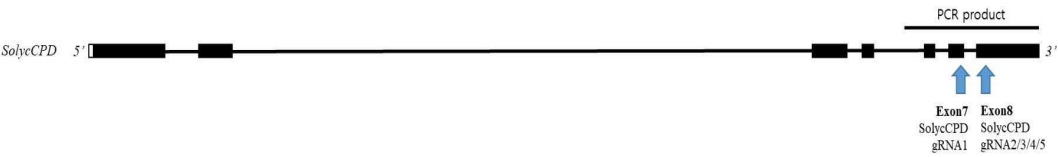
[0120] 실시예 4에서 수득한 T2세대의 #3-14 라인(D100-3-14) 및 #7-1 라인(D100-7-1)의 표현형 및 종자 생성 여부를 확인하였다. 구체적으로, 과실에서 주로 발현되는 DWF5(서열번호 1, DWF5-1) 유전자(도 33)만 결손시켰을 경우, 야생형과 비교하여 과실, 잎 및 가지줄기의 표현형에서 큰 차이가 관찰되지 않았다. 하지만 DWF5-1(서열번호 1) 및 DWF5-2(서열번호 3)을 모두 결손시켰을 경우, 토마토의 생장이 저해되고 종자가 생성되지 않았다. 또한, 상술한 바와 같이 과실에서 DWF5-2 유전자의 발현량이 DWF5-1에 비하여 현저히 낮기 때문에(도 33), DWF5-2 유전자만을 결손시킨 변이 식물체는 정상적인 성장을 보이지만 과실에 농축된 프로비타민 D3 함량은 DWF5-1 유전자 변이 식물체에 비하여 현저히 낮을 것으로 예상되었다. 따라서, 본 발명에서는 DWF5-1(서열번호 1) 유전자만 결손시킴으로써 프로비타민 D3의 농축을 유도하면서 종자 생성의 문제점을 보완할 수 있는 토마토를 생산할 수 있었다.

도면

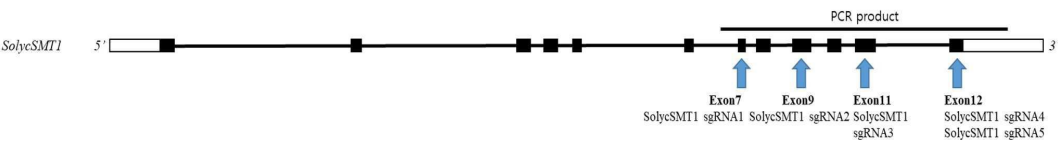
도면1



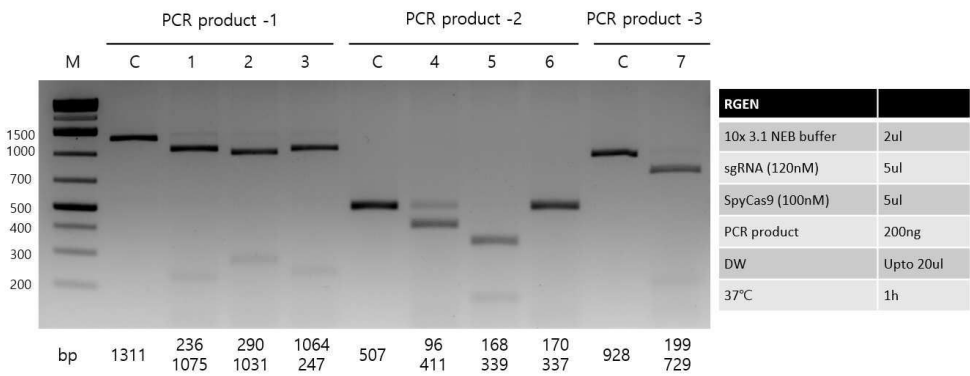
도면2



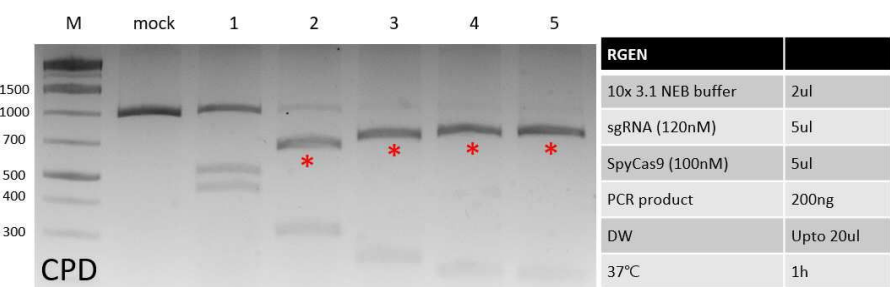
도면3



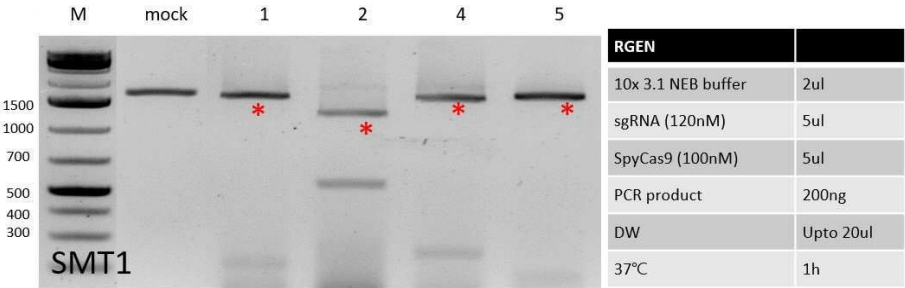
도면4



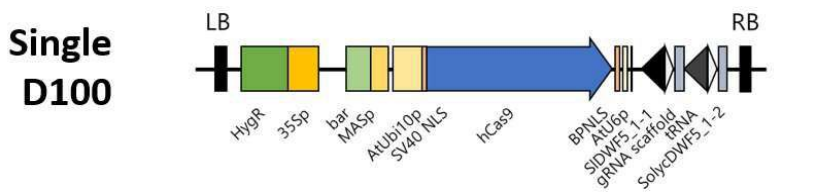
도면5



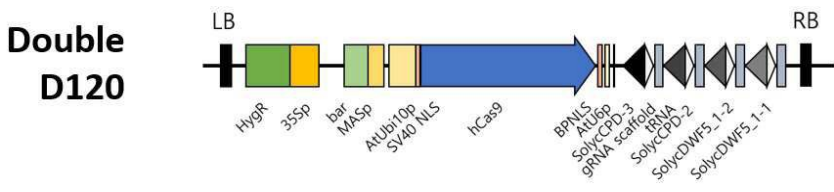
도면6



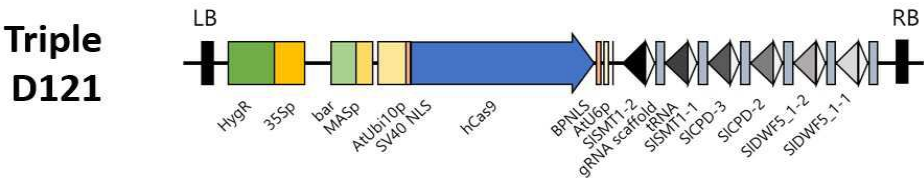
도면7



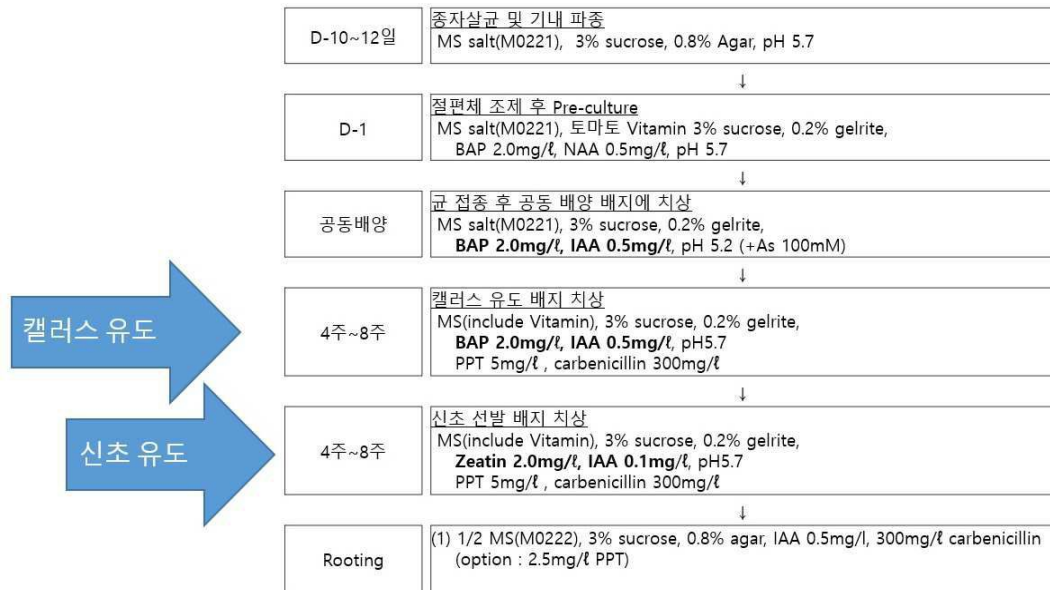
도면8



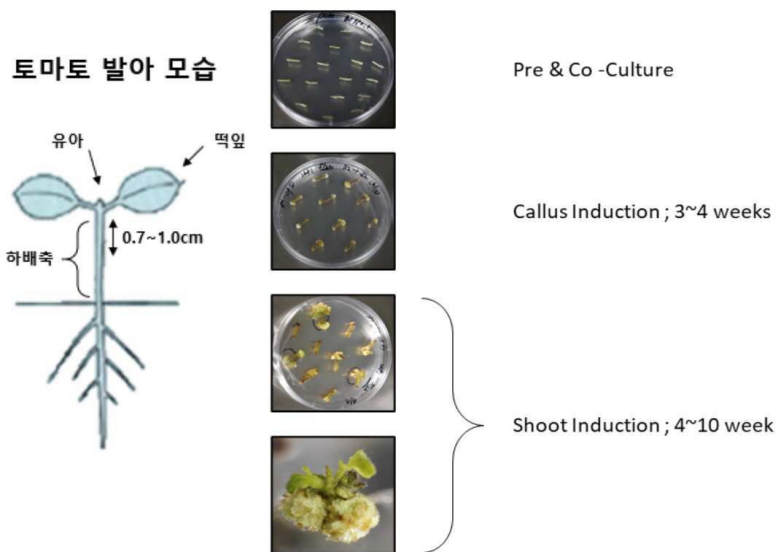
도면9



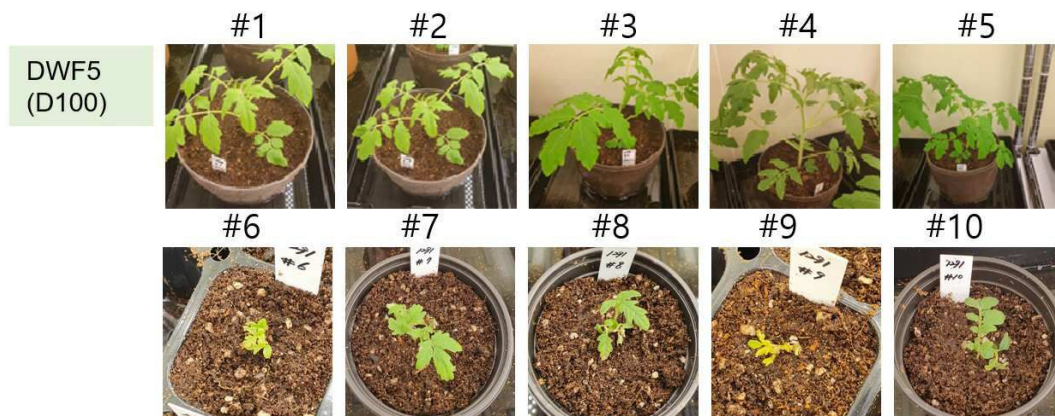
도면10



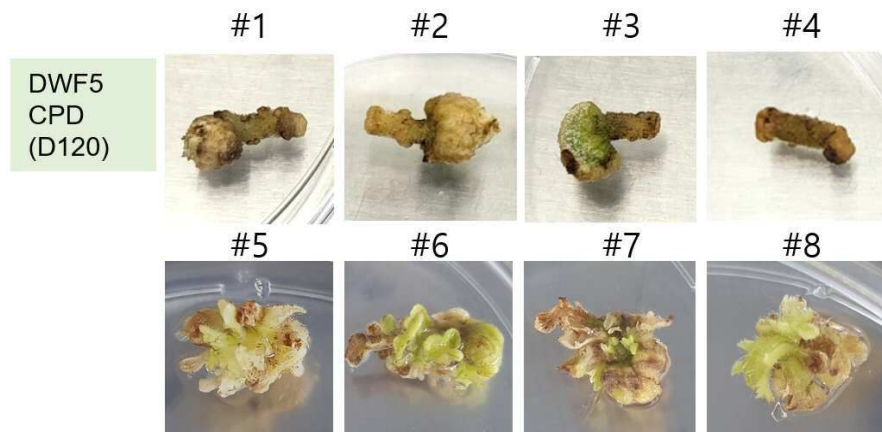
도면11



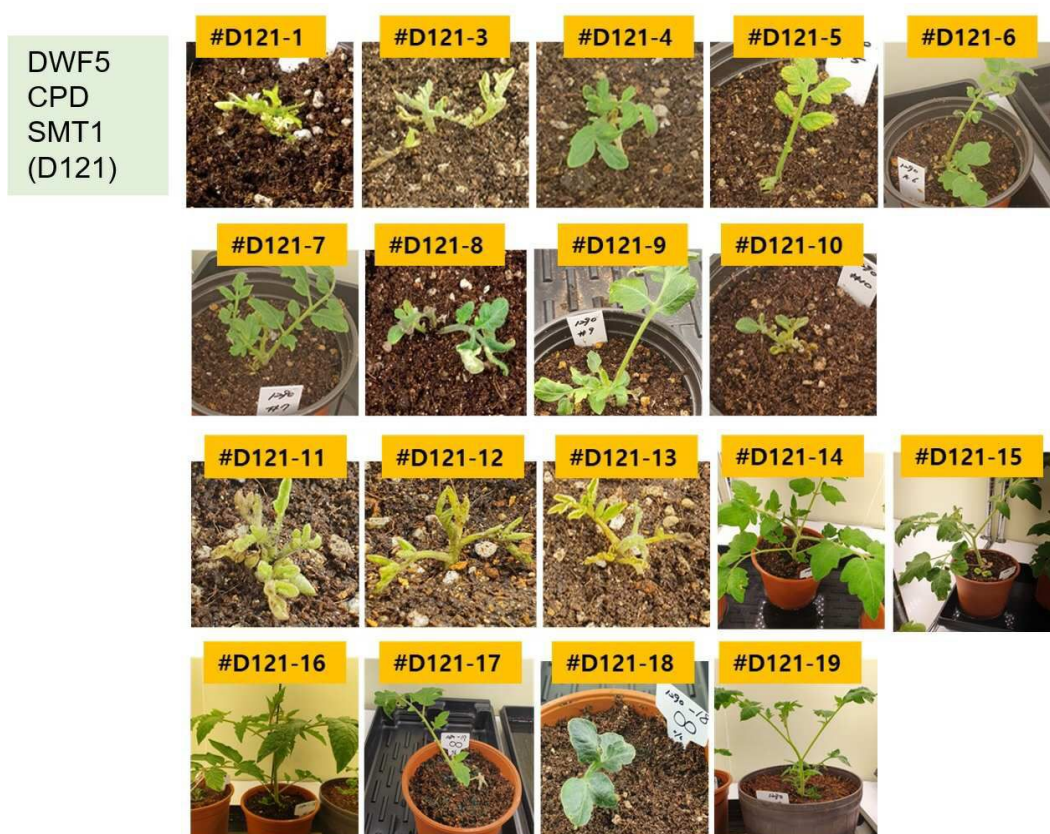
도면12



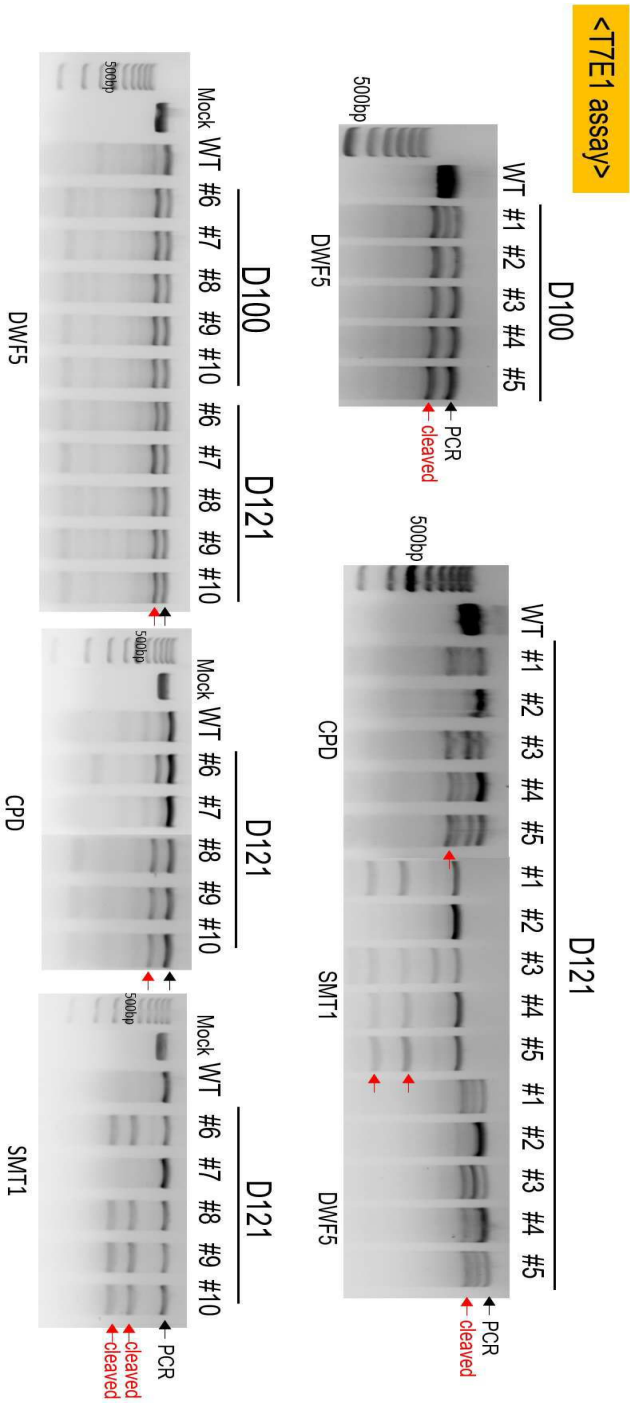
도면13



도면14



도면15



도면16

		D100-1	D100-2	D100-3	D100-4	D100-5	D100-6	D100-7
DWF5-1	Mutation	0/-159 +1/-4 -0/-11	+1/-4,-5/-1 -8/-1,-8/-4	+1/-4 -2/-1	+1/-4 -2/-1	+1/-4 -6/-1	+1/-4, 0/-1, -4/-1,-5/-1, -33/-1,-11/-1	0/-1,+1/-4 -4/-1,-17/-1
	Zygosity	Homo	Homo	Homo	Homo	Homo	Homo	Homo

		D100-8	D100-10	D100-12	D100-13	D100-14	D100-15	D100-16
DWF5-1	Mutation	+1/-4 0/-1	+1/-1 +1/-4 -5/-1	+1/-4 -12/-1 -5/-1	+1/-4 -8/-1	+1/-4 -5/-1	+1/4 -5/-1	+1/-4 -8/-1
	Zygosity	Homo	Homo	Homo	Homo	Homo	Homo	Homo

도면17

		D121-4	D121-5	D121-6	D121-7	D121-8	D121-9	D121-10
DWF5-1	Mutation	0/-158; 2 WT; 18	-157; 3 -5/-5; 1 -5/-3; 1 -5/0; 4 WT; 11	-1; 4 WT; 5	+1/-4; 6 -13/1, 0/-1, -8/-1,-3/-1; 1	-158; 5 -7,-4/-7; 1 -6; 1 WT; 5	-659; 1 -8/0,-6/0,- 4/0; 1 WT; 2	-4; 1 -6; 3 +1; 1 WT; 7
	Zygosity	Hetero	Hetero	Hetero	Homo	Hetero	Hetero	Hetero
CPD	Mutation	-136	-136 WT; 2	-2; 7 -5; 4	none	-2; 2 -3; 1 WT; 6	-2; 2 -3; 6 WT; 2	-2; 14 -3; 2 -4; 3
	Zygosity	homo	Hetero	homo	WT	Hetero	Hetero	Homo
SMT1	Mutation	-4; 1 -5/-4; 1 WT; 9	-8; 3 -4; 2 WT; 7	-4; 1 -2; 3 WT; 2	none	-4; 4 -2; 1 WT; 2	-1; 5 -2; 6 -7; 9 -91, WT; 1	-2; 5 -4; 5 WT; 1
	Zygosity	Hetero	Hetero	Hetero	WT	Hetero	Hetero	Hetero
Knockout		CPD		CPD	DWF5			CPD

		D121-13	D121-14	D121-15	D121-16	D121-17	D121-18	D121-19
DWF5-1	Mutation	-8; 1 WT; 9	-8; 2 WT; 8	-4; 4 +1; 1 WT; 3	-5; 5 -5/-15; 1 -5/-1; 1	-7; 2 WT; 5	-4; 5 -6; 2 -8/-7; 1 +1/-6; 1	-1; 1 -4; 3 -8; 1 WT; 4
	Zygosity	Hetero	Hetero	Hetero	Homo	Hetero	Homo	Hetero
CPD	Mutation	-8; 1 -4; 2 WT; 7	none	None	-136; 8 -6; 2	none	none	none
	Zygosity	Hetero	WT	WT	Homo	WT	WT	WT
SMT1	Mutation	-45; 2 WT; 4	-1; 2 -2; 6 -3; 1	-4; 7 -2; 3	-4; 3 -8; 5	-1; 3 -2; 6 -4; 1	-1; 2 -2; 5 -4; 2	-1; 2 -2; 6 -4; 2
	Zygosity	Hetero	Homo	Homo	Homo	Homo	Homo	Homo
Knockout			SMT1	SMT1	DWF5 CPD SMT1	SMT1	DWF5 SMT1	SMT1

도면18

T0	T1	Total reads (raw fastq)	SgRNA 1			SgRNA 2		
			Insertions	Deletions	Indel freq.	Insertions	Deletions	Indel freq.
D100-1	D100_1-1	87344	86370	340	86710 (99.8%)	0	86913	86913 (99.9%)
	D100_1-2	54277	53495	29	53524 (99.9%)	0	53629	53629 (100.0%)
	D100_1-4	54192	53220	51	53271(99.9%)	0	53357	53357 (100.0%)
	D100_1-5	69409	68498	167	68665 (99.9%)	0	68821	68821 (100.0%)
	D100_3-5	50717	44	50256	50300 (99.9%)	0	50376	50376 (100.0%)
	D100_3-6	71229	70528	65	70593 (99.8%)	0	70775	70775 (99.9%)
D100-3	D100_3-11	54964	17	54531	54548 (100.0%)	0	54615	54615 (100.0%)
	D100_3-14	44027	16	42986	43002 (100.0%)	0	43042	43042 (100.0%)
	D100_3-15	53998	53473	70	53543 (100.0%)	0	53663	53663 (100.0%)
	D100_3-18	53317	49	52868	52917 (100.0%)	0	52978	52978 (100.0%)
	D100_4-4	83511	386	82710	83096 (100.0%)	0	82938	82938 (100.0%)
	D100_4-8	83297	65	82819	82884 (100.0%)	0	82757	82757 (100.0%)
D100-4	D100_4-9	79034	108	78516	78624 (100.0%)	0	78519	78519 (100.0%)
D100-5	D100_5-4	60994	60531	60	60591 (99.9%)	0	60737	60737 (100.0%)
	D100_5-9	71134	70473	134	70607 (99.9%)	0	70699	70699 (99.9%)
	D100_7-1	89757	89183	54	89237(99.9%)	0	89421	89421 (100.0%)
D100-7	D100_7-5	99082	58	98701	98759 (100.0%)	0	98803	98803 (100.0%)
D100-8	D100_8-1	71913	37	39	76 (0.1%)	0	71658	71658 (100.0%)
	D100_8-3	76121	75656	81	75737 (99.9%)	0	75867	75867 (100.0%)
	D100_8-4	113584	85	111676	111761 (100.0%)	0	111869	111869 (100.0%)
D100-12	D100_12-3	88658	88001	163	88164 (99.9%)	0	88331	88331 (100.0%)
	D100_12-7	110408	109131	623	109754 (99.9%)	0	109946	109946 (100.0%)
	D100_12-8	7532	7377	13	7390 (99.9%)	0	7412	7412 (100.0%)

Label	Total reads	SgRNA 1			SgRNA 2			T2 종자량(g)
		Insertions	Deletions	Indel freq.	Insertions	Deletions	Indel freq.	
D100-1-4	54192	53220	51	99.9%	0	53357	100%	0.37
D100-7-1	89757	89183	54	99.9%	0	89421	100%	0.35 1.06
D100-12-8	7532	7377	13	99.9%	0	7412	100%	0.34
D100-3-11	54964	17	54531	100%	0	54615	100%	0.13
D100-3-14	44027	16	42986	100%	0	43042	100%	0.32 0.45
D100-4-8	83297	65	82819	100%	0	82757	100%	0

SolyCDWF5_1-1

SolyCDWF5_1-2

WT GGTATGGAGCTTTATCCTCGTA-TG**GG**AAACACTTCGATATTAAAGCTTCACAACTGCAGATTGGTATGATGCT**TGG**CACTTC

#3-11 GGTATGGAGCTTTATCCTC-A-TGGGAAACACTTCGATATTAAAGCTTCACAACTGCAGATTGGTATGAT-TCTTGGGCACTTC (-2/-1)

#3-14 GGTATGGAGCTTTATCCTC-A-TGGGAAACACTTCGATATTAAAGCTTCACAACTGCAGATTGGTATGAT-TCTTGGGCACTTC (-2/-1)

#4-8 GGTATGGAGCTTTATCCTC-A-TGGGAAACACTTCGATATTAAAGCTTCACAACTGCAGATTGGTATGAT-TCTTGGGCACTTC (-2/-1)

#1-4 GGTATGGAGCTTTATCCTCGTAATCGGGAAACACTTCGATATTAAAGCTTCACAACTGCAGATTGGTAT-TCTTGGGCACTTC (+1/-4)

#7-1 GGTATGGAGCTTTATCCTCGTAATCGGGAAACACTTCGATATTAAAGCTTCACAACTGCAGATTGGTAT-TCTTGGGCACTTC (+1/-4)

#12-8 GGTATGGAGCTTTATCCTCGTAATCGGGAAACACTTCGATATTAAAGCTTCACAACTGCAGATTGGTAT-TCTTGGGCACTTC (+1/-4)



도면19

도면20

시료 전처리 방법

시료 일정량을 칭량한 뒤 EtOH 30 mL, 10% 에탄올성 피로갈로 1 mL 를 넣는다.



90% KOH 3 mL 를 넣고 80℃ 수욕조에서 30 분간 반응시키며 10분 마다 vortexing 한다.



실온으로 방냉한 후 증류수 30 mL, n-Hexane 30 mL 를 넣고 1시간 진탕한다.



상층액을 분액여두에 옮기고 물층에 n-Hexane 30 mL 넣고 vortexing 한다.



상층액을 분액여두에 옮기고 Hexane 층에 증류수 100 mL 를 넣고 수세한다. (2회)



Hexane 층을 무수황산나트륨으로 탈수 한 뒤 농축한다.



MeOH 5 mL로 정용한 후 PTFE 필터로 여과하여 시험용액으로 한다.

* 기존 Vitamin D 전처리 방법을 응용하여 작성

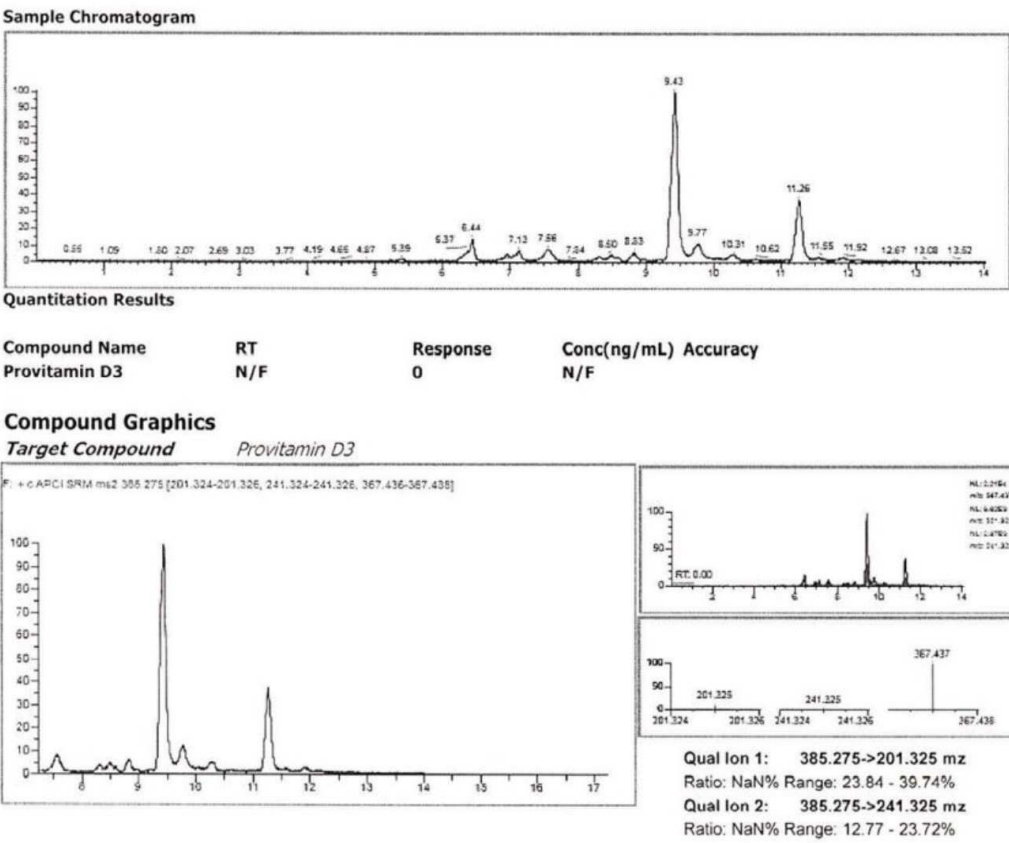
도면21

기기 분석 조건

Instrument	HPLC-MS/MS		
Ion source Type	APCI		
Injection volume	20 μL		
Column	Waters HSS T3 (1.8μm, 2.1×100mm)		
Mobile Phase	A : 0.1% Formic acid in D.W		
	B : 0.1% Formic acid in MeOH		
	시간(min)	A(%)	B(%)
	0	90	10
	1	0	100
	10	0	100
	10.1	90	10
14	90	10	
Flow rate	0.2 mL/min		
Run time	14 min		
Oven Temperature	30 °C		
Pos ion discharge current	4 (μA)		
Sheath gas pressure	50 (Arb)		
Aux gas pressure	10 (Arb)		
Sweep gas pressure	0 (Arb)		
Ion transfer tube temperature	325 °C		
SRM	Parent MS - 385.275		
	Product MS - 정량이온 - 367.437		
	정성이온 - 201.325 / 241.325		

도면22

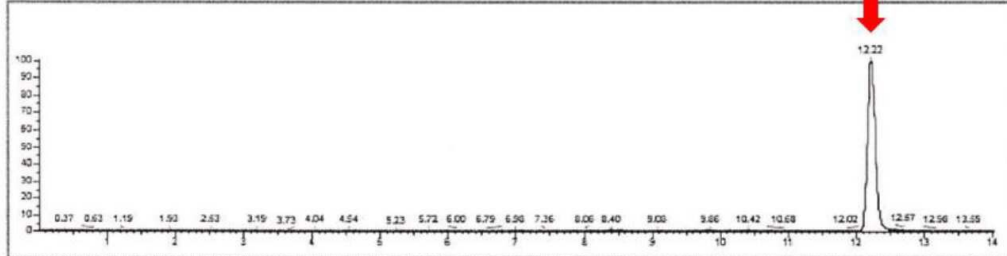
WT 토마토 (서광)



도면23

7-dehydrocholesterol : Provitamin D3 standard

Sample Chromatogram

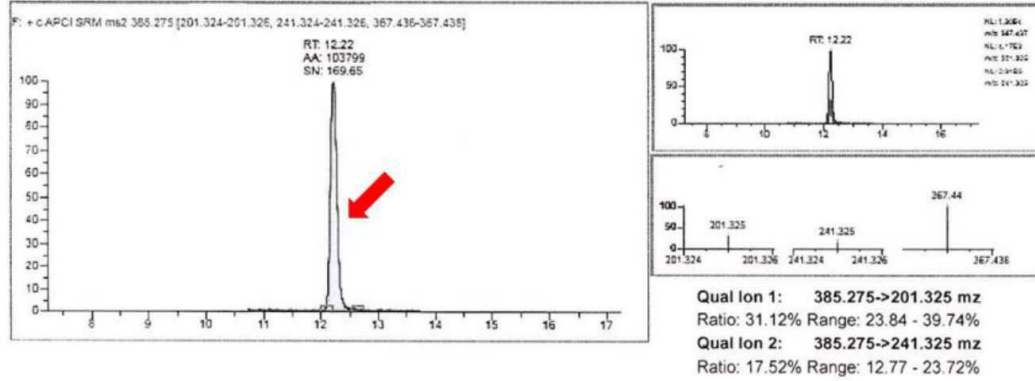


Quantitation Results

Compound Name	RT	Response	Conc(ng/mL)	Accuracy
Provitamin D3	12.22	103799	1065.1220	99.17

Compound Graphics

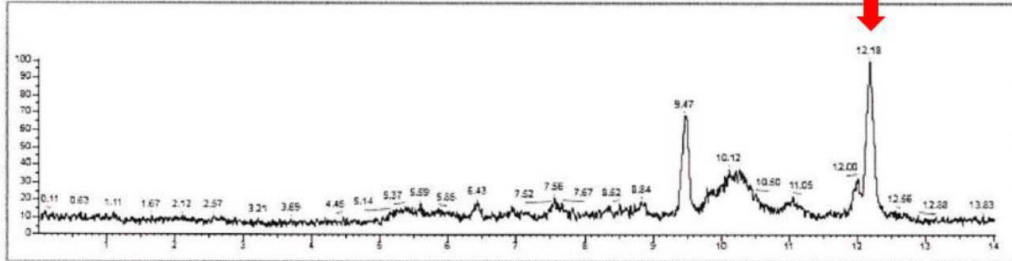
Target Compound Provitamin D3



도면24

Dwf5 토마토 : #3-14-2 (성숙과실)

Sample Chromatogram

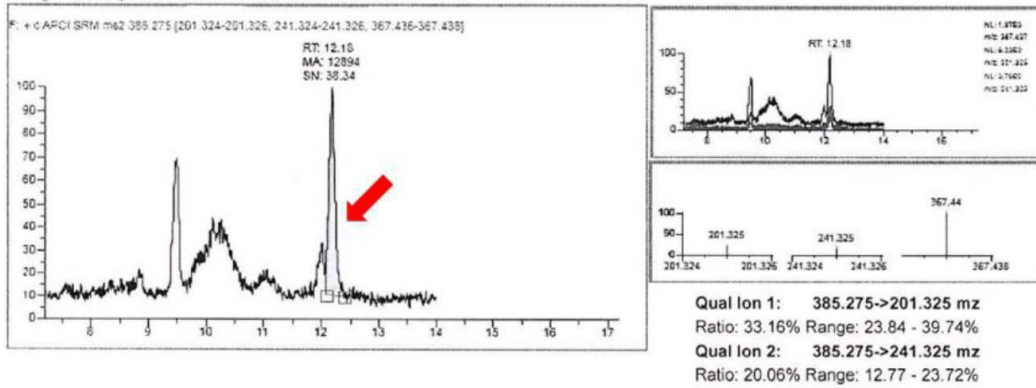


Quantitation Results

Compound Name	RT	Response	Conc(ng/mL)	Accuracy
Provitamin D3	12.18	12894	115.4740	

Compound Graphics

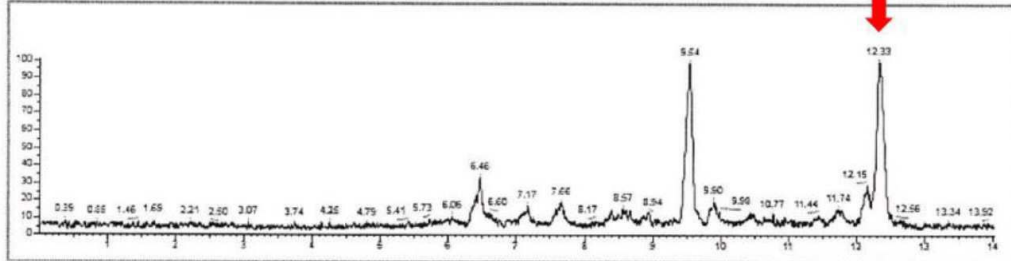
Target Compound Provitamin D3



도면25

Dwf5 토마토 : #3-14-13 (성숙과실)

Sample Chromatogram

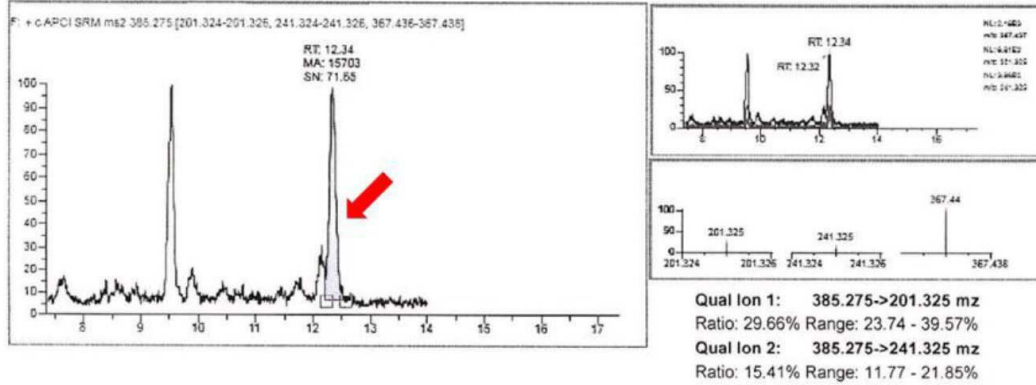


Quantitation Results

Compound Name	RT	Response	Conc(ng/mL)	Accuracy
Provitamin D3	12.34	15703	140.8500	

Compound Graphics

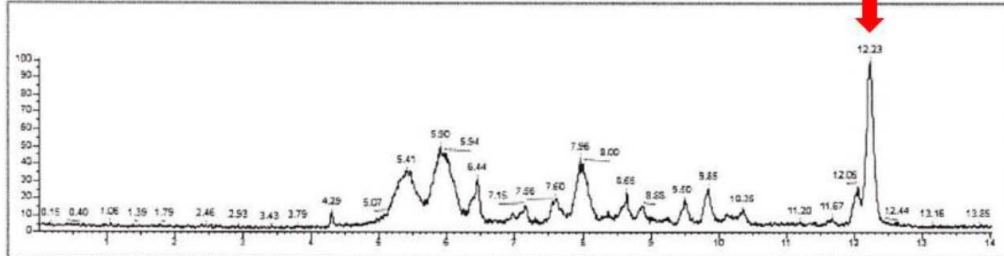
Target Compound Provitamin D3



도면26

Dwf5 토마토 : #7-1-12 (미성숙과실 포함)

Sample Chromatogram

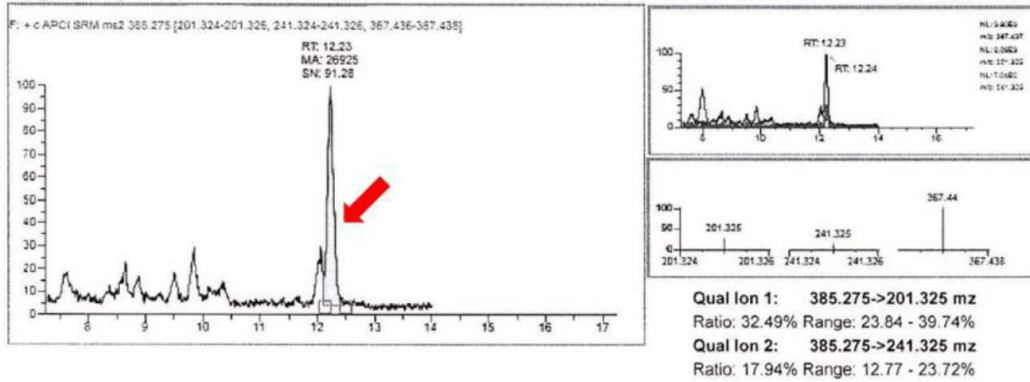


Quantitation Results

Compound Name	RT	Response	Conc(ng/mL)	Accuracy
Provitamin D3	12.23	26925	262.0450	

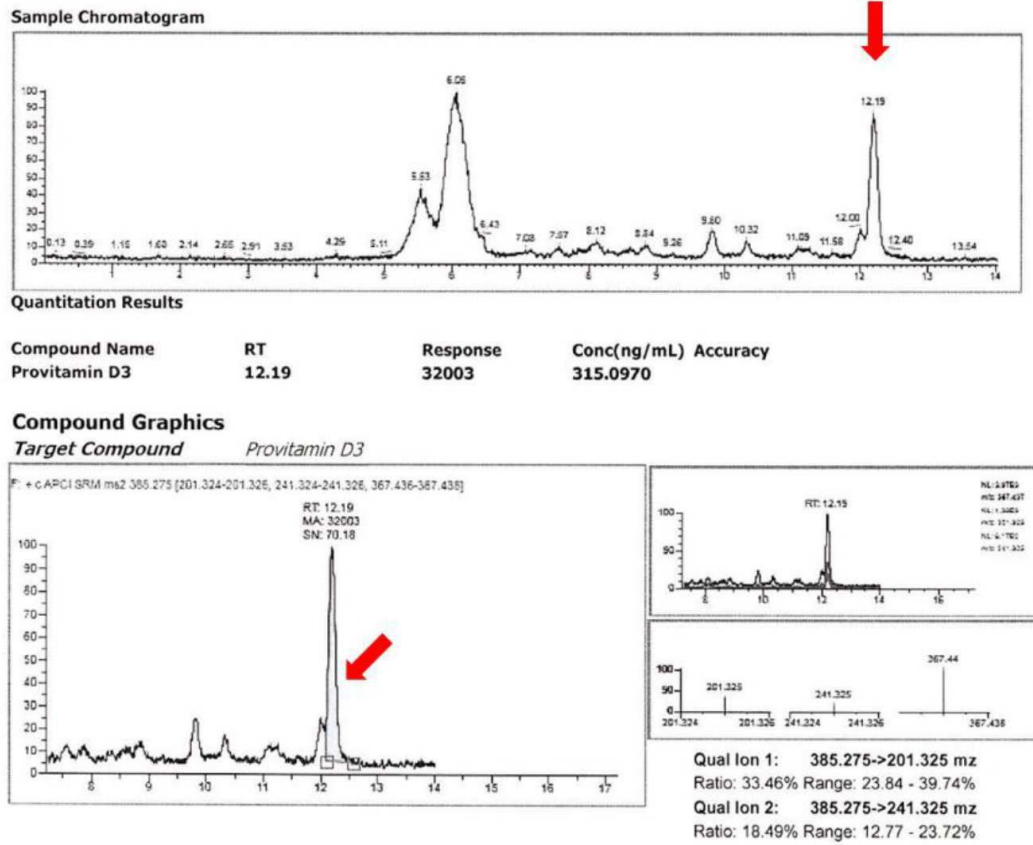
Compound Graphics

Target Compound Provitamin D3

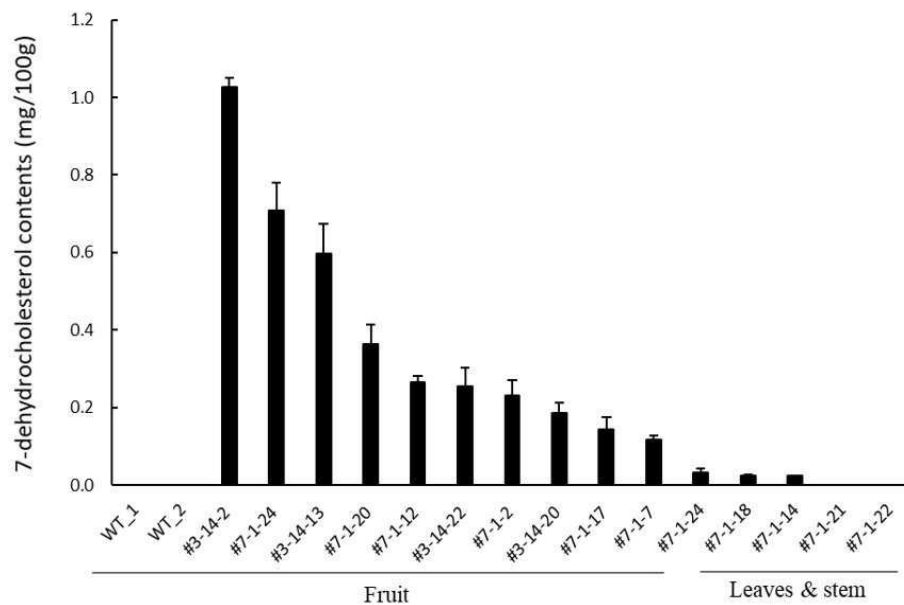


도면27

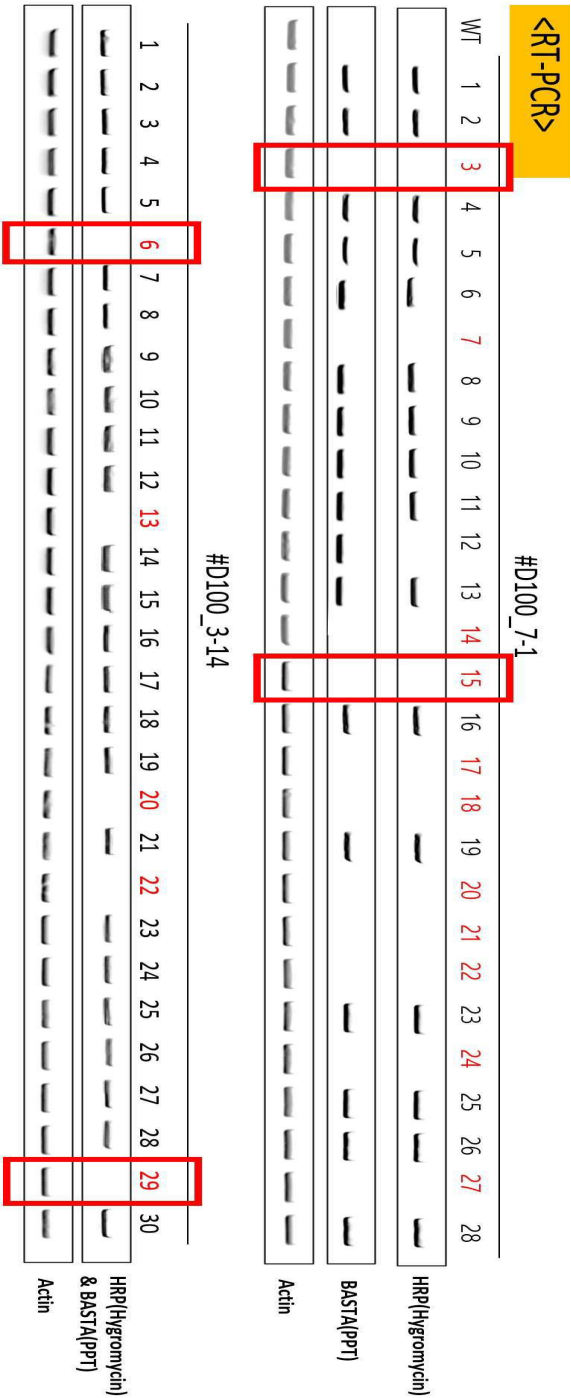
Dwf5 토마토 : #7-1-24 (미성숙과실)



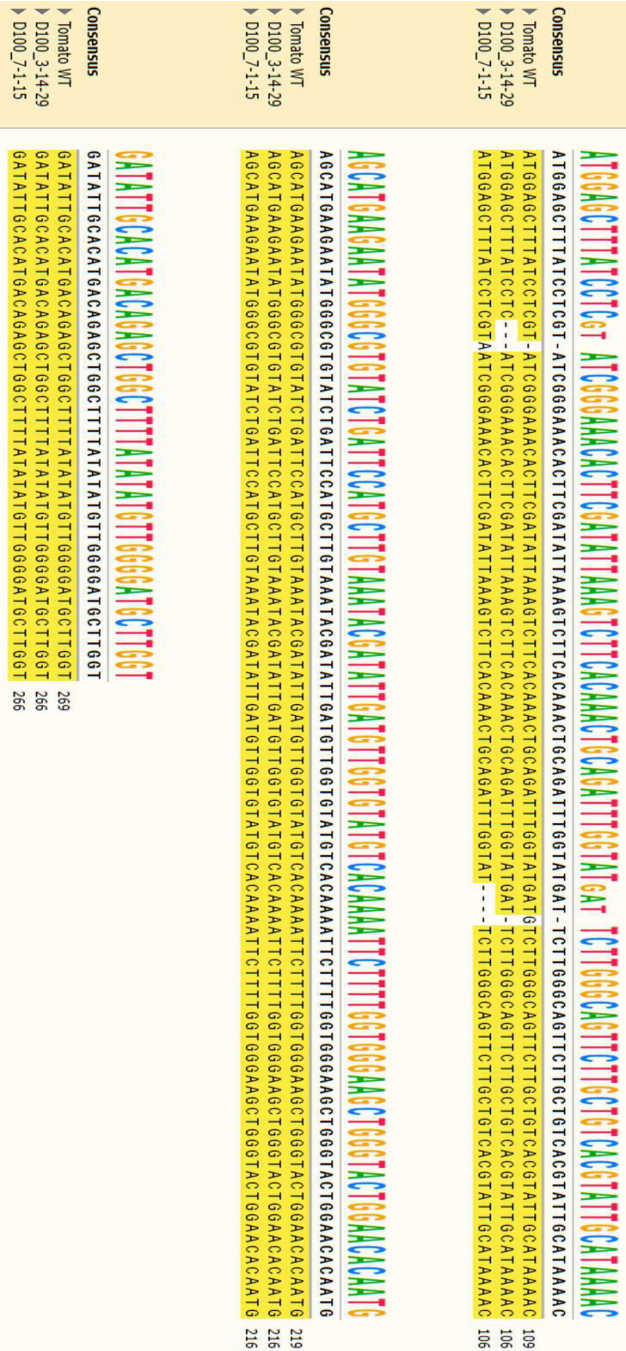
도면28



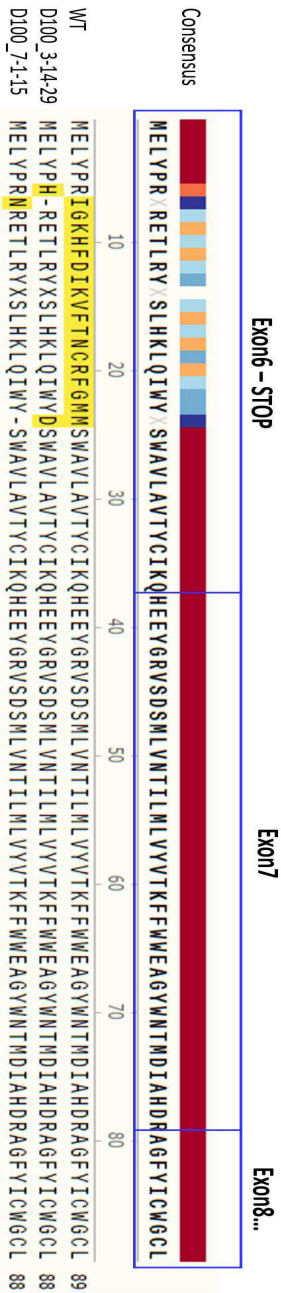
도면29



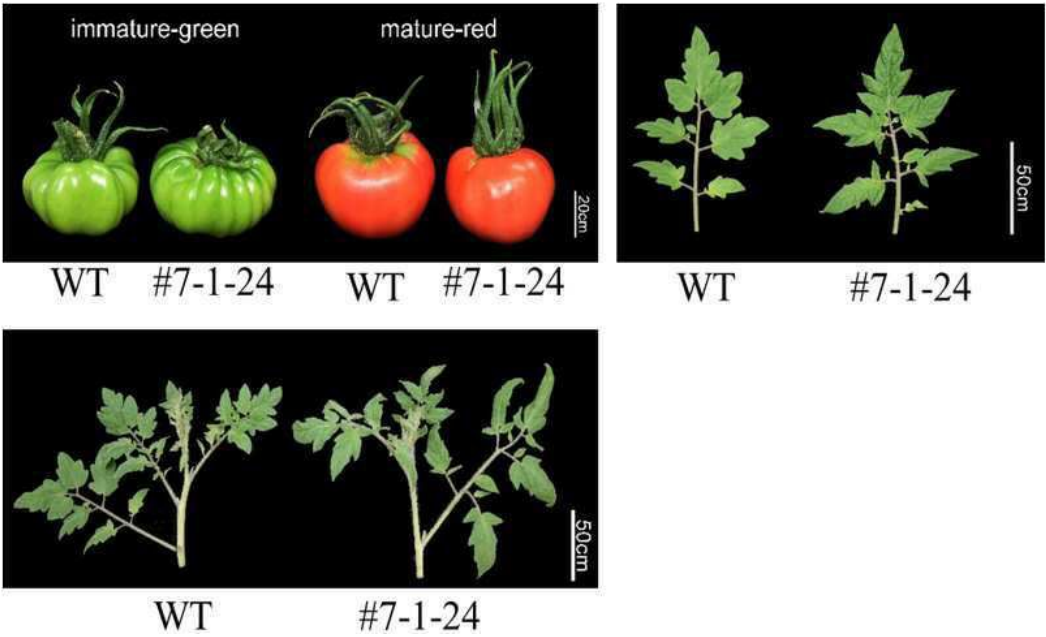
도면30



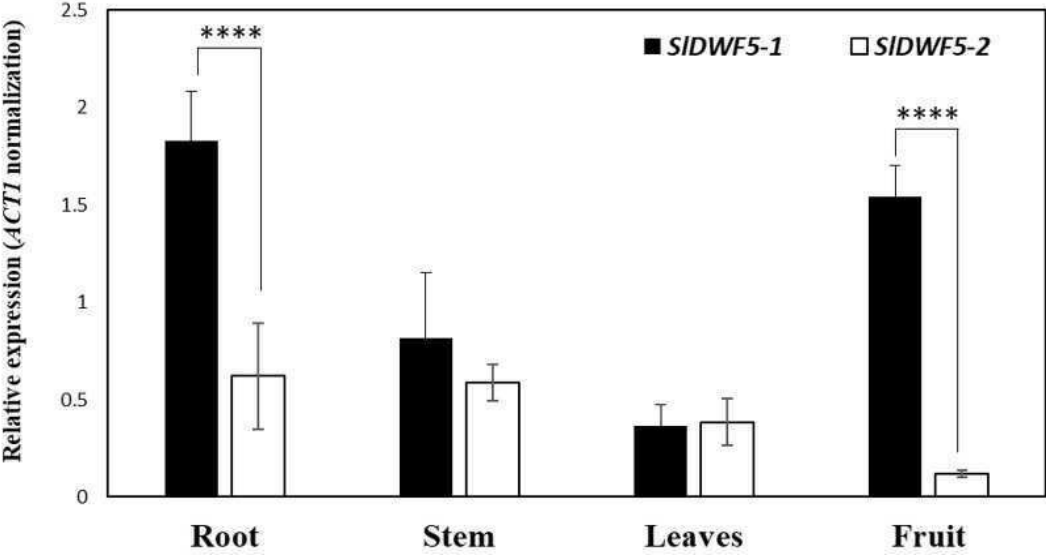
도면31



도면32



도면33



서열목록

- <110> G+FLAS Life Sciences
- <120> TOMATO CONTAINING HIGH CONCENTRATION OF 7-DEHYDROCHOLESTEROL AND METHOD FOR PRODUCING THE SAME
- <130> SPD21-127-GFS
- <160> 39
- <170> KoPatent In 3.0
- <210> 1
- <211> 1678

<212> DNA

<213> Artificial Sequence

<220><223> nucleotides coding DWF5 of *S. lycopersicum*

<400> 1

ttagtcagta caacagattt ttataattga gagaatctga gaagaggacc tataaccactg	60
tggagagagt ggagtaaaca cagagcacta aaaatcagtg aattaaagtc tacatataag	120
tcaagattcc atattgggcg gagacttctg tagttccagt tttttcttc aaaaccccaa	180
aatcagcaaa agctgaattg attgaggaag aagaagaaga agaaaaaatg gtggagaaca	240
agttggtaca ctctccatta atcacttatg gatcaatgct cagtcttcta tctttcacac	300
caccttttgt aatcctcatg tggatataca atgtgcacgc tgatgggtct atcttgaaaa	360
cgttcaatca tctaaggga aatggtttgc aaggattaat tgatatttgg ccaaaccta	420
ctgcaattgc tgggaagcta atcatttgtt atgctctatt tgaggcggca ctacagcttc	480
tattgcctgg taaaacagtc gaaggcccaa tatctccaac tggacatagg cctgtctata	540
aggctaattg catggcagca tatgcagtga cactgatcac atatatcagt ctctggtggt	600
ttggaatatt taatcctgca attgtttatg accatctggg cgagatttct tctacactta	660
tttttgggag tttagttctg tgtgttttat tgtacattaa aggtcatgtt gcaccatctt	720
ccactgattc cggctcatca gggaacataa tagttgactt ctattggggg atggagcttt	780
atcctcgtat cgggaaacac ttcgatatta aagtcttcac aaactgcaga ttggtatga	840
tgtcttgggc agttcttgct gtcacgtatt gcataaaaca gcatgaagaa tatgggcgtg	900
tatctgattc catgcttgta aatacgatat tgatgttggg gtatgtcaca aaattctttt	960
ggtgggaagc tgggtactgg aacacaatgg atattgcaca tgacagagct ggcttttata	1020
tatgttgggg atgcttgggt tgggttccaa gcatatatac gtctcctgga atgtatcttg	1080
ttaaacaacc tgtgaatctt gggcttcagc ttcccttta catactagtg gctggtcttc	1140
tctgcatata tatcaactat gattgtgaca ggcagaggca agagtccgg agaacaatg	1200
gaaaatgcac tgcttgggga aagactccat ctaagatagt cgccgcatac actactacct	1260
ctggtgagaa aaagaccagc cttctgctta catcaggatg gtggggctta gctcgccatt	1320
tccactatgt tccagaaata ttagcggcat ttttctggag tgtaccagct cttttcaacc	1380
acttcattcc ctacttctac gtcacttttc taataatcct ccttctcgat cgagctaaaa	1440
gggacgatga ccgatgcaag gcaagtatg gcaagtactg gaaattgtat tgtgaaaagg	1500
ttccttacag agttatccca ggaatttact agtgattttt ggacttctaa gaatgtttta	1560
ctataaagtt tgatgtattt tgtagcctct tcatgttggt gtttgtatgt acacttatta	1620

gttccaacag ttttgagaaa cttgttttta tataaaaata gtgaagacgc gagttgag 1678

<210> 2

<211> 434

<212> PRT

<213> Artificial Sequence

<220><223> DWF5 of *S. lycopersicum*

<400>

> 2

Met Val Glu Asn Lys Leu Val His Ser Pro Leu Ile Thr Tyr Gly Ser

1 5 10 15

Met Leu Ser Leu Leu Ser Phe Thr Pro Pro Phe Val Ile Leu Met Trp

20 25 30

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

35 40 45

Leu Arg Glu Asn Gly Leu Gln Gly Leu Ile Asp Ile Trp Pro Lys Pro

50 55 60

Thr Ala Ile Ala Gly Lys Leu Ile Ile Cys Tyr Ala Leu Phe Glu Ala

65 70 75 80

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

85 90 95

Pro Thr Gly His Arg Pro Val Tyr Lys Ala Asn Gly Met Ala Ala Tyr

100 105 110

Ala Val Thr Leu Ile Thr Tyr Ile Ser Leu Trp Trp Phe Gly Ile Phe

115 120 125

Asn Pro Ala Ile Val Tyr Asp His Leu Gly Glu Ile Phe Ser Thr Leu

130 135 140

Ile Phe Gly Ser Leu Val Phe Cys Val Leu Leu Tyr Ile Lys Gly His

145 150 155 160

Val Ala Pro Ser Ser Thr Asp Ser Gly Ser Ser Gly Asn Ile Ile Val

165 170 175

Asp Phe Tyr Trp Gly Met Glu Leu Tyr Pro Arg Ile Gly Lys His Phe

180 185 190

Asp Ile Lys Val Phe Thr Asn Cys Arg Phe Gly Met Met Ser Trp Ala

195 200 205
Val Leu Ala Val Thr Tyr Cys Ile Lys Gln His Glu Glu Tyr Gly Arg

210 215 220
Val Ser Asp Ser Met Leu Val Asn Thr Ile Leu Met Leu Val Tyr Val
225 230 235 240

Thr Lys Phe Phe Trp Trp Glu Ala Gly Tyr Trp Asn Thr Met Asp Ile
245 250 255

Ala His Asp Arg Ala Gly Phe Tyr Ile Cys Trp Gly Cys Leu Val Trp
260 265 270

Val Pro Ser Ile Tyr Thr Ser Pro Gly Met Tyr Leu Val Lys Gln Pro
275 280 285

Val Asn Leu Gly Leu Gln Leu Ser Leu Tyr Ile Leu Val Ala Gly Leu
290 295 300

Leu Cys Ile Tyr Ile Asn Tyr Asp Cys Asp Arg Gln Arg Gln Glu Phe
305 310 315 320

Arg Arg Thr Asn Gly Lys Cys Thr Val Trp Gly Lys Thr Pro Ser Lys
325 330 335

Ile Val Ala Ala Tyr Thr Thr Thr Ser Gly Glu Lys Lys Thr Ser Leu
340 345 350

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

355 360 365
Pro Glu Ile Leu Ala Ala Phe Phe Trp Ser Val Pro Ala Leu Phe Asn

370 375 380
His Phe Ile Pro Tyr Phe Tyr Val Ile Phe Leu Ile Ile Leu Leu Leu

385 390 395 400
Asp Arg Ala Lys Arg Asp Asp Asp Arg Cys Lys Ala Lys Tyr Gly Lys

405 410 415
Tyr Trp Lys Leu Tyr Cys Glu Lys Val Pro Tyr Arg Val Ile Pro Gly

420 425 430

Ile Tyr

<210> 3

<211> 1735

<212> DNA

<213> Artificial Sequence

<220><223> nucleotides coding DWF5 of *S. lycopersicum*

<400> 3

tagaagaatc agcggagagt tggagaacca cgtagcagt atacgacaag gaagattacc	60
agtgtccagt gtgcgtgtgt gtaatttata atcaacaaaa ttagcagtgg ccgagttatc	120
ttttagtttt tcaaaaacaa gaagaggtgc agagaacagg gagggaggaa gaagcaatgg	180
cggagtcctca gttaggtcac cctcctttat tcacctatat ctcaatgac actctgctca	240
ctttagtacc cccttttggt attctcatgt ggtatacaaa tgttcatgct gatggatctg	300
tattgcaaac atataattat ctaaaggaga atggcctgca aggactaatt gacatttggc	360
caagaccac tgcaattgcg ggaaaaataa taatttgcta tgctctatct gaggtacac	420
ttcagctttt gttgcctggt aaaagggttg aaggaccgat atctcctact ggacacaggc	480
ctgtctataa ggcaaatggc atggcagcat atacagtac actaattacc tatctcagcc	540
tttggtgggt tggaatatc aatcctacaa tcgtgtatga tcatctggga gagattctct	600
ctacactaaa ttttgaagc ttaatcttct gtctcttttt atacataaaa ggtcatgttg	660
caccatcttc cactgatcac ggttcatcag ggaacataat agttgactac tattggggga	720
tggagtgtga tcctcgcat ggcaaacact ttgatatcaa ggtcttcaca aactgcagat	780
ttggcatgat atcttgggga ctctctcccta ttacctactg tataaagcag tatgaagaat	840
atggaagtct ctctgattcc atgcttatac ataccataat aacgttgggt tatgtcacga	900
aattcttttg gtgggaagca ggttactgga acaccatgga tattgctcat gacagagccg	960
gtttttacat atgctgggggt tgcttgggtat ttcttccatg tatgtacact tctcctggca	1020
tgtacctgtt caaacatctt gtaaatcttg gacctcagct tgcaatttct atcctagtag	1080
ctggtattct ctgcgtatac ataaactatg attgtgacag acagaggcaa gagttccgca	1140
gaacaaatgg caaagccctt gtctggggaa aggctccttc aaagattgtc gcctcataca	1200
ctactacaac tggcgaaact aaaaccagca ttctcctgac atctggatgg tggggcttat	1260
cgcggcattt ccattatgtt ccagaaatac tagcttcgtt tttctggagt gtgccagctc	1320
ttctcaacca tttcatgccc tacatctatg taatctatct gacgggcctc cttttggatc	1380
gagccaaaag ggatgacgaa aggtgcaagt caaagtatgg caaatactgg aaaaagtatt	1440
gtgaaaaggt gccttaccgg gtagtccttg gaatttatta gtgattcttg agcttcaact	1500

gatgaactgt ctttaatagt gctgcttcta gtattgtagt agtatTTaga ccatgaaaaa 1560

ttttcttctt tcaataaaagc acttttagtt gaaccaacat ctcatttcta tggttgtaag 1620

ttgttacacg atgaaaaaat cagcttgaaa tagatatgcc gtgaaattag ttgaggtgtg 1680

tgaaaactcg gaccagtttg aaatattgtt ccaattaaaa tgaacactaa aacag 1735

<210> 4

<211> 434

<212> PRT

<213> Artificial Sequence

<220><223> DWF5 of *S. lycopersicum*

<400> 4

Met Ala Glu Ser Gln Leu Val His Pro Pro Leu Phe Thr Tyr Ile Ser

1 5 10 15

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

20 25 30

Tyr Thr Asn Val His Ala Asp Gly Ser Val Leu Gln Thr Tyr Asn Tyr

35 40 45

Leu Lys Glu Asn Gly Leu Gln Gly Leu Ile Asp Ile Trp Pro Arg Pro

50 55 60

Thr Ala Ile Ala Gly Lys Ile Ile Ile Cys Tyr Ala Leu Phe Glu Ala

65 70 75 80

Thr Leu Gln Leu Leu Leu Pro Gly Lys Arg Val Glu Gly Pro Ile Ser

85 90 95

Pro Thr Gly His Arg Pro Val Tyr Lys Ala Asn Gly Met Ala Ala Tyr

100 105 110

Thr Val Thr Leu Ile Thr Tyr Leu Ser Leu Trp Trp Phe Gly Ile Phe

115 120 125

Asn Pro Thr Ile Val Tyr Asp His Leu Gly Glu Ile Leu Ser Thr Leu

130 135 140

Asn Phe Gly Ser Leu Ile Phe Cys Leu Phe Leu Tyr Ile Lys Gly His

145 150 155 160

Val Ala Pro Ser Ser Thr Asp His Gly Ser Ser Gly Asn Ile Ile Val

165	170	175
Asp Tyr Tyr Trp Gly Met Glu Leu Tyr Pro Arg Ile Gly Lys His Phe		
180	185	190
Asp Ile Lys Val Phe Thr Asn Cys Arg Phe Gly Met Ile Ser Trp Gly		
195	200	205
Leu Leu Pro Ile Thr Tyr Cys Ile Lys Gln Tyr Glu Glu Tyr Gly Ser		
210	215	220
Leu Ser Asp Ser Met Leu Ile His Thr Ile Ile Thr Leu Val Tyr Val		
225	230	235
		240
Thr Lys Phe Phe Trp Trp Glu Ala Gly Tyr Trp Asn Thr Met Asp Ile		
245	250	255
Ala His Asp Arg Ala Gly Phe Tyr Ile Cys Trp Gly Cys Leu Val Phe		
260	265	270
Leu Pro Cys Met Tyr Thr Ser Pro Gly Met Tyr Leu Val Lys His Pro		
275	280	285
Val Asn Leu Gly Pro Gln Leu Ala Ile Ser Ile Leu Val Ala Gly Ile		
290	295	300
Leu Cys Val Tyr Ile Asn Tyr Asp Cys Asp Arg Gln Arg Gln Glu Phe		
305	310	315
Arg Arg Thr Asn Gly Lys Ala Leu Val Trp Gly Lys Ala Pro Ser Lys		
325	330	335
Ile Val Ala Ser Tyr Thr Thr Thr Thr Gly Glu Thr Lys Thr Ser Ile		
340	345	350
Leu Leu Thr Ser Gly Trp Trp Gly Leu Ser Arg His Phe His Tyr Val		
355	360	365
Pro Glu Ile Leu Ala Ser Phe Phe Trp Ser Val Pro Ala Leu Leu Asn		
370	375	380
His Phe Met Pro Tyr Ile Tyr Val Ile Tyr Leu Thr Gly Leu Leu Leu		
385	390	395
Asp Arg Ala Lys Arg Asp Asp Glu Arg Cys Lys Ser Lys Tyr Gly Lys		
405	410	415
Tyr Trp Lys Lys Tyr Cys Glu Lys Val Pro Tyr Arg Val Val Pro Gly		

	420	425	430
Ile Tyr			
<210>	5		
<211>	1742		
<212>	DNA		
<213>	Artificial Sequence		
<220><223>	nucleotides coding CPD of <i>S. lycopersicum</i>		
<400>	5		
gcagagcggtt ttgcatacaa aaaaaaatta attaacatgg ataccatcga tctttttctc			60
tatctttttc tctctgtttt caccttttat ctctccggt gtatggcggc ggcccatctt			120
cgcgggcgta aaactcgact cccgccggga acccttggcc tcccgtttat cggagaaacc			180
ctccagttaa tttctgcgta caaacggaa aatcctgaac cgttcacga tgaccgtgtt			240
tcaaaatacg gcagcatttt cagcacccat gtttttgggtg aaccgacggt ttctctgct			300
gacccggaaa cgaaccggtt tattttgcag aatgaaggtc ggctttttga gtcgagtat			360
cccgttcga tacagaattt acttgggaga tactctctgt tgcttatgag aggaagtcta			420
cacaaacgaa tgcattcatt aactatgagt ttgtctaatt ctccattct taaagatcat			480
ctgttggcg atatagatcg attggttagg cttaatttgg attcatggac cggcagggtt			540
ttctcatgg acgaggctaa gaagataacg ttaatactaa cagtgaagca gttgatgagt			600
tttgatccat gtgagtggac agagaatctg atgaaagagt atatgcttgt tattgaaggt			660
ttcttctgca ttcttttgc tattttctca tccacctatc gcaaggccat tcaagcgaga			720
acgaaagtag cggaggcggt gggattggta gtgagagatc ggagaaaaga acgagacgga			780
ggagaacgaa agaatgatat gttggaggca ttgttcgaag gagacggagt tgaaggagta			840
ggattttccg atgaggaaat tgttgatttt atactggcgt tgcttgttgc tggctatgaa			900
actacctcca ccatcatgac ctttctgtga aaattcctca cggagacacc ccgtgctcta			960
tcactactta aggaagagca cgaggagatc aggttaagaa aaggatgaagt taagtcttta			1020
caatgggaag attacaagtc aatgcccttc actcaatgcg ttgttaatga aactctcgca			1080
atcgctaaca taattagtgg agtgttcagg agagccatga ctgacataaa tatcaaaggt			1140
tataccattc caaaagggtg gaaggttttt gcttctctcc gagctgttca tctagaccat			1200
gaacatttca aagatgcgcg tacatttgat ccatggaggt ggacagtag tgcaggaccg			1260
acaagctcac ctaacgtatt cacaccattt ggtgggtggac ctctcgatg tcctggttac			1320

gagcttgcta gagtggaaact ctctgttttc cttcaccact tggtgactcg ttttagttgg 1380
gttcggctg aggagataa gttagttttc ttcccaacga caaggatgct gaagcgatat 1440
ccaattaacg tccaacatcg aagcttggtt gagcagaaag aagaagagaa gggttcgtga 1500
gattgtagaa tgcagagtgg agggagggtg atgaaagggtg ttaaataaga caatagtata 1560
taggtgttag catatactta gaaaattaaa ttattttag tagaattgagc agagattaac 1620
attcctctga tctgcaaggg gaatcaggaa gatatttattc tgagcaattg gttattatat 1680

aatagtaatg cacagttttt aatgtatgtt aattagcatg attaaagatt taattgcctt 1740
ca 1742

<210> 6

<211> 487

<212> PRT

<213> Artificial Sequence

<220><223> CPD of *S. lycopersicum*

<400> 6

Met Asp Thr Ile Asp Leu Phe Leu Tyr Leu Phe Leu Ser Val Phe Thr

1 5 10 15

Phe Tyr Leu Leu Arg Cys Met Ala Ala Ala His Leu Arg Gly Arg Lys

20 25 30

Thr Arg Leu Pro Pro Gly Thr Leu Gly Leu Pro Phe Ile Gly Glu Thr

35 40 45

Leu Gln Leu Ile Ser Ala Tyr Lys Thr Glu Asn Pro Glu Pro Phe Ile

50 55 60

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

65 70 75 80

Gly Glu Pro Thr Val Phe Ser Ala Asp Pro Glu Thr Asn Arg Phe Ile

85 90 95

Leu Gln Asn Glu Gly Arg Leu Phe Glu Ser Ser Tyr Pro Gly Ser Ile

100 105 110

Gln Asn Leu Leu Gly Arg Tyr Ser Leu Leu Leu Met Arg Gly Ser Leu

115 120 125

His Lys Arg Met His Ser Leu Thr Met Ser Phe Ala Asn Ser Ser Ile

130 135 140

Leu Lys Asp His Leu Leu Gly Asp Ile Asp Arg Leu Val Arg Leu Asn
 145 150 155 160
 Leu Asp Ser Trp Thr Gly Arg Val Phe Leu Met Asp Glu Ala Lys Lys
 165 170 175

 Ile Thr Phe Asn Leu Thr Val Lys Gln Leu Met Ser Phe Asp Pro Cys
 180 185 190
 Glu Trp Thr Glu Asn Leu Met Lys Glu Tyr Met Leu Val Ile Glu Gly
 195 200 205
 Phe Phe Cys Ile Pro Leu Pro Ile Phe Ser Ser Thr Tyr Arg Lys Ala
 210 215 220
 Ile Gln Ala Arg Thr Lys Val Ala Glu Ala Leu Gly Leu Val Val Arg
 225 230 235 240
 Asp Arg Arg Lys Glu Arg Asp Gly Gly Glu Arg Lys Asn Asp Met Leu

 245 250 255
 Glu Ala Leu Phe Glu Gly Asp Gly Val Glu Gly Val Gly Phe Ser Asp
 260 265 270
 Glu Glu Ile Val Asp Phe Ile Leu Ala Leu Leu Val Ala Gly Tyr Glu
 275 280 285
 Thr Thr Ser Thr Ile Met Thr Leu Ala Val Lys Phe Leu Thr Glu Thr
 290 295 300
 Pro Arg Ala Leu Ser Leu Leu Lys Glu Glu His Glu Glu Ile Arg Leu
 305 310 315 320

 Arg Lys Gly Glu Val Lys Ser Leu Gln Trp Glu Asp Tyr Lys Ser Met
 325 330 335
 Pro Phe Thr Gln Cys Val Val Asn Glu Thr Leu Arg Ile Ala Asn Ile
 340 345 350
 Ile Ser Gly Val Phe Arg Arg Ala Met Thr Asp Ile Asn Ile Lys Gly
 355 360 365
 Tyr Thr Ile Pro Lys Gly Trp Lys Val Phe Ala Ser Leu Arg Ala Val
 370 375 380
 His Leu Asp His Glu His Phe Lys Asp Ala Arg Thr Phe Asp Pro Trp

385					390					395					400
Arg	Trp	Gln	Ser	Ser	Ala	Gly	Pro	Thr	Ser	Ser	Pro	Asn	Val	Phe	Thr
				405				410				415			
Pro	Phe	Gly	Gly	Gly	Pro	Arg	Arg	Cys	Pro	Gly	Tyr	Glu	Leu	Ala	Arg
				420				425				430			
Val	Glu	Leu	Ser	Val	Phe	Leu	His	His	Leu	Val	Thr	Arg	Phe	Ser	Trp
				435				440				445			
Val	Pro	Ala	Glu	Ala	Asp	Lys	Leu	Val	Phe	Phe	Pro	Thr	Thr	Arg	Met
450				455				460							

Leu Lys Arg Tyr Pro Ile Asn Val Gln His Arg Ser Leu Phe Glu Gln
465 470 475 480
Lys Glu Glu Glu Lys Gly Ser

485

<210> 7

<211> 1782

<212> DNA

<213> Artificial Sequence

<220><223> nucleotides coding SMT1 of *S. lycopersicum*

<400> 7

tagggggtga aggagacgtg gtataatatg tgacgtggcc acgaaccccc acttctgttt 60

atggtgtctt tcaatTTTTg agtgctggca ttaggcgtat atatatatat gagttgttgc 120

tgtagtgtgt tacagcatat ttgctttttc taaggatccg tctccgtttt ggcattttct 180

tttccgtttt tttccatctt aaagtgttaa gctttctact tttctttttt tcttctgcat 240

tgtaaatctt tacttgattt catcatggggg ttgttttaat cttgtttata gatcttgtga 300

tttgggaaaa gagatgtcaa aacaaggggc ttttgatctg gcatttgggg ttggaggaag 360

aattggcaag gatgaagttc tctctgctgt tgacaaatat gagaagtacc atggttatta 420

tggaaggatgaa gaagacgaga gaaagaataa ctacactgac atggtaaaca aatactatga 480

tctttgcacc agctttctatg aatacggatg gggagagtcc ttccattttg cgcccagatg 540

gaaagaagaa tcactccagg aaagcattaa gaggcattgag cacttccttg ccttgcaact 600

gggattgaaa ccaggacaaa aggtcttggg ttaggatgt ggaattggag ggccgttaag 660

agaaattgct cgattcagct ctacatcagt tacaggcctc aacaacaatg aatatcagat 720

atctcgagga caggtgttga accgcaaagt gggattggat caaacttgca actttgtaaa 780

gggtgatttc atgaaaatgc cattccctga caatagcttt gatgcagtat atgcaataga 840
agctacctgc catgcaccag atccggttgg atgctataga gagatctata ggggtgctgaa 900
gcctgggtcag tgctttgctg tgtatgagtg gtgcatgact gatgcataca atccaaataa 960
tgaagagcag aaaagaatca aggaagaaat tgagctcgga aatggtctcc cggagattag 1020

atcaacacaa cagtccttag aagcagcgag aaaagctggt tttgaagttg tgtgggataa 1080
ggatctcgct gaagactcac cagtttcatg gtacatgcct ttagatacaa gtcacttctc 1140
actcagtagc ttccgcctaa cagcagttgg cagacttttc accagaaatc tggtttcggc 1200
acttgaatat gtgggtgtag ctctctaaagg tagtcaaagg gttcaagctt tcttagaaaa 1260
agccgcagaa ggctttgttg gtggtgccaa gtaagataac aactctactt catacataca 1320
tattagtaga gtatcgaacg ttataatgat acattaactt tttaaacctg tcgttgcctt 1380
ttcaggaaag ggatcttcac accgatgtat ttcttcttgg ttcgcaagcc aatttcagac 1440

tcacagtaat atttgactta gctcttcatt gcttttagtta gcaaactctt gagattcttg 1500
gccttgaact tatatggcca ggctaataca aggttatctt gattttcggt tagggatttg 1560
gggctatttg tttttcttgc attgcttggt tgcaccatt tttgccattt attgttcttc 1620
tcaactccta attgtttgct attttgccat tgaattgatt gttgtcctgg aaaatatgtt 1680
gttgataaag gtcagcattt tgctaaatat atgactaatg acattatggt gatacatctt 1740
cttcttcttc ttgagagttg ataaatatta attattatgt ca 1782

<210> 8
<211> 326
<212> PRT

<213> Artificial Sequence

<220><223> SMT1 of *S. lycopersicum*

<400> 8

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

1 5 10 15

Ile Gly Lys Asp Glu Val Leu Ser Ala Val Asp Lys Tyr Glu Lys Tyr

20 25 30

His Gly Tyr Tyr Gly Gly Glu Glu Asp Glu Arg Lys Asn Asn Tyr Thr

35 40 45

Asp Met Val Asn Lys Tyr Tyr Asp Leu Cys Thr Ser Phe Tyr Glu Tyr

50 55 60

Gly	Trp	Gly	Glu	Ser	Phe	His	Phe	Ala	Pro	Arg	Trp	Lys	Glu	Glu	Ser
65					70					75					80
Leu	Gln	Glu	Ser	Ile	Lys	Arg	His	Glu	His	Phe	Leu	Ala	Leu	Gln	Leu
				85					90					95	
Gly	Leu	Lys	Pro	Gly	Gln	Lys	Val	Leu	Asp	Val	Gly	Cys	Gly	Ile	Gly
			100					105					110		
Gly	Pro	Leu	Arg	Glu	Ile	Ala	Arg	Phe	Ser	Ser	Thr	Ser	Val	Thr	Gly
		115					120					125			
Leu	Asn	Asn	Asn	Glu	Tyr	Gln	Ile	Ser	Arg	Gly	Gln	Val	Leu	Asn	Arg
130						135					140				
Lys	Val	Gly	Leu	Asp	Gln	Thr	Cys	Asn	Phe	Val	Lys	Gly	Asp	Phe	Met
145					150					155					160
Lys	Met	Pro	Phe	Pro	Asp	Asn	Ser	Phe	Asp	Ala	Val	Tyr	Ala	Ile	Glu
				165					170					175	
Ala	Thr	Cys	His	Ala	Pro	Asp	Pro	Val	Gly	Cys	Tyr	Arg	Glu	Ile	Tyr
		180						185					190		
Arg	Val	Leu	Lys	Pro	Gly	Gln	Cys	Phe	Ala	Val	Tyr	Glu	Trp	Cys	Met
	195						200					205			
Thr	Asp	Ala	Tyr	Asn	Pro	Asn	Asn	Glu	Glu	Gln	Lys	Arg	Ile	Lys	Glu
210						215					220				
Glu	Ile	Glu	Leu	Gly	Asn	Gly	Leu	Pro	Glu	Ile	Arg	Ser	Thr	Gln	Gln
225					230					235					240
Cys	Leu	Glu	Ala	Ala	Arg	Lys	Ala	Gly	Phe	Glu	Val	Val	Trp	Asp	Lys
				245					250					255	
Asp	Leu	Ala	Glu	Asp	Ser	Pro	Val	Ser	Trp	Tyr	Met	Pro	Leu	Asp	Thr
		260						265					270		
Ser	His	Phe	Ser	Leu	Ser	Ser	Phe	Arg	Leu	Thr	Ala	Val	Gly	Arg	Leu
		275					280					285			
Phe	Thr	Arg	Asn	Leu	Val	Ser	Ala	Leu	Glu	Tyr	Val	Gly	Val	Ala	Pro
	290						295				300				
Lys	Gly	Ser	Gln	Arg	Val	Gln	Ala	Phe	Leu	Glu	Lys	Ala	Ala	Glu	Gly
305					310					315					320

Leu Val Gly Gly Ala Lys

325

<210> 9

<211> 114

<212> DNA

<213> Artificial Sequence

<220><223> nucleotides coding exon 6 of DWF5 of *S. lycopersicum*

<400> 9

ggatatggagc ttatcctcg tatcgggaaa cacttcgata ttaaagtctt cacaaactgc 60

agatttggtg tgatgtcttg ggcagttctt gctgtcacgt attgcataaa acag 114

<210> 10

<211> 127

<212> DNA

<213> Artificial Sequence

<220><223> nucleotides coding exon 7 of DWF5 of *S. lycopersicum*

<400> 10

catgaagaat atgggcgtgt atctgattcc atgcttgtaa atacgatatt gatgttggtg 60

tatgtcacia aattcttttg gtgggaagct gggtactgga acacaatgga tattgcacat 120

gacagag 127

<210> 11

<211> 126

<212> DNA

<213> Artificial Sequence

<220><223> nucleotides coding exon 9 of DWF5 of *S. lycopersicum*

<400> 11

ctttcccttt acatactagt ggctgggtctt ctctgcatat atatcaacta tgattgtgac 60

aggcagaggc aagagttccg gagaacaaat ggaaaatgca ctgtctgggg aaagactcca 120

tctaag 126

<210> 12

<211> 65

<212> DNA

<213> Artificial Sequence

<220><223> nucleotides coding exon 10 of DWF5 of *S. lycopersicum*

<400> 12

atagtcgccg catacactac tacctctggt gagaaaaaga ccagccttct gcttacatca 60

ggatg 65

<210> 13

<211> 82

<212> DNA

<213> Artificial Sequence

<220><223> nucleotides coding exon 11 of DWF5 of *S. lycopersicum*

<400> 13

gtggggctta gctgccatt tccactatgt tccagaaata ttagcggcat ttttctggag 60

tgtaccagct cttttcaacc ac 82

<210> 14

<211> 131

<212> DNA

<213> Artificial Sequence

<220><223> nucleotides coding exon 7 of CPD of *S. lycopersicum*

<400> 14

agtagtgcag gaccgacaag ctcacctaac gtattcacac catttggtgg tggacctcgt 60

cgatgtcctg gttacgagct tgctagagt gaactctctg ttttccttca ccacttggtg 120

actcgtttta g 131

<210> 15

<211> 255

<212> DNA

<213> Artificial Sequence

<220><223> nucleotides coding exon 8 of CPD of *S. lycopersicum*

<400>

15

ttgggttccg gctgaggcag ataagttagt tttcttccca acgacaagga tgctgaagcg 60

atatccaatt aacgtccaac atcgaagcct gtttgagcag aaagaagaag agaagggttc 120

gtgagattgt agaatgcaga gtggagggag gtgtatgaaa ggtgttaaat aagacaatag 180

tatatagggtg ttagcatata cttagaaaat taaattatct gtagtgaatt gagcagagat 240

taacattcct ctgat 255

<210> 16
 <211> 48
 <212> DNA
 <213> Artificial Sequence
 <220><223> nucleotides coding exon 7 of SMT1 of *S. lycopersicum*

<400> 16
 gtgttgaacc gcaaagtggg attggatcaa acttgcaact ttgttaaag 48

<210> 17
 <211> 117
 <212> DNA
 <213> Artificial Sequence
 <220><223> nucleotides coding exon 8 of SMT1 of *S. lycopersicum*

<400> 17
 gtgggatgct atagagagat ctataggggtg ctgaagcctg gtcagtgcctt tgctgtgtat 60
 gagtggtgca tgactgatgc atacaatcca aataatgaag agcagaaaag aatcaag 117

<210> 18
 <211> 126
 <212> DNA
 <213> Artificial Sequence
 <220><223> nucleotides coding exon 11 of SMT1 of *S. lycopersicum*

<400> 18
 gttgtgtggg ataaggatct cgctgaagac tcaccagttt catggtacat gccttttagat 60
 acaagtcact ttcactcag tagcttccgc ctaacagcag ttggcagact tttcaccaga 120
 aatctg 126

<210> 19
 <211> 98
 <212> DNA
 <213> Artificial Sequence
 <220><223> nucleotides coding exon 12 of SMT1 of *S. lycopersicum*

<400> 19
 gtttcggcac ttgaatatgt ggggttagct cctaaaggta gtcaaagggt tcaagctttc 60
 ttagaaaaag ccgcagaagg tcttgttggg ggtgccaa 98

<210>	20	
<211>	23	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	D5-1 sgRNA	
<400>	20	
	tggagcttta tcctcgatc ggg	23
<210>	21	
<211>	23	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	D5-2 sgRNA	
<400>	21	
	gcagatttgg tatgatgtct tgg	23
<210>	22	
<211>	23	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	D5-3 sgRNA	
<400>	22	
	tgtaaatacg atattgatgt tgg	23
<210>	23	
<211>	23	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	D5-4 sgRNA	
<400>	23	
	gaccagccac tagtatgtaa agg	23
<210>	24	
<211>	23	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	D5-5 sgRNA	
<400>	24	

tgcatTTTcc attgtttctc cgg	23
<210> 25	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> D5-6 sgRNA	
<400> 25	
cctctggtga gaaaaagacc agc	23
<210> 26	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> D5-7 sgRNA	
<400> 26	
agaaagatga cgtagaagta ggg	23
<210> 27	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> C1 sgRNA	
<400> 27	
aacgagtcac caagtgtga agg	23
<210> 28	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> C2 sgRNA	
<400> 28	
cgcttcagca tccttgctgt tgg	23
<210> 29	
<211> 23	
<212> DNA	
<213> Artificial Sequence	

<220><223>	C3 sgRNA	
<400>	29	
tgagcagaaa gaagaagaga agg		23
<210>	30	
<211>	23	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	C4 sgRNA	
<400>	30	
tgagattgta gaatgcagag tgg		23
<210>	31	
<211>	23	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	C5 sgRNA	
<400>	31	
gtagaatgca gagtggaggg agg		23
<210>	32	
<211>	23	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	S1 sgRNA	
<400>	32	
gttgaaccgc aaagtgggat tgg		23
<210>	33	
<211>	23	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	S2 sgRNA	
<400>	33	
gatgctatag agagatctat agg		23
<210>	34	
<211>	23	
<212>	DNA	

<213> Artificial Sequence

<220><223> S4 sgRNA

<400> 34

gtttcggcac ttgaatatgt ggg 23

<210> 35

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> S5 sgRNA

<400> 35

gccgcagaag gtcttgttgg tgg 23

<210> 36

<211> 15954

<212> DNA

<213> Artificial Sequence

<220><223> nucleotides coding D100

<400> 36

aagcttgata atttatttga aaattcataa gaaaagcaaa cgttacatga attgatgaaa 60

caatacaaaag acagataaag ccacgcacat ttaggatatt ggccgagatt actgaatatt 120

gagtaagatc acggaatttc tgacaggagc atgtcttcaa ttcagcccaa atggcagttg 180

aaataactcaa accgccccat atgcaggagc ggatcattca ttgtttgttt ggttgccttt 240

gcccaacatgg gagtccaaga ttctgcagtc aaatctcggt gacgggcagg accggacggg 300

gcggtaccgg caggctgaag tccagctgcc agaaaccac gtcatgccag ttcccgtgct 360

tgaagccggc cggccgcagc atgccgcggg gggcatatcc gagcgcctcg tgcattgcga 420

cgctcgggtc gttgggcagc ccgatgacag cgaccacgct cttgaagccc tgtgcctcca 480

gggacttcag caggtgggtg tagagcgtgg agcccagtc cgtccgctgg tggcgggggg 540

agacgtacac ggtcgactcg gccgtccagt cgtaggcggt gcgtgccttc caggggcccc 600

cgtaggcgat gccggcgacc tcgccgtcca cctcggcgac gagccaggga tagcgtccc 660

gcagacggac gaggtcgtcc gtccactcct gcggttcctg cggctcggta cggaagtga 720

ccgtgcttgt ctgatgtag tggttgacga tggtcagac cgccggcatg tccgcctcgg 780

tggcacggcg gatgtcgcc gggcgtcgtt ctgggctcat cgattcgatt tgggtgatcg 840

agattggtta tgaattcag atgctagtgt aatgtattgg taatttgga agatataata 900

ggaagcaagg ctatttatcc atttctgaaa aggcgaaatg gcgtcaccgc gagcgtcacg	960
cgcatccgt tcttgctgta aagcgttgtt tggtagactt ttgactagcg aggcttggcg	1020
tgtcagcgta tctattcaaa agtcgttaat ggctgcggat caagaaaaag ttggaataga	1080
aacagaatac ccgcgaaatt caggccccgt tgccatgtcc tacacgccga aataaacgac	1140
caaatagta gaaaaataaa aactgactcg gatacttacg tcacgtcttg cgcactgatt	1200
tgaaaaatct cagccatggc tgcaggtcga cgagtcagta ataaacggcg tcaaagtgg	1260
tgcagccggc acacacgagt cgtgtttatc aactcaaagc acaataactt ttcctcaacc	1320
taaaaataag gcaattagcc aaaaacaact ttgcgtgtaa acaacgtca atacacgtgt	1380
cattttatta ttagctattg cttcaccgcc ttagctttct cgtgacctag tcgtcctcgt	1440
ctttcttct tcttcttcta taaaacaata cccaaagagc tcttcttctt cacaattcag	1500
atttcaattt ctcaaatct taaaaacttt ctctcaattc tctctaccgt gatcaaggta	1560
aatttctgtg ttccttattc tctcaaaatc ttcgattttg ttttcgttcg atcccaattt	1620
cgtatatgtt ctttggttta gattctgtta atcttagatc gaagacgatt ttcctgggtt	1680
gatcgtaga tatcatctta attctcgatt aggggtttcat agatatacgc cgatttgttc	1740
aaataatttg agttttgtcg aataattact cttcgatttg tgatttctat ctagatctgg	1800
tgtagtttc tagtttgtgc gatcgaattt gtcgattaat ctgagttttt ctgattaaca	1860
gccatggact ataaggacca cgacggagac tacaaggatc atgatattga ttacaaagac	1920
gatgacgata agatggcccc aaagaagaag cggaaggtcg gtatccacgg agtcccagca	1980
gccgacaaga agtacagcat cggcctggac atcggcacca actctgtggg ctgggccgtg	2040
atcacgcagc agtacaaggt gcccgcaag aaattcaagg tgcctgggcaa caccgaccgg	2100
cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag	2160
gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccgcatctgc	2220
tatctgcaag agatcttcag caacgagatg gccaaagtg acgacagctt cttccacaga	2280
ctggaagagt ctttcttggg ggaagaggat aagaagcacg agcggcaccc catcttcggc	2340
aacatcgtgg acgaggtggc ctaccacgag aagtaccca ccatctacca cctgagaaag	2400
aaactggtgg acagcaccga caagccgac ctgcggctga tctatctggc cctggccac	2460
atgatcaagt tccggggcca cttctgatc gagggcgacc tgaaccccga caacagcgac	2520
gtggacaagc tgttcatcca gctgtgtcag acctacaacc agctgttcga ggaaaacccc	2580
atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga	2640

cggttgaaa atctgatcgc ccagctgccc ggcgagaaga agaattggcct gttcggaac	2700
ctgattgccc ttagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag	2760
gatgcaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc	2820
cagatcggcg accagtagcg cgacctgttt ctggccgcca agaacctgtc cgacgccatc	2880
ctgctgagcg acatcctgag agtgaacacc gagatcacca agggccccct gagcgctctt	2940
atgatcaaga gatacgacga gcaccaccag gacctgaccc tgcagaaagc tctcgtgcgg	3000
cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc	3060
ggctacattg acggcggagc cagccaggaa gatttctaca agttcatcaa gcccatcctg	3120
gaaaagatgg acggcacca ggaactgttc gtgaagctga acagagagga cctgctgcgg	3180
aagcagcgga cttcgacaa cggcagcatc cccaccaga tccacctggg agagctgcac	3240
gccattctgc ggcgcgagga agatttttac ccattctga aggacaaccg gaaaaagatc	3300
gagaagatcc tgaccttcg catccctac tacgtgggcc ctctggccag gggaaacagc	3360
agattcgctt ggatgaccag aaagagcgag gaaacatca cccctggaa cttcgaggaa	3420
gtggtggaca agggcgcttc cgcccagagc ttcacgagc ggatgacaa cttcgataag	3480
aacctgccc acgagaaggt gctgccaag cacagcctgc tgtacagta cttcacctg	3540
tataacgagc tgaccaaagt gaaatcgtg accgagggaa tgagaaagcc cgccttctg	3600
agcggcgagc agaaaaagc catcgtggac ctgctgttca agaccaaccg gaaagtgacc	3660
gtgaagcagc tgaagagga ctacttcaag aaatcgagt gcttcgactc cgtggaaatc	3720
tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt	3780
atcaaggaca aggacttctt ggacaatgag gaaaacgagg acattctgga agatatcgtg	3840
ctgacctga cactgtttga ggacagagag atgatcagg aacggctgaa aacctatgcc	3900
cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc	3960
aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg	4020
gatttctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac	4080
agcctgacct ttaaagagga catccagaaa gccaggtgt cggccaggc cgatagcctg	4140
cacgagcaca ttgccaatct ggccggcagc cccgccatta agaaggcat cctgcagaca	4200
gtgaaggtgg tggacgagct cgtgaaagt atgggcccgc acaagcccga gaacatcgtg	4260
atcgaatgg ccagagagaa ccagaccacc cagaaggac agaagaacag ccgcgagaga	4320
atgaagcgga tcgaagagg catcaaagag ctgggcagcc agatcctgaa agaacacccc	4380

gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg	4440
gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat	4500
atcgtgcctc agagctttct gaaggacgac tccatcgaca acaagggtgt gaccagaagc	4560
gacaagaacc ggggcaagag cgacaacgtg cctccgaag aggtcgtgaa gaagatgaag	4620
aactactggc ggcagctgct gaacccaag ctgattaccc agagaaagt t cgacaatctg	4680
accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggttcat caagagacag	4740
ctggtggaaa cccggcagat cacaagcac gtggcacaga tccctggactc ccggtgaac	4800
actaagtacg acgagaatga caagctgac cggaagtga aagtgtcac cctgaagtcc	4860
aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgca gatcaacaac	4920
taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag	4980
taccctaagc tggaaagcga gtctgtgtac ggcgactaca aggtgtacga cgtgcggaag	5040
atgatcgcca agagcgagca ggaaatcggc aaggctaccg ccaagtactt ctctacagc	5100
aacatcatga actttttcaa gaccgagatt acctggcca acggcgagat ccggaagcgg	5160
cctctgatcg agacaaacgg cgaaaccggg gagatcgtgt gggataaggg ccgggatttt	5220
gccaccgtgc ggaaagtgt gagcatgccc caagtgaata tcgtgaaaa gaccgaggtg	5280
cagacaggcg gcttcagcaa agagtctatc ctgcccaaga ggaacagcga taagctgatc	5340
gccagaaaga aggactggga ccctaagaag tacggcggct tcgacagccc caccgtggcc	5400
tattctgtgc tgggtggtgc caaagtggaa aaggcgaagt ccaagaaact gaagagtgtg	5460
aaagagctgc tggggatcac catcatggaa agaagcagct tcgagaagaa tcccatcgac	5520
tttctggaag ccaagggtc caaagaagtg aaaaaggacc tgatcatcaa gctgcctaag	5580
tactcctgt tcgagctgga aaacggccgg aagagaatgc tggcctctgc cggcgaactg	5640
cagaaggga acgaactggc cctgccctcc aaatatgtga acttctgtga cctggccagc	5700
cactatgaga agctgaaggg ctccccgag gataatgagc agaaacagct gtttgtggaa	5760
cagcacaagc actacctgga cgagatcatc gagcagatca gcgagttctc caagagagtg	5820
atcctggccg acgctaactt ggacaaagtg ctgtccgcct acaacaagca ccgggataag	5880
cccatcagag agcaggccga gaatatcatc cacctgttta cctgaccaa tctgggagcc	5940
cctgccgcct tcaagtactt tgacaccacc atcgaccgga agaggtagac cagcaccaaa	6000
gaggtgctgg acgccacct gatccaccag agcatcaccg gcctgtacga gacacggatc	6060
gacctgtctc agctgggagg cgacaaaagg ccggcgcca cgaaaaaggc cggccaggca	6120
aaaaagaaaa agtaagaatt gcctcttctt cgttgaacaa cggaactcg acttgccttc	6180
cgcacaatac atcatttctt cttagctttt tttcttctt ttcgttcata cagttttttt	6240

ttgtttatca gcttacattt tcttgaaccg tagctttcgt tttcttcttt ttaactttcc	6300
attcggagtt ttgtatctt gtttcatagt ttgtccagg attagaatga ttaggcacg	6360
aaccttcaag aatttgattg aataaaacat cttcattctt aagatatgaa gataatcttc	6420
aaaaggcccc tgggaatctg aaagaagaga agcaggccca tttatatggg aaagaacaat	6480
agtatttctt atataggccc atttaagttg aaaacaatct tcaaaagtcc cacatcgctt	6540
agataagaaa acgaagctga gtttatatac agctagagtc gaagtagtga ttgtgagacc	6600
gaggtctcgg ttttagagct agaaatagca agttaaata aggctagtcc gttatcaact	6660
tgaaaaagtg gcaccgagtc ggtgcaacaa agcaccagtg gtctagtggg agaatagtag	6720
cctgccacgg tacagaccgg ggttcgattc ccggctggg cagcagattt ggtatgatgt	6780
ctgttttaga gctagaaata gcaagttaaa ataaggctag tccgttatca acttgaaaaa	6840
gtggcaccga gtcggtgcaa caaagcacca gtggtctagt ggtagaatag taccctgcca	6900
cggtacagac ccgggttcga ttcccggctg gtgcatggag ctttatcctc gtatcgtttt	6960
agagctagaa atagcaagtt aaaataaggc tagtccgtta tcaacttgaa aaagtggcac	7020
cgagtcggtg ctttttttct agaggcgcaa ttcactggcc gtcgttttac aacgtcgtga	7080
ctgggaaaac ctggcgta cccaacttaa tcgcccttga gcacatcccc ctttcgccag	7140
ctggcgtaat agcgaagagg ccgcaccga tcgcccttcc caacagttgc gcagcctgaa	7200
tggcgaatgc tagagcagct tgagcttgga tcagattgtc gtttccgcc ttcagtttaa	7260
actatcagtg tttagacagga tatattggcg ggtaaaccta agagaaaaga gcgtttatta	7320
gaataacgga tatttaaaag ggcgtgaaaa ggtttatccg ttcgtccatt tgtatgtgca	7380
tgccaaaccac agggttcccc tcgggatcaa agtactttga tccaaccctt ccgctgctat	7440
agtgcagtcg gcttctgacg ttcagtgcag ccgtcttctg aaaacgacat gtgcacaag	7500
tcctaagtta cgcgacagcg tgccgccctg cctttttcct ggcgttttct tgtcgcgtgt	7560
tttagtcgca taaagtagaa tacttgcgac tagaaccgga gacattacgc catgaacaag	7620
agcgccgccg ctggcctgct gggctatgcc cgcgtcagca ccgacgacca ggacttgacc	7680
aaccaacggg ccgaactgca cgcggccggc tgcaccaagc tgttttcga gaagatcacc	7740
ggcaccaggc gcgaccgccc ggagctggcc aggatgcttg accacctacg ccctggcgac	7800
gttgtgacag tgaccaggct agaccgcctg gccgcagca cccgcgacct actggacatt	7860
gccgagcgca tccaggaggc cggcgcgggc ctgcgtagcc tggcagagcc gtgggcccag	7920
accaccacgc cggccggccg catggtgttg accgtgttcg ccggcattgc cgagttcgag	7980
cgttccctaa tcatcgaccg caccgggagc gggcgcgagg ccgccaaggc ccgaggcgtg	8040
aagtttggcc ccgcacctac cctacccccg gcacagatcg cgcacgcccg cgagctgatc	8100

gaccaggaag gccgcacgt gaaagaggcg gctgcactgc ttggcgtgca tcgctcgacc	8160
ctgtaccgcg cacttgagcg cagcgaggaa gtgacgcca ccgaggccag gcggcgcggt	8220
gccttccgtg aggacgcatt gaccgaggcc gacgccctgg cggccgccga gaatgaacgc	8280
caagaggaac aagcatgaaa ccgcaccagg acggccagga cgaaccgttt ttcattaccg	8340
aagagatcga ggccggagatg atcgcgggcg ggtacgtgtt cgagccgcc gcgcacgtct	8400
caaccgtgcg gctgcatgaa atcctggccg gtttgtctga tgccaagctg gcggcctggc	8460
cggccagctt ggccgctgaa gaaaccgagc gccgccgtct aaaaaggtga tgtgtatttg	8520
agtaaacacag cttgcgtcat gcggtcgctg cgtatatgat gcgatgagta aataaacaaa	8580
tacgcaaggg gaacgcatga aggttatcgc tgtacttaac cagaaaggcg ggtcaggcaa	8640
gacgaccatc gcaaccatc tagcccgcc cctgcaactc gccggggccg atgttctgtt	8700
agtcgattcc gatccccagg gcagtgcccg cgattggcg gccgtgcggg aagatcaacc	8760
gctaaccgtt gtcggcatcg accgcccac gattgaccgc gacgtgaagg ccatcgcccg	8820
gcgcgacttc gtagtgatcg acggagcgcc ccaggcgccg gacttggtg tgtcccgat	8880
caaggcagcc gacttcgtgc tgattccggt gcagccaagc cttacgaca tatgggccac	8940
cgccgacctg gtggagctgg ttaagcagcg cattgaggtc acggatggaa ggctacaagc	9000
ggcctttgtc gtgtcgcggg cgatcaaagg cacgcgcatc ggcggtgagg ttgccgaggc	9060
gttgccgggg tacgagctgc ccattcttga gtcccgtatc acgcagcgcg tgagctaccc	9120
aggcactgcc gccgccgga caaccgttct tgaatcagaa cccgagggcg acgctgcccg	9180
cgaggtcag gcgctggccg ctgaaattaa atcaaaactc atttgagtta atgaggtaaa	9240
gagaaaatga gaaaagcac aaacacgcta agtgccggcc gtccgagcgc acgcagcagc	9300
aaggctgcaa cgttggccag cctggcagac acgccagcca tgaagcgggt caactttcag	9360
ttgccggcgg aggatcacac caagctgaag atgtacgagg tacgccaagg caagaccatt	9420
accgagctgc tatctgaata catcgcgag ctaccagagt aaatgagcaa atgaataaat	9480
gagtagatga attttagcgg ctaaaggagg cgcatggaa aatcaagaac aaccaggcac	9540
cgacccgtg gaatgcccc tgttggagg aacgggcgggt tggccaggcg taagcggctg	9600
ggttgtctgc cgccctgca atggcactgg aacccccaa cccaggaat cggcgtgacg	9660
gtcgcaaac atccggccc gtacaaatcg gcgcggcgct gggtagtgac ctggtggaga	9720
agttgaaggc cgcgcaggcc gccagcggc aacgcatcga ggcaagaagca cggccggctg	9780
aatcgtggca agcggccgct gatcgaatcc gaaaagaatc ccggaaccg ccggcagccg	9840

gtgcgccgtc gattaggaag ccgccaag ggcagagca accagatttt ttcgttccga	9900
tgctctatga cgtgggcacc cgcgatagtc gcagcatcat ggacgtggcc gttttccgtc	9960
tgctgaagcg tgaccgacga gctggcgagg tgatccgcta cgagcttcca gacgggcacg	10020
tagaggtttc cgcagggccg gccggcatgg ccagtgtgtg ggattacgac ctggtactga	10080
tggcggtttc ccatctaacc gaatccatga accgataccg ggaagggaag ggagacaagc	10140
ccggccgcgt gttccgtcca cacgttgcgg acgtactcaa gttctgccgg cgagccgatg	10200
gcggaaagca gaaagacgac ctggtagaaa cctgcattcg gttaaacc acgcacgttg	10260
ccatgcagcg tacgaagaag gccaaagacg gccgcctggt gacggtatcc gaggtgaag	10320
ccttgattag ccgtacaag atcgtaaaga gcgaaaccgg gcggccggag tacatcgaga	10380
tcgagctagc tgattggatg taccgcgaga tcacagaagg caagaaccg gacgtgctga	10440
cggttcacc cgattacttt ttgatcgatc ccggcatcgg ccgttttctc taccgcctgg	10500
cacgccgcgc cgcaggcaag gcagaagcca gatggttgtt caagacgac tacgaacgca	10560
gtggcagcgc cggagagttc aagaagttct gtttcaccgt gcgcaagctg atcgggtcaa	10620
atgacctgcc ggagtacgat ttgaaggagg aggcggggca ggctggcccg atcctagtca	10680
tgcgctaccg caacctgac gagggcgaag catccgccgg ttcctaattg acggagcaga	10740
tgctagggca aattgccta gcaggggaaa aaggtcgaaa agcactcttt cctgtggata	10800
gcacgtacat tgggaaccca aagccgtaca ttgggaaccg gaaccgtac attgggaacc	10860
caaagccgta cattgggaac cggtcacaca tgtaagtac tgatataaaa gagaaaaag	10920
gcgatttttc cgcctaaaac tctttaaaac ttattaaaac tcttaaaacc cgcctggcct	10980
gtgcataact gtctggccag cgcacagccg aagagctgca aaaagcgcct acccttcggt	11040
cgctgcgtc cctacgcccc gccgcttcgc gtccgcctat cgcggccgct ggccgctcaa	11100
aatggctgg cctacggcca ggcaatctac cagggcgcgg acaagccgcg ccgtgccac	11160
tcgaccgccg gcgccacat caaggcacc tgccctgcgc gtttcggtga tgacgtgaa	11220
aacctctgac acatgcagct cccggagacg gtcacagctt gtctgtaagc ggatgccggg	11280
agcagacaag cccgtcaggc cgcgtcagcg ggtgttggcg ggtgtcgggg cgcagccatg	11340
accagtcac gtagcgatag cggagtgtat actggcttaa ctatcggca tcagagcaga	11400
ttgtactgag agtgcaccat atgcggtgtg aaataccgca cagatgcgta aggagaaaat	11460
accgcatcag gcgtcttcc gcttctcgc tcaactgactc gctgcgtcg gtcgttcggc	11520
tgccggcagc ggtatcagct cactcaaagg cgtaataacg gttatccaca gaatcagggg	11580
ataacgcagg aaagaacatg tgagcaaaa ggcagcaaaa ggccaggaac cgtaaaaagg	11640
ccgcgttgct ggcgtttttc cataggtccc gccccctga cgagcatcac aaaaatcgac	11700

gctcaagtca gaggtggcga aacccgacag gactataaag ataccaggcg ttccccctg	11760
gaagtcacct cgtgcgtctt cctgttccga cctgcccgt taccggatac ctgtccgct	11820
tttccccctt gggaagcgtg gcgctttctc atagctcacg ctgtaggat ctgagttcgg	11880
tgtaggctgt tcgtccaag ctgggctgtg tgcacgaacc ccccgttcag cccgaccgt	11940
gcgccttacc cggtaactat cgtcttgagt ccaacccggt aagacacgac ttatcgccac	12000
tggcagcagc cactggtaac aggattagca gagcgaggta ttaggcggt gctacagagt	12060
tcttgaagtg gtggcctaac tacggctaca ctagaaggac agtatattgt atctgcgtc	12120
tgttgaagcc agttacctt ggaaaaagag ttggtagctc ttgatccggc aaacaaacca	12180
ccgctggtag cgggtggttt tttgtttgca agcagcagat tacgcgcaga aaaaaaggat	12240
ctcaagaaga tcctttgatc tttctacgg ggtctgacgc tcagtgaac gaaaactcac	12300
gttaagggat tttggtcatg cattctaggt actaaaacaa ttcattccagt aaaatataat	12360
atattatttt ctccaatca ggcttgatcc ccagtaagtc aaaaaatagc tcgacatact	12420
gttcttcccc gatatactcc ctgatcgacc ggacgcagaa ggcaatgtca taccacttgt	12480
ccgcctgcc gcttctccca agatcaataa agccacttac tttgccatct ttcacaaaga	12540
tgttgctgtc tcccaggctg ccgtgggaaa agacaagttc ctcttcgggc tttccgtct	12600
ttaaaaaatc atacagctcg cgcggatctt taaatggagt gtcttcttcc cagttttcgc	12660
aatccacatc ggccagatcg ttattcagta agtaatccaa ttcggctaag cggctgtcta	12720
agctattcgt atagggacaa tccgatatgt cgatggagtg aaagagcctg atgcactccg	12780
catacagctc gataatcttt tcagggtttt gttcatcttc ataactttcc gagcaaagga	12840
cgccatcggc ctactcatg agcagattgc tccagccatc atgccgttca aagtgcagga	12900
cctttggaac aggcagcttt cttccagcc atagcatcat gtccttttcc cgttccacat	12960
catagtggt cctttatac cggctgtccg tcatttttaa atataggttt tcattttctc	13020
ccaccagctt atatactta gcaggagaca ttccttccgt atcttttacg cagcgttatt	13080
tttcgatcag tttttcaat tccggtgata ttctatttt agccatttat tatttccttc	13140
ctcttttcta cagtatttaa agatacceca agaagctaata tataacaaga cgaactccaa	13200
ttcactgttc cttgcattct aaaaccttaa ataccagaaa acagcttttt caaagtgtt	13260
ttcaaagttg gcgtataaca tagtatcgac ggagccgatt ttgaaaccgc ggtgatcaca	13320
ggcagcaacg ctctgtcatc gttacaatca acatgctacc ctccgcgaga tcatccgtgt	13380
ttcaaaccgg gcagcttagt tgccgttctt ccgaatagca tcggtaacat gagcaaagtc	13440
tgcgcctta caacggtct cccgtgacg ccgtcccga ctgatgggt gcctgtatcg	13500
agtgttgatt ttgtccgag ctgccggtcg gggagctgtt ggctggctgg tggcaggata	13560

tattgtggtg taaacaaatt gacgcttaga caacttaata acacattgcg gacgttttta	13620
atgtactgaa ttaacgccga attaatcgg gggatctgga ttttagtact ggattttggt	13680
tttaggaatt agaaatttta ttgatagaag tattttaca atacaatac atactaaggg	13740
tttcttataat gctcaacaca tgagcgaaac cctataggaa ccctaattcc cttatctggg	13800
aactactcac acattattat ggagaaactc gagcttgctg atcgacagat ccggtcggca	13860
tctactctat ttctttgcc tccgacgagt gctggggcgt cggtttcac tatcggcgag	13920
tacttctaca cagccatcgg tccagacggc cgcgcttctg cgggcgattt gtgtacgcc	13980
gacagtcccc gctccggatc ggacgattgc gtcgcatcga ccctgcgccc aagctgcatc	14040
atcgaaattg ccgtcaacca agctctgata gagttggtca agaccaatgc ggagcatata	14100
cggccggagt cgtggcgatc ctgcaagctc cggatgcctc cgctcgaagt agcgcgtctg	14160
ctgctccata caagccaacc accggcctcca gaagaagatg ttggcgacct cgtattggga	14220
atccccgaac atcgccctgc tccagtcaat gaccgctgtt atcgggccat tgtccgtcag	14280
gacattgttg gagccgaat ccgctgcac gaggtgccgg acttcggggc agtcctcggc	14340
ccaaagcatc agctcatcga gagcctgcgc gacggacgca ctgacggtgt cgtccatcac	14400
agtttgccag tgatacacat ggggatcagc aatcgcgcat atgaaatcac gccatgtagt	14460
gtattgaccg attccttgcg gtccgaatgg gccgaaccg ctcgtctggc taagatcggc	14520
cgcagcgatc gcatccatag cctccgcgac cggttgtaga acagcgggca gttcggtttc	14580
aggcaggtct tgcaacgtga caccctgtgc accggcgggag atgcaatagg tcaggctctc	14640
gctaaactcc ccaatgtcaa gcatttccgg aatcgggagc gcggccgatg caaagtgccg	14700
ataaacataa cgatctttgt agaaaccatc ggcgacgcta ttaccgcga ggacatatcc	14760
acgccctcct acatcgaagc tgaaagcacg agattcttcg ccctccgaga gctgcatcag	14820
gtcggagacg ctgtcgaact ttctgatcag aaacttctcg acagacgtcg cggtagttc	14880
aggctttttc atatctcatt gcccccgagg atctgcgaaa gctcgagaga gatagatttg	14940
tagagagaga ctggtgatit cagcgtgtcc tctccaaatg aaatgaactt ctttatatag	15000
aggaaggctt tgcgaaggat agtgggattg tgcgtcatcc cttacgtcag tggagatc	15060
acatcaatcc acttgctttg aagacgtggt tggaacgtct tctttttcca cgatgctcct	15120
cgtgggtggg ggtccatctt tgggaccact gtcggcagag gcattctgaa cgatagcctt	15180
tcctttatcg caatgatggc attttaggt gccaccttcc ttttctactg tccttttgat	15240
gaagtgcagc atagctgggc aatggaatcc gaggaggttt cccgatatta ccctttgttg	15300

aaaagtcctca atagcccttt ggtcttctga gactgtatct ttgatattct tggagtagac 15360
gagagtgtcg tgctccacca tgttatcaca tcaatccact tgctttgaag acgtggttgg 15420
aacgtcttct ttttccacca tgctcctcgt ggggtgggggt ccatctttgg gaccactgtc 15480
ggcagaggca tcttgaacga tagcctttcc tttatcgcaa tgatggcatt ttaggtgcc 15540
accttcttt tctactgtcc ttttgatgaa gtgacagata gctgggcaat ggaatccgag 15600
gaggtttccc gatattacce tttgttga aa agtctcaata gccctttggt ctctgagac 15660
tgtatctttg atattcttgg agtagacgag agtgtcgtgc tccaccatgt tggcaagctg 15720

ctctagccaa tacgcaaacc gcctctcccc gcgcgttggc cgattcatta atgcagctgg 15780
cacgacaggt ttcccgactg gaaagcgggc agtgagcgca acgcaattaa tgtgagttag 15840
ctcactcatt aggcacccca ggctttacac tttatgcttc cggtcgtat gttgtgtgga 15900
attgtgagcg gataacaatt tcacacagga aacagctatg accatgatta cgcc 15954

<210> 37

<211> 16300

<212> DNA

<213> Artificial Sequence

<220><223> nucleotides coding D120

<400> 37

aagcttgata atttatttga aaattcataa gaaaagcaaa cgttacatga attgatgaaa 60

caatacaaaag acagataaag ccacgcacat ttaggatatt ggccgagatt actgaatatt 120
gagtaagatc acggaatttc tgacaggagc atgtcttcaa ttcagcccaa atggcagttg 180
aaatactcaa accgccccat atgcaggagc ggatcattca ttgtttgttt ggttgccctt 240
gccaacatgg gagtccaaga ttctgcagtc aaatctcggt gacgggcagg accggacggg 300
gcggtaccgg caggctgaag tccagctgcc agaaaccac gtcatgccag ttcccgtgct 360
tgaagccggc cgcccgagc atgcccgggg gggcatatcc gagcgcctcg tgcatgcgca 420
cgctcgggtc gttgggcagc ccgatgacag cgaccacgt cttgaagccc tgtgcctcca 480

gggacttcag caggtgggtg tagagcgtgg agcccagtcc cgtccgctgg tggcgggggg 540
agacgtacac ggtcgactcg gccgtccagt cgtaggcgtt gcgtgccttc caggggcccc 600
cgtaggcgat gccggcgacc tcgccgtcca cctcggcgac gagccaggga tagcgtccc 660
gcagacggac gaggtcgtcc gtccactcct gcggttcctg cggctcggta cggaagtga 720
ccgtgcttgt ctgatgtag tggttgacga tggtcagac cgccggcatg tccgcctcgg 780
tggcacggcg gatgtcgccc gggcgtcgtt ctgggctcat cgattcgatt tgggtgatcg 840

agattggtta tgaattcag atgctagtgt aatgtattgg taatttggga agatataata	900
ggaagcaagg ctatttatcc atttctgaaa aggcgaaatg gcgtcaccgc gagcgtcacg	960
cgcatccgt tcttctgta aagcgttgtt tggtagactt ttgactagcg aggcttggcg	1020
tgtcagcgta tctattcaaa agtcgttaat ggctgcggat caagaaaaag ttggaataga	1080
aacagaatac ccgcgaaatt caggcccggt tgccatgtcc tacacgccga aataaacgac	1140
caaattagta gaaaaataaa aactgactcg gatacttacg tcacgtcttg cgcactgatt	1200
tgaaaaatct cagccatggc tgcaggtcga cgagtcagta ataacggcg tcaaagtgg	1260
tgcagccggc acacacgagt cgtgtttatc aactcaaagc acaataactt ttcctcaacc	1320
taaaaataag gcaattagcc aaaaacaact ttgcgtgtaa acaacgtca atacacgtgt	1380
cattttatta ttagctattg cttcaccgcc ttagctttct cgtgacctag tcgtcctcgt	1440
cttttcttct tcttcttcta taaaacaata ccaaagagc tcttcttctt cacaattcag	1500
atttcaattt ctcaaatct taaaacttt cttcaattc tctctaccgt gatcaaggta	1560
aatttctgtg ttccttattc tctcaaatc ttcgattttg ttttcgttcg atcccaattt	1620
cgtatatgtt ctttggttta gattctgtta atcttagatc gaagacgatt tcttgggttt	1680
gatcgtaga tatcatctta attctcgatt aggggttcat agatatcatc cgatttgttc	1740
aaataatttg agttttgtcg aataattact cttcgatttg tgatttctat ctagatctgg	1800
tgtagtttc tagtttgtc gatcgaattt gtcgattaat ctgagttttt ctgattaaca	1860
gccatggact ataaggacca cgacggagac tacaaggatc atgatattga ttacaaagac	1920
gatgacgata agatggcccc aaagaagaag cggaaggtcg gtatccacgg agtcccagca	1980
gccgacaaga agtacagcat cggcctggac atcggcacca actctgtggg ctgggccgtg	2040
atcaccgacg agtacagggt gccagcaag aaattcaagg tgctgggcaa caccgaccgg	2100
cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag	2160
gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccgatctgc	2220
tatctgcaag agatcttcag caacgagatg gccagggtgg acgacagctt cttccacaga	2280
ctggaagagt ctttcttgtt ggaagaggat aagaagcacg agcggcaccc catcttcggc	2340
aacatcgtgg acgaggtggc ctaccacgag aagtaccca ccatctacca cctgagaaag	2400
aaactggtgg acagaccga caaggccgac ctgcggctga tctatctggc cctggccac	2460
atgatcaagt tccggggcca cttctgatac gagggcgacc tgaaccccga caacagcgac	2520
gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc	2580

atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga	2640
cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct gttcggaac	2700
ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag	2760
gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc	2820
cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc	2880
ctgctgagcg acatcctgag agtgaacacc gagatcacca agggccccct gagcgctct	2940
atgatcaaga gatacgacga gcaccaccag gacctgaccc tgcigaaagc tctcgtgcgg	3000
cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggtacgcc	3060
ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccacctg	3120
gaaaagatgg acggcaccga ggaactgtc gtgaagctga acagagagga cctgctgcgg	3180
aagcagcgga ctttcgacaa cggcagcatc cccccaccga tccacctggg agagctgcac	3240
gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc	3300
gagaagatcc tgaccttccg catccctac tacgtgggcc ctctggccag gggaaacagc	3360
agattcgct ggatgaccag aaagagcgag gaaacatca cccctggaa cttcgaggaa	3420
gtggtggaca agggcgcttc cgcccagagc ttcacgagc ggatgaccaa cttcgataag	3480
aacctgccc acgagaaggt gctgcccag cacagcctgc tgtacgagta cttaccctg	3540
tataacgagc tgaccaaagt gaaatcgtg accgagggaa tgagaaagcc cgccttctg	3600
agcggcgagc agaaaaagc catcgtggac ctgctgttca agaccaaccg gaaagtgacc	3660
gtgaagcagc tgaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc	3720
tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt	3780
atcaaggaca aggacttctt ggacaatgag gaaaacgagg acattctgga agatatctg	3840
ctgacctga cactgtttga ggacagagag atgatcagg aacggctgaa aacctatgcc	3900
cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc	3960
aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg	4020
gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac	4080
agcctgacct ttaaagagga catccagaaa gccaggtgt cggccaggc cgatagcctg	4140
cacgagcaca ttgccaatct ggccggcagc cccgccatta agaaggcat cctgcagaca	4200
gtgaagggtg tggacgagct cgtgaaagt atgggccggc acaagcccga gaacatcgtg	4260
atcgaatgg ccagagagaa ccagaccacc cagaaggac agaagaacag ccgagagaga	4320
atgaagcgga tcgaagagg catcaaagag ctgggcagcc agatcctgaa agaacacccc	4380
gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatggcg	4440

gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat	4500
atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc	4560
gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag	4620
aactactggc ggcagctgct gaacgccaag ctgattaccc agagaaagt t cgacaatctg	4680
accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag	4740
ctggtggaaa cccggcagat cacaagcac gtggcacaga tcttggaact ccgcatgaac	4800
actaagtacg acgagaatga caagctgac cggaagtga aagtgtcac cctgaagtcc	4860
aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac	4920
taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgcct gatcaaaaag	4980
taccctaagc tggaaagcga gtctgtgtac ggcgactaca aggtgtacga cgtgcggaag	5040
atgatcgcca agagcgagca ggaaatcggc aaggctaccg ccaagtactt ctctacagc	5100
aacatcatga actttttcaa gaccgagatt accctggcca acggcgagat ccggaagcgg	5160
cctctgatcg agacaaacgg cgaaaccggg gagatcgtgt gggataaggg ccgggatttt	5220
gccacgtgc ggaaagtgt gagcatgcc caagtgaata tcgtgaaaaa gaccgaggtg	5280
cagacaggcg gcttcagcaa agagtctatc ctcccaaga ggaacagcga taagctgatc	5340
gccagaaaga aggactggga ccctaagaag tacggcggct tcgacagccc caccgtggcc	5400
tattctgtgc tgggtgtggc caaagtggaa aaggcgaagt ccaagaaact gaagagtgtg	5460
aaagagctgc tggggatcac catcatggaa agaagcagct tcgagaagaa tccatcgac	5520
tttctggaag ccaagggeta caaagaagtg aaaaaggacc tgatcatcaa gctgcctaag	5580
tactccctgt tcgagctgga aaacggccgg aagagaatgc tggcctctgc cggcgaactg	5640
cagaagggaa acgaactggc cctgccctcc aaatatgtga acttctgtga cctggccagc	5700
cactatgaga agctgaaggg ctccccgag gataatgagc agaaacagct gtttgtggaa	5760
cagcacaagc actacctgga cgagatcatc gagcagatca gcgagtctc caagagagtg	5820
atcctggccg acgctaactt ggacaaagtg ctgtccgctt acaacaagca ccgggataag	5880
cccatcagag agcaggccga gaatatcatc cacctgttta ccctgaccaa tctgggagcc	5940
cctgccgct tcaagtactt tgacaccacc atcgaccgga agaggtacac cagcaccaaa	6000
gaggtgctgg acgccacct gatccaccag agcatcaccg gcctgtacga gacacggatc	6060
gacctgtctc agctgggagg cgacaaaagg ccggcgccca cgaaaaaggc cggccaggca	6120
aaaaagaaaa agtaagaatt cgcccttctt cgttgaacaa cggaactcg acttgccttc	6180
cgacaatac atcatttctt cttagctttt tttcttctt ttcgttcata cagttttttt	6240
ttgtttatca gcttacattt tcttgaaccg tagctttcgt tttcttctt ttaactttcc	6300

attcggagtt tttgtatctt gtttcatagt ttgtcccagg attagaatga ttaggcatcg	6360
aaccttcaag aatttgattg aataaaacat ctfcattctt aagatatgaa gataatcttc	6420
aaaaggcccc tgggaatctg aaagaagaga agcaggccca tttatatggg aaagaacaat	6480
agtatttctt atataggccc atttaagttg aaaacaatct tcaaaagtcc cacatcgctt	6540
agataagaaa acgaagctga gtttatatac agctagagtc gaagtagtga ttgtgagacc	6600
gaggtctcgg ttttagagct agaaatagca agttaaata aggctagtcc gttatcaact	6660
tgaaaaagtg gcaccgagtc ggtgcaaca agcaccagtg gtctagtggg agaatagtac	6720
cctgccacgg tacagaccgg ggttcgattc cggctgggtg catgagcaga aagaagaaga	6780
gagttttaga gctagaaata gcaagttaaa ataaggctag tccgttatca acttgaaaaa	6840
gtggcaccga gtcggtgcaa caaagcacca gtggtctagt ggtagaatag taccctgcca	6900
cggtagacac ccgggttcga ttcccggctg gtgcacgctt cagcatcctt gtcgtgtttt	6960
agagctagaa atagcaagtt aaaataaggc tagtccgtta tcaacttgaa aaagtggcac	7020
cgagtgggtg caacaaagca ccagtgggtc agtggtagaa tagtaccctg ccacggtaca	7080
gacccgggtt cgattcccgg ctggtgcagc agatttggta tgatgtctgt tttagagcta	7140
gaaatagcaa gttaaaataa ggctagtccg ttatcaactt gaaaaagtgg caccgagtcg	7200
gtgcaacaaa gcaccagtg tctagtggta gaatagtacc ctgccacggt acagaccgg	7260
gttcgattcc cggctgggtc atggagcttt atcctcgtat cgttttagag ctagaaatag	7320
caagttaaaa taaggctagt ccgttatcaa ctfgaaaaag tggcaccgag tcggtgcttt	7380
ttttctagag ggcaattca ctggccgtcg ttttacaacg tcgtgactgg gaaaaccctg	7440
gcgttaccga acttaatcgc cttgcagcac atccccctt cgcagctgg cgtaatagcg	7500
aagaggcccc caccgatcgc ctttcccaac agttgcgcag cctgaatggc gaatgctaga	7560
gcagcttagc cttggatcag attgtcgttt cccgccttca gtttaacta tcagtgtttg	7620
acaggatata ttggcgggta aacctaagag aaaagagcgt ttattagaat aacggatatt	7680
taaaaggcgg tgaaaagggt taccgttcg tccatttga tgtgcatgcc aaccacagg	7740
ttcccctcgg gatcaaagta ctttgatcca acccctccgc tgctatagtg cagtcggctt	7800
ctgacgttca gtgcagccgt cttctgaaaa cgacatgtcg cacaagtcct aagttacgcg	7860
acaggtgcc gccctgccct tttctggcg tttcttgc gcgtgtttta gtcgcataaa	7920
gtagaatact tgcgactaga accggagaca ttacccatg aacaagagcg ccgccgtgg	7980
cctgctgggc tatgcccgcg tcagcaccga cgaccaggac ttgaccaacc aacgggccga	8040

actgcacgcg gccggctgca ccaagctgtt ttccgagaag atcaccggca ccaggcgca	8100
ccgcccgag ctggccagga tgttgacca cctacgcct ggccagcttg tgacagtac	8160
caggctagac cgcctggccc gcagaccgcg cgacctactg gacattgccg agcgcatcca	8220
ggaggccggc gcgggcctgc gtagcctggc agagccgtgg gccgacacca ccacgccggc	8280
cggccgcatg gtgttgaccg tgttcgccg cattgccgag ttcgagcgtt ccctaactcat	8340
cgaccgcacc cggagcgggc gcgagggcgc caaggcccg ggctgaagt ttggccccc	8400
ccctaccctc accccggcac agatcgcgca cggccgcgag ctgatcgacc aggaaggccg	8460
caccgtgaaa gaggcggctg cactgcttgg cgtgcatgcg tcgaccctgt accgcgcaact	8520
tgagcgagc gaggaagtga cgcccaccga ggccaggcgg cgcggtgcct tccgtgagga	8580
cgcattgacc gaggcgcagc ccctggcggc cgccgagaat gaacccaag aggaacaagc	8640
atgaaaccgc accaggacgg ccaggacgaa ccgtttttca ttaccgaaga gatcgaggcg	8700
gagatgatcg cggccgggta cgtgttcgag cggccgcgc acgtctcaac cgtgcggctg	8760
catgaaatcc tggccggtt gtctgatgcc aagctggcgg cctggccggc cagcttgcc	8820
gctgaagaaa ccgagcgccg ccgtctaaaa agtgatgtg tatttgagta aaacagcttg	8880
cgatcatcgg tcgtgcgta tatgatgcga tgagtaaata aacaatacg caagggaac	8940
gcatgaaggt tatcgtgta cttaaccaga aaggcgggtc aggcaagacg accatcgaa	9000
cccatctagc ccgcgcctg caactcgccg gggccgatgt tctgttagtc gattccgac	9060
cccaggcgag tgcccgcat tggcgccggc tgcgggaaga tcaaccgcta accgtgtcg	9120
gcatcgaccg ccgacgatt gaccgcgacg tgaaggccat cggccggcgc gacttcgtag	9180
tgatcgacgg agcgcgccg gcggcggaact tggctgtgtc cgcgatcaag gcagccgact	9240
tcgtgtgat tccggtgcag ccaagccctt acgacatatg ggccaccgcc gacctggtg	9300
agctggttaa gcagcgcat gaggtcacgg atggaaggct acaagcgcc tttgtcgtgt	9360
cgcggcgcat caaaggcacg cgcatcggc gtgaggttg cgaggcgctg gccgggtacg	9420
agctgcccatt tcttgagtc cgtatcacgc agcgcgtgag ctaccagcg actgccgcg	9480
ccggcacaac cgttcttgaa tcagaaccgc agggcgacgc tgcccgcgag gtccaggcgc	9540
tggccgctga aattaaatca aaactcattt gagttaatga ggtaagaga aatgagcaa	9600
aagcacaaac acgctaagt cggccgtcc gagcgacgc agcagcaagg ctgcaacgtt	9660
ggccagcctg gcagacgcg cagccatgaa gcgggtcaac tttcagttgc cggcgaggga	9720
tcacaccaag ctgaagatgt acgcggtacg ccaaggcaag accattaccg agctgctatc	9780
tgaatacatc gcgcagctac cagagtaaat gagcaaatga ataaatgagt agatgaattt	9840
tagcggttaa aggaggcggc atggaaaatc aagaacaacc aggcaccgac gccgtggaat	9900

gccccatgtg tggaggaacg ggcggttggc caggcgtaag cggctgggtt gtctgccggc	9960
cctgcaatgg cactggaacc cccaagcccc aggaatcggc gtgacggtcg caaacatcc	10020
ggccccgtac aaatcggcgc ggcgctgggt gatgacctgg tggagaagtt gaaggccgcg	10080
caggccgccc agcggcaacg catcgaggca gaagcacgcc ccggtgaatc gtggcaagcg	10140
gcccgtgac gaatccgcaa agaatcccgg caaccgccgg cagccggtgc gccgtcgatt	10200
aggaagccgc ccaagggcga cgagcaacca gatTTTTTcg ttccgatgct ctatgacgtg	10260
ggcaccgcg atagtgcgac catcatggac gtggccgttt tccgtctgtc gaagcgtgac	10320
cgacgagctg gcgaggtgat ccgctacgag ctccagacg ggcacgtaga ggtttccgca	10380
gggccggccg gcatggccag tgtgtgggat tacgacctgg tactgatggc ggtttcccat	10440
ctaaccgaat ccatgaaccg ataccgggaa gggaaggag acaagcccgg ccgcgtgttc	10500
cgtccacacg ttccggacgt actcaagttc tgcggcgag ccgatggcgg aaagcagaaa	10560
gacgacctgg tagaaacctg cattcggtta aacaccacgc acgttgccat gcagcgtacg	10620
aagaaggcca agaacggccg cctggtgacg gtatccgagg gtgaagcctt gattagccgc	10680
tacaagatcg taaagagcga aaccgggcgg ccggagtaca tcgagatcga gctagctgat	10740
tggatgtacc gcgagatcac agaaggcaag aaccggacg tgctgacggt tcacccgat	10800
tactttttga tcgatcccg catcgccgt tttctctacc gcctggcacg ccgcgccga	10860
ggcaaggcag aagccagatg gttgttcaag acgatctacg aacgcagtgg cagcgcgga	10920
gagttcaaga agttctgttt caccgtgcgc aagctgatcg ggtcaaatga cctgccggag	10980
tacgatttga aggaggaggc ggggcaggct ggcccgatcc tagtcatgcg ctaccgcaac	11040
ctgatcgagg gcgaagcatc cgccggttcc taatgtacgg agcagatgct agggcaaatt	11100
gccctagcag gggaaaaagg tcgaaaagca ctctttctg tggatagcac gtacattggg	11160
aacccaaagc cgtacattgg gaaccggaac ccgtacattg ggaacccaaa gccgtacatt	11220
gggaaccggt cacacatgta agtgactgat ataaaagaga aaaaaggcga tttttccgcc	11280
taaaactctt taaaacttat taaaactctt aaaaccgcc tggcctgtgc ataactgtct	11340
ggccagcgca cagccgaaga gctgcaaaaa gcgcctaccc ttcggtcgct gcgctcccta	11400
cgccccgccg cttcgctcg gcctatcgcg gccgtggcc gctcaaaaat ggctggccta	11460
cggccaggca atctaccagg gcgcggacaa gccgcgccgt cgccactcga ccgccggcgc	11520
ccacatcaag gcacctgcc tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat	11580
gcagctccc gagacggta cagcttgtct gtaagcggat gccgggagca gacaagccc	11640
tcaggcgcg tcagcgggtg ttggcgggtg tcggggcgca gccatgacct agtcacgtag	11700
cgatagcgga gtgtatactg gcttaactat gcggcatcag agcagattgt actgagagt	11760

caccatatgc gggtgtgaaat accgcacaga tgcgtaagga gaaaataccg catcaggcgc	11820
tcttccgctt cctcgtcac tgactcgtg cgctcggtcg ttcggctgcg gcgagcggta	11880
tcagctcact caaaggcggc aatacggta tccacagaat caggggataa cgcaggaaag	11940
aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc gttgctggcg	12000
tttttccata ggctccgcc cctgacgag catcacaaaa atcgacgtc aagtcagagg	12060
tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag ctccctcgtg	12120
cgctctcctg ttccgacct gccgcttacc ggatacctgt ccgcctttct ccttcggga	12180
agcgtggcgc tttctcatag ctacgctgt aggtatctca gttcgggtga ggtcgttcgc	12240
tccaagctgg gcgtgtgca cgaaccccc gttcagcccg accgtgcgc cttatccggt	12300
aactatcgtc ttgagtcaa cccgtaaga cagcacttat cgccactggc agcagccact	12360
ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagtctt gaagtggtag	12420
cctaactacg gctacactag aaggacagta ttgggtatct gcgctctgct gaagccagtt	12480
accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc tggtagcggc	12540
ggtttttttg ttgcaagca gcagattacg cgcagaaaaa aaggatctca agaagatcct	12600
ttgatcttt ctacggggtc tgacgtcag tggaaacgaa actcacgtta agggattttg	12660
gtcatgcatt ctaggtacta aaacaattca tccagtaaaa tataatat ttatctctcc	12720
caatcaggct tgatccccag taagtaaaaa aatagctcga catactgttc ttccccgata	12780
tcctccctga tcgaccggac gcagaaggca atgtcatacc acttgccgc cctgccgctt	12840
ctccaagat caataaagcc acttactttg ccacttttca caaagatgtt gctgtctccc	12900
aggtcgccgt gggaaaagac aagttcctct tcgggctttt ccgtctttaa aaaatcatac	12960
agctcgccgc gatctttaa tggagtgtct tcttccagc ttctgcaatc cacatcgcc	13020
agatcgttat tcagtaagta atccaattcg gctaagcggc tgtctaagct attcgtatag	13080
ggacaatccg atatgtcgat ggagtgaag agcctgatgc actccgcata cagctcgata	13140
atcttttcag ggctttgttc atcttcatac tcttccgagc aaaggacgcc atcgccctca	13200
ctcatgagca gattgtcca gccatcatgc cgttcaaagt gcaggacctt tggaaacaggc	13260
agctttcctt ccagccatag catcatgtcc ttttccggt ccacatcata ggtggtcctt	13320
ttataccggc tgtccgtcat ttttaatat aggttttcat ttttccac cagcttatat	13380
accttagcag gagacattcc ttccgtatct ttacgcagc ggtattttc gatcagtttt	13440
ttcaattccg gtgatattct cattttagcc atttattatt tcttctct tttctacagt	13500

atttaaagat accccaagaa gctaattata acaagacgaa ctccaattca ctgttccttg	13560
cattctaaaa ccttaaatat cagaaaacag ctttttcaaa gttgttttca aagttggcgt	13620
ataacatagt atcgacggag ccgattttga aaccgcggtg atcacaggca gcaacgctct	13680
gtcatcgta caatcaacat gctaccctcc gcgagatcat ccgtgtttca aaccggcag	13740
cttagttgcc gttcttccga atagcatcgg taacatgagc aaagtctgcc gccttacaac	13800
ggctctcccg ctgacgccgt cccggactga tgggctgcct gtatcgagtg gtgattttgt	13860
gccgagctgc cggtcgggga gctgttggct ggctgggtggc aggatatatt gtggtgtaaa	13920
caaattgacg cttagacaac ttaataacac attgcggacg tttttaatgt actgaattaa	13980
cgccgaatta attcggggga tctggatttt agtactggat tttggtttta ggaattagaa	14040
attttattga tagaagtatt ttacaaatac aaatacatat taagggtttc ttatatgctc	14100
aacacatgag cgaaaccta taggaacct aattccctta tctgggaact actcacacat	14160
tattatggag aaactcgagc ttgtcgatcg acagatccgg tcggcatcta ctctatttct	14220
ttgcctcgg acgagtgtg gggcgtcgg ttcactatc ggcgagtact tctacacagc	14280
catcggtcca gacggccg cttctgcggg cgatttgtgt acgcccgaca gtcccggctc	14340
cggatcggac gattgcgtcg catcgacct gcgccaagc tgcatcatcg aaattgccgt	14400
caaccaagct ctgatatagt tggtaagac caatgcggag catatacgcc cggagtcgtg	14460
gcgatcctgc aagctccgga tgcctccgt cgaagtagcg cgtctgctgc tccatacaag	14520
ccaaccagg cctccagaag aagatgttgg cgacctgta ttgggaatcc ccgaacatcg	14580
cctcgtcca gtcaatgacc gctgttatgc ggccattgtc cgtcaggaca ttgttggagc	14640
cgaatccgc gtgcacgagg tgccggactt cggggcagtc ctgggcccagc agcatcagct	14700
catcgagagc ctgcgcgacg gacgcactga cgtgtcgtc catcacagtt tgccagtgt	14760
acacatgggg atcagcaatc gcgcatatga aatcaccca tgtagtgtat tgaccgattc	14820
cttgcggtcc gaatgggccc aaccgctcg tctggctaag atcggccga gcgatcgc	14880
ccatagcctc cgcgaccggt tgtagaacag cgggcagttc ggtttcagc aggtcttgca	14940
acgtgacacc ctgtgcacgg cgggagatgc aataggtcag gctctcgta aactcccaa	15000
tgtcaagcac ttccggaatc gggagcgcgg ccgatgcaa gtgccgataa acataacgat	15060
ctttgtagaa accatggcg cagctattta ccgcaggac atatccacgc cctctacat	15120
cgaagctgaa agcacgagat tcttcgccct ccgagagctg catcaggtcg gagacgtgt	15180
cgaacttttc gatcagaaac ttctcgacag acgtcgcggt gagttcagc tttttcatat	15240
ctcattgccc cccgggatct gcgaaagctc gagagagata gattttaga gagagactgg	15300
tgatttcagc gtgtcctctc caaatgaaat gaacttcctt atatagagga aggtcttgcg	15360

aaggatagtg ggattgtgcg tcatccctta cgtcagtgga gatatcacat caatccactt 15420
gctttgaaga cgtgggttga acgtcttctt ttccacgat gctcctcgtg ggtgggggtc 15480
catctttggg accactgtcg gcagaggcat cttgaacgat agcctttcct ttatcgcaat 15540
gatggcattt gtaggtgcc a ctttctttt ctactgtcct tttgatgaag tgacagatag 15600

ctgggcaatg gaatccgagg aggtttcccg atattaccct ttgttgaaaa gtctcaatag 15660
ccctttggtc ttctgagact gtatctttga tattcttga gtagacgaga gtgtcgtgct 15720
ccaccatgtt atcacatcaa tccacttgct ttgaagacgt ggttggaaacg tcttctttt 15780
ccacgatgct cctcgtgggt ggggttccat ctttgggacc actgtcggca gaggcattct 15840
gaacgatagc ctttcttta tcgcaatgat ggcatttgta ggtgccacct tccttttcta 15900
ctgtcctttt gatgaagtga cagatagctg ggcaatggaa tccgaggagg tttcccgata 15960
ttaccctttg ttgaaaagtc tcaatagccc ttgggtcttc tgagactgta tctttgatat 16020

tcttggagta gacgagagtg tcgtgctcca ccatgttggc aagctgctct agccaatacg 16080
caaaccgcct ctccccgcgc gttggccgat tcattaatgc agctggcacg acaggtttcc 16140
cgactggaaa gcgggcagtg agcgcaacgc aattaatgtg agttagctca ctcataggc 16200
acccaggct ttacacttta tgcttccggc tcgtatgttg tgtggaattg tgagcggata 16260
acaatttcac acaggaaaca gctatgacca tgattacgcc 16300

<210> 38

<211> 16477

<212> DNA

<213> Artificial Sequence

<220><223> nucleotides coding D121

<

400> 38

aagcttgata atttattga aaattcataa gaaaagcaaa cgttacatga attgatgaaa 60
caatacaaag acagataaag ccacgcacat ttaggatatt ggccgagatt actgaatatt 120
gagtaagatc acggaatttc tgacaggagc atgtcttcaa ttcagcccaa atggcagttg 180
aaatactcaa accgccccat atgcaggagc ggatcattca ttgtttgttt ggttgccttt 240
gccaacatgg gattccaaga ttctgcagtc aaatctcggg gacgggcagg accggacggg 300
gcggtaccgg caggctgaag tccagctgcc agaaaccac gtcattgccag ttcccgtgct 360
tgaagccggc cgcccgagc atgccgcggg gggcatatcc gagcgcctcg tgcattgcga 420

cgctcgggtc gttgggcagc ccgatgacag cgaccacgt cttgaagccc tgtgcctcca 480
gggacttcag caggtgggtg tagagcgtgg agcccagtc cgtccgctgg tggcgggggg 540

agacgtacac ggctgactcg gccgtccagt cgtaggcggt gcgtgccttc caggggccccg	600
cgtaggcgat gccggcgacc tcgccgtcca cctcgcgac gagccagga tagcgctccc	660
gcagacggac gaggtcgcc gtccactcct gcggttcctg cggctcgga cggaagtga	720
ccgtgcttgt ctcgatgtag tggttgacga tggtcagac cgccggcatg tccgcctcgg	780
tggcacggcg gatgtcgcc gggcgctgt ctgggctcat cgattcgatt tgggtatcg	840
agattggta tgaattcag atgctagtgt aatgtattgg taatttggga agatataata	900
ggaagcaagg ctatttatcc atttctgaaa aggcgaaatg gcgtcaccgc gagcgtcacg	960
cgcattccgt tcttgcgtga aagcgttgtt tggtagactt ttgactagcg aggcttggcg	1020
tgtcagcgt tctattcaaa agtcgttaat ggctgcggat caagaaaaag ttggaataga	1080
aacagaatac ccgcgaaatt caggccccgt tgccatgtcc tacacgccga aataaacgac	1140
caaatagta gaaaaataaa aactgactcg gatacttacg tcacgtcttg cgcaactgatt	1200
tgaaaaatct cagccatggc tgcaggtcga cgagtcagta ataacggcg tcaaagtgg	1260
tgcagccggc acacacgagt cgtgtttatc aactcaaagc acaataactt ttcctcaacc	1320
taaaaataag gcaattagcc aaaaacaact ttgcgtgtaa acaacgtca atacacgtgt	1380
cattttatta ttagctattg cttcaccgcc ttagctttct cgtgacctag tcgtcctcgt	1440
cttttcttct tcttcttcta taaaacaata cccaaagagc tcttcttctt cacaattcag	1500
atttcaattt ctcaaatct taaaaacttt ctctcaattc tctctaccgt gatcaaggta	1560
aatttctgtg ttccttatc tctcaaatc ttcgattttg ttttcgttcg atcccaattt	1620
cgtatatgtt ctttggttta gattctgtta atcttagatc gaagacgatt tcttgggttt	1680
gatcgttaga tatcatctta attctcgatt agggtttcat agatatac cgtattgttc	1740
aaataatttg agttttgtcg aataattact cttcgatttg tgatttctat ctagatctgg	1800
tgtagtttc tagtttgtgc gatcgaattt gtcgattaat ctgagttttt ctgattaaca	1860
gccatggact ataaggacca cgacggagac tacaaggatc atgatattga ttacaagac	1920
gatgacgata agatggcccc aaagaagaag cggaaggtcg gtatccacgg agtcccagca	1980
gccgacaaga agtacagcat cggcctggac atcggcacca actctgtggg ctgggccgtg	2040
atcaccgacg agtacaaggt gccagcaag aaattcaagg tgctgggcaa caccgaccgg	2100
cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag	2160
gccacccggc tgaagagaac cgccagaaga agataacca gacggaagaa cggatctgc	2220
tatctgcaag agatcttcag caacgagatg gccaaagtgg acgacagctt cttccacaga	2280
ctggaagagt ctttcttgtt ggaagaggat aagaagcacg agcggcaccc catcttcggc	2340
aacatcgtgg acgaggtggc ctaccacgag aagtaccca ccatctacca cctgagaaag	2400

aaactggtgg acagcacga caagccgac ctgcgctga tctatctggc cctggccac	2460
atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaacccga caacagcgac	2520
gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaaaccc	2580
atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga	2640
cggctggaaa atctgatcg ccagctgccc ggcgagaaga agaatggcct gttcggaac	2700
ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag	2760
gatgccaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc	2820
cagatcggcg accagtacg cgacctgttt ctggccgcca agaacctgtc cgacgccatc	2880
ctgctgagcg acatcctgag agtgaacacc gagatcacca aggccccct gagcgctct	2940
atgatcaaga gatacgacga gcaccaccag gacctgacct tgctgaaagc tctcgtgcgg	3000
cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc	3060
ggctacattg acggcggagc cagccaggaa gagttctaca agttcatcaa gcccattctg	3120
gaaaagatgg acggcaccga ggaactgtc gtgaagctga acagagagga cctgctgcgg	3180
aagcagcgga ctttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac	3240
gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc	3300
gagaagatcc tgaccttccg catcccctac tacgtgggcc ctctggccag gggaaacagc	3360
agattcgctt ggatgaccag aaagagcgag gaaacatca ccccttgaa cttcgaggaa	3420
gtggtggaca agggcgcttc cgcccagagc ttcacgagc ggatgacaa cttcgataag	3480
aacctgccc acgagaaggt gctgcccag cacagcctgc tgtacgagta cttaccctg	3540
tataacgagc tgaccaaagt gaaatcgtg accgaggga tgagaaagcc cgccttcctg	3600
agcggcgagc agaaaaagc catcgtggac ctgctgttca agaccaaccg gaaagtgacc	3660
gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc	3720
tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt	3780
atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg	3840
ctgacctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc	3900
cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc	3960
aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg	4020
gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac	4080
agcctgacct ttaaagagga catccagaaa gcccggtgt cggccaggg cgatagcctg	4140

cacgagcaca ttgccaatct ggccggcagc cccgccatta agaagggcat cctgcagaca	4200
gtgaaggtgg tggacgagct cgtgaaagt atgggccggc acaagcccga gaacatcgtg	4260
atcgaaatgg ccagagagaa ccagaccacc cagaagggaac agaagaacag ccgcgagaga	4320
atgaagcgga tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc	4380
gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg	4440
gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat	4500
atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct gaccagaagc	4560
gacaagaacc ggggcaagag cgacaactg ccctccgaag aggtcgtgaa gaagatgaag	4620
aactactggc ggcagctgct gaacccaag ctgattacc agagaaagt cgacaatctg	4680
accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggttcat caagagacag	4740
ctggtggaaa cccggcagat cacaagcac gtggcacaga tcctggactc ccggtgaac	4800
actaagtacg acgagaatga caagctgac cggaagtga aagtatcac cctgaagtcc	4860
aagctggtgt ccgatttccg gaagatttc cagttttaca aagtgcgca gatcaacaac	4920
taccaccag cccacgacg ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag	4980
taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag	5040
atgatcgcca agagcgagca ggaaatcggc aaggtaccg ccaagtactt ctctacagc	5100
aacatcatga actttttcaa gaccgagatt accctggcca acggcgagat ccggaagcgg	5160
cctctgatcg agacaaacgg cgaaaccggg gagatcgtgt gggataaggg ccgggatttt	5220
gccaccgtgc ggaaagtgt gagcatgcc caagtgaata tcgtgaaaaa gaccgaggtg	5280
cagacaggcg gcctcagcaa agagtctatc ctgcccaaga ggaacagcga taagctgatc	5340
gccagaaaga aggactggga ccctaagaag tacggcggct tcgacagccc caccgtggcc	5400
tattctgtgc tgggtgtggc caaagtggaa aagggaagt ccaagaaact gaagagtgtg	5460
aaagagctgc tggggatcac catcatggaa agaagcagct tcgagaagaa tccatcgac	5520
tttctggaag ccaagggcta caaagaagtg aaaaaggacc tgatcatcaa gctgcctaag	5580
tactccctgt tcgagctgga aaacggccgg aagagaatgc tggcctctgc cggcgaactg	5640
cagaaggga acgaactggc cctgccctcc aaatatgtga acttctgtg cctggccagc	5700
cactatgaga agctgaaggg cteccccgag gataatgagc agaaacagct gtttgtgga	5760
cagcacaagc actacctgga cgagatcatc gagcagatca gcgagttctc caagagagtg	5820
atcctggccg acgctaactt ggacaaagt ctgtccgct acaacaagca ccgggataag	5880

cccatcagag agcaggccga gaatatcatc cacctgttta ccctgaccaa tctgggagcc	5940
cctgccgcct tcaagtactt tgacaccacc atcgaccgga agaggtagac cagcaccaaa	6000
gaggtgctgg acgccaccct gatccaccag agcatcaccg gcctgtacga gacacggatc	6060
gacctgtctc agctgggagg cgacaaaagg cggcgggcca cgaaaaaggc cggccaggca	6120
aaaaagaaaa agtaagaatt cgcccttctt cgttgaacaa cggaaactcg acttgccttc	6180
cgcacaatac atcatttctt cttagctttt tttcttcttc ttcgttcata cagttttttt	6240
ttgtttatca gcttacattt tcttgaaccg tagctttcgt tttcttcttt ttaactttcc	6300
attcggagtt tttgtatctt gtttcatagt ttgtcccagg attagaatga ttaggcatcg	6360
aaccttcaag aatttgattg aataaaacat cttcattctt aagatatgaa gataatcttc	6420
aaaaggcccc tgggaatctg aaagaagaga agcaggccca tttatatggg aaagaacaat	6480
agtatttctt atataggccc atttaagttg aaaacaatct tcaaaagtcc cacatcgctt	6540
agataagaaa acgaagctga gtttatatac agctagagtc gaagtagtga ttggttgaac	6600
cgcaaagtgg gatgttttag agctagaaat agcaagttaa aataaggcta gtccgttacc	6660
aacttgaaaa agtggcaccg agtcggtgca acaaagcacc agtgggtctag tggtagaata	6720
glaccctgcc acggtacaga cccgggttcg attcccggct ggtgcagatg ctatagagag	6780
atctatgttt tagagctaga aatagcaagt taaaataagg ctagtccgtt atcaacttga	6840
aaaagtggca ccgagtcggt gcaacaaagc accagtggtc tagtggtaga atagtaccct	6900
gccacggtac agaccgggtg tcgattcccg gctgggtcat gagcagaaag aagaagagag	6960
tttttagagct agaaatagca agttaaataa aggctagtcc gttatcaact tgaaaaagtg	7020
gcaccgagtc ggtgcaacaa agcaccagtg gtctagtggg agaatagtag cctgccacgg	7080
tacagaccgg ggttcgattc ccggctgggtg cacgcttcag catccttgtc gtgtttttaga	7140
gctagaaata gcaagttaaa ataaggctag tccgttatca acttgaaaaa gtggcaccga	7200
gtcgggtgcaa caaagcacca gtgggtctagt ggtagaatag taccctgccg cggtagagac	7260
ccgggttcga ttcccggctg gtgcagcaga ttgggtatga tgtctgtttt agagctagaa	7320
atagcaagtt aaaataaggc tagtccgtta tcaacttgaa aaagtggcac cgagtcggtg	7380
caacaaagca ccagtgggtc agtggttagaa tagtaccctg ccacggtaca gaccggggtt	7440
cgattcccgg ctgggtcatg gagctttatc ctggtatcgt ttttagagcta gaaatagcaa	7500
gttaaaaaaa ggctagtcgg ttatcaactt gaaaaagtgg caccgagtcg gtgctttttt	7560
tctagagggc gaattcactg gccgtcgttt tacaacgtcg tgactgggaa aaccctggcg	7620
ttaccaact taatgcctt gcagcacatc cccctttcgc cagctggcgt aatagcgaag	7680
aggcccgcac cgatgcacct tccaacagc tgcgcagcct gaatggcgaa tgctagagca	7740

gcttgagctt ggatcagatt gtcgtttccc gccttcagtt taaactatca gtgtttgaca	7800
ggatatattg gcgggtaaac ctaagagaaa agagcggtta ttagaataac ggatatatta	7860
aaggcgctga aaaggtttat ccgttcgtcc atttgtatgt gcatgccaac cacagggttc	7920
ccctcgggat caaagtactt tgatccaacc cctccgtgc tatagtgcag tcggcttctg	7980
acgttcagtg cagccgtctt ctgaaaacga catgtcgac aagtcctaag ttacgcgaca	8040
ggctgccgcc ctgccctttt cctggcggtt tcttgtcgcg tgttttagtc gcataaagta	8100
gaatacttgc gactagaacc ggagacatta cgccatgaac aagagcgccg ccgctggcct	8160
gctgggctat gcccgcgtca gcaccgacga ccaggacttg accaaccaac gggccgaact	8220
gcacgcggcc ggctgcacca agctgttttc cgagaagatc accggcacca ggcgcgaccg	8280
cccggagctg gccaggatgc ttgaccacct acgccctggc gacgttgtga cagtgaccag	8340
gctagaccgc ctggcccga gcacccgca cctactggac attgccgagc gcatccagga	8400
ggccggcgcg ggcttgcgta gcctggcaga gccgtgggcc gacaccacca cgccggccgg	8460
ccgcatggtg ttgaccgtgt tcgccggcat tgccgagttc gagcgttccc taatcatcga	8520
ccgcaccggg agcgggcgcg aggccgcca ggcccaggc gtgaagtttg gccccgccc	8580
tacctcacc ccggcacaga tcgcgcacgc ccgcgagctg atcgaccagg aaggccgcac	8640
cgtgaaagag gcggctgcac tgcctggcgt gcatcgctcg acctgtacc gcgcacttga	8700
gcgcagcgag gaagtgcgc ccaccgaggc caggcggcgc ggtgccttcc gtgaggacgc	8760
attgaccgag gccgacgccc tggcgccgc cgagaatgaa cgccaaggagg aacaagcatg	8820
aaaccgcacc aggacggcca ggacgaaccg tttttcatca ccgaagagat cgaggcgag	8880
atgatcgcg ccgggtacgt gttcgagccg ccgcgcacg tctcaaccgt gcggctgcat	8940
gaaatcctgg ccggtttgtc tgatgccaag ctggcgccct ggccggccag cttggccgct	9000
gaagaaaccg agcgcgcg ccctaaaaagg tgatgtgtat ttgagtaaaa cagcttgcgt	9060
catgcggtcg ctgcgtatat gatgcgatga gtaataaac aaatcgcaa ggggaacgca	9120
tgaaggttat cgtgtactt aaccagaaag gcgggtcagg caagacgacc atcgcaacc	9180
atctagcccc cgccctgcaa ctgcggggg ccgatgttct gttagtcgat tccgatcccc	9240
agggcagtc ccgcgattgg gcggccgtgc gggaagatca accgtaacc gttgtcgga	9300
tcgaccgccc gacgattgac gcgcagtgga aggccatcgg ccggcgcgac ttcgtagtga	9360
tcgacggagc gcccaggcg gcggacttgg ctgtgtccgc gatcaaggca gccgacttcg	9420
tgctgattcc ggtgcagcca agcccttacg acatatgggc caccgccgac ctggtggagc	9480
tggttaagca gcgcattgag gtcacggatg gaaggctaca agcgcccttt gtcgtgtcg	9540
ggcgatcaa aggcacgcgc atcgcggtg aggttgccga ggcgctggcc gggtacgagc	9600

tgccatttct tgagtcctgt atcacgcagc gcgtgagcta cccaggcact gccgccgccg	9660
gcacaaccgt tcttgaatca gaacccgagg gcgacgtgc ccgcgaggtc caggcgtgg	9720
ccgtgaaat taaatcaaaa ctcatcttgag ttaatgaggt aaagagaaaa tgagcaaaag	9780
cacaacacg ctaagtgccg gccgtccgag cgcacgcagc agcaaggctg caacgttggc	9840
cagcttgga gacacgccag ccatgaagcg ggtcaacttt cagttgccgg cggaggatca	9900
caccaagctg aagatgtacg cggtagcca aggcaagacc attaccgagc tgctatctga	9960
atacatcgcg cagctaccag agtaaatgag caaatgaata aatgagtaga tgaattttag	10020
cggctaagg aggcggcatg gaaaatcaag aacaaccagg caccgacgcc gtggaatgcc	10080
ccatgtgtgg aggaacgggc ggttggccag gcgtaagcgg ctgggttgtc tgccggccct	10140
gcaatggcac tggaaacccc aagcccagg aatcggcgtg acggtcgaa accatccggc	10200
ccgttataaa tggcgccgcg gctgggtgat gacctgttg agaagttaga ggccgcgag	10260
gccgccagc ggcaacgcat cgaggcagaa gcacgccccg gtgaatcgtg gcaagcgcc	10320
gctgatcgaa tccgcaaaga atcccggcaa ccgccgcgag ccggtgcgcc gtcgattagg	10380
aagccgcccc agggcgacga gcaaccagat ttttctgtc cgatgctcta tgacgtgggc	10440
acccgcgata gtcgcagcat catggacgtg gccgttttcc gtcgtcgaa gcgtgaccga	10500
cgagctggcg aggtgatccg ctacagctt ccagacgggc acgtagaggt ttccgcaggg	10560
ccggccggca tggccagtg gtgggattac gacctgttac tgatggcggg ttcccatcta	10620
accgaatcca tgaaccgata ccgggaaggg aaggagaca agccggccg cgtgttccgt	10680
ccacacgttg cggacgtact caagtctgc cggcgagccg atggcgaaa gcagaaagac	10740
gacctgtag aaacctgcat tcggttaaacc accacgcacg ttgcatgca gcgtacgaag	10800
aaggccaaga acggccgctt ggtgacggt tccgaggtg aagccttgat tagccgtac	10860
aagatcgtaa agagcgaac ccggcgccg gagtacatcg agatcgagct agctgattgg	10920
atgtaccgag agatcacaga aggcaagaac ccggacgtgc tgacggttca cccgattac	10980
tttttgatcg atcccggcat cggccgtttt ctctaccgcc tggcacgccg cgccgcaggc	11040
aaggcagaag ccagatggtt gttcaagacg atctacgaac gcagtggcag cgccggagag	11100
ttcaagaagt tctgtttcac cgtgcgcaag ctgatcgggt caaatgacct gccggagtag	11160
gatttgaagg aggagggcgg gcagctggc ccgatctag tcatcgcta ccgcaacctg	11220
atcgaggcgg aagcatccgc cggttcctaa tgtacggagc agatgctagg gcaaattgcc	11280
ctagcagggg aaaaaggtcg aaaagcactc tttctgtgg atagcacgta cattgggaac	11340

ccaaagccgt acattgggaa ccggaacccg tacattggga acccaaagcc gtacattggg	11400
aaccggtcac acatgtaagt gactgatata aaagagaaaa aaggcgattt ttccgcctaa	11460
aactctttaa aacttattaa aactcttaaa acccgcttgg cctgtgcata actgtctggc	11520
cagcgcacag ccgaagagct gcaaaaagcg cctacccttc ggtcgtgcg ctccctacgc	11580
cccgccgctt cgcgtcgcc tatecgggcc gctggccgct caaaaatggc tggcctacgg	11640
ccaggcaatc taccagggcg cggacaagcc gcgccgtcgc cactcgaccg ccggcgccca	11700
catcaaggca cctgcctcg cgcgtttcgg tgatgacggt gaaaacctct gacacatgca	11760
gctcccgag acggtcacag cttgtctgta agcggatgcc gggagcagac aagcccgta	11820
ggcgcgta cggggtgtg gcgggtgtcg gggcgagcc atgaccagt cacgtagcga	11880
tagcgagtg tatactggct taactatgcg gcatcagagc agattgtact gagagtgcac	11940
catatcggt gtgaaatacc gcacagatgc gtaaggagaa aataccgcat caggcgctct	12000
tccgttctt cgtcactga ctgcgtgcgc tcggtcgttc ggctgcggcg agcggatca	12060
gctcactcaa aggcggtaac acggttatcc acagaatcag gggataacgc aggaaagaac	12120
atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa aggcgcgctt gctggcgctt	12180
ttccataggc tccgcccc tgacgagcat cacaaaaatc gacgtcaag tcagaggtgg	12240
cgaaacccga caggactata aagataccag gcgtttcccc ctggaagctc cctcgtgcgc	12300
tctcgttgc cgacctgcc gcttaccgga tacctgtccg cttttctccc ttcgggaagc	12360
gtggcgctt ctcatagctc acgtgtagg tatctcagtt cgggttaggt cgttcgctcc	12420
aagctgggct gtgtgcacga accccccgtt cagcccgacc gctgcgcctt atccggtaac	12480
tatcgtcttg agtccaacct ggtaagacac gacttatcgc cactggcagc agccactggt	12540
aacaggatta gcagagcgag gtatgtaggc ggtgctacag agttcttgaa gtggtggcct	12600
aactacggct aactagaag gacagtattt ggtatctgcg ctctgctgaa gccagttacc	12660
ttcgaaaaa gatttggtag ctcttgatcc ggcaaaaaa ccaccgctgg tagcggtggt	12720
ttttttgtt gcaagcagca gattacgcgc agaaaaaaag gatctcaaga agatcctttg	12780
atcttttcta cggggtctga cgctcagtgg aacgaaaact cacgttaagg gattttggtc	12840
atgcattcta ggtactaaaa caattcatcc agtaaaatat aatattttat tttctccaa	12900
tcaggcttga tccccagtaa gtcaaaaaat agctcgacat actgttcttc cccgatatcc	12960
tccctgatcg accggacgca gaaggcaatg tcataccact tgtccgcct gccgcttctc	13020
ccaagatcaa taaagccact tactttgcca tctttcacia agatgttgct gtctcccagg	13080
tcgccgtggg aaaagacaag ttctctcttc ggcttttccg tctttaaaaa atcatacagc	13140
tcgcgcggat ctttaaatgg agtgtcttct tcccagtttt cgcaatccac atcgccaga	13200

tcgttattca gtaagtaatc caattcggct aagcggctgt ctaagctatt cgtataggga	13260
caatccgata tgtcgatgga gtgaaagagc ctgatgcact ccgcatacag ctcgataatc	13320
ttttcagggc tttgttcacg ttcatactct tccgagcaaa ggacgccatc ggcctcactc	13380
atgagcagat tgctccagcc atcatgccgt tcaaagtga ggacctttgg aacaggcagc	13440
tttcccttcca gccatagcat catgtccttt tcccggttcca catcataggt ggtcccttta	13500
taccggctgt ccgtcatttt taaatatagg ttttcatttt cteccaccag cttatatacc	13560
ttagcaggag acattccttc cgtatctttt acgcagcggg atttttcgat cagttttttc	13620
aattccgggt atatttctat tttagccatt tattatttcc ttcctctttt ctacagtatt	13680
taaagatacc ccaagaagct aattataaca agacgaactc caattcactg ttccttgcat	13740
tctaaaacct taaataccag aaaacagctt tttcaaagt gttttcaaag ttggcgata	13800
acatagtatc gacggagccg attttgaaac cgcggtgatc acaggcagca acgctctgtc	13860
atcgttacaa tcaacatgct accctccgcg agatcatccg tgtttcaaac ccggcagctt	13920
agttgccgtt cttccgaata gcacggtaa catgagcaaa gtctgccgc ttacaacggc	13980
tctccgctg acgccgtcc ggactgatgg gctgcctgta tcgagtgggtg attttgtgcc	14040
gagctgccgg tcggggagct gttggctggc tggtaggcagg atatattgtg gtgtaacaa	14100
attgacgctt agacaactta ataacacatt gcggacgttt ttaatgtact gaattaacgc	14160
cgaattaatt cgggggatct ggattttagt actggatttt ggttttagga attagaaatt	14220
ttattgatag aagtatttta caaatacaaa tacatactaa gggtttctta tatgtcaac	14280
acatgagcga aaccctatag gaaccctaatt tcccttatct gggaactact cacacattat	14340
tatggagaaa ctcgagcttg tcgatcgaca gatccggtcg gcatctactc tatttctttg	14400
ccctcggacg agtgctgggg cgctcggttt cactatcggc gactacttct acacagccat	14460
cgggtccagac ggccgcgctt ctgcgggcga tttgtgtacg cccgacagtc ccggtccgg	14520
atcggacgat tgcgtcgcat cgacctgcg cccaagctgc atcatcgaaa ttgccgtcaa	14580
ccaagctctg atagagttgg tcaagacaa tgcggagcat atacgcccg agtcgtggcg	14640
atcctgcaag ctccggatgc ctccgctcga agtagcgcgt ctgctgtcc atacaagcca	14700
accacggcct ccagaagaag atgttggcga cctcgatttg ggaatccccg aacatcgctt	14760
cgctccagtc aatgaccgt gttatcgggc cattgtccgt caggacattg ttggagccga	14820
aatcccgctg cagcaggctg cggacttcgg ggcagtcctc ggcccaaagc atcagctcat	14880
cgagagcctg cgcgacggac gactgacgg tgcgtccat cacagtttgc cagtataca	14940
catggggatc agcaatcgcg catatgaaat cagccatgt agtgtattga ccgattcctt	15000
gcggtccgaa tgggccgaac ccgctcgtct ggctaagatc ggccgcagcg atcgcatcca	15060

tagcctccgc gaccggttgt agaacagcgg gcagttcggg ttcaggcagg tcttgcaacg 15120

tgacaccctg tgcacggcgg gagatgcaat aggtcaggct ctcgctaaac tccccaatgt 15180

caagcacttc cggaatcggg agcgcggccg atgcaaagtg ccgataaaca taacgatctt 15240

tgtagaaacc atcggcgcag ctatttaccg gcaggacata tccacgccct cctacatcga 15300

agctgaaagc acgagattct tcgccctccg agagctgcat caggctggag acgtgtcga 15360

acttttcgat cagaaacttc tcgacagacg tcgcggtgag ttcaggcttt ttcatatctc 15420

attgcccccc gggatctgcg aaagctcgag agagatagat ttgtagagag agactggtga 15480

tttcagcgtg tctctcctaa atgaaatgaa ctctcttata tagaggaagg tcttgcaag 15540

gatatggga ttgtgcgtca tcccttacgt cagtggagat atcacatcaa tccacttgct 15600

ttgaagacgt ggttggaacg tcttcttttt ccacgatgct cctcgtgggt ggggggtccat 15660

ctttgggacc actgtcgga gaggcattct gaacgatagc ctttcttta tcgcaatgat 15720

ggcatttgta ggtgccacct tctttttcta ctgtccttt gatgaagtga cagatagctg 15780

ggcaatggaa tccgaggagg ttcccgata ttacccttg ttgaaaagtc tcaatagccc 15840

tttggcttct tgagactgta tctttgatat tcttgagta gacgagagt tcgtgtcca 15900

ccatgttctc acatcaatcc acttgctttg aagacgtggg tggaacgtct tctttttcca 15960

cgatgtcctc cgtgggtggg ggtccatctt tgggaccact gtcggcagag gcattctgaa 16020

cgatagcctt tcttttatcg caatgatggc attttaggt gccaccttcc ttttctactg 16080

tccttttgat gaagtgcagc atagctgggc aatggaatcc gaggaggttt cccgatatta 16140

ccctttgttg aaaagtctca atagcccttt ggtcttctga gactgtatct ttgatattct 16200

tggagtagac gagagtgtcg tgctccacca tgttggaag ctgctctagc caatacgcaa 16260

accgcctctc cccgcgcgtt ggccgattca ttaatgcagc tggcacgaca ggtttccga 16320

ctggaagcgg ggcagtgagc gcaacgcaat taatgtgagt tagctcactc attagcacc 16380

ccaggcttta cactttatgc ttccggctcg tatgttgtgt ggaattgtga gcggataaca 16440

atttcacaca ggaacagct atgaccatga ttacgcc 16477

<210> 39

<211> 15734

<212> DNA

<213> Artificial Sequence

<220><223> nucleotides coding pNGPJ0014

<400> 39

aagcttgata atttattga aaattcataa gaaaagcaaa cgttacatga attgatgaaa 60

caatacaaag acagataaag ccacgcacat ttaggatatt ggccgagatt actgaatatt	120
gagtaagatc acggaatttc tgacaggagc atgtcttcaa ttcagcccaa atggcagttg	180
aaatactcaa accgccccat atgcaggagc ggatcattca ttgtttgttt ggttgccttt	240
gccaacatgg gatlccaaga ttctgcagtc aaatctcggg gacgggcagg accggacggg	300
gcggtaccgg caggctgaag tccagctgcc agaaacccac gtcatgccag ttcccgtgct	360
tgaagccggc cgccgcagc atgccgcggg gggcatatcc gagcgctcg tgcattcgca	420
cgctcgggtc gttaggcagc ccgatgacag cgaccacgtt cttgaagccc tgtgcctcca	480
gggacttcag caggtaggtg tagagcgtgg agcccagtc cgtccgtgg tggcgggggg	540
agacgtacac ggtcgactcg gccgtccagt ctaggcgtt gcgtgccttc caggggcccc	600
cgtaggcgat gccggcgacc tcgccgtcca ctcggcgac gagccaggga tagcgtccc	660
gcagacggac gaggctgtcc gtccactcct gcggttcttg cggtcggta cggagttga	720
ccgtgcttgt ctgatgtag ttgttgacga ttgtgcagac cgccggcatg tccgcctcgg	780
tggcacggcg gatgtcgcc gggcgtcgtt ctgggtcat cgattcgatt ttgtgtatcg	840
agattggta tgaaattcag atgctagtgt aatgtattgg taatttggga agatataata	900
ggaagcaagg ctatttatcc atttctgaaa aggcgaaatg gcgtcaccgc gagcgtcacg	960
cgcattccgt tcttctgta aagcgttgtt ttgtacactt ttgactagcg aggcctggcg	1020
tgtcagcgta tctattcaaa agtcgttaat ggctgcggat caagaaaaag ttggaataga	1080
aacagaatac ccgcgaaatt caggcccggt tgccatgtcc tacacgccga aataaacgac	1140
caaattagta gaaaaataaa aactgactcg gatcttacg tcacgtcttg cgactgatt	1200
tgaaaaatct cagccatggc tgcaggtcga cgagtcagta ataacggcg tcaaagtgg	1260
tgcagccggc acacacaggt cgtgtttatc aactcaaagc acaataactt ttcctcaacc	1320
taaaaataag gcaattagcc aaaaacaact ttgcgtgtaa acaacgtca atacacgtgt	1380
cattttatta ttagctattg cttaccgcc ttagctttct cgtgacctag tcgtcctcgt	1440
ctttcttct tcttcttcta taaaacaata ccaaagagc tcttcttct cacaattcag	1500
atttcaattt ctcaaatct taaaacttt ctctcaattc tctctaccgt gatcaaggta	1560
aatttctgtt ttcttattc tctcaaatc ttcgattttg ttttcgttcg atcccaattt	1620
cgtatatgtt ctttggttta gattctgtta atcttagatc gaagacgatt ttctgggttt	1680
gatcgttaga tatcatctta attctcgatt agggtttcat agatatcatc cgatttggtc	1740
aaataatttg agttttgtcg aataattact cttcgatttg tgatttctat ctagatctgg	1800

tgtagtttc tagtttgtgc gatcgaattt gtcgattaat ctgagttttt ctgattaaca	1860
gccatggact ataaggacca cgacggagac tacaaggatc atgatattga ttacaaagac	1920
gatgacgata agatggcccc aaagaagaag cggaaggtcg gtatccacgg agtcccagca	1980
gccgacaaga agtacagcat cggcctggac atcggcacca actctgtggg ctgggccgtg	2040
atcaccgacg agtacaaggt gccagcaag aaattcaagg tgcgtggcaa caccgaccgg	2100
cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga aacagccgag	2160
gccacccggc tgaagagaac cgccagaaga agataacca gacggaagaa ccggatctgc	2220
tatctgcaag agatcttcag caacgagatg gccaaaggtgg acgacagctt cttccacaga	2280
ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc	2340
aacatcgtgg acgaggtggc ctaccacgag aagtaccca ccatctacca cctgagaaag	2400
aaactggtgg acagcacga caaggccgac ctgcggtga tctatctggc cctggccac	2460
atgatcaagt tccggggcca ctctctgac gagggcgacc tgaacccga caacagcgac	2520
gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggaaaacccc	2580
atcaacgcca gcggcgtgga cgccaaggcc atcctgtctg ccagactgag caagagcaga	2640
cggctggaaa atctgatcg ccagctgccc ggcgagaaga agaatggcct gttcggaac	2700
ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttga cctggccgag	2760
gatgccaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc	2820
cagatcggcg accagtacg cgacctgttt ctggccgcca agaacctgtc cgacgccatc	2880
ctgctgagcg acatcctgag agtgaacacc gagatcacca agccccct gagcgctct	2940
atgatcaaga gatagcaga gcaccaccag gacctgacc tgcigaaagc tctctgcgg	3000
cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc	3060
ggctacattg acggcggagc cagccaggaa gatttctaca agttcatcaa gcccacctg	3120
gaaaagatgg acggcaccga ggaactgtc gtgaagctga acagagagga cctgctgcgg	3180
aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac	3240
gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc	3300
gagaagatcc tgaccttccg catccctac tacgtgggcc ctctggccag gggaaacagc	3360
agattcgctt ggaatgaccg aaagagcgag gaaacatca cccctggaa cttcgaggaa	3420
gtggtggaca agggcgcttc cgcccagagc ttcacgagc ggatgacaa cttcgataag	3480
aacctgcca acgagaaggt gctgccaag cacagcctgc tgtacgagta cttaccgtg	3540

tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc cgccttcctg	3600
agcggcgagc agaaaaagc catcgtggac ctgctgttca agaccaaccg gaaagtgacc	3660
gtgaagcagc tgaagagga ctacttcaag aaaatcgagt gcttcgactc cgtggaaatc	3720
tccggcgtgg aagatcggtt caacgcctcc ctgggcacat accacgatct gctgaaaatt	3780
atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg	3840
ctgaccctga cactgtttga ggacagagag atgatcgagg aacggctgaa aacctatgcc	3900
cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc	3960
aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg	4020
gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac	4080
agcctgacct ttaaagagga catccagaaa gcccgaggtgt ccggccaggc cgatagcctg	4140
cacgagcaca ttgccaatct ggccggcagc cccgccatta agaaggcat cctgcagaca	4200
gtgaagggtg tggacgagct cgtgaaagtg atgggccggc acaagcccga gaacatcgtg	4260
atcgaatgg ccagagagaa ccagaccacc cagaaggagc agaagaacag ccgcgagaga	4320
atgaagcggg tcgaagaggg catcaaagag ctgggcagcc agatcctgaa agaacacccc	4380
gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaatgggcgg	4440
gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga tgtggaccat	4500
atcgtgcctc agagctttct gaaggacgac tccatcgaca acaagggtgt gaccagaagc	4560
gacaagaacc ggggcaagag cgacaacgtg ccctccgaag aggtcgtgaa gaagatgaag	4620
aactactggc ggcagctgct gaacccaag ctgattaccc agagaaagtt cgacaatctg	4680
accaaggccg agagaggcgg cctgagcgaa ctggataagg ccggcttcat caagagacag	4740
ctggtggaaa cccggcagat cacaagcac gtggcacaga tcctggactc ccgatgaac	4800
actaagtacg acgagaatga caagctgac cgggaagtga aagtgatcac cctgaagtec	4860
aagctggtgt ccgatttccg gaaggatttc cagttttaca aagtgcgcga gatcaacaac	4920
taccaccacg cccacgacgc ctacctgaac gccgtcgtgg gaaccgccct gatcaaaaag	4980
taccctaagc tggaaagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcggaag	5040
atgatcgcca agagcgagca ggaaatcggc aaggctaccg ccaagtactt cttctacagc	5100
aacatcatga actttttcaa gaccgagatt accctggcca acggcgagat ccggaagcgg	5160
cctctgatcg agacaaacgg cgaaaccggg gagatcgtgt gggataaggg ccgggatttt	5220
gccaccgtgc ggaagtgct gagcatgcc caagtgaata tcgtgaaaaa gaccgaggtg	5280
cagacaggcg gcttcagcaa agagtctatc ctgcccaaga ggaacagcga taagctgac	5340
gccagaaaga aggactggga ccctaagaag tacggcggct tcgacagccc caccgtggcc	5400

tattctgtgc tgggtgtggc caaagtggaa aagggaagt ccaagaaact gaagagtgtg	5460
aaagagctgc tgggatcac catcatggaa agaagcagct tcgagaagaa tcccatcgac	5520
tttctggaag ccaagggcta caaagaagtg aaaaaggacc tgatcatcaa gctgcctaag	5580
tactccctgt tcgagctgga aaacggccgg aagagaatgc tggcctctgc cggcgaactg	5640
cagaaggga acgaactggc cctgccctcc aaatatgtga acttcctgta cctggccagc	5700
cactatgaga agctgaaggg ctccccgag gataatgagc agaaacagct gtttgtgga	5760
cagcacaagc actacctgga cgagatcatc gagcagatca gcgagtcttc caagagagtg	5820
atcctggccg acgctaactt ggacaaagtg ctgtccgctt acaacaagca cgggataag	5880
cccatcagag agcaggccga gaatatcatc cacctgttta ccctgaccaa tctgggagcc	5940
cctgccgctt tcaagtactt tgacaccacc atcgaccgga agaggtacac cagcaccaaa	6000
gaggtgctgg acgccacctt gatccaccag agcatcaccg gcctgtacga gacacggatc	6060
gacctgtctc agctgggagg cgacaaaagg cggcgccca cgaaaaaggc cggccaggca	6120
aaaaagaaaa agtaagaatt cgcccttctt cgttgaacaa cggaaactcg acttgccttc	6180
cgcacaatac atcatctctt cttagctttt ttctctcttc ttcttctata cagttttttt	6240
ttgtttatca gcttacattt tcttgaaccg tagctttcgt ttctctcttt ttaactttcc	6300
attcggagtt tttgtatctt gtttcatagt ttgtcccagg attagaatga ttaggcacg	6360
aaccttcaag aatttgattg aataaaacat cttcattctt aagatatgaa gataatcttc	6420
aaaaggcccc tgggaatctg aaagaagaga agcaggccca tttatatggg aaagaacaat	6480
agtatttctt atataggccc atttaagttg aaaacaatct tcaaaagtcc cacatcgctt	6540
agataagaaa acgaagctga gtttatatac agctagagtc gaagtagtga ttgtgagacc	6600
gagagagtca tcgagggtga tctcaacctt gataactctg atgtggataa gttgttcac	6660
cagctcgtgc agacctaca ccagcttttc gaagagaacc ctatcaacgc tggctcagct	6720
tttagagcta gaaatagcaa gttaaataa ggctagtcg ttatcaactt gaaaaagtgg	6780
caccgagtcg gtgctttttt tctagagggc gaattcactg gccgtcgttt tacactggcc	6840
gtcgttttac aacgtcgtga ctgggaaaac cctggcgtaa cccaacttaa tcgccttgca	6900
gcacatcccc ctttcgccag ctggcgtaat agcgaagagg cccgcaccga tcgcccttcc	6960
caacagttgc gcagcctgaa tggcgaatgc tagagcagct tgagcttgga tcagattgtc	7020
gtttcccgcc ttcagtttaa actatcagtg ttgacagga tatattggcg ggtaaaccta	7080
agagaaaaga gcgtttatta gaataacgga tatttaaaag ggcgtgaaaa ggtttatccg	7140
ttcgtccatt tgatgtgca tgccaaccac aggggttcccc tcgggatcaa agtactttga	7200
tccaaccctt ccgtgctat agtgacgtcg gcttctgacg ttcagtgcag ccgtcttctg	7260

aaaacgacat gtcgcacaag tcctaagtta cgcgacaggc tgccgccctg cccttttcct	7320
ggcgttttct tgtcgcgtgt tttagtcgca taaagtagaa tacttgcgac tagaaccgga	7380
gacattacgc catgaacaag agcgccgccg ctggcctgct gggtatgcc cgcgtcagca	7440
ccgacgacca ggacttgacc aaccaacggg ccgaactgca cgcggccggc tgcaccaagc	7500
tgttttcga gaagatcacc ggcaccaggc gcgaccgcc ggagctggcc aggatgcttg	7560
accacctacg cctggcgac gttgtgacag tgaccaggct agaccgcctg gcccgagca	7620
cccgcgacct actggacatt gccgagcgca tccaggaggc cggcgccggc ctgcgtagcc	7680
tggcagagcc gtgggccgac accaccacgc cggccggccg catggtgttg accgtgttcg	7740
ccggcattgc cgagttcgag cgttcctaa tcatcgaccg caccggagc gggcgcgagg	7800
ccgccaaggc ccgaggcgtg aagtttgcc ccgcctac cctaccccg gcacagatcg	7860
cgcaccccc cgagctgac gaccaggaag gccgcaccgt gaaagaggcg gctgcactgc	7920
ttggcgtgca tcgctcgacc ctgtaccgcg cacttgagcg cagcgaggaa gtgacgcca	7980
ccgaggccag gcggcgcggt gccttcctg aggacgcatt gaccgaggcc gacgccctgg	8040
cggcccgca gaatgaacgc caagaggaac aagcatgaaa ccgcaccagg acggccagga	8100
cgaaccgttt ttattaccg aagagatcga ggcggagatg atcgccggcg ggtacgtgtt	8160
cgagccgcc gcgcacgtct caaccgtcg gctgcatgaa atcctggccg gtttgtctga	8220
tgccaagctg gcggcctggc cggccagctt ggccgctgaa gaaaccgagc gccgcgtct	8280
aaaaaggtga tgtgtatttg agtaaaacag ctgctcat gcggtcgtg cgtatatgat	8340
gcgatgagta aataaacaac tacgcaaggg gaacgcata aggttatcgc tgtactaac	8400
cagaaaggcg ggtcaggcaa gacgaccatc gcaaccatc tagcccgcg cctgcaactc	8460
gccggggccg atgttctgtt agtcgattcc gatcccagg gcagtgccg cgattggcg	8520
gccgtgcggg aagatcaacc gctaaccgtt gtcggcatcg accgccgac gattgaccgc	8580
gacgtgaagg ccatcgccg gcgcgacttc gtagtgatcg acggagcgcc ccaggcgcg	8640
gacttgctg tgtccgcat caaggcagcc gacttcgtgc tgattccgt gcagccaagc	8700
ctttacgaca tatgggccac cgccgacctg gtggagctgg ttaagcagcg cattgaggtc	8760
acggatggaa ggctacaagc gccctttgtc gtgtcgcggg cgatcaaagg cacgcgcatc	8820
ggcgtgagg ttgccgagc gctggccggg tacgagctgc ccattcttga gtcccgtatc	8880
acgcagcgcg tgagctacc aggcactgcc gccgccgca caaccgttct tgaatcagaa	8940
cccgaggcg acgtgcccg cgaggtccag gcgtggccg ctgaaattaa atcaaaactc	9000

at ttgag tta atgagg taaa gagaaa atga gcaaa agcac aaacac gcta agtgccggcc	9060
gtccgagcgc acgcagcagc aaggctgcaa cgttgccag cctggcagac acgccagcca	9120
tgaagcgggt caactttcag ttgccggcgg aggatcacac caagctgaag atgtacgcgg	9180
tacgccaagg caagaccatt accgagctgc tatctgaata catcgcgag ctaccagagt	9240
aatgagcaa atgaataaat gagtagatga attttagcgg ctaaaggagg cggcatggaa	9300
aatcaagaac aaccaggcac cgacgccgtg gaatgcccc tgtgtggagg aacgggcgg	9360
tggccaggcg taagcggctg ggttgtctgc cggccctgca atggcactgg aacccccaag	9420
cccgaggaat cggcgtgacg gtgcgaaacc atccggcccc gtacaaatcg gcgcggcgct	9480
gggtgatgac ctggtggaga agttgaaggc gcgcgagcc gccagcggc aacgcattga	9540
ggcagaagca cggcccggtg aatcgtggca agcggccgct gatcgaatcc gcaaagaatc	9600
ccggcaaccg cggcgagcgc gtgcgccgtc gattaggaag ccgccaagg gcgacgagca	9660
accagat ttt ttcgttccga tgctctatga cgtgggcacc cgcgatagtc gcagcatcat	9720
ggacgtggcc gttttccgtc tgctgaagcg tgaccgacga gctggcgagg tgatccgcta	9780
cgagcttcca gacgggcacg tagaggtttc cgcagggccg gccggcatgg ccagtgtgtg	9840
ggattacgac ctggtactga tggcggtttc ccatctaacc gaatccatga accgataccg	9900
ggaagggaag ggagacaagc cggccgcgt gtccgtcca cacgttgcgg acgtactcaa	9960
gttctgccgg cgagccgatg gcggaaagca gaaagacgac ctggtagaaa cctgcattcg	10020
gttaaacacc acgcacgttg ccatgcagcg tacgaagaag gccaagaacg gccgcctggt	10080
gacggtatcc gaggggtgaag ccttgattag ccgtacaag atcgtaaaga gcgaaaccgg	10140
gcggccggag tacatcgaga tcgagctagc tgattggatg taccgcgaga tcacagaagg	10200
caagaaccg gacgtgctga cggttcacc cgattacttt ttgatcgatc ccggcatcgg	10260
ccgttttttc taccgcctgg cacgccgcgc cgcaggcaag gcagaagcca gatggttgtt	10320
caagacgac tacgaacgca gtggcagcgc cggagagttc aagaagtctt gtttaccgt	10380
gcgcaagctg atcgggtcaa atgacctgcc ggagtacgat ttgaaggagg aggcggggca	10440
ggctggcccc atcctagtc tgcgtaccg caacctgac gagggcgaag catccgccgg	10500
ttcctaattg acggagcaga tgctagggca aattgcccta gcaggggaaa aaggtcgaaa	10560
agcactcttt cctgtggata gcacgtacat tgggaacca aagccgtaca ttgggaaccg	10620
gaacccttac attgggaacc caaagccgtc cattgggaac cggtcacaca tgtaagtac	10680
tgataaaaa gagaaaaaag gcgatttttc cgcctaaaac tctttaaaac ttattaaaac	10740
tcttaaaacc cgctggcct gtgcataact gtctggccag cgcacagccg aagagctgca	10800
aaaagcgct acccttcggt cgctgcgtc cctacgccc gccgcttcgc gtcggcctat	10860

cgcgccgct ggccgctcaa aaatggctgg cctacggcca ggcaatctac cagggcgcg	10920
acaagccgcg ccgtcgccac tcgaccgccc gcgcccacat caaggcacc tcctcgcg	10980
gtttcggtga tgacggtgaa aacctctgac acatgcagct cccggagacg gtcacagctt	11040
gtctgtaagc ggatgccggg agcagacaag cccgtcaggg cgcgtcagcg ggtgttggcg	11100
gggtgcgggg cgagccatg acccagtcac gtagcgatag cggagtgtat actggcttaa	11160
ctatcgggca tcagagcaga ttgtactgag agtgcacat atgcggtgtg aaataccgca	11220
cagatgcgta aggagaaaat accgcatcag gcgctcttc gcttctcgc tactgactc	11280
gctgcgctcg gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg cgtaatacg	11340
gttatccaca gaatcagggg ataacgcagg aaagaacatg tgagcaaaag gccagcaaaa	11400
ggccaggaac cgtaaaaagg ccgctgtgct ggcgttttc cataggctcc gccccctga	11460
cgagcatcac aaaaatcgac gctcaagtca gaggtggcga aaccgcagag gactataaag	11520
ataccaggcg tttccccctg gaagctccct cgtgcgctct cctgttccga cctgcccgt	11580
taccgatac ctgtccgct tttcccttc gggaagcgtg gcgctttctc atagctcacg	11640
ctgtaggtat ctgattcgg tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc	11700
ccccgttcag cccgaccgt gcgccttacc cggtaactat cgtcttgagt ccaaccgggt	11760
aagacacgac ttatcgccac tggcagcagc cactggtaac aggattagca gagcgaggta	11820
tgtaggcggt gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaaggac	11880
agtatttggt atctgcgctc tgctgaagcc agttacctc ggaaaaagag ttggtagctc	11940
ttgatccggc aaacaaacca ccgctggtag cgggtgtttt tttgtttgca agcagcagat	12000
tacgcgcaga aaaaaaggat ctcaagaaga tcttttgatc ttttctacgg ggtctgacgc	12060
tcagtggaac gaaaactcac gttaagggat ttgggtcatg cattctaggt actaaaacaa	12120
ttcatccagt aaaatataat attttatatt ctcccaatca ggcttgatcc ccagtaagtc	12180
aaaaaatagc tcgacatact gttcttcccc gatatcctcc ctgatcgacc ggacgcagaa	12240
ggcaatgtea taccacttgt ccgcccctgc gcttctccca agatcaataa agccacttac	12300
tttgccatct ttcacaaaga tgttgctgtc tcccaggtcg ccgtgggaaa agacaagttc	12360
ctcttcgggc ttttcgctc ttaaaaaatc atacagctcg cgcggatctt taaatggagt	12420
gtcttcttc cagttttcgc aatccacatc ggccagatcg ttattcagta agtaatccaa	12480
ttcggctaag cggtgtgcta agctattcgt atagggacaa tccgatatgt cgatggagtg	12540
aaagagcctg atgcactccg catacagctc gataatcttt tcagggtttt gttcatcttc	12600
atactcttc gagcaaaagg cgccatcggc ctactcatg agcagattgc tccagccatc	12660
atgccgttca aagtgaggga cctttggaac aggcagcttt ccttcagcc atagcatcat	12720

gtccttttcc cgttccacat cataggtggg ccctttatac cggtgtccg tcatttttaa	12780
atataaggttt tcattttctc ccaccagctt atatacctta gcaggagaca ttccttcctg	12840
atcttttacg cagcggatatt ttctgatcag ttttttcaat tccggtgata ttctcathtt	12900
agccatttat tatttccttc ctcttttcta cagtatttaa agatacccca agaagcta	12960
tataacaaga cgaactccaa ttactgttc ctgtcattct aaaaccttaa ataccagaaa	13020
acagcttttt caaagtgtt ttcaaagttg gcgtataaca tagtatcgac ggagccgatt	13080
ttgaaaccgc gggtatcaca ggcagcaacg ctctgtcatc gttacaatca acatgctacc	13140
ctccgcgaga tcacccgtgt ttcaaaccg gcagcttagt tgccgttctt ccgaatagca	13200
tcggtaacat gagcaaagtc tgccgcctta caacggctct cccgtgacg cgttcccgga	13260
ctgatgggct gcctgtatcg agtgggtatt ttgtgccgag ctgccggtcg gggagctgtt	13320
ggctggctgg tggcaggata tattgtgggtg taacaaatt gacgcttaga caacttaata	13380
acacattgcg gacgttttta atgtactgaa ttaacgccga attaattcgg gggatctgga	13440
tttttagtact ggattttgtt tttaggaatt agaaatttta ttgatagaag tattttacaa	13500
atacaaatat atactaagggt ttctttatat gctcaacaca tgagcgaaac cctataggaa	13560
ccctaattcc ctatctggg aactactcac acattattat ggagaaactc gagcttgtcg	13620
atcgacagat ccggtcggca tctactctat ttctttgccc tcggacgagt gctggggcgt	13680
cggtttccac tatcggcgag tacttctaca cagccatcgg tccagacggc cgcgcttctg	13740
cgggcgattt gtgtacgcc gacagtcctg gctccggatc ggacgattgc gtcgcatcga	13800
ccctgcgcc aagctgcac atcgaaattg ccgtcaacca agctctgata gagttggtea	13860
agaccaatgc ggagcatata cgcccgaggt cgtggcgatc ctgcaagctc cggatgcctc	13920
cgctcgaagt agcgcgtctg ctgtccata caagccaacc acggcctcca gaagaagatg	13980
ttggcgacct cgtattggga atccccgaac atgcctcgc tccagtcaat gaccgctgtt	14040
atgcggccat tgtccgtcag gacattgttg gagccgaaat ccgctgcac gaggtgccgg	14100
acttcggggc agtcctcggc ccaaagcatc agtcatcga gagcctgcgc gacggacgca	14160
ctgacggtgt cgtccatcac agtttgccag tgataacat ggggatcagc aatcgcgcat	14220
atgaaatcac gccatgtagt gtattgaccg attccttgcg gtccgaatgg gccgaaccg	14280
ctcgtctggc taagatcggc cgcagcgatc gcatccatag cctccgcgac cggttgtaga	14340
acagcgggca gticggtttc aggcaggtct tgcaactga caccctgtgc acggcgggag	14400
atgcaatagg tcaggctctc gctaaactcc ccaatgtcaa gcacttccgg aatcgggagc	14460

gcgccgatg caaagtgccg ataaacataa cgatctttgt agaaaccatc ggcgagcta	14520
tttaccgcga ggacatatcc acgccctcct acatcgaagc tgaaagcacg agattcttcg	14580
ccctccgaga gctgcatcag gtcggagacg ctgtcgaact tttcgatcag aaacttctcg	14640
acagacgtcg cggtagagttc aggctttttc atatctcatt gccccccggg atctgcgaaa	14700
gctcgagaga gatagatttg tagagagaga ctggtgattt cagcgtgtcc tctccaatg	14760
aaatgaactt cttatatag aggaaggtct tgcgaaggat agtgggattg tgcgtcatcc	14820
cttacgtcag tggagatatc acatcaatcc acttgctttg aagacgtggt tggacgtct	14880
tctttttcca cgatgctcct cgtgggtggg ggtccatctt tgggaccact gtcggcagag	14940
gcatcttgaa cgatagcctt tcctttatcg caatgatggc attttaggt gccaccttcc	15000
ttttctactg tccttttgat gaagtgcag atagctgggc aatggaatcc gaggaggttt	15060
cccgatatta ccctttgttg aaaagtctca atagcccttt ggtcttctga gactgtatct	15120
ttgatattct tggagtagac gagagtgtcg tgcaccacca tgttatcaca tcaatccact	15180
tgctttgaag acgtggttgg aacgtcttct tttccacga tgcctctcgt ggggtgggggt	15240
ccatctttgg gaccactgtc ggcagaggca tcttgaacga tagcctttcc tttatcgcaa	15300
tgatggcatt ttaggtgcc accttccttt tctactgtcc tttgatgaa gtgacagata	15360
gctgggcaat ggaatccgag gaggtttccc gatattacc tttgttgaaa agtctcaata	15420
gcccccttgg ctctgagac tgtatctttg atattcttgg agtagacgag agtgtcgtgc	15480
tccaccatgt tggcaagctg ctctagccaa tacgcaaacc gcctctcccc gcgcgttggc	15540
cgattcatta atgcagctgg cagcacaggt ttcccactg gaaagcgggc agtgagcgca	15600
acgcaattaa tgtgagttag ctactcatt aggcacccca ggctttacac tttatgcttc	15660
cggctcgtat gttgtgtgga attgtgagcg gataacaatt tcacacagga aacagctatg	15720
accatgatta cgcc	15734