(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0022749
(43) 공개일자 2023년02월16일

(51) 국제특허분류(Int. Cl.)
C12N 15/10 (2017.01) *A01H 1/06* (2006.01)
A01H 5/10 (2018.01) *C07K 14/415* (2006.01)
C12N 15/113 (2010.01) *C12N 15/82* (2006.01)
C12N 9/22 (2006.01)
(52) CPC특허분류
C12N 15/102 (2013.01)
A01H 1/06 (2021.01)
(21) 출원번호 10-2021-0104895
(22) 출원일자 2021년08월09일
심사청구일자 2021년08월09일

(71) 출원인
경상국립대학교산학협력단
경상남도 진주시 진주대로 501 (가좌동)
(72) 발명자
김민철
경상남도 진주시 하대로 142, 105동 2001호(하대동, 현대아파트)
전현진
경상남도 진주시 하대로 142, 105동 2001호(하대동, 현대아파트)
(뒷면에 계속)
(74) 대리인
최규환

전체 청구항 수 : 총 12 항

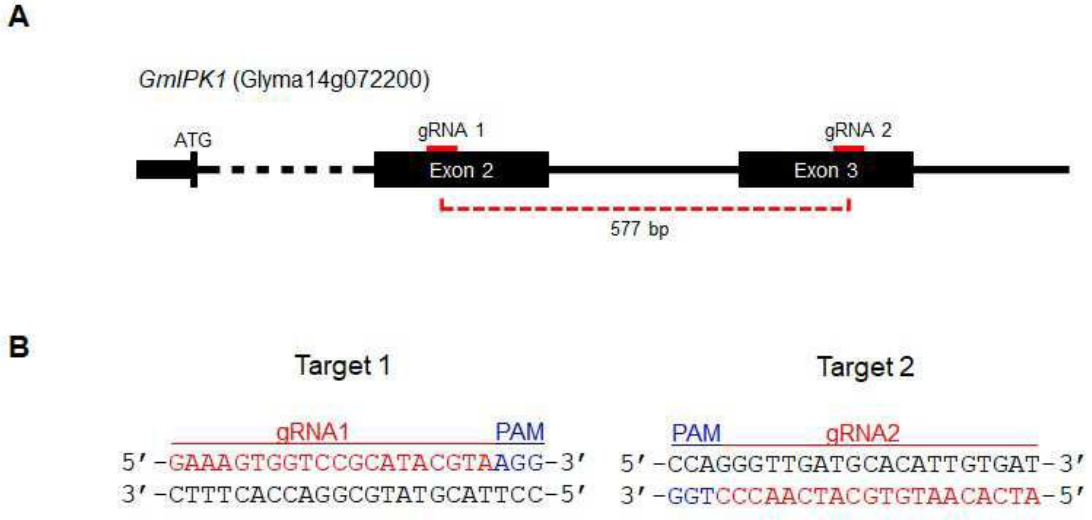
(54) 발명의 명칭 **GmIPK1 유전자교정 시스템 및 이의 용도**

(57) 요 약

본 발명은 콩에서 피틴산 생합성을 조절하는 *GmIPK1* 및 이의 상동체 유전자를 동시에 교정시키는 CRISPR/Cas 시스템을 구축하였고, 이를 이용한 유전자교정 기반 기술을 통해 피틴산 함량이 낮은 유전자교정 콩 형질전환체 개발하였다. 본 발명에 따른 CRISPR/Cas 기반의 콩 유전자교정 시스템 또는 벡터, 이를 포함하는 콩 유전자교정용 조성물은 콩에 존재하는 IPK1 단백질의 발현을 동시에 억제하는 효과를 가진다.

본 발명은 콩 식물체에서 *GmIPK1* 유전자를 특이적으로 교정함으로써, 항영양인자인 피틴산 함량이 감소된 콩을 생산하는 새로운 식물체를 제공할 수 있으며, 콩 종자에서 피틴산 함량이 감소된 콩을 육성함으로써 식용 및 사료용으로써의 콩의 이용성 및 가치를 증진시키고, 동시에 깨끗한 농업생태계를 보존하는데 기여할 수 있을 것이다.

대 표 도 - 도1



(52) CPC특허분류
A01H 5/10 (2022.05)
C07K 14/415 (2013.01)
C12N 15/113 (2013.01)
C12N 15/8242 (2013.01)
C12N 9/22 (2013.01)
C12N 2310/20 (2017.05)
(72) 발명자
조현민
경상남도 진주시 가호로 26, 213동 303호(가좌동, 진주가좌주공아파트)
박미숙
경상남도 진주시 초전북로 148-3(초전동)
정춘희
경상남도 진주시 평거로139번길 11-3, 203동 505호(신안동, 신안흥한타운2차아파트)

이 발명을 지원한 국가연구개발사업

과제고유번호	1395067797
과제번호	PJ014769012021
부처명	농촌진흥청
과제관리(전문)기관명	농촌진흥청
연구사업명	차세대농작물신육종기술개발(R&D)
연구과제명	유전자교정 기반 가치증진 식량작물 개발(1주관)
기 여 율	1/2
과제수행기관명	경상국립대학교산학협력단
연구기간	2021.01.01 ~ 2021.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1395073209
과제번호	PJ016533022022
부처명	농촌진흥청
과제관리(전문)기관명	농촌진흥청
연구사업명	차세대농작물신육종기술개발(R&D)
연구과제명	유용소재 및 농업형질 개선 유전자교정 콩 개발 및 시장경쟁력 확보(2공동)
기 여 율	1/2
과제수행기관명	경상국립대학교산학협력단
연구기간	2022.01.01 ~ 2023.12.31

명 세 서

청구범위

청구항 1

Cas 엔도뉴클레아제(endonuclease); 및 콩의 *GmIPK1* 유전자 내 2 곳의 표적 부위를 각각 표적하는 gRNA1 및 gRNA2;를 포함하는, CRISPR/Cas 기반의 콩 유전자교정 시스템.

청구항 2

제1항에 있어서,

상기 gRNA1 및 gRNA2는 이를 암호화하는 핵산 서열이 각각 서열번호 1의 염기서열 및 서열번호 2의 염기서열인 것을 특징으로 하는, CRISPR/Cas 기반의 콩 유전자교정 시스템.

청구항 3

제2항에 있어서,

상기 Cas 엔도뉴클레아제(endonuclease)는 코돈-최적화된 Cas9 핵산을 포함하는 것으로, 서열번호 5 내지 7 중 어느 하나의 염기서열을 포함하는 것을 특징으로 하는, CRISPR/Cas 기반의 콩 유전자교정 시스템.

청구항 4

제1항에 있어서,

상기 콩 유전자교정 시스템은 리보핵산단백질(Ribonucleoprotein, RNP)이고, 상기 gRNA1 및 gRNA2는 이를 암호화하는 핵산 서열이 각각 서열번호 1의 염기서열 및 서열번호 2의 염기서열인 것을 특징으로 하는, CRISPR/Cas 기반의 콩 유전자교정 시스템.

청구항 5

코돈-최적화된 Cas9 핵산; 및 콩의 *GmMIPS1* 유전자 및 이의 상동체(homolog) 유전자를 동시에 표적하는 gRNA1 및 gRNA2를 각각 암호화하는 핵산;이 작동 가능하게 연결된, 콩 유전자교정을 위한 벡터.

청구항 6

제5항에 있어서,

상기 벡터는 작동 가능하게 연결된 하나 이상의 프로모터, 식물선별마커 및 핵위치 신호(Nuclear localization signal, NLS)가 더 포함되는 것을 특징으로 하는, 콩 유전자교정을 위한 벡터.

청구항 7

제5항에 있어서,

상기 코돈-최적화된 Cas9 핵산은 서열번호 5 내지 7 중 어느 하나의 염기서열을 포함하는 것이고, 상기 gRNA1 및 gRNA2를 암호화하는 핵산은 각각 서열번호 1의 염기서열 및 서열번호 2의 염기서열로 이루어진 것; 또는 서

열번호 3의 염기서열 및 서열번호 4의 염기서열로 이루어진 것;을 특징으로 하는, 콩 유전자교정을 위한 벡터.

청구항 8

제1항의 CRISPR/Cas 기반의 콩 유전자교정 시스템 또는 제5항의 벡터를 포함하는, 콩 유전자교정용 조성물.

청구항 9

제8항에 있어서,

gRNA1 및 gRNA2를 암호화하는 핵산이 각각 서열번호 1의 염기서열 및 서열번호 2의 염기서열로 이루어진 것이고, 콩의 *GmIPK1* 유전자를 결실시키는 것을 특징으로 하는, 콩 유전자교정용 조성물.

청구항 10

제1항의 CRISPR/Cas 기반의 콩 유전자교정 시스템 또는 제5항의 벡터를 이용하여 콩을 형질전환하는 단계를 포함하고, 콩의 *GmIPK1* 유전자를 결실시키는 것을 특징으로 하는, 콩 유전자교정 방법.

청구항 11

제10항의 콩 유전자교정 방법으로 형질전환된, 콩에서 피틴산 생합성을 조절하는 효소 *GmIPK1* 유전자가 교정된 콩 형질전환체.

청구항 12

제10항의 방법으로 형질전환된 콩 형질전환체의 종자로써, 피틴산 생합성을 조절하는 효소 *GmIPK1* 유전자가 교정된 콩 종자.

발명의 설명

기 술 분 야

[0001] 본 발명은 콩의 항영양인자인 피틴산 함량이 낮은 유전자교정 콩 형질전환체를 개발하기 위한, 콩의 *GmIPK1* 유전자를 교정하는 유전자교정 시스템에 관한 것이다. 구체적으로 콩에서 *GmIPK1* 유전자를 결실시키는 유전자교정 시스템, 이를 포함하는 콩 유전자 교정용 조성물, 이를 이용하는 콩 유전자 교정 방법, 피탄산 함량이 낮은 콩 형질전환체 및 이의 종자에 관한 것이다.

배 경 기 술

[0003] 콩은 단백질, 지방 및 식이섬유 성분이 풍부하며, 이소플라본, 안토시아닌, 레시틴, 루테인, 사포닌 등 다양한 건강 기능성 물질을 함유하고 있어 식품으로써 높은 영양학적 가치를 가지고 있을뿐만 아니라, 사료 및 가공용 등의 산업적 가치도 큰 작물이다.

[0004] 하지만 콩에는 피틴산, 난소화성 올리고 당, 알러지 유발원 등과 같은 항영양인자들(anti-nutritional factors)이 함유되어 있어 식품으로써 콩의 이용성을 저해하고 있다.

[0005] 콩 종자에는 피틴산(phytic acid, phytate)이 건물중 기준 약 2 %로 다량 함유되어 있는데, 이러한 피틴산(phytic acid, phytate)은 칼슘, 철, 아연과 같은 미네랄들과 결합하여 침전물을 형성함으로써 미네랄들의 흡수를 방해하는 대표적인 항영양인자(anti-nutritional factor)이다. 피틴산은 식용으로써의 콩의 가치를 감소시킬

뿐만 아니라 콩의 사료적 가치 또한 감소시키는 것으로 알려져 있다. 특히 피틴산이 다량으로 함유된 가축의 분뇨는 부영양화에 의한 수질오염의 주요 요인으로 알려져 있다. 이런 이유로, 피틴산 함량이 감소된 콩 개발은 식용 및 사료용으로써의 콩의 이용성 및 가치를 증진 시키고 농업환경을 보존하는 등 경제적 및 환경적 가치가 크다.

[0006]

[0007] 박테리아, 인간을 포함하는 동물 및 식물 등 다양한 개체에서 유전자의 도입, 결실 또는 변이를 야기할 수 있는 CRISPR/Cas 시스템을 이용한 유전자교정(gene editing) 기술은 유전자의 기능을 밝히는 연구 방법나 질병의 치료, 새로운 식물 품종의 개발 등을 위한 도구로 활발히 연구되고 이용되고 있다. 최근 CRISPR/Cas 시스템을 이용한 유전자교정(gene editing) 연구가 식물에서도 다양하게 적용되고 있다.

[0008] CRISPR/Cas 시스템은 Cas 엔도뉴클레아제(endonuclease) 단백질과 표적 유전자 서열을 인식하는 CRISPR RNA(crRNA)와 crRNA와 결합하면서 상기 Cas 엔도뉴클리아제에 결합하는 transactivating CRISPR RNA(tracrRNA)가 복합체를 형성한 것이다. 여기서, crRNA와 tracrRNA는 링커로 연결한 단일 가닥 가이드 RNA(single guide RNA, sgRNA) 형태가 주로 이용되고, 이 단일 가닥 가이드 RNA(sgRNA)는 Cas 엔도뉴클리아제(endonuclease)를 잘라야 할 표적 유전자의 이중가닥 DNA 염기서열로 안내한다. 표적 유전자 부위에 위치한 Cas 엔도뉴클리아제(endonuclease)는 표적 유전자 서열과 이웃하고 있는 프로토스페이서 인접 모티프(protospacer adjacent motif, PAM)를 인식한 후, 표적 핵산 서열의 내부 또는 외부 염기서열(base pair, bp)에서 DNA 이중가닥 절단(double strand breaks, DSB)이 일어나게 된다(Jinek et al., 2012).

[0009] 유전자가위(CRISPR/Cas) 시스템에 의해 절단된 표적 핵산은 오류가 발생하기 쉬운 DNA 복구기작(error-prone DSB repair)이 일어나는 동안에 상동재조합(Homology directed repair, HDR) 또는 비상동말단연결(non-homologous end joining, NHEJ) 과정의 DNA 복구기작을 통해 복구가 일어난다. 비상동말단연결(NHEJ)의 DNA 복구기작을 통해서는 절단된 DNA 부위 사이에 무작위적 염기의 삽입(insertion) 또는 결실(deletion)의 인델(insertion and deletion, indel) 돌연변이가 일어나게 된다. 그 결과 유전자의 코딩 부분에서 틀이동 변이(frameshift mutation) 또는 조기종결 변이(premature mutation)가 발생하여 표적 유전자가 제거(Knock-out)된다. 반면에 상동재조합(HDR)의 DNA 복구기작은 절단된 DNA를 복구하기 위하여 공여자 DNA(Donor DNA)를 필요로 하는데, 이 공여자 DNA의 서열을 주형으로 하여 내부의 목적 유전자의 서열이 정교하게 교체됨으로써, 유전자 편집이 완성된다.

[0011] 유전자교정(gene editing)은 생명체의 유전정보를 자유롭게 편집하는 기술로써, 인간을 포함하여 동물, 식물, 미생물의 유전정보를 변화시켜 그 활용범위를 획기적으로 확장시킬 수 있다. 현재 많은 연구자들이 유전자교정 연구를 추진하고 있으며 유전자교정 원천특허 확보, 유전자교정 벡터 개발, 유전자교정 인터넷 플랫폼 구축 등 기반 기술 관련 연구를 진행 중이다. CRISPR 기반의 유전자교정 시장은 매년 빠르게 성장하고 있으며, 향후 10년 내에 현재의 약 20배 이상의 규모로 증가 될 것으로 예상되고 있다(Inkwood Research, 2019).

[0012] 그러나, CRISPR 기술을 이용한 작물 개량 연구는 초기 단계이며, 특히 콩은 다른 작물에 비하여 형질전환체를 제조하는 기간이 길고 상대적으로 효율이 낮기 때문에, 형질전환이 어려운 콩의 경우 유전자교정 기술 기반의 콩 육종소재 개발이 일부 연구실에서만 제한적으로 이루어지고 있는 실정이다.

[0014] 현재 식용 및 사료용으로써의 콩의 이용성 및 가치를 증진 시키고 농업환경을 보존하기 위해서 피틴산 함량이 감소된 콩 형질전환체 및 이의 종자를 개발하기 위한 다양한 연구가 지속적으로 진행되고 있다.

[0015] 하지만, 유전자가위(CRISPR/Cas) 시스템을 도입하여 피탄산 함량이 감소된 콩 형질전환체 및 이의 종자를 개발하기 위한 끊임없는 노력 및 연구에도 불구하고, 그 성과는 미미한 실정이다. 이에, 피탄산 함량이 감소된 콩 형질전환체 및 이의 종자를 생산하기 위한 최적의 효과적인 유전자가위(CRISPR/Cas) 시스템이 절실이 요구되는 상황이다.

선행기술문헌

특허문헌

[0017] (특허문헌 0001) KR 10-2015-0016588 A

(특허문헌 0002) US 2020/0190494 A1

(특허문헌 0003) KR 10-2021-0039306 A

(특허문헌 0004) KR 10-2018-0117785 A

비특허문헌

[0018] (비특허문헌 0001) Jinek, M. et al., A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity, Science, Vol. 337, 816-821(2012)

(비특허문헌 0002) Inkwood Research, GLOBAL CRISPR MARKET FORECAST 2019-2027(2019)

(비특허문헌 0003) Jacobs, TB. et al., Targeted genome modifications in soybean with CRISPR/Cas9, BMC Biotechnology, Vol. 15, No. 16, pp. 1-10(2015)

발명의 내용

해결하려는 과제

[0019] 본 발명은 상기와 같은 요구를 해결하고 종래기술의 문제점을 극복하기 위한 것으로, 콩의 항영양인자인 피틴산의 함량이 낮은 유전자교정 콩 형질전환체 개발을 위해서, 콩에서 *GmIPK1* 유전자를 결실시키는 CRISPR/Cas 기반의 콩 유전자교정 시스템을 제공하는 것을 목적으로 한다.

[0021] 본 발명은 또한, 상기 콩 유전자교정 시스템을 포함하는 벡터, 콩 유전자교정 조성물 및 이를 이용하는 콩 유전자교정 방법을 제공하는 것을 목적으로 한다.

[0023] 또한, 본 발명은 상기 콩 유전자교정 시스템 또는 콩 유전자교정 조성물로 형질전환되어 콩 종자에서 피틴산 생합성을 조절하는 효소 *GmIPK1* 유전자가 교정된 콩 형질전환체 및 이의 종자를 제공하는 것을 목적으로 한다.

[0025] 본 출원의 다른 목적 및 본 발명이 이루고자 하는 기술적 과제는 이상에서 언급한 목적에 제한되지 않으며, 본 명세서에 기재된 청구범위 및 도면과 함께 하기의 발명의 설명에 의해 보다 명확해질 것이다. 또한 본 명세서에 기재되지 않은 본 발명이 이루고자 하는 기술적 과제는 본 발명의 기술 분야에서 통상의 지식을 가진 자(이하 '통상의 기술자'라 함)가 명확하게 이해하고 유추할 수 있는 것을 포함한다.

과제의 해결 수단

[0027] 상기 과제를 해결하기 위해, 다음의 발명을 제공한다.

[0029] 본 발명은 콩의 항영양인자인 피틴산 함량이 낮은 유전자교정 콩 형질전환체 개발을 위해서 *GmIPK1* 유전자를 결실시키는, CRISPR/Cas 기반의 콩 유전자교정 시스템을 제공한다. 여기서 상기 CRISPR/Cas 기반의 콩 유전자교정 시스템은 Cas 엔도뉴클레아제(endonuclease); 및 콩의 *GmIPK1* 유전자를 특이적으로 표적하는 gRNA1 및 gRNA2; 를 포함할 수 있다.

[0030] 본 발명의 일 구체예로, 상기 CRISPR/Cas 기반의 콩 유전자교정 시스템에서 상기 gRNA1 및 gRNA2는 이를 암호화

하는 핵산 서열이 각각 서열번호 1의 염기서열 및 서열번호 2의 염기서열인 것일 수 있다.

[0031] 본 발명의 다른 구체예로, 상기 Cas 엔도뉴클레아제(endonuclease)는 식물 코돈-최적화된 Cas9 핵산(pcoCas9), 애기장대 코돈-최적화된 Cas9 핵산(acoCas9) 또는 GC 함량이 높은 식물 코돈-최적화된 Cas9 핵산(pcohCas9)을 포함하는 것일 수 있고, 구체적으로 서열번호 5 내지 서열번호 7 중 어느 하나의 염기서열을 포함하는 것일 수 있다.

[0032] 본 발명의 또다른 구체예로, 상기 콩 유전자교정 시스템은 리보핵산단백질(Ribonucleoprotein, RNP)이고, 상기 gRNA1 및 gRNA2는 이를 암호화하는 핵산 서열이 각각 서열번호 1의 염기서열 및 서열번호 2의 염기서열인 것일 수 있다.

[0034] 또한, 본 발명은 코돈-최적화된 Cas9 핵산; 및 콩의 *GmIPK1* 유전자를 특이적으로 표적하는 gRNA1 및 gRNA2를 암호화하는 핵산;이 작동 가능하게 연결된, 콩 유전자교정을 위한 벡터를 제공한다.

[0035] 본 발명의 일 구체예로, 상기 벡터는 작동 가능하게 연결된 하나 이상의 프로모터, 식물 선별마커 및 하나 이상의 핵 위치 신호(Nuclear localization sequence, NLS)가 더 포함되는 것일 수 있다.

[0036] 본 발명의 다른 구체예로, 상기 벡터에서 상기 gRNA1 및 gRNA2를 암호화하는 핵산 서열은 각각 서열번호 1의 염기서열 및 서열번호 2의 염기서열인 것일 수 있다. 또한, 코돈-최적화된 Cas9 핵산은 식물 코돈-최적화된 Cas9 핵산(pcoCas9), 애기장대 코돈-최적화된 Cas9 핵산(acoCas9) 또는 GC 함량이 높은 식물 코돈-최적화된 Cas9 핵산(pcohCas9)을 포함하는 것일 수 있고, 구체적으로 서열번호 5 내지 서열번호 7 중 어느 하나의 염기서열을 포함하는 것일 수 있다.

[0038] 또한, 본 발명은 상기 CRISPR/Cas 기반의 콩 유전자교정 시스템 또는 상기 벡터를 포함하는, 콩 유전자교정용 조성물을 제공한다.

[0039] 본 발명의 일 구체예로, 상기 콩 유전자교정용 조성물에서 상기 gRNA1 및 gRNA2를 암호화하는 핵산 서열이 각각 서열번호 1의 염기서열 및 서열번호 2의 염기서열인 것이고, 콩의 *GmIPK1* 유전자를 결실시키는 것을 특징으로 할 수 있다.

[0041] 또한, 본 발명은 상기 CRISPR/Cas 기반의 콩 유전자교정 시스템 또는 상기 벡터를 이용하여 콩을 형질전환하는 단계를 포함하고, 콩의 *GmIPK1* 유전자를 결실시키는 것을 특징으로 하는, 콩 유전자교정 방법을 제공한다.

[0042] 본 발명의 일 구체예로, 상기 콩 유전자교정 방법은 상기 CRISPR/Cas 기반의 콩 유전자교정 시스템 또는 상기 벡터를 콩 식물세포에 도입하여 표적 유전자를 교정하는 단계 및 상기 표적 유전자가 교정된 콩 식물세포로부터 콩 식물체를 재분화하는 단계를 포함하는 것일 수 있다.

[0044] 본 발명은 또한, 상기 콩 유전자교정 방법으로 형질전환된, 콩에서 피틴산 생합성을 조절하는 효소 *GmIPK1* 유전자가 교정된 콩 형질전환체를 제공한다.

[0046] 또한, 본 발명은 상기 콩 유전자교정 방법으로 형질전환된 콩 형질전환체의 종자로써, 피틴산 생합성을 조절하는 효소 *GmIPK1* 유전자가 교정된 콩 종자를 제공한다.

발명의 효과

[0048] 본 발명에 따른 CRISPR/Cas 기반의 콩 유전자교정 시스템 또는 벡터, 이를 포함하는 콩 유전자교정용 조성물은 콩에 존재하는 IPK1 단백질의 발현을 억제하는 효과를 가진다. 또한, 본 발명에 따른 콩 유전자교정 시스템을 이용하여 제작된 콩 형질전환체 및 이의 종자는 피틴산 생합성을 조절하는 *GmIPK1* 유전자가 교정되어 피틴산 함량이 감소된 콩을 생산하는 장점을 가지고 있다.

[0049] 그러므로 본 발명은 콩 식물체에서 표적 유전자를 교정함으로써 항영양인자가 감소한 새로운 식물체를 제공할 수 있으며, 콩 종자에서 피틴산 함량이 감소된 콩을 육성함으로써 식용 및 사료용으로써의 콩의 이용성 및 가치를 증진시키고, 동시에 깨끗한 농업생태계를 보존하는데 기여할 수 있다.

도면의 간단한 설명

[0051] 도 1은 콩의 *GmIPK1* 유전자를 특이적으로 표적할 수 있는 본 발명에 따른 gRNA1 및 gRNA2의 표적 유전자 내 표적 부위를 나타내는 도면이다. (A)는 최종 선발된 표적(target) 부위 및 이에 각각 결합하는 gRNA1 및 gRNA2의 염색체 상 위치를 나타낸다. (B)는 *GmIPK1* 유전자의 표적(target) 부위 중에서 엑손(Exon) 2 내의 표적 서열(좌) 및 엑손(Exon) 3 내의 표적 서열(우)의 염기서열; 및 이에 각각 결합하는 gRNA1 및 gRNA2를 암호하는 염기서열;을 나타낸다.

도 2는 pMDC123_35S$_{pro}$ 벡터의 예시적 제작 모식도를 나타낸다.

도 3은 본 발명에 따른 선별된 gRNA1 및 gRNA2와 코돈-최적화된 Cas9 유전자를 포함하는 예시적인 벡터의 모식도를 나타낸다.

도 4는 아그로박테리움을 이용한 콩 형질전환체 제조 및 유전자교정이 일어난 형질전환체를 검증한 결과이다. (A)는 콩 형질전환체 제조 과정을 보여주고 있고, (B)는 상기 콩 형질전환체를 본 발명에서 이용한 벡터의 선별 마커 및 gRNA 발현 카세트에 대한 PCR 증폭 반응으로 형질전환을 확인한 결과를 보여준다.

도 5는 본 발명에 따른 콩 형질전환체에서 *GmIPK1* 유전자가 교정되었는지를 확인한 결과를 나타낸다. (A)는 InDel PCR을 이용하여 상기 콩 형질전환체에서 *GmIPK1* 유전자가 결실되었는지 확인한 것이고, (B)는 *GmIPK1* 유전자가 결실이 확인된 본 발명에 따른 콩 형질전환체를 보여주고 있다.

도 6은 본 발명에 따른 콩 형질전환체에서 *GmIPK1* 유전자 및 단백질의 표적 서열 부위 결실을 나타내는 결과이다. (A)는 *GmIPK1* 유전자의 표적 서열 부위 577 bp 결실(deletion)을 나타내고, (B)는 IPK1 단백질의 표적 서열 부위 95개 아미노산 결실을 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0052] 본 발명은 CRISPR/Cas9(Clustered Regularly-Interspaced Short Palindromic Repeats/CRISPR associated protein 9) 시스템을 이용하여, 콩에서 항영양인자로 알려진 피틴산 생합성을 조절하는 효소 *GmIPK1* 유전자를 특이적으로 인델(Indel) 돌연변이를 일으키고, 그 결과 콩의 IPK1 발현이 결실된 콩 형질전환체 및 이의 종자를 제조하였다. 상기 콩 형질전환체 및 이의 종자는 targeted deep sequencing을 통해서 *GmIPK1* 유전자가 결실되는 유전자교정이 야기된 것을 확인함으로써, 본 발명을 완성하였다.

[0054] 이하, 본 발명을 용어의 설명, 특정한 예시 및 실시예를 통해서 상세하게 설명한다. 상기 특정한 예시 및 실시예는 발명의 일부 구현예를 포함하는 것으로 모든 구현예를 포함하고 있지는 않다는 점에 유의해야 한다. 본 명세서에 의해 개시되는 발명의 내용은 여기에서 설명되는 특정 실시예로 제한되지 않고, 본 발명이 속한 기술 분야에 있어 통상의 기술자가 다양하게 구현할 수 있다는 것을 포함하는 것은 자명하다. 따라서 본 명세서에 개시된 발명의 내용은 여기에 기재된 특정 실시예로 제한되지 않으며, 이에 대한 변형 및 다른 구현예들도 청구범위 내에 포함되는 것으로 이해되어야 한다.

[0056] 발명에서 사용되는 용어에 대한 정의는 이하와 같다.

[0058] **[용어의 설명]**

[0059] CRISPR/Cas 시스템

[0060] 본 발명에서 사용되는 용어 “CRISPR/Cas 시스템”은 Cas 엔도뉴클레아제(Cas endonuclease) 및 상기 핵산분해 효소에 대응하는 가이드 분자가 포함된 복합체로써, 표적 유전자 또는 표적 핵산에 결합할 수 있는 복합체를 의

미한다. 여기서 가이드 분자는 가이드 RNA(gRNA)로 대표되는 표적화 핵산일 수 있으나, 이에 제한되지는 않는다.

[0061] 상기 CRISPR/Cas 시스템은 Cas 엔도뉴클레아제(Cas endonuclease) 및 상기 가이드 RNA(gRNA)에 따라 다양한 종류가 가능하며, Cas 엔도뉴클레아제(Cas endonuclease)는 상기 대응되는 가이드 분자를 이용하여 표적 핵산 또는 표적 유전자에 특이적으로 결합하여 DNA 이중 가닥 절단 등의 유전자교정(gene editing)을 수행한다.

[0062] CRISPR/Cas 시스템을 이용하는 유전자교정(gene editing) 기술은 인간이나 동식물 등 다양한 개체의 유전자에 변이를 도입할 수 있고, 이를 통해 유전자의 기능을 밝히는 연구, 새로운 형질을 가지는 형질전환체의 제조 및 유전자 관련 질병의 치료 방법으로 사용되고 있다. 상기 CRISPR/Cas 시스템은 특정 서열을 표적화하기 위해 맞춤형 단백질의 생성을 필요로 하지 않고, 가이드 분자 내의 표적 부위 결합 서열인 스페이서 서열의 변경을 통해서 다양한 레퍼토리의 CRISPR/Cas 시스템을 제작할 수 있다.

[0064] Cas 엔도뉴클레아제(Cas endonuclease)

[0065] 본 발명에서 사용되는 용어 “Cas 엔도뉴클레아제(Cas endonuclease)"는 표적하는 핵산인 DNA 또는 RNA, 또는 표적 유전자 내의 프로토스페이서 인접 모티프(protospacer adjacent motif, PAM)를 인식한 후, 표적 핵산 서열의 내부 또는 외부 염기서열(base pair, bp)에서 DNA 이중가닥 절단(double strand breaks, DSB)이 일어나게 편집할 수 있는 핵산분해 효소를 의미한다. 본 명세서에서 Cas 엔도뉴클레아제(endonuclease)는 CRISPR/Cas 시스템을 구성하는 효과기(effector) 단백질을 지칭한다. 여기서 효과기(effector) 단백질은 CRISPR 단백질, 가이드 RNA(gRNA)에 결합할 수 있는 핵산분해 단백질, 또는 표적 핵산 또는 표적 유전자에 결합할 수 있는 올리고핵산에 결합 가능한 펩티드 단편일 수 있다. 구체적으로 Cas9 또는 Cpf1, 또는 변형된 핵산분해 단백질일 수 있지만, 이들로 제한되지 않는다. 또한, CRISPR/Cas 기술 분야에서 통상의 기술자가 인식할 수 있는 의미를 모두 포함한다.

[0067] Cas9 단백질

[0068] 본 발명에 따른 “Cas9 단백질"은 표적 핵산 또는 표적 유전자 내의 표적 부위 서열 근처의 프로토스페이서 인접 모티프(protospacer adjacent motif, PAM)를 인식한 후, 표적 핵산 서열의 내부 또는 외부 염기서열(base pair, bp)에서 DNA 이중가닥 절단(double strand breaks, DSB)하여 표적 유전자에서 인델(insertion and/or deletion, Indel) 변이(mutation)을 야기할 수 있는 모든 Cas9들 중에서 선택된 1종 이상일 수 있다. 구체적으로, 상기 Cas9 단백질은 스트렙토코커스(Streptococcus) 속, 캄필로박터(Campylobacter) 속, 네이세리아(Neisseria) 속, 파스테우렐라(Pasteurella) 속, 프란시셀라(Francisella) 속 유래의 Cas9 단백질 등일 수 있고, 더욱 구체적으로, 스트렙토코커스 써모필러스(Streptococcus thermophiles), 스트렙토코커스 아우레우스(Streptocuccus aureus) 또는 스트렙토코커스 피요게네스(Streptococcus pyogenes) 유래의 Cas9 단백질; 캄필로박터 제주니(Campylobacter jejuni) 유래의 Cas9 단백질; 네이세리아 메닝기디티스(Neisseria meningitidis) 유래의 Cas9 단백질; 파스테우렐라 물토시다 (Pasteurella multocida) 유래의 Cas9 단백질; 프란시셀라 노비시다(Francisella novicida) 유래의 Cas9 단백질 등일 수 있으나, 이에 제한되는 것은 아니다. 또한, CRISPR/Cas 기술 분야에서 통상의 기술자가 인식할 수 있는 의미를 모두 포함한다. 상기 Cas9 단백질 또는 유전자 정보는 NCBI(National Center for Biotechnology Information)의 GenBank와 같은 공지의 데이터 베이스에서 얻을 수 있다.

[0070] 가이드 RNA(guide RNA)

[0071] 본 발명에 따른 용어 "가이드 RNA(guide RNA)"는 Cas 엔도뉴클레아제(Cas endonuclease)와 복합체를 형성할 수 있고, 표적 핵산 서열과 혼성화할 수 있으며, 표적 핵산 서열에 대한 복합체의 서열-특이적 결합을 하기에 충분한 표적 핵산 서열과의 상보성을 갖는 가이드 서열을 포함하는 RNA를 의미한다. 본 명세서에서 가이드 분자 또는 가이드 RNA는 상호호환 가능하게 사용된다.

[0072] 본 명세서에 가이드 RNA(guide RNA)는 표적 유전자 서열을 인식하는 CRISPR RNA(crRNA)와 crRNA와 결합하면서 상기 Cas 엔도뉴클리아제(Cas endonuclease)에 결합하는 transactivating CRISPR RNA(tracrRNA)를 포함한다. 가이드 RNA(guide RNA)는 crRNA와 tracrRNA가 링커로 연결된 단일 가닥 형태일 수 있고, crRNA와 tracrRNA가

상보적 결합하여 이중가닥 형태를 형성할 수도 있다. 본 명세서에 가이드 RNA는 하나 이상의 화학적 변형 예를 들어, 2개의 리보뉴클레오타이드의 화학적 연결에 의해 또는 하나 이상의 리보뉴클레오타이드의 하나 이상의 데옥시리보뉴클레오타이드로의 대체에 의한 변형을 갖는 RNA-기반 분자일 수 있다.

[0073] Cas 엔도뉴클레아제(Cas endonuclease)에 따른 가이드 RNA(guide RNA)의 특이성은 직접 반복 서열인 스케폴드(Scafold)에에 의해 좌우되므로, Cas 엔도뉴클레아제(Cas endonuclease)에 최적의 가이드 RNA(guide RNA)를 설계하고자 할 때는 상기 직접 반복 서열을 Cas의 유래에 따라 설계할 수 있다. 본 발명에 따른 가이드 RNA(guide RNA)는 Cas 엔도뉴클레아제(Cas endonuclease)에 특이적으로 최대의 활성을 나타내도록 하기 위해 자연에서 발견되는 가이드 RNA(guide RNA) 또는 그 길이나 서열의 변형을 가지는 것일 수 있다. 또한, CRISPR/Cas 기술 분야에서 통상의 기술자가 인식할 수 있는 의미를 모두 포함한다.

[0075] 스캐폴드(Scafold) 영역

[0076] 본 발명에 따른 용어 “스캐폴드(Scafold) 영역”은 가이드 RNA(guide RNA) 중 핵산분해 단백질과 상호작용할 수 있는 부분을 통틀어 지칭하며, 자연계에서 발견되는 가이드 RNA의 부분 중 스페이서(Spacer)를 제외한 나머지 부분을 지칭할 수 있다. 상기 스캐폴드(Scafold) 영역은 tracrRNA 및 crRNA의 일부를 포함하며, 반드시 한 분자의 RNA를 지칭하는 것은 아니다.

[0078] 스페이서 서열(spacer sequence)

[0079] 본 발명에 따른 용어 "스페이서 서열(spacer sequence)"은 본 발명에 따른 CRISPR/Cas 시스템에서 표적 서열 부분과 혼성화되는 폴리뉴클레오타이드를 의미한다. 상기 스페이서 서열(spacer sequence)은 본 발명에 따른 유전자가위 시스템에서 가이드 RNA(guide RNA)의 crRNA의 3'-말단 부근의 약 20개의 연속된 뉴클레오타이드를 지칭하며, 본 발명에서 "gRNA"라고도 표현된다.

[0080] 본 발명에 따른 "gRNA" 즉, 상기 스페이서 서열은 상기 유전자가위 시스템을 사용하여 유전자 교정하고자 하는 표적 핵산 또는 표적 유전자의 표적 서열에 대응하여 설계된다. 즉, 상기 스페이서는 표적 핵산의 표적 서열에 따라 다양한 서열을 가질 수 있다. 본 명세서에서 스페이서 서열은 상기 유전자가위 시스템에 포함된 crRNA 내의 스페이서 서열을 의미하고, 가이드 RNA(guide RNA)에서 표적 서열과 결합하는 부위로써 표적 부위를 표적화하는 중요한 부분으로써, 달리 명시되지 않는 한, 가이드 RNA(guide RNA)와 구분되는 개념으로 "gRNA"라고 표현된다. 또한, CRISPR/Cas 기술 분야에서 통상의 기술자가 인식할 수 있는 의미를 모두 포함한다.

[0082] CRISPR RNA(crRNA)

[0083] 본 발명에 따른 용어 "CRISPR RNA(crRNA)"은 본 발명에 따른 CRISPR/Cas 시스템 내의 가이드 RNA(guide RNA)를 구성하는 중요한 구성요소이다. 본 명세서에서 상기 crRNA는 표적 유전자 서열을 인식하는 스페이서(Spacer) 서열을 포함하는 것으로, 표적 유전자 또는 표적 핵산에 결합하는 가이드 RNA(guide RNA)의 부분이다.

[0085] transactivating CRISPR RNA(tracrRNA)

[0086] 본 발명에 따른 "tracrRNA"는 crRNA와 함께 CRISPR/Cas 시스템의 가이드 RNA(guide RNA)를 구성하는 중요한 구성요소이다. 본 명세서에서 상기 tracrRNA의 적어도 일부는 crRNA의 적어도 일부와 상보적으로 결합하여 이중가닥을 형성할 수 있고, Cas 엔도뉴클리아제에 결합할 수 있는 가이드 RNA 부분이다. 더 구체적으로, 상기 tracrRNA의 일부 서열은 상기 crRNA에 포함된 CRISPR RNA 반복 서열의 전부 또는 일부와 상보적 서열을 가진다.

[0088] 싱글 가이드 RNA(sgRNA)

[0089] 본 발명에 따른 "싱글 가이드 RNA(sgRNA)"는 가이드 RNA(guide RNA) 형성이 쉽도록 상기 tracrRNA 및 상기 crRNA를 링커로 연결한 한 가닥의 RNA 분자이다. 본 명세서에서 제공되는 가이드 RNA(guide RNA)는 싱글 가이드 RNA일 수 있고, 이 경우 상기 싱글 가이드 RNA는 상기 tracrRNA 및 스페이서 서열을 포함하는 crRNA를 모두 포

함하는 한 분자의 RNA이다.

[0091] 표적 유전자(Target gene)

[0092] 본 발명에서 사용되는 용어 "표적 유전자(target gene)"는 본 발명에 따른 CRISPR/Cas 시스템에 의한 유전자 편집의 대상이 되는 세포 내 유전자 또는 핵산을 의미한다. 표적 핵산 또는 표적 유전자는 혼용될 수 있으며, 서로 동일한 대상을 지칭할 수 있다. 상기 표적 유전자는 달리 기재되지 않은 한, 대상 세포가 가진 고유한 유전자 또는 핵산 혹은 외부 유래의 유전자 또는 핵산, 또는 인위적으로 합성된 핵산 또는 유전자일 수 있고, 단일가닥 DNA, 이중가닥 DNA 및/또는 RNA 모두를 의미할 수 있다. 상기 표적 유전자 또는 표적 핵산은 CRISPR/Cas 시스템에 의한 유전자 편집의 대상이 될 수 있다면 특별히 제한되지 않는다.

[0094] 표적 부위(Target region) 또는 표적 서열(Target sequence)

[0095] 본 발명에 따른 "표적 부위(Target region)" 또는 "표적 서열(Target sequence)"은 표적 핵산 또는 표적 유전자 내에 존재하는 서열로, 본 발명에 따른 CRISPR/Cas 시스템이 표적 유전자 또는 표적 핵산을 절단하기 위해 인식하는 특정 서열을 의미한다. 상기 표적 부위 또는 표적 서열은 그 목적에 따라 적절히 선택될 수 있다. 구체적으로, 표적 부위 또는 표적 서열은 본 발명의 단일가닥 가이드 RNA(sgRNA)에 포함된 스페이서 서열과 상보성을 가지는 서열을 의미한다. 상기 스페이서 서열은 표적 유전자 또는 표적 핵산의 서열 및 CRISPR/Cas 시스템의 효과기(effector) 단백질이 인식하는 PAM 서열을 고려하여 결정된다. 상기 표적 부위 또는 표적 서열은 유전자가위 시스템의 가이드 RNA와 상보적으로 결합하는 특정 가닥만을 지칭할 수 있으며, 상기 특정 가닥 부분을 포함하는 표적 이중 가닥 전체를 지칭할 수도 있으며, 이는 문맥에 따라 적절히 해석된다. 또한, 상기 용어는 당업계 통상의 기술자가 인식할 수 있는 의미를 모두 포함하며, 문맥에 따라 적절히 해석될 수 있다.

[0097] 상동체(homolog)

[0098] 본 발명에서 사용되는 용어 “상동체(homolog)”는 어떠한 형질이 진화의 과정 동안 보존된 것을 말하며, 이는 형태적 형질, 분자적 형질 또는 유전자의 염기서열에서의 보존을 의미할 수 있다. 진화가 일종의 개조 과정이라면, 그 과정을 거치면서 보존되는 특징들이 존재할 것이라 예측할 수 있는 것을 의미한다. 여기서 상동성이 보존되었다는 것은 정확히 일치한다는 것이 아니라, 어떠한 공통적 조상 형질로부터 유래된 것이 소실되지 않고 남아 있어 진화적 관계를 유추할 수 있게 하는 것을 의미한다.

[0099] 또한, 유전자의 염기서열 또는 단백질의 아미노산 서열을 기준으로 상동체는 진화적으로 가까운 종일수록 보존된 서열의 비율이 높고, 생물의 기본적 기능에 필수적인 요소일수록 보존되어 있을 가능성이 높으므로, 본 발명에서 의미하는 상동체는 기능적 등가물을 포함한다. 기능적 등가물은 기준 서열로부터 하나 또는 그 이상의 치환, 결실 또는 부가 서열을 가지는 것으로 다양한 기능적 비유사성을 가지지 않는 실질적으로 동일하거나 유사한 기능을 가지는 변화된 돌연변이 서열의 뉴클레오티드와 핵산서열 모두를 의미한다.

[0101] 형질전환(transformation)

[0102] 본 발명에서 사용되는 용어 “형질전환(transformation)”은 세포의 유전적 특성 변화, 즉 유전형질을 변화시키는 현상을 의미한다. 세포가 외부의 새로운 DNA 또는 RNA 등의 물질이 도입됨으로써 천연 상태로부터 유전적으로 변형되는 것을 의미하며, 형질변환, 형전환 또는 형변환이라고도 한다. 형질감염 또는 형질도입후, 외부의 새로운 DNA 또는 RNA 등은 세포의 염색체 내로 생리학적으로 통합됨으로써 세포의 것과 재결합될 수 있거나, 복제과정 없이 에피솜 요소로서 일시적으로 유지될 수 있거나, 또는 플라스미드로서 독립적으로 복제될 수 있다.

[0104] 벡터(Vector)

[0105] 본 발명에 따른 "벡터"는 달리 특정되지 않는 한, 유전 물질을 세포 내로 운반할 수 있는 모든 물질을 통틀어 일컫는다. 예를 들어, 벡터는 대상이 되는 유전 물질, 예를 들어 CRISPR/Cas 시스템의 이펙터 단백질을 암호화하는 핵산 및/또는 가이드 RNA를 암호화하는 핵산을 포함하는 DNA 분자일 수 있으나, 이에 제한되는 것은 아니

다. 상기 용어는 "벡터" 이 기술분야에서 통상의 기술자가 인식할 수 있는 의미를 모두 포함하며, 문맥에 따라 적절히 해석될 수 있다.

[0106] 또한, 본 발명에서 "벡터"는 삽입된 유전자가 정상적으로 발현되도록 작동가능하게 연결된 필수적인 조절 요소를 포함하는 "발현벡터" 일 수 있고, 플라스미드 벡터, 코스미드 벡터, 박테리오파지 벡터, 및 바이러스 벡터 등을 포함할 수 있다. 구체적으로, Ti 플라스미드 pGA3383, pCAMBIA3301, pCAMBIA3300, pGA3426, pGA3780, pGWB12, pGWB14, 레트로바이러스, 아데노바이러스 또는 백시니아바이러스와 같은 동물 바이러스, 배큘로바이러스와 같은 곤충 바이러스 또는 식물 바이러스가 사용될 수 있으며, pPZP, pGA, pCAMBIA, pMDC123, pECO100, pECO200 및 pECO300 계열과 같은 바이너리 벡터를 사용할 수 있다.

[0107] 또한, 상기 벡터의 목적이 CRISPR/Cas 시스템 각 구성요소를 세포 내에서 발현되도록 하는 것이므로, 상기 벡터의 서열은 CRISPR/Cas 시스템의 각 구성요소를 암호화하는 핵산 서열 중 하나 이상을 필수적으로 포함해야 한다. 구체적으로, 상기 벡터의 서열은 발현하고자 하는 CRISPR/Cas 시스템에 포함된 가이드 RNA, 및/또는 Cas 단백질을 암호화하는 핵산 서열을 포함한다. 이때, 상기 벡터의 서열은 야생형의 가이드 RNA 및 야생형의 Cas 단백질을 암호화하는 핵산 서열 뿐 아니라, 코돈 최적화된 Cas 단백질을 암호화하는 핵산 서열을 포함할 수 있다.

[0108] 상기 벡터는 또한 플라스미드, 선형의 PCR 엠플리콘 또는 바이러스 벡터일 수 있다. 여기서, 상기 바이러스 벡터는 레트로바이러스 벡터(retrovirus vector), 렌티바이러스 벡터(lentivirus vector), 아데노바이러스 벡터(adenovirus vector), 아데노-연관 바이러스 벡터(adeno-associated virus(AAV) vector), 백시니아바이러스 벡터(vaccinia virus vector), 폭스바이러스 벡터(poxvirus vector) 및 단순포진 바이러스 벡터(herpes simplex virus vector)로 구성된 군에서 선택되는 적어도 하나의 바이러스 벡터일 수 있다.

[0109] 또한 상기 벡터는 발현 벡터일 수 있고, 선형, 또는 원형 벡터 형태로 설계될 수 있다. 상기 Cas9 유전자가 발현될 수 있도록, 발현조절 서열과 기능적으로 연결되어 있다. 예를 들어, 벡터는 프로모터, 오퍼레이터, 개시코돈, 종결코돈, 폴리아데닐화 시그널, 인핸서 같은 발현 조절 요소 외에도 막 표적화 또는 분비를 위한 신호서열 또는 리더 서열을 포함하며 목적에 따라 다양하게 제조될 수 있다. 또한, 벡터는 선택성 마커를 포함할 수 있으며, 벡터는 자가 복제하거나 숙주 DNA에 통합될 수 있다.

[0110] 본 발명의 벡터는 당해 기술 분야에서 잘 알려진 유전자 재조합 기술을 이용하여 제조할 수 있으며, 부위-특이적 DNA 절단 및 연결은 당해 기술 분야에서 일반적으로 알려진 효소 등을 사용한다.

[0112] 프로모터

[0113] 본 발명에서 사용되는 용어 "프로모터"는 벡터의 발현 대상을 세포 내에서 발현시키기 위한 각 구성 요소를 암호화하는 서열에 작동적으로 연결되어 세포 내에서 RNA 전사 인자가 활성화될 수 있도록 하는 것을 의미한다. 상기 프로모터 서열은 대응하는 RNA 전사 인자, 또는 발현 환경에 따라 달리 설계할 수 있으며, CRISPR/Cas 시스템의 구성 요소를 세포 내에서 적절히 발현시킬 수 있는 것이라면 제한되지 않는다.

[0114] 본 발명의 벡터가 식물 세포에 적용되는바, 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), 컬리플라워 모자이크 바이러스(CaMV) 35S 프로모터, 노팔린 씬타아제(nos) 프로모터, 피그워트 모자이크 바이러스 35S 프로모터, 수가크레인 바실리폼 바이러스 프로모터, 콤멜리나 엘로우 모틀 바이러스 프로모터, 리불로오스-1,5-비스-포스페이트 카르복실라아제 스몰 서브유티트(ssRUBISCO)의 광유도성 프로모터, 벼 사이토졸 트리오스포스페이트 이소머라아제(TPI) 프로모터, 아라비돕시스의 아데닌 포스포리보실트랜스퍼라아제(APRT) 프로모터 및 옥토파인 신타아제 프로모터를 포함할 수 있다. 본 발명의 프로모터는 구체적으로 상시발현 프로모터이며, 더욱 구체적으로 35S 프로모터인 것을 특징으로 할 수 있다.

[0115] 또한, 상기 벡터는 상기 엔지니어링 된 가이드 RNA를 암호화하는 핵산을 위한 프로모터를 더 포함할 수 있다. 구체적으로, 상기 프로모터는 AtU6 프로모터, H1 프로모터 또는 7SK 프로모터일 수 있다.

[0117] 선별 마커

[0118] 본 발명의 용어 "선별 마커"는 화학적인 방법으로 선택될 수 있는 특성을 갖는 핵산 서열을 의미하며, 형질전

환된 세포를 비형질전환 세포로부터 구별할 수 있는 모든 유전자가 이에 해당된다. 상기 선택성 마커는 본 발명에 따른 벡터에 포함될 수 있다. 구체적으로 상기 선택성 마커는 글리포세이트(glyphosate), 글루포시네이트암모늄(glufosinate ammonium) 또는 포스피노트리신(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. 글리포세이트(glyphosate), 글루포시네이트암모늄(glufosinate ammonium) 또는 포스피노트리신(phosphinothricin)과 같은 제초제 저 항성 유전자, 암피실린(ampicillin), 카나마이신(kanamycin), G418, 블레오마이신 (Bleomycin), 하이그로마이신(hygromycin), 클로람페니콜(chloramphenicol)과 같은 항생제 내성 유전자일 수 있으나, 이에 제한되는 것은 아니다.

[0120] 작동 가능하게 연결된(operably linked)

[0121] 본 발명에서 사용되는 "작동 가능하게 연결된"이라는 용어는 유전자 발현 기술에 있어서, 특정 구성이 다른 구성과 연결되어, 상기 특정 구성이 의도된 방식대로 기능할 수 있도록 연결되어 있는 것을 의미한다. 예를 들어, 프로모터 서열이 암호화 서열과 작동적으로 연결되었다고 할 때, 상기 프로모터가 상기 암호화 서열의 세포 내에서의 전사 및/또는 발현에 영향을 미칠 수 있도록 연결된 것을 의미한다. 또한, 상기 용어는 이 기술분야의 통상의 기술자가 인식할 수 있는 의미를 모두 포함하며, 문맥에 따라 적절히 해석될 수 있다.

[0123] 핵 위치 신호 서열(Nuclear Localization Signal, NLS)

[0124] 본 발명에 따른 "핵 위치 신호(NLS)"라 함은, 핵 수송(nuclear transport) 작용으로 세포 핵 외부의 물질을 핵 내부로 수송할 때, 수송 대상인 단백질에 붙어 일종의 "태그" 역할을 하는 일정 길이의 펩티드 또는 그 서열을 의미한다. 본 발명에 따른 CRISPR/Cas 시스템을 식물 세포에서의 적용을 위해 Cas 엔도뉴클레아제에 NLS가 연결될 수 있다. 여기서 NLS는 SV40 바이러스 대형 T-항원의 NLS, 뉴클레오플라스민(nucleoplasmin)으로부터의 NLS, c-myc NLS일 수 있다. 또한 hRNPA1 M9 NLS 서열; 임포틴-알파로부터의 IBB 도메인의 NLS 서열, 마이오 마(myoma) T 단백질의 NLS 서열 및 인간 p53의 NLS 서열, 마우스 c-abl IV의 NLS 서열; 인플루엔자 바이러스 NS1의 NLS 서열, 간염 바이러스 델타 항원의 NLS 서열, 마우스 Mx1 단백질의 NLS 서열, 인간 폴리(ADP-리보스) 중합효소의 NLS 서열 또는 스테로이드 호르몬 수용체(인간) 글루코코르티코이드의 NLS 서열로부터 유래된 NLS 서열일 수 있으나, 이에 제한되지 않는다.

[0125] 일 예로, 본 발명에 따른 벡터에 포함되는 사용되는 NLS 서열은 서열번호 8 내지 10 중 어느 하나 이상의 SV40 바이러스 대형 T-항원의 NLS 서열 및/또는 서열번호 11 내지 12 중 어느 하나 이상의 뉴클레오플라스민(nucleoplasmin)으로부터의 NLS 서열일 수 있다.

[0127] 약

[0128] 본 발명에 따른 “약”이라는 용어는 참조 양, 수준, 값, 수, 빈도, 퍼센트, 치수, 크기, 중량 또는 길이에 대해 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 또는 1% 정도로 변하는 양, 수준, 값, 수, 빈도, 퍼센트, 치수, 크기, 중량 또는 길이를 의미한다.

[0130] 본 발명에서 사용되는 모든 기술용어는, 달리 정의되지 않는 이상, 본 발명의 관련 분야에서 통상의 당업자가 일반적으로 이해하는 바와 같은 의미로 사용된다. 또한 본 명세서에는 바람직한 방법이나 시료가 기재되나, 이와 유사하거나 동등한 것들도 본 발명의 범주에 포함된다.

[0132] 이하, 본 발명을 상세히 설명한다.

[0134] **[재료 및 방법]**

[0135] **1. 표적 유전자의 유전자교정을 위한 고효율 gRNA 선발**

[0136] 본 발명에서 사용할 모든 gRNA들은 유전자교정 Web Tool을 이용하여 후보 gRNA를 선발하고, 이를 콩의 형질전환에 이용하였다.

[0137] 먼저, 유전자교정을 실시할 표적 유전자를 선정한다. 그 후, 콩 유전자편집을 위한 gRNA를 디자인하기 위하여 CRISPR-P2.0 (http://crispr.hzau.edu.cn/ CRISPR2/), RGEN(http://www.rgenome.net/), Phytozome (https://phytozome.jgi. doe.gov/pz/portal.html) 등의 web site와 Phytozome 내에 있는 JBrowse tool을 이용하여 비표적유전자를 교정하는 '오프-타겟 효과(off-target effect)'가 적고 유전자교정 효율이 높을 것으로 예측되는 gRNA들을 선발한다.

[0139] CRISPR-P는 목적하는 염기서열로 가이드 RNA(guide RNA)를 검색할 수 있는 web site로서 후보가 되는 gRNA들이 효율에 따라 랭킹되어있고, off-target의 종류와 genome상에서 exon, intron, intergenic, promoter 등의 off-target의 위치도 확인할 수 있다. 하지만 CRISPR-P2.0 에서는 off target을 제한적으로 보여주는 단점이 있다.

[0140] 또한, RGEN은 목표로 하는 유전자의 염기서열 중 CRISPR/Cas9시스템이 인식하는 PAM서열을 선정한 후 PAM으로부터 완전한 매치(perfect match) 또는 1bp 미스매치(mismatch) 또는 2bp 미스매치(mismatch)되는 20bp의 gRNA를 제시해준다. 그리고 RGEN에서는 out-of-frame score가 제공되어 틀 이동 변이(frame shift mutation)가 일어날 가능성을 예측해주므로 스코어(score)가 높은 것을 선정하면 이후에 염기에 변이가 일어날 가능성이 높다. 다만, RGEN은 많은 수의 off-target을 제공해 주지만 각 off-target의 genome상의 위치만 제공해 주기때문에 off-target의 상세 정보는 직접 확인하는 과정이 반드시 필요하다.

[0141] 따라서 본 발명에서는 콩에서 표적 유전자에 특이적으로 결합하는 고효율의 gRNA를 디자인하기 위하여, CRISPR-P2.0 와 RGEN tool을 함께 사용하여 상기 각 tool의 단점을 보완함으로써 콩 유전자교정을 위한 고효율의 gRNA를 디자인할 수 있다. 또한, 하나의 표적 유전자 내에서 2 개의 표적 부위를 특이적으로 인식하는 2 개의 gRNA 조합를 선별함으로써 향후 세대를 진전할 때 NGS없이 InDel PCR만으로 유전자 편집된 라인을 선택하는데 유용하게 하였다.

[0143] 본 발명은 구체적으로, 목적하는 식물의 유전자 정보를 'phytozome' 에서 검색하고, 목적 유전자의 genomic DNA 염기서열을 획득하였다. gRNA의 표적 특이성 및 효율성을 고려하여 1000 bp 이내로 genomic DNA 염기서열을 취득 한 후 'CRISPR-P' 에 입력하여 해당 염기서열 내의 gRNA를 검색하였다.

[0144] 'CRISPR-P' 에서 검색 된 gRNA 목록 중 On-target score가 높으면서, Thymine (T) 염기가 4개 이상 반복되지 않고, 해당 gRNA의 off-target 이 발생하는 위치가 타 유전자의 엑손(exon)에 해당하는 경우가 가장 적은 것을 후보군으로 선발하였다. 후보군 gRNA들을 선발한 후, 표적 유전자에 적용할 2개 gRNA 거리는, 형질전환체 획득 후 targeted deep sequencing으로 indel을 확인할 때 sequencing의 기술적인 한계로 최대 300bp까지만 가능하므로 가능한 두 target 사이의 거리를 맞추기 위해서 약 250 bp 내에 위치하는 쌍을 선택하였다. 상기 조건을 만족하는 gRNA 염기서열을 'RGEN' 의 Cas-OFFinder 를 이용해 한번 더 해당 gRNA로 인해 발생할 수 있는 off-target을 검색하였다.

[0145] 'CRISPR-P'는 On-score 정보로 인해 효율이 높은 gRNA를 선발하기 용이하나, off-target 정보는 제한적으로 보여주므로(최대 20개), 'RGEN'에서 한번 더 교차 검증하여 콩 유전자교정을 위한 고효율의 gRNA을 선별하였다. 'RGEN'에서 검색된 genome 상의 위치한 off-target 정보를 이용하여 'JBrowse' 에서 선발한 gRNA의 off-target 이 발생되는 부분이 어딘지를 검색하였다.

[0146] 'JBrowse'는 n번 chromosome의 전체 genome 정보를 보여주므로 off-target이 발생하는 부위가 exon 인지 그 외 다른 부분인지를 파악할 수 있으며, 이를 통해 gRNA의 적절성 여부를 확인할 수 있다. 상기 과정을 반복하여 On-target score가 높고, Off-target이 가장 적은 gRNA를 최종 선발하고, 이를 검증하였다.

[0147]

[0148] **2. 선발된 고효율 gRNA을 포함하는 CRISPR/Cas9 벡터의 제조**

[0149] 목적에 따라 표적 유전자를 표적으로 하는 gRNA들을 최종 선발하여 CRISPR/Cas9 벡터를 제작할 수 있다. 본 발명에서는 콩 유전자편집을 위하여 pMDC123_pcoCas9_g1g2(plant codon-optimized Cas9), pMDC123_acoCas9_g1g2(Arabidopsis codon-optimized Cas9), pMDC123_pcohCas9_g1g2(plant codon- optimized

Cas9 with high GC content)와 카이스트 김상규 교수님으로부터 분양받은 pECO201 vector를 이용하였다. 상기 벡터들은 형질전환된 식물체를 선발하기 위하여 선발 마커로 *Bar* 유전자를 가지며, 재조합을 위하여 attR1과 attR2 서열을 가지는 gateway용 벡터이다.

[0150] pECO201 벡터는 pGRNA 벡터와 식물 바이너리 벡터를 골든 게이트 조립(Golden gate assembly)하여 변형시킨 벡터이고(Oh Y. et al., Plant Methods, Vol. 16, No.37(2020)), Arabidopsis codon optimized SpCas9 및 35S promoter을 포함하고 있다. pMDC123_Cas9_g1g2는 bar유전자가 enhanced 35S CaMV 프로모터에 의해서 제어되는 반면, pECO201은 *Agrobacterium*에서 유래한 nos 프로모터 (nopalin synthase promoter)에 의해서 제어된다.

[0152] pMDC123 벡터는 목적하는 유전자의 발현을 제어하는 프로모터가 없기때문에, pMDC32벡터로부터 2X 35S CaMV 프로모터를 절단하여 pMDC123 벡터의 *HindIII/Asc*I site에 연결하여 pMDC123_35S_{pro} 벡터를 제작하였다. 제작된 pMDC123_35S_{pro} 벡터는 최종 Multisite Gateway recombination에 이용하였다. 도 2는 pMDC123_35S_{pro} 벡터의 제작 모식도를 나타낸다.

[0154] 상기 선별된 gRNA1 및 gRNA2와 상기 코돈-최적화된 Cas9 유전자(codon-optimized SpCas9)를 가지는 벡터를 제작하기 위하여 두 번의 Golden Gate assembly와 한 번의 Multisite Gateway recombination의 과정을 진행하였다. 먼저, gRNA가 AtU6 프로모터에 작동될 수 있도록 연결하는 과정이다. AtU6 프로모터를 가지는 golden gate entry vector인 pYPQ131A와 pYPQ132A vector를 제한효소 *ESP3I(BsmBI)*으로 절단한 후, oligo annealing 과정을 거친 gRNA1과 gRNA2를 각각 pYPQ131A와 pYPQ132A 벡터에 Golden Gate assembly 방법으로 연결하였다. 다음으로, 각각의 AtU6 프로모터에 연결된 gRNA1(pYPQ131A_gRNA1)과 gRNA2(pYPQ132A_gRNA2)를 제한효소 *BsaI*을 이용하여 pYPQ142 벡터(Golden Gate recipient and Gateway entry vector)에 Golden Gate assembly 방법으로 연결하였다. 그 후, 각각의 Cas9 유전자와 nos Terminator를 가지는 Gateway entry vector(pYPQ150(plant codon optimized), pYPQ154(Arabidopsis codon optimized), pYPQ167(plant codon optimized: high GC content at 5' region))와 두 번째 과정에서 만들어진 gRNA1과 gRNA2를 모두 가지는 pYPQ 142 vector 및 식물 발현용 벡터인 pMDC123_35S $_{pro}$를 이용하여 multisite gateway assembly 방법을 이용하여 최종 연결한다. 본 실험에 사용한 pYPQ131A, pYPQ132A, pYPQ142, pYPQ150, pYPQ154, pYPQ167 plasmid DNA는 Addgene으로부터 구입하여 사용할 수 있다.

[0156] 본 발명에서 최종 선별된 gRNA들의 염기서열에 대한 센스(sense) 및 안티센스(antisense) 프라이머를 제작하여 annealing 시킨 후, 식물발현용 벡터 내의 *AtU6* 프로모터(Arabidopsis U6 promter) 뒤에 작동가능하게 클로닝한다. Golden gateway 방법을 이용하여 각 gRNA들의 발현 카세트를 하나의 벡터에 클로닝한다. *Streptococcus pyogenes* 유래의 *SpCas9* 유전자와 gRNA 발현카세트, 35S 프로모터와 ccdB recombination system을 가지는 T-DNA 벡터를 재조합하여 최종적으로 선발마커를 가지는 식물형질전환용 multiplex CRISPR/Cas9 벡터를 제작할 수 있다.

[0157] 이 때, multiplex CRISPR/Cas9 벡터에서 *SpCas9* 유전자는 코돈-최적화된 핵산일 수 있고, 선발마커유전자는 Bar 유전자일 수 있으며, 상기 벡터는 NLS 서열을 한 개 이상 포함할 수 있다. 하나의 표적유전자 당 2개씩의 gRNA를 포함하는 multiplex 벡터를 제조하여 Indel PCR을 이용한 유전자교정(Gene Editing, GE) 형질전환체 선발을 용이하게 하였다.

[0159] **3. 콩 형질전환 및 형질전환체 선발**

[0160] 콩 형질전환에 사용할 상기 식물형질전환용 binary vector를 *Agrobacterium tumefaciens* strain EHA105에 일렉트로포레이션(electroporation) 방식으로 형질전환하였다. 형질전환이 확인된 하나의 콜로니(single colony)는 항생제가 첨가된 LB 액체 배지(LB, Ripampicin 25mg/L, kanamycin 50mg/L 또는 spetinomycin 50mg/L)에서 이틀 동안 28°C에서 배양한 후 glycerol(최종 30%)을 첨가하여 -70°C에 보관해 두었다. 형질전환 이틀 전에 -70°C에 보관해 두었던 cell stock의 일부를 항생제가 첨가된 LB 액체 배지에 접종하여 30시간 동안 28°C에서 배양하였

다.

[0161] 잘 자란 *Agrobacterium* cell 400ul를 항생제가 첨가된 AB 고체 배지(100㎖ 당 glucose 0.5g, agar 1.5g/90ml water, 5㎖ 20x AB buffer(K2HPO4 60g/L, NaH2PO4 20g/L), 5㎖ 20x Stock of AB salts(NH4Cl 20g/L, MgSO4.7H2O 6g/L, KCl 3g/L, CaCl 0.2g/L, FeSO4.7H2O 50mg/L), Ripampicin 25mg/L, kanamycin 50mg/L)에 도말하여 배양하였다. 형질전환 당일 AB고체 배지에 도말하여 잘 자란 *Agrobacterium* cell은 LCM배지(액체 CCM 배지, 1ℓ당 B5 medium 3.2g, Sucrose 30g, Sodium thiosulfate 158mg, DTT 154mg, L-cystein 580mg, MES 4.26g, pH5.4 with 1M KOH)를 이용하여 AB배지의 표면으로부터 긁어 낸다. 이를 일부는 OD_{600}이 0.6(저농도)이 되도록 희석하고 일부는 OD_{600}이 2(고농도)가 되도록 희석한 뒤 Acetosyringon (최종 200uM)을 첨가하여 *Agrobacterium* infection 전까지 28°C에서 배양하였다.

[0163] 또한, 본 발명에서 상기 벡터로 식물체를 형질전환하는 것은 이 기술분야에서 기술자에게 공지된 형질전환기술에 의해 수행될 수 있다. 구체적으로, 아그그로박테리움을 이용한 형질전환방법 이외에도, 미세사출법(microprojectile bombardment), 일렉트로포레이션(electroporation), PEG-매개 융합법(PEG-mediated fusion), 미세주입법(microinjection), 리포좀 매개법(liposome-mediated method), 인-플란타 형질전환법(In planta transformation), 진공 침윤법(Vacuum infiltration method), 화아침지법(floral meristem dipping method), 및 아그로박테리아 분사법(Agrobacterium spraying method)을 이용할 수 있으며, 더 구체적으로는 아그로박테리움을 이용한 형질전환방법 또는 아그로박테리아 분사법(Agrobacterium spraying method)을 이용할 수 있다.

[0165] 다음으로, 콩 종자를 흐르는 수도물에 30분 동안 세척한 뒤 정상적인 종피를 가지는 종자를 고른다. 준비된 콩 종자를 70% ethanol에서 30초 동안 살균하고 멸균수로 3회 세척한다. 이후 1% sodium hypochlorite 용액에서 30분 동안 shaking하면서 살균한 뒤 5분 간격으로 5회 세척한다. 여분의 물기를 제거한 콩 종자를 배꼽이 아래로 가도록 발아 배지(Germination medium(GM), 1 ℓ당 B5 medium 3.2g, Sucrose 20g, MES 0.5g, agar(sigma) 8g, pH5.7 with 1M KOH)에 파종 후 26°C 배양기에서 20시간 동안 암조건으로 발아시킨다.

[0166] 20시간 동안 GM배지에서 발아한 콩 종자를 scalpel(#11 blade)을 이용하여 종자 길이를 따라 절반으로 쪼갠다. 형질전환을 위하여 embryonic axis를 가지는 cothyledon을 선택하고, embryonic axis tip과 axial bud를 제거한다. 미리 준비해 둔 고농도(OD_{600}=2)의 *Agrobacterium* 용액을 칼끝에 묻혀 axial bud를 제거한 부위를 6-8회 상처를 준다. 상처를 준 half seed는 저농도의 *Agrobacterium* 용액 (OD_{600}=0.6)에 담궈 sonication 20초, vaccum 3분 처리한 뒤 30분 동안 접종 시킨다. *Agrobacterium*용액을 제거한 절편체를 멸균 여과지가 올려진 CCM배지에 치상하여 (7 explant) 4일 동안 공동 배양(23°C, 16h light, 8h dark)한다.

[0168] 4일간 공동 배양한 절편체는 *Agrobacterium*을 제거하기 위하여 항생제가 첨가된 멸균수(250mg/L cefotaxim, 100mg/L Timetin, 50mg/L vancomycin)로 5분씩 3회 세척하고 동일한 항생제가 첨가된 액체 순(shoot) 유도 배지(SIMIP0 배지, 1ℓ 당 B5 medium 3.2g, Sucrose 30g, Sodium thiosulfate 158mg, DTT 154mg, MES 0.6g agar(sigma) 8g, pH5.7 with 1M KOH)에서 15분간 3회 더 세척하였다. 세척 후 절편체로부터 여분의 물기를 제거한 뒤 항생제가 첨가되지않은 SIMIP0 배지에 하배축 부분을 45도 정도 비스듬하게 세워 치상하였다. 2주 후 half seed를 준비하는 가정에서 axial bud가 재대로 제거되지 않아 발생한 큰 shoot는 완전히 제거하고 나머지 작은 순(shoot)는 항생제 선발이 용이하도록 순(shoot)의 위쪽 부분만 제거하여 항생제가 첨가된 SIMP10 배지에 계대 배양하였다.

[0169] 2주 후 PPT에 의해서 갈변한 shoot는 가위를 이용하여 제거하고 하배축 부분에 형성된 큰 캘러스 덩어리는 다듬어서 SEMP5 배지(shoot elogtion medium_PPT5, 1ℓ당 MS B5 medium 4.4g, Sucrose 30g, Sodium thiosulfate 158mg, DTT 154mg, MES 0.6g, Asparagin 50mg, Pyroglutamic acid 100mg, agar(sigma) 8g, pH5.7 with 1M KOH, Zeatin-ribose 1mgL, GA3 1mg, IAA 0.1mg, Cefotaxim 125mg, Timetin 50mg, Vancomycin 50mg)에 치상하였다.

[0170] 2주마다 동일한 SEMP5 배지로 계대 배양하여 줄기가 4cm이상 신장하였을 경우 콩 형질전환체의 줄기 부분을 절

단하여 항생제가 첨가되지 않은 뿌리(Root) 유도 배지(RIMP0, 1ℓ당 MS B5 medium 4.4g, Sucrose 30g, Sodium thiosulfate 158mg, DTT 154mg, MES 0.6g, Asparagin 50mg, Pyroglutamic acid 100mg, agar(sigma) 8g, pH5.7 with 1M KOH, IBA 1mg, Cefotaxim 50mg, Timetin 50mg, Vancomycin 25mg) 배지에서 뿌리를 유도하였다. 줄기의 절단면에서 뿌리가 발생한 형질전환체는 작은 화분으로 옮겨 플라스틱 통에서 2주동안 순화한 후 큰 화분으로 옮겨 온실에서 재배하였다.

[0172] 재분화된 콩 식물체를 순화시킨 후, pot에 이식하여 생장시키고, 재분화된 식물체의 T-DNA 도입을 확인하기 위하여 개체별로 genomic DNA를 추출하고 genomic PCR을 통하여 도입한 선발마커 유전자 및 gRNA 발현 카세트의 도입을 확인함. 또한 Bar-strip을 이용하여 재분화된 식물체에서의 선발마커 유전자(*Bar*)의 발현을 확인하였다.

[0173] 또한, T-DNA 도입이 확인된 T_0 식물체로부터 추출한 genomic DNA를 주형으로 하여 InDel PCR을 통하여 유전자 결실(deletion)이 발생하였는지를 확인하였다. 이후 각 형질전환 라인들의 deletion 위치와 패턴, genome-editing 효율을 targeted deep sequencing을 통하여 확인하였다.

[0175] **[실시예]**

[0176] 이하, 실시예를 통하여 본 발명을 더욱 상세히 설명하고자 한다. 본 발명의 실시예는 오로지 본 발명을 예시하기 위한 것으로서, 본 발명의 범위가 이들 실시예에 의해 제한되는 것으로 해석되지는 않는 것은 이 기술 분야에서 통상의 지식을 가진 자에게 있어서 자명할 것이다.

[0178] **실시예 1. 고효율의 gRNA 선발**

[0179] 식물에서 피틴산(phytic acid, Ins(1,2,3,4,5,6)P_6) 생합성은 많은 효소들이 관여하는 과정으로, 이 생합성 경로에서 IPK1 (inositol 1,3,5,6-pentakisphosphate 2-kinase) 효소는 피틴산 생합성 경로의 마지막 단계를 촉매하는 것으로 알려져 있기 때문에, 유전자교정을 통한 저피틴산 콩 제조를 위해서 *GmIPK1* 유전자를 CRISPR/Cas9 시스템을 이용하여 결실시킬 표적 유전자로 선정하였다.

[0180] 이에, *GmIPK1(Glyma14g072200)* 유전자에 대한 표적 서열에 특이적으로 표적할 수 있는 gRNA를 선발하고자 하였다. 여기서 "gRNA"는 가이드 RNA(guide RNA)와 다른 개념으로, 가이드 RNA(guide RNA)에서 표적 핵산 또는 유전자 내의 표적 서열에 결합하는 약 20 개의 뉴클레오티드(20nt)에 해당하는 스페이서(spacer) 서열을 의미한다.

[0182] 먼저, 콩 유전자교정 CRISPR/Cas9 시스템을 위한 고효율의 gRNA1 및 gRNA2의 조합을 선별하기 위해서, 유전자교정을 위한 표적 유전자에 대한 염기서열 정보를 Phytozome database에서 확보하였다.

[0183] 'CRISPR-P' 또는 'CRISPR RGEN Tools' 등의 유전자교정 Web Tool을 이용하여 비표적유전자를 교정하는 오프-타겟 효과(off-target effect)가 적고, 유전자교정 효율이 높을 것으로 예측되는 *GmIPK1* 유전자 내의 표적 서열(target sequence)을 선정하여, 이 표적 서열(target sequence)에 결합하는 고효율의 gRNA1 및 gRNA2의 조합을 선별하였다. 이를 위해, 형질전환 모본으로 사용한 Willimas 82 품종의 염기서열을 바탕으로 표적 유전자 내의 표적 서열을 확인하였다.

[0184] 이 때, CRISPR-P에 비해 RGEN이 보다 많은 off-target 유전자 후보군(candidate)들을 제시해 주기 때문에, 두 가지 프로그램을 모두 사용하여 오프-타겟(off-target) 가능성이 적은 *GmIPK1* 유전자 내의 표적 서열을 선정하였다. 또한, 오프-타겟(off-target) 유전자 후보(candidate)들이 최종 선발되면 Jbrowse site를 이용하여 오프-타겟(off-target) 유전자 내에서의 gRNA 결합 위치 등을 최종 확인하여 유전자교정 효율이 높고 오프-타겟(off-target) 가능성이 적은 표적 유전자 내의 표적 서열(target sequence)을 선발하였다.

[0186] 그 결과, IPK1 단백질을 암호화하는 *GmIPK1* 유전자를 특이적으로 표적할 수 있는 2 개의 gRNA 조합을 최종 선발하였다. 상기 최종 선발된 2 개의 gRNA가 표적하는 표적 서열(target sequence) 및 이의 염색체 상 위치를 도 1에 나타내었다. 또한, 상기 표적 서열(target sequence)에 중에서 엑손(Exon) 2 내의 표적 서열(Target 1) 및

엑손(Exon) 3 내의 표적 서열(Target 2)에 각각 결합하는 gRNA를 각각 gRNA1 및 gRNA2라고 명명하고, gRNA1 및 gRNA2를 각각 암호화하는 핵산 서열을 도 1에 도시하였다.

[0188] 또한, 아래 표 1에 gRNA1 및 gRNA2를 각각 암호화하는 핵산 서열을 나타내었다.

표 1

[0189]

Name	서열(5'→3')	서열번호
gRNA1	5'-GAAAGTGGTCCGCATACGTA-3'	1
gRNA2	5'-ATCACAATGTGCATCAACCC-3'	2

[0191] **실시예 2. 코돈-최적화된 CRISPR/Cas9 벡터 제조**

[0192] 다음으로, 본 발명은 콩 유전자교정의 표적 유전자인 *GmIPK1* 유전자에 대해서, 각 표적 유전자 당 선정된 2 개의 표적 서열(target sequence)에 각각 결합하는 gRNA1 및 gRNA2를 암호화하는 핵산을 하나의 벡터에 도입한 multiplex 벡터를 제조하기로 하였다. 이것은 하나의 표적 유전자 내에서 넓은 범위에 위치하고 있는 2 개의 표적 서열(target sequence) 부위를 절단함으로써, 표적 유전자의 확실한 인델(indel) 변이를 야기하고, 형질전환체를 확보한 후에 Indel PCR을 이용하여 유전자교정이 일어난 형질전환체를 보다 효율적으로 선별할 수 있는 장점을 가지는 시스템을 제조하기 위함이다.

[0193]

[0194] 상기 선정된 2개의 표적 부위에 특이적으로 결합하는 gRNA1 및 gRNA2를 포함하고, 콩에서 코돈-최적화된 Cas9 유전자를 포함하는 multiplex 벡터를 제조하였다. 이때, gRNA의 발현을 목적으로 AtU6 프로모터를 사용할 경우 gRNA를 암호화하는 핵산은 'G' 로 시작하는 것이 효율이 좋기 때문에, gRNA1 및 gRNA2를 각각 암호화하는 핵산인 서열번호 1(5'-GAAAGTGGTCCGCATACGTA-3') 및 서열번호 2(5'-ATCACAATGTGCATCAACCC-3') 각각의 5'-말단 앞에 염기 'G' 를 첨가하여, 상기 multiplex 벡터 제조시 이를 사용하였다.즉, AtU6 프로모터를 가지는 벡터에 포함되는 gRNA1 및 gRNA2를 암호화하는 핵산은 각각 서열번호 3(5'-**G**GAAAGTGGTCCGCATACGTA-3') 및 서열번호 4(5'-**G**ATCACAATGTGCATCAACCC-3')를 포함한다.

[0196] 콩 형질전환에 사용한 기본 벡터는 pMDC123 벡터(vector) 또는 pECO201 벡터이다. 상기 벡터들은 선발 마커로 *Bar* 유전자를 가지며, 재조합을 위하여 attR1과 attR2 서열을 가지는 gateway용 벡터이다.

[0197] 상기 pECO201 벡터에 본 발명에 따라 선정된 상기 2 개의 표적 서열(target sequence)에 결합하는 gRNA1 및 gRNA2를 암호화하는 핵산을 클로닝하여 multiplex 벡터를 제조하였다. pECO201 벡터는 2개 이상의 gRNA를 동시에 도입하기 위한 golden gateway entry vector로, 선발 마커로 *Bar* 유전자를 가지며, 35S 프로모터의 조절하에 애기장대 코돈-최적화된 SpCas9(*Arabidopsis* codon optimized SpCas9)이 발현될 수 있도록 제작되어져 있다. 두 번의 과정에 의해서 pECO201 벡터로 두 개의 gRNA를 조합하였다. 첫 번째 oligo annealing 과정을 거친 gRNA1과 gRNA2는 *Bsa* I으로 절단한 pECO400_1,2 및 pECO404_2,e와 각각 ligation하여 pGRNA1과 pGRNA2e plasmid를 획득하였다. 두 번째 각각의 gRNA가 도입된 pGRNA1 및 pGRNA2e와 pECO201 벡터를 golden gateway assembly를 통하여 pECO201_g1_g2 vector를 최종 제작하였다.

[0199] 본 발명에 따른 선별된 gRNA1 및 gRNA2를 각각 암호화하는 핵산과 코돈-최적화된 Cas9 유전자를 포함하는 예시적인 벡터의 모식도는 도 3에 나타냈다.

[0201] **실시예 3. 콩 형질전환체의 유전자교정 검증**

[0202] 상기 실시예 2에서 제작한, 선별된 gRNA1 및 gRNA2를 각각 암호화하는 핵산 및 코돈-최적화된 Cas9 유전자를 포함하는 콩 유전자교정을 위한 CRISPR/Cas9 벡터를 아그로박테리움 EHA105에 형질전환 한 후, Willimas 82 품

종 콩의 절편체와 공동배양하여 형질전환시켰다.

[0203] 재분화된 콩 식물체를 순화시킨 후, pot에 이식하여 생장시켰다. 재분화된 식물체의 T-DNA 도입을 확인하기 위하여 개체별로 genomic DNA를 추출하고 genomic PCR을 통하여 도입한 선발마커 유전자 및 gRNA 발현 카세트의 도입을 확인하였다(도 4).

[0204] 도 4에서 나타내는 바와 같이, *GmIPK1* 유전자 내 표적 부위 2 곳을 특이적으로 표적으로 하는 gRNA1 및 gRNA2를 발현하는 T_0 계통을 확보하였다(Bar PCR positive).

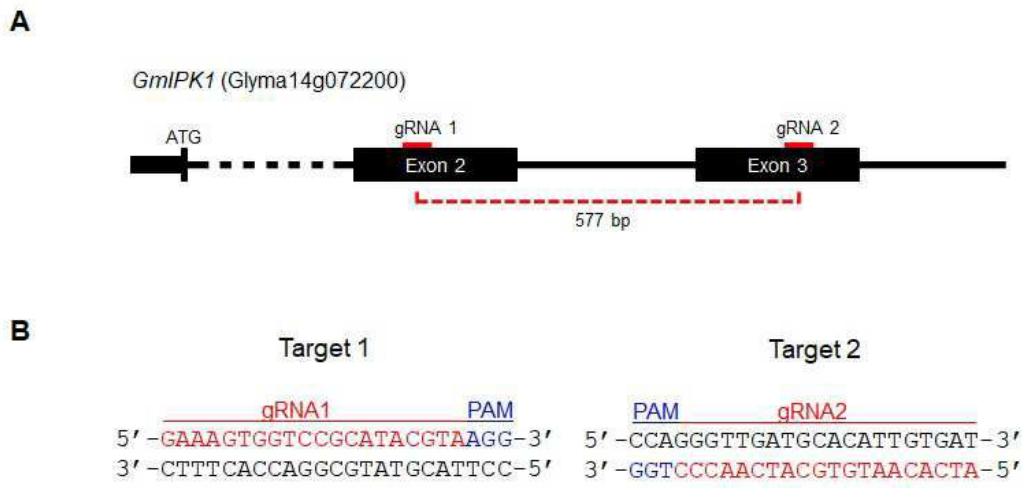
[0206] 또한, 도 5에서 나타내는 바와 같이, Bar 저항성 유전자에 대한 PCR에서 positive 반응을 보인 식물체를 이용한 InDel PCR을 통하여 두 gRNA1 및 gRNA2 표적 사이의 염기서열이 deletion된 형질전환체를 확인하였다.

[0207] 상기 확보된 콩 형질전환체는 *GmIPK1* 유전자 내의 표적 서열 부위 2 곳에서 모두 인델변이(indel mutation)가 발생하여 577 bp 결실(deletion)이 일어난 것을 Targeted deep sequenceing을 통해서 확인하였다(도 6A). 상기 확보된 콩 형질전환체가 IPK1 단백질에서 95개의 아미노산이 결실되어 야생형 또는 정상기능을 하는 IPK1 단백질이 발현되지 못하는 것을 도 6B에 도시하였다.

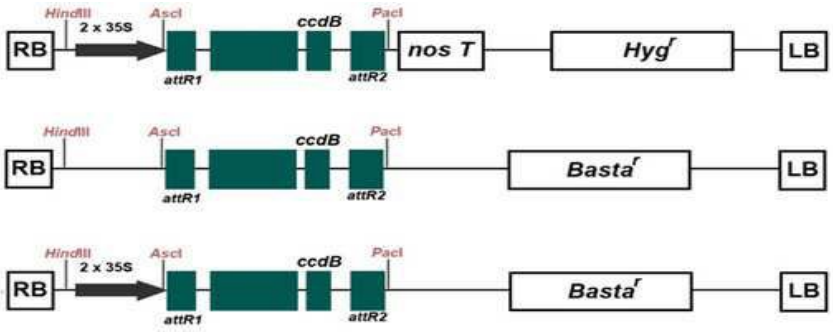
[0209] 상기 결과를 통해서, 본 발명에 따른 CRISPR/Cas 기반의 콩 유전자 교정시스템을 이용하여, 본 발명은 *GmIPK1* 유전자 내의 표적 부위 2 곳을 동시에 표적으로 하는 유전자교정이 콩 식물체에서 효율적으로 일어났음을 확인할 수 있었다. 또한 본 발명은 콩의 *GmIPK1* 유전자가 결실된 콩 형질전환체를 제조하였고 이의 종자를 확보하였다. 또한 본 발명에 따른 콩의 *GmIPK1* 유전자가 교정된 콩 형질전환체 및 이의 종자는 다음 세대에서 다양한 조합의 유전자교정 형태를 가지는 식물체를 제조하는데 이용될 수 있다.

도면

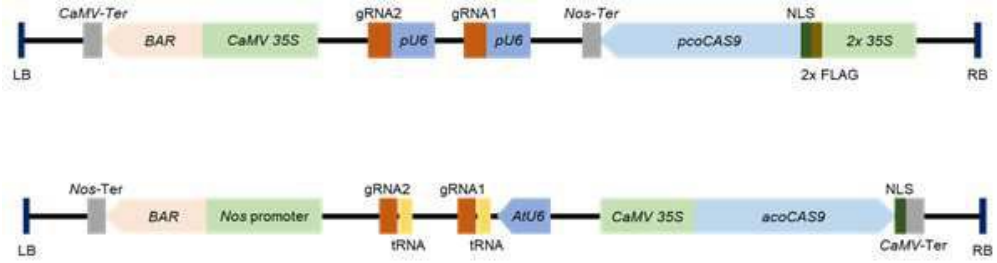
도면1



도면2

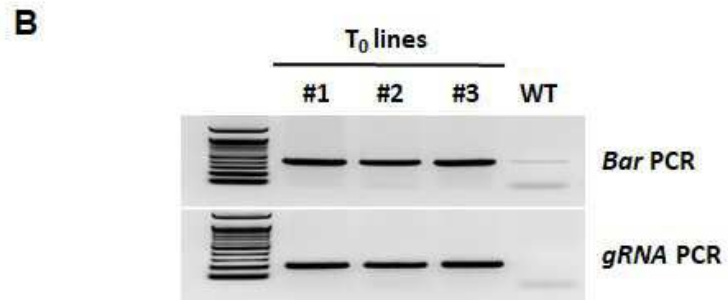

HindIII
2 x 35S
AscI
ccdB
PacI
RB
attR1
attR2
nos T
Hyg^r
LB
Basta^r

도면3

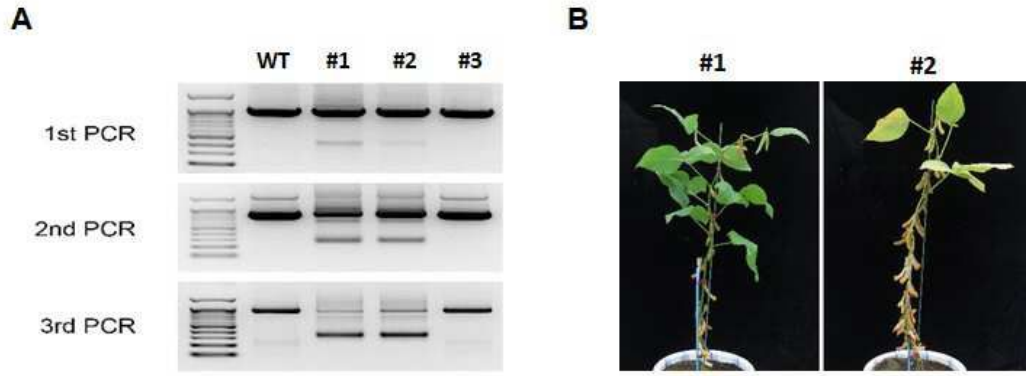

CaMV-Ter
BAR
CaMV 35S
gRNA2
pU6
gRNA1
Nos-Ter
pcoCAS9
NLS
2x 35S
2x FLAG
LB
RB
Nos promoter
tRNA
AtU6
acoCAS9

도면4

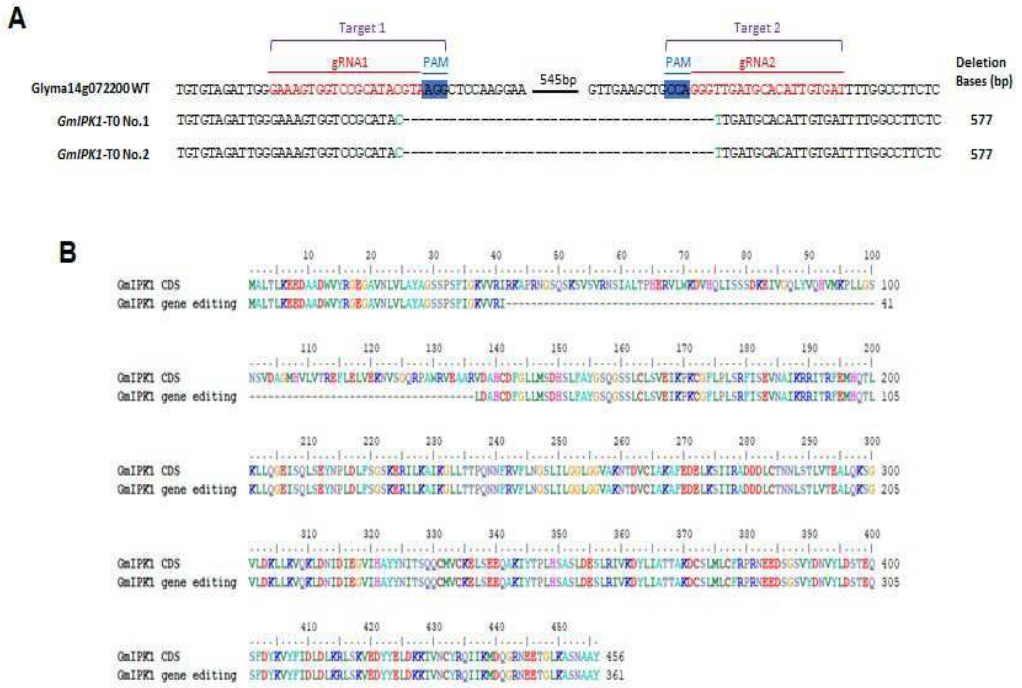

A


B
T0 lines
#1
#2
#3
WT
Bar PCR
gRNA PCR

도면5



도면6



서 열 목 록

<110> INDUSTRY-ACADEMIC COOPERATION FOUNDATION GYEONGSANG NATIONAL UNIVERSITY

<120> Gene editing system for GmIPK1 gene editing and uses thereof

<130> PN21249

<160> 12

<170> KoPatentIn 3.0

<210> 1

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> nucleotide sequence encoding gRNA1

<400> 1

gaaagtggtc cgcatacgta 20

<210> 2

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223>

nucleotide sequence encoding gRNA2

<400> 2

atcacaatgt gcatcaaccc 20

<210> 3

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> nucleotide sequence encoding gRNA1

<400> 3

ggaaagtggt ccgcatacgt a 21

<210> 4

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> nucleotide sequence encoding gRNA2

<400> 4

gatcacaatg tgcatcaacc c 21

<210> 5

<211> 4293

<212> DNA

<213> Artificial Sequence

<220><223> pcoCas9

<400> 5

atggataaga agtactctat cggacttgac atcggaacca actctgttgg atgggctgtt 60

atcaccgatg agtacaaggt tccatctaag aagttcaagg ttcttggaaa caccgataga 120

cactctatca agaagaacct tatcggtgct cttcttttcg attctggaga gaccgctgag 180

gctaccagat tgaagagaac cgctagaaga agatacacca gaagaaagaa cagaatctgc 240

taccttcagg aaatcttctc taacgagatg gctaaggttg atgattcttt cttccacaga 300

cttgaggagt ctttccttgt tgaggaggat aagaagcacg agagacaccc aatcttcgga 360

aacatcgttg atgaggttgc ttaccacgag aagtacccaa ccatctacca ccttagaaag 420

aagttggttg attctaccga taaggctgat cttagactta tctaccttgc tcttgctcac 480

atgatcaagt tcagaggaca cttccttatc gagggagacc ttaacccaga taactctgat 540

gttgataagt tgttcatcca gcttgttcag acctacaacc agcttttcga ggagaaccca 600

atcaacgctt ctggagttga tgctaaggct atcctttctg ctagactttc taagtctcgt 660

agacttgaga accttatcgc tcagcttcca ggagagaaga agaacggact tttcggaaac 720

cttatcgctc tttctcttgg acttacccca aacttcaagt ctaacttcga tcttgctgag 780

gatgctaagt tgcagctttc taaggatacc tacgatgatg atcttgataa ccttcttgct 840

cagatcggag atcagtacgc tgatcttttc cttgctgcta agaacctttc tgatgctatc 900

cttctttctg acatccttag agttaacacc gagatcacca aggctccact ttctgcttct 960

atgatcaaga gatacgatga gcaccaccag gatcttaccc ttttgaaggc tcttgttaga 1020

cagcagcttc cagagaagta caaggaaatc ttcttcgatc agtctaagaa cggatacgct 1080

ggatacatcg atggaggagc ttctcaggag gagttctaca agttcatcaa gccaatcctt 1140

gagaagatgg atggaaccga ggagcttctt gttaagttga acagagagga tcttcttaga 1200

aagcagagaa ccttcgataa cggatctatc ccacaccaga tccaccttgg agagcttcac 1260

gctatccttc gtagacagga ggatttctac ccattcttga aggataacag agagaagatc 1320

gagaagatcc ttaccttcag aatcccatac tacgttggac cacttgctag aggaaactct 1380

cgtttcgctt ggatgaccag aaagtctgag gagaccatca ccccttggaa cttcgaggag 1440

gtaagtttct gcttctacct ttgatatata tataataatt atcattaatt agtagtaata 1500

taatatttca aatatttttt tcaaaataaa agaatgtagt atatagcaat tgcttttctg 1560

tagtttataa gtgtgtatat tttaatttat aacttttcta atatatgacc aaaatttgtt 1620

gatgtgcagg ttgttgataa gggagcttct gctcagtctt tcatcgagag aatgaccaac 1680

ttcgataaga accttccaaa cgagaaggtt cttccaaagc actctcttct ttacgagtac 1740

ttcaccgttt acaacgagct taccaaggtt aagtacgtta ccgagggaat gagaaagcca 1800

gctttccttt ctggagagca gaagaaggct atcgttgatc ttcttttcaa gaccaacaga 1860

aaggttaccg ttaagcagtt gaaggaggat tacttcaaga agatcgagtg cttcgattct 1920

gttgaaatct ctggagttga ggatagattc aacgcttctc ttggaaccta ccacgatctt 1980

ttgaagatca tcaaggataa ggatttcctt gataacgagg agaacgagga catccttgag 2040

gacatcgttc ttacccttac ccttttcgag gatagagaga tgatcgagga gagactcaag 2100

acctacgctc acctttttcga tgataaggtt atgaagcagt tgaagagaag aagatacacc 2160

ggatggggta gactttctcg taagttgatc aacggaatca gagataagca gtctggaaag 2220

accatccttg atttcttgaa gtctgatgga ttcgctaaca gaaacttcat gcagcttatc 2280

cacgatgatt ctcttacctt caaggaggac atccagaagg ctcaggtttc tggacaggga 2340

gattctcttc acgagcacat cgctaacctt gctggatctc cagctatcaa gaagggaatc 2400

cttcagaccg ttaaggttgt tgatgagctt gttaaggtta tgggtagaca caagccagag 2460

aacatcgtta tcgagatggc tagagagaac cagaccaccc agaagggaca gaagaactct 2520

cgtgagagaa tgaagagaat cgaggaggga atcaaggagc ttggatctca aatcttgaag 2580

gagcacccag ttgagaacac ccagcttcag aacgagaagt tgtaccttta ctaccttcag 2640

aacggaagag atatgtacgt tgatcaggag cttgacatca acagactttc tgattacgat 2700

gttgatcaca tcgttccaca gtctttcttg aaggatgatt ctatcgataa caaggttctt 2760

acccgttctg ataagaacag aggaaagtct gataacgttc catctgagga ggttgttaag 2820

aagatgaaga actactggag acagcttctt aacgctaagt tgatcaccca gagaaagttc 2880

gataaccttа ccaaggctga gagaggagga ctttctgagc ttgataaggc tggattcatc 2940

aagagacagc ttgttgagac cagacagatc accaagcacg ttgctcagat ccttgattct 3000

cgtatgaaca ccaagtacga tgagaacgat aagttgatca gagaggttaa ggttatcacc 3060

ttgaagtcta agttggtttc tgatttcaga aaggatttcc agttctacaa ggttagagag 3120

atcaacaact accaccacgc tcacgatgct taccttaacg ctgttgttgg aaccgctctt 3180

atcaagaagt acccaaagtt ggagtctgag ttcgtttacg gagattacaa ggtttacgat 3240

gttagaaaga tgatcgctaa gtctgagcag gagatcggaa aggctaccgc taagtacttc 3300

ttctactcta acatcatgaa cttcttcaag accgagatca cccttgctaa cggagagatc 3360

agaaagagac cacttatcga gaccaacgga gagaccggag agatcgtttg ggataaggga 3420

agagatttcg ctaccgttag aaaggttctt tctatgccac aggttaacat cgttaagaaa 3480

accgaggttc agaccggagg attctctaag gagtctatcc ttccaaagag aaactctgat 3540

aagttgatcg ctagaaagaa ggattgggac ccaaagaagt acggaggatt cgattctcca 3600

accgttgctt actctgttct tgttgttgct aaggttgaga agggaaagtc taagaagttg 3660

aagtctgtta aggagcttct tggaatcacc atcatggagc gttcttcttt cgagaagaac 3720

ccaatcgatt tccttgaggc taagggatac aaggaggtta agaaggatct tatcatcaag 3780

ttgccaaagt actctctttt cgagcttgag aacggaagaa agagaatgct tgcttctgct 3840

ggagagcttc agaagggaaa cgagcttgct cttccatcta agtacgttaa cttcctttac 3900

cttgcttctc actacgagaa gttgaaggga tctccagagg ataacgagca gaagcagctt 3960

ttcgttgagc agcacaagca ctaccttgat gagatcatcg agcaaatctc tgagttctct 4020

aagagagtta tccttgctga tgctaacctt gataaggttc tttctgctta caacaagcac 4080

agagataagc caatcagaga gcaggctgag aacatcatcc acctttcac ccttaccaac 4140

cttggtgctc cagctgcttt caagtacttc gataccacca tcgatagaaa aagatacacc 4200

tctaccaagg aggttcttga tgctaccctt atccaccagt ctatcaccgg actttacgag 4260

accagaatcg atctttctca gcttggagga gat 4293

<210> 6

<211> 4116

<212> DNA

<213> Artificial Sequence

<220><223> acoCas9

<400> 6

atggataaga agtactctat cggactcgat atcggaacta actctgtggg atgggctgtg 60

atcaccgatg agtacaaggt gccatctaag aagttcaagg ttctcggaaa caccgatagg 120

cactctatca agaaaaacct tatcggtgct ctcctcttcg attctggtga aactgctgag 180

gctaccagac tcaagagaac cgctagaaga aggtacacca gaagaaagaa caggatctgc 240

tacctccaag agatcttctc taacgagatg gctaaagtgg atgattcatt cttccacagg 300

ctcgaagagt cattcctcgt ggaagaagat aagaagcacg agaggcaccc tatcttcgga 360

aacatcgttg atgaggtggc ataccacgag aagtacccta ctatctacca cctcagaaag 420

aagctcgttg attctactga taaggctgat ctcaggctca tctacctcgc tctcgctcac 480

atgatcaagt tcagaggaca cttcctcatc gagggtgatc tcaaccctga taactctgat 540

gtggataagt tgttcatcca gctcgtgcag acctacaacc agcttttcga agagaaccct 600

atcaacgctt caggtgtgga tgctaaggct atcctctctg ctaggctctc taagtcaaga 660

aggcttgaga acctcattgc tcagctccct ggtgagaaga agaacggact tttcggaaac 720

ttgatcgctc tctctctcgg actcacccct aacttcaagt ctaacttcga tctcgctgag 780

gatgcaaagc tccagctctc aaaggatacc tacgatgatg atctcgataa cctcctcgct 840

cagatcggag atcagtacgc tgatttgttc ctcgctgcta agaacctctc tgatgctatc 900

ctcctcagtg atatcctcag agtgaacacc gagatcacca aggctccact ctcagcttct 960

atgatcaaga gatacgatga gcaccaccag gatctcacac ttctcaaggc tcttgttaga 1020

cagcagctcc cagagaagta caaagagatt ttcttcgatc agtctaagaa cggatacgct 1080
ggttacatcg atggtggtgc atctcaagaa gagttctaca agttcatcaa gcctatcctc 1140
gagaagatgg atggaaccga ggaactcctc gtgaagctca atagagagga tcttctcaga 1200
aagcagagga ccttcgataa cggatctatc cctcatcaga tccacctcgg agagttgcac 1260
gctatcctta gaaggcaaga ggatttctac ccattcctca aggataacag ggaaaagatt 1320
gagaagattc tcaccttcag aatcccttac tacgtgggac ctctcgctag aggaaactca 1380
agattcgctt ggatgaccag aaagtctgag gaaaccatca ccccttggaa cttcgaagag 1440

gtggtggata agggtgctag tgctcagtct ttcatcgaga ggatgaccaa cttcgataag 1500
aaccttccaa acgagaaggt gctccctaag cactctttgc tctacgagta cttcaccgtg 1560
tacaacgagt tgaccaaggt taagtacgtg accgagggaa tgaggaagcc tgctttttg 1620
tcaggtgagc aaaagaaggc tatcgttgat ctcttgttca agaccaacag aaaggtgacc 1680
gtgaagcagc tcaaagagga ttacttcaag aaaatcgagt gcttcgattc agttgagatt 1740
tctggtgttg aggataggtt caacgcatct ctcggaacct accacgatct cctcaagatc 1800
attaaggata aggatttctt ggataacgag gaaaacgagg atatcttgga ggatatcgtt 1860

cttaccctca ccctctttga agatagagag atgattgaag aaaggctcaa gacctacgct 1920
catctcttcg atgataaggt gatgaagcag ttgaagagaa gaagatacac tggttgggga 1980
aggctctcaa gaaagctcat taacggaatc agggataagc agtctggaaa gacaatcctt 2040
gatttcctca agtctgatgg attcgctaac agaaacttca tgcagctcat ccacgatgat 2100
tctctcacct ttaaagagga tatccagaag gctcaggttt caggacaggg tgatagtctc 2160
catgagcata tcgctaacct cgctggatct cctgcaatca agaagggaat cctccagact 2220
gtgaaggttg tggatgagtt ggtgaaggtg atgggaaggc ataagcctga gaacatcgtg 2280

atcgaaatgg ctagagagaa ccagaccact cagaagggac agaagaactc tagggaaagg 2340
atgaagagga tcgaggaagg tatcaaagag cttggatctc agatcctcaa agagcaccct 2400
gttgagaaca ctcagctcca gaatgagaag ctctacctct actacctcca gaacggaagg 2460
gatatgtatg tggatcaaga gttggatatc aacaggctct ctgattacga tgttgatcat 2520
atcgtgccac agtcattctt gaaggatgat tctatcgata acaaggtgct caccaggtct 2580
gataagaaca ggggtaagag tgataacgtg ccaagtgaag aggttgtgaa gaaaatgaag 2640
aactattgga ggcagctcct caacgctaag ctcatcactc agagaaagtt cgataacttg 2700

actaaggctg agaggggagg actctctgaa ttggataagg caggattcat caagaggcag 2760

cttgtggaaa ccaggcagat cactaagcac gttgcacaga tcctcgattc taggatgaac 2820

accaagtacg atgagaacga taagttgatc agggaagtga aggttatcac cctcaagtca 2880

aagctcgtgt ctgatttcag aaaggatttc caattctaca aggtgaggga aatcaacaac 2940

taccaccacg ctcacgatgc ttaccttaac gctgttgttg gaaccgctct catcaagaag 3000

tatcctaagc tcgagtcaga gttcgtgtac ggtgattaca aggtgtacga tgtgaggaag 3060

atgatcgcta agtctgagca agagatcgga aaggctaccg ctaagtattt cttctactct 3120

aacatcatga atttcttcaa gaccgagatt accctcgcta acggtgagat cagaaagagg 3180

ccactcatcg agacaaacgg tgaaacaggt gagatcgtgt gggataaggg aagggatttc 3240

gctaccgtta gaaaggtgct ctctatgcca caggtgaaca tcgttaagaa aaccgaggtg 3300

cagaccggtg gattctctaa agagtctatc ctccctaaga ggaactctga taagctcatt 3360

gctaggaaga aggattggga ccctaagaaa tacggtggtt tcgattctcc taccgtggct 3420

tactctgttc tcgttgtggc taaggttgag aagggaaaga gtaagaagct caagtctgtt 3480

aaggaacttc tcggaatcac tatcatggaa aggtcatctt tcgagaagaa cccaatcgat 3540

ttcctcgagg ctaagggata caaagaggtt aagaaggatc tcatcatcaa gctcccaaag 3600

tactcactct tcgaactcga gaacggtaga aagaggatgc tcgcttctgc tggtgagctt 3660

caaaagggaa acgagcttgc tctcccatct aagtacgtta actttcttta cctcgcttct 3720

cactacgaga agttgaaggg atctccagaa gataacgagc agaagcaact tttcgttgag 3780

cagcacaagc actacttgga tgagatcatc gagcagatct ctgagttctc taaaagggtg 3840

atcctcgctg atgcaaacct cgataaggtg ttgtctgctt acaacaagca cagagataag 3900

cctatcaggg aacaggcaga gaacatcatc catctcttca cccttaccaa cctcggtgct 3960

cctgctgctt tcaagtactt cgatacaacc atcgatagga agagatacac ctctaccaaa 4020

gaagtgctcg atgctaccct catccatcag tctatcactg gactctacga gactaggatc 4080

gatctctcac agctcggtgg tgattcaagg gctgat 4116

<210> 7

<211> 4104

<212> DNA

<213> Artificial Sequence

<220><223> pcohCas9

<400> 7

atggacaaga agtactccat cggcctcgac atcggcacca acagcgtcgg ctgggcggtg 60

atcaccgacg agtacaaggt cccgtccaag aagttcaagg tcctgggcaa caccgaccgc 120

cactccatca agaagaacct catcggcgcc ctcctcttcg actccggcga gacggcggag 180

gcgacccgcc tcaagcgcac cgcccgccgc cgctacaccc gccgcaagaa ccgcatctgc 240

tacctccagg agatcttctc caacgagatg gcgaaggtcg acgactcctt cttccaccgc 300

ctcgaggagt ccttcctcgt ggaggaggac aagaagcacg agcgccaccc catcttcggc 360

aacatcgtcg acgaggtcgc ctaccacgag aagtacccca ctatctacca ccttcgtaag 420

aagcttgttg actctactga taaggctgat cttcgtctca tctaccttgc tctcgctcac 480

atgatcaagt tccgtggtca cttccttatc gagggtgacc ttaaccctga taactccgac 540

gtggacaagc tcttcatcca gctcgtccag acctacaacc agctcttcga ggagaaccct 600

atcaacgctt ccggtgtcga cgctaaggcg atcctttccg ctaggctctc caagtccagg 660

cgtctcgaga acctcatcgc ccagctccct ggtgagaaga agaacggtct tttcggtaac 720

ctcatcgctc tctccctcgg tctgacccct aacttcaagt ccaacttcga cctcgctgag 780

gacgctaagc ttcagctctc caaggatacc tacgacgatg atctcgacaa cctcctcgct 840

cagattggag atcagtacgc tgatctcttc cttgctgcta agaacctctc cgatgctatc 900

ctcctttcgg atatccttag ggttaacact gagatcacta aggctcctct ttctgcttcc 960

atgatcaagc gctacgacga gcaccaccag gacctcaccc tcctcaaggc tcttgttcgt 1020

cagcagctcc ccgagaagta caaggagatc ttcttcgacc agtccaagaa cggctacgcc 1080

ggttacattg acggtggagc tagccaggag gagttctaca agttcatcaa gccaatcctt 1140

gagaagatgg atggtactga ggagcttctc gttaagctta accgtgagga cctccttagg 1200

aagcagagga ctttcgataa cggctctatc cctcaccaga tccaccttgg tgagcttcac 1260

gccatccttc gtaggcagga ggacttctac cctttcctca aggacaaccg tgagaagatc 1320

gagaagatcc ttactttccg tattccttac tacgttggtc ctcttgctcg tggtaactcc 1380

cgtttcgctt ggatgactag gaagtccgag gagactatca ccccttggaa cttcgaggag 1440

gttgttgaca agggtgcttc cgcccagtcc ttcatcgagc gcatgaccaa cttcgacaag 1500

aacctcccca acgagaaggt cctccccaag cactccctcc tctacgagta cttcacggtc 1560

tacaacgagc tcaccaaggt caagtacgtc accgagggta tgcgcaagcc tgccttcctc 1620

tccggcgagc agaagaaggc tatcgttgac ctcctcttca agaccaaccg caaggtcacc 1680

gtcaagcagc tcaaggagga ctacttcaag aagatcgagt gcttcgactc cgtcgagatc 1740

agcggcgttg aggaccgttt caacgcttct ctcggtacct accacgatct cctcaagatc 1800

atcaaggaca aggacttcct cgacaacgag gagaacgagg acatcctcga ggacatcgtc 1860

ctcactctta ctctcttcga ggatagggag atgatcgagg agaggctcaa gacttacgct 1920

catctcttcg atgacaaggt tatgaagcag ctcaagcgtc gccgttacac cggttggggt 1980

aggctctccc gcaagctcat caacggtatc agggataagc agagcggcaa gactatcctc 2040

gacttcctca agtctgatgg tttcgctaac aggaacttca tgcagctcat ccacgatgac 2100

tctcttacct tcaaggagga tattcagaag gctcaggtgt ccggtcaggg cgactctctc 2160

cacgagcaca ttgctaacct tgctggttcc cctgctatca agaagggcat ccttcagact 2220

gttaaggttg tcgatgagct tgtcaaggtt atgggtcgtc acaagcctga gaacatcgtc 2280

atcgagatgg ctcgtgagaa ccagactacc cagaagggtc agaagaactc gagggagcgc 2340

atgaagagga ttgaggaggg tatcaaggag cttggttctc agatccttaa ggagcaccct 2400

gtcgagaaca cccagctcca gaacgagaag ctctacctct actacctcca gaacggtagg 2460

gatatgtacg ttgaccagga gctcgacatc aacaggcttt ctgactacga cgtcgaccac 2520

attgttcctc agtctttcct taaggatgac tccatcgaca acaaggtcct cacgaggtcc 2580

gacaagaaca ggggtaagtc ggacaacgtc ccttccgagg aggttgtcaa gaagatgaag 2640

aactactgga ggcagcttct caacgctaag ctcattaccc agaggaagtt cgacaacctc 2700

acgaaggctg agaggggtgg cctttccgag cttgacaagg ctggtttcat caagaggcag 2760

cttgttgaga cgaggcagat taccaagcac gttgctcaga tcctcgattc taggatgaac 2820

accaagtacg acgagaacga caagctcatc cgcgaggtca aggtgatcac cctcaagtcc 2880

aagctcgtct ccgacttccg caaggacttc cagttctaca aggtccgcga gatcaacaac 2940

taccaccacg ctcacgatgc ttaccttaac gctgtcgttg gtaccgctct tatcaagaag 3000

taccctaagc ttgagtccga gttcgtctac ggtgactaca aggtctacga cgttcgtaag 3060

atgatcgcca agtccgagca ggagatcggc aaggccaccg ccaagtactt cttctactcc 3120

aacatcatga acttcttcaa gaccgagatc accctcgcca acggcgagat ccgcaagcgc 3180

cctcttatcg agacgaacgg tgagactggt gagatcgttt gggacaaggg tcgcgacttc 3240

gctactgttc gcaaggtcct ttctatgcct caggttaaca tcgtcaagaa gaccgaggtc 3300

cagaccggtg gcttctccaa ggagtctatc cttccaaaga gaaactcgga caagctcatc 3360

gctaggaaga aggattggga ccctaagaag tacggtggtt tcgactcccc tactgtcgcc 3420

tactccgtcc tcgtggtcgc caaggtggag aagggtaagt cgaagaagct caagtccgtc 3480

aaggagctcc tcggcatcac catcatggag cgctcctcct tcgagaagaa cccgatcgac 3540

ttcctcgagg ccaagggcta caaggaggtc aagaaggacc tcatcatcaa gctccccaag 3600

tactctcttt tcgagctcga gaacggtcgt aagaggatgc tggcttccgc tggtgagctc 3660

cagaagggta acgagcttgc tcttccttcc aagtacgtga acttcctcta cctcgcctcc 3720

cactacgaga agctcaaggg ttcccctgag gataacgagc agaagcagct cttcgtggag 3780

cagcacaagc actacctcga cgagatcatc gagcagatct ccgagttctc caagcgcgtc 3840

atcctcgctg acgctaacct cgacaaggtc ctctccgcct acaacaagca ccgcgacaag 3900

cccatccgcg agcaggccga gaacatcatc cacctcttca cgctcacgaa cctcggcgcc 3960

cctgctgctt tcaagtactt cgacaccacc atcgacagga agcgttacac gtccaccaag 4020

gaggttctcg acgctactct catccaccag tccatcaccg gtctttacga gactcgtatc 4080

gacctttccc agcttggtgg tgat 4104

<210> 8

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> SV40 NLS

<400> 8

ccaaagaaga agagaaaggt t 21

<210> 9

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> SV40 NLS

<400> 9

cctaagaaga agaggaaggt t 21

<210> 10

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> SV40 NLS

<400> 10

cctaagaaga agcggaaggt t 21

<210> 11

<211> 47

<212> DNA

<213> Artificial Sequence

<220><223> nucleoplasmin NLS

<400> 11

aagagaccag ctgctaccaa gaaggctgga caggctaaga agaagaa 47

<210> 12

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> nucleoplasmin NLS

<400> 12

aagcgtcctg ctgccaccaa aaaggccgga caggctaaga aaaagaag 48