



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2016-0103953
(43) 공개일자 2016년09월02일

- (51) 국제특허분류(Int. Cl.)
C12N 15/10 (2006.01) C12N 15/113 (2010.01)
C12N 9/22 (2006.01) C12Q 1/68 (2006.01)
- (52) CPC특허분류
C12N 15/10 (2013.01)
C12N 15/113 (2013.01)
- (21) 출원번호 10-2016-0022810
(22) 출원일자 2016년02월25일
심사청구일자 2016년02월25일
- (30) 우선권주장
1020150026203 2015년02월25일 대한민국(KR)

- (71) 출원인
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
- (72) 발명자
방두희
서울특별시 동대문구 회기로25가길 22-9 (이문동)
이지원
서울특별시 서초구 방배선행길 1, 106동 603호 (방배동, 방배우성아파트)
- 임현섭
충청남도 아산시 배방읍 호서로 460
- (74) 대리인
특허법인다나

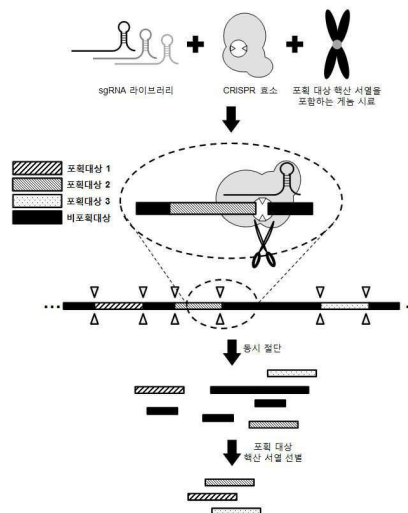
전체 청구항 수 : 총 16 항

(54) 발명의 명칭 CRISPR 시스템을 이용한 다중 위치 염기서열의 동시 포획 방법

(57) 요약

본 발명은 게놈 시퀀싱에 있어서 사용되는 염기서열 포획 방법에 관한 것이다. 본 발명에 따르면 복수 개의 CRISPR 시스템을 사용함으로써 게놈 내 다중 위치의 포획 대상 핵산 서열을 동시에 포획할 수 있게 된다.

대표도 - 도1



(52) CPC특허분류

C12N 9/22 (2013.01)

C12Q 1/6806 (2013.01)

C12Q 2521/301 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 HI13C2163

부처명 보건복지부

연구관리전문기관 한국보건산업진흥원

연구사업명 한국보건산업진흥원-유전체 임상적용 기반기술

연구과제명 종양 유전체 진단의 임상 적용을 위한 유전자 포획 IT-BT-NT 융합 기술

기 여 율 1/2

주관기관 연세대학교 산학협력단

연구기간 2014.12.01 ~ 2015.11.30이 발명을 지원한 국가연구개발사업

과제고유번호 C0217133

부처명 중소기업청

연구관리전문기관 사단법인 한국산학연합회

연구사업명 정부-산업통상자원부-중소기업청-산학협력기술 개발사업-첫걸음기술개발사업

연구과제명 표적 유전자 포획 효율성 향상을 위한 패널 개발

기 여 율 1/2

주관기관 연세대학교 산학협력단

연구기간 2014.08.01 ~ 2015.07.31

명세서

청구범위

청구항 1

게놈 시퀀싱에 있어서,

포획 대상 핵산 서열을 포함하는 게놈 시료에 포획 대상 핵산 서열의 양 말단 위치를 절단할 수 있거나, 또는 포획 대상 핵산 서열 내 타겟 서열에 상보적으로 결합할 수 있는 복수 개의 CRISPR 시스템을 처리하고,

게놈 시료의 단편들 또는 이들의 PCR 증폭산물로부터 포획 대상 핵산 서열을 선별하는 것을 포함하며,

게놈 내 하나 이상의 포획 대상 핵산 서열을 동시에 포획하는 것을 특징으로 하는

포획 대상 핵산 서열의 포획 방법.

청구항 2

제1항에 있어서,

포획 대상 핵산 서열을 포함하는 게놈 시료에 포획 대상 핵산 서열의 양 말단 위치를 절단할 수 있는 복수 개의 CRISPR 시스템을 처리하고,

게놈 시료의 단편들 또는 이들의 PCR 증폭산물로부터 포획 대상 핵산 서열을 선별하는 것을 포함하며,

게놈 내 하나 이상의 포획 대상 핵산 서열을 동시에 포획하는 것을 특징으로 하는

포획 대상 핵산 서열의 포획 방법.

청구항 3

제2항에 있어서,

포획 대상 핵산 서열을 포함하는 핵산 시료에, 포획 대상 핵산 서열의 양 말단 위치를 절단할 수 있는 복수개의 CRISPR 시스템과 함께 추가로 포획 대상 핵산 서열 내의 소정의 위치 중 한 곳 이상을 절단할 수 있는 하나 이상의 CRISPR 시스템을 처리하고,

게놈 시료의 단편들 또는 이들의 PCR 증폭산물로부터 포획 대상 핵산 서열을 선별하는 것을 포함하며,

게놈 내 하나 이상의 포획 대상 핵산 서열을 동시에 포획하는 것을 특징으로 하는

포획 대상 핵산 서열의 포획 방법.

청구항 4

제1항에 있어서,

포획 대상 핵산 서열을 포함하는 게놈 시료에, 포획 대상 핵산 서열 내 타겟 서열에 상보적으로 결합할 수 있는 복수 개의 CRISPR 시스템을 처리하고,

게놈 시료의 단편들 또는 이들의 PCR 증폭산물로부터 포획 대상 핵산 서열을 선별하는 것을 포함하며,

게놈 내 하나 이상의 포획 대상 핵산 서열을 동시에 포획하는 것을 특징으로 하는

포획 대상 핵산 서열의 포획 방법.

청구항 5

제1항 내지 제4항 중 어느 한 항에 있어서,
상기 CRISPR 시스템은
sgRNA 및 CRISPR 효소; 또는
crRNA, tracrRNA 및 CRISPR 효소를 포함하는
포획 대상 핵산 서열의 포획 방법.

청구항 6

제1항 내지 제4항 중 어느 한 항에 있어서,
상기 CRISPR 시스템은
sgRNA 및 CRISPR 효소인
포획 대상 핵산 서열의 포획 방법.

청구항 7

제6항에 있어서,
상기 sgRNA는 주형 DNA로부터 인 비트로 전사에 의해 얻어진 sgRNA 라이브러리인 포획 대상 핵산 서열의 포획 방법.

청구항 8

제7항에 있어서,
상기 주형 DNA는 RNA 폴리머라아제와 결합하여 전사를 시작할 수 있는 프로모터와 sgRNA를 코딩하는 DNA 서열을 포함하는 것인 포획 대상 핵산 서열의 포획 방법.

청구항 9

제5항에 있어서,
상기 CRISPR 효소는 타입 II CRISPR 시스템 효소인 포획 대상 핵산 서열의 포획 방법.

청구항 10

제5항에 있어서,
상기 CRISPR 효소는 Cas9 효소인 포획 대상 핵산 서열의 포획 방법.

청구항 11

제10항에 있어서,
상기 Cas9 효소는 *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Flaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifractor*,

Mycoplasma 및 *Campylobacter*으로 이루어지는 군으로부터 선택되는 미생물 속으로부터 유래한 Cas9의 오소로그(ortholog)인 포획 대상 핵산 서열의 포획 방법.

청구항 12

제1항 내지 제4항 중 어느 한 항에 있어서,
포획 대상 핵산 서열은 DNA, RNA 또는 PNA 인 포획 대상 핵산 서열의 포획 방법.

청구항 13

제1항 내지 제4항 중 어느 한 항에 있어서,
포획 대상 핵산 서열은 동물 또는 식물 유래인 포획 대상 핵산 서열의 포획 방법.

청구항 14

제2항 내지 제4항 중 어느 한 항에 있어서,
상기 CRISPR 효소는 야생형 CRISPR 효소인 포획 대상 핵산 서열의 포획 방법.

청구항 15

제5항에 있어서,
상기 CRISPR 효소는 변이 CRISPR 효소인 포획 대상 핵산 서열의 포획 방법.

청구항 16

제1항 내지 제4항 중 어느 한 항에 있어서,
포획 대상 핵산 서열의 선별은 핵산의 크기에 따른 분리 또는 프로브를 이용하여 수행되는 것인 포획 대상 핵산 서열의 포획 방법.

발명의 설명

기술 분야

[0001] 본 명세서에 개시된 기술은 일반적으로 게놈 시퀀싱에 있어서 사용되는 염기서열 포획 방법에 관한 것이다.

배경 기술

[0002] 일반적으로 게놈 시퀀싱에 있어서 사용되는 염기서열 포획은 다음과 같은 방법들로 수행되고 있다. 첫 번째로 단일가닥 DNA인 올리고뉴클레오타이드를 이용한 선택적 증폭 방법, 두 번째로 제한효소(Restriction Enzyme)를 이용한 염기서열의 절단 방법, 세 번째로 Molecular Inversion Probe(MIP)를 이용한 선택적 증폭의 방법, 그리고 마지막으로 RNA hybridization을 이용한 포획 방법이다.

[0003] 그 중 올리고뉴클레오타이드를 이용한 선택적 증폭 방법은 증폭하고자 하는 서열의 양 끝 부분과 같은 서열을 갖는 프라이머(primer)로 불리는 올리고뉴클레오타이드를 만들어 핵산중합효소(DNA polymerase), dNTP(dATP, dTTP, dCTP, dGTP)와 함께 중합반응을 하여 중간의 포획할 염기서열 부분만 선택적으로 증폭하는 방법이다. 이 방법은 포획 영역의 개수가 적을 때에는 매우 쉽게 이용할 수 있지만 포획 영역의 개수가 많아지면 필요한 올리고뉴클레오타이드의 종류가 많아진다. 이 때 개별 올리고뉴클레오타이드가 상호간섭을 일으켜 모두가 증폭이 잘 되지 않는 단점이 있다. 또한 증폭하고자 하는 영역에 따라 프라이머의 서열이 다르기 때문에 중합 반응 시 프라이머와 포획할 DNA의 결합력도 다르게 된다. 따라서 증폭하고자 하는 영역별로 증폭 효율이 달라져 균등하게 증폭하는 것은 불가능하다.

[0004] 그 다음으로 제한효소를 이용하여 원하는 염기서열을 포획하는 방법은 제한효소가 특정 염기서열만을 인식하여

정해진 서열을 자르는 특성을 이용한 것이다. 따라서 포획하고자 하는 영역에 제한효소가 인식하는 서열이 있으면 포획하고자 하는 부분만을 잘라낼 수 있다. 하지만 이 방법은 포획하려는 서열 부근에 제한효소 인식서열이 없으면 사용할 수 없다. 그리고 두 가지 이상의 제한효소를 사용할 경우 각 효소가 작동용액(working buffer)에 따라 활성이 다르기 때문에 그 용액을 맞춰줘야 한다. 따라서, 이 방법 또한 포획하는 영역이 많을수록 사용하기 어렵다.

[0005] 비교적 최근에 개발된 방법인 MIP을 이용한 선택적 증폭방법은 긴 올리고뉴클레오타이드를 가운데 부분이 뒤집힌 상태로 포획하고자 하는 염기서열 양쪽에 결합을 시키고 그 사이를 증폭하는 방법이다. 이 방법은 기존 방법들과 달리 포획 과정 중에 사용되는 올리고뉴클레오타이드가 수 천, 수 만개의 다른 종류를 이용하더라도 거의 상호 간섭 없이 각각의 염기서열을 포획할 수 있다. 하지만 이 방법 역시 MIP의 결합 서열에 따라 DNA와의 결합력이 달라져 포획하는 영역별로 그 효율이 각각 달라진다. 이에 따라 포획하는 부분에 따른 효율의 차이가 생겨 고르게 포획되지 않는다.

[0006] 마지막으로 RNA hybridization 방법은 DNA-DNA 간 결합보다 DNA-RNA 간 결합이 더 강한 것을 이용하여 미리 biotin을 결합시킨 RNA를 포획하고자 하는 DNA에 결합시킨 후 이 biotin을 이용하여 다시 분리해내는 방법이다. 현재 개발된 방법 중 가장 포획 효율이 좋은 방법이지만 그 포획 과정이 복잡하고 포획 영역이 작을수록 포획이 잘 안 된다.

[0007] 한편, CRISPR 시스템은 원핵 생물, 고세균의 면역 시스템으로서 최근 유전자 가위의 하나로 그 활용성에 대한 연구가 급증하고 있으나(Jinek *et al*, A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity, *Science*, 2012), (Zalatan *et al*, Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds, *Cell*, 2014), 이를 게놈 시퀀싱에 있어서 사용되는 염기서열 포획 방법에 활용하고자 하는 시도는 없었다.

발명의 내용

해결하려는 과제

[0008] 따라서, 본 발명은 게놈 시퀀싱에 있어서 다중 위치의 염기서열을 동시에 효율적으로 포획하기 위한 새로운 방법을 제공하고자 한다.

과제의 해결 수단

[0009] 이에, 본 발명은 CRISPR 시스템을 이용하여 게놈 내 다중 위치 염기서열을 동시에 포획하기 위한 방법을 제공한다.

[0010] CRISPR(Clustered Regularly Interspaced Short Palindromic Repeats) 시스템은 대부분 바이러스 같은 외부 침입자에 대한 저항성을 제공하는 원핵 생물, 고세균의 면역 시스템으로서, 보통 타입 I, 타입 II, 타입 III 및 타입 U 등으로 분류된다.

[0011] 이 중 가장 잘 알려진 타입 II CRISPR 시스템을 예로 들면, CRISPR RNA(crRNA)와 transactivating crRNA(tracrRNA)로 이루어진 RNA 복합체에 CRISPR 효소가 결합한 CRISPR-Cas 복합체가 타겟서열(CRISPR 복합체 결합 서열을 의미한다)을 인식하고, 특정 위치를 절단한다. 이러한 CRISPR-Cas 복합체는 PAM 이라 불리는 특정 서열 앞의 20 bp 정도의 타겟서열을 인식하고 이러한 타겟서열 내부 혹은 인근 서열을 절단하는 것으로 알려져 있다. 또한, crRNA와 tracrRNA를 하나의 형태로 만든 sgRNA(single guide RNA) 역시 crRNA와 tracrRNA의 복합체와 동일한 역할을 하는 것으로 밝혀져, sgRNA와 CRISPR 효소의 복합체가 타겟서열을 절단할 수 있음에 대해서도 잘 알려져 있다.

[0012] 이러한 CRISPR 효소의 여러 도메인 중 염기 절단에 관여하는 두 개의 도메인에 특정한 변이가 도입되면 그 염기 절단 능력을 상실하는 것도 알려져 있다. 그 예로 *Streptococcus pyogenes*의 Cas9 단백질의 경우 아미노산 10번과 840번을 각각 알라닌으로 변이시키면(D10A 및 H840A 변이), DNA 절단 능력을 상실하게 되고 이를 보통 dead Cas9(dCas9)이라고 한다. 또한 10번, 840번 중 어느 하나의 아미노산을 알라닌으로 변이(D10A 또는 H840A 변이)시킨 Cas9 효소는 이중 가닥 염기 중 한 가닥만을 절단하게 된다고 이미 알려져 있다.

[0013] 본 발명자들은 CRISPR 시스템이 타겟서열에 대한 sgRNA의 설계 과정만 거치면 비교적 자유롭게 특정 서열을 절단하거나 상보결합 할 수 있다는 점을 주목하고, 복수 개의 CRISPR 시스템을 이용하면 원하는 영역의 서열을 절

단 혹은 상보결합함으로써 동시에 다중 위치의 핵산서열을 포획(capture)할 수 있음에 착안하였다.

- [0014] 이에 본 발명은,
- [0015] 게놈 시퀀싱에 있어서,
- [0016] 포획 대상 핵산 서열을 포함하는 게놈 시료에
- [0017] 포획 대상 핵산 서열의 양 말단 위치를 절단할 수 있거나, 또는 포획 대상 핵산 서열 내 타겟 서열에 상보적으로 결합할 수 있는 복수 개의 CRISPR 시스템을 처리하고,
- [0018] 게놈 시료의 단편들 또는 이들의 증폭산물로부터 포획 대상 핵산 서열을 선별하는 것을 포함하며,
- [0019] 게놈 내 하나 이상의 포획 대상 핵산 서열을 동시에 포획하는 것을 특징으로 하는
- [0020] 포획 대상 핵산 서열의 포획 방법을 제공한다.
- [0021] 'CRISPR 시스템'은 타입 I, 타입 II, 타입 III 및 타입 U 등 각 유형마다 구성요소가 조금씩 상이하나, CRISPR 효소 및 이와 결합하는 RNA 등을 공통적으로 포함하고 있다.
- [0022] 본 명세서에 있어서, 'CRISPR 시스템'은 CRISPR 효소들 혹은 변이 CRISPR 효소들과 이 효소들과 결합하는 단백질, RNA의 조합, 또는 이들이 작동하기 위해 필요한 추가적 요소를 포함하는 조합을 의미한다.
- [0023] 본 명세서에 있어서, CRISPR 효소는 CRISPR Associated (Cas) 효소라고도 칭한다. 같은 맥락에서, 'CRISPR 시스템'은 'CRISPR 복합체' 또는 'CRISPR-Cas 복합체'와 상호교환적으로 사용된다.
- [0024] CRISPR 시스템에서 사용될 수 있는 CRISPR 효소는 포획 대상 핵산 서열내의 타겟서열과 혼성화되는 crRNA/tracrRNA 복합체 혹은 sgRNA와 복합체를 이루는 효소를 의미한다. CRISPR 효소는 CRISPR 시스템의 유래 생물에 따라 CRISPR 효소 이외의 명칭으로 지칭되기도 한다. 또 CRISPR 효소는 기능적 변이를 가진 효소들이 알려져 있는데 그 예로 leading strand 혹은 lagging strand 중 한 쪽 가닥만 특별히 절단할 수 있는 nickase 기능이 있는 CRISPR 효소와 DNA 절단능력을 상실하여 상보결합만 할 수 있는 dCas 효소들이 있다.
- [0025] 본 발명에 있어서, 상기 'CRISPR 효소'는 '야생형 CRISPR 효소'와 '변이 CRISPR 효소'를 모두 포함하는 개념이다. 야생형 CRISPR 효소는 타겟 서열과 결합하여 타겟 서열 내부 혹은 인근 서열을 절단하는 활성을 갖고 있는 것을 말하며, 반면 변이 CRISPR 효소는 타겟 서열과 결합은 하되 절단능력의 전부 또는 일부가 상실된 효소를 말한다. 본 발명에서는, 야생형 CRISPR 효소의 일례로 'Cas9 효소'를, 그리고 변이 CRISPR 효소의 일례로 'Cas9 효소'를 사용하였다.
- [0026] CRISPR 시스템은 그것이 유래한 종마다 그 타입이 I, II, III, U 등으로 나뉘며 같은 타입에서도 CRISPR 효소의 아미노산 서열이 다르다. 또한 crRNA, tracrRNA의 염기서열도 다르며 그에 따라 이를 융합시킨 sgRNA의 염기서열 역시 달라진다.
- [0027] 당업자는 포획의 효율성 및 정확성 등을 고려하여 다양한 미생물의 CRISPR 시스템 중 적합한 것을 선택하여 사용할 수 있다. 또한, 하나의 생물 내에서 유래한 CRISPR 시스템이 아니라고 하더라도 효율적이고 정확한 포획을 가능하게 하는 CRISPR 시스템의 구동을 가능하게 한다면 다양한 미생물 유래의 crRNA, tracrRNA 혹은 이를 융합시킨 sgRNA 및 CRISPR 효소를 조합하여 사용하는 것도 가능하다. 마찬가지로 포획의 목적과 효율 등에 따라 변이 CRISPR 효소나 변형 RNA들로 구성된 CRISPR 시스템을 사용하는 것도 가능하다.
- [0028] 본 발명은 하나 이상의 포획 대상 핵산 서열에 대해 복수 개의 CRISPR 시스템 혹은 CRISPR 복합체를 이용하여 다중위치의 포획 대상 핵산 서열을 동시에 포획하는 것을 특징으로 한다.
- [0029] 본 발명에 있어서, 포획 대상 핵산 서열을 포획하기 위해 사용하는 CRISPR 시스템은 복수 개의 crRNA/tracrRNA 복합체 세트 또는 복수 개의 sgRNA 혹은 이들의 변형체와 CRISPR 효소를 사용할 수 있다.
- [0030] 본 발명에서 '포획 대상 핵산 서열'은 '타겟 서열'과는 구별되는 용어로 사용된다. '타겟 서열'은 CRISPR 시스템이 인식하여 상보결합하는 특정 서열을 의미하는 반면, '포획 대상 핵산 서열'은 복수개의 CRISPR 복합체를 이용하여 '타겟 서열'의 특정 위치를 절단하거나 타겟 서열의 특정 위치에 상보결합 후 선별에 따라 얻게 되는 핵산 서열을 의미한다.
- [0031] 본 발명에 따른 CRISPR 시스템을 이용한 포획 대상 핵산 서열의 포획 방법은 크게 1) CRISPR 시스템에 의한 핵산 서열의 절단과 2) CRISPR 시스템의 타겟 서열에 대한 상보적 결합에 근거하여 구분된다.

- [0032] CRISPR 시스템에 의한 핵산 서열의 절단에 근거한 포획 대상 핵산 서열의 포획 방법과 관련하여, 본 발명의 한 구체예는
- [0033] 게놈 시퀀싱에 있어서,
- [0034] 포획 대상 핵산 서열을 포함하는 게놈 시료에, 포획 대상 핵산 서열의 양 말단 위치를 절단할 수 있는 복수 개의 CRISPR 시스템을 처리하고,
- [0035] 게놈 시료의 단편들 또는 이들의 PCR 증폭산물로부터 포획 대상 핵산 서열을 선별하는 것을 포함하며,
- [0036] 게놈 내 하나 이상의 포획 대상 핵산 서열을 동시에 포획하는 것을 특징으로 하는
- [0037] 포획 대상 핵산 서열의 포획 방법을 제공한다.
- [0038] 상기 구체예에 대한 이해를 돕기 위해 도 1의 모식도를 예로 들어 설명하면, 도 1은 sgRNA 라이브러리와 CRISPR 효소, 그리고 포획 대상 핵산 서열을 포함하는 게놈 시료를 반응시켜 게놈 시료 내 다중 위치의 핵산 서열을 동시에 절단한 후 원하는 포획 대상 핵산 서열만을 선별하는 과정의 모식도이다. 포획 대상 핵산 서열을 포함하는 게놈 시료에서 포획 대상 핵산 서열만을 선별해내기 위해 sgRNA 라이브러리와 CRISPR 효소를 섞어주면 CRISPR 복합체가 형성되고 이 복합체들은 개별적인 sgRNA가 인식하여 상보적으로 결합하는 타겟 서열 내 특정 서열 위치를 절단한다.
- [0039] 도 2는 두 개의 CRISPR 복합체(I, II)가 폴리뉴클레오타이드 내 특정 서열 두 곳을 절단하는 것을 모식화한 것이다. sgRNA가 상보적으로 결합하는 부위가 CRISPR 복합체의 타겟서열이고, “번개”에 의해 잘리는 위치로 표시된 a와 b 부분이 CRISPR 효소에 의해 타겟서열 내에서 절단되는 특정 서열 위치를 표현한 것이다. 도 2에 있어서, 본 발명에서 언급하는 “포획 대상 핵산 서열”은 타겟서열 내 절단되는 위치 사이의 영역, 즉 도 2 상의 a와 b 사이의 영역을 의미한다.
- [0040] 다른 구체예에서, 본 발명은
- [0041] 포획 대상 핵산 서열을 포함하는 게놈 시료에, 포획 대상 핵산 서열의 양 말단 위치를 절단할 수 있는 복수개의 CRISPR 시스템과 함께 추가로 포획 대상 핵산 서열 내의 소정의 위치 중 한 곳 이상을 절단할 수 있는 하나 이상의 CRISPR 시스템을 처리하고,
- [0042] 게놈 시료의 단편들 또는 이들의 PCR 증폭산물로부터 포획 대상 핵산 서열을 선별하는 것을 포함하며,
- [0043] 게놈 내 하나 이상의 포획 대상 핵산 서열을 동시에 포획하는 것을 특징으로 하는
- [0044] 포획 대상 핵산 서열의 포획 방법을 제공한다.
- [0045] 상기 구체예와 관련하여 도 3을 예로 들어 설명하면, 본 발명에 따른 핵산 서열의 포획 방법은 게놈 시퀀싱을 위해 사용되는데, 이 경우 핵산 서열은 시퀀싱 기기에서 분석하기에 적합한 크기, 예컨대, 300 내지 500 bps 정도로 절단되어 서열분석이 수행된다. 만일 포획 대상 서열이 시퀀싱 기기에 바로 투입하기 적합하지 않은 경우, 예를 들어 포획 대상 서열이 너무 길 경우, 도 3과 같은 형태로 셋 또는 그 이상의 CRISPR 복합체를 사용하여 포획 대상 서열을 포획해 낼 수 있다. 이 경우, 세 개의 CRISPR 복합체(III, IV, V)가 각각 p, q, r 위치를 절단하여 결과적으로 p-r에 이르는 포획 대상 핵산을 획득할 수 있게 된다.
- [0046] 본 발명은 또한 CRISPR 시스템의 타겟 서열에 대한 상보적 결합에 근거한 포획 대상 핵산 서열의 포획 방법을 제공한다.
- [0047] 이와 관련하여 본 발명의 한 구체예는
- [0048] 게놈 시퀀싱에 있어서,
- [0049] 포획 대상 핵산 서열을 포함하는 게놈 시료에, 포획 대상 핵산 서열 내 타겟 서열에 상보적으로 결합할 수 있는 복수 개의 CRISPR 시스템을 처리하고,
- [0050] 게놈 시료의 단편들 또는 이들의 PCR 증폭산물로부터 포획 대상 핵산 서열을 선별하는 것을 포함하며,
- [0051] 게놈 내 하나 이상의 포획 대상 핵산 서열을 동시에 포획하는 것을 특징으로 하는
- [0052] 포획 대상 핵산 서열의 포획 방법을 제공한다.
- [0053] 상기 구체예에 대한 이해를 돕기 위해 도 4 및 도 5의 모식도를 예로 들어 설명하면, 도 4는 변이 Cas 단백질을

포함하는 CRISPR 복합체가 포획 대상 핵산 서열을 포함하는 게놈 시료 내 타겟 서열에 상보적으로 결합하고, 게놈 시료의 단편들로부터 CRISPR 복합체가 상보적으로 결합된 핵산 서열을 선별해 냄으로써, 포획 대상 핵산 서열을 포획하는 방법을 모식화한 것이다. 또한, 도 5는 두 개의 변이 CRISPR 효소를 포함하는 CRISPR 복합체(VI, VII)가 게놈 시료 단편들 중 포획 대상 서열 VI, VII을 포함한 폴리뉴클레오타이드만을 포획해내는 방법을 모식도로 보여준다.

[0054] 도 4 및 도 5와 같은 경우, 변이 CRISPR 효소는 CRISPR 복합체가 타겟 서열을 인식하고 상보적 결합을 형성하긴 하나 절단 능력은 상실되어 있는 상태이기 때문에 타겟 서열 내 특정 위치를 절단하지는 않는다. 그러나, 타겟 서열에 상보적으로 결합되어 있는 CRISPR 복합체를 선별함으로써, 결과적으로 “포획 대상 핵산 서열”을 선별해 낼 수 있게 된다. 이 경우에 있어서, 본 발명에서 언급하는 “포획 대상 핵산 서열”은 이 경우 절단된 DNA 중 타겟서열을 포함하는 도 4 하단의 선별된 염기들을 의미한다.

[0055] 이에 제한되는 것은 아니나, 포획 대상 핵산 서열을 포함하고 있는 게놈 시료에 CRISPR 복합체를 처리하기 전 또는 후에, 음파처리(sonication), 전이인자(Transposon)등의 DNA를 절단할 수 있는 여러 방법을 이용하여 CRISPR-Cas 복합체가 게놈 시료의 핵산서열과 상보적 결합을 하기 전 또는 후에 포획 대상 핵산 서열을 포함하고 있는 게놈 시료를 무작위로 파편화할 수 있다. 이에 제한되는 것은 아니나, 상보적 결합 전 미리 절단하는 방법은 초음파분쇄 등의 방법을 사용할 수 있고 상보적 결합 후 절단하는 방법은 전이인자를 이용한 방법 등을 이용할 수 있다. 도 4는 포획 대상 핵산 서열을 포함하고 있는 게놈 시료에 CRISPR 복합체를 처리하기 전에 게놈 시료를 무작위로 파편화한 CRISPR 복합체를 처리한 경우를 모식화한 것이다.

[0056] 한편, CRISPR 효소 중 대표적인 잘 알려진 것은 *streptococcus pyogenes*에서 유래한 Type II의 Cas9 효소이다. Cas9 효소도 그것이 유래한 생물의 종에 따라 약간씩 다르다. 본 발명에서 다양한 종에서 유래한 CRISPR 효소들과 함께 Cas9 효소의 오소로그(ortholog)들과 그 변이체들을 포함한다. 이러한 Cas9 효소의 예로는 이에 제한되는 것은 아니나, *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Flaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor*, *Mycoplasma* 및 *Campylobacter*으로 이루어지는 군으로부터 선택되는 미생물 속으로부터 유래한 Cas9의 오소로그(ortholog)일 수 있다. 본 발명에 있어서, CRISPR 효소는 야생형이거나 또는 하나 이상의 돌연변이를 포함할 수 있고 이 변이 CRISPR 효소는 이에 제한되는 것은 아니나 DNA의 한쪽 가닥만을 절단하는 nickase CRISPR 효소, DNA절단 능력을 상실한 dead CRISPR 효소 등이 있다.

[0057] 한 구체예에서, 본 발명에서 사용되는 CRISPR 효소는 Cas9 효소일 수 있다.

[0058] 이러한 CRISPR 효소 또는 변이 CRISPR 효소는 당업자에게 공지된 통상의 단백질 합성, 분리, 정제 방법에 의해 획득하여 사용할 수 있다. 그 예로 이에 제한되는 것은 아니나 CRISPR 효소는 대장균에서의 과발현법, 고체상 합성 등의 단백질 제조 방법을 활용하여 생산할 수 있다.

[0059] 또한, 본 발명에 따른 포획 대상 핵산 서열의 포획에 있어 CRISPR 효소 즉, CRISPR 효소 혹은 변이 CRISPR 효소 등이 포함된 CRISPR 시스템이 활성을 가져 절단, 상보결합 등을 하기 위해서는 그에 적합한 '작동 용액'을 사용할 필요가 있다. CRISPR 시스템에 따른 작동 용액의 조건에 대해서는 당업계에 잘 알려져 있다.

[0060] 한편, CRISPR 효소와 결합하여 CRISPR 복합체를 형성하는 crRNA 및 tracrRNA; 또는 sgRNA 혹은 이들의 변이체는 CRISPR 효소의 종류에 따라 결정될 수 있다. CRISPR 복합체가 인식하는 타겟서열은 PAM 이라 불리는 특정 서열의 앞의 약 10개 이상의 서열이다. 이러한 PAM 서열은 CRISPR 복합체가 유래한 생물 종에 따라 그 서열과 길이가 다르며, 이들의 구체적인 서열은 당업계에 잘 알려져 있다(Shah, et al, Protospacer recognition motifs: mixed identities and functional diversity, *RNA biology*, 2013). crRNA와 sgRNA 그리고 이들의 변이체는 타겟서열과 상보결합하는 서열이 존재한다. 이에 다양한 생물의 CRISPR 시스템 중 적합한 것을 선택하면, 그에 따라 PAM 서열도 결정되며, 이러한 PAM 서열과 상호작용하여 타겟서열을 절단 혹은 상보결합 할 수 있는 crRNA, sgRNA 혹은 그 변이체의 서열이 결정된다. 이렇게 결정된 crRNA와 tracrRNA의 세트 또는 sgRNA 혹은 이들의 변이 RNA를 사용하면 된다.

[0061] 한편, tracrRNA와 그 변이 RNA는 crRNA 혹은 변이 crRNA와 CRISPR 효소를 연결해 주는 역할을 한다. 또한 tracrRNA와 그 변이 RNA는 crRNA 혹은 변이 crRNA와 변이 CRISPR 효소를 연결해 주는 역할을 할 수도 있다. 이러한 tracrRNA와 그 변이체에 대한 서열 정보는 CRISPR 복합체가 유래한 생물 종에 따라 다르며 일부 알려져 있다.

- [0062] 또한, crRNA와 tracrRNA를 하나의 서열로 만들어 놓은 sgRNA는 타겟서열과 crRNA, 그리고 tracrRNA의 역할을 하는 스캐폴드 영역을 포함한다. CRISPR 복합체의 유래에 따른 이러한 스캐폴드 영역에 대한 정보 역시 대부분 공지되어 있으므로, 당업자는 적합한 서열 정보를 선택하여 sgRNA 혹은 이로부터 유래한 변이체를 이용할 수 있다.
- [0063] 본 발명에 따른 다중 위치 염기서열의 동시 포획 방법은 한 개부터 수 개, 수십개, 수백개, 수천개, 수 만개, 수십만개, 수백만개 또는 그 이상의 포획 대상 서열을 동시적으로 포획할 수 있다. 이를 위해 본 발명에서는 다양한 포획 대상 서열에 대한 개별적인 crRNA와 sgRNA 혹은 그 변이체를 포함하고 있는 RNA 풀을 사용할 수 있다.
- [0064] 이에 제한되는 것은 아니나, 한 구체예에서, sgRNA는 주형 DNA로부터 인 비트로 전사에 의해 얻을 수 있다. sgRNA를 얻기 위해 사용한 주형 DNA는 RNA 폴리머라아제와 결합하여 전사를 시작할 수 있는 프로모터와 sgRNA를 전사하는 DNA 서열, 즉, 타겟서열 및 sgRNA 스캐폴드를 포함한다. 프로모터, sgRNA 스캐폴드는 sgRNA 풀 내에 포함되는 모든 sgRNA에 대해 공통적인 사항이므로, 타겟서열만 달리하여 주형 DNA를 합성하면 된다.
- [0065] 이에 제한되는 것은 아니나, 예를 들어, 주형 DNA는 마이크로어레이 올리고뉴클레오티드 합성 방법으로 제작할 수 있다. 구체적으로, 얻고자 하는 sgRNA 라이브러리에 상응하는 주형 DNA의 라이브러리를 마이크로어레이 올리고뉴클레오티드 합성 방법으로 마이크로칩상에 고정하여 합성한 후 이를 절단하여 얻을 수 있다. 이렇게 합성된 주형 DNA로부터 체외 전사를 통해 sgRNA 라이브러리를 획득한다.
- [0066] 도 1은 다양한 crRNA 혹은 sgRNA 라이브러리를 구성한 후 포획 대상 핵산을 포함하고 있는 핵산 시료에 CRISPR 효소와 함께 처리하고 인큐베이션함으로써 이들이 형성하는 CRISPR 복합체가 포획 대상 핵산을 절단하여 포획하는 과정을 모식도로 보여준다.
- [0067] 도2에서는 다양한 crRNA 혹은 sgRNA 라이브러리를 구성한 후 포획 대상 핵산을 포함하고 있는 핵산 시료에 변이 CRISPR 효소와 함께 처리하고 인큐베이션함으로써 이들이 형성하는 CRISPR 복합체가 포획 대상 핵산과 상보결합하여 포획하는 과정을 모식도로 보여준다.
- [0068] 본 발명을 적용하여 특정 핵산 서열을 포획함에 있어서 포획 대상 핵산 서열은 이에 제한되는 것은 아니나, DNA, RNA 또는 PNA 등 그 종류가 제한되지 않으며 또한 이에 제한되는 것은 아니나, 동물, 식물, 균류 등 그 유래가 특별히 제한되지 않는다.
- [0069] 앞서 설명한 바와 같이, CRISPR 시스템의 특정 서열 절단을 이용하고자 하는 경우 상기 CRISPR 효소는 야생형 CRISPR 효소일 수 있다. 반대로, CRISPR 시스템의 절단능을 배제하고 타겟 서열에 대한 상보적 결합만을 이용하고자 하는 경우, 상기 CRISPR 효소는 변이 CRISPR 효소일 수 있다.
- [0070] 또한, 본 발명에 따른 포획 대상 핵산 서열의 포획 방법은 게놈 시료의 단편들 또는 이들의 PCR 증폭산물로부터 포획 대상 핵산 서열을 선별하는 단계를 포함한다.
- [0071] 포획 대상 핵산 서열이 존재하는 풀(pool)은 게놈 시료의 단편들 또는 이들의 PCR 증폭산물이다. 포획 대상 핵산 서열의 증폭을 위해서는 게놈 시료의 단편들을 PCR을 통해 증폭하는 과정을 거치는 것이 바람직할 수 있다.
- [0072] 포획 대상 핵산 서열의 선별은 이에 제한되는 것은 아니나 크기에 따른 분리 또는 프로브에 의한 분리를 통해 수행될 수 있다. 크기에 따른 분리 방법은 아가로스 젤 분리법과 같이 염기를 크기에 따라 분리할 수 있는 통상의 방법들을 활용할 수 있으며, 이렇게 분리된 핵산 서열들은 PCR 또는 리가아제(ligase) 등의 공지된 방법을 이용하여 어댑터 서열을 결합한 후 시퀀싱 함으로써 정확한 포획 여부를 확인 할 수 있다.
- [0073] 프로브에 의한 분리 방법의 첫번째 예로는 프로브가 연결된 crRNA, tracrRNA, sgRNA 혹은 그 변이체를 만든 후 이 프로브를 이용하여 정제하는 방법이 있다. 그 예로 이에 제한되는 것은 아니나, biotin 표지자가 결합된 RNA와 CRISPR 효소나 변이 CRISPR 효소들로 CRISPR 복합체를 만든 후 포획에 이용하면 절단 혹은 상보결합이 일어난 후 많은 수의 복합체가 타겟서열과 결합한 채로 남아있는데 이 때 RNA의 프로브인 biotin을 이용하여 자성 비드인 streptavidin으로 프로브가 있는 RNA만을 골라내면 그와 함께 복합체가 선별되어 포획 대상 서열만을 포획 할 수 있다.
- [0074] 또 다른 방법으로는 프로브가 연결된 CRISPR 효소 혹은 변이 CRISPR 효소를 이용하여 이 프로브를 이용하여 정제하는 방법이 있다. 그 예로 이에 제한되는 것은 아니나, 히스티딘 태그를 CRISPR 효소들과 그 변이체들에 달아주어 CRISPR 복합체가 포획 대상 서열에 절단, 상보결합을 일으킨 후 이 표지자를 Ni-NTA 비드를 이용하여 선별하는 방법이 있다. Ni-NTA에 붙은 효소들만을 골라내면 그와 함께 CRISPR 복합체가 선별되어 포획 대상 서열

만을 포획할 수 있다. 이 때 Ni-NTA는 Ni-NTA 아가로스 비드, Ni-NTA 자성 비드 이외에도 표지지가 결합할 수 있는 모든 종류가 가능하다.

[0075] 프로브를 이용한 분리 방법의 경우, CRISPR 복합체와 상보적으로 결합된 핵산 서열이 선별되면 CRISPR 복합체와 상보적으로 결합된 핵산 서열을 해리시키는 과정이 뒤따르게 된다. 이에 제한되는 것은 아니나, 예를 들어, CRISPR 복합체와 상보적으로 결합된 핵산 서열이 담긴 용액에 0.2% Sodium Dodecyl Sulfate 용액을 첨가하여 반응시키면 CRISPR 효소의 기능이 제거되어 결합된 핵산 서열이 해리되는 방법을 이용할 수 있다.

[0076] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 것이며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다.

발명의 효과

[0077] 본 발명에 따르면 복수 개의 CRISPR 시스템을 사용함으로써 게놈 내 다중 위치의 포획 대상 핵산 서열을 동시에 포획할 수 있게 된다.

도면의 간단한 설명

[0078] 도 1은 crRNA/tracrRNA 혹은 sgRNA 라이브러리와 CRISPR 효소 포획 대상 핵산 서열을 포함하는 핵산을 반응시켜 다중 위치 DNA를 동시에 절단한 후 원하는 포획 대상 핵산 서열만을 선별하는 과정의 모식도이다.

도 2는 두 개의 CRISPR 복합체(I, II)가 폴리뉴클레오타이드의 특정 서열 중 두 곳(a, b 위치)을 절단하여 포획 대상 서열(a-b 사이의 서열)을 포획해 내는 것 나타낸 모식도이다.

도 3은 세 개의 CRISPR 복합체(III, IV, V)가 폴리뉴클레오타이드의 특정 서열 중 세 곳(p, q, r 위치)을 절단하여 포획 대상 서열(p-r 사이의 서열)하는 것을 나타낸 모식도이다.

도 4는 crRNA/tracrRNA 혹은 sgRNA 라이브러리와 CRISPR 효소 포획 대상 핵산 서열을 포함하는 핵산을 반응시켜 다중 위치 DNA에 동시에 상보결합 시킨 후 결합된 포획 대상 핵산 서열만을 선별하는 과정의 모식도이다.

도 5는 두 개의 변이 CRISPR 효소를 포함하는 CRISPR 복합체(VI, VII)가 게놈 시료 단편들 중 포획 대상 서열 VI, VII을 포함한 폴리뉴클레오타이드만을 포획해내는 방법을 모식도로 보여준다.

발명을 실시하기 위한 구체적인 내용

[0079] [실시예]

[0080] I. 절단을 통한 다중위치 염기서열 포획

[0081] 제조예 I-1: 다중 위치 염기서열 포획을 위한 sgRNA 설계 및 제조

[0082] 본 발명에서 CRISPR 복합체를 이용하여 포획 대상 핵산 서열을 절단하는데 사용한 모든 RNA는 포획하고자 하는 영역의 염기 PAM 서열 앞의 18bp를 인식하도록 설계되어 있다. 본 실시예에서는 PAM 서열로서 'NGG' (N = A, T, C, G 중 한 개)가 사용되었으며, 이 NGG 서열은 *streptococcus pyogenes*가 특이적으로 인식하는 PAM 서열로 GG 염기 앞에 A, T, C, G 중 임의의 염기 한 개가 오면 된다.

[0083] 이렇게 결합부분이 설계된 sgRNA는 주형(template) DNA로부터 채취(*in vitro*) 전사시켜서 획득하였고 이를 위해 주형 DNA에는 T7 RNA polymerase와 결합하여 전사를 시작할 수 있는 T7 프로모터와 sgRNA 주형 서열을 결합시켜서 전사시켰다. 이 때 사용한 T7 프로모터는 'GGATTCTAATACGACTCACTATAGG'(서열번호 1)서열을 가지며, sgRNA 주형 서열 중 포획 대상 핵산과 결합하는 18bp의 서열을 제외한 나머지 sgRNA scaffold는 다음과 같은 서열을 가진다: 'GTTTGTAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGTGCTTTT'(서열번호 3)

[0084] 서열번호 1의 T7 프로모터 서열과 서열번호 3의 sgRNA scaffold 사이에는 'NNNNNNNNNNNNNNNNNN'(N = A, T, C, G 중 한 개)(서열번호 2)에 해당하는 18bp의 타겟서열이 위치한다. 이러한 타겟서열은 절단하고자 하는 염기서열의 위치에 따라 달라진다.

- [0085] 결과적으로, 합성되는 주형 DNA의 서열은 T7 프로모터와 타겟서열 및 sgRNA 주형 서열이 순차적으로 결합된 서열번호 4와 같다.
- [0086] 'GGATTCTAATACGACTCACTATAGGNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA GTGGCACCAGTCGGTGCCTTTTTT'(서열번호 4)
- [0087] 원하는 영역을 각각 타겟하는 sgRNA를 만들기 위해 주형 DNA 라이브러리로 체외전사를 수행하였다. 전사시킨 RNA들을 LiCl를 이용하여 침전시켰고 원심분리(13000rpm, 5min, 4℃)를 통해 침전으로 만들었다. 이 침전을 70% 에탄올로 세척해준 후 다시 한번 원심분리(13000rpm, 5min, 4℃)하여 가라앉혔다. 이후 에탄올을 완벽하게 말려 준 뒤 핵산분해효소가 없는 물에 녹여 보관하였다. 포획 능력 확인 시에는 500nmol 농도로 사용하였고 다중서열 동시포획 시에는 3 μg을 사용하였는데 포획 직전 sgRNA가 들어있는 용액의 온도를 95℃까지 올렸다가 37℃까지 1초당 0.1℃씩 내려서 리폴딩(re-folding)시킨 후 사용하였다.
- [0088] 위와 같은 과정을 거쳐 합성된 sgRNA 풀 내에 포함되어 있는 sgRNA 중 일부를 예로 들어 설명하면 다음과 같다.
- [0089] 제조예 I-1-1: 염색체 1번의 1448014-1448256 위치를 포획하기 위한 두 개의 sgRNA 합성
- [0090] 염색체 1번의 1448014-1448256 위치(서열번호 5)를 포획하기 위해, 앞쪽으로는 1448011-1448028 위치인 'GGAGGATCGGACTCTTTC'(서열번호 6)을 인식하는 sgRNA인 'GAAAGAGTCCGATCCTCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCCTTTTTT T'(서열번호 7)와 뒤쪽으로는 1448254-1448271 위치인 'CGTAACAAGGAAGCGTA'(서열번호 8)를 인식하는 sgRNA인 'TACGCTTCCCTTGTTACGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCCTTTTTT T'(서열번호 9)를 합성하였다.
- [0091] 제조예 I-1-2: 염색체 1번의 55537908-55538174 위치를 포획하기 위한 두 개의 sgRNA의 합성
- [0092] 염색체 1번의 55537908-55538174 위치(서열번호 10)를 포획하기 위해, 앞 부분에 55537893-55537910 위치인 'TCATACCTCTCTTCTCAG'(서열번호 11)을 인식하는 sgRNA인 'TCATACCTCTCTTCTCAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCCTTTTTT T'(서열번호 12)와 뒷부분에 55538160-55538177 위치인 'TTAAAAGCATCCCAAGTA'(서열번호 13)를 인식하는 sgRNA인 'TTAAAAGCATCCCAAGTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCCTTTTTT T'(서열번호 14)를 합성하였다.
- [0093] 제조예 I-1-3: 염색체 10번의 38406959-38407462 위치를 포획하기 위한 세 개의 sgRNA 합성
- [0094] 염색체 10번의 38406959-38407462 위치(서열번호 15)를 포획하기 위해, 38406946-38406963 위치인 'TCAGAGAACACACACAGG'(서열번호 16)을 인식하는 sgRNA인 'TCAGAGAACACACACAGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCCTTTTTT T'(서열번호 17)와 중간 부분에 38407195-38407212 위치인 'GCATCAGAAAACACACAC'(서열번호 18)을 인식하는 sgRNA 'GCATCAGAAAACACACACGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCCTTTTTT T'(서열번호 19)와 뒷부분에 38407447-38407464 위치인 'ACATCTGAGAAGACACAC'(서열번호 20)을 인식하는 sgRNA인 'ACATCTGAGAAGACACACGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCCTTTTTT T'(서열번호 21)을 합성하였다.
- [0095] 제조예 I-1-4: 염색체 12번의 9580101-9580360 위치를 포획하기 위한 두 개의 sgRNA 합성
- [0096] 염색체 12번의 9580101-9580360 위치(서열번호 22)를 포획하기 위해, 앞 부분에 9580087-9580104 위치인 'ACAGGCGTGTTCGTTAA'(서열번호 23)을 인식하는 sgRNA인 'ACAGGCGTGTTCGTTAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCCTTTTTT T'(서열번호 24)와 뒷부분에 9580357-9580374 위치인 'AGGGTTAAGCTCGGAAGT'(서열번호 25)을 인식하는 sgRNA인 'ACTTCCGAGCTTAACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCCTTTTTT T'(서열번호 26)를 합성하였다.
- [0097] 제조예 I-2: 다중 위치 염기서열 포획을 위한 Cas9 단백질의 제조

- [0098] *Streptococcus pyogenes*의 Cas9 유전자를 대장균 단백질 발현 벡터에 일종인 pET28a 벡터에 삽입하였다. 이 때 벡터의 서열 중 단백질 발현에 관련된 부분은 T7 프로모터, Cas9 그리고 정제를 위한 히스티딘-태그를 발현시키는 DNA 서열로 이루어져있다. 이 벡터는 T7 RNA 중합효소와 lac repressor에 의해 발현이 조절되며 일반적으로 T7 RNA 중합효소가 존재해야만 발현이 되고 Isopropyl beta-D-1-thiogalactopyranoside(IPTG)와 함께 배양하면 그 발현량이 매우 증가하는 벡터이다. 이렇게 제작한 벡터를 T7 RNA 중합효소를 가지고 있는 대장균(T7 Express Competent *E. coli*, NEB 사)에 도입하여 Cas9 단백질을 과발현 시킨 후 따로 정제하였다.
- [0099] Cas9 단백질 정제 시 먼저 단백질을 과발현 시킨 대장균을 원심분리(3900rpm, 10min)로 모은 후 세포 배양액을 모두 버린다. 그 후 세포 용해액(lysis buffer, 20mM Tris-HCl at pH8.0, 300mM NaCl, 1mg/mL lysozyme, 1x Phenylmethylsulfonyl fluoride(PMSF))을 1mL 세포용해액/100mL 세포 배양액의 비율로 넣고 대장균을 재부유(resuspension)시켜서 초음파(40% 진폭(amplitude)로 10초 파쇄, 30초 휴식을 한 주기로 총 10분)로 분쇄하였다. 이렇게 파쇄한 용액을 원심분리하여(13000rpm, 10min) 상층액만을 얻어낸 후 Ni-NTA 아가로스 레진(resin)에 통과시켜 히스티딘태그를 가지고 있는 단백질을 레진 상에 남겨두었다. 이후 이 레진에 비정상적으로 결합하여있는 원하지 않는 단백질들을 없애주기 위해 세척을 세 차례 5mL씩 세척용액(washing buffer, 20mM Tris-HCl at pH8.0, 300mM NaCl, 20mM imidazole 1x Phenylmethylsulfonyl fluoride(PMSF))으로 레진을 세척하였다. 그런 후 단백질을 다시 얻어내기 위하여 용리용액(elution buffer, 20mM Tris-HCl at pH8.0, 300mM NaCl, 250mM imidazole, 1x Phenylmethylsulfonyl fluoride(PMSF)) 500를 8번 레진에 통과시켜 원하는 단백질을 선별해내었다.
- [0100] 이렇게 정제한 단백질을 염기서열 포획에 이용하기 위해서는 먼저 단백질이 작동하는 작동용액(working buffer, 50mM Tris-HCl at pH 8.0, 200mM KCl, 0.1mM EDTA, 1mM DTT, 0.5mM PMSF, 20% glycerol)로 용액을 교체해줘야 한다. 이는 용리용액에 많이 들어있는 이미다졸을 제거하는 동시에 단백질이 더 안정한 상태로 유지될 수 있는 용액으로 옮겨주는 과정으로 투석(dialysis) 방법을 이용하였다. 여덟 번에 나눠 용리한 용액 중 단백질이 용리된 세 개의 용액 총 1.5mL를 투석 카세트에 넣어준 후 1L의 작동용액으로 4℃에서 16시간 투석해 주었다. 용액 조성을 바꾼 단백질들은 Bradford assay로 정량하였다.
- [0101] **제조예 I-3: 다중 위치 염기서열 포획을 위한 대상 핵산 서열의 정제**
- [0102] 다중 위치 염기서열 포획을 위한 대상 핵산 서열은 인간 배아 신장세포 293(HEK293)을 배양한 후 정제하였다. 배양 조건은 37℃, 5% CO₂ 배양기에서 우태혈청이 10% 함유된 Dulbecco Modified Eagle Medium을 배양액으로 배양하였다. 배양한 세포는 배양접시에 붙은 채로 자라는데 이를 Trypsin/EDTA 용액으로 떼어낸 후 원심분리(3000rpm, 10min)으로 세포만을 모았다. 그 후 QIAGEN사의 DNeasy 96 Blood & Tissue Kit를 이용하여 게놈만을 정제하였다.
- [0103] **실험예 I-1: Cas9 단백질의 포획 능력 확인**
- [0104] 정제한 단백질의 포획능력을 확인하기 위하여 먼저 1080bp 크기의 이중가닥 DNA를 pUC19 벡터에서 증폭하여 그 중간을 자르는 실험을 하였다. 1080base pair 크기의 절단용 DNA는 절단 시 약 630과 450bp로 절단되는데 이를 실험하기 위하여 위에서 명시한 농도의 Cas9 단백질과 sgRNA 그리고 300ng의 절단용 DNA를 넣고 완충용액 (20μl에서의 최종 농도: 50mM Tris-HCl, 100mM NaCl, 10mM MgCl₂, 1mM DTT, pH 7.9)과 물을 함께 넣어주어 총 20μl 용량을 만들었다. 이외에도 Cas9 단백질을 과량 넣은 용액, sgRNA를 과량 넣은 혼합용액을 37℃에서 1, 8, 16시간 동안 각각 반응시켜 그 절단능력을 확인하였다. 그 결과 500nmol의 sgRNA는 충분량이며 반응은 Cas9 단백질의 양이 가장 중요하다는 것을 알 수 있다. 또한 대부분의 절단 반응은 1시간 내에 일어나는 것을 알 수 있다.
- [0105] **실시예 I-1: 포획대상 핵산 서열의 양 말단 절단을 이용한 다중위치 염기서열 동시 포획**
- [0106] 제조예 I-1을 통해 제작한 sgRNA 라이브러리 1000ng과 함께 기 언급한 Cas9 작동용액 조건에서 제조예 I-2를 통해 제조한 Cas9 단백질을 3000ng 함께 넣고 실험예 1의 최종 농도를 갖도록 20 μL 부피에 맞춘 용액을 37℃에서 한 시간 반응을 진행하여 다중위치 염기서열의 동시 포획을 수행하였다.
- [0107] 다중위치 염기서열 동시 포획이 이루어졌는지 확인하기 위해 포획한 서열의 시퀀싱을 수행하였다. 구체적으로, 반응 진행 후 반응 용액 전체를 QIAGEN 사의 MinElute PCR Purification kit를 이용하여 정제하였고 다음 과정에 바로 Illumina 사의 차세대 시퀀싱 기계를 이용하기 위한 어댑터 DNA 서열을 Enzymatics 사의 SPARK DNA sample prep kit를 이용하여 붙였다. 어댑터를 붙인 DNA 조각들은 USER 효소로 어댑터 DNA에 있는 uracil을 잘라준 뒤 Illumina 사에서 판매하는 Index 서열과 Universal 서열 프라이머를 이용하여 포획한 서열들을 증폭하였다. 증폭한 서열들을 아가로스 젤로 크기 별로 구분하였고 이 때 원하는 크기인 300-500bp 영역만 선별하여

QIAGEN 사의 QIAquick Gel Extraction Kit를 이용하여 정제한 후 Illumina 사의 Hiseq 2500 차세대 시퀀싱 기계를 이용하여 서열 정보를 얻어내었다.

[0108] 얻어낸 서열 정보를 자체 제작한 Python 프로그램과 BWA 등의 프로그램을 이용하여 원하는 서열이 포획되었는지 분석하였으며 원하는 염기 서열들이 동시에 포획된 것을 확인하였다.

[0109] 그 일부를 예로 들면, 시퀀싱 결과 중 'GGATCGGACTCTTCCGTCACCCGTTTGCACCTCTGCAGCTGTCAGGAGCGGGTCAGGTGCGGAAAGCGGTGCGGAGGTGGCGCTCATAGGTTACAGGGTCAGGGTCTGGGGCTGGCCGTGGTCTTCAGTTACCGCCGAGCGTGCGGGAT'(서열번호 27)과 'TACAGGGGTGAGGTCTGGGGCTGGCCGTGGTCTTCAGTTACCGCCGAGCGTGCGGGATCCTTCTGCGCTTGCCGCTCCACGTGGCACAGGCCAAGGCGTGGCCAGATGGGTAGATGGGTTTGTGGTGGTTGCTAGCAGTTTCCACGT'(서열번호 28)의 두 시퀀싱 결과로부터 제조예 I-1-1의 서열번호 7 및 서열번호 9의 두 개의 sgRNA에 의해, 전체 유전체 중 1번 염색체의 1448014-1448256 위치인 서열번호 5의 염기서열이 포획되었음을 확인하였다.

[0110] 또한, 'CAGAGGTTGCAGTTTCTGAGAAACACACTGAAAATCCTCCATAAGTGATTTAGACCACGCAAAAACAAGAGACAACCTCTCACCTGAGCTGAAATGGTTTCGCTGAAAGGTTTTTCCAGTTGATGTTTCATTAGAGACATTACTCTGTGGTGT'(서열번호 29)와 'GTTGATGTTTCATTAGAGACATTACTCTGTGGTGTCCAGTAATGTTCTGACATCTGAGATGAAAGGTCAAAAATGCCATCAGAGGTGACAAAATAGCCCCATGGGTTACAGTTTCTACCATTAGATATTGAGTCTTAAAGCATCCCCAA'(서열번호 30)의 두 가지 시퀀싱 결과로부터 제조예 I-1-2의 서열번호 12 및 서열번호 14의 두 개의 sgRNA에 의해, 염색체 1번의 55537908-55538174 위치(서열번호 10)가 정확하게 포획된 것을 확인하였다.

[0111] 서열번호 17, 19 및 21의 3개의 sgRNA에 의해 포획하고자 했던 염색체 10번의 38406959-38407462 위치(서열번호 15) 또한 'AGGGGGAAAACCTATGAATGTCATGAATGTGGGAAGACCTTCTATAAGAATTCAGACCTATTAAACATCAAAGAATTCATACAGGGGAGAGACCTTATGATGTCATGAATGTGGGAAATCCTTCAGTGAAGTCAACCCCTTACTCAA'(서열번호 31), 'TGGATGTCATGAATGTGGGAAATCCTTCAGTGAAGTCAACCCCTTACTCAACATCAAAGAACGCACACAGGGGAGAAACCATATGAATGTCATGAATGTGGAAAACCTTCTCATTTAAGTCAGTCTTACTGTGCATCAGAAAACACAC'(서열번호 32), 'ACAGGGGAGAAGCCCTATGAATGCTATGCATGTGGGAAAGCCTTCTCAGAAAATCAGACCTCATTAAACATCAAAGAATACACACAGGTGAAAAACCTTATGAATGAATGAATGTGGGAAGTCATTCTCTGAGAAGTCAACCCCTTACTA'(서열번호 33), 'ATGAATGAATGAATGTGGGAAGTCATTCTCTGAGAAGTCAACCCCTTACTAAACATCTAAGAACTCACACAGGTGAGAAACCTTATGAATGTATTTCAGTGTGGAAAATTTTCTGCTACTACTCCGGTTTCAGAACATCTGAGAAGACA'(서열번호 34) 이렇게 네 가지 시퀀싱 결과에 의해 해당 영역이 정확하게 포획된 것을 확인하였다.

[0112] 또 다른 포획 영역인 염색체 12번의 9580101-9580360(서열번호 22) 위치의 경우 'TAAGGGTTAAGTAATTACACATCTGTTTGTCTTTTCTCCTTCTATAGTCTTAACATAGTACTCTACCCACAGGTGGTGACAGGAAGGAAATTTGGATGTGCAATGTGGAAGGTGGAACCTTACCTTGAACAGGTGATGTTGTGCGAT'(서열번호 35)와 'GGAAGGTGGAACCTTACCTTGAACAGGTGATGTTGTGCGATCTGGCTCTGGAAGAGAAAGTCGTTGATAGTCTTTCAGTCCATCCCTGAGAACAAACACATGAAGGGCCTTGGGAGCTTACCCTAAGCCTCAGGTTTCAGTCCCAGG'(서열번호 36) 두 개의 시퀀싱 결과로부터 원하는 영역이 포획된 것을 확인하였고 12번 염색체의 9580202번 염기가 레퍼런스로 사용한 human genome 19의 염기서열과 실험에 이용한 HEK293T의 유전체 사이의 차이(G→C)도 발견할 수 있었다.

[0113] 이러한 결과와 같이 우리가 원하는 다양한 염기서열을 동시에 포획하는데 성공하였다.

[0114] II. 상보결합을 통한 다중위치 염기서열의 포획

[0115] 제조예 II-1: 다중 위치 염기서열 포획을 위한 sgRNA 설계 및 제조

[0116] 본 발명에서 변이 CRISPR 효소를 가진 CRISPR 복합체를 이용하여 포획 대상 핵산 서열과 상보결합 하는데 사용한 모든 RNA는 포획하고자 하는 영역의 염기 PAM 서열 앞의 20bp를 인식하도록 설계되어 있다. 본 실시예에서는 PAM 서열로서 'NGG'(N = A, T, C, G 중 한 개)가 사용되었으며, 이 NGG 서열은 *streptococcus pyogenes*가 특이적으로 인식하는 PAM 서열로 GG 염기 앞에 A, T, G, G 중 임의의 염기 한 개가 오면 된다.

[0117] 이렇게 결합부분이 설계된 sgRNA는 주형(template) DNA로부터 체외(*in vitro*) 전사시켜서 획득하였고 이를 위해 주형 DNA에는 T7 RNA polymerase와 결합하여 전사를 시작할 수 있는 T7 프로모터와 sgRNA 주형 서열을 결합시켜서 전사시켰다. 이 때 사용한 T7 프로모터는 'GGATTCTAATACGACTCACTATAGG'(서열번호 1)서열을 가지며, sgRNA 주형 서열 중 포획 대상 핵산과 결합하는 20bp의 서열을 제외한 나머지 sgRNA scaffold는 다음과 같은 서열을

가진다: 'GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCCTTTTTT'(서열번호 3)

[0118] 서열번호 1의 T7 프로모터 서열과 서열번호 3의 sgRNA scaffold 사이에는 'NNNNNNNNNNNNNNNNNNNN'(N = A, T, C, G 중 한 개)(서열번호 37)에 해당하는 20bp의 타겟 서열이 위치한다. 이러한 타겟 서열은 절단하고자 하는 염기서열의 위치에 따라 달라진다.

[0119] 결과적으로, 합성되는 주형 DNA의 서열은 T7 프로모터와 타겟 서열 및 sgRNA 주형 서열이 순차적으로 결합된 서열번호 38와 같다.

[0120] 'GGATTCTAATACGACTCACTATAGGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAGTGGCACCGAGTCGGTGCCTTTTTT'(서열번호 38)

[0121] 원하는 영역을 각각 타겟하는 sgRNA를 만들기 위해 주형 DNA 라이브러리로부터 *in vitro* transcription을 수행하였다. 전사시킨 RNA들을 Ambion사의 Turbo DNase를 이용하여 주형 DNA를 제거하였다. 이후 Zymoresearch사의 Oligo Clean & Concentrator를 이용하여 RNA만 선택적으로 filter에 모았다. 이후 같은 kit에 포함된 washing buffer를 이용해 세척한 후 핵산분해효소가 없는 물에 녹여 보관하였다. 다중서열 동시포획 시에는 480.7ng를 사용하였는데 포획 직전 sgRNA가 들어있는 용액의 온도를 95℃까지 올렸다가 37℃까지 1초당 0.1℃씩 내려서 리폴딩(re-folding)시킨 후 사용하였다.

[0122] 위와 같은 과정을 거쳐 합성된 sgRNA 풀 내에 포함되어 있는 sgRNA 중 일부를 예로 들어 설명하면 다음과 같다.

[0123] 제조예 II-1-1: bla 유전자와 상보결합 하기 위한 열 한 개의 sgRNA 합성

[0124] EcNR2 대장균 게놈의 bla 유전자(서열번호 39, ATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAAGCTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTACCAAGTCACAGAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGACAACATGGGGGATCATGTAACCTCGCCTTGATCGTTGGGAACCGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACCTGGCGAACTACTTACTCTAGCTTCCCGGCAACAA TTAATAGACTGGATGGAGGCGGATAAAGTTGAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTTGGTAA)를 포획하기 위해, bla 유전자의 인근 영역(서열번호 40, GAAAGCGAGATACCAAGGTAGTCGGCAATAATGTCTAACAATTCGTTCAAGCCGACGGATATCGAGCTCGCTTGGACTCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTGGATTGTTTCAAGC)까지 범위를 넓혀서 bla 유전자가 충분히 포획되도록 하고 이 영역에서 CRISPR 복합체가 결합하도록 열 한 개의 서열을 합성하였다. 817623-817642 위치의 AAACAACCTAAATGTGAAAG(서열번호 41)를 인식하는 sgRNA인 AAACAACCTAAATGTGAAAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCCTTTTTT(서열번호 42), 817708-817727 위치의 TGCTTCAATAATATTGAAAA(서열번호 43)를 인식하는 sgRNA인 TGCTTCAATAATATTGAAAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCCTTTTTT(서열번호 44), 817799-817818 위치의 TTTTGCTCACCAGAAACGC(서열번호 45)를 인식하는 sgRNA인 TTTTGCTCACCAGAAACGCCTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCCTTTTTT(서열번호 46), 817916-817935 위치의 CGAAGAAGCTTTTCCATGA(서열번호 47)를 인식하는 sgRNA인 CGAAGAAGCTTTTCCAATGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCCTTTTTT(서열번호 48), 818012-818031 위치의 CATACACTATTCTCAGAATG(서열번호 49)를 인식하는 sgRNA인 CATACACTATTCTCAGAATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCCTTTTTT(서열번호 50), 818110-818129 위치의 TAACCATGAGTGATAA(서열번호 51)를 인식하는 sgRNA인 TAACCATGAGTGATAA(서열번호 52), 818216-818235 위치의 TGATCGTTGGGAACCGAGC(서열번호 53)를 인식하는 sgRNA인 TGATCGTTGGGAACCGAGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCCTTTTTT(서열번호 54), 818295-818314 위치의 ACGTTGCGCAAACTATTAAC(서열번호 55)를 인식하는 sgRNA인 ACGTTGCGCAAACTATTAACGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCCTTTTTT(서열번호 56), 818409-818428 위치의 GCTGGCTGGTTATTGCTGA(서열번호 57)를 인식하는 sgRNA인 GCTGGCTGGTTATTGCTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCCTTTTTT

T(서열번호 58), 818501-818520 위치의 TATCGTAGTTATCTACACGA(서열번호 59)를 인식하는 sgRNA인 TATCGTAGTTATCTACACGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTTTTT T(서열번호 60), 818606-818625 위치의 CTACGTGAAAGGCGAGATCA(서열번호 61)를 인식하는 sgRNA인 CTACGTGAAAGGCGAGATCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTTTTT T(서열번호 62)를 합성하였다.

[0125] 제조예 II-1-2: cat 유전자와 상보결합 하기 위한 아홉 개의 sgRNA 합성

[0126] EcNR2 대장균 게놈의 cat 유전자(서열번호 63), ATGGAGAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTCAGTCAGTTGCTCAATGTACCTATAAC CAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCTGATGAAT GCTCATCCGAATTTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTACCCTTGTACACCGTTTTCCATGAGCAAAGTAAACGTTTT TCATCGCTCTGGAGTGAATACCACGACGATTCCCGCAGTTTCTACACATATATTGCAAGATGTGGCGTGTACGGTGAAAAACCTGGCCTATTTCCCTAAA GGGTTTATTGAGAATATGTTTTCTGCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACCTCTTCGCCCCCGTT TTCACCATGGGCAATATTATACGCAAGGCGACAAGGTGCTGATGCCGTGGCGATTTCAGGTTTCATGCGCTGTGTATGGCTTCCATGTCGGCAGAATG CTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAA)를 포획하기 위해, cat 유전자의 인근 영역(서열번호 64, CGCGGAATTCATGCTATCGACGTCGATATCTGGCGAAAAATGAGACGTTGATCGGCACGTAAGAGGTTCCAACCTTCACCATAATGAAATAAGATCACTACCG GGCCTATTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAATGGAGAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGT AAAGAACATTTTGAGGCATTTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAG CACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCTGATGAATGCTCATCCGAATTTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGAT AGTGTTCACCTTGTTCACCGTTTTCCATGAGCAAAGTAAACGTTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCGGCAGTTTCTACACATATAT TCGCAAGATGTGGCGTGTACGGTGAAAAACCTGGCCTATTTCCCTAAAGGGTTATTGAGAATATGTTTTCTGCTCAGCCAATCCCTGGGTGAGTTTCACC AGTTTTGATTTAAACGTGGCCAATATGGACAACCTTCTTCGCCCCCGTTTTACCATGGGCAATATTATACGCAAGGCGACAAGGTGCTGATGCCGTGGCG ATTCAGGTTTCATGATGCCGTCTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAATTTGAT ATCGAGCTCGTCAGCAGGCGCGCTGTAATCACACTGGCTCACCTTCGGGTGGGCTTTCTGCGTTTAAAAAAGCGGCGCGCAACGCCGCCGCCGCG CCGCCACCCAGCTTTTGTCCCTTTAGCGTCAGGCGCTGGAG)까지 범위를 넓혀서 cat 유전자가 충분히 포획되도록 하고 이 영 역에서 CRISPR 복합체가 결합하도록 아홉 개의 서열을 합성하였다. 2864476-2864495 위치의 GGCAGAAATGAGACGTTGAT (서열번호 65)를 인식하는 sgRNA인 GGCAGAAATGAGACGTTGATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTTTTT T(서열번호 66), 2864576-2864595 위치의 AGGAGCTAAGGAAGCTAAAA (서열번호 67)를 인식하는 sgRNA인 AGGAGCTAAGGAAGCTAAAAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTTTTT T(서열번호 68), 2864692-2864711 위치의 ATAACCAGACCGTTTCAGCTG (서열번호 69)를 인식하는 sgRNA인 ATAACCAGACCGTTTCAGCTGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTTTTT T(서열번호 70), 2864792-2864811 위치의 GATGAATGCTCATCCGAAT (서열번호 71)를 인식하는 sgRNA인 GATGAATGCTCATCCGAATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTTTTT T(서열번호 72), 2864882-2864901 위치의 TGAGCAAAGTAAACGTTTT (서열번호 73)를 인식하는 sgRNA인 TGAGCAAAGTAAACGTTTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTTTTT T(서열번호 74), 2864987-2865006 위치의 GGCCTATTTCCCTAAAGGGT (서열번호 75)를 인식하는 sgRNA인 GGCCTATTTCCCTAAAGGGTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTTTTT T(서열번호 76), 2865079-2865098 위치의 ATATGGACAACCTCTTCGCC (서열번호 77)를 인식하는 sgRNA인 ATATGGACAACCTCTTCGCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTTTTT T(서열번호 78), 2865178-2865197 위치의 TCTGTGATGGCTTCCATGTC(서열번호 79)를 인식하는 sgRNA인 TCTGTGATGGCTTCCATGTCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTTTTT T(서열번호 80), 2865256-2865275 위치의 TTGATATCGAGCTCGTCAGC(서열번호 81)를 인식하는 sgRNA인 TTGATATCGAGCTCGTCAGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTTTTT T(서열번호 82)를 합성하였다.

[0127] 제조예 II-2: 다중 위치 염기서열 포획을 위한 dCas9 단백질의 제조

[0128] *Streptococcus pyogenes*의 Cas9 유전자에 변이를 도입한 dCas9 유전자를 대장균 단백질 발현 벡터에 일종인 pET28a 벡터에 삽입하였다. 이 때 벡터의 서열 중 단백질 발현에 관련된 부분은 T7 프로모터, dCas9 유전자 그 리고 정제를 위한 히스티딘-태그를 발현시키는 DNA 서열로 이루어져있다. 이 벡터는 T7 RNA 중합효소와 lac repressor에 의해 발현이 조절되며 일반적으로 T7 RNA 중합효소가 존재해야만 발현이 되고 Isopropyl beta-D-1-

thiogalactopyranoside(IPTG)와 함께 배양하면 그 발현량이 매우 증가하는 벡터이다. 이렇게 제작한 벡터를 T7 RNA 중합효소를 가지고 있는 대장균(T7 Express Competent *E. coli*, NEB 사)에 도입하여 dCas9 단백질을 과발현시킨 후 따로 정제하였다.

[0129] dCas9 단백질을 정제 시 먼저 단백질을 과발현 시킨 대장균을 원심분리(3900rpm, 10min)로 모은 후 세포 배양액을 모두 버린다. 그 후 세포 용해액(lysis buffer, 20mm Tris-HCl at pH8.0, 300mM NaCl, 1mg/mL lysozyme, 1x Phenylmethylsulfonyl fluoride(PMSF))을 1mL 세포용해액/100mL 세포 배양액의 비율로 넣고 대장균을 재부유(resuspension)시켜서 초음파(40% 진폭(amplitude)로 10초 파쇄, 30초 휴식을 한 주기로 총 10분)로 분쇄하였다. 이렇게 파쇄한 용액을 원심분리하여(13000rpm, 10min) 상층액만을 얻어낸 후 Ni-NTA 아가로스 레진(resin)에 통과시켜 히스티딘태그를 가지고 있는 단백질만을 레진 상에 남겨두었다. 이후 이 레진에 비정상적으로 결합하여있는 원하지 않는 단백질들을 없애주기 위해 세척을 세 차례 5mL씩 세척용액(washing buffer, 20mm Tris-HCl at pH8.0, 300mM NaCl, 20mm imidazole 1x Phenylmethylsulfonyl fluoride(PMSF))으로 레진을 세척하였다. 그런 후 단백질을 다시 얻어내기 위하여 용리용액(elution buffer, 20mm Tris-HCl at pH8.0, 300mM NaCl, 250mM imidazole, 1x Phenylmethylsulfonyl fluoride(PMSF)) 500를 8번 레진에 통과시켜 원하는 단백질을 선별해내었다.

[0130] 이렇게 정제한 단백질을 염기서열 포획에 이용하기 위해서는 먼저 단백질이 작동하는 작동용액(working buffer, 50mM Tris-HCl at pH 8.0, 200mM KCl, 0.1mM EDTA, 1mM DTT, 0.5mM PMSF, 20% glycerol)로 용액을 교체해줘야 한다. 이는 용리용액에 많이 들어있는 이미다졸을 제거하는 동시에 단백질이 더 안정한 상태로 유지될 수 있는 용액으로 옮겨주는 과정으로 투석(dialysis) 방법을 이용하였다. 여덟 번에 나눠 용리한 용액 중 단백질이 용리된 세 개의 용액 총 1.5mL를 투석 카세트에 넣어준 후 1L의 작동용액으로 4℃에서 16시간 투석해 주었다. 용액 조성을 바꿔준 단백질들은 Bradford assay로 정량 하였다.

[0131] **제조예 II-3: 다중 위치 염기서열 포획을 위한 대상 핵산 서열의 정제**

[0132] 다중 위치 염기서열 포획을 위한 대상 핵산 서열은 대장균 EcNR2 균주를 배양한 후 정제하였다. 배양 조건은 30℃에서 Luria Broth(LB) media를 배양액으로 배양하였다. 이를 원심분리(3600rpm, 10min)으로 세포만 모은 후 Geneall 사의 Exgen Cell SV mini Kit를 이용하여 게놈만을 정제하였다.

[0133] **실시예 II-1: 기 절단된 게놈 핵산 서열에 상보결합을 통한 다중위치 염기서열 동시 포획**

[0134] 제조예 II-1를 통해 제작한 sgRNA 라이브러리 480.7ng과 함께 기 언급한 Cas9 작동용액 조건에서 제조예 II-2를 통해 제조한 dCas9 단백질을 2248.3ng 함께 넣고 20 μL 부피로 맞춘 후 37℃에서 한 시간 반응을 진행하여 다중 위치 염기서열의 동시 포획을 수행하였다. 다중위치 염기서열 동시 포획이 이루어졌는지 확인하기 위해 포획한 서열의 시퀀싱을 수행하였다.

[0135] 구체적으로, Sonication으로 절단한 genome을 바로 Illumina 사의 차세대 시퀀싱 기계를 이용하기 위한 어댑터 DNA 서열을 Enzymatics 사의 SPARK DNA sample prep kit를 이용하여 붙였다. 어댑터를 붙인 DNA 조각들은 USER 효소로 어댑터 DNA에 있는 uracil을 잘라준 뒤 Illumina 사에서 판매하는 Index 서열과 Universal 서열 프라이머를 이용하여 포획한 서열들을 증폭하였다. 증폭한 서열들을 아가로스 젤로 크기별로 구분하였고 이 때 원하는 크기만 선별하여 QIAGEN 사의 spin column을 이용하여 정제하였다. 이후 sgRNA 라이브러리와 dCas9을 섞어 CRISPR 복합체를 만들어준 후 어댑터 DNA 서열을 가지고 있는 절단된 genome과 섞어 CRISPR 복합체가 절단된 genome 서열에 붙도록 하였다.

[0136] 이렇게 절단된 genome, sgRNA 라이브러리, dCas9 효소를 결합시킨 후 포획 대상 핵산만을 선별하기 위하여 dCas9 효소 말단에 위치한 히스티딘 태그를 이용하여 CRISPR 복합체를 정제하였다. CRISPR 복합체 정제 후 포?된 DNA들을 어댑터 서열로 다시 증폭한 후 아가로스 젤로 크기를 다시 확인한 후 정제하였다.

[0137] 그리고 Illumina사의 NextSeq 차세대 시퀀싱 기계를 이용하여 서열 정보를 얻어내었다. 얻어낸 서열 정보를 자체 제작한 Python 프로그램과 BWA 등의 프로그램을 이용하여 원하는 서열이 포획되었는지 분석하였으며 원하는 염기 서열들이 동시에 포획된 것을 확인하였다.

[0138] 그 일부를 예로 들면, 시퀀싱 결과 중 bla 유전자 영역(서열번호 39) 중 'CACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTCGCCCCGAAGAACGTTTCCATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGCA'(서열번호 83)의 시퀀싱 결과로부터 제조예 II-1-1의 서열번호 48(CGAAGAAGCTTTTCCATGAGTTTGTAGAGCTAGAAAATAGCAAGTAAAAATAAGGCTAGTCCGTTATCAACTGAAAAAGTGGACCGAGTCGGTGCTTTT) sgRNA에 의해, EcNR2 유전체 중 817855-817993 위치인 서열번호 84

'CACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCA'의 염기서열이 포획되었음을 확인하였다.

[0139]

또한,

'CTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAA GCCCTCCCGTATCGTAGTTATCTACACGAC'(서열번호 85)의 시퀀싱 결과로부터 제조예 II-1-1의 서열번호 58(GCTGGCTGGTTTATTGCTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTT TTTT)의 sgRNA 혹은, 서열번호 60(TATCGTAGTTATCTACACGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTT TTTT)의 sgRNA에 의해,

[0140]

EcNR2 유전체의 818391-818521 위치인 서열번호 86 'CTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAA GCCCTCCCGTATCGTAGTTATCTACACGAC'가 정확하게 포획된 것을 확인하였다.

[0141]

또 다른 일부를 예로 들면, 시퀀싱 결과 중 cat 유전자 영역 중 'CGTAAAGAACATTTTGAGGCATTTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTAAAGACCGTAAAGAAAAA TAAGCACAAGTTTATCCGGCC'(서열번호 87)의 시퀀싱 결과로부터 제조예 II-1-2의 서열번호 70(ATAACCAGACCGTTCAGCTGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTT TTTT) sgRNA에 의해, EcNR2 유전체 중 2864646-2864768 위치인 서열번호 88 'CGTAAAGAACATTTTGAGGCATTTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTAAAGACCGTAAAGAAAAA TAAGCACAAGTTTATCCGGCC'의 염기서열이 포획되었음을 확인하였다.

[0142]

또한, 서열번호 89 'GCTCTGGAGTGAATACCACGACGATTTCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGT TTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACC'의 시퀀싱 결과로부터 제조예 II-1-2의 서열번호 76(GGCCTATTTCCCTAAAGGGTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTT TTTT)의 sgRNA에 의해,

[0143]

EcNR2 유전체의 2864906-2865056 위치인 서열번호 90 'GCTCTGGAGTGAATACCACGACGATTTCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGT TTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACC'가 정확하게 포획된 것을 확인하였다.

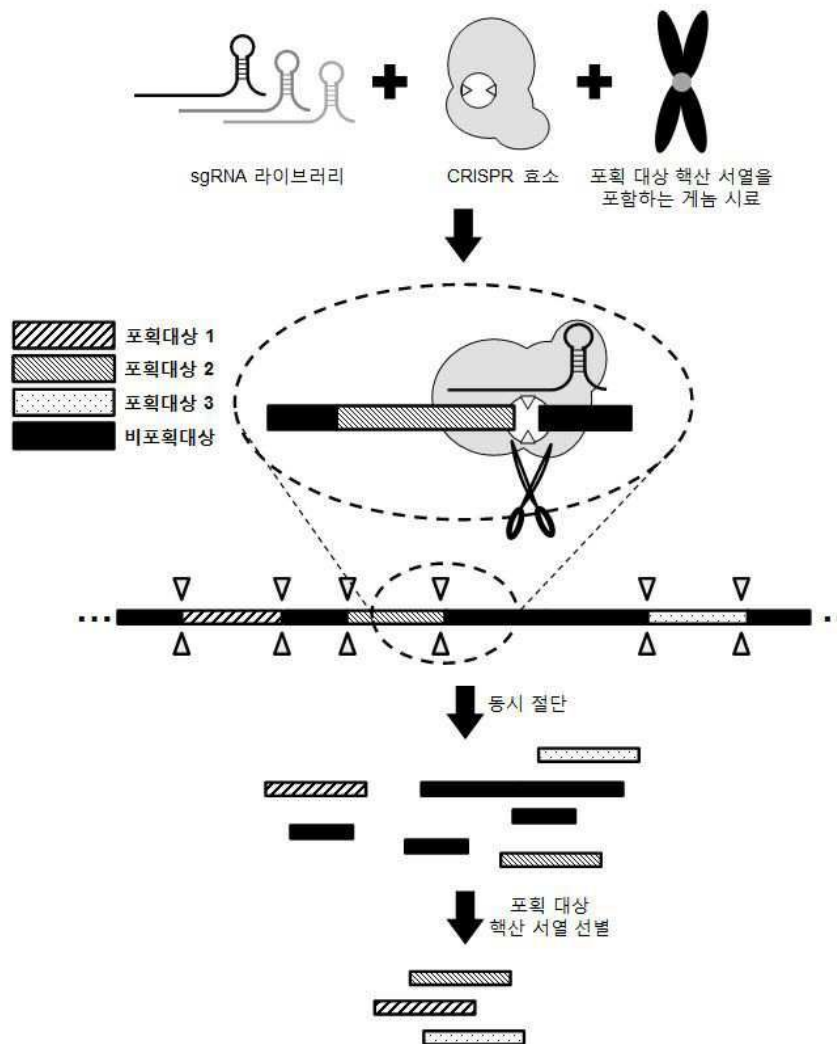
[0144]

[0145]

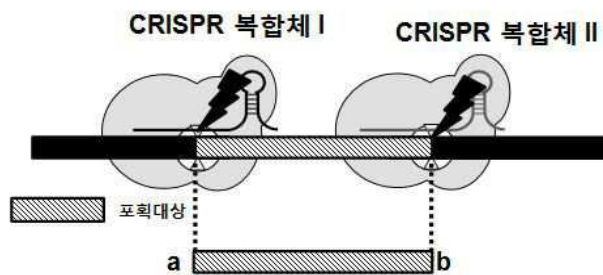
이러한 결과와 같이 우리가 원하는 다양한 염기서열을 동시에 포획하는데 성공하였다.

도면

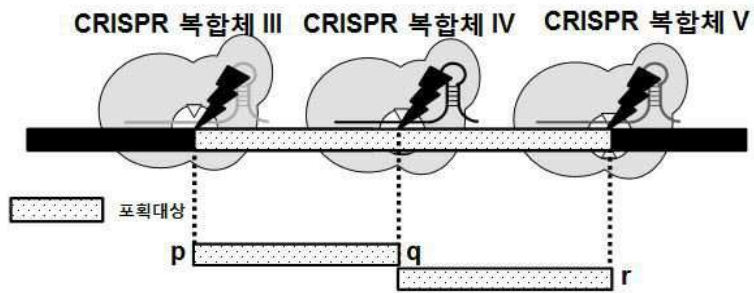
도면1



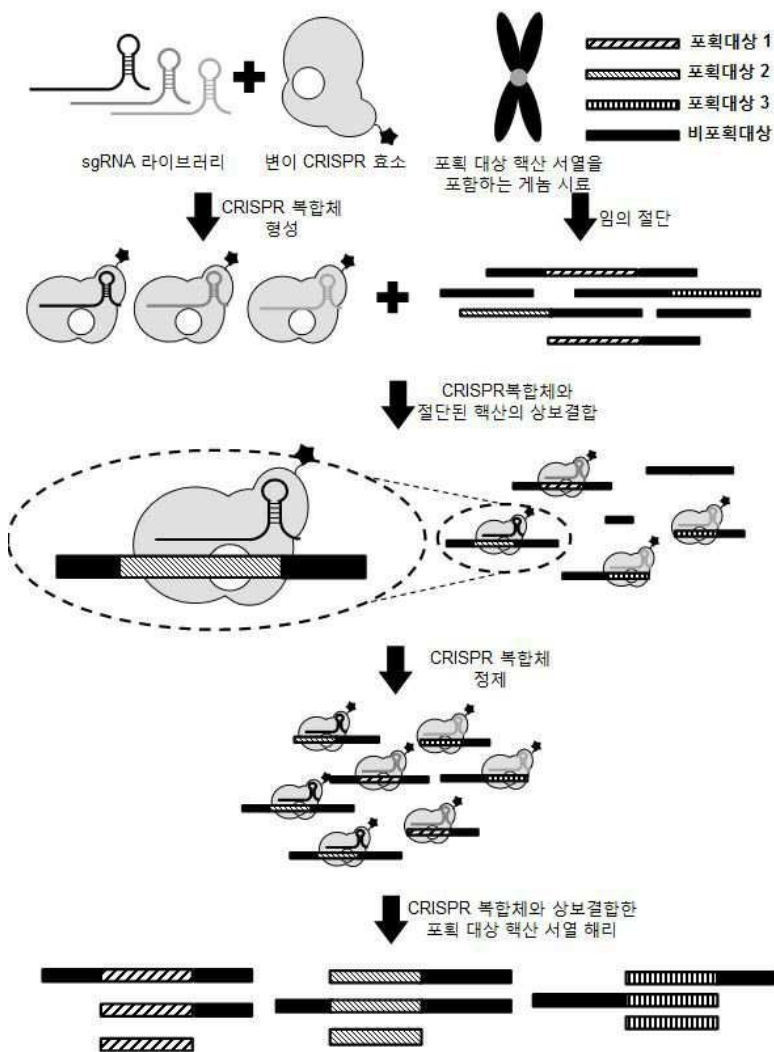
도면2



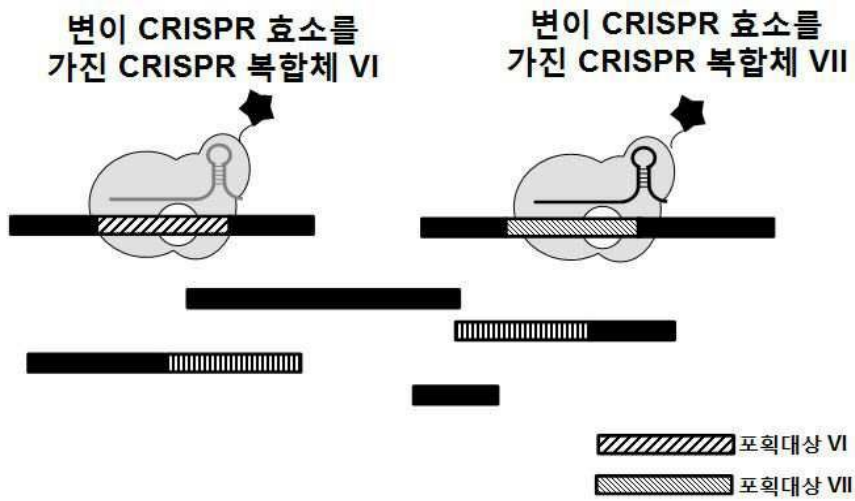
도면3



도면4



도면5



서열 목록

SEQUENCE LISTING

<110> UNIVERSITY-INDUSTRY FOUNDATION, YONSEI UNIVERSITY

<120> Method for target DNA enrichment using CRISPR system

<130> G16U16C0040P/US

<150> KR10-2015-0026203

<151> 2015-02-25

<160> 90

<170> PatentIn version 3.2

<210> 1

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> T7 promoter

<400> 1

ggatttctaatacgactcactatagg

25

<210> 2

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> CRISPR complex-binding sequence1, N is one selected from A, T, C, and G.

<220><221> misc_feature

<222> (1)..(18)

<223> n is a, c, g, or t

<400> 2

nnnnnnnnnn nnnnnnnn 18

<210> 3

<211> 83

<212> DNA

<213> Artificial Sequence

<220><223> sgRNA scaffold

<400> 3

gttttagagc tagaaatagc aagttaaat aaggctagtc cgttatcaac ttgaaaaagt 60

ggcaccgagt cggtagctttt ttt 83

<210>

> 4

<211> 126

<212> DNA

<213> Artificial Sequence

<220><223> template DNA sequence, N is one selected from A, T, C, and G.

<220><221> misc_feature

<222> (26)..(43)

<223> n is a, c, g, or t

<400> 4

ggattctaatac acgactcact ataggnnnnnn nnnnnnnnnn nnngtttttag agctagaaat 60

agcaagttaa aataaggcta gtccgttatc aacttgaaaa agtggcaccg agtcggtgct 120

ttttttt 126

<210> 5

<211> 243

<212> DNA

<213> Homo sapiens

<400>

5

ggatcggact ctttccgtca cccgtttgca cctctgcagc tgtcaggagc gggtcaggtg 60

cggaaagcgg tgcggagggtg gcgctcatag gttacagggg tcagggtctg gggctggccc 120

tggctcttcag ttaccgccga gcgtgcggga tccttctgcg cttgccgcct ccacgtggca 180

caggccaagg cgtggccaga tgggtagatg ggtttgttgg gtggttgcta gcagtttcca 240
cgt 243
<210> 6
<211> 18
<212> DNA
<213> Homo sapiens
<400> 6
ggaggatcgg actctttc 18

<210> 7
<211> 101
<212> DNA
<213> Artificial Sequence
<220><223> sgRNA for recognizing the CRISPR complex-binding sequence of
position
1448011-1448028 of chromosome 1
<400> 7
gaaagagtcc gatcctccgt tttagagcta gaaatagcaa gttaaaataa ggctagtcg 60
ttatcaactt gaaaaagtgg caccgagtcg gtgctttttt t 101
<210> 8
<211> 18
<212> DNA
<213> Homo sapiens
<400> 8
cgtaacaagg gaagcgta 18
<210> 9

<211> 101
<212> DNA
<213> Artificial Sequence
<220><223> sgRNA for recognizing the CRISPR complex-binding sequence of
position
1448254-1448271 of chromosome 1
<400> 9
tacgcttccc ttgttacggt tttagagcta gaaatagcaa gttaaaataa ggctagtcg 60

ttatcaactt gaaaaagtgg caccgagtcg gtgctttttt t 101

<210> 10

<211> 267

<212> DNA

<213> Homo sapiens

<400> 10

cagaggttgc agtttctgag aaacacactg aaaatcctcc ataagtgatt tagaccacgc 60

aaaaacaaga gacaactctc acctgagctg aaatgggtcg ctgaaagggtt tttccagttg 120

atgtttcatt agagacatta ctctgtgggtg tccagtaatg ttctgacatc tgagatgaaa 180

ggtcaaaaat gccatcagag gtgacaaata agcccccattg gggttcacagt ttctaccatt 240

agatatgtgag tcttaaaagc atcccaa 267

<210> 11

<211> 18

<212> DNA

<213> Homo sapiens

<400> 11

tcatacctct cttctcag 18

<210> 12

<211> 101

<212> DNA

<213> Artificial Sequence

<220><223> sgRNA for recognizing the CRISPR complex-binding sequence of

position

55537893-55537910 of chromosome 1

<400> 12

tcatacctct cttctcaggt tttagagcta gaaatagcaa gttaaaataa ggctagtccg 60

ttatcaactt gaaaaagtgg caccgagtcg gtgctttttt t 101

<210> 13

<211> 18

<212> DNA

<213> Homo sapiens

<400> 13

ttaaaagcat cccaagta 18

<210> 14
 <211> 101
 <212> DNA
 <213> Artificial Sequence
 <220><223> sgRNA for recognizing the CRISPR complex-binding sequence of

position

55538160-55538177 of chromosome 1

<400> 14
 ttaaaagcat cccaagtagt tttagagcta gaaatagcaa gttaaataa ggctagtccg 60
 ttatcaactt gaaaaagtgg caccgagtcg gtgctttttt t 101

<210> 15
 <211> 504
 <212> DNA
 <213> Homo sapiens
 <400> 15
 acagggggaa aaccctaiga atgtcatgaa tggggaaga cttctataa gaattcagac 60
 ctcatataac atcaaagaat tcatacaggg gagagacctt atggatgtca tgaatgtggg 120
 aaatccttca gtgaaaagtc aacccttact caacatcaaa gaacgcacac aggggagaaa 180

ccatatgaat gtcatgaatg tgggaaaacc ttctcattta agtcagtcct tactgtgcat 240
 cagaaaaacac acacagggga gaagccctat gaatgctatg catgtgggaa agcctttctc 300
 agaaaatcag acctcattaa acatcaaaga atacacacag gtgaaaaacc ttatgaatgt 360
 aatgaatgtg ggaagtcatt ctctgagaag tcaaccctta ctaaaccatct aagaactcac 420
 acaggtgaga aaccttatga atgtattcag tgtggaaaat tttctgcta ctactccggt 480
 ttcacagaac atctgagaag acac 504

<210> 16
 <211> 18
 <212> DNA
 <213> Homo sapiens
 <400>
 > 16
 tcagagaaca cacacagg 18

<210> 17
 <211> 101

<212> DNA

<213> Artificial sequence

<220><223> sgRNA for recognizing the CRISPR complex-binding sequence of position 38406946-38406963 of chromosome 10

<400> 17

tcagagaaca cacacagggt tttagagcta gaaatagcaa gttaaataa ggctagtcg 60

ttatcaactt gaaaaagtgg caccgagtcg gtgctttttt t 101

<210> 18

<211> 18

<212> DNA

<213> Homo sapiens

<400> 18

gcatcagaaa acacacac 18

<210> 19

<211> 101

<212> DNA

<213> Artificial sequence

<220><223> sgRNA for recognizing the CRISPR complex-binding sequence of position 38407195-38407212 of chromosome 10

<400> 19

gcatcagaaa acacacacgt tttagagcta gaaatagcaa gttaaataa ggctagtcg 60

ttatcaactt gaaaaagtgg caccgagtcg gtgctttttt t 101

<210> 20

<211> 18

<212> DNA

<213> Homo sapiens

<400> 20

acatctgaga agacacac 18

<210> 21

<211> 101

<212> DNA

<213> Artificial Sequence

<220><223> sgRNA for recognizing the CRISPR complex-binding sequence of

position

38407447-38407464 of chromosome 10

<400> 21

acatctgaga agacacacgt tttagagcta gaaatagcaa gttaaaataa ggctagtcg 60
ttatcaactt gaaaaagtgg caccgagtcg gtgctttttt t 101

<210> 22

<211> 260

<212> DNA

<213> Homo sapiens

<400> 22

ttaagggtta agtaattaca catctgtttt gctttttctt ccttctatag tcttaacata 60
gtactctacc cacaggtggg gacaggaagg aaattggatg tggaatgtgg aaaggtggaa 120

acctctacct tgaacaggtt gatgtgtctg atctggctct ggaagagaaa gtcgttgata 180
gtcttcagct ccatccctga gaacaaacac atgaagggcc ttgggagctt caccctaagc 240
ctcaggtttc agtcccaggg 260

<210> 23

<211> 18

<212> DNA

<213> Homo sapiens

<400> 23

acaggcgtgt tgcgttaa 18

<210> 24

<211> 101

<212> DNA

<213> Artificial Sequence

<220><223> sgRNA for recognizing the CRISPR complex-binding sequence of

position

9580087-9580104 of chromosome 12

<400> 24

acaggcgtgt tgcgttaagt tttagagcta gaaatagcaa gttaaaataa ggctagtcg 60
ttatcaactt gaaaaagtgg caccgagtcg gtgctttttt t 101

<210> 25

<211> 18

<212> DNA

<213> Homo sapiens

<400> 25

agggttaagc tcggaagt 18

<210> 26

<211> 101

<212> DNA

<213> Artificial Sequence

<220><223> sgRNA for recognizing the CRISPR complex-binding sequence of

position

9580357-9580374 of chromosome 12

<400> 26

acttccgagc ttaacctgt ttagagcta gaaatagcaa gttaaataa ggctagtccg 60

ttatcaactt gaaaaagtgg caccgagtcg gtgctttttt t 101

<210> 27

<211> 151

<212> DNA

<213> Artificial Sequence

<220><223> sequencing result1 for SEQ ID NO.5

<400> 27

ggatcggact ctttccgtca cccgtttgca cctctgcagc tgtcaggagc gggtcagggtg 60

cggaaagcgg tgcggagggtg gcgctcatag gttacagggg tcagggtctg gggctggccg 120

tggctctcag ttaccgccga gcgtgcggga t 151

<210> 28

<211> 151

<212> DNA

<213> Artificial Sequence

<220><223> sequencing result2 for SEQ ID NO.5

<400> 28

tacaggggtc agggctctggg gctggccgtg gtcttcagtt accgccgagc gtgcggggtc 60

cttctgcgct tgccgctcc acgtggcaca ggccaaggcg tggccagatg ggtagatggg 120

tttgttgggt ggttgctagc agtttccacg t 151

<210> 29

<211> 151
 <212> DNA
 <213> Artificial Sequence
 <220><223> sequencing result1 for SEQ ID NO.10
 <400> 29
 cagaggttgc agtttctgag aaacacactg aaaatcctcc ataagtgatt tagaccacgc 60

 aaaaacaaga gacaactctc acctgagctg aaatgggtcg ctgaaagggt tttccagttg 120
 atgtttcatt agagacatta ctctgtggtg t 151
 <210> 30
 <211> 151
 <212> DNA
 <213> Artificial Sequence
 <220><223> sequencing result2 for SEQ ID NO.10
 <400> 30
 gttgatgttt cattagagac attactctgt ggtgtccagt aatgttctga catctgagat 60
 gaaagggtcaa aaatgccatc agaggtgaca aataagcccc catgggttca cagtttctac 120
 cattagatat tgagtcttaa aagcatccca a 151
 <210> 31

 <211> 151
 <212> DNA
 <213> Artificial Sequence
 <220><223> sequencing result1 for SEQ ID NO.15
 <400> 31
 agggggaaaa ccctatgaat gtcatgaatg tgggaagacc ttctataaga attcagacct 60
 cattaacat caaagaattc atacagggga gagaccttat ggatgtcatg aatgtgggaa 120
 atccttcagt gaaaagtcaa cccttactca a 151
 <210> 32
 <211> 151
 <212> DNA
 <213> Artificial Sequence
 <220><223> sequencing result2 for SEQ ID NO.15
 <400> 32

tggatgtcat gaatgtggga aatccttcag tgaaaagtca acccttactc aacatcaaag 60

aacgcacaca ggggagaaac catatgaatg tcatgaatgt gggaaaacct tctcatttaa 120

gtcagtcctt actgtgcatc agaaaacaca c 151

<210> 33

<211> 151

<212> DNA

<213> Artificial Sequence

<220><223> sequencing result3 for SEQ ID NO.15

<400> 33

acaggggaga agccctatga atgctatgca tgtgggaaag cttttctcag aaaatcagac 60

ctcattaaac atcaaagaat acacacaggt gaaaaacctt atgaatgtaa tgaatgtggg 120

aagtcattct ctgagaagtc aacccttact a 151

<210> 34

<211> 151

<212> DNA

<213> Artificial Sequence

<220><223> sequencing result4 for SEQ ID NO.15

<400> 34

atgaatgtaa tgaatgtggg aagtcattct ctgagaagtc aacccttact aaacatctaa 60

gaactcacac aggtgagaaa ctttatgaat gtattcagtg tggaaaattt ttctgtact 120

actccggttt cacagaacat ctgagaagac a 151

<210> 35

<211> 151

<212> DNA

<213> Artificial Sequence

<220><223> sequencing result1 for SEQ ID NO.22

<400> 35

taagggttaa gtaattacac atctgttttg ctttttcttc cttctatagt cttaacatag 60

tactctaccc acaggtggtg acaggaagga aattggatgt gcaatgtgga aaggtggaaa 120

cctctacctt gaacaggttg atgttgtcga t 151

<210> 36

<211> 151

<212> DNA

<213> Artificial Sequence

<220><223> sequencing result2 for SEQ ID NO.22

<400> 36

ggaaaggtgg aaacctctac cttgaacagg ttgatgttgt cgatctggct ctggaagaga 60

aagtcgttga tagtcttcag ctccatccct gagaacaaac acatgaaggg ccttgggagc 120

ttcacccctaa gcctcaggtt tcagtcaccag g 151

<210> 37

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> CRISPR complex-binding sequence2, N is one selected from A, T, C, and G.

<220><221> misc_feature

<222> (1)..(20)

<223> n is a, c, g, or t

<400> 37

nnnnnnnnnn nnnnnnnnnn 20

<210> 38

<211> 128

<212> DNA

<213> Artificial sequence

<220><223> Template DNA sequence 2

<220><221> misc_feature

<222> (26)..(45)

<223> n is a, c, g, or t

<400> 38

ggattctaat acgactcact ataggnnnnn nnnnnnnnnn nnnnngtttt agagctagaa 60

atagcaagtt aaaataaggc tagtccgtta tcaacttgaa aaagtggcac cgagtcggtg 120

cttttttt 128

<210> 39

<211> 861

<212> DNA

<213> Escherichia coli

<400> 39

atgagtattc aacatttccg tgtcgccctt attccctttt ttgcggcatt ttgccttct	60
gtttttgctc acccagaaac gctggtgaaa gtaaaagatg ctgaagatca gttgggtgca	120
cgagtgggtt acatcgaact ggatctcaac agcggtaaga tccttgagag ttttcgcccc	180
gaagaacgtt ttccaatgat gagcactttt aaagtctcgc tatgtggcgc ggtattatcc	240
cgtattgacg ccgggcaaga gcaactcggc cgccgcatac actattctca gaatgacttg	300
gttgagtact caccagtcac agaaaagcat cttacggatg gcatgacagt aagagaatta	360
tgcagtgtcg ccataacatc gagtgataac actgcggcca acttacttct gacaacgac	420
ggaggaccga aggagctaac cgcttttttg cacaacatgg gggatcatgt aactcgcctt	480
gatcgttggg aaccggagct gaatgaagcc ataccaaagc acgagcgtga caccacgatg	540
cctgtagcaa tggcaacaac gttgcgcaaa ctattaactg gcgaactact tactctagct	600
tcccggcaac aattaataga ctggatggag gcggataaag ttgcaggacc acttctgcgc	660
tcggcccttc cgcttggtcg gttttattgt gataaatctg gagccggtga gcgtgggtct	720
cgcggtatca ttgcagcact ggggccagat ggtaagccct cccgtatcgt agttatctac	780
acgacgggga gtcaggcaac tatggatgaa cgaaatagac agatcgctga gatagggtgcc	840
tcactgatta agcatttgta a	861

<210> 40

<211> 1161

<212> DNA

<213> Escherichia coli

<400> 40

ttattcggcc ttgaattgat catatgcgga ttagaaaaac aacttaaatg tgaaagtggg	60
tcctaacagt tcctggatat ccggatgaag gcacgaaccc agtggacata accctgataa	120
atgcttcaat aatattgaaa aaggaagagt atgagtattc aacatttccg tgtcgccctt	180
attccctttt ttgcggcatt ttgccttctt gtttttgctc acccagaaac gctggtgaaa	240
gtaaaagatg ctgaagatca gttgggtgca cgagtgggtt acatcgaact ggatctcaac	300
agcggtaaga tccttgagag ttttcgcccc gaagaacgtt ttccaatgat gagcactttt	360
aaagtctcgc tatgtggcgc ggtattatcc cgtattgacg ccgggcaaga gcaactcggc	420
cgccgcatac actattctca gaatgacttg gttgagtact caccagtcac agaaaagcat	480
cttacggatg gcatgacagt aagagaatta tgcagtgtcg ccataacatc gagtgataac	540

actgcggcca acttacttct gacaacgac ggaggaccga aggagctaac cgcttttttg 600
 cacaacatgg gggatcatgt aactcgcctt gatcgttggg aaccggagct gaatgaagcc 660
 ataccaaacg acgagcgtga caccacgatg cctgtagcaa tggcaacaac gttgcgcaaa 720
 ctattaactg gcgaactact tactctagct tcccggcaac aattaataga ctggatggag 780
 gcggataaag ttgcaggacc acttctgcgc tcggcccttc cggtcggctg gtttattgct 840
 gataaatctg gagccggtga gcgtgggtct cgcggtatca ttgcagcact ggggccagat 900
 ggtaagccct cccgtatcgt agttatctac acgacgggga gtcaggcaac tatggatgaa 960

cgaaatagac agatcgctga gatagtgcc tctactgatta agcatttgta atttgtccac 1020
 tactgaaag gcgagatcac caaggtagtc ggcaaataat gtctaacaat tcgttcaagc 1080
 cgacggatat cgagctcgtc tggactcctg ttgatagatc cagtaatgac ctcagaactc 1140
 catctggatt tggtcagaac g 1161

<210> 41

<211> 20

<212> DNA

<213> Escherichia coli

<400> 41

aaacaactta aatgtgaaag 20

<210> 42

<211> 103

<212> DNA

<213> Artificial sequence

<220><223> sgRNA recognizing position 817623-817642 of bla gene

<400> 42

aaacaactta aatgtgaaag gtttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cgtgctttt ttt 103

<210> 43

<211> 20

<212> DNA

<213> Escherichia coli

<400> 43

tgcttcaata atattgaaaa 20

<210> 44

<211> 103

<212> DNA

<213> Artificial sequence

<220><223> sgRNA recognizing position 817708-817727 of bla gene

<400> 44

tgcttcaata atattgaaaa gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggtagctttt ttt 103

<210> 45

<211> 20

<212> DNA

<213> Escherichia coli

<400> 45

ttttgtcac ccagaaacgc 20

<210> 46

<211> 103

<212> DNA

<213> Artificial sequence

<220><223> sgRNA recognizing position 817799-817818 of bla gene

<400> 46

ttttgtcac ccagaaacgc gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggtagctttt ttt 103

<210> 47

<211> 20

<212> DNA

<213> Escherichia coli

<400> 47

cgaagaacgt tttccaatga 20

<210> 48

<211> 103

<212> DNA

<213> Artificial sequence

<220><223> sgRNA recognizing position 817916-817935 of bla gene

<400> 48

cgaagaacgt tttccaatga gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt ttt 103

<210> 49

<211> 20

<212> DNA

<213> Escherichia coli

<400> 49

catacactat tctcagaatg 20

<210> 50

<211> 103

<212> DNA

<213> Artificial sequence

<220><223> sgRNA recognizing position 818012-818031 of bla gene

<400> 50

catacactat tctcagaatg gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt ttt 103

<210> 51

<211> 20

<212> DNA

<213> Escherichia coli

<400>

> 51

taaccatgag tgataacact 20

<210> 52

<211> 103

<212> DNA

<213> Artificial sequence

<220><223> sgRNA recognizing position 818110-818129 of bla gene

<400> 52

taaccatgag tgataacact gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt ttt 103

<210> 53

<211> 20

<212> DNA

<213> Escherichia coli

<400> 53

tgatcggttg gaaccggagc

20

<210> 54

<211> 103

<212> DNA

<213> Artificial sequence

<220><223> sgRNA recognizing position 818216-818235 of bla gene

<400> 54

tgatcggttg gaaccggagc gtttagagc tagaaatagc aagttaaat aaggctagtc

60

cgttatcaac ttgaaaaagt ggcaccgagt cggcgctttt ttt

103

<210> 55

<211> 20

<212> DNA

<213> Escherichia coli

<400> 55

acgttgcgca aactattaac

20

<210> 56

<211> 103

<212> DNA

<213> Artificial sequence

<220><223> sgRNA recognizing position 818295-818314 of bla gene

<400> 56

acgttgcgca aactattaac gtttagagc tagaaatagc aagttaaat aaggctagtc

60

cgttatcaac ttgaaaaagt ggcaccgagt cggcgctttt ttt

103

<210> 57

<211> 20

<212> DNA

<213> Escherichia coli

<400> 57

gctggctggt ttattgctga

20

<210> 58

<211> 103

<212> DNA

<213> Artificial sequence

<220><223> sgRNA recognizing position 818409-818428 of bla gene

<400> 58

gctggctggt ttattgctga gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggcgctttt ttt 103

<210> 59

<211> 20

<212> DNA

<213> Escherichia coli

<400> 59

tatcgtagtt atctacacga 20

<210> 60

<211> 103

<212> DNA

<213> Artificial sequence

<220><223> sgRNA recognizing position 818501-818520 of bla gene

<400> 60

tatcgtagtt atctacacga gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggcgctttt ttt 103

<210> 61

<211> 20

<212> DNA

<213> Escherichia coli

<400> 61

ctacgtgaaa ggcgagatca 20

<210> 62

<211> 103

<212> DNA

<213> Artificial sequence

<220><223> sgRNA recognizing position 818606-818625 of bla gen

<400> 62

ctacgtgaaa ggcgagatca gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggtagctttt ttt 103

<210> 63

<211> 660

<212> DNA

<213> Escherichia coli

<400> 63

atggagaaaa aaatcactgg atataccacc gttgatatat cccaatggca tcgtaagaa 60
cattttgagg catttcagtc agttgctcaa tgtacctata accagaccgt tcagctggat 120
attacggcct ttttaaagac cgtaaagaaa aataagcaca agttttatcc ggcctttatt 180
cacattcttg cccgcctgat gaatgctcat ccggaatttc gtatggcaat gaaagacggt 240
gagctggtga tatgggatag tgttcacctt ttttacaccg ttttccatga gcaaactgaa 300
acgttttcat cgctctggag tgaataccac gacgatttcc ggcagtttct acacatatat 360

tcgcaagatg tggcgtgtta cggtgaaaac ctggcctatt tccctaaagg gtttattgag 420
aatatgtttt tcgtctcagc caatccctgg gtgagtttca ccagttttga tttaaacgtg 480
gccaatatgg acaacttctt cgccccgtt ttcacatgg gcaaatatta tacgcaaggc 540
gacaaggtgc tgatgccgtt ggcgattcag gttcatcatg ccgtctgtga tggtttccat 600
gtcggcagaa tgcttaatga attacaacag tactgcatg agtggcaggg cggggcgtaa 660

<210> 64

<211> 960

<212> DNA

<213> Escherichia coli

<400> 64

cgcggaattc atgctatcga cgtcgataac tggcgaaaat gagacgttga tcggcacgta 60

agaggttcca actttcacca taatgaaata agatcactac cgggcgtatt ttttgagtta 120
tcgagatttt caggagctaa ggaagctaaa atggagaaaa aaatcactgg atataccacc 180
gttgatatat cccaatggca tcgtaagaa cattttgagg catttcagtc agttgctcaa 240
tgtacctata accagaccgt tcagctggat attacggcct ttttaaagac cgtaaagaaa 300
aataagcaca agttttatcc ggcctttatt cacattcttg cccgcctgat gaatgctcat 360
ccggaatttc gtatggcaat gaaagacggt gagctggtga tatgggatag tgttcacctt 420
tgttacaccg ttttccatga gcaaactgaa acgttttcat cgctctggag tgaataccac 480

gacgatttcc ggcagtttct acacatatat tcgcaagatg tggcgtgtta cggtgaaaac 540

ctggcctatt tccctaaagg gtttattgag aatatgtttt tcgtctcagc caatccctgg 600
 gtgagtttca ccagttttga ttttaacgtg gccaatatgg acaacttctt cgcctccgtt 660
 ttcacatagg gcaaatatta tacgcaaggc gacaaggatg tgatgccgct ggcgattcag 720
 gtatcatcatg ccgtctgtga tggcttccat gtcggcagaa tgcttaatga attacaacag 780
 tactgcgatg agtggcaggg cggggcgtaa tttgatatcg agctcgtcag caggcgcgcc 840
 tgtaatcaca ctggctcacc ttcgggtggg cttttctgcg tttaaaaaaa acgggccggc 900

gcgaacgccg gcccgcgccc gccacccagc tttgttccc tttagcgtca ggcgctggag 960

<210> 65
 <211> 20
 <212> DNA
 <213> Escherichia coli
 <400> 65

ggcgaaaatg agacgttgat 20

<210> 66
 <211> 103
 <212> DNA
 <213> Artificial sequence

<220><223> sgRNA recognizing position 2864476-2864495 of cat gene

<400> 66

ggcgaaaatg agacgttgat gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
 cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt ttt 103

<210> 67
 <211> 20
 <212> DNA
 <213> Escherichia coli
 <400> 67

aggagctaag gaagctaaaa 20

<210> 68
 <211> 103
 <212> DNA
 <213> Artificial sequence

<220><223> sgRNA recognizing position 2864576-2864595 of cat gene

<400> 68

aggagctaag gaagctaaaa gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
 cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt ttt 103
 <210> 69
 <211> 20
 <212> DNA
 <213> Escherichia coli
 <400>
 > 69
 ataaccagac cggtcagctg 20
 <210> 70
 <211> 103
 <212> DNA
 <213> Artificial sequence
 <220><223> sgRNA recognizing position 2864692-2864711 of cat gene
 <400> 70
 ataaccagac cggtcagctg gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
 cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt ttt 103
 <210> 71
 <211> 20
 <212> DNA
 <213> Escherichia coli
 <400> 71
 gatgaatgct catccggaat 20
 <210> 72
 <211> 103
 <212> DNA
 <213> Artificial sequence
 <220><223> sgRNA recognizing position 2864792-2864811 of cat gene
 <400> 72
 gatgaatgct catccggaat gttttagagc tagaaatagc aagttaaaat aaggctagtc 60
 cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt ttt 103
 <210> 73
 <211> 20
 <212> DNA

<213> Escherichia coli

<400> 73

tgagcaaact gaaacgtttt 20

<210> 74

<211> 103

<212> DNA

<213> Artificial sequence

<220><223> sgRNA recognizing position 2864882-2864901 of cat gene

<400> 74

tgagcaaact gaaacgtttt gtttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cgtgctttt ttt 103

<210> 75

<211> 20

<212> DNA

<213> Escherichia coli

<400> 75

ggcctatttc cctaaagggt 20

<210> 76

<211> 103

<212> DNA

<213> Artificial sequence

<220><223> sgRNA recognizing position 2864987-2865006 of cat gene

<400> 76

ggcctatttc cctaaagggt gtttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cgtgctttt ttt 103

<210> 77

<211> 20

<212> DNA

<213> Escherichia coli

<400> 77

atatggacaa cttcttcgcc 20

<210> 78

<211> 103

<212> DNA

<213> Artificial sequence

<220><223> sgRNA recognizing position 2865079-2865098 of cat gene

<400> 78

atatggacaa cttcttcgcc gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggcgctttt ttt 103

<210> 79

<211> 20

<212> DNA

<213> Escherichia coli

<400> 79

tctgtgatgg cttccatgtc 20

<210> 80

<211> 103

<212> DNA

<213> Artificial sequence

<220><223> sgRNA recognizing position 2865178-2865197 of cat gene

<400> 80

tctgtgatgg cttccatgtc gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggcgctttt ttt 103

<210> 81

<211> 20

<212> DNA

<213> Escherichia coli

<400> 81

ttgatatcga gctcgtcagc 20

<210> 82

<211> 103

<212> DNA

<213> Artificial sequence

<220><223> sgRNA sequence position 2865256-2865275 of cat gene

<400> 82

ttgatatcga gctcgtcagc gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt ttt 103

<210> 83

<211> 139

<212> DNA

<213> Artificial sequence

<

220><223> sequencing result1 of bla gene

<400> 83

cacgagtggg ttacatcgaa ctggatctca acagcggtaa gatccttgag agttttcgcc 60

ccgaagaacg ttttccaatg atgagcactt ttaaagtctt gctatgtggc gcggtattat 120

cccgtattga cgccgggca 139

<210> 84

<211> 139

<212> DNA

<213> Escherichia coli

<400> 84

cacgagtggg ttacatcgaa ctggatctca acagcggtaa gatccttgag agttttcgcc 60

ccgaagaacg ttttccaatg atgagcactt ttaaagtctt gctatgtggc gcggtattat 120

cccgtattga cgccgggca 139

<210> 85

<211> 131

<212> DNA

<213> Artificial sequence

<220><223> sequencing result2 of bla gene

<400> 85

ctgcgctcgg cccttccggc tggctggttt attgctgata aatctggagc cggtgagcgt 60

gggtctcgcg gtatcattgc agcactgggg ccagatggta agccctcccg tatcgtagtt 120

atctacacga c 131

<210> 86

<211> 131

<212> DNA

<213> Escherichia coli

<400> 86

ctgcgctcgg cccttccggc tggctggttt attgctgata aatctggagc cggtgagcgt 60

gggtctcgcg gtatcattgc agcactgggg ccagatggta agccctcccg tatcgtagtt 120

atctacacga c 131

<210> 87

<211> 123

<212> DNA

<213> Artificial sequence

<220><223> sequencing result1 of cat gene

<400> 87

cgtaaagaac atttgaggc atttcagtca gttgctcaat gtacctataa ccagaccgtt 60

cagctggata ttacggcctt tttaaagacc gtaaagaaaa ataagcacia gttttatccg 120

gcc 123

<210> 88

<211> 123

<212> DNA

<213> Escherichia coli

<400> 88

cgtaaagaac atttgaggc atttcagtca gttgctcaat gtacctataa ccagaccgtt 60

cagctggata ttacggcctt tttaaagacc gtaaagaaaa ataagcacia gttttatccg 120

gcc 123

<210> 89

<211> 151

<212> DNA

<213> Artificial sequence

<220><223> sequencing result2 of cat gene

<400> 89

gtcttgaggt gaataccacg acgatttccg gcagtttcta cacatatatt cgcaagatgt 60

ggcgtgttac ggtgaaaacc tggcctatit ccctaaaggg tttattgaga atatgttttt 120

cgtctcagcc aatccctggg tgagtttcac c 151

<210> 90

<211>

> 151

<212> DNA

<213> Escherichia coli

<400> 90

gctctggagt gaataccacg acgatttccg gcagtttcta cacatatatt cgcaagatgt	60
ggcgtgttac ggtgaaaacc tggcctatit ccctaaaggg tttattgaga atatgttttt	120
cgtctcagcc aatccctggg tgagtttcac c	151