DBpia

논문 20-45-08-15 The Journal of Korean Institute of Communications and Information Sciences '20-08 Vol.45 No.08
https://doi.org/10.7840/kics.2020.45.8.1447

분자육종 연구에서 Big Data를 활용한 가이드 RNA(gRNA) 제작의 중요성

정 찬 영◆, 홍 정 원*, 유 정 희*, 김 수 배*, 강 상 국*,
김 성 완*, 김 남 숙*, 김 기 영*, 박 종 우°

Importance of Guide-RNA Design Using Big Data in Insect Molecular Breeding Research

Chan Young Jeong◆, Jeong Won Hong*, Jeong Hee Yu*, Su-Bae Kim*, Sang Kuk Kang*,
Seong-Wan Kim*, Nam-Suk Kim*, Kee Young Kim*, Jong Woo Park°

요 약

본 논문에서는 특이 종에 대한 분자육종 연구시 생물 게놈 big data를 활용 필요성 및 효과를 분석하였다. 게놈 데이터베이스 기반의 웹 서버에 편집이 필요한 영역의 서열을 입력하고 게놈 big data와 상호 검증을 통하여 유전자 편집을 위한 가이드 RNA를 선발하고 제작할 수 있다. 제작된 가이드 RNA는 누에에 적용하여 돌연변이를 제작함으로써 효율을 입증하였고, 돌연변이의 염색체 서열분석을 통하여 표적 서열에 대한 정확성을 검증하였다. 그 결과 big data 기반의 CRISPRdirect 서버를 통해 제작된 가이드 RNA는 표적 부위 특이적인 반응성을 나타냄으로써 성공적인 유전자 편집을 가능하게 함을 확인하였다.

키워드 : 유전자 편집, 크리스퍼, 가이드 RNA, 빅데이터
Key Words : Genome editing, CRISPR/Cas9, guide-RNA, Big data

ABSTRACT

In this paper, we analyzed the necessity and effect of using genome big data when studying molecular breeding for various species. A guide RNA for gene editing was selected and produced by inputting the sequence of the region that needs to be edited into a genome data base web server and cross-verifying it with genome big data. The efficiency was verified by making the mutant by applying the prepared guide RNA to the silkworm, and the accuracy of the target sequence was verified through chromosomal sequencing of the mutant. As a result, it was confirmed that the guide RNA produced through the CRISPRdirect server based on the big data enables specific gene editing by exhibiting specific reactivity to the target site.

※ 본 연구는 2020년도 농촌진흥청 연구사업(세부과제번호: PJ01270902) 지원사업에 의해 이루어진 것임
◆ First Author : Department of Agricultural Biology, National Institute of Agricultural Science, duckbb@korea.kr, 정회원
° Corresponding Author : Department of Agricultural Biology, National Institute of Agricultural Science, jwpark0824@korea.kr, 정회원
* Department of Agricultural Biology, National Institute of Agricultural Science, corea97@korea.kr, yjh0817@rda.go.kr; subaekim@korea.kr; wkdudghl@korea.kr; tarupa@korea.kr; nskim@korea.kr; applekky@korea.kr
논문번호 : 202004-076-0-SE, Received March 30, 2020; Revised May 24, 2020; Accepted June 3, 2020

www.dbpia.co.kr

Ⅰ. 서 론

식량의 안정적인 생산 및 고품질 원료 생산을 위한 육종의 역사는 매우 깊으며, 현대에 이르러서는 기후나 환경변화에 대응하기 위한 육종 또한 증가 추세이다. 하지만 전통적인 교배육종은 새로운 형질보다는 우수한 형질 사이의 교배를 통해 더 나은 형질을 선발하거나 교배에 의한 잡종강세를 이용하는 방법을 택해왔으며, 이와 같은 교배육종은 긴 육종 기간과 노동력의 투입에 따른 비용손실이 크다는 문제점 외에도 육종 가능한 형질에 한계가 존재한다고 지적받아왔다[1].

현대사회에 이르러 생명공학 기술 및 정보처리 기술의 급속한 발달에 따라 생물체의 유전자 정보가 데이터베이스화 되고, 많은 유전자의 기능이 밝혀지게 되었으며, 이를 이용하여 교배육종의 한계를 극복하고자 유전공학 기술을 이용한 분자육종 연구가 활발하게 이루어지고 있다. 하지만 대표적 유전공학 기술인 형질전환을 통해 생산된 생물체들이 GMO로 분류되면서 안전성에 대한 논쟁이 끊이지 않게 되었었고 부정적인 인식의 확산에 따라 개발된 동·식물자원의 이용에 제약을 받고 있다. 이에 따라 수많은 연구자들이 GMO에 대한 제약에서 벗어날 수 있는 분자육종 방법에 대하여 고민하게 되었고, 최근 유전자가위를 이용한 유전자 편집기술이 등장하면서 새로운 분자육종 분야에서 큰 호응을 얻고 있다.

유전자 편집기술은 염색체 DNA를 부위 특이적으로 조작함으로써 분자육종을 가능하게 하는 유망한 기술이다. 제3세대 유전자가위로 통칭되는 Clustered, regularly interspaced short palindromic repeat (CRISPR)/CRISPR associated protein (Cas) 시스템은 원핵생물인 Streptococcus pyogenes의 외부에서 침입한 바이러스에 대응하기 위한 적응면역 시스템에서 발견되었다. 이는 감염된 바이러스에 대한 유전정보가 규칙적인 간격을 가진 클러스터 형태로 S. pyogenes 유전체에 삽입되고, 클러스터 내부의 서열들이 머리핀을 닮은 짧은 RNA 단편으로 전사된 후 S. pyogenes의 핵산가수분해효소(Cas9)를 바이러스 게놈으로 인도하여 제거하게 된다. 즉 목적 유전자를 표적화하여 핵산가수분해효소를 작동시킬 수 있기 때문에 CRISPR/Cas 시스템은 유전자 편집 도구로서 효용성을 가진다[2]. 최근 연구에 따르면 표적에 상보적인 20개의 뉴클레오타이드와 머리핀 구조를 형성하는 100개가량의 뉴클레오타이드가 결합 된 가이드 RNA는 Cas9 단백질을 표적 부위로 인도하여 게놈 내부에 절단을 유발 시키고 생물체의 DNA 회복기작에 의하여 절단 부위의 복구가 일어나면서 일부 염기의 결실이 발생 됨이 밝혀졌다[3]. 이러한 절단-회복 반응을 효율적으로 이용하는 것이 유전자 편집기술이라 할 수 있다.

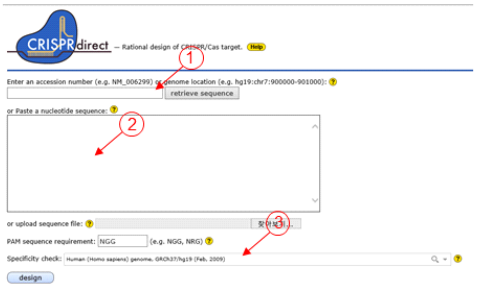
CRISPR/Cas 시스템은 Cas9 단백질과 가이드 RNA만을 필수 요소로 필요하기 때문에 운용이 매우 간단하며, 가이드 RNA의 20개의 염기서열을 변경함으로써 다양한 표적 유전자에 대한 표적화가 가능하다. 그뿐만 아니라 특정 유전자를 표적화하기 위한 가이드 RNA는 서열의 길이가 짧기 때문에 합성이 간단하다. 이처럼 짧은 DNA 서열에 의한 표적 결정은 사용 편의성에서 매우 큰 장점이 있지만, 게놈 내에 유사서열의 존재 확률이 증가한다는 단점을 가진다[4]. 즉, 유사서열의 존재 여부는 원하지 않는 부위 유전자의 불활성화 및 표현 형질 예측에 어려움을 겪게 할 수 있다. 따라서 유전자 편집을 진행에 있어 안정적인 변이를 유발하기 위해서는 의도하지 않은 off-target 유전자의 절단을 최소화하는 것이 중요한 문제로 제기되고 있다[5].

본 연구에서 우리는 CRISPR/Cas 시스템을 이용한 곤충 육종을 위하여 필요한 가이드 RNA 제작을 위해 big data 기반의 CRISPRdirect (http://crispr.dbcls.jp/)을 활용하고, 표적 서열 외 염기서열의 길이가 축소된 잠재적인 off-target 서열을 분석 후 얻어지는 CRISPR/Cas 표적 부위 선발 및 효과에 대해 분석하였다.

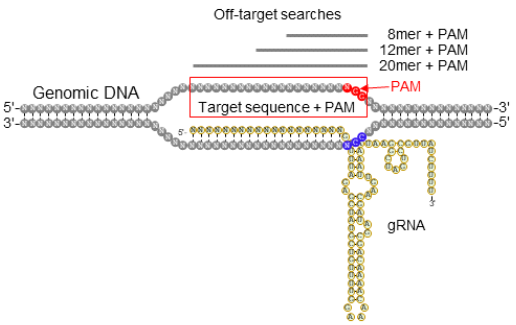
Ⅱ. 시스템의 구성 및 특징

CRISPRdirect는 이용자가 가지고 있는 최대 10 kb의 염기서열을 그림 1(a)-①에 입력하고 내부에서 Protospacer Adjacent Motif (PAM; NGG) 서열과 인접한 20 bp의 염기서열을 검출할 수 있다. 그뿐만 아니라 NCBI의 염기서열 데이터베이스에 등록된 유전자 서열번호를 그림 1(a)-②에 입력 후 표적 서열에 대한 분석이 가능하도록 구성되어있다, 또한 그림 1(a)-③을 통해 검출된 표적 서열들의 표적 특이성 분석기능을 제공한다. 이는 목적 생물 종의 게놈 big data 내에서 그림 1(b)에 나타낸 PAM 측면의 표적 서열(20 mer) 및 off-target 사이트인 시드 서열(12 또는 8 mer)의 세부 목록 탐색이 가능하며, 인간, 생쥐, 쥐, 돼지, 닭, 개구리, 초파리, 누에, 애기장대, 쌀, 및 효모 등의 다양한 게놈 데이터베이스를 이용할 수 있도록 구성되어있다[6]. 대표적인 가이드 RNA 제작 서버인

www.dbpia.co.kr



(a) Screenshot from the CRSPRdirect web



(b) Illustrated overview of specificity check

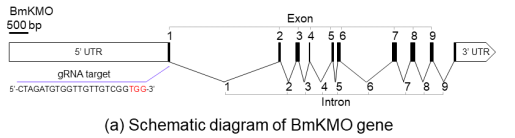
그림 1. CRSPRdirect 툴의 특성
Fig. 1. Features of th CRSPRdirect web tool

CRISPR Design[7]의 경우 최대 250 bp의 서열에서 표적을 선발하여 잠재적인 off-target을 기반으로 점수를 부여하고, E-CRISP[8]의 경우 표적 내 특이성 및 수에 따라 가이드 RNA의 순위를 결정한다. 이와 같은 서버 역시 가이드 RNA를 설계하는 데 유용하지만 많은 수의 입력 서열을 처리하기 위해서는 프로세스 및 제공되는 종 데이터베이스에 한계가 존재한다. 하지만 CRISPRdirect에서는 결과를 신속하게 처리할 수 있으며 데이터 내보내기 및 자동화된 인터페이스를 제공한다.

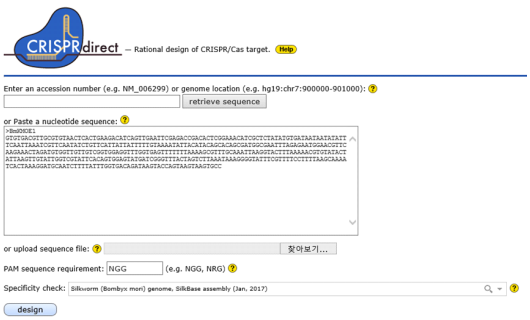
Ⅲ. 시스템의 활용 및 검증

3.1 분석대상

CRISPRdirect 시스템의 효용성을 분석하기 위하여 산업 곤충인 누에의 kynurenine 3-monooxygenase (KMO) 유전자를 이용하였다[9]. KMO 유전자는 그림 2(a)와 같이 엑손 및 인트론이 각각 9개로 구성되어있으며, 유전자 편집 분석을 위해 엑손 1번에 대한 가이드 RNA 표적 서열 및 잠재적인 off-target 서열을 검출하였다.



(a) Schematic diagram of BmKMO gene



(b) Analysis of KMO gene using CRSPRdirect web tool

그림 2. KMO 유전자 편집을 위한 가이드 RNA 디자인
Fig. 2. Guide RNA design for BmKMO gene editing

3.2 곤충모델을 이용한 가이드 RNA 검증

그림 2(b)의 온라인 시스템에 KMO 유전자의 엑손 1번 서열을 입력 후 분석대상을 누에의 게놈으로 선택한 후 분석을 수행하였다. 표 1의 표적 서열 중 off-target 가능성이 가장 낮은 1종의 가이드 RNA를 선발 후 논문 [10]의 방법에 따라 제작하고, Cas9 단백질과 혼합하여 누에의 배아에 도입 후 사육하였다. KMO 유전자 편집에 따라 발생 된 돌연변이 개체를 선발하고 게놈을 추출하여 잠재적인 off-target 부위에

표 1. KMO 유전자에 대한 가이드 RNA 표적 서열 및 유사서열 출현 빈도
Table 1. Analysis of gRNA target sequences of the BmKMO gene and potential off-target in the genome

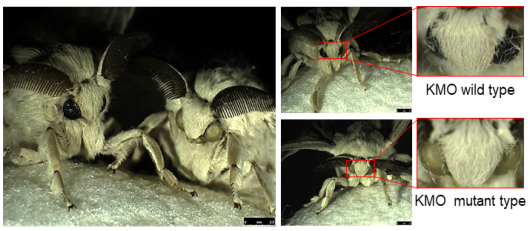
Position	(Strand) Sequence	Hit (mer)		
		20	12	8
36-58	(+) TTGAATTCGAGACCGACACT**CGG**	0	2	289
48-70	(-) **CCG**ACACTCGGAAACATCGCTCT	0	1	202
151-173	(+) GCGATGGCGAATTTAGAGAA**TGG**	1	4	439
174-196	(+) AACGTTCAAGAAACTAGATGTGG	1	4	377
184-206	(+) AAACTAGATGTGGTTGTTGTCGG	1	5	894
187-209	(+) CTAGATGTGGTTGTTGTCGGTGG	1	2	248
190-212	(+) GATGTGGTTGTTGTCGGTGGAGG	1	3	930
195-217	(+) GGTTGTTGTCGGTGGAGGTTTGG	1	3	419
227-249	(+) TTAAAAGCGTTTGCAAATTAAGG	1	9	494
263-285	(+) TGTATACTATTAAGTTGTATTGG	1	8	595
278-300	(+) TGTATTGGTCGTATTCACAGTGG	0	1	1864
289-311	(+) TATTCACAGTGGAGTATGATCGG	0	1	274
290-312	(+) ATTCACAGTGGAGTATGATCGGG	0	3	162
311-333	(+) GGTTTACTAGTCTTAAATAAAGG	0	6	4957
347-369	(-) CCTTTTAAGCAAAATCACTAAAG	0	10	2804
348-370	(+) CTTTTAAGCAAAATCACTAAAGG	0	4	425
366-388	(+) AAAGGATGCAATCTTTTATTTGG	1	22	1320

www.dbpia.co.kr

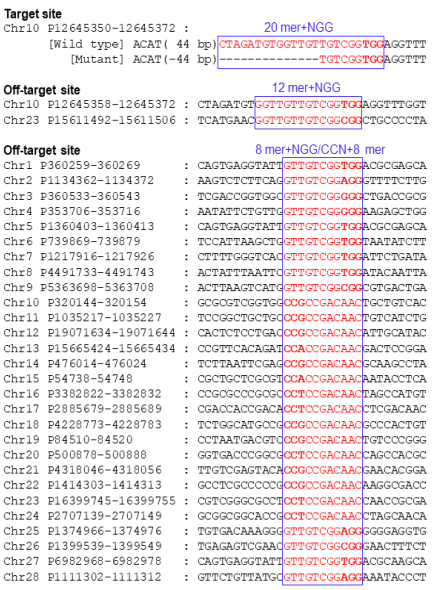
대한 염기서열을 분석하였다. 표적서 열 187-209 부위의 누에 게놈 내 존재하는 12 mer 중복된 2 부위와 8 mer 중복부위 248개 중 28곳에 대하여 염기서열분석을 통하여 off - target의 영향에 대하여 분석하였다.

Ⅳ. 결 과

누에의 배아에 micro-injection을 통하여 선발된 가이드 RNA와 Cas9 효소를 주입하고, 부화한 유충을 성충까지 사육한 결과 성충 나방의 눈색이 변화된 돌연변이 나방을 획득하였다. 정상 나방의 눈은 흑색인 반면 KMO 유전자의 변이가 유도된 나방의 경우 진한 회색을 나타내었다 (그림 3(a)). 이러한 결과를 통해 CRISPRdirect 서버를 통하여 선발된 가이드 RNA가 유전자 편집에 성공적으로 이용될 수 있음을 확인하였다.



(a) Eye color of wild type and mutant *B. mori*



(b) Sequence analysis of *B. mori*

그림 3. BmKMO 유전자 편집에 의한 돌연변이 발생
Fig. 3. Guide RNA design for BmKMO gene editing mutations of BmKMO gene induced by CRISPR/Cas9 system

선발된 가이드 RNA의 표적 선택 특이성을 검증하기 위하여 그림 3(b)에서 누에의 게놈 내 표적 영역과 잠재적인 off-target으로 예상되는 영역에 대하여 염기서열을 분석하였다. 누에 10번 염색체상에 존재하는 KMO 유전자 내부 표적 서열인 20 mer+TGG 영역의 염기서열 분석 결과 돌연변이 누에에서 58개의 염기가 결실됨이 확인되었다. 반면 off-target 가능성이 존재하는 10번 및 23번 염색체에 12 mer+NGG 영역에서는 정상 누에와 동일한 염기서열이 확인되었으며, 1번부터 28번 염색체 내에 나타나는 8 mer+NGG 영역에 대하여 염기서열을 분석한 결과 역시 정상 누에와 일치하여 off-target이 발생하지 않았음을 확인하였다. 이러한 결과는 CRISPRdirect 서버에 의해 선발된 가이드 RNA가 off-target 효과를 최소화함으로써 효율적인 유전자 편집을 가능하게 한다는 것을 시사한다.

Ⅴ. 결 론

본 논문에서는 유전자 편집을 통한 분자육종에 이용되는 가이드 RNA를 개발에 있어 생물체의 게놈 big data의 유용성에 대하여 실험적 검증을 수행하였다. 가이드 RNA 제작은 상대적으로 간단하지만, 거대한 게놈 내에 유사한 염기서열의 존재 확률이 높아 off-target 효과에 의해 제어되지 못한 돌연변이 발생 가능성이 증가한다. 하지만 유전자 분석기술의 발달에 따라 생물체의 게놈 데이터베이스의 양적 확장이 이루어 짐에 따라 이를 적절히 활용한다면 비특이적 돌연변이 발생에 대한 손실 없이 분자육종이 가능하다고 판단된다.

실제로 본연구에서 big data를 활용한 CRISPRdirect 서버를 통하여 가이드 RNA를 제작하고, 누에 유전자 편집에 적용시켜 본 결과 잠재적인 off-target 부위에서의 돌연변이는 관찰되지 않았다. 따라서 big data를 활용한 육종연구는 연구자가 가진 제한된 유전자 정보를 최대한으로 활용할 수 있게 해주며, 흔하게 사용되는 육종 소재가 아니라 할지라도 단시간에 연구를 수행하게 하는 유용한 도구라고 할 수 있다.

References

[1] P. D. Kang, et al., "Breeding of new silkworm variety golden silk, a yellow cocoon color for spring rearing season," *Korean J. Seric. Sci.*, vol. 49, pp. 14-17, 2007.

www.dbpia.co.kr

[2] S. W. Cho, et al., “Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease,” *Nat. Biotechnol.*, vol. 31, pp. 230-232, 2013.
[3] M. H. Harrison, et al., “A CRISPR view of development,” *Genes Dev.*, vol. 28, pp. 1859-1872, 2014.
[4] J. A. Doudna and E. Charpentier, “The new frontier of genome engineering with CRISPR-Cas,” *Science*, vol. 346, pp. 1078-1086, 2014.
[5] Y. Fu, et al., “Improving CRISPR-Cas nuclease specificity using truncated guide RNAs,” *Nat Biotechnol.*, vol. 32, pp. 279-284, 2014.
[6] Y. Naito, et al., “CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites,” *Bioinformatics*, vol. 31, pp. 1120-1123, 2015.
[7] P. D. Hsu, et al., “DNA targeting specificity of RNA-guided Cas9 nucleases,” *Nat. Biotechnol.*, vol. 31, pp. 827-832, 2013.
[8] F. Heigwer, et al., “E-CRISP: fast CRISPR target site identification,” *Nat. Methods*, vol. 11, pp. 122-123, 2014.
[9] H. Kikkawa, “Mechanism of pigment formation in Bombyx and Drosophila,” *Genetics*, vol. 26, pp. 587-607, 1941.
[10] K. Y. Kim, et al., “Biogenesis of lysosome-related organelle mutant silkworms by direct injection of a Cas9 protein-guided RNA complex into Bombyx mori embryos,” *J. Life Sci.*, vol. 29, pp. 537-544, 2019.

정 찬 영 (Chan Young Jeong)

2019년 8월 : 고려대학교 바이오시스템공학과 박사
2020년 2월~현재 : 국립농업과학원 연구원

홍 정 원 (Jeong Won Hong)

2015년 2월 : 조선대학교 바이오신약개발학과 박사
2018년 8월~현재 : 국립농업과학원 연구원

유 정 희 (Jeong Hee Yu)
2015년 8월 : 전북대학교 의과학과 석사
2015년 9월~현재 : 국립농업과학원 연구원

김 수 배 (Su Bae Kim)
2014년 2월 : 전남대학교 임산공학 석사
2014년 3월~현재 : 국립농업과학원 연구원

강 상 국 (Sang Kuk Kang)
2007년 2월 : 전남대학교 농학과
2014년 10월~현재 :국립농업과학원 연구원

김 성 완 (Seong Wan Kim)
2009년 2월 : 충남대학교 해부학 박사
2009년 3월~현재 : 국립농업과학원 연구원

김 남 숙 (Nam Suk Kim)
2000년 02월 : 방송통신대학교 농학과
2019년 01월~현재 : 국립농업과학원 연구원

김 기 영 (Kee Young Kim)
2004년 02월 : 동아대학교 식물생명공학 박사
2004년 11월~현재 : 국립농업과학원 연구원

박 종 우 (Jong Woo Park)

2014년 8월 : 조선대학교 의생명과학과 박사
2015년 3월~현재 : 국립농업과학원 연구원
<관심분야> 생물정보학, 분자생물학, 단백질공학
[ORCID:0000-0003-0649-9343]

www.dbpia.co.kr