

CRISPR 간섭에 필요한 sgRNA 표적 인식 서열 길이의 결정

김범준, 김병찬, 이호중, 이상준*
중앙대학교 시스템생명공학과

Received: November 8, 2021 / Revised: November 24, 2021 / Accepted: November 25, 2021

Determination of the Length of Target Recognition Sequence in sgRNA Required for CRISPR Interference

Bumjoon Kim, Byeong Chan Kim, Ho Joung Lee, and Sang Jun Lee*

Department of Systems Biotechnology, Chung-Ang University, Anseong 17546, Republic of Korea

Single-molecular guide RNA (sgRNA) plays a role in recognizing the DNA target sequence in CRISPR technology for genome editing and gene expression control. In this study, we systematically compared the length of the target recognition sequence in sgRNAs required for genome editing using Cas9-NG (an engineered Cas9 recognizing 5'-NG as PAM sequence) and gene expression control using deactivated Cas9-NG (dCas9-NG) by targeting the *gal* promoter in *E. coli*. In the case of genome editing, the truncation of three nucleotides in the target recognition sequence (TRS) of sgRNA was allowed. In gene expression regulation, we observed that target recognition and binding were possible even if eleven nucleotides were deleted from twenty nucleotides of the TRS. When 4 or more nucleotides are truncated in the TRS of the sgRNA, it is thought that the sgRNA/Cas9-NG complex can specifically bind to the target DNA sequence, but lacks endonuclease activity to perform genome editing. Our study will be helpful in the development of artificial transcription factors and various CRISPR technologies in the field of synthetic biology.

Keywords: CRISPR interference, single-molecular guide RNA, *gal* promoter, D-galactose

서 론

세균의 적응성 면역시스템으로 밝혀진 CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)는 그 단어가 의미하는 바와 같이 세균과 고세균의 유전체에서 발견되는 일정한 간격을 두고 주기적으로 분포하는 짧은 회문 반복서열로 처음 발견되었고[1], 세균의 유전체는 보호하면서 박테리오파지의 유전체 DNA 또는 RNA의 특정 서열을 인식하여 잘라내어 미생물을 보호할 수 있다[2].

여러 가지 CRISPR/Cas 시스템 중에서 지금까지 CRISPR/Cas9 시스템이 가장 활발하게 연구되었으며[3], 20개의 DNA 염기서열을 표적으로 인식하는 sgRNA (single-molecular guide RNA)와 서로 다른 가닥의 DNA를 각각 자를 수 있는 RuvC와 HNH 도메인을 가지고 있는 Cas9 핵산 분해 단백질 효소가 sgRNA/Cas9 복합체를 이루어 표적 DNA의 이중

가닥을 절단할 수 있다[4].

CRISPR/Cas system은 특정 서열의 DNA를 자를 수 있어서, 10여 년 전 유전체 공학의 도구로서 가능성이 제시된 이후, 꾸준히 발전해 왔다[5]. 적절한 Donor DNA를 제공한다면, 원하는 서열로 완벽하게 유전체를 편집할 수 있는 방법이 미생물에서 발전해 왔다[6]. 하지만, 면역시스템이 갖춘 자기(self)와 비자기(non-self)를 구분할 수 있도록 진화되어 온 PAM (protospacer adjacent motif) 서열의 존재로 인해 표적서열로 지정 가능한 범위가 제한될 수 있다[7]. PAM에 의한 표적서열 제한을 완화시키기 위해 Cas9 단백질의 7개 아미노산의 점 돌연변이를 일으켜 원래 Cas9의 PAM 서열인 5'-NGG의 3번째 구아닌(guanine)에 대한 의존도를 없애서 PAM 서열을 5'-NG로 인식하도록 하여 표적 DNA 서열을 확장한 Cas9-NG가 개발되었다[8].

이중가닥 절단없이 원하는 표적 DNA에 결합하여 유전자의 발현을 억제할 수 있는 방법인 CRISPRi (CRISPR interference)가 개발되었다[9]. 구체적으로, Cas9의 핵산 분해 효소 활성을 점 돌연변이를 도입하여 제거한 dCas9 (deactivated Cas9)을 사용해 원하는 특정 유전자의 발현을

*Corresponding author

Tel.: +82-31-670-3356, Fax: +82-31-675-3108
E-mail: sangjlee@cau.ac.kr

억제할 수 있다[10]. 표적 DNA 서열에 결합된 sgRNA/dCas9 복합체는 전사인자 또는 RNA 중합효소의 결합을 경쟁적으로 방해하고, 해당 유전자의 발현을 억제할 수 있다 [11].

이전 연구들에 의하면 Cas9의 경우 표적 DNA의 절단을 위해 sgRNA에서 17-18개 뉴클레오티드로 이루어진 표적인 식서열이 필요하였다[12, 13]. Cas9-NG는 Cas9과 비교하여 동일한 표적 DNA의 절단에 필요한 sgRNA 표적인식서열 길이의 차이를 보였다[13]. 한편, dCas9의 경우에는 14개의 뉴클레오티드로 이루어진 표적인식서열로 전사억제가 가능하다는 보고가 있었다[14].

본 연구에서는 확장된 PAM 서열을 갖는 dCas9-NG를 이용하여 대장균 모델에서 갈락토스 오페론(galactose operon) 프로모터인 P_{gal} 의 -10 영역을 표적서열로 인식하는 다양한 길이의 sgRNA들을 제작하여 전사억제를 하는지를 분석하였고, 또한, 동일한 표적인식서열에 sgRNA/Cas9-NG 복합체가 유전체 편집을 할 수 있는지를 비교하였다. dCas9-NG가 표적서열에 결합하는데 필요한 sgRNA의 길이와 Cas9-NG가 표적서열을 자르는데 필요한 sgRNA의 길이를 체계적으로 비교하였다.

재료 및 방법

균주와 배양 조건

Escherichia coli K-12 DH5 α 균주는 클로닝 숙주로 이용하였고, MG1655 균주는 유전체 편집 및 CRISPRi를 비교하기 위한 실험에 사용하였다(Table 1). DH5 α 균주는 LB 배지에서 배양되었으며 sgRNA 벡터 제작에 사용하였다. CRISPRi 실험에서 MG1655 균주는 LB 배지에 전배양 한 후 속신산(Sodium succinate dibasic, Sigma-Aldrich, Cat. No. 14160, USA)이 유일탄소원으로 첨가된 M9 최소배지로 옮겨 본배양하였다.

D-갈락토스(D-galactose, Samchun chemicals, Cat. No. G0476, Korea)를 0.5% 첨가한 맥콘키(MacConkey, BD Difco, Cat. No. 281810, USA) 아가배지에 도말을 통해 균주가 유일한 탄소원인 D-갈락토스를 섭취할 수 있는지 확인하였다. 또한, 본배양시 L-아라비노스(L-arabinose, TCI, Cat. No. A0515, Japan) 1 mM을 첨가하여 균주의 유전체에 있는 *cas9-NG* 또는 *dcas9-NG* 유전자를 각각 발현시켰다. 배지의 목적에 따라 항생제 암피실린(ampicillin) 50 μ g/ml, 카나마이신(kanamycin) 25 μ g/ml, 그리고 스펙티노마이신

Table 1. Strains and plasmids used in this study.

Name	Characteristics	Source/reference
Strain		
DH5 α	<i>fhuA2 lac(del)U169 phoA glnV44 Φ80' lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17</i>	Laboratory stock
MG1655	F ⁻ <i>ilvG rfb-50 rph-1</i>	S. Adhya
HK1159	MG1655, Δ <i>araBAD::P_{BAD}-cas9-NG-KmR</i>	[13]
HK1160	MG1655, Δ <i>araBAD::P_{BAD}-dcas9-NG-KmR</i>	[15]
Plasmid		
pHK463	pSC101 <i>ori^{ts}, araC, λ bet gene, AmpR</i>	[16]
pBJ005	pBR322 <i>ori</i> , sgRNA target length: 20 nt(⁻³⁰ TTCGCATCTTTGTTATGCTA ⁻¹¹ in P_{gal}), SpR	[17]
pBJ049	pBR322 <i>ori</i> , sgRNA target length: 19 nt(⁻²⁹ TCGCATCTTTGTTATGCTA ⁻¹¹ in P_{gal}), SpR	This study
pBJ048	pBR322 <i>ori</i> , sgRNA target length: 18 nt(⁻²⁸ CGCATCTTTGTTATGCTA ⁻¹¹ in P_{gal}), SpR	This study
pBJ047	pBR322 <i>ori</i> , sgRNA target length: 17 nt(⁻²⁷ GCATCTTTGTTATGCTA ⁻¹¹ in P_{gal}), SpR	This study
pBJ046	pBR322 <i>ori</i> , sgRNA target length: 16 nt(⁻²⁶ CATCTTTGTTATGCTA ⁻¹¹ in P_{gal}), SpR	This study
pBJ045	pBR322 <i>ori</i> , sgRNA target length: 15 nt(⁻²⁵ ATCTTTGTTATGCTA ⁻¹¹ in P_{gal}), SpR	This study
pBJ044	pBR322 <i>ori</i> , sgRNA target length: 14 nt(⁻²⁴ TCTTTGTTATGCTA ⁻¹¹ in P_{gal}), SpR	This study
pBJ043	pBR322 <i>ori</i> , sgRNA target length: 13 nt(⁻²³ CTTTGTTATGCTA ⁻¹¹ in P_{gal}), SpR	This study
pBJ042	pBR322 <i>ori</i> , sgRNA target length: 12 nt(⁻²² TTGTTATGCTA ⁻¹¹ in P_{gal}), SpR	This study
pBJ041	pBR322 <i>ori</i> , sgRNA target length: 11 nt(⁻²¹ TGTATGCTA ⁻¹¹ in P_{gal}), SpR	This study
pBJ040	pBR322 <i>ori</i> , sgRNA target length: 10 nt(⁻²⁰ TGTTATGCTA ⁻¹¹ in P_{gal}), SpR	This study
pBJ039	pBR322 <i>ori</i> , sgRNA target length: 9 nt(⁻¹⁹ GTTATGCTA ⁻¹¹ in P_{gal}), SpR	This study
pBJ038	pBR322 <i>ori</i> , sgRNA target length: 8 nt(⁻¹⁸ TTATGCTA ⁻¹¹ in P_{gal}), SpR	This study
pBC006	pBR322 <i>ori</i> , sgRNA target length: 7 nt(⁻¹⁷ TATGCTA ⁻¹¹ in P_{gal}), SpR	This study
pBC005	pBR322 <i>ori</i> , sgRNA target length: 6 nt(⁻¹⁶ ATGCTA ⁻¹¹ in P_{gal}), SpR	This study

Table 2. Primers used in this study.

Name	Sequence (5'→3')	Description
P1	TCGCATCTTTGTTATGCTAGTTTTAGAGCTAGAAATAGCAAG	pBJ049 construction
P2	TAGCATAACAAAGATGCGAACTAGTATTATACCTAGGACTG	
P3	CGCATCTTTGTTATGCTAGTTTTAGAGCTAGAAATAGCAAG	pBJ048 construction
P4	TAGCATAACAAAGATGCGACTAGTATTATACCTAGGACTG	
P5	GCATCTTTGTTATGCTAGTTTTAGAGCTAGAAATAGCAAG	pBJ047 construction
P6	TAGCATAACAAAGATGCACTAGTATTATACCTAGGACTG	
P7	CATCTTTGTTATGCTAGTTTTAGAGCTAGAAATAGCAAG	pBJ046 construction
P8	TAGCATAACAAAGATGACTAGTATTATACCTAGGACTG	
P9	ATCTTTGTTATGCTAGTTTTAGAGCTAGAAATAGCAAG	pBJ045 construction
P10	TAGCATAACAAAGATACTAGTATTATACCTAGGACTG	
P11	TCTTTGTTATGCTAGTTTTAGAGCTAGAAATAGCAAG	pBJ044 construction
P12	TAGCATAACAAAGAACTAGTATTATACCTAGGACTG	
P13	CTTTGTTATGCTAGTTTTAGAGCTAGAAATAGCAAG	pBJ043 construction
P14	TAGCATAACAAAGACTAGTATTATACCTAGGACTG	
P15	TTTGTATGCTAGTTTTAGAGCTAGAAATAGCAAG	pBJ042 construction
P16	TAGCATAACAAAAGCTAGTATTATACCTAGGACTG	
P17	TTGTTATGCTAGTTTTAGAGCTAGAAATAGCAAG	pBJ041 construction
P18	TAGCATAACAAAGCTAGTATTATACCTAGGACTG	
P19	TGTTATGCTAGTTTTAGAGCTAGAAATAGCAAG	pBJ040 construction
P20	TAGCATAACAAGCTAGTATTATACCTAGGACTG	
P21	GTTATGCTAGTTTTAGAGCTAGAAATAGCAAG	pBJ039 construction
P22	TAGCATAACACTAGTATTATACCTAGGACTG	
P23	TTATGCTAGTTTTAGAGCTAGAAATAGCAAG	pBJ038 construction
P24	TAGCATAAACTAGTATTATACCTAGGACTG	
P25	TATGCTAGTTTTAGAGCTAGAAATAGCAAGTAAAATAAGG	pBC006 construction
P26	TAGCATAACTAGTATTATACCTAGGACTGAGCTAGCTGTCA	
P27	ATGCTAGTTTTAGAGCTAGAAATAGCAAGTAAAATAAGGC	pBC005 construction
P28	TAGCATACTAGTATTATACCTAGGACTGAGCTAGCTGTCAA	
P29	GATACTGGCCGGCAGGCGCTCCATTGCC	Construction of sgRNA plasmids
P30	GCAATGGAGCGCCTGCCGGCCAGTATCAG	
P31	AACGCCTGGTATCTTTATAGTCCTGTCG	Sequencing primers for sgRNA
P32	CAGCAGCCGCGTAATAC	
P33	ACCAGGGTATCTAATCCTGT	qPCR primers for 16S rRNA
P34	CCACCGTTTATGGCGATCAG	
P35	GTTCCACCATCAGCTTGCTT	qPCR primers for <i>galE</i>
P36	ACTGGCAAATTGTCGGGCCA	
P37	TGTCACACTTTTCGCATCTTACGGTACCAATGGTTATTCATACCATAAG	Mutagenic oligonucleotides for <i>Pgal</i>

(spectinomycin) 75 µg/ml을 첨가해서 사용하였다.

플라스미드 제작

본 연구에 사용된 프라이머는 Table 2에 정리하였다. 갈락토스 오페론의 프로모터를 표적서열로 갖는 sgRNA 발현

플라스미드는 다음과 같이 제작하였다. 플라스미드 pBJ005를 주형으로 표적서열의 길이에 따라 설계된 프라이머 쌍을 사용해 양 끝에 서로 겹치는 서열이 존재하는 DNA 두 조각을 PCR에 의해 증폭시킨 후, Gibson assembly Master Mix (NEB, Cat. No. E2611, USA)를 이용하여 DNA 두 조각을

연결시켜 여러 가지 sgRNA 발현 플라스미드를 제작하였다. 이후 제작된 sgRNA 플라스미드는 유전체 편집과 CRISPRi를 위해 각각 제작된 대장균 균주 HK1159와 HK1160으로 도입되었다.

유전체 편집

HK1159 균주에 pHK463 플라스미드를 형질전환한 균주를 LB 배지에서 배양하면서 OD_{600nm}가 0.3에서 L-아라비노스를 배지에 첨가하여 Cas9-NG 단백질과 람다 Bet 단백질을 과발현시켜 유전체 편집을 위한 competent 세포를 준비하였다[15]. 제한효소 KpnI 서열을 갖는 돌연변이 유발 올리고뉴클레오티드와 여러 가지 표적인식서열의 길이를 가진 sgRNA 플라스미드를 함께 전기천공을 통하여 위에 준비한 competent 세포에 도입하였다. 형질전환된 세포들은 D-갈락토스를 함유한 맥콘키 아가 배지에 도말하여 흰색 콜로니의 비율과 생존한 콜로니(log₁₀CFU/μg DNA) 측정을 통해 유전체 편집 효율을 확인하였다. 흰색 콜로니에서 표적서열을 PCR 후 KpnI 제한효소를 처리하여 유전체 편집되었음을 확인하였다.

유전자 발현조절

표적인식서열의 길이가 다른 sgRNA 플라스미드를 HK1160 균주에 형질전환시킨 후, 스펙티노마이신(최종 75 μg/ml)과 D-갈락토스(최종 0.5%)를 첨가한 맥콘키 아가 배지에 선상도말(streaking)을 하였다. 또한, L-아라비노스(최종 1 mM)을 추가적으로 첨가한 배지에도 같은 세포를 선상도말을 하였다. dCas9-NG는 L-아라비노스에 의해 유도 발현이 되며, sgRNA는 항시발현 프로모터에 의해 발현된다. 따라서, L-아라비노스가 존재하면 세포내에서 sgRNA/dCas9-NG의 복합체가 만들어져 유전체의 표적서열인 gal 프로모터 -10 부위에 특이적으로 결합할 수 있도록 디자인되었다. 갈락토스 대사유전자의 전사가 억제될 경우, D-갈락토스를 대사할 수 없어서 맥콘키 D-갈락토스 배지에서 흰색 콜로니가 형성될 것이고, 갈락토스 대사유전자의 전사억제를 하지 못할 경우에는 붉은색 콜로니가 형성되도록 고안하였다.

대사물질 분석

맥콘키 아가 배지에서 형성된 콜로니의 색깔로써 gal 프로모터에 대한 전사억제 가능 유무를 먼저 판별한 균주를 LB-스펙티노마이신 액체배지에서 37°C, 15시간동안 전배양한 후, 250 ml 플라스크에 숙신산(최종 20 mM)과 스펙티노마이신을 첨가한 M9 최소배지 25 ml에 전배양액 0.25 ml을 첨가(최종 1%)하여 본배양을 진행하였다. 또 다른 본배양 플라스크에는 L-아라비노스(최종 1 mM)를 첨가하여 *dcas9-NG* 유전자 발현을 유도하였다. 세포 성장을 분광광도계(Ultrospec

8000 spectrophotometer, GE Healthcare, Sweden)를 사용하여 3시간 간격으로 측정하였다. 본배양 시작 후 6시간에 D-갈락토스(최종 20 mM)를 첨가하여 gal 프로모터의 전사억제 여부에 의한 갈락토스 대사여부를 모니터링하였다. 소모된 D-갈락토스의 농도는 액체크로마토그래피(liquid chromatography, RID-10A RI monitor, Shimadzu, Japan)와 컬럼(Aminex HPX-87H column 300 × 7.8 mm, BioRad)을 사용하여 측정하였으며 D-갈락토스의 대사율은 배양을 시작한 후 9시간과 12시간 사이의 섭취량을 시간으로 나누어 계산하였다[16].

전사 분석

본배양 시작 후 9시간에 Total RNA purification kit (Nanohelix, Cat. No. RPT100, Korea)를 사용하여 배양액으로부터 RNA를 추출하였다. 전사 분석에 필요한 galE 프라이머는 Universal Probe Library Assay Design Center (<http://lifescience.roche.com>)에서 디자인하였다. 추출한 RNA와 5 ng의 프라이머를 혼합한 후에 RealHelix qPCR kit (Nanohelix, Cat. No. QP2-S500, Korea)를 사용하여 galE 유전자의 전사를 CFX96 Touch (Bio-Rad, USA)를 통해 측정하였다. RT-qPCR 반응 조건은 다음과 같다. cDNA 합성(50°C, 40분); 변성(95°C, 12분); 증폭(95°C, 20초; 60°C, 1분); 녹는 곡선 측정 (65°C-95°C: 0.5°C씩 증가, 5초). 형광 데이터 결과는 16S 리보솜 RNA 발현 정도에 의해 정규화되었으며, L-아라비노스가 존재하는 조건의 galE(시작 코돈으로부터 374-453 염기서열 정도 떨어진 위치)의 mRNA의 양과 L-아라비노스가 존재하지 않는 조건의 mRNA의 양을 나눠서 계산하였다.

결과 및 고찰

Cas9-NG의 유전체 편집에 필요한 sgRNA 표적인식서열의 길이

sgRNA의 표적인식서열이 짧아지더라도 sgRNA/Cas9-NG 복합체의 엔도뉴클레아제 활성은 유지되어 유전체 편집이 가능하다는 것이 알려져 있어[12], Cas9-NG가 DNA 절단에 필요한 표적서열 길이의 최솟값을 분석하였다. gal 프로모터 P₁, P₂의 중복되는 -10 영역에서 절단이 발생하도록 표적인식서열을 갖는 sgRNA를 디자인하였다. sgRNA의 표적인식서열의 5'-말단부터 뉴클레오티드가 1개씩 제거하여 최대 14 개의 뉴클레오티드가 제거되도록 설계하였다. 유전체 편집 효율을 확인하기 위해 제한효소 KpnI 서열이 포함된 돌연변이 올리고뉴클레오티드와 여러 가지 sgRNA 플라스미드를 HK1159 균주에 전기천공법을 이용하여 도입하였다(Fig. 1A). 결과로, sgRNA의 표적인식서열의 5'-말단에서 뉴클레

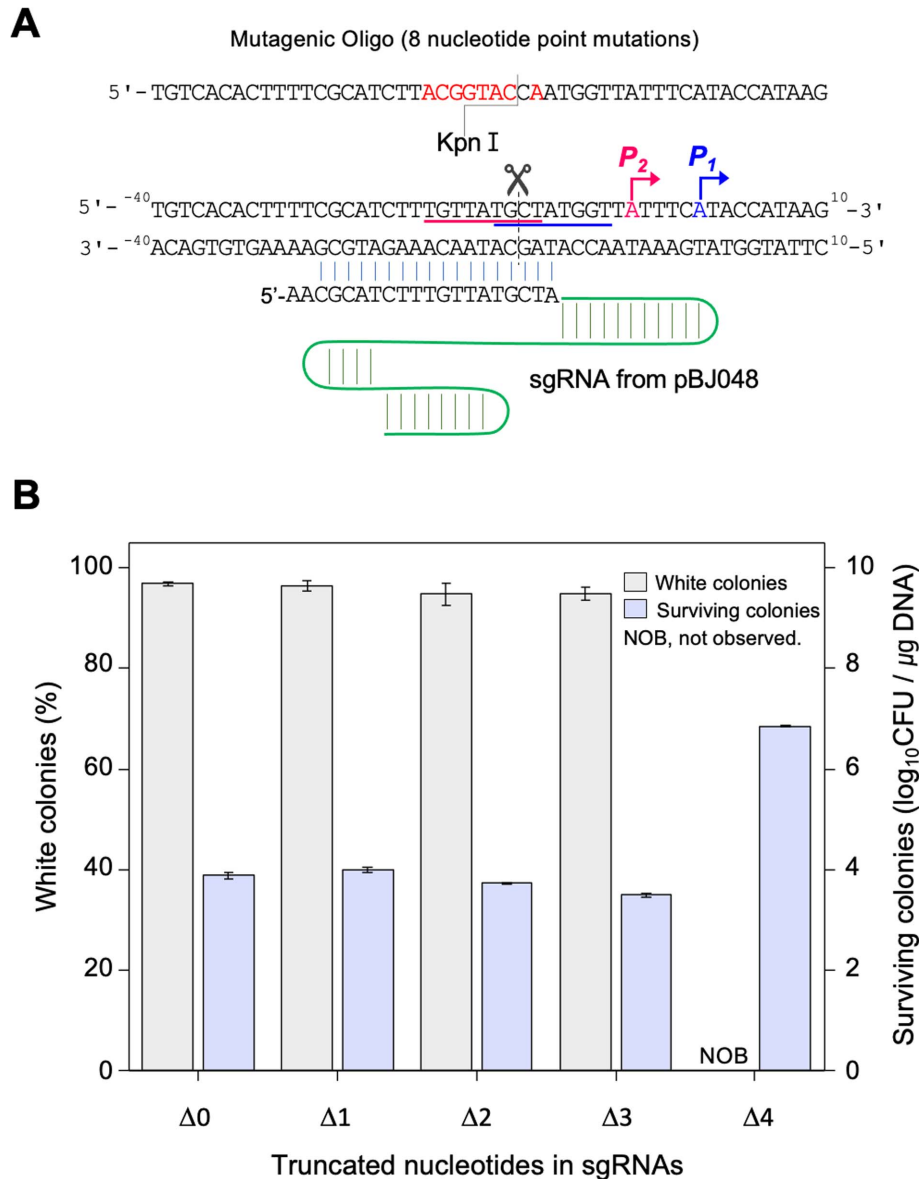


Fig. 1. Genome editing with CRISPR/Cas9-NG negative selection using various length of 5'-truncated sgRNAs. (A) Nucleotide sequences of mutagenic oligonucleotide and target DNA region for genome editing. The mutagenic oligonucleotide harbors the KpnI site. Red and blue underlines indicate -10 regions of *gal* P_1 and P_2 promoters, respectively. (B) Genome editing efficiencies using Cas9-NG with untruncated sgRNA ($\Delta 0$) and 5'-truncated sgRNAs ($\Delta 1$ to $\Delta 4$). The white colonies indicate cells with the *gal* promoter inactivated by genome editing.

오티드가 1, 2, 3개까지 결손되었을 경우(sgRNA $\Delta 1$, $\Delta 2$, $\Delta 3$)에는 뉴클레오티드가 결손되지 않은 sgRNA와 같이 돌연변이의 결과로 *gal* 프로모터가 망가져서 D-갈락토스의 대사 유전자의 발현이 정상적으로 일어나지 않아서 맥콘키(MacConkey) D-갈락토스 아가배지에 형성되는 흰색 콜로니가 90% 이상의 비율로 관찰되었다. 하지만, sgRNA의 표적 인식서열의 5'-말단에서 뉴클레오티드가 4개가 결손되었을

경우(sgRNA $\Delta 4$)에는, 돌연변이 올리고뉴클레오티드를 이용하여 *gal* 프로모터에 대한 유전체 편집을 할 수 없었다. 또한 sgRNA $\Delta 4$ 경우에는 맥콘키(MacConkey) D-갈락토스 배지에 살아남은 미생물의 콜로니 개수가 sgRNA $\Delta 1$, $\Delta 2$, $\Delta 3$ 의 경우에서 관찰된 $10^4/\mu\text{g DNA}$ 이하 수준에서 $10^6/\mu\text{g DNA}$ 이상으로 크게 증가하는 것으로 관찰되었다. 또한, sgRNA $\Delta 4$ 가 사용되었을 경우에는 유전체 편집되었음을 알 수 있는 표

현형인 흰색 콜로니가 전혀 관찰되지 않았다. 이 결과로, sgRNA Δ 1, Δ 2, Δ 3는 Cas9-NG과 복합체를 이루어 유전체 편집에 필요한 엔도뉴클레아제 활성을 유지할 수 있지만, sgRNA Δ 4는 Cas9-NG 복합체는 엔도뉴클레아제 활성이 현저하게 줄어들어 세포내에서 유전체편집을 할 수 없음을 알

수 있다(Fig. 1B).

dCas9-NG의 유전자 발현 조절이 가능한 sgRNA 표적 인식 서열의 길이

dCas9을 사용한 CRISPRi에 필요한 sgRNA의 최소 뉴클

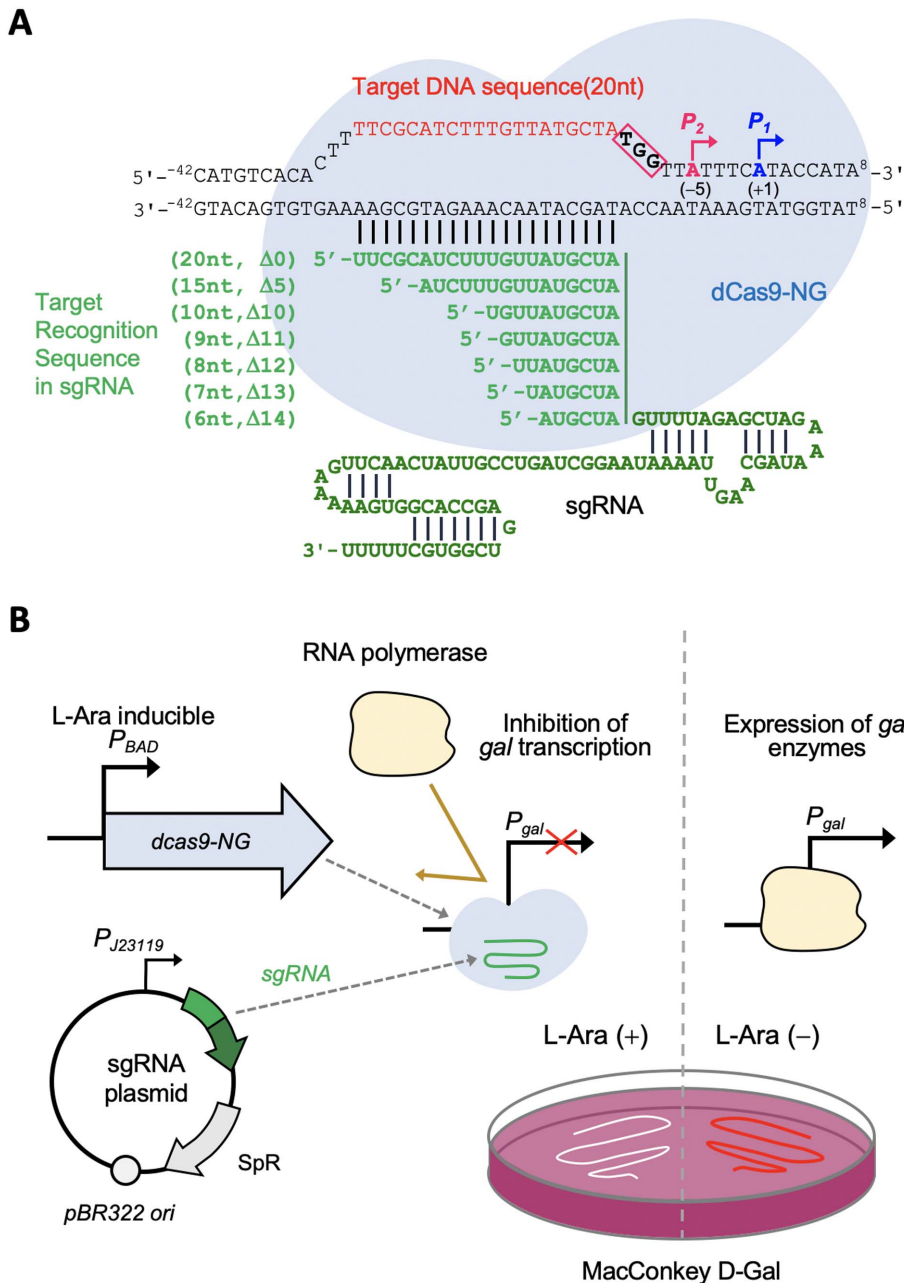


Fig. 2. Design of regulation of *gal* promoter by CRISPR interference with 5'-truncated sgRNAs. (A) dCas9-NG protein with various lengths of 5'-truncated sgRNAs targeting -10 regions of *gal* P_1 and P_2 promoters. (B) The expression of dCas9-NG in the chromosome of HK1160 strain is induced by L-arabinose. The sgRNA/dCas9-NG complex can bind to the -10 region of the *gal* promoter to inhibit the expression of D-galactose metabolizing enzymes. If D-galactose is not metabolized, white colonies form on the MacConkey agar plates.

레오티드의 개수는 12개로 알려졌다[11]. 우리는 dCas9-NG를 활용한 유전자의 전사 억제를 할 수 있는 CRISPRi 시스템에서 표적서열을 인식하는 기능을 하는 sgRNA의 표적인식길이 표적서열에 결합에 미치는 영향을 조사하기 위한 방법을 디자인하였다(Fig. 2). 우선, 표적인식서열 20개를 갖는 sgRNA 플라스미드와 표적인식서열이 뉴클레오티드1개씩 짧아진 sgRNA 플라스미드를 HK1160 균주에 형질전환하였다(Fig. 2A). 여러 가지 형질전환된 미생물 균주들을 맥콘키 D-갈락토스 배지에 선상도말하여 흰색과 붉은색 콜로니 표현형을 각각 전사억제의 유무로 판단하였다(Fig. 2B). dCas9-NG 발현의 유도인자인 L-아라비노스가 함유되지 않은 맥콘키 D-갈락토스 배지에서 성장한 미생물 균주들은 활발하게 D-갈락토스를 대사하여 붉은색 콜로니를 형성하였다. 하지만, L-아라비노스를 함유한 맥콘키 D-갈락토스 배지에서 성장한 미생물 균주들의 콜로니 색깔은 sgRNA의 표적인식서열의 길이에 따라 차이가 있었다. sgRNA의 표적인식서열에서 뉴클레오티드 결손이 1개(sgRNA Δ 1)에서 11개(sgRNA Δ 11)까지 이루어졌음에도 D-갈락토스의 대사에 필요한 유전자가 발현되지 못하도록 *gal* 프로모터를 억제할 수 있었다(Fig. 3). 하지만, sgRNA의 표적인식서열의 뉴클레오티드의 결손 개수가 12개(sgRNA Δ 12) 이상이 되면, *gal* 프로모터가 억제되지 않았을 때 보이는 붉은색 콜로니로 관찰되

었다(Fig. 3).

세포의 대사 조절을 할 수 있는 CRISPRi의 sgRNA 표적인식서열의 길이

sgRNA의 표적인식서열 길이에 따른 HK1160 균주의 *gal* 프로모터의 전사 억제가 D-갈락토스의 대사에 미치는 영향을 확인하기 위해서, 숙신산이 함유된 M9 최소배지에 배양하던 세포에 D-갈락토스를 추가 탄소원으로 첨가했을 때 D-갈락토스를 소모하는 지를 모니터링하였다(Fig. 4). dCas9-NG의 발현을 위해서 L-아라비노스는 배양 초기에 첨가하였다. L-아라비노스의 첨가 여부, 즉 CRISPRi의 작동 여부에 따른 세포 성장속도(cell growth rate; OD_{600nm}/h)를 비교하였다(Fig. 4B). 본배양 후 6시간 뒤인 초기 지수기인 배양액의 OD_{600nm}가 0.3이 되는 시점에서 D-갈락토스를 최종 농도가 20 mM이 되도록 첨가하였다. 그 결과, L-아라비노스를 첨가하지 않은 배지에서는 sgRNA의 표적인식서열의 길이와 상관없이 추가로 첨가한 D-갈락토스를 탄소원으로 이용하여 본배양 15시간째에 모두 OD_{600nm}가 3이상으로 세포가 자라는 것을 관찰하였다(Fig. 4A). 반면에 L-아라비노스를 첨가한 배지에서 sgRNA의 표적인식서열에서 뉴클레오티드를 11개 결손시킨 경우(sgRNA Δ 11)까지 HK1160 균주가 D-갈락토스를 소비하지 못하였다(Fig. 4C). 하지만, sgRNA의 표적인식서열에서 12개의 뉴클레오티드 이상을 결손시킨 경우(sgRNA Δ 12, Δ 13, Δ 14)에는 L-arabinose 유무와 관계없이 0.4(OD_{600nm}/h) 정도의 비슷한 세포성장 속도를 관찰하였다(Fig. 4B). D-갈락토스 대사속도(D-galactose consumption rate; mM/h)를 비교한 결과, L-아라비노스를 첨가한 배지에서 sgRNA의 표적인식서열에서 뉴클레오티드를 11개 결손시킨 경우(sgRNA Δ 11)까지 D-갈락토스가 대사되지 않고 배지에 그대로 남아 있음을 확인하였다(Fig. 4C).

무세포 전사-번역(cell-free transcription-translation) 시스템을 이용한 CRISPRi 연구에서는 sgRNA 표적인식서열에서 뉴클레오티드를 10개 결손시킨 경우까지는 결손이 없을 때와 비슷한 전사억제 수준을 보였으나 뉴클레오티드 12개를 결손시킨 경우부터 전사억제 수준이 50% 이하로 크게 감소하였다[18]. 해당 연구는 8-12 nt의 sgRNA 표적인식서열이 dCas9의 DNA 결합을 위한 구조적 변화(conformational shift)의 체크포인트인 것으로 설명하였다. 본 연구에서는 표적인식서열에서 염기 1개가 더 짧아진 sgRNA Δ 11까지 dCas9-NG와 복합체를 이루어 세포내에서 효과적으로 *gal* 프로모터의 전사를 억제할 수 있었으며, 전체 표적인식서열 20개의 염기서열에서 12개의 뉴클레오티드를 결손시킨 경우에는 표적서열을 인식할 수 없어 결합에 의한 전사억제를 할 수 없었다.

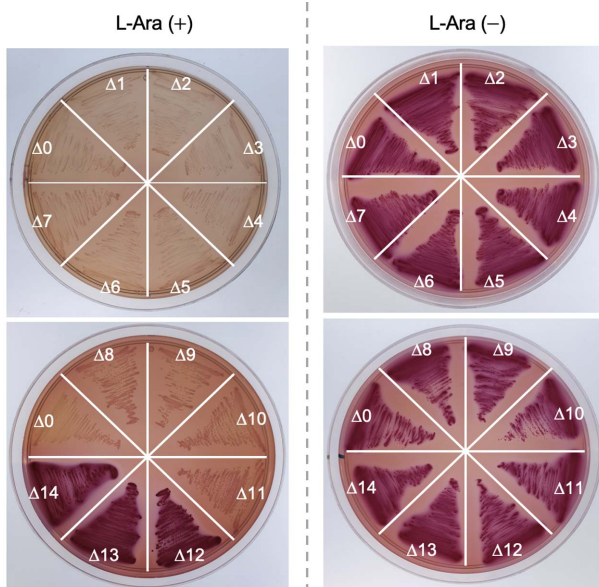


Fig. 3. Visualized CRISPR interference using various 5'-truncated sgRNAs on MacConkey agar. If CRISPRi works properly with 5'-truncated sgRNAs (Δ 1 to Δ 14), white colonies are formed on the MacConkey agar plates, otherwise red colonies are formed. The presence or absence of L-arabinose turns CRISPRi on or off, respectively.

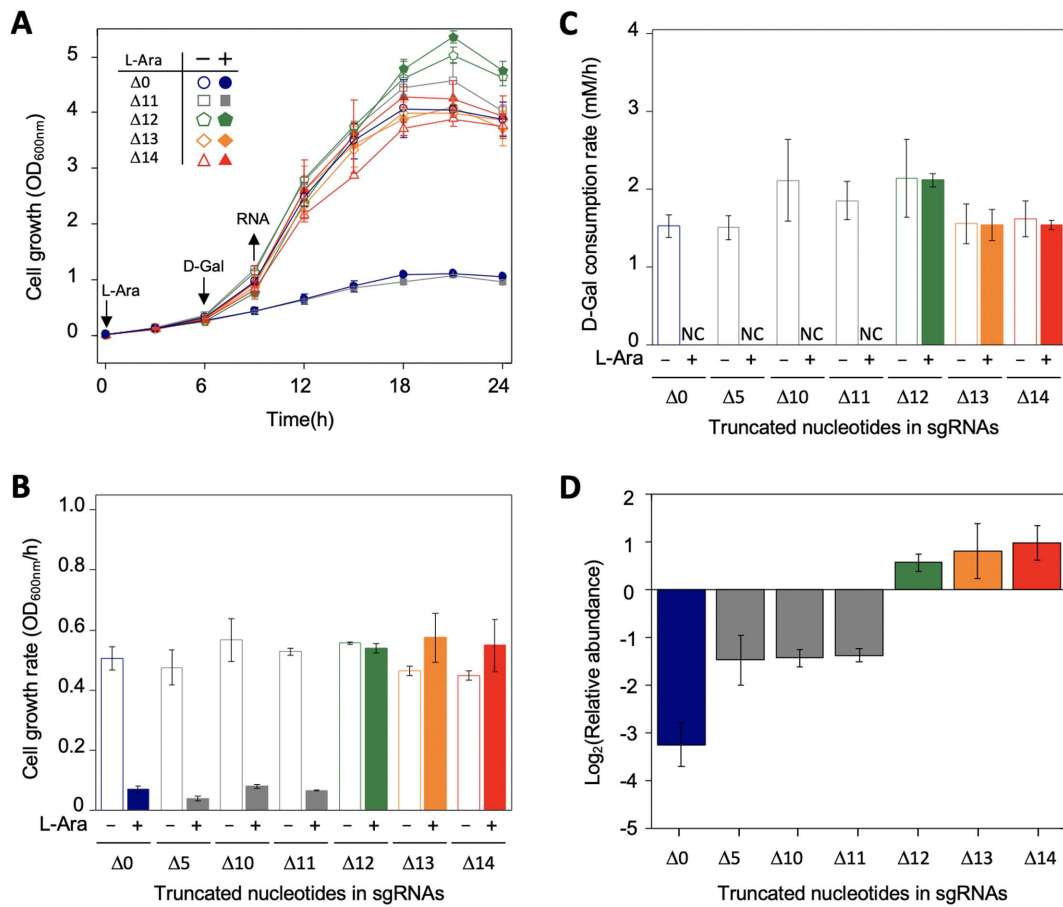


Fig. 4. Transcriptional and metabolic inhibition by various 5'-truncated sgRNA/dCas9-NG complexes. (A) Growth curves of HK1160 cells harboring different lengths of sgRNAs ($\Delta 0$, $\Delta 11$, $\Delta 12$, $\Delta 13$, and $\Delta 14$). (B) Cell growth rates, and (C) D-galactose consumption rates of HK1160 cells harboring different lengths of sgRNAs ($\Delta 0$, $\Delta 5$, $\Delta 10$, $\Delta 11$, $\Delta 12$, $\Delta 13$, and $\Delta 14$). L-arabinose induces the expression of dCas9-NG protein. sgRNAs are constitutively transcribed in the cell. (D) Relative *in vivo* expression levels of *galE* transcripts with and without L-arabinose. Values less than zero on the vertical axis indicate transcriptional repression by the 5'-truncated sgRNA/dCas9-NG complex.

sgRNA의 표적인식서열의 길이에 따른 dCas9-NG의 유전자 전사 조절

sgRNA의 표적인식서열의 길이가 *gal* 프로모터의 전사를 얼마나 억제하는 지를 정량적으로 분석하기 위해서, 본배양에서 D-갈락토스 첨가한 후 3시간을 더 배양한 후에 세포로부터 RNA를 추출해서 RT-qPCR 방법을 이용하여 *galE* 유전자의 전사된 mRNA 양을 비교하였다. 16S 리보솜 RNA의 발현 정도와 비교하여 정규화(normalization) 한 후 L-아라비노스 첨가에 따른 mRNA의 상대적 양(Relative abundance)을 계산하였다. 그 결과, sgRNA의 표적인식서열의 뉴클레오티드가 결손되지 않는 경우의 log₂ (Relative abundance) 값이 -3 이하로 *gal* 프로모터의 전사 억제를 가장 잘 하는 것으로 보였다. sgRNA의 표적인식서열의 뉴클레오티드의 결손이 $\Delta 5$, $\Delta 10$, $\Delta 11$ 일 경우에는 log₂ (Relative abundance)

값이 모두 -1.5 정도로 뉴클레오티드의 결손이 없는 sgRNA보다 약한 전사억제를 보였다. 하지만, sgRNA의 표적인식서열의 뉴클레오티드의 결손이 $\Delta 12$, $\Delta 13$, $\Delta 14$ 인 경우에는 L-아라비노스 유무에 따라 *galE* 유전자의 전사량의 변화가 없는 것으로 관찰되었다. 이것은 sgRNA의 표적인식서열의 뉴클레오티드의 결손이 11개까지 dCas9-NG와 복합체를 이루어 *gal* 프로모터의 전사를 억제할 수 있고, 뉴클레오티드의 결손이 12개 이상부터는 전사 억제를 할 수 없음을 보여준다(Fig. 4D). 결론적으로, 유전자 발현을 조절할 수 있는 CRISPRi에 필요한 sgRNA의 표적인식서열 길이를 결정할 수 있었고, 유전체 편집을 위한 sgRNA의 표적인식서열의 길이보다 훨씬 짧은 것을 알게 되었다. 표적인식서열이 짧아짐에 따라 두 가닥의 표적 DNA를 자를 수 있는 엔도뉴클레아제 활성을 먼저 잃었고, sgRNA와 표적 DNA 서열 사이의 특

이적인 결합은 유지할 수 있음을 밝혔다. 본 연구 결과는 CRISPRi 기술을 이용하는 데 도움이 될 뿐만 아니라, 앞으로 합성세포에서 작동하는 인공전사인자를 개발하는 데에도 중요한 정보를 제공할 것이다.

요약

CRISPR/Cas를 이용한 유전체 편집과 유전자 발현 조절을 위한 기술에서 sgRNA는 표적서열을 인식하는 역할을 한다. *gal* 프로모터를 표적서열로 하여 유전체 편집에 필요한 sgRNA의 표적인식서열의 길이와 유전자 발현 조절에 필요한 sgRNA의 표적인식서열의 길이를 Cas9-NG에서 체계적으로 비교하였다. 유전체 편집의 경우, sgRNA의 표적인식서열을 구성하는 20개의 뉴클레오티드에서 3개의 뉴클레오티드의 결손만을 허용하였다. 하지만, 유전자 발현 조절에는 표적인식서열에서 11개의 뉴클레오티드가 결손되어도 표적서열을 인식하고 결합할 수 있다는 것을 밝혔다. 따라서, sgRNA의 표적인식서열에서 4개 이상의 뉴클레오티드의 결손이 있는 경우에 sgRNA/Cas9-NG는 표적 DNA 서열에 특이적으로 결합을 하지만, 엔도뉴클레아제의 활성을 갖지 못하기 때문에 유전체 편집을 할 수 없는 것으로 판단된다. 이 결과는 인공전사인자 개발과 합성생물학 분야의 다양한 CRISPR 기술 발전에 도움을 줄 것이다.

Acknowledgments

This study was supported by National Research Foundation of Korea (NRF-2021R1A2C1013606), and Rural Development Administration (Project No. PJ015001032021), Republic of Korea.

Conflict of Interest

The authors have no financial conflicts of interest to declare.

References

- Jansen R, van Embden JDA, Gaastra W, Schouls LM. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**: 1565-1575.
- Horvath P, Barrangou R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**: 167-170.
- Lee HJ, Lee SJ. 2021. Advances in accurate microbial genome-editing CRISPR technologies. *J. Microbiol. Biotechnol.* **31**: 903-911.
- Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, et al. 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**: 1247997.
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816-821.
- Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**: 233-239.
- Cong L, Ran FA, Cox D, Lin SL, Barretto R, Habib N, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819-823.
- Nishimasu H, Shi X, Ishiguro S, Gao L, Hirano S, Okazaki S, et al. 2018. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**: 1259-1262.
- Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA. 2013. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41**: 7429-7437.
- Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. 2013. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8**: 2180-2196.
- Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**: 1173-1183.
- Fu YF, Sander JD, Reyon D, Cascio VM, Joung JK. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32**: 279-284.
- Lee HJ, Kim HJ, Lee SJ. 2021. Mismatch intolerance of 5'-truncated sgRNAs in CRISPR/Cas9 enables efficient microbial single-base genome editing. *Int. J. Mol. Sci.* **22**: 6457.
- Moghadam F, LeGraw R, Velazquez JJ, Yeo NC, Xu C, Park J, et al. 2020. Synthetic immunomodulation with a CRISPR super-repressor *in vivo*. *Nat. Cell Biol.* **22**: 1143-1154.
- Kim B, Kim HJ, Lee SJ. 2020. Effective blocking of microbial transcriptional initiation by dCas9-NG-mediated CRISPR interference. *J. Microbiol. Biotechnol.* **30**: 1919-1926.
- Lee HJ, Kim HJ, Lee SJ. 2020. CRISPR-Cas9-mediated pinpoint microbial genome editing aided by target-mismatched sgRNAs. *Genome Res.* **30**: 768-775.
- Kim B, Kim HJ, Lee SJ. 2020. Regulation of microbial metabolic rates using CRISPR interference with expanded PAM sequences. *Front. Microbiol.* **11**: 282.
- Khakimzhan A, Garenne D, Tickman B, Fontana J, Carothers J, Noireaux V. 2021. Complex dependence of CRISPR-Cas9 binding strength on guide RNA spacer lengths. *Phys. Biol.* **18**: 056003.