Development of CRISPR Tools for Metabolic Engineering of Specialized Metabolite Production in Streptomyces

Whitford, Christopher Martin

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

DTU

The Novo Nordisk Foundation
Center for Biosustainability

Development of CRISPR Tools for Metabolic Engineering of Specialized Metabolite Production in *Streptomyces*

Ph.D. Thesis

Christopher M. Whitford

Development of CRISPR Tools for Metabolic Engineering of Specialized Metabolite Production in *Streptomyces*

Ph.D. Thesis

Christopher M. Whitford

The Novo Nordisk Foundation Center for Biosustainability

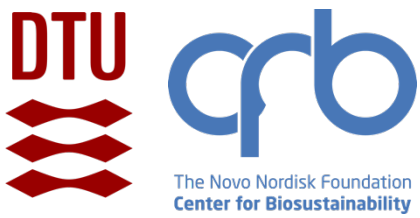
Technical University of Denmark

May 2023

Development of CRISPR Tools for Metabolic Engineering of Specialized Metabolite Production in *Streptomyces*

Ph.D. thesis written by Christopher M. Whitford
Supervised by Prof. Tilmann Weber
Research conducted in the Natural Products Genome Mining group

Novo Nordisk Foundation Center for Biosustainability
Technical University of Denmark



"Science is magic that works"
Kurt Vonnegut

"The imagination of nature is far,
far greater than the imagination of man"
Richard P. Feynman

"What is the use of a house
if you haven't got a tolerable planet to put it on?"
Henry David Thoreau

ASSESSMENT COMMITTEE

Professor Byung-Kwan Cho
Department of Biological Sciences and Institute of the BioCentury
Korean Advanced Institute of Science and Technology

Associate Professor Thomas Tørring
Department of Biological and Chemical Engineering
Aarhus University

Senior Researcher Michael Krogh Jensen
Novo Nordisk Foundation Center for Biosustainability
Technical University of Denmark

PREFACE

This thesis was written in partial fulfillment of the requirements to obtain the degree of Doctor of Philosophy at the Technical University of Denmark. The work presented in this thesis was carried out between January 2020 and May 2023 in the Natural Product Genome Mining group at the Novo Nordisk Center for Biosustainability. The work was supervised by Prof. Tilmann Weber, and co-supervised by Assoc Prof. Yaojun Tong from Shanghai Jiao Tong University, as well as Microbial Cell Factory Engineer Dr. Tetiana Gren from the Strain Design Team at the Novo Nordisk Center for Biosustainability. Additional work was carried out as part of the external stay at the Joint Bioenergy Institute of the Lawrence Berkeley National Lab, Berkeley, U.S.A., under the supervision of Prof. Jay Keasling.

This Ph.D. project was funded by the Novo Nordisk Foundation as part of the iimena challenge grant project, grant no. NNF16OC0021746. The external stay was further supported through grants from the Otto Mønsted and William Demant Foundations.

Christopher M. Whitford
Kgs. Lyngby, May 2023

ABSTRACT

Soil microorganisms make up the second largest fraction of biomass on Earth. The soil microcosm is incredibly diverse and encompasses a vast diversity of species, often participating in complex interspecies interactions, even across kingdoms of life. These interactions are often facilitated through specialized metabolites, molecules of fascinating complexity and chemical diversity. Specialized metabolites are of great importance for modern society, for example as antibiotics, antifungals, anticancer compounds, or plant growth-promoting agents. Of all soil bacteria, members of the genus *Streptomyces* were found to encode the largest biosynthetic potential by far. This biosynthetic potential is encoded in genes clustered on the genomes, and thus referred to as biosynthetic gene clusters (BGCs). However, under laboratory conditions, the majority of BGCs are not expressed, complicating the assignment of specialized metabolites to predicted BGCs. Activating, studying, and modifying BGCs, as well as strain engineering of production strains, relies heavily on our ability to efficiently engineer *Streptomyces* species. However, introduction of just one chromosomal modification can require months of laboratory work using classical methods. This Ph.D. thesis aimed to improve our metabolic engineering capabilities of *Streptomyces* species through the development of novel CRISPR-based tools, enabling engineering with unprecedented speed, efficiency, and throughput.

As a contextual starting point, this thesis provides a review of the Design-Build-Test-Learn cycle for metabolic engineering of *Streptomyces* species. For each cycle stage, important resources and considerations are highlighted. Furthermore, a protocol for all existing CRISPR tools for streptomycetes, including CRISPR-Cas9, CRISPRi, and CRISPR-BEST, is provided. Given the slow throughput of current engineering methods, multiplexed cytosine base editing was studied at various scales, revealing key performance parameters, and highlighting important bottlenecks. The system achieved the highest number of simultaneously edited targets in *Streptomyces* species using any technology, highlighting its great promise. Given the frequent need for full deletions of genomic regions, a novel CRISPR tool based on CASCADE-Cas3 was developed with superior efficiency to CRISPR-Cas9. This tool, the first developed for *Streptomyces* based on a type I CRISPR system, was shown to work efficiently in several species and was used for the streamlined

construction of a genome-minimized expression host. Finally, the first CRISPR-Prime tool for *Escherichia coli* was developed, serving as a stepping stone towards the establishment of CRISPR-Prime editing in *Streptomyces* species. This tool enables the introduction of complicated mutations and represents a platform for further tool development. The developed tools greatly expand the engineering capabilities for *Streptomyces* species, bringing us one step closer to closing the gap between computationally predicted, and experimentally characterized BGCs. This will soon allow us to greatly expand our access to Nature's biosynthetic potential, and ultimately urgently needed solutions for medicine, agriculture, and sustainability.

DANSK SAMMENFATNING

Mikroorganismer i jord udgør den næststørste biomasse på jorden efter planter. Biotoper i jord er utroligt mangfoldige og besidder en enorm artsdiversitet, der ofte deltager i komplekse interaktioner, selv på tværs af livets tre riger. Disse interaktioner udgøres ofte af specialiserede metabolitter, molekyler af fascinerende kompleksitet og med stor kemisk mangfoldighed. Specialiserede metabolitter er af stor betydning for det moderne samfund, f.eks. som antibiotika, svampemidler, kræftmidler eller plantevækstfremmende stoffer. Blandt jordbakterier er det blevet dokumenteret, at medlemmer af slægten *Streptomyces* koder for langt større biosyntetisk potentiale end nogen anden slægt. Dette biosyntetiske potentiale er indkodet i gener, der er grupperet i klynger i genomerne og derfor kaldes biosyntetiske gen-clusters (BGC'er). Under laboratorieforhold udtrykkes størstedelen af BGC'erne imidlertid ikke, hvilket gør det vanskeligt at identificere hvilke BGC'er der er ansvarlige for produktionen af hvilke kemiske stoffer. Aktivering, eksperimentering og genmodifikation af BGC'er samt udvikling af produktionsstammer afhænger i høj grad af vores evne til effektivt at arbejde med arter af *Streptomyces*. Introduktion af blot én kromosomal modifikation kan imidlertid kræve måneders laboratoriearbejde ved hjælp af klassiske metoder. Denne ph.d.-afhandling har til formål at forbedre vores muligheder for metabolisk engineering af Streptomyces-arter gennem udvikling af nye CRISPR-baserede værktøjer, der muliggør engineering med en hidtil uset hastighed og effektivitet.

Som et kontekstuelt udgangspunkt giver denne afhandling en gennemgang af Design-Build-Test-Learn-cyklussen for metabolisk engineering af *Streptomyces*-arter. For hver fase af cyklussen fremhæves vigtige ressourcer og overvejelser. Desuden gives en protokol for alle eksisterende CRISPR-værktøjer til streptomyceter, herunder CRISPR-Cas9, CRISPRi og CRISPR-BEST. Med afsæt i den langsomme hastighed som klassiske genmodificeringsmetoder har, blev multiplexed cytosinbaseediting undersøgt i forskellige skalaer, hvilket afdækkede vigtige præstationsparametre og fremhævede flaskehalse. Med metoder udviklet i denne afhandling har jeg opnået et større antal samtidige genmodifikationer i arter af *Streptomyces* end nogen tidligere metode har formået, hvilket viser det store potentiale for de udviklede metoder.

Da det i forbindelse med stammeudvikling ofte er nødvendigt at lave deletioner af hele genomiske regioner har jeg udviklet et nyt CRISPR-værktøj baseret på CASCADE-Cas3 med en effektivitet, der er bedre end CRISPR-Cas9. Dette værktøj, som er det første, der er udviklet til Streptomyces baseret på et type I CRISPR-system, viste sig at fungere effektivt i flere arter og blev anvendt til effektiv konstruktion af en genom-minimeret vært til heterolog ekspression. Endelig har jeg udviklet det første CRISPR-Prime-værktøj til Escherichia coli, hvilket har været et springbræt til etablering af CRISPR-Prime-redigering i arter af *Streptomyces*. Dette værktøj gør det muligt at indføre komplicerede mutationer og udgør en platform for yderligere værktøjsudvikling. De udviklede værktøjer udvider i høj grad de tekniske muligheder for Streptomyces-arter og mindsker afstanden mellem computerannoterede og eksperimentelt karakteriserede BGC'er. Dette vil i fremtiden gøre det muligt for os at øge vores adgang til naturens biosyntetiske potentiale og i sidste ende at finde løsninger indenfor medicin, landbrug og bæredygtighed, som der er presserende behov for.

ACKNOWLEDGMENTS

During my time as a Ph.D. student, I got to meet, work together, and learn from so many fantastic people. I can say with certainty that I would not have been able to make it this far without the continuous support, encouragement, and help from all of you. I once had a great conversation about whether life happens to you, or you happen to life. While this thesis is the result of 3 ½ years of hard work, it certainly would have been way harder, if not impossible, to achieve this without so many great people "happening to me". It is thanks to you that I got to grow both scientifically and personally during these last couple of years.

I would like to start by thanking my wonderful supervisors, Tilmann, Yaojun, and Tanya. Thank you Tilmann for giving me the chance to work on this Ph.D. project in your group. I greatly appreciate all the freedom you offered me to explore my own ideas, and all your support and trust over the last couple of years. Thank you for all the great feedback and input, for always having an open door and always finding time to answer questions, and for being such a supportive supervisor. Thank you for giving me the chance to prove myself, to attend inspiring conferences and meetings around the world, and to work together with so many fantastic people in your group!

Thank you Yaojun for all the fun collaborations and for teaching me so much about CRISPR and technology development. I am very glad we got to keep working together, even after you started your own group in Shanghai! Thank you for teaching me about the Chinese concept of "Liùbǎi", I thought about it a lot during the final stretches of my Ph.D.

Thank you, Tanya, for all your kind support over these last couple of years, for sharing your vast *Streptomyces* knowledge with me, and for always being willing to help with experiments! And of course, thank you for teaching me all the original *Streptomyces* methods and tricks!

While this Ph.D. thesis only covers the work of the last 3 ½ years, I do believe that this thesis is also a testament to all the fantastic support I received from previous supervisors and mentors at Bielefeld University and the Center for Biotechnology, at the Massachusetts Institute of Technology, and iGEM.

Therefore, I would like to thank Volker Wendisch, Elvira Sgobba, Anthony Sinskey, Christian Rückert, Boas Pucker & Julian Droste, and most importantly Jörn Kalinowski. Thank you Jörn for opening so many doors during my studies, and for ultimately being the one who started my Danish adventure!

Thank you to all current and former members of the New Bioactive Compounds / Natural Products Genome Mining group for the wonderful time. Working in such an interdisciplinary team made every day exciting and provided so many learning opportunities. But most importantly, working together with all of you truly made this Ph.D. the exciting and fun experience it was. I'd like to thank especially the sequencing team, aka Tue, Maria, Eva, Oliwia, and David for their help with sequencing so many of my strains, teaching me how to analyze the data, and being just fantastic, fun, and very helpful colleagues. Tue, thank you for being often the only one to chuckle at my bad jokes. Thank you to Pep for all the brainstorming and tough questions, to Kai and Simon for all the help with antiSMASH and bioinformatics-related problems, to Thom and Matiss for all the fun and an "eggcellent" time in the spookiest hotel ever, to Omkar and Matin for great help and feedback anything genome analytics related, to Efi for all the support and help in fighting my nemesis, aka metabolomics data, and Agnes for all the great presentation feedback and German jokes. Thank you to my students Olga and Dennis, I hope you learned as much from me as I did from supervising you! Thank you to Renata, Tetiana, Peter, and Zhijie for the collaboration and for being eager early adopters of my new tools. I'd also like to thank former group members Julie and Emil for all the great times! Having a beer after work with you was always equally as dangerous as it was fun!

Thank you to all my Danish colleagues for being willing to speak Danish with me at work, and for teaching me many new phrases, words, and even poems! Tusind tak til Tue, David, Eva, Anna-Sophie, og Julie! Det var en stor fornøjelse!

A big research center like CfB only runs smoothly because so many people work "behind the scenes". Therefore, I would like to thank all the support staff at CfB! Especially Morten and Rebeca for all Ph.D. school-related tasks, as well as Birte and Darko for all the help with administrative paperwork for my external stay and grants! A special thank you to the reception team Jesper, Annette, and Lise, the IT guys Kasper, August, and Sebastian, the goods

reception team Etem, and Torben, and our media and dishwasher team Lise-Lotte and Ann! Thank you for your work, which allows all the lab folks to focus on the science! I would like to extend my thanks to Tune and Daniella for their help in running metabolomics and proteomics samples, and for taking time to answer all my questions.

I had the wonderful opportunity to join the Joint Bioenergy Institute in Emeryville for my external stay. A big thank you to Jay Keasling for hosting me, and to Robert Bertrand and Kevin Yin for the inspiring collaboration. I am very grateful to all Keasling lab members for the fantastic time, especially Aiden, Matthias, Allie, Jacob, Graham, Namil, Peter, Matt, Thomas, Leah, Michael, Lucas, and Isaac. I would also like to thank Pablo for helping me set up my external stay and for putting me in contact with the right people.

Thank you to all current and former collaborators, especially Julian Schmidt, Rhena Decker, and Stephanie Grond, as well as Felipe Guerrero Garzon.

Outside of my Ph.D. life, I got the work together with the fantastic REBBLS bunch, exploring the intersection of entrepreneurship and biotech. I am especially proud of what we have accomplished with the REBBLS Primer over the last three years. Therefore, I would like to extend my special thanks to all current and former members of the organizing committee, namely Lasse, Max, Vivienne, Inge, Lasse B., Birgit, Keenan, Emma, Kamil, Petar, Christoffer, Paloma, Sabrina, and Shelley. I learned so much from all of you and can't wait to see what comes next!

Starting my Ph.D. in January 2020, little did I know that world-changing events were just around the corner. I had the pleasure to join an inspiring group of volunteers from CfB and other departments to help scale up COVID-19 testing during the first wave. Thank you to Jeremy and Lachlan for all the organization, and all the volunteers, especially the ones from station 1 for the hard but rewarding work!

During these last couple of years, I got to befriend so many wonderful, inspiring, and beautifully crazy people. You truly are what made this journey worthwhile and added endless beautiful memories to my Ph.D. experience.

Thank you, Efi, for your incredible kindness, your support, and for being the friend to call if things went south. Sorry you had to tolerate my dad jokes for so many years! It's been wonderful going through the Ph.D. experience with you, and you definitely made the whole journey much more enjoyable, in the good and the bad times! Omkar, thank you for the countless inspiring and thought-provoking philosophical and spiritual conversations, for all the fun parties, and for the great food! Having you as a friend certainly made me a more reflective person. I cannot wait to visit you in San Diego soon, and to see where your journey will take you. Thank you, Jonas, for being my first roommate & friend in Copenhagen! Thank you for the shared love for The Office, for all the bad jokes, your terrible ways of explaining board game rules, and for the tough squash matches! Alberto, thanks for making this strict rule-abiding German loosen up a bit, for teaching me how to make proper pasta, for being the best Disco bartender, and for all the fun at home. Vivienne, thanks for all the great conversations, forcing me to admit that things are actually going well sometimes, for all the great fun at REBBLS, for Banana Club, and for FINALLY admitting that Disco music is actually great. Bram, thank you for all your jokes and pranks, for being such a positive spirit, for all the creepy looks from the coffee machine, and for making me run the M3 city run. Max, thank you for all the fun at REBBLS, for all the great dinners & shared love for tofu, & I even forgive you for the body role. Thank you, Jasper, the ultimate Dutch wild spirit. Thank you for bringing so much fun and creative craziness to pretty much everything you do, for all the great (themed)parties, and of course for the many beers at Kompasset. Kate, thank you for the shared love for CRISPR, for making music together (why did we stop??), for the great times at Banana Club, and for the crazy themed parties and dinners! Arrate, thank you for constantly distracting me in the lab in the worst possible moments, for teaching me all about strange Basque sports, your eagerness to learn the German language, and generally, all the fun! Lucas, I forgive you for constantly bringing out the worst The Office fan side of me! Thank you for all the inappropriate jokes in the worst possible moments, your help with fixing bikes, and for all the great parties. By the way, I still do not believe you actually work here. Thank you, Philip, for all the nice dinners I almost always missed, for making COVID Christmas enjoyable, for the great times in the US, especially for making us go to Seattle by train, an interesting climb in Yosemite, Stanford football, and for great parties in the Berkeley hills. Thanks, Nico G. for a great time at JBEI, for the fun trips to SF and Seattle with Philip, walking 10 blocks in 20 minutes, fun

parties, and for being such a great lab buddy! Pippo, thank you for the shared love for nice shirts, for playing the finest tunes both as DiscoPippo and BrohainPippo, for your constant bad jokes, and spacy parties. Thank you Matiss for being such a fun and positive guy, for bringing Milo, my favorite stress relief dog to the office, and for the nice hikes!

Anja, Lara, Nico, Justine, Carolyn, Sophia, and everyone else from the "Beach Business" group, thank you for the wonderful times at CfB, the beach, or Copenhagen's finest Ølbars!

Thank you to the entire CfB community for the great time. It truly has been a privilege to work in such a diverse environment with so many wonderful and inspiring people! I also would like to thank everyone in the Ph.D. club and the CfB Goes Green initiative for all the great events and activities!

Alla mia famiglia italiana, Alberto, Silvia, and Dedo, thank you for being such wonderful roommates and for all the fun, chaos, enjoyed cultural differences, and great parties. Having you as roommates certainly prevented me from going insane during this Ph.D. at times! Also thank you to my former roommates Niklas, Anika, and Sophia for a great time, and for making it through the lockdowns together.

I would also like to thank my study friends from Germany, especially Markus, Svenja, Melinda, Thilo, Heiko, and Jess. It is wonderful to see where all of you ended up, and I am looking forward to the next big reunion!

Marie-Sophie, I got so lucky meeting you at the beginning of my Ph.D. and just before the pandemic started. Thank you for all your love and support, for your encouragements, and for being so understanding when I had a lot do to or was very stressed. You not just made these last couple of years so so much better, but your support, love, and encouragement are what helped me cross the finish line. Thank you for all the wonderful trips over the last couple of years, for the hyggelige times in Copenhagen, our camping and road trip adventures. I am so grateful to have you by my side!

Last but definitely not least I would like to thank my family for their tremendous support over the last 10 years.

My grandparents were always very interested in what I was doing and studying, helped and supported me, and enabled many of the opportunities I had the privilege to enjoy over the last 10 years. Unfortunately, they are not with us anymore to witness the fruits of their support. I hope you would have been proud! Thank you, Hermann, Elfriede, and “Captain” Nell, for all your love and support!

Marianne and Leo, thank you for everything you have done for me. For all your support, interest, for enabling so much, and for always being so interested and caring. Thank you!

My wonderful sisters, Johanna and Susanne, thank you for all your support, for always having an open ear, for in many regards leading the way, and for your encouragements. Thank you for always being there for me! You are both such great inspirations!

It might be my name on the thesis, but this Ph.D. thesis would not exist without the incredible continuous support from my parents, who were always willing to make sacrifices so that my sisters and I had the best possible opportunities. Thank you, Mom and Dad, for all your love, for the many years of support, for your encouragements, for your understanding, and for opening so many doors! Thank you for believing in me, for being there in the good and the bad times, and for being so supportive, even when I moved far away. This thesis is just as much your accomplishment as it is mine!

From the bottom of my heart, thank you all! I feel very proud to be finally able to deliver this thesis, but above all, I feel incredibly thankful and blessed.

Thank you!

PUBLICATIONS

Publications included in this thesis:

Chapter 2

Whitford, C. M.; Cruz-Morales, P.; Keasling, J. D.; Weber, T. The Design-Build-Test-Learn Cycle for Metabolic Engineering of Streptomycetes. *Essays Biochem.* **2021**, *65* (2), 261–275. https://doi.org/10.1042/EBC20200132.

Chapter 3

Tong, Y.; **Whitford, C. M.**; Blin, K.; Jørgensen, T. S.; Weber, T.; Lee, S. Y. CRISPR–Cas9, CRISPRi and CRISPR-BEST-Mediated Genetic Manipulation in Streptomycetes. *Nat. Protoc.* **2020**, *15* (8), 2470–2502. https://doi.org/10.1038/s41596-020-0339-z.

Chapter 4

Whitford, C. M.; Gren, T.; Palazzotto, E.; Lee, S.Y.; Tong, Y.; Weber, T. Systems Analysis of Highly Multiplexed CRISPR-Base Editing in Streptomycetes. *Under Review at ACS Synthetic Biology*

Chapter 5

Whitford, C. M.; Gockel, P.; Faurdal, D.; Gren, T.; Sigrist, R.; Weber, T. CASCADE-Cas3 Enables Highly Efficient Genome Engineering in *Streptomyces* Species
Published as a preprint biorxiv: https://doi.org/10.1101/2023.05.09.539971

Chapter 6

Tong, Y.; Jørgensen, T. S.; **Whitford, C. M.**; Weber, T.; Lee, S. Y. A Versatile Genetic Engineering Toolkit for E. Coli Based on CRISPR-Prime Editing. *Nat. Commun.* **2021**, *12* (1), 5206. https://doi.org/10.1038/s41467-021-25541-3.

Other publications:

Tong, Y.; **Whitford, C. M.**; Robertsen, H. L.; Blin, K.; Jørgensen, T. S.; Klitgaard, A. K.; Gren, T.; Jiang, X.; Weber, T.; Lee, S. Y. Highly Efficient DSB-Free Base Editing for Streptomycetes with CRISPR-BEST. *Proc. Natl. Acad. Sci.* **2019**, *116* (41), 20366–20375. https://doi.org/10.1073/pnas.1913493116.

Gren, T.; Jørgensen, T. S.; **Whitford, C. M.**; Weber, T. High-Quality Sequencing, Assembly, and Annotation of the *Streptomyces Griseofuscus* DSM 40191 Genome. *Microbiol. Resour. Announc.* **2020**, *9* (47), 19–20. https://doi.org/10.1128/mra.01100-20.

Gren, T.; **Whitford, C. M.**; Mohite, O. S.; Jørgensen, T. S.; Kontou, E. E.; Nielsen, J. B.; Lee, S. Y.; Weber, T. Characterization and Engineering of *Streptomyces Griseofuscus* DSM 40191 as a Potential Host for Heterologous Expression of Biosynthetic Gene Clusters. *Sci. Rep.* **2021**, *11* (1), 18301. https://doi.org/10.1038/s41598-021-97571-2.

Yin, K.; Cruz-Morales, P.; **Whitford, C. M.**; Keasling, J. D. Heterologous Production of Polycyclopropanated Fatty Acids and Their Methyl Esters in *Streptomyces*. *STAR Protoc.* **2023**, *4* (2), 102190. https://doi.org/10.1016/j.xpro.2023.102190.

AIM AND OUTLINE OF THIS THESIS

The aim of this Ph.D. project was to develop novel CRISPR tools and perturbation technologies, which can accelerate the exploration of the biosynthetic space of streptomycetes by enabling engineering with unprecedented precision, throughput, and efficiency.

Chapter 1 of the Ph.D. thesis provides a general introduction to specialized metabolites, *Streptomyces*, CRISPR, and challenges towards accessing the incredible microbial biosynthetic potential found in Nature. The chapter aims to provide the necessary background information needed for the following chapters.

Chapter 2 builds on top of Chapter 1 and provides a comprehensive overview of the Design-Build-Test-Learn cycle for metabolic engineering of streptomycetes. Therefore, important resources for each of the stages of the cycle are described and highlighted. Finally, the chapter describes some of the existing challenges that have been hampering the widespread implementation of the Design-Build-Test-Learn cycle for metabolic engineering of streptomycetes. This chapter was published as a review in Essays in Biochemistry.

Chapter 3 offers a detailed overview of the existing streptomycete CRISPR tools and protocols. The step-by-step instructions provide a great understanding of the complex workflows associated with engineering streptomycetes. This chapter was published as a protocol in Nature Protocols.

In **Chapter 4**, multiplexed cytosine base editing is employed to demonstrate the simultaneous editing of the highest number of genomic targets in *Streptomyces* to date. Multi-omics technologies are used to investigate the editing performance of differently sized sgRNA arrays, to analyze the scale of off-target effects, and to determine the systems-wide consequences of such deep base editing mediated genomic perturbations. This chapter is currently under review at ACS Synthetic Biology.

In **Chapter 5**, the first type I CRISPR system-based genome engineering tool for streptomycetes is described. The system, pCRISPR-Cas3, enables highly

efficient genome engineering through the introduction of random-sized deletions, targeted deletions, and substitutions. pCRISPR-Cas3 was applied in various *Streptomyces* strains of importance and shown to allow genome engineering with superior efficiencies compared to pCRISPR-Cas9. This manuscript was published as a preprint on biorxiv.

In **Chapter 6**, the development of a CRISPR-Prime genome engineering tool for *Escherichia coli* is described. The tool, the first of its kind, highlights the capabilities of the next generation of multi-component CRISPR tools for sophisticated engineering of cell factories. Adaptation of CRISPR-Prime for *Streptomyces* is still ongoing, and likely to become a powerful tool in the growing toolbox. This chapter was published in Nature Communications.

Finally, **Chapter 7** provides a discussion of the previous chapters as well as an outlook into the future of specialized metabolite discovery and the exploration of Nature's biosynthetic potential.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

The Dawn of Synthetic Biology Powered Exploration of Nature's Biosynthetic Potential

Christopher M. Whitford[1]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby 2800, Denmark.

A MICROBIAL WORLD

Biology operates in continuous cycles of life and death, synthesis and decomposition. All biological cycles are the result of intricate interactions between countless species, ecosystems, and all kingdoms of life. In these complex networks, microbial species play leading roles, and their functions are just as diverse as the ecological niches they occupy. Microbial life can be found all over Earth, from the barren polar regions to the dark and high-pressure environments of the deep sea, from steaming hot hydrothermal vents to the moderate environments of forests. One of the most important ecosystems for microbial life on Earth is soil. An incredible diversity of life is present here, and it is estimated that at least a quarter of the total biodiversity on Earth can be found in soil [1]. This biodiversity is made up of millions of species of prokaryotes, fungi, archaea, viruses, and microeukaryotes, such as protists, nematodes, and tardigrades [1,2]. In fact, soil microorganisms make up the second largest fraction of biomass on Earth [3]. This truly is a microbial world.

THE SOIL MICROBIOME

The soil microbiome is not only incredibly diverse but also plays a crucial role in most biogeochemical cycles. The major building blocks of life, namely hydrogen, carbon, nitrogen, oxygen, sulfur, and phosphorus, are all cycled in Earth's biogeochemical cycles, which at least in part are catalyzed by microbial conversions [3,4]. As such, soil represents one of the most important catalytic zones for Earth's biogeochemical cycles.

Important examples include nitrogen fixation, nitrification, denitrification, or phosphate solubilization [3,5,6]. Fixation of atmospheric nitrogen is a crucial step for soil nitrogen availability and is performed by diazotrophs, either legume-associated or other soil bacteria [7,8]. Nitrification describes the stepwise conversion of ammonia to nitrate, whereas denitrification refers to the stepwise conversion of nitrate into molecular nitrogen, effectively closing the nitrogen cycle [9].

Given that most dead terrestrial organic matter eventually falls to the ground, many of the species present in soil are so-called saprophytes, which play an important role in breaking down dead organic matter into simple building blocks. These are then used for building up new organic matter, such as microbial, fungal, or plant biomass. Among the most important saprophytes are certain fungi and bacteria, which are well known for their ability to secrete enzymes to break down complex and recalcitrant biomass such as lignin or chitin and have hence been used as sources for many industrial enzymes [10]. The complex soil environment, in which decomposition and synthesis are so closely intertwined, has resulted in and is equally made possible by, highly sophisticated interactions between species across all kingdoms of the tree of life. For example, the majority of nitrogen and phosphorus needed for plant growth is provided by soil microorganisms [3]. The symbiosis between legumes and rhizobia represents one of the most well-studied such interactions. Rhizobia fix atmospheric N_2 and secrete ammonia in return for energy and carbon in the form of dicarboxylic acids [11]. An even more complex example is leaf-cutting ants which use soil-dwelling bacterial species to protect their fungus garden against other pathogenic fungi [12]. However, given the abundance of substrates in soil, and the diversity of species therein, interactions between species can come in many different forms. In fact, mutualistic symbiotic interactions appear to be rather rare, and negative interactions, such as competition and exploitation are dominant [13]. However, how are these complicated interactions coordinated, especially given their sophistication and specificity?

CHEMISTRY IS THE LANGUAGE OF INTERACTION IN THE SOIL MICROBIOME

Most complex interspecies interactions go far beyond simple competition for nutrients and are facilitated through the exchange of signals in the form of complex molecules [14–16]. These chemical mediators can be both volatile and soluble and can thus act both proximally and distally. Especially volatile molecules can modulate species and communities beyond their direct proximity [16]. Interestingly, isolates from the same location appear to be better at inhibiting each other compared to isolates from different locations, suggesting that these interactions are the result of and shaped by the local

environment [17]. Many of these chemically mediated interactions are highly specific and targeted at specific organisms within the shared environment. For example, many chemical mediators function as antibiotics, targeting competing species within the same ecological niche. Antibiotics are usually produced in response to signals indicating the presence of a competing species and allow the producing species to directly modulate the abundance of the competing species [18,19]. Other molecules function as antifungals. The previously mentioned leaf-cutting ants protect their fungus garden from pathogenic fungi by employing symbiotic soil bacteria which produce the potent antifungal compound candicidin [12]. This antifungal is highly specific and does not significantly affect growth of the desired farmed fungus, thus enabling a highly specialized symbiotic interaction. Chemical compounds produced by soil bacteria can further have direct beneficial effects on plants and plant growth. Certain strains of *Bacillus subtilis*, a rhizobacterium frequently found in soil, produce 3-hydroxy-2-butanone and 2,3-butanediol, which were shown to result in significant plant growth promotion in the model plant *Arabidopsis thaliana* [20]. Finally, geosmin is a great example of the functional diversity of these complex molecules, but also of the intricate interkingdom interactions they facilitate. The molecule, responsible for the "after-rain soil" smell, was shown to act as an attractant for arthropods. These feed on the spores of the producing bacteria, and thus become a vehicle for spore dispersal both through excretion and by dispersing spores stuck to the arthropod's cuticle [21]. By producing a complex chemical mediator, these bacteria are essentially calling themselves an arthropod taxi.

All these highly complex and specific interactions are enabled by the structural diversity and complexity of the employed molecules. These molecules, referred to as specialized metabolites, are the cornerstone of intricate interspecies interactions.

SPECIALIZED METABOLITES

All essential building blocks for an organism's basic biological functions, such as amino acids or nucleotides, are products of its primary metabolism [22]. In contrast, specialized metabolites, commonly also referred to as secondary metabolites or natural products, are complex metabolites not directly essential

for basic biological functions such as growth and replication [23]. However, specialized metabolites allow species to adapt and persist in competitive ecological niches and can hence indirectly be very important for a species' survival in a given environment.

Specialized metabolites have been used by humans for thousands of years, as they are the active ingredients in most plant extracts [24,25]. In fact, this Ph.D. project was only made possible through the powerful contributions of caffeine, a potent specialized metabolite produced by plants of the genus *Caffea*. Other popular examples include the specialized metabolites produced by hops, giving beers their complex flavors, or those produced by species of the genus *Cannabis*, which have been used for thousands of years for recreational and medicinal purposes [26].

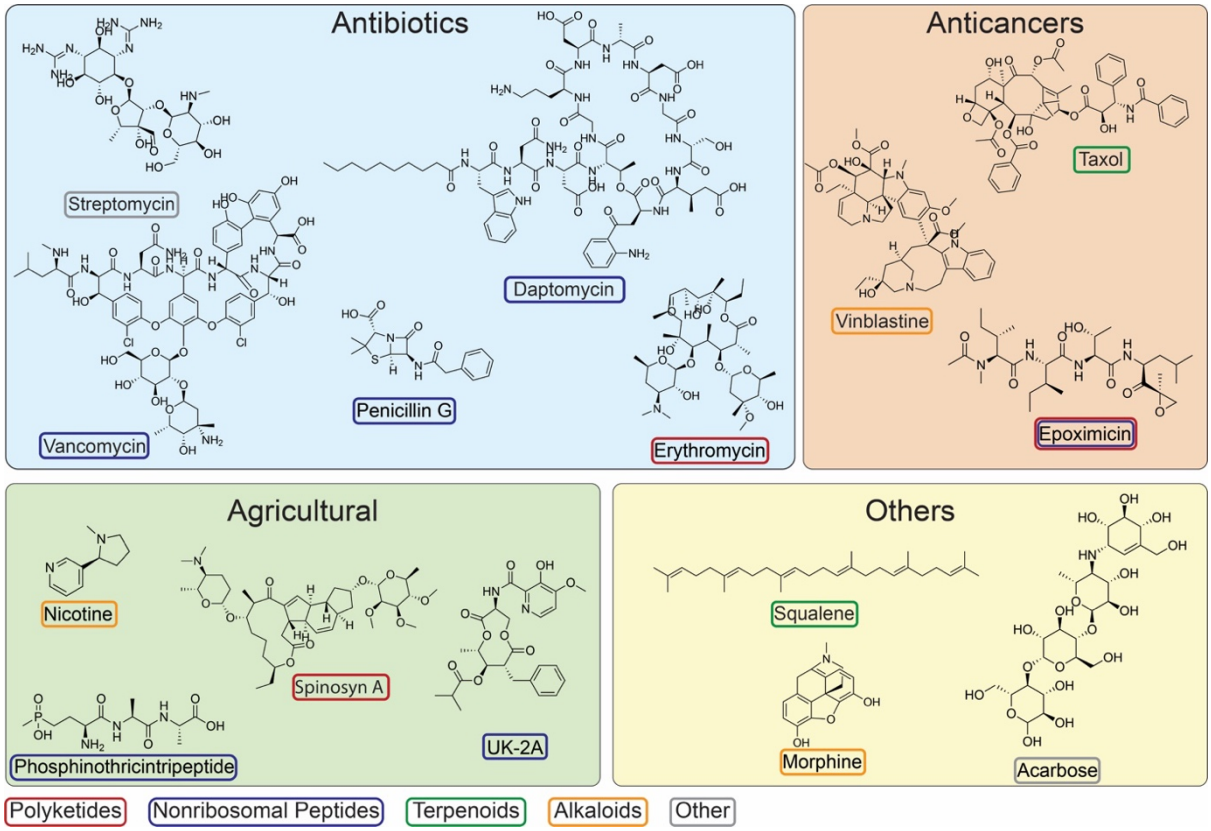


Figure 1: Examples of specialized metabolites, their applications, structures, and compound classes. Specialized metabolites can display an astonishing structural diversity, which results in a wide range of complex bioactivities. All shown specialized metabolites are marked with their compound classes, and it must be highlighted that specialized metabolites can fall into many more compound classes. Compounds with applications besides antibiotics, anticancerous, and agricultural agents, are used for example as vaccine adjuvants and cosmetic ingredients (Squalene) [27], analgesics (morphine), or anti-diabetic drugs (acarbose) [28].

In bacteria, the genetic information needed to produce these specialized metabolites is typically encoded in genes clustered together in biosynthetic gene clusters (BGCs) [29]. BGCs are made up of biosynthetic and accessory genes, encoding tailoring enzymes, transporters, or self-resistance mechanisms (Fig. 2). The biosynthetic enzymes responsible for the production of the molecular scaffolds are among the most fascinating and complex enzymatic machineries known [30,31]. Common classes of compounds produced from BGCs include polyketides [30,32], nonribosomal peptides [33,34], or terpenes [35,36].

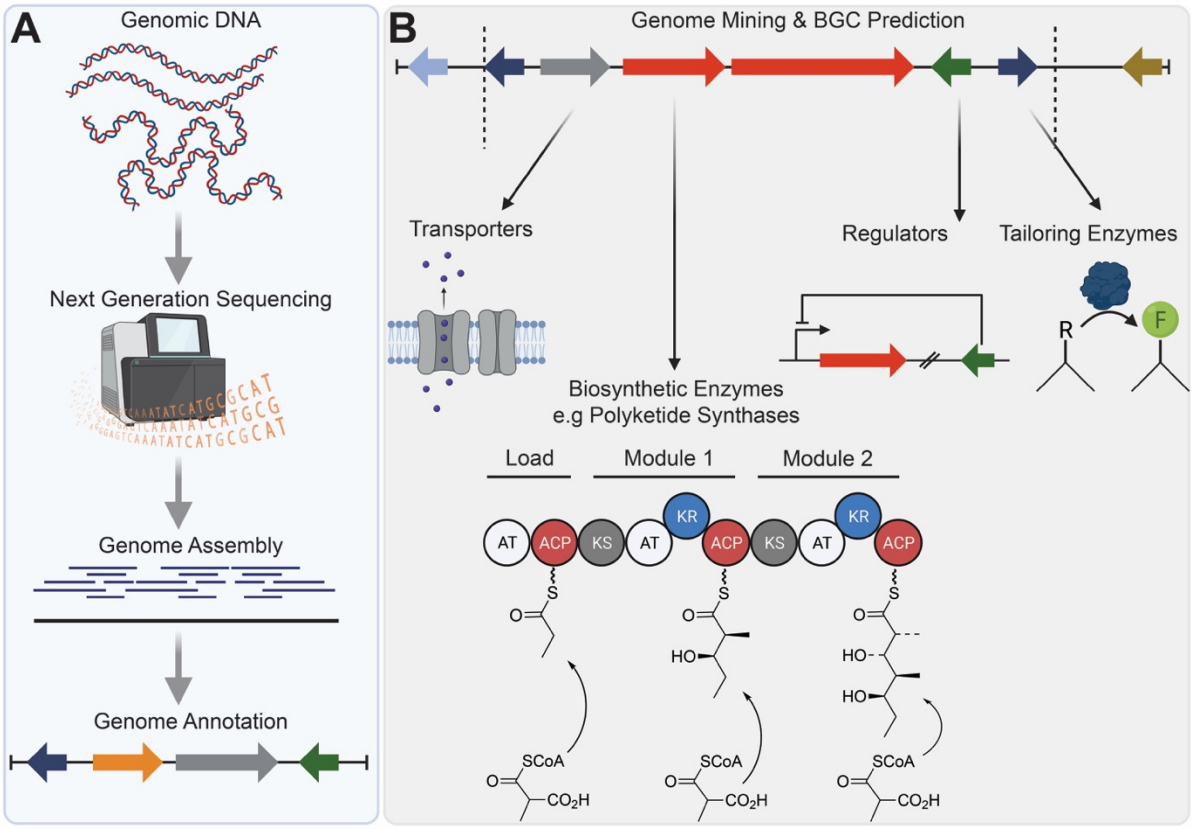


Figure 2: Overview of the workflow from genomic DNA to functional annotation of biosynthetic gene clusters. A. Genomic DNA from isolates or metagenomic samples is isolated and sequenced using next-generation sequencing technologies such as Illumina or Oxford Nanopore sequencing. The genomes are subsequently assembled and annotated, assigning specific DNA sequences (putative) functions and products. **B.** Annotated genomes can subsequently be mined for biosynthetic gene clusters using software such as antiSMASH [50]. Biosynthetic gene clusters typically consist of biosynthetic enzymes, transporters for the export of the product, regulators controlling the expression of the respective genes, and tailoring enzymes, which can install additional modifications on the compound scaffold.

Especially polyketide and nonribosomal peptide synthases are well known for their modular structure and biosynthetic logic, and therefore often referred to as biosynthetic assembly lines [30,34]. In addition to biosynthetic genes, BGCs typically also encode for regulators controlling the lower BGC-specific levels of regulatory cascades [37]. Following the advent of next-generation sequencing and the ability to sequence even complex organisms [38–40], the clustered organization of most specialized metabolite pathways, as well as the modular assembly-line logic of many, has allowed comprehensive bioinformatic analysis of diversity, distribution, and novelty of BGCs [41–45]. Tools like antiSMASH for BGC prediction [46], MIBiG, a database of experimentally validated BGCs [29], BiG-SCAPE and CORASON for exploring BGC diversity [47], or the antibiotic target seeker ARTS [48], have since become standard tools. The increasing number of sequenced microbial genomes, together with sophisticated computational tools, has since helped to revitalize the investigation of specialized metabolites from natural origins as potential drug candidates [49]. Performing bioinformatic analysis at large scale further allows to direct experimental work toward species or ecological niches with large untapped biosynthetic potential. One extensive bioinformatic analysis performed by Gavriilidou et al. in 2022 on around 170,000 bacterial genomes and 47,000 metagenome assembled genomes, identified members of the genus *Streptomyces* to encode the largest biosynthetic diversity by far [41]. As it turns out, *Streptomyces* are Nature's greatest chemists.

STREPTOMYCES: NATURE'S GREATEST CHEMISTS

Streptomyces are members of the phylum Actinomycetota (formerly Actinobacteria). The name, loosely translating to "twisted fungus", was originally given in confusion over the correct classification as filamentous fungal or bacterial species [51]. However, *Streptomyces* are Gram-positive, aerobe, non-motile, and primarily soil-dwelling bacteria [52,53]. This clash of bacterial and seemingly fungal features already hints at the complex biology of these fascinating bacteria.

Their most striking feature is the filamentous lifestyle (Fig. 3). Starting from single germinated spores, substrate mycelium is formed. In the next phase, aerial hyphae start to develop, which expand to form reproductive aerial

mycelium, which gives *Streptomyces* their intriguing phenotypes [54,55]. Specialized metabolites, especially antibiotics, are primarily produced during the formation of aerial mycelium. During this phase, streptomycetes undergo programmed cell death of the vegetative mycelium to provide substrates for the formation of the aerial mycelium. Classically, it has been hypothesized that antibiotics are produced during this stage of the life cycle to fend off motile competitors from feeding on this nutrient pool [54,56,57].

During the following phase, chromosome segregation and cell septation premediate spore maturation. Once these are fully matured, they can be dispersed to complete the life cycle [54,55]. Unlike textbook unicellular bacteria, *Streptomyces* do not replicate through fission, that is the division of one parental cell into two daughter cells. Instead, *Streptomyces* grow through tip extension and branching of their hyphae [53,58].

Another striking feature of *Streptomyces is* their large linear chromosomes (Fig. 3C). These are typically between 6 and 12 Mb in size, have a high GC content of around 72 %, and feature on average around 30 BGCs [59,60]. Another feature of *Streptomyces* genomes is their high plasticity, with extensive deletions and rearrangements being rather common [61].

Given their incredible biosynthetic potential, *Streptomyces* are extensively studied and screened for bioactive compounds. This started with the discovery of the antibiotic streptomycin from *Streptomyces griseus* in 1944 by Selman Waksman, Albert Schatz, and Elizabeth Bugie [62], a discovery that would lead to a Nobel Price for Selman Waksman in 1952. During the following "Golden Age" of antibiotic discovery, *Streptomyces,* along with other Actinobacteria as well as fungal species, were extensively screened for new antibiotics, leading to a wave of new compounds entering the clinics [63–65].

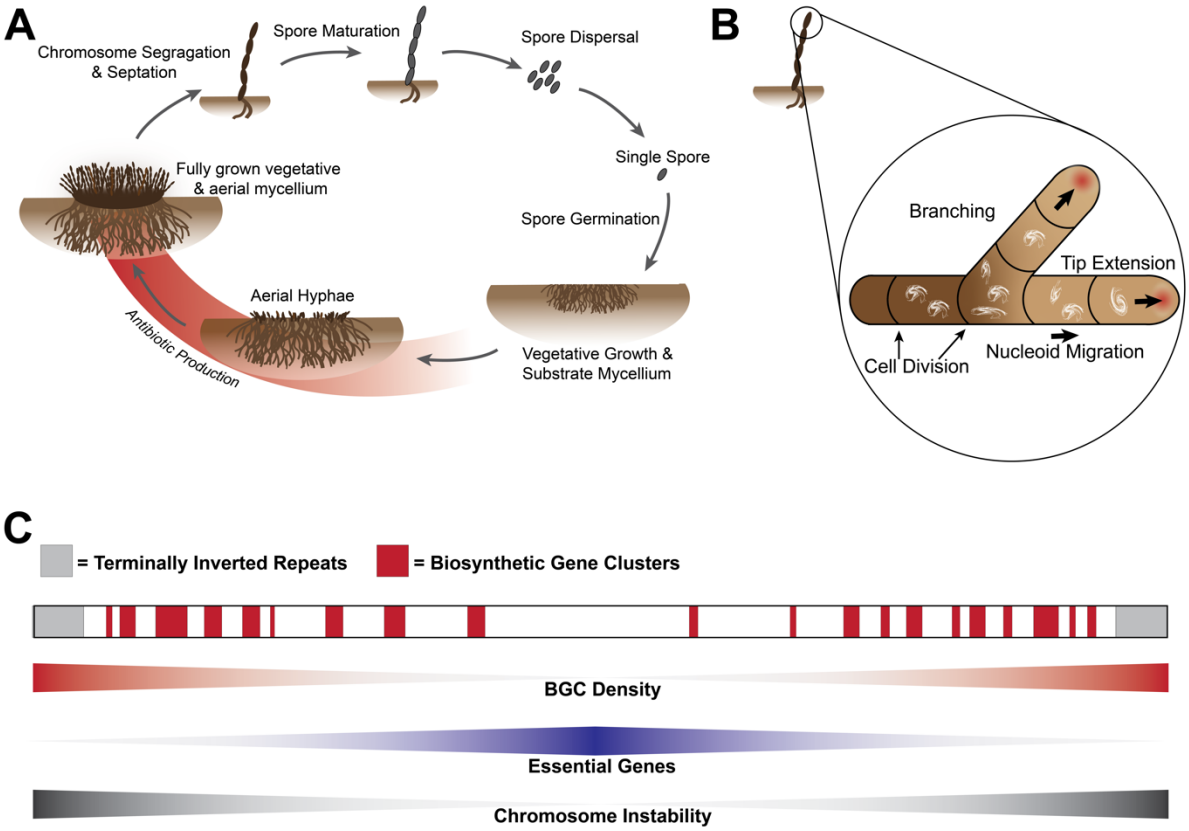


Figure 3: Schematic overview of the life cycle, growth characteristics, and genome organization of streptomycetes. A. The life cycle of streptomycetes. Starting from a single spore, complex colonies form through spore germination and the formation of substrate and aerial mycelium. Following chromosome segregation and septation, spores mature and are finally dispersed, closing the life cycle. Figure adapted from Barka et. al [54]. **B.** Growth characteristics of streptomycetes. Growth of streptomycetes is characterized not by fission, but by filamentous growth along hyphae. These grow through branching and tip extension, accompanied by migration of the nucleoids. Figure adapted from Flärdh [58]. **C.** Streptomycetes have a linear chromosome with terminally inverted repeats. The majority of BGCs are present on the chromosomal arms, and essential genes are concentrated in the core region. Chromosome instability increases towards the termini.

Following the sequencing of the first whole genomes of *Streptomyces* in the early 2000s, namely *Streptomyces coelicolor* A3(2) [66], *Streptomyces avermitilis* [67], and *Streptomyces griseus* [68], it became apparent that despite decades of research and screening, a majority of the biosynthetic potential of streptomycetes had not yet been explored. For all analyzed species a large number of BGCs was identified, surpassing that of identified specialized metabolites from these species significantly [69]. This large untapped biosynthetic potential has since revitalized the search for novel bioactive compounds from natural sources [49]. But why, despite decades of research, has so much of the biosynthetic potential of streptomycetes remained unexplored?

BREAKING THROUGH THE BARRIER OF ECOLOGICAL CONTEXT LOSS

Historically, strains were isolated from diverse environments by growing samples on complex media and picking individual colonies for propagation and analysis. This sampling approach has resulted in strain collections of incredible size and diversity. It is estimated that up to 20 million isolates were analyzed for readily produced specialized metabolites between 1950 and 2000 [23]. To screen for new bioactive compounds, isolates were typically cultivated under few conditions and screened for bioactivity against target strains [23]. While this approach initially yielded many clinically used antibiotics, it managed to capture only a fraction of the existing biosynthetic diversity. This was mainly due to the inability to culture the majority of species [63], and due to what I will refer to as the *barrier of ecological context loss*.

Isolated species are sampled from highly complex ecological contexts in terms of abiotic and biotic parameters. These environments represent complex microbial consortia, with intricate interspecies interactions shaping these ecological niches. Here, the expression of BGCs for the production of specialized metabolites is tightly controlled and activated by specific environmental triggers. Upon sampling, isolates are usually cultivated in the laboratory in monocultures in complex media that do not mimic the natural environment. In this context, most environmental triggers are absent, leading to BGCs remaining unexpressed, so-called silent BGCs [69]. During the golden age of antibiotic discovery, which started decades before the development of recombinant DNA technology or next-generation sequencing, variations in culture conditions and untargeted mutagenesis were among the only tools available to access a broader breadth of the biosynthetic potential of *Streptomyces* species. But how can we access the so far inaccessible biosynthetic potential of streptomycetes under laboratory conditions, almost 80 years after the discovery of streptomycin?

Accessing the Biosynthetic Potential of Streptomycetes

The ultimate problem towards accessing more of the biosynthetic potential of streptomycetes can be described as a challenge to overcome the often very

tight and complex transcriptional regulation controlling the expression of BGCs (Fig. 4). While advances in metabolomics have allowed the identification and characterization of previously undetectable compounds produced in very low amounts, most BGCs remain virtually silent until today, either because they are not expressed, or because the product titers are below the level of detection.

Given the low dimensionality of most cultivation setups, one simple way to activate more BGCs is to systematically diversify the culturing conditions. This approach has been termed 'One Strain Many Compounds' (OSMAC) and relies on the alteration of parameters such as media composition, aeration, cultivation vessels, pH, or temperature, to activate the production of previously cryptic specialized metabolites [70]. While this approach allows the identification of a much higher number of compounds from individual strains, the scalability of this approach is certainly limited. Another widely used method is elicitor screening. Here, libraries of small molecules are screened for their ability to activate silent BGCs [71]. However, both OSMAC and elicitor screening represent untargeted approaches, with no control over which BGCs are activated.

The ever-increasing number of whole genome sequences of *Streptomyces* species allows extensive characterizations and *in silico* predictions with regard to the potential novelty of compounds produced from silent biosynthetic gene clusters. Therefore, targeted approaches have become increasingly popular.

These were made possible by the development of recombinant DNA technology in the 1970s, which opened up the possibility to directly engineer strains of interest. This can mean both the wild isolates, as well as heterologous expression hosts. Important advancements for genetic engineering of *Streptomyces* were for example the establishment of plasmid vector systems [72,73], the characterization of promoters [74], or the development of efficient conjugation methods [75,76]. Cloning of bacterial artificial chromosome (BAC) libraries is a classical (semi-)targeted method and allows the capturing of large chromosomal fragments, including entire BGCs [77]. These can then be transferred to other, well-characterized and established *Streptomyces* hosts, such as *S. coelicolor* or *S. albus,* for heterologous expression [78–80].

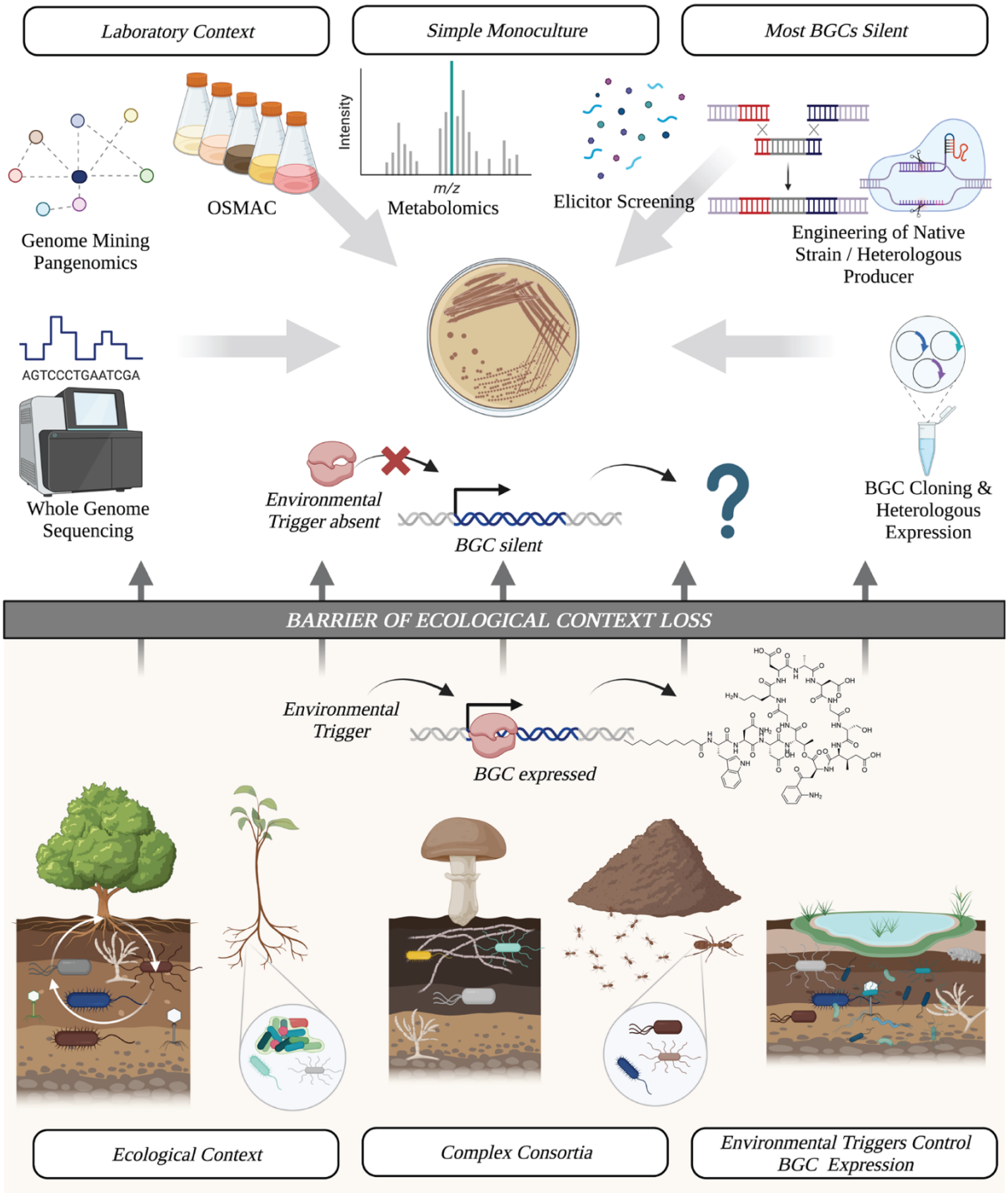


Figure 4: The challenge of ecological context loss for exploration of the biosynthetic potential of streptomycetes. Strains isolated from soil are usually isolated from highly complex environments, shaped by a diverse array of biotic and abiotic parameters. On top of that, many ecological niches are characterized by intricate interspecies interaction across the tree of life. Upon cultivation of isolates in the laboratory, normally done in monocultures, most BGCs remain silent due to the absence of environmental triggers normally controlling the expression of the respective biosynthetic gene clusters. Within the laboratory context, several methods can be deployed alone or in combination, to overcome the barrier of ecological context loss and to activate silent biosynthetic gene clusters.

Unfortunately, obtaining and verifying BACs is not only tedious and time-consuming, but their expression in heterologous hosts is often plagued by low or no expression of the encoded BGCs. This can be due to necessary higher-level regulators missing in the heterologous host, repressors on the BAC that are transferred as well, or missing precursors the heterologous host cannot provide. Nonetheless, by randomly trying different heterologous hosts, the expression of entire BACs has been shown to lead to the activation of silent BGCs and the identification of the corresponding products in many cases [81,82].

Advancements in genetic engineering tools have since enabled entirely new possibilities to access the hidden biosynthetic potential of streptomycetes, such as refactoring of entire BGCs, promoter knock-in, or inactivation of repressor genes [69,83,84]. Direct engineering of native producers and heterologous hosts thus represents the only targeted approach, which allows not only activation of BGCs but also pathway elucidation and modification. However, most classical engineering methods in streptomycetes take substantial time, require ordered cosmid, fosmid, or BAC libraries, and remain very tedious, with experimental timeframes easily stretching over several months. Therefore, tools that would allow highly precise and high throughput engineering of domesticated as well as previously genetically inaccessible streptomycetes were urgently needed. In 2012, such a tool published in the journal *Science* would usher in a new era, not just for streptomycetes, but for life science in general.

CRISPR: A NEW ERA

In 2012, a team of scientists led by Jennifer Doudna and Emmanuel Charpentier published a paper titled "A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity" [85], work which just ten years later would win Doudna and Charpentier the Nobel Prize in chemistry. Their paper described for the first time how a bacterial adaptive "immune system" called CRISPR-Cas could be reprogrammed to cut and edit any DNA sequence. In the almost 11 years since, CRISPR-Cas has revolutionized biomedical research, diagnostics, strain engineering, treatments for inherited diseases, and crop breeding, to name just a few application areas.

The term CRISPR was first coined by Francisco Mojica in 2001 and stands for Clustered Regularly Interspaced Short Palindromic Repeats [86,87]. The first observations of CRISPR sequences were made as early as 1987 [88], but their functions remained unknown until 2003, when Mojica identified CRISPR-Cas as an adaptive bacterial "immune system", results that were published in 2005 [89]. In 2002, Ruud Jansen also coined the term "CRISPR-associated" genes/enzymes, or Cas, together making up CRISPR-Cas. But how does CRISPR-Cas work and why has it become such a game changer?

Viruses are the most abundant biological entity on Earth, and those attacking bacteria are called bacteriophages. Bacteria and bacteriophages have been battling each other for billions of years [90,91]. It is therefore not surprising that complicated systems evolved to provide an edge over the respective foe. CRISPR-Cas is such a system that allows bacteria (and archaea) to fight off invading bacteriophages.

By storing genetic sequences from attacking bacteriophages, bacteria can keep a genetic memory, so-called protospacers, in their CRISPR locus. By using the stored memory as a guide, bacteria can fight off the same bacteriophages in the future by using the Cas enzymes to cut and destroy the genetic material of the invading bacteriophage (Fig. 5). To target a sequence, a protospacer adjacent motif (PAM) must be present, e.g. NGG for Cas9 [85]. These PAM motifs enable the scanning of long DNA sequences for potential target sequences and play an important role in activation of the catalytic activity of the Cas enzyme [92].

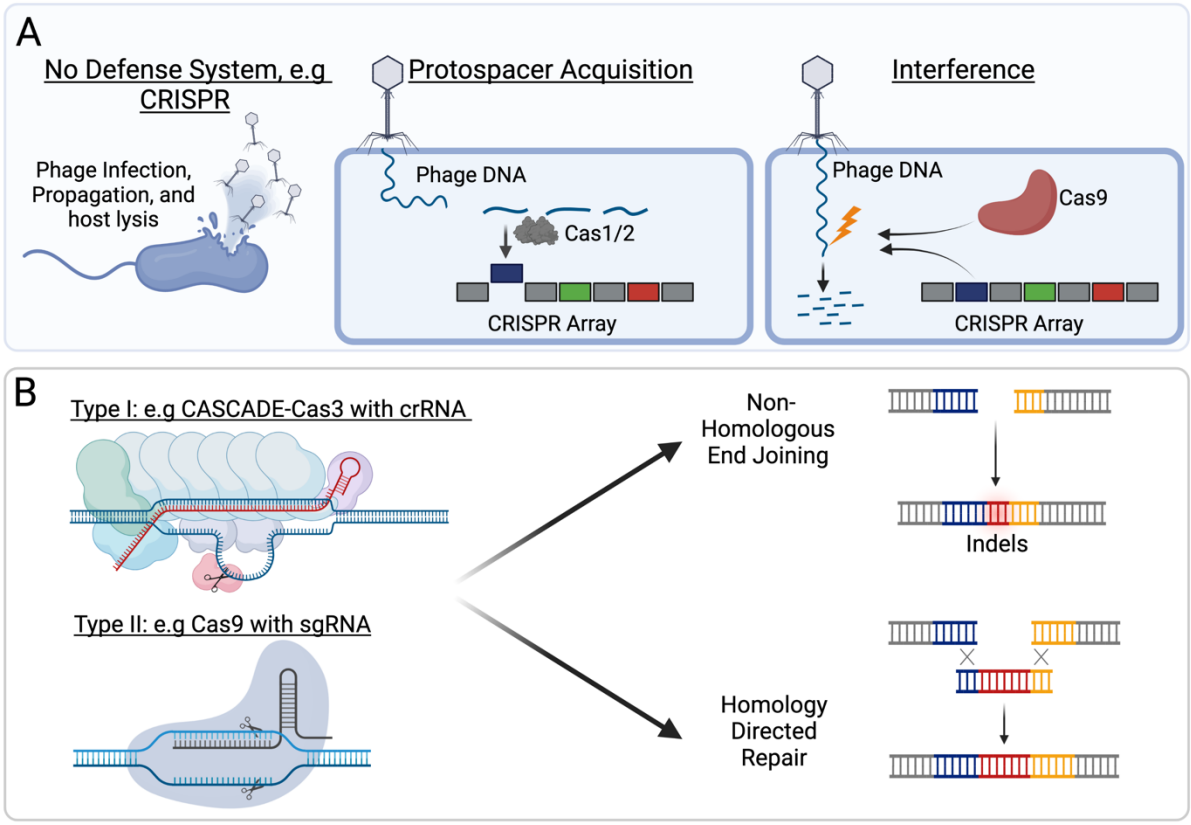


Figure 5: Overview of the natural function of CRISPR-Cas and its application for genome engineering. A. Bacteria and phages are in a constant battle. Without defense systems such as CRISPR, bacteria are infected by phages, which use the hosts' machinery for propagation, which finally results in lysis of the host cell. Using CRISPR, bacteria can store a genetic snippet of the phage DNA, called the protospacer, in their CRISPR locus, and use it to guide CRISPR effectors such as Cas9 to cut the invading phage DNA. **B**. Type I CRISPR systems require multiple CRISPR effectors to recruit the crRNA and to bind and cut the target sequence. This complex is called CASCADE. The hallmark effector of type I systems, Cas3, furthermore is of processive nature and starts degrading the DNA usually in a directional matter after introduction of a double-strand break. Type II CRISPR systems, such as Cas9, require only a single effector to bind and cut the target sequence. For genome engineering, double-strand breaks can be either repaired through non-homologous end joining, frequently resulting in random integrations and deletions (indels), or through homology-directed repair utilizing a repair template.

Utilizing this find-and-cut mechanism, Doudna and Charpentier showed that by providing rationally designed protospacers, this machinery could be reprogrammed to target any DNA sequence and hence be potentially applied for genome engineering. Furthermore, it was demonstrated that a synthetic chimera of the CRISPR RNA (crRNA), which carries the protospacer sequence, and the trans-activating crRNA (tracrRNA), the two RNA molecules making up

the guiding complex, could be used to direct the Cas complex, and was referred to as single guide RNA (sgRNA).

CRISPR systems are classified based on their structures, functional components, and mechanisms, and the associated Cas genes are usually simply numbered. In type II CRISPR systems, only a single Cas protein is needed to recruit the crRNA:tracrRNA or sgRNA, and cut the target sequence. All the ground-breaking experiments by Doudna and Charpentier were carried out using the CRISPR-Cas9 system from *Streptococcus pyogenes*, the most prominent member of type II CRISPR systems, and the most widely used CRISPR system in general. In contrast, in type I CRISPR systems, instead of a single Cas protein, a so-called CASCADE made up of several effectors facilitates the processing, recruiting, and targeting of the target locus. The hallmark nuclease in type I systems, Cas3, usually is of processive nature and starts degrading the DNA in a directional manner after the introduction of a double-strand break (DSB) [93,94]. In recent years ever-increasing numbers of new CRISPR systems and effectors have been identified, characterized, and utilized for genome engineering, and several reviews have summarized these advances in great detail [95,96].

Based on the mechanisms of CRISPR systems, a plethora of genome engineering approaches were developed. Classically, CRISPR-Cas9 (or a CRISPR-CASCADE) can be directed to introduce a DSB at a specific locus in the genome, which the cell can then either repair through non-homologous end joining (NHEJ) or homology-directed repair (HDR) (Fig. 5B). By providing a repair template made up from flanking homologous sequences from the genome and potentially additional modification, instructions for how to repair the DSB can be given to the cell, enabling site-directed genome engineering. Without a repair template, the cell is forced to either repair the double-strand break through NHEJ, usually resulting in integrations or deletions of small sequences (indels), or cell death in the absence of a NHEJ pathway [97,98].

CRISPR-Cas9 has since been turned into an incredibly versatile platform on top of which new genome engineering tools can be developed. By inactivating the two nuclease domains of Cas9, RuvC and HNH, with single amino acid mutations (D10A and H840A), Cas9 was turned into a catalytically dead nuclease. This mutated enzyme, referred to as dCas9, can still recruit sgRNAs

and target the target locus, but will not cut DNA. This allows dCas9 to be used to block the transcription of a target gene by sterically blocking the RNA polymerase, a method referred to as CRISPR interference [99]. Furthermore, proteins such as transcriptional activators can be guided to a specific target locus by fusing them to dCas9 [100,101].

Innovative technologies out of the David Liu Lab at the Broad Institute of Harvard University and the Massachusetts Institute of Technology took this approach even one step further. Combining a nicking Cas9, that is a Cas9 version with only one inactivated nuclease domain, and fusing it to a cytosine deaminase or an evolved adenosine deaminase, the team created base editors [102,103]. These allow the modification of single nucleotides within a specific editing window, without the introduction of potentially detrimental DSBs. A similar approach was used for the establishment of Prime editing, which utilizes a reverse transcriptase fused to a nicking Cas9 to enable modification or introduction of longer sequences without the need for DSBs [104]. These tools far surpass the capabilities of classical tools, and have since opened up entirely new possibilities for engineering of cellular systems.

THE DAWN OF SYNTHETIC BIOLOGY POWERED EXPLORATION OF NATURE'S BIOSYNTHETIC POTENTIAL

The last two decades have seen incredible advances in the field of synthetic biology, largely driven by enormous technological leaps. Synthetic biology describes a methodology rooted in the view of biology as an engineering discipline [105]. As such, synthetic biology relies heavily on standardization in the form of well-characterized genetic parts and modules, experimental workflows centered around iterations of the Design-Build-Test-Learn (DBTL) cycle, and the use of these methods and tools for forward engineering of cellular systems [106]. Synthetic biology has resulted in remarkable achievements in model organisms such as *Escherichia coli* and *Saccharomyces cerevisiae* [107–110]. However, despite a strong tradition of microbiology and molecular biology, widespread adaptation of synthetic biology methodologies has been very slow within the *Streptomyces* community, and synthetic biology applications in *Streptomyces* remain sparse. Still, the few applications using methodologies of synthetic biology in *Streptomyces*, especially modular designs

based on standardized parts and the DBTL cycle, demonstrated great promise [84,111–113]. More importantly, most enabling technologies of synthetic biology were first developed for and demonstrated in well-established and domesticated organisms such as *E. coli* and *S. cerevisiae*. The slow modification and adoption of these tools and technologies for complex *Streptomyces* species, as well as the trend to work with strain collections instead of single species, has delayed the widespread adoption of synthetic biology methodologies for *Streptomyces* research.

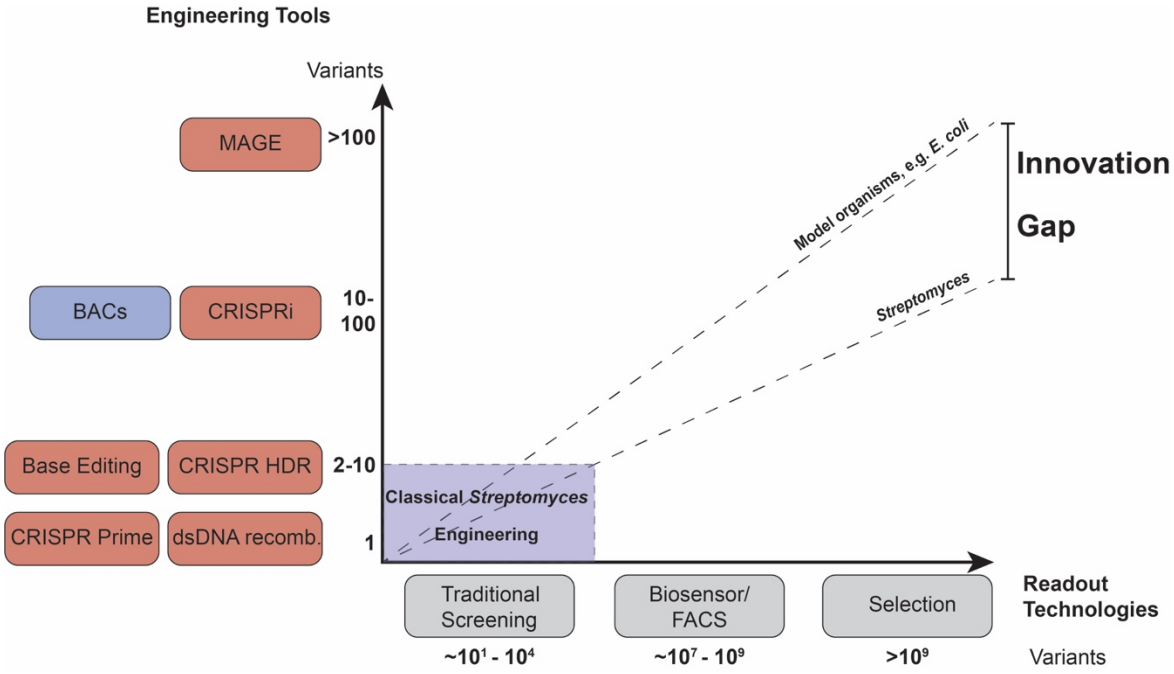


Figure 6: The innovation gap of *Streptomyces* metabolic engineering. The highest throughput of strain engineering experiments can be achieved by combining the engineering tools allowing the highest number of perturbations and variations with the appropriate readout technologies. Approximations for the highest number of variants for each tool and technology are shown. While BAC cloning and transfer is not a classical technique, it is of great importance for *Streptomyces* research and was therefore added to the plot. The underlying numbers for BACs are based on Libis et al. [112]. For model organisms, both the engineering tools and readout technologies were established allowing high throughput experiments. For *Streptomyces*, most experiments rely on both low throughput engineering and readout experiments, as indicated by the blue rectangle. Higher content experiments would be possible (indicated by the lower dashed line) but still fall short of the possibilities in model organisms. This reveals a large innovation gap for metabolic engineering of *Streptomyces*. The numbers in the figure were adapted from Yilmaz et al. [114].

Arguably the biggest bottleneck however remains the inability to engineer *Streptomyces* and their pathways at scale. Their slow growth results in slow iterations of the DBTL cycle, and the filamentous growth hampers widespread adoption of automated workflows. Therefore, engineering tools that enable highly efficient, high content or multiplexed engineering of *Streptomyces* species are urgently needed (Fig. 6). As the discrepancy between engineering capabilities, readout technologies, and the number of computationally identified BGCs keeps growing, our limited engineering capabilities are increasingly turning into the major bottleneck.

Once such engineering tools are established for *Streptomyces* and synthetic biology methodologies are applied, the combination of the incredible advances in sequencing and synthesis of DNA, computational tools, analytical and screening methods, automation, cultivation setups, and most importantly, our ability to edit DNA at scale and with unprecedented efficiency, even in previously inaccessible strains, will usher in a new era for specialized metabolite discovery.

This will come just in time, as the antimicrobial resistance crisis threatens the foundations of modern medicine [115–117], as climate change exacerbates the spread of fungal infections [118], as agriculture faces increased pressure to move towards biobased & sustainable solutions [119,120], as the increased risk for future pandemics requires novel antivirals [121,122], and as biobased production of complex chemicals becomes increasingly necessary [123]. If we can develop and establish the technologies needed, and deploy them at scale, nature's large untapped biosynthetic potential will be a hoard full of solutions to many of these problems.

REFERENCES

1. Sokol, N. W. *et al.* Life and death in the soil microbiome: how ecological processes influence biogeochemistry. *Nat Rev Microbiol* **20**, 415–430 (2022).
2. Bardgett, R. D. & van der Putten, W. H. Belowground biodiversity and ecosystem functioning. *Nature* **515**, 505–511 (2014).
3. Banerjee, S. & van der Heijden, M. G. A. Soil microbiomes and one health. *Nat Rev Microbiol* **21**, 6–20 (2023).
4. Falkowski, P. G., Fenchel, T. & Delong, E. F. The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science* **320**, 1034–1039 (2008).
5. Fierer, N. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat Rev Microbiol* **15**, 579–590 (2017).
6. Jones, C. M. *et al.* Recently identified microbial guild mediates soil N2O sink capacity. *Nature Clim Change* **4**, 801–805 (2014).
7. Mus, F. *et al.* Symbiotic Nitrogen Fixation and the Challenges to Its Extension to Nonlegumes. *Appl Environ Microbiol* **82**, 3698–3710 (2016).
8. Bueno Batista, M. & Dixon, R. Manipulating nitrogen regulation in diazotrophic bacteria for agronomic benefit. *Biochemical Society Transactions* **47**, 603–614 (2019).
9. Kuypers, M. M. M., Marchant, H. K. & Kartal, B. The microbial nitrogen-cycling network. *Nat Rev Microbiol* **16**, 263–276 (2018).
10. Thapa, S. *et al.* Microbial cellulolytic enzymes: diversity and biotechnology with reference to lignocellulosic biomass degradation. *Rev Environ Sci Biotechnol* **19**, 621–648 (2020).
11. Poole, P., Ramachandran, V. & Terpolilli, J. Rhizobia: from saprophytes to endosymbionts. *Nat Rev Microbiol* **16**, 291–303 (2018).
12. Haeder, S., Wirth, R., Herz, H. & Spiteller, D. Candicidin-producing *Streptomyces* support leaf-cutting ants to protect their fungus garden against the pathogenic fungus *Escovopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 4742–4746 (2009).
13. Palmer, J. D. & Foster, K. R. Bacterial species rarely work together. *Science* **376**, 581–582 (2022).
14. Hughes, D. T. & Sperandio, V. Inter-kingdom signalling: communication between bacteria and their hosts. *Nat Rev Microbiol* **6**, 111–120 (2008).
15. Scherlach, K. & Hertweck, C. Chemical Mediators at the Bacterial-Fungal Interface. *Annu. Rev. Microbiol.* **74**, 267–290 (2020).

16. Tyc, O., Song, C., Dickschat, J. S., Vos, M. & Garbeva, P. The Ecological Role of Volatile and Soluble Secondary Metabolites Produced by Soil Bacteria. *Trends in Microbiology* **25**, 280–292 (2017).
17. Vaz Jauri, P. & Kinkel, L. L. Nutrient overlap, genetic relatedness and spatial origin influence interaction-mediated shifts in inhibitory phenotype among *Streptomyces* spp. *FEMS Microbiol Ecol* **90**, 264–275 (2014).
18. Maan, H., Itkin, M., Malitsky, S., Friedman, J. & Kolodkin-Gal, I. Resolving the conflict between antibiotic production and rapid growth by recognition of peptidoglycan of susceptible competitors. *Nat Commun* **13**, 431 (2022).
19. Abrudan, M. I. *et al.* Socially mediated induction and suppression of antibiosis during bacterial coexistence. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11054–11059 (2015).
20. Ryu, C.-M. *et al.* Bacterial volatiles promote growth in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4927–4932 (2003).
21. Becher, P. G. *et al.* Developmentally regulated volatiles geosmin and 2-methylisoborneol attract a soil arthropod to *Streptomyces* bacteria promoting spore dispersal. *Nat Microbiol* **5**, 821–829 (2020).
22. Dewick, P. M. *Medicinal natural products: a biosynthetic approach*. (Wiley, A John Wiley and Sons, Ltd., Publication, 2009).
23. Katz, L. & Baltz, R. H. Natural product discovery: past, present, and future. *Journal of Industrial Microbiology and Biotechnology* **43**, 155–176 (2016).
24. Hardy, K. Paleomedicine and the Evolutionary Context of Medicinal Plant Use. *Rev. Bras. Farmacogn.* **31**, 1–15 (2021).
25. Petrovska, B. Historical review of medicinal plants' usage. *Phcog Rev* **6**, 1 (2012).
26. Ren, M. *et al.* The origins of cannabis smoking: Chemical residue evidence from the first millennium BCE in the Pamirs. *Sci. Adv.* **5**, eaaw1391 (2019).
27. Kim, S.-K. & Karadeniz, F. Biological Importance and Applications of Squalene and Squalane. in *Advances in Food and Nutrition Research* vol. 65 223–233 (Elsevier, 2012).
28. Tsunoda, T., Samadi, A., Burade, S. & Mahmud, T. Complete biosynthetic pathway to the antidiabetic drug acarbose. *Nat Commun* **13**, 3455 (2022).
29. Medema, M. H. *et al.* Minimum Information about a Biosynthetic Gene cluster. *Nat Chem Biol* **11**, 625–631 (2015).

30. Grininger, M. Enzymology of assembly line synthesis by modular polyketide synthases. *Nat Chem Biol* (2023) doi:10.1038/s41589-023-01277-7.
31. Carretero-Molina, D. *et al.* Discovery of gargantulides B and C, new 52-membered macrolactones from *Amycolatopsis* sp. Complete absolute stereochemistry of the gargantulide family. *Org. Chem. Front.* **9**, 462–470 (2022).
32. Nivina, A., Yuet, K. P., Hsu, J. & Khosla, C. Evolution and Diversity of Assembly-Line Polyketide Synthases: Focus Review. *Chem. Rev.* **119**, 12524–12547 (2019).
33. Süssmuth, R. D. & Mainz, A. Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew. Chem. Int. Ed.* **56**, 3770–3821 (2017).
34. McErlean, M., Overbay, J. & Van Lanen, S. Refining and expanding nonribosomal peptide synthetase function and mechanism. *Journal of Industrial Microbiology and Biotechnology* **46**, 493–513 (2019).
35. Oldfield, E. & Lin, F.-Y. Terpene Biosynthesis: Modularity Rules. *Angew. Chem. Int. Ed.* **51**, 1124–1137 (2012).
36. Helfrich, E. J. N., Lin, G.-M., Voigt, C. A. & Clardy, J. Bacterial terpene biosynthesis: challenges and opportunities for pathway engineering. *Beilstein J. Org. Chem.* **15**, 2889–2906 (2019).
37. Rigali, S., Anderssen, S., Naômé, A. & van Wezel, G. P. Cracking the regulatory code of biosynthetic gene clusters as a strategy for natural product discovery. *Biochemical Pharmacology* **153**, 24–34 (2018).
38. National Human Genome Research Institute. The Cost of Sequencing a Human Genome. https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost.
39. Wick, R. R., Judd, L. M. & Holt, K. E. Assembling the perfect bacterial genome using Oxford Nanopore and Illumina sequencing. *PLoS Comput Biol* **19**, e1010905 (2023).
40. Alvarez-Arevalo, M. *et al.* Extraction and Oxford Nanopore sequencing of genomic DNA from filamentous Actinobacteria. *STAR Protocols* **4**, 101955 (2023).
41. Gavriilidou, A. *et al.* Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat Microbiol* **7**, 726–735 (2022).
42. Medema, M. H. *et al.* antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial

and fungal genome sequences. *Nucleic Acids Research* **39**, W339–W346 (2011).
43. González-Salazar, L. A. *et al.* Biosynthetic novelty index reveals the metabolic potential of rare actinobacteria isolated from highly oligotrophic sediments. *Microbial Genomics* **9**, (2023).
44. Cimermancic, P. *et al.* Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. *Cell* **158**, 412–421 (2014).
45. Doroghazi, J. R. & Metcalf, W. W. Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics* **14**, 611 (2013).
46. Blin, K. *et al.* antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Research* **45**, W36–W41 (2017).
47. Navarro-Muñoz, J. C. *et al.* A computational framework to explore large-scale biosynthetic diversity. *Nat Chem Biol* **16**, 60–68 (2020).
48. Alanjary, M. *et al.* The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. *Nucleic Acids Research* **45**, W42–W48 (2017).
49. Harvey, A. L., Edrada-Ebel, R. & Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov* **14**, 111–129 (2015).
50. Blin, K. *et al.* antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research* **49**, W29–W35 (2021).
51. Lechevalier, H. A. & Lechevalier, M. P. Biology of Actinomycetes. *Annu. Rev. Microbiol.* **21**, 71–100 (1967).
52. Hopwood, D. A. *Streptomyces in nature and medicine: the antibiotic makers*. (Oxford University Press, 2007).
53. Flärdh, K. & Buttner, M. J. *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat Rev Microbiol* **7**, 36–49 (2009).
54. Barka, E. A. *et al.* Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol Mol Biol Rev* **80**, 1–43 (2016).
55. Claessen, D., Rozen, D. E., Kuipers, O. P., Søgaard-Andersen, L. & van Wezel, G. P. Bacterial solutions to multicellularity: a tale of biofilms, filaments and fruiting bodies. *Nat Rev Microbiol* **12**, 115–124 (2014).

56. Manteca, A., Mäder, U., Connolly, B. A. & Sanchez, J. A proteomic analysis *coelicolor* programmed cell death. *Proteomics* **6**, 6008–6022 (2006).
57. Manteca, Á., Fernández, M. & Sánchez, J. A death round affecting a young compartmentalized mycelium precedes aerial mycelium dismantling in confluent surface cultures of *Streptomyces antibioticus*. *Microbiology* **151**, 3689–3697 (2005).
58. Flärdh, K. Growth polarity and cell division in *Streptomyces*. *Current Opinion in Microbiology* **6**, 564–571 (2003).
59. Lee, N. *et al.* Thirty complete *Streptomyces* genome sequences for mining novel secondary metabolite biosynthetic gene clusters. *Sci Data* **7**, 55 (2020).
60. Caicedo-Montoya, C., Manzo-Ruiz, M. & Ríos-Estepa, R. Pan-Genome of the Genus *Streptomyces* and Prioritization of Biosynthetic Gene Clusters With Potential to Produce Antibiotic Compounds. *Front. Microbiol.* **12**, 677558 (2021).
61. Hoff, G., Bertrand, C., Piotrowski, E., Thibessard, A. & Leblond, P. Genome plasticity is governed by double strand break DNA repair in *Streptomyces*. *Sci Rep* **8**, 5272 (2018).
62. Schatz, A., Bugle, E. & Waksman, S. A. Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria.*. *Experimental Biology and Medicine* **55**, 66–69 (1944).
63. Lewis, K. Recover the lost art of drug discovery. *Nature* **485**, 439–440 (2012).
64. Lewis, K. The Science of Antibiotic Discovery. *Cell* **181**, 29–45 (2020).
65. Demain, A. L. & Sanchez, S. Microbial drug discovery: 80 years of progress. *J Antibiot* **62**, 5–16 (2009).
66. Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). **417**, 7 (2002).
67. Ikeda, H. *et al.* Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol* **21**, 526–531 (2003).
68. Ohnishi, Y. *et al.* Genome Sequence of the Streptomycin-Producing Microorganism *Streptomyces griseus* IFO 13350. *J Bacteriol* **190**, 4050–4060 (2008).
69. Rutledge, P. J. & Challis, G. L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Microbiol* **13**, 509–523 (2015).

70. Bode, H. B., Bethe, B., Höfs, R. & Zeeck, A. Big Effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity. *ChemBioChem* **3**, 619 (2002).
71. Seyedsayamdost, M. R. High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7266–7271 (2014).
72. Muth, G. The pSG5-based thermosensitive vector family for genome editing and gene expression in actinomycetes. *Appl Microbiol Biotechnol* **102**, 9067–9080 (2018).
73. Aubry, C., Pernodet, J.-L. & Lautru, S. Modular and Integrative Vectors for Synthetic Biology Applications in *Streptomyces spp*. *Appl Environ Microbiol* **85**, e00485-19 (2019).
74. Myronovskyi, M. & Luzhetskyy, A. Native and engineered promoters in natural product discovery. *Nat. Prod. Rep.* **33**, 1006–1019 (2016).
75. *Practical Streptomyces genetics*. (Innes, 2000).
76. Mazodier, P., Petter, R. & Thompson, C. Intergeneric conjugation between *Escherichia coli* and *Streptomyces* species. *J Bacteriol* **171**, 3583–3585 (1989).
77. Malpartida, F. & Hopwood, D. A. Molecular cloning of the whole biosynthetic pathway of a *Streptomyces antibiotic* and its expression in a heterologous host. *Nature* **309**, 462–464 (1984).
78. Gomez-Escribano, J. P. & Bibb, M. J. *Streptomyces coelicolor* as an Expression Host for Heterologous Gene Clusters. in *Methods in Enzymology* vol. 517 279–300 (Elsevier, 2012).
79. Thanapipatsiri, A., Claesen, J., Gomez-Escribano, J.-P., Bibb, M. & Thamchaipenet, A. A *Streptomyces coelicolor* host for the heterologous expression of Type III polyketide synthase genes. *Microb Cell Fact* **14**, 145 (2015).
80. Myronovskyi, M. *et al.* Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. *Metabolic Engineering* **49**, 316–324 (2018).
81. Xu, M. *et al.* Functional Genome Mining for Metabolites Encoded by Large Gene Clusters through Heterologous Expression of a Whole-Genome Bacterial Artificial Chromosome Library in *Streptomyces spp*. *Appl Environ Microbiol* **82**, 5795–5805 (2016).

82. Zhang, B. *et al.* Discovery, Biosynthesis, and Heterologous Production of Streptoseomycin, an Anti-Microaerophilic Bacteria Macrodilactone. *Org. Lett.* **20**, 2967–2971 (2018).
83. Lee, N. *et al.* Synthetic Biology Tools for Novel Secondary Metabolite Discovery in *Streptomyces*. *Journal of Microbiology and Biotechnology* **29**, 667–686 (2019).
84. Smanski, M. J. *et al.* Synthetic biology to access and expand nature's chemical diversity. *Nat Rev Microbiol* **14**, 135–149 (2016).
85. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).
86. Isaacson, W. *The code breaker: Jennifer Doudna, gene editing, and the future of the human race*. (Simon & Schuster, 2021).
87. Jansen, Ruud., Embden, Jan. D. A. van, Gaastra, Wim. & Schouls, Leo. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**, 1565–1575 (2002).
88. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol* **169**, 5429–5433 (1987).
89. Mojica, F. J. M., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J Mol Evol* **60**, 174–182 (2005).
90. Ackermann, H.-W. Tailed Bacteriophages: The Order Caudovirales. *Advances in Virus Research* vol. 51 135–201 (Elsevier, 1998).
91. Kaliniene, L. *et al.* Molecular Analysis of Arthrobacter Myovirus vB_ArtM-ArV1: We Blame It on the Tail. *J Virol* **91**, e00023-17 (2017).
92. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
93. Zheng, Y. *et al.* Endogenous Type I CRISPR-Cas: From Foreign DNA Defense to Prokaryotic Engineering. *Front. Bioeng. Biotechnol.* **8**, 62 (2020).
94. Yoshimi, K. & Mashimo, T. Genome editing technology and applications with the type I CRISPR system. *Gene and Genome Editing* **3–4**, 100013 (2022).
95. Liu, G., Lin, Q., Jin, S. & Gao, C. The CRISPR-Cas toolbox and gene editing technologies. *Molecular Cell* **82**, 333–347 (2022).

96. Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol* **18**, 67–83 (2020).
97. Wang, H., La Russa, M. & Qi, L. S. CRISPR/Cas9 in Genome Editing and Beyond. *Annu. Rev. Biochem.* **85**, 227–264 (2016).
98. Jiang, F. & Doudna, J. A. CRISPR–Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* **46**, 505–529 (2017).
99. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).
100. Maeder, M. L. *et al.* CRISPR RNA–guided activation of endogenous human genes. *Nat Methods* **10**, 977–979 (2013).
101. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nat Methods* **10**, 973–976 (2013).
102. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
103. Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
104. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
105. Church, G. M., Elowitz, M. B., Smolke, C. D., Voigt, C. A. & Weiss, R. Realizing the potential of synthetic biology. *Nat Rev Mol Cell Biol* **15**, 289–294 (2014).
106. Cameron, D. E., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. *Nat Rev Microbiol* **12**, 381–390 (2014).
107. Kutyna, D. R. *et al.* Construction of a synthetic *Saccharomyces cerevisiae* pan-genome neo-chromosome. *Nat Commun* **13**, 3628 (2022).
108. Ro, D.-K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).
109. Fredens, J. *et al.* Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **569**, 514–518 (2019).
110. Nyerges, A. *et al.* A swapped genetic code prevents viral infections and gene transfer. *Nature* **615**, 720–727 (2023).
111. Ayikpoe, R. S. *et al.* A scalable platform to discover antimicrobials of ribosomal origin. *Nat Commun* **13**, 6135 (2022).
112. Libis, V. *et al.* Multiplexed mobilization and expression of biosynthetic gene clusters. *Nat Commun* **13**, 5256 (2022).

113. Hsu, S.-Y., Lee, J., Sychla, A. & Smanski, M. J. Rational search of genetic design space for a heterologous terpene metabolic pathway in *Streptomyces*. *Metabolic Engineering* **77**, 1–11 (2023).
114. Yilmaz, S., Nyerges, A., van der Oost, J., Church, G. M. & Claassens, N. J. Towards next-generation cell factories by rational genome-scale engineering. *Nat Catal* **5**, 751–765 (2022).
115. O'Neill, J. Tackling Drug-Resistant Infections Globally: Final Report and Recommendations the Review on Antimicrobial Resistance. https://amr-review.org/sites/default/files/160525_Final%20paper_with%20cover.pdf (2016).
116. Naylor, N. R. *et al.* Estimating the burden of antimicrobial resistance: a systematic literature review. *Antimicrob Resist Infect Control* **7**, 58 (2018).
117. World Health Organization. *Antimicrobial resistance: global report on surveillance*. (World Health Organization, 2014).
118. Casadevall, A., Kontoyiannis, D. P. & Robert, V. On the Emergence of *Candida auris*: Climate Change, Azoles, Swamps, and Birds. **10**, (2019).
119. Clark, M. A. *et al.* Global food system emissions could preclude achieving the 1.5° and 2°C climate change targets. *Science* **370**, 705–708 (2020).
120. Ivanovich, C. C., Sun, T., Gordon, D. R. & Ocko, I. B. Future warming from global food consumption. *Nat. Clim. Chang.* **13**, 297–302 (2023).
121. Adamson, C. S. *et al.* Antiviral drug discovery: preparing for the next pandemic. *Chem. Soc. Rev.* **50**, 3647–3655 (2021).
122. Carlson, C. J. *et al.* Climate change increases cross-species viral transmission risk. *Nature* **607**, 555–562 (2022).
123. Lee, S. Y. *et al.* A comprehensive metabolic map for production of bio-based chemicals. *Nat Catal* **2**, 18–33 (2019).

Chapter 2

The Design-Build-Test-Learn Cycle for Metabolic Engineering of Streptomycetes

Christopher M. Whitford[1], Pablo Cruz-Morales[2,3,4], Jay D. Keasling[1,2,3,4,5,6], Tilmann Weber[1*]

[1] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[2] Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, CA, 94608, United States
[3] Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States.
[4] QB3 Institute, University of California, Berkeley, Berkeley, CA 94720, United States
[5] Department of Chemical & Biomolecular Engineering and Department of Bioengineering, University of California, Berkeley, Berkeley, CA 94720, United States.
[6] Center for Synthetic Biochemistry, Shenzhen Institutes for Advanced Technologies, Shenzhen 518055, P. R. China

*Correspondence address
Email: tiwe@biosustain.dtu.dk

Published in Essays in Biochemistry: doi.org/10.1042/EBC20200132

ABSTRACT

Streptomycetes are producers of a wide range of specialized metabolites of great medicinal and industrial importance, such as antibiotics, antifungals, or pesticides. Having been the drivers of the golden age of antibiotics in the 50s and 60s, technological advancements over the last two decades have revealed that very little of their biosynthetic potential has been exploited so far. Given the great need for new antibiotics due to the emerging antimicrobial resistance crisis, as well as the urgent need for sustainable biobased production of complex molecules, there is a great renewed interest in exploring and engineering the biosynthetic potential of streptomycetes. Here, we describe the Design-Build-Test-Learn cycle for metabolic engineering experiments in streptomycetes and how it can be used for the discovery and production of novel specialized metabolites.

SUMMARY POINTS

- Streptomycetes are prolific producers of complex specialized metabolites with great importance for modern society
- Their biosynthetic space has not yet been fully explored and exploited, as revealed by recent technological advancements
- Here, we described the most important technologies and methods of the synthetic biology Design-Build-Test-Learn cycle in streptomycetes
- In the future, tight integration of the described technologies and methods with laboratory automation is likely to rapidly increase the number of newly discovered bioactive compounds and *Streptomyces* cell-factories

INTRODUCTION

Streptomyces are ubiquitous Gram positive bacteria that produce a broad diversity of specialized metabolites (also referred to as natural products or secondary metabolites) which are widely used as antibiotic, antifungal, anti-cancer, and immunosuppressant drugs, or pesticides, and herbicides [1]. Members of this genus are slow-growing, filamentous organisms with a complex life cycle including spore germination, vegetative and aerial hyphae formation, and spore maturation [2]. Their genomes are rich in G+C and organized in linear chromosomes often displaying genomic instability. The functions required for the production of specialized metabolites, including biosynthesis, regulation, transport, and self-resistance, are encoded in genes that are usually found clustered in the genome, and consequently referred to as biosynthetic gene clusters (BGCs) [3].

The emerging antimicrobial resistance (AMR) crisis has revitalized the natural products discovery field, propelled by recent technological advances such as next generation sequencing and sophisticated genome mining. Since the golden age of antibiotics in the 60s and 70s, the antibiotics development pipeline has slowly dried out, resulting in a critical discovery void [4]. Analysis of the first whole genome sequences (WGS) from streptomycetes in the early 2000s revealed that their biosynthetic potential remained largely underexplored [5,6]. Traditional chemical screening approaches were insufficient to capture the entire potential, as only a small fraction of the BGCs are readily expressed under standard laboratory conditions [7].

Therefore, much of the engineering efforts are focused on pathway activation, elucidation, and linking compounds to BGCs and *vice versa*. Further developments that have revitalized the natural products discovery field are advances in synthetic biology, gene synthesis, advanced cultivation, automation and co-Adaptive Laboratory Evolution setups [8–12]. Other widely used methods to induce or optimize the expression of silent or poorly expressed known BGCs are media optimization [13] and elicitor screening [12]. These methods are relatively simple to implement as they do not require genetic manipulation of the native producer strains. However, to access the vast biosynthetic repertoire found in ever growing genomic databases without dependence on strain collections, or from earth's microbiomes [14] for which

isolates might not even be available, natural product research largely relies upon synthetic biology approaches.

The Design-Build-Test-Learn (DBTL) cycle is a key concept of synthetic biology borrowed from traditional engineering disciplines. By repeated iteration of the DBTL cycle, initial designs can be improved until the desired outcome or phenotype is achieved. So far, the synthetic biology DBTL cycle has been successfully applied in a number of important industrial microorganisms [15–18]. Furthermore, an increasing number of commercial and non-commercial biofoundries are being established based on the DBTL cycle coupled with extensive laboratory automation [19] .

In this review, we describe the DBTL cycle for metabolic engineering of streptomycetes with a focus on its application to discovery and production of novel specialized metabolites as well as the refactoring of biosynthetic pathways.

THE DESIGN-TEST-BUILD-LEARN CYCLE

The DBTL cycle (Figure 1) starts with the **Design** stage, in which the engineering target is defined and translated into cellular designs. This stage involves genome mining for bioparts and genome-scale metabolic model (GEM) analysis for strain engineering. In the **Build** stage, the cellular designs are implemented by engineering the target strains and pathways to obtain the desired output. During the **Test** stage, the performance of the prototypes is evaluated, typically using omics, biosensors and bioactivity screenings. In the **Learn** stage, the collected data are analysed, compared to the expected outcomes, and used to refine the cellular design for the next iteration. While the cycle is conceptually separated into four stages, in practice assignment of certain steps to a single stage is not always possible, and overlaps and sub iterations between different stages are common.

The Design Stage

The Design stage starts by defining the engineering goal. For streptomycetes, the goal usually is (over)production of a specialized metabolite, pathway

activation and elucidation, production of derivatives, or construction of *de novo* pathways and assembly lines (Figure 2). Based on the engineering goal, bioinformatic analysis and models are used to identify engineering targets in native or host strains and to mine for required parts. The Design stage also involves planning the later stages and designing the required experiments. Importantly, considerations about the amount and kind of data that will be gathered and how it will be analyzed need to be taken into account, e.g. through Design of Experiments [20]. The Design stage heavily relies on computational tools for parts mining, metabolic model analysis, pathway prediction and optimization, as well as for modelling gene cloning, genome editing and pathway assembly strategies [21–24].

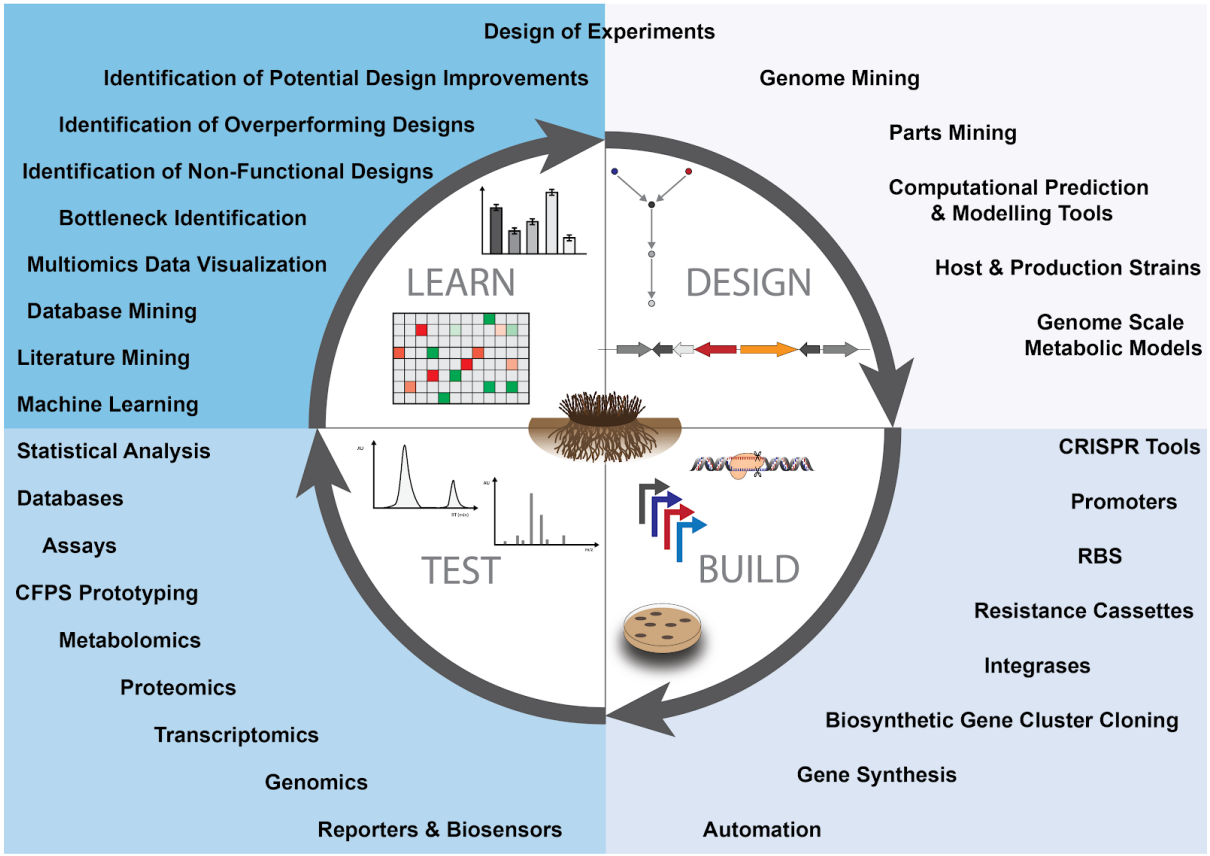


Figure 1: Overview of the DBTL-cycle for metabolic engineering of streptomycetes. The cycle starts with the Design stage and continues until the Learn stage, where bottlenecks and potential improvements are identified and subsequently used as the basis for the next iteration of the cycle.

The rapidly decreasing cost and increasing throughput of next generation sequencing has resulted in an ever growing number of available high quality streptomycete genomes (there are 2138 draft and complete genomes in the

RefSeq database as of Feb. 9, 2021) that can be mined for BGCs and parts using tools such as antiSMASH [25] and PRISM 4 [26], and databases such as MIBiG [22], IMG-ABC [27], and Streptome DB [28]. The antiSMASH database contains over 147.000 BGCs precomputed from over 25.000 unique species [29]. Both antiSMASH and the antiSMASH database link to entries in the MIBiG database, which contains curated BGCs with known function [30]. Many other tools for part identification and analysis are based on antiSMASH data, such as ARTS [31] or ClusterCAD, a database and web tool for molecular design using type I modular polyketide synthase [32]. The large number of predicted BGCs available in public databases has further given rise to tools to analyze BGC diversity, hence adding another method to identify promising BGCs and parts (Table 1). Such tools and databases include BiG-SCAPE and CORASON [33], BiG-SLiCE [34], and BiG-FAM [35].

If the desired output of a DBTL cycle is production of a given specialized metabolite, four options are available after BGC selection: (i) direct alteration of the BGC in the producer strain, (ii) engineering of the producer strain, (iii) expression of the native BGC in a heterologous host, and (iv) heterologous expression of a refactored BGC.

Over the years, streptomycete host strains for heterologous BGC expression have been developed by deletion of the native BGCs and introduction of mutations known to favor production of antibiotics [36]. These strains have low endogenous metabolic background, high precursor pools for production of target compounds, genomic stability, fast and dispersed growth, and are easy to engineer with established methods (Table 1) [36,37]. Commonly used engineered hosts include derivatives of *S. coelicolor* A3(2) [38,39], *S. albus* J1074 [40,41], as well as *S. lividans* [42,43], *S. venezuelae* [44–46], and *S. avermitilis* strains [47,48]. Like *S. avermitilis, S. fradiae, S. ambofaciens, and S. roseosporus* are important industrial producer strains of polyketide compounds and non-ribosomal peptides [49,50]. Therefore, engineering these strains into production hosts for heterologous BGC expression, as well as for combinatorial biosynthesis, has also proved promising [51,52]

Given that the potential biosynthetic space of streptomycetes can be directly explored using genome mining tools, genome-scale metabolic models (GEMs) represent a promising tool for the design of metabolic engineering

experiments in streptomycetes. GEMs can be used to predict strain engineering targets through constraint-based flux analysis and have been applied to multiple important industrial microorganisms [53]. In recent years, an increasing number of streptomycete GEMs were constructed (Table 1) [54], including for model strains such as *S. coelicolor* [55] and *S. lividans [56],* and for some non-model streptomycetes [57–59]. GEMs can be used for a wide range of predictions for metabolic engineering, summarized comprehensively by Kim et al. [53]. For host engineering, a typical strategy is to enforce the flux into the target pathway by overexpression of precursor pathways and by blocking flux into competing pathways. This can be predicted *in silico* using GEMs with tools such as OptKnock and OptFlux [60,61].

Other omics, such as transcriptomics, translatomics, interactomics, proteomics, and metabolomics can also provide valuable insights for synthetic biology experiments, as summarized extensively by Lee et al. [62], and can help in the identification of engineering targets for production of specialized metabolites [63]. In recent years, an increasing amount of omics data for streptomycetes has been published, further fueling systems biology-based insights and engineering approaches [62,64–67].

Table 1: A selection of commonly used resources for the Design stage.

Tool/Resource/Database	Description	Reference
BGC prediction / Parts Mining:		
antiSMASH	Tool for identification, annotation and analysis of biosynthetic gene clusters. Interlinks with many other bioinformatic tools such as the antiSMASH or MIBiG databases	[25]
antiSMASH database	Database with approximately 147.000 putative BGCs precomputed from over 25.000 unique species	[29]
MIBiG database	Database containing approximately 2000 manually curated BGCs	[30]
ClusterCAD	Webtool and database to guide combinatorial biosynthesis of type I polyketide synthases	[32]
BiG-SCAPE / CORASON	Tools for the analysis and exploration of BGC diversity through clustering based on sequence similarity networks and definition of gene cluster families	[33]

BiG-SLiCE	Tool for clustering and grouping into gene cluster families of massive numbers of BGCs	[34]
BiG-FAM	Database containing information on approximately 30.000 gene cluster families computed from over 1.2 million BGCs	[35]
Genome Scale Metabolic Models		
Streptomyces coelicolor iKS1317	GEM including 1317 genes, 2119 reactions, and 1581 metabolites	[55]
Streptomyces lividans iJV710 and iJV1220	Based on earlier models for *S. lividans* (iIL708 and web-published model from SurreyFBA),and integration of more recent data for *S. coelicolor* (iMK1208). iJV1220 includes 1220 genes, 1446 reactions, and 1867 metabolites	[56]
Streptomyces clavuligerus iLT1021	GEM including 1021 genes, 1494 reactions and 1360 metabolites	[57]
Streptomyces tsukubaensis NRRL 18488	Initial GEM integrated 865 reactions and 621 metabolites, and was later updated by integration of 199 additional reactions.	[58,59]
Computational Design Tools		
RetroPATH 2.0	Workflow for automated retrobiosynthesis for identification of alternative biosynthetic route designs	[68]
ClusterCAD	Web-based tool for retrosynthetic design of polyketide synthase design. It proposes domain exchanges and bioparts given a small molecule target	[27]
GEM-Path	Pathway prediction tool that combines searching the biochemical space, thermodynamic calculations, and feasibility analysis for each reaction by integration into a GEM.	[69]
PathPRED	Tool for prediction of possible pathways leading to a target compound using data from the KEGG RPAIR database.	[70]
OptKnock	Tool for identification of gene knockouts for overproduction of a target compound by decoupling growth and production.	[61]
OptFlux	Modular software platform combining several existing methods, including identification of engineering targets, OptKnock, or phenotype predictions.	[60]
Fluxer	Web based tool for computation, visualization, and analysis of genome scale metabolic model flux networks	[71]
CRISPy-web	Automatic CRISPR protospacer prediction for Cas9 and C2C2/Cas12a, also supports predictions of protospacers suitable for base editing	[21,72]
J5 / DIVA Device Editor	A web-based tool, which automates the design of multipart DNA assembly protocols including SLIC, Gibson, CPEC, and Golden Gate	[24,73]

Host Strains		
Streptomyces coelicolor M145, M1146, M1152, M1154	*S. coelicolor* M145: Plasmid free *S. coelicolor* A3(2). Most, widely studied and used laboratory strain. *S. coelicolor* M1146: Deletion of the actinorhodin, undecylprodigiosin, Coelimycin P1, and the calcium-dependent antibiotic gene clusters *S. coelicolor* M1152: Derivative of *S. coelicolor* M1146 with insertion of antibiotic production enhancing mutation rpoB[C1298T] *S. coelicolor* M1154: *S. coelicolor* M1152 rpsL[A262G]	[38,39]
Streptomyces albus J1074 , DEL14, B4	*Streptomyces albus* J1074: a tractable strain with a naturally small BGC repertoire S. albus Δ*pfk+Crp*: engineered strain with a *pfk* deletion and expression of the global regulator *crp* gene from *S. coelicolor* *S. albus* DEL14: cluster-free *S. albus* J1074 with 15 deleted BGCs, resulting in a genome minimization by 500 kb *S. albus* B4: derivative of *S. albus* DEL14 with a total of four canonical *attB* integration sites	[40,41,74]
Streptomyces lividans	*S lividans* Tk24*:* A stable, plasmid-free strain derived from S. lividans 1326 with low endogenous production of specialized metabolites that accepts methylated DNA *S. lividans* LJ1018: derivative of *S. lividans* SBT5 with a *wblA* deletion and integration of three global regulatory genes and two multi-drug efflux pump genes *S. lividans* ΔYA11: derivative of *S. lividans* TK24 with 11 deleted clusters (180 kb) and three *attB* sites	[42,43,75]
Streptomyces avermitilis	*S. avermitilis* SUKA22: genome reduced derivative of the SAP2 plasmid free *S. avermitilis* K139, with almost 20 % of the genome removed	[47,48]
Streptomyces venezuelae	*S. venezuelae* has a doubling time below 1 h and grows dispersedly, making it a very interesting host. In recent years, great focus on cell free systems	[44,46,76]

The Build Stage

During the Build stage, the cellular design is implemented by assembly of genetic constructs and engineering of the target strain. In biofoundries, many different construct versions can be easily assembled by means of automation. Production of specialized metabolites involves high metabolic costs and toxicity, and their biosynthesis is controlled by complex regulatory systems. Therefore, heterologous expression or activation of BGCs requires orthogonal regulatory elements that can be induced or that are constitutively expressed.

The most widely used promoters for BGC heterologous expression in *Streptomyces* are natural: *kasOp* from the coelimycin P1 BGC in *S. coelicolor* [77], *ermE**p from the erythromycin BGC in *Saccharopolyspora erythraea* [78], and SF14p from the i19 phage [79]. Early efforts to develop a broad range of promoters exploited some of these classic elements. *kasO*p was engineered to produce *kasO*p*, one of the strongest promoters known to date [80]. *ermE*p was used to generate conveniently short semisynthetic promoters (41 bp) [81]. In a similar fashion the promoter for ACTII-orf4, the actinorhodin positive regulator in *S. coelicolor,* was used to generate and select semisynthetic strong constitutive promoters [82].

The mining of transcriptomes for strongly expressed house-keeping genes represents an alternative approach to procure new promoters. This rather simple method has yielded strong constitutive promoters from *S. coelicolor* [83] and *S. albus*. Promoters with 200-1300 % strength increase compared with *ermE**p were identified in *S. albus* and shown to be functional in multiple streptomycetes [84]. This approach has the disadvantage that the DNA sequences obtained can be long (100s of bp) and are often not well characterized, but they are usually easy to PCR-amplify from common laboratory strains.

Shorter, synthetic regulatory RBSs and promoters have been developed using selection systems for sequences with randomized RBS, and bases flanking the -10 and -35 regions, as well as in the spacer region between them [85,86]. Recently, a panel of short constitutive and divergent promoters of various strengths has been obtained using this approach. Among them, A26 is the strongest promoter reported so far. Some of these promoters have been used to refactor and express the actinorhodin BGC in *S. albus* [87].

For inducible expression, classic systems such as the thiostrepton inducible *tipA* promoter are widely used, however these systems are not efficient and depend on toxic molecules. Expression systems that can be strictly controlled have been developed, such as the RolR-based Resorcinol inducible system from *C. glutamicum*, the Cumate inducible CymR system from *Pseudomonas putida* [88] and the tightly controlled xylose responsive XylR system from *S. avermitilis,* which has been successfully implemented in various *Streptomyces*

hosts [89]. Furthermore, theophylline-dependent riboswitches were successfully implemented in *S. coelicolor* for inducible gene expression [90].

Given that BGCs usually include all genes required for biosynthesis of the corresponding specialized metabolite, cloning of entire BGCs is of great interest for heterologous expression. However, this is still challenging, due to the high G+C content, high sequence similarities, and the size, in many cases reaching over 100 kb. Therefore, several methods were developed for targeted and untargeted BGC cloning (Table 2), already comprehensively summarized elsewhere [91,92]. Selection of the right method is crucial and depends on the size and complexity of the BGC, whether refactoring is desired, and the host selection. Combining BGC cloning and heterologous expression frequently results in the detection of new BGC encoded compounds [93–96]

The most common method for the stable introduction of heterologous BGCs in streptomycetes is the use of phage-derived serine integrases. An advantage of this strategy over CRISPR-Cas9 is that the latter is inefficient for manipulating large DNA fragments such as full BGCs [97]. In recent years, the integrase genetic toolkit has incorporated methods for multiplex integration of up to three *loci* using plasmids carrying orthogonal integrases [98]. Alternatively, a single *locus* can be integrated up to 4 times using a single plasmid with multiple orthogonal integrases [99] or five times by introducing multiple integration sites in the host strain for a single integrase [100]. Important considerations are the presence of integration sites, their location, and the fidelity of the integration. It has been shown that integrases can target pseudo sites [101] leading to off-target events and that some integration events may have negative effects, as in the case of PhiC31 mediated integration in *S. ambofaciens*, which targets the conserved *pirA* gene and causes significant changes in primary and secondary metabolism [102].

BGCs can be refactored in the native producer or for production in a heterologous host, typically either by engineering regulation, catalytic domains, or tailoring enzymes. A CRISPR-Cas9 mediated knock-in of the *kasO*p* was shown to be an efficient method for activation of silent BGCs in several streptomycetes [103]. Shao et al. successfully heterologously expressed the spectinabilin BGC using a set of heterologous promoters in a plug-and-play scaffold [104]. PCR and Gibson Assembly-based cloning and refactoring of a

streptophenazine BGC resulted in detection of over 100 streptophenazines upon expression in *S. coelicolor* M1146 [93]. Several comprehensive reviews cover the pathway refactoring workflow, available tools, as well as recent advances [105,106].

Engineering of streptomycetes benefited immensely from the development of CRISPR (Clustered regularly interspaced short palindromic repeats) tools for streptomycetes, enabling and simplifying an unprecedented level of engineering [107]. CRISPR-Cas9 was adapted by several groups for application in streptomycetes [108–111]. A nuclease deficient Cas9 further enables CRISPR interference experiments [109]. CRISPR-Cas9-based base editing systems were developed for streptomycetes, allowing double strand break-free editing of base pairs in single or multiplexed experiments [112]. Comprehensive protocols for these tools were recently published [113]. Besides CRISPR-Cas9, other nucleases have been utilized for genome engineering as well, such as Cpf1 (Cas12a) [114].

Fast and cheap methods to read and write DNA have been a major driver of synthetic biology. The applications of DNA synthesis for synthetic biology applications are manifold, ranging from synthesis of libraries, codon optimized genes, to synthesis of coding sequences for *de novo* designed proteins [115]. Given the high GC content of streptomycetes and the high number of repeats in biosynthetic genes, synthesis of large and complex genes for streptomycetes still remains challenging for many companies. Recent advances in enzymatic DNA synthesis are likely to allow synthesis of larger and more complex genes, possibly greatly benefiting synthetic biology in streptomycetes [116].

Table 2: Commonly used parts and methods used for the Build section.

Parts/ Tools / Methods	Description	Reference
Promoter		
*ermE**p	Promoter from *S. erythraea*. Was used for the construction of a panel of shorter semi-synthetic promoters	[81]
*kasO*p	From *S. coelicolor*. Engineered derivative kasOp* represents one of the strongest constitutive promoters	[80]
*tipA*p	Widely used thiostrepton inducible promoter from *S. lividans*	[117]
Synthetic promoters	Short, synthetic elements of variable strength with randomized RBS, and bases flanking the -10 and -35 regions, as well as in the spacer region in between them	[85,86]
Transcriptomics-mined promoters	Natural promoter elements located upstream of genes which transcriptional strength is assessed by transcriptomics	[83,84]
RolR	RolR-based Resorcinol inducible system from *C. glutamicum*, reaches maximum expression when induced with 400nM of Resorcinol	[88]
CymR	CymR-based inducible system from *P. putida*, responds linearly to induction with Cumate	[88]
XylR	Tightly controlled XylR system from *S. avermitilis*, strictly repressed and induced by xylose	[89]
Integrases		
PhiC31, PhiBT, PhiVWB	Most widely used serine integrases with broad host specificity, and conserved integration sites	[102,118,119]
Multiplexed integration of multiple plasmids	Multiplexed conjugation and integration using three plasmids with orthogonal integrases PhiOZJ, PhiC31, and PhiBT1	[98]
Multicopy integration of the same plasmid	MSGE: multiplexed site-specific genome engineering, requires introduction of multiple PhiC31 integration sites into the genome, integration of up to five copies aMSGE: advanced MSGE, integration of up to four plasmid copies using four different integrases	[99,100]
Cloning		
Cosmid / BAC libraries	Method for cloning of complete BGCs. Average BAC sizes are often around 100 kb	[120–124]
DiPAC	BGC cloning method based on long amplicon high fidelity PCR and Hifi DNA Assembly. Allows simultaneous refactoring and assembly into any vector. Suitable for BGCs of up to 55 kb	[125]

TAR cloning	Method utilizing high homologous recombination levels in *S. cerevisiae* for cloning of up to 300 kb. Efficiencies can be improved by pretreatment with CRISPR-Cas9	[126,127]
LLHR/RecE	Highly efficient linear-linear homologous recombination mediated by full-length RecE and RecT. For BGCs of up to 52 kb.	[128]
ExoCET	LLHR/RecE-based cloning which additionally utilizes the 3' exonuclease activity of T4 polymerase. For BGCs of up to 100 kb	[129]
CATCH	Cloning of fragments of up to 100 kb by targeted in-gel *in vitro* Cas9 digestion of lysed bacterial cells and subsequent Gibson Assembly into a cloning vector.	[130]
iCATCH	Introduction of homing endonuclease sites I-SceI and PI-PspI flanking the target BGC, digestion of gDNA in a low-melting agarose plug, followed by ligation and transformation in *E. coli*.	[131]
Integrase-based cloning	Flanking of a target BGC with integration sites and integrase catalyzed excision of the BGC, resulting in an extrachromosomal plasmid. A BGC can be knocked out by temperature-based plasmid curation.	[132]
CAPTURE	Cas12a-assisted precise targeted cloning using *in vivo* Cre-lox recombination. Uses Cas12a digestion followed by T4 polymerase exo+fill-in DNA assembly and subsequent Cre-lox mediated *in vivo* circularization.	[94]
ANT cloning	Automatic natural transformation cloning utilizing the natural competence of *Acinetobacter baylyi* ADP1. Demonstration of automated cloning of 21 small sized BGCs of up to 19 kb	[133]
CRISPR-based genome editing		
CRISPR-Cas9	pCRISPR-Cas9: pSG5 replicon, codon optimized *spcas9,* one sgRNA expression cassette, up to 100 % efficiency pCRISPR-Cas9-ScaligD: plasmid version carrying the ligase D *ligD* gene for NHEJ experiments pCRISPR-dCas9: nuclease inactivated version for CRISPRi experiments pCRISPomyces-2: pSG5 replicon, codon optimized *spcas9*, two sgRNA cassettes, efficiencies between 70 % to 100 % pCMU-4 and pCM4.4 : Plasmids based on pCRISPomyces-2 with fine-tuned Cas9 expression pKCcas9dO: pKC1139 based, codon optimized *spcas9,* one sgRNA cassette, editing efficiencies between 45 % and 54 %	[108–111]
CRISPR-BEST	CRISPR-cBEST: Cytosine deaminase mediated base editing CRISPR-aBEST: Adenine deaminase mediated base editing CRISPR-mcBEST: Csy4-based multiplexed cytosine base editing using a single synthetic sgRNA array	[112]

CRISPR-Cas12a / Cpf1	pKCCpf1: plasmid carrying a codon-optimized *cpf1* gene for single or double gene deletions. Addition of *ligD and KU* allows NHEJ experiments pSETddCpf1: nuclease deficient version for single or multiplexed CRISPRi experiments	[114]

The Test Stage

During the Test stage of a DBTL cycle focused on specialized metabolites, the performance of the cellular designs is analyzed. Strains are commonly tested for increased production of a target metabolite, activation of a BGC, or detection of metabolite derivatives. Testing can also include bioassays and basic strain characterizations (Table 3).

For metabolomic analysis, the engineered strains are grown under the target conditions, followed by extractions for metabolomic analysis. Since the methods involved can differ greatly depending on the properties of the target compound, we will refer to several reviews here, which cover metabolomics of specialized metabolites in depth [62,134]. Manual data dereplication remains a big challenge in metabolomics. In recent years, many computational tools have become available that allow automatic annotation and dereplication of metabolomics data, such as DEREPLICATOR+ [135], CSI:FingerID [136], ZODIAC [137], SIRIUS 4 [138], or CliqueMS [139]. Commercial databases such as AntiBase or the Dictionary of Natural Products contain tens of thousands and hundreds of thousands of entries, respectively [140,141]. Comparing metabolomics results against these databases is of great value for dereplication efforts, especially for specialized metabolite discovery. Open access databases such as GNPS, a community driven resource for sharing, annotation, and knowledge dissemination of MS/MS data, as well as for direct analysis and dereplication of raw data; StreptomeDB, a database for specialized metabolites produced by streptomycetes; or the Natural Products Atlas will likely become invaluable resources as more data are deposited [28,142,143]. A comprehensive overview of relevant databases was recently published by Sorokina et al. [144].

Besides metabolomics, biosensors can be used for testing of engineered strains. Given the structural diversity of specialized metabolites and the complex regulatory networks controlling biosynthesis, the number of available biosensor systems remains limited, but several groups reported functioning biosensors [145–148].

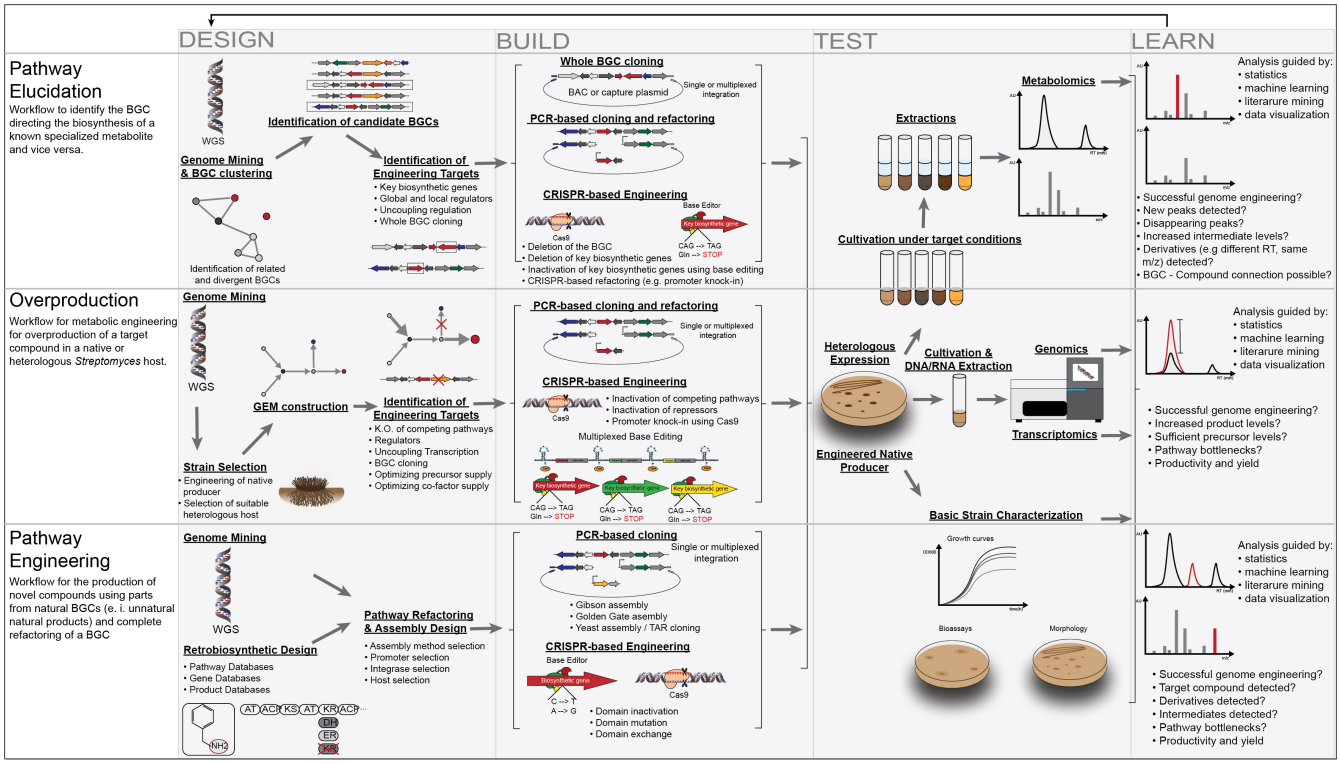


Figure 2: Exemplary engineering workflows for different experiments. After each iteration, the DBTL cycle can be repeated using the results from the Learn stage as input for the next iteration. The shown workflows are simplified examples and can differ greatly from real experiments depending on the goal.

Especially for antibiotics and other bioactive compounds, bioassays and co-cultivations are well established methods that date back to the early days of antibiotic discovery. Bioactivity-guided fractionation is also common. Here, chromatographic fractions are tested for bioactivity, followed by further analysis and purification of active fractions. An emerging technology for biosynthetic pathway prototyping is cell free protein synthesis (CFPS) through *in vitro* transcription and translation (TX-TL). CFPS systems were developed for *S. venezuelae* [45,46], *S. lividans* and *S. coelicolor* [149], and further expanded to species such as *S. rimosus* B2659, *S. roseosporus* NRRL11379, and *S. sp.* F4474 [150].

The Learn Stage

A typical DBTL project involves multiple rounds of experimentation until the desired specifications are achieved. Due to the large gaps in knowledge about the physiology, natural roles, regulation, and biosynthesis of natural products, as well as the inherent issues with reproducibility in experimental biology this stage is widely de-emphasized and the rationalization of subsequent DBTL cycles (e. g. learning) is rather limited to obvious adjustments to methods and experimental design [152].
Machine learning holds the promise of data-driven empirical calculations to guide the design step in subsequent cycles, even without deep mechanistic understanding of the biological systems. Tools such as the Automated Recommendation Tool (ART) use machine learning with a Bayesian approach with the assumption that the inputs of an experimental set up (genetic makeup, promoter combinations, etc) is predictive of the response (production of a metabolite, titer). ART can take the results form a typical synthetic biology experiment (100's of data points) and generate recommendations useful for the design stage of a subsequent cycle iteration [153].

The advent of automation has allowed the development of pipelines that integrate large genomic datasets, genome scale metabolic model simulations, fast construction of genetic parts at low cost and high throughput platforms for analytical chemistry. Rationalization of such large experimental outputs can be efficiently done by taking advantage of machine learning, an excellent guide for the use of multi-omics datasets to analyze, learn and improve the performance of bioengineered strains using machine learning has been recently published [152]. Furthermore, a recent example of the amalgamation of mechanistic and machine learning models in a large synthetic biology experiment was applied to the engineering of the tryptophan production in *Saccharomyces cerevisiae* leading to significant increase in production [154], application of such methods in streptomycetes is yet to be done.

CHALLENGES AND CONCLUSIONS

In practice, the implementation of the DBTL cycle to streptomycetes has been rather challenging. Their life cycle hampers the pace of experiments, their genetic instability and phenotypic diversity impacts reproducibility, filamentous growth complicates the establishment of automated workflows, while complex regulation hinders *in silico* predictions and modelling. Tackling these serious hurdles is a major challenge as they are traits intimately linked to the outstanding ability of streptomycetes to produce complex metabolites [155–158]. Nevertheless, the desire to exploit this ability, which has driven about 80 years of fundamental and applied research by industrial and academic scientists [159], is still fueled by the antimicrobial resistance crisis and the need for sustainable production methods for increasingly complex chemicals and materials. Finally, recent advances in the understanding of the biology of *Streptomyces*, genome editing, automation, machine learning and analytical chemistry as well as the exponential growth of high quality entries in publicly available genetic and chemical databases will eventually lead to better automated pipelines for the design and construction of heterologous pathways, host engineering and analytical methods to put molecules into vials at an unprecedented pace.

Author Contributions

CMW, PCM, JDK and TW conceptualized the review. CMW and PCM wrote the review. All authors edited, corrected and approved the manuscript.

Declaration of Interest

The authors declare no conflicting interests.

Acknowledgments and Funding

The work of the authors is funded by grants from the Novo Nordisk Foundation [NNF16OC0021746 to TW; NNF20CC0035580 to JDK, TW] and by the U.S. Department of Energy (the Co-Optimization of Fuels & Engines project sponsored by the U.S. DOE Office of Energy Efficiency and Renewable Energy, Bioenergy Technologies and Vehicle Technologies Offices and the Joint BioEnergy Institute, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research [Contract DEAC02-05CH11231 between DOE and Lawrence Berkeley National Laboratory To JDK]).

REFERENCES

1. Katz L, Baltz RH. Natural product discovery: past, present, and future. J Ind Microbiol Biotechnol. 2016;43: 155–176.
2. Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, Klenk H-P, et al. Taxonomy, Physiology, and Natural Products of Actinobacteria. Microbiol Mol Biol Rev. 2016;80: 1–43.
3. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, et al. Minimum Information about a Biosynthetic Gene cluster. Nat Chem Biol. 2015;11: 625–631.
4. The Science of Antibiotic Discovery. Cell. 2020;181: 29–45.
5. Bentley SD, Chater KF, Cerdeño-Tárraga A-M, Challis GL, Thomson NR, James KD, et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature. 2002;417: 141–147.
6. Ōmura S, Ikeda H, Ishikawa J, Hanamoto A, Takahashi C, Shinose M, et al. Genome sequence of an industrial microorganism *Streptomyces avermitilis*: Deducing the ability of producing secondary metabolites. Proc Natl Acad Sci U S A. 2001;98: 12215–12220.
7. Hoskisson PA, Seipke RF. Cryptic or Silent? The Known Unknowns, Unknown Knowns, and Unknown Unknowns of Secondary Metabolism. mBio. 2020. doi:10.1128/mbio.02642-20
8. Charusanti P, Fong NL, Nagarajan H, Pereira AR, Li HJ, Abate EA, et al. Exploiting adaptive laboratory evolution of *Streptomyces clavuligerus* for antibiotic discovery and overproduction. PLoS One. 2012;7: e33727.
9. Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP, et al. A new antibiotic kills pathogens without detectable resistance. Nature. 2015;517: 455–459.
10. Berdy B, Spoering AL, Ling LL, Epstein SS. In situ cultivation of previously uncultivable microorganisms using the ichip. Nat Protoc. 2017;12: 2232–2242.
11. Wilson MC, Mori T, Rückert C, Uria AR, Helf MJ, Takada K, et al. An environmental bacterial taxon with a large and distinct metabolic repertoire. Nature. 2014;506: 58–62.
12. Xu F, Wu Y, Zhang C, Davis KM, Moon K, Bushin LB, et al. A genetics-free method for high-throughput discovery of cryptic microbial metabolites. Nat Chem Biol. 2019;15: 161–168.

13. Gao H, Liu M, Liu J, Dai H, Zhou X, Liu X, et al. Medium optimization for the production of avermectin B1a by *Streptomyces avermitilis* 14-12A using response surface methodology. Bioresour Technol. 2009;100: 4012–4016.
14. Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, et al. A genomic catalog of Earth's microbiomes. Nat Biotechnol. 2020. doi:10.1038/s41587-020-0718-6
15. Opgenorth P, Costello Z, Okada T, Goyal G, Chen Y, Gin J, et al. Lessons from Two Design–Build–Test–Learn Cycles of Dodecanol Production in *Escherichia coli* Aided by Machine Learning. ACS Synthetic Biology. 2019. pp. 1337–1351. doi:10.1021/acssynbio.9b00020
16. Carbonell P, Jervis AJ, Robinson CJ, Yan C, Dunstan M, Swainston N, et al. An automated Design-Build-Test-Learn pipeline for enhanced microbial production of fine chemicals. Commun Biol. 2018;1: 66.
17. Miskovic L, Alff-Tuomala S, Soh KC, Barth D, Salusjärvi L, Pitkänen J-P, et al. A design–build–test cycle using modeling and experiments reveals interdependencies between upper glycolysis and xylose uptake in recombinant *S. cerevisiae* and improves predictive capabilities of large-scale kinetic models. Biotechnology for Biofuels. 2017. doi:10.1186/s13068-017-0838-5
18. Feith A, Schwentner A, Teleki A, Favilli L, Blombach B, Takors R. Streamlining the Analysis of Dynamic C-Labeling Patterns for the Metabolic Engineering of as l-Histidine Production Host. Metabolites. 2020;10. doi:10.3390/metabo10110458
19. Hillson N, Caddick M, Cai Y, Carrasco JA, Chang MW, Curach NC, et al. Building a global alliance of biofoundries. Nat Commun. 2019;10: 2040.
20. Gilman J, Walls L, Bandiera L, Menolascina F. Statistical Design of Experiments for Synthetic Biology. ACS Synth Biol. 2021. doi:10.1021/acssynbio.0c00385
21. Blin K, Pedersen LE, Weber T, Lee SY. CRISPy-web: An online resource to design sgRNAs for CRISPR applications. Synthetic and Systems Biotechnology. 2016. pp. 118–121. doi:10.1016/j.synbio.2016.01.003
22. Carbonell P, Currin A, Jervis AJ, Rattray NJW, Swainston N, Yan C, et al. Bioinformatics for the synthetic biology of natural products: integrating across the Design–Build–Test cycle. Natural Product Reports. 2016. pp. 925–932. doi:10.1039/c6np00018e

23. Ren H, Shi C, Zhao H. Computational Tools for Discovering and Engineering Natural Product Biosynthetic Pathways. iScience. 2020;23: 100795.
24. Hillson NJ, Rosengarten RD, Keasling JD. j5 DNA assembly design automation software. ACS Synth Biol. 2012;1: 14–21.
25. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Research. 2019. pp. W81–W87. doi:10.1093/nar/gkz310
26. Skinnider MA, Johnston CW, Gunabalasingam M, Merwin NJ, Kieliszek AM, MacLellan RJ, et al. Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. Nat Commun. 2020;11: 6058.
27. Palaniappan K, Chen I-MA, Chu K, Ratner A, Seshadri R, Kyrpides NC, et al. IMG-ABC v.5.0: an update to the IMG/Atlas of Biosynthetic Gene Clusters Knowledgebase. Nucleic Acids Res. 2019;48: D422–D430.
28. Moumbock AFA, Gao M, Qaseem A, Li J, Kirchner PA, Ndingkokhar B, et al. StreptomeDB 3.0: an updated compendium of streptomycetes natural products. Nucleic Acids Res. 2021;49: D600–D604.
29. Blin K, Shaw S, Kautsar SA, Medema MH, Weber T. The antiSMASH database version 3: increased taxonomic coverage and new query features for modular enzymes. Nucleic Acids Res. 2020. doi:10.1093/nar/gkaa978
30. Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hooft JJJ, et al. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic Acids Res. 2020;48: D454–D458.
31. Mungan MD, Alanjary M, Blin K, Weber T, Medema MH, Ziemert N. ARTS 2.0: feature updates and expansion of the Antibiotic Resistant Target Seeker for comparative genome mining. Nucleic Acids Res. 2020;48: W546–W552.
32. Eng CH, Backman TWH, Bailey CB, Magnan C, García Martín H, Katz L, et al. ClusterCAD: a computational platform for type I modular polyketide synthase design. Nucleic Acids Res. 2018;46: D509–D515.
33. Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI, et al. A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol. 2020;16: 60–68.

34. Kautsar SA, van der Hooft JJJ, de Ridder D, Medema MH. BiG-SLiCE: A Highly Scalable Tool Maps the Diversity of 1.2 Million Biosynthetic Gene Clusters. GigaScience. 2021;10: giaa154.
35. Kautsar SA, Blin K, Shaw S, Weber T, Medema MH. BiG-FAM: the biosynthetic gene cluster families database. Nucleic Acids Res. 2020. doi:10.1093/nar/gkaa812
36. Myronovskyi M, Luzhetskyy A. Heterologous production of small molecules in the optimized *Streptomyces* hosts. Nat Prod Rep. 2019;36: 1281–1294.
37. Huo L, Hug JJ, Fu C, Bian X, Zhang Y, Müller R. Heterologous expression of bacterial natural product biosynthetic pathways. Nat Prod Rep. 2019;36: 1412–1436.
38. Gomez-Escribano JP, Bibb MJ. Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: from genome mining to manipulation of biosynthetic pathways. Journal of Industrial Microbiology & Biotechnology. 2014. pp. 425–431. doi:10.1007/s10295-013-1348-5
39. Gomez-Escribano JP, Bibb MJ. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. Microb Biotechnol. 2011;4: 207–215.
40. Kallifidas D, Jiang G, Ding Y, Luesch H. Rational engineering of *Streptomyces albus* J1074 for the overexpression of secondary metabolite gene clusters. Microb Cell Fact. 2018;17: 25.
41. Myronovskyi M, Rosenkränzer B, Nadmid S, Pujic P, Normand P, Luzhetskyy A. Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. Metab Eng. 2018;49: 316–324.
42. Ahmed Y, Rebets Y, Estévez MR, Zapp J, Myronovskyi M, Luzhetskyy A. Engineering of *Streptomyces lividans* for heterologous expression of secondary metabolite gene clusters. Microb Cell Fact. 2020;19: 5.
43. Peng Q, Gao G, Lü J, Long Q, Chen X, Zhang F, et al. Engineered *Streptomyces lividans* Strains for Optimal Identification and Expression of Cryptic Biosynthetic Gene Clusters. Frontiers in Microbiology. 2018. doi:10.3389/fmicb.2018.03042
44. Yin S, Li Z, Wang X, Wang H, Jia X, Ai G, et al. Heterologous expression of oxytetracycline biosynthetic gene cluster in *Streptomyces venezuelae*

WVR2006 to improve production level and to alter fermentation process. Appl Microbiol Biotechnol. 2016;100: 10563–10572.

45. Moore SJ, Lai H-E, Needham H, Polizzi KM, Freemont PS. *Streptomyces venezuelae* TX-TL - a next generation cell-free synthetic biology tool. Biotechnol J. 2017;12. doi:10.1002/biot.201600678
46. Moore SJ, Lai H-E, Chee S-M, Toh M, Coode S, Chengan K, et al. A *Streptomyces venezuelae* Cell-Free Toolkit for Synthetic Biology. ACS Synth Biol. 2021. doi:10.1021/acssynbio.0c00581
47. Komatsu M, Komatsu K, Koiwai H, Yamada Y, Kozone I, Izumikawa M, et al. Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. ACS Synth Biol. 2013;2: 384–396.
48. Komatsu M, Uchiyama T, Omura S, Cane DE, Ikeda H. Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. Proc Natl Acad Sci U S A. 2010;107: 2646–2651.
49. Baltz RH. *Streptomyces* and *Saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters. J Ind Microbiol Biotechnol. 2010;37: 759–772.
50. Baltz RH. Genetic manipulation of secondary metabolite biosynthesis for improved production in *Streptomyces* and other actinomycetes. J Ind Microbiol Biotechnol. 2016;43: 343–370.
51. Baltz RH, Brian P, Miao V, Wrigley SK. Combinatorial biosynthesis of lipopeptide antibiotics in *Streptomyces roseosporus*. J Ind Microbiol Biotechnol. 2006;33: 66–74.
52. Baltz RH. Combinatorial biosynthesis of cyclic lipopeptide antibiotics: a model for synthetic biology to accelerate the evolution of secondary metabolite biosynthetic pathways. ACS Synth Biol. 2014;3: 748–758.
53. Kim B, Kim WJ, Kim DI, Lee SY. Applications of genome-scale metabolic network model in metabolic engineering. Journal of Industrial Microbiology & Biotechnology. 2015. pp. 339–348. doi:10.1007/s10295-014-1554-9
54. Mohite OS, Weber T, Kim HU, Lee SY. Genome-Scale Metabolic Reconstruction of Actinomycetes for Antibiotics Production. Biotechnol J. 2019;14: e1800377.
55. Kumelj T, Sulheim S, Wentzel A, Almaas E. Predicting Strain Engineering Strategies Using iKS1317: A Genome-Scale Metabolic Model of *Streptomyces coelicolor*. Biotechnol J. 2019;14: e1800180.

56. Valverde JR, Gullón S, Mellado RP. Modelling the metabolism of protein secretion through the Tat route in *Streptomyces lividans*. BMC Microbiology. 2018. doi:10.1186/s12866-018-1199-3
57. Toro L, Pinilla L, Avignone-Rossa C, Ríos-Estepa R. An enhanced genome-scale metabolic reconstruction of *Streptomyces clavuligerus* identifies novel strain improvement strategies. Bioprocess Biosyst Eng. 2018;41: 657–669.
58. Wang C, Liu J, Liu H, Liang S, Wen J. Combining metabolomics and network analysis to improve tacrolimus production in *Streptomyces tsukubaensis* using different exogenous feedings. Journal of Industrial Microbiology & Biotechnology. 2017. pp. 1527–1540. doi:10.1007/s10295-017-1974-4
59. Huang D, Li S, Xia M, Wen J, Jia X. Genome-scale metabolic network guided engineering of *Streptomyces tsukubaensis* for FK506 production improvement. Microb Cell Fact. 2013;12: 52.
60. Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, Pinto JP, et al. OptFlux: an open-source software platform for in silico metabolic engineering. BMC Syst Biol. 2010;4: 45.
61. Burgard AP, Pharkya P, Maranas CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng. 2003;84: 647–657.
62. Lee N, Hwang S, Kim W, Lee Y, Kim JH, Cho S, et al. Systems and synthetic biology to elucidate secondary metabolite biosynthetic gene clusters encoded in *Streptomyces* genomes. Natural Product Reports. 2021. doi:10.1039/d0np00071j
63. Palazzotto E, Weber T. Omics and multi-omics approaches to study the biosynthesis of secondary metabolites in microorganisms. Curr Opin Microbiol. 2018;45: 109–116.
64. Kim W, Hwang S, Lee N, Lee Y, Cho S, Palsson B, et al. Transcriptome and translatome profiles of Streptomyces species in different growth phases. Sci Data. 2020;7: 138.
65. Lee Y, Lee N, Hwang S, Kim W, Jeong Y, Cho S, et al. Genome-scale determination of 5′ and 3′ boundaries of RNA transcripts in *Streptomyces* genomes. Sci Data. 2020;7: 436.
66. Lee Y, Lee N, Jeong Y, Hwang S, Kim W, Cho S, et al. The Transcription Unit Architecture of *Streptomyces lividans* TK24. Frontiers in Microbiology. 2019. doi:10.3389/fmicb.2019.02074

67. Lee Y, Lee N, Hwang S, Kim K, Kim W, Kim J, et al. System-level understanding of gene expression and regulation for engineering secondary metabolite production in *Streptomyces*. Journal of Industrial Microbiology & Biotechnology. 2020. pp. 739–752. doi:10.1007/s10295-020-02298-0
68. Delépine B, Duigou T, Carbonell P, Faulon J-L. RetroPath2.0: A retrosynthesis workflow for metabolic engineers. Metab Eng. 2018;45: 158–170.
69. Campodonico MA, Andrews BA, Asenjo JA, Palsson BO, Feist AM. Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm, GEM-Path. Metab Eng. 2014;25: 140–158.
70. Moriya Y, Shigemizu D, Hattori M, Tokimatsu T, Kotera M, Goto S, et al. PathPred: an enzyme-catalyzed metabolic pathway prediction server. Nucleic Acids Res. 2010;38: W138–43.
71. Hari A, Lobo D. Fluxer: a web application to compute, analyze and visualize genome-scale metabolic flux networks. Nucleic Acids Res. 2020;48: W427–W435.
72. Blin K, Shaw S, Tong Y, Weber T. Designing sgRNAs for CRISPR-BEST base editing applications with CRISPy-web 2.0. Synth Syst Biotechnol. 2020;5: 99–102.
73. Chen J, Densmore D, Ham TS, Keasling JD, Hillson NJ. DeviceEditor visual biological CAD canvas. J Biol Eng. 2012;6: 1.
74. Chater KF, Wilde LC. Restriction of a bacteriophage of *Streptomyces albus* G involving endonuclease SalI. J Bacteriol. 1976;128: 644–650.
75. Cruz-Morales P, Vijgenboom E, Iruegas-Bocardo F, Girard G, Yáñez-Guerra LA, Ramos-Aboites HE, et al. The genome sequence of *Streptomyces lividans* 66 reveals a novel tRNA-dependent peptide biosynthetic system within a metal-related genomic island. Genome Biol Evol. 2013;5: 1165–1175.
76. Phelan RM, Sachs D, Petkiewicz SJ, Barajas JF, Blake-Hedges JM, Thompson MG, et al. Development of Next Generation Synthetic Biology Tools for Use in *Streptomyces venezuelae*. ACS Synth Biol. 2017;6: 159–166.
77. Takano E, Kinoshita H, Mersinias V, Bucca G, Hotchkiss G, Nihira T, et al. A bacterial hormone (the SCB1) directly controls the expression of a pathway-specific regulatory gene in the cryptic type I polyketide

biosynthetic gene cluster of *Streptomyces coelicolor*. Mol Microbiol. 2005;56: 465–479.

78. Bibb MJ, Janssen GR, Ward JM. Cloning and analysis of the promoter region of the erythromycin resistance gene (ermE) of *Streptomyces erythraeus*. Gene. 1985;38: 215–226.
79. Labes G, Bibb M, Wohlleben W. Isolation and characterization of a strong promoter element from the *Streptomyces ghanaensis* phage I19 using the gentamicin resistance gene (aacC1) of Tn 1696 as reporter. Microbiology. 1997;143 (Pt 5): 1503–1512.
80. Wang W, Li X, Wang J, Xiang S, Feng X, Yang K. An engineered strong promoter for streptomycetes. Appl Environ Microbiol. 2013;79: 4484–4492.
81. Siegl T, Tokovenko B, Myronovskyi M, Luzhetskyy A. Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. Metab Eng. 2013;19: 98–106.
82. Sohoni SV, Fazio A, Workman CT, Mijakovic I, Lantz AE. Synthetic promoter library for modulation of actinorhodin production in *Streptomyces coelicolor* A3(2). PLoS One. 2014;9: e99701.
83. Li S, Wang J, Li X, Yin S, Wang W, Yang K. Genome-wide identification and evaluation of constitutive promoters in streptomycetes. Microb Cell Fact. 2015;14: 172.
84. Luo Y, Zhang L, Barton KW, Zhao H. Systematic Identification of a Panel of Strong Constitutive Promoters from *Streptomyces albus*. ACS Synth Biol. 2015;4: 1001–1010.
85. Horbal L, Siegl T, Luzhetskyy A. A set of synthetic versatile genetic control elements for the efficient expression of genes in Actinobacteria. Sci Rep. 2018;8: 1–13.
86. Bai C, Zhang Y, Zhao X, Hu Y, Xiang S, Miao J, et al. Exploiting a precise design of universal synthetic modular regulatory elements to unlock the microbial natural products in *Streptomyces*. Proc Natl Acad Sci U S A. 2015;112: 12181–12186.
87. Ji C-H, Kim H, Kang H-S. Synthetic Inducible Regulatory Systems Optimized for the Modulation of Secondary Metabolite Production in *Streptomyces*. ACS Synth Biol. 2019;8: 577–586.
88. Horbal L, Fedorenko V, Luzhetskyy A. Novel and tightly regulated resorcinol and cumate-inducible expression systems for *Streptomyces* and other actinobacteria. Appl Microbiol Biotechnol. 2014;98: 8641–8655.

89. Noguchi Y, Kashiwagi N, Uzura A, Ogino C, Kondo A, Ikeda H, et al. Development of a strictly regulated xylose-induced expression system in *Streptomyces*. Microb Cell Fact. 2018;17: 151.
90. Rudolph MM, Vockenhuber M-P, Suess B. Conditional control of gene expression by synthetic riboswitches in *Streptomyces coelicolor*. Methods Enzymol. 2015;550: 283–299.
91. Nah H-J, Pyeon H-R, Kang S-H, Choi S-S, Kim E-S. Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in *Streptomyces* Species. Frontiers in Microbiology. 2017. doi:10.3389/fmicb.2017.00394
92. Lin Z, Nielsen J, Liu Z. Bioprospecting Through Cloning of Whole Natural Product Biosynthetic Gene Clusters. Front Bioeng Biotechnol. 2020;8: 526.
93. Bauman KD, Li J, Murata K, Mantovani SM, Dahesh S, Nizet V, et al. Refactoring the Cryptic Streptophenazine Biosynthetic Gene Cluster Unites Phenazine, Polyketide, and Nonribosomal Peptide Biochemistry. Cell Chem Biol. 2019;26: 724–736.e7.
94. Enghiad B, Huang C, Guo F, Jiang G, Wang B, Tabatabaei SK, et al. Cas12a-assisted precise targeted cloning using in vivo Cre-lox recombination. Nat Commun. 2021;12: 1171.
95. Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V, et al. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. Proc Natl Acad Sci U S A. 2014;111: 1957–1962.
96. Bilyk O, Sekurova ON, Zotchev SB, Luzhetskyy A. Cloning and Heterologous Expression of the Grecocycline Biosynthetic Gene Cluster. PLoS One. 2016;11: e0158682.
97. Smanski MJ, Zhou H, Claesen J, Shen B, Fischbach MA, Voigt CA. Synthetic biology to access and expand nature's chemical diversity. Nat Rev Microbiol. 2016;14: 135–149.
98. Ko B, D'Alessandro J, Douangkeomany L, Stumpf S, deButts A, Blodgett J. Construction of a new integrating vector from actinophage ϕOZJ and its use in multiplex *Streptomyces* transformation. J Ind Microbiol Biotechnol. 2020;47: 73–81.
99. Li L, Wei K, Liu X, Wu Y, Zheng G, Chen S, et al. aMSGE: advanced multiplex site-specific genome engineering with orthogonal modular recombinases in actinomycetes. Metab Eng. 2019;52: 153–167.

100. Li L, Zheng G, Chen J, Ge M, Jiang W, Lu Y. Multiplexed site-specific genome engineering for overproducing bioactive secondary metabolites in actinomycetes. Metab Eng. 2017;40: 80–92.
101. Combes P, Till R, Bee S, Smith MCM. The streptomyces genome contains multiple pseudo-attB sites for the (phi)C31-encoded site-specific recombination system. J Bacteriol. 2002;184: 5746–5752.
102. Talà A, Damiano F, Gallo G, Pinatel E, Calcagnile M, Testini M, et al. Pirin: A novel redox-sensitive modulator of primary and secondary metabolism in *Streptomyces*. Metab Eng. 2018;48: 254–268.
103. Zhang MM, Wong FT, Wang Y, Luo S, Lim YH, Heng E, et al. CRISPR-Cas9 strategy for activation of silent *Streptomyces* biosynthetic gene clusters. Nat Chem Biol. 2017. doi:10.1038/nchembio.2341
104. Shao Z, Rao G, Li C, Abil Z, Luo Y, Zhao H. Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold. ACS Synth Biol. 2013;2: 662–669.
105. Tan G-Y, Liu T. Rational synthetic pathway refactoring of natural products biosynthesis in actinobacteria. Metab Eng. 2017;39: 228–236.
106. Luo Y, Enghiad B, Zhao H. New tools for reconstruction and heterologous expression of natural product biosynthetic gene clusters. Nat Prod Rep. 2016;33: 174–182.
107. Mitousis L, Thoma Y, Musiol-Kroll EM. An Update on Molecular Tools for Genetic Engineering of Actinomycetes—The Source of Important Antibiotics and Other Valuable Compounds. Antibiotics. 2020. p. 494. doi:10.3390/antibiotics9080494
108. Cobb RE, Wang Y, Zhao H. High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. ACS Synth Biol. 2015;4: 723–728.
109. Tong Y, Charusanti P, Zhang L, Weber T, Lee SY. CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. ACS Synth Biol. 2015;4: 1020–1029.
110. Huang H, Zheng G, Jiang W, Hu H, Lu Y. One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. Acta Biochim Biophys Sin. 2015;47: 231–243.
111. Ye S, Enghiad B, Zhao H, Takano E. Fine-tuning the regulation of Cas9 expression levels for efficient CRISPR-Cas9 mediated recombination in *Streptomyces*. J Ind Microbiol Biotechnol. 2020;47: 413–423.

112. Tong Y, Whitford CM, Robertsen HL, Blin K, Jørgensen TS, Klitgaard AK, et al. Highly efficient DSB-free base editing for streptomycetes with CRISPR-BEST. Proc Natl Acad Sci U S A. 2019;116: 20366–20375.
113. Tong Y, Whitford CM, Blin K, Jørgensen TS, Weber T, Lee SY. CRISPR-Cas9, CRISPRi and CRISPR-BEST-mediated genetic manipulation in streptomycetes. Nat Protoc. 2020;15: 2470–2502.
114. Li L, Wei K, Zheng G, Liu X, Chen S, Jiang W, et al. CRISPR-Cpf1-Assisted Multiplex Genome Editing and Transcriptional Repression in *Streptomyces*. Appl Environ Microbiol. 2018;84. doi:10.1128/AEM.00827-18
115. Huang P-S, Boyken SE, Baker D. The coming of age of de novo protein design. Nature. 2016;537: 320–327.
116. Palluk S, Arlow DH, de Rond T, Barthel S, Kang JS, Bector R, et al. De novo DNA synthesis using polymerase-nucleotide conjugates. Nat Biotechnol. 2018;36: 645–650.
117. Murakami T, Holt TG, Thompson CJ. Thiostrepton-induced gene expression in *Streptomyces lividans*. J Bacteriol. 1989;171: 1459–1466.
118. Baltz RH. *Streptomyces* temperate bacteriophage integration systems for stable genetic engineering of actinomycetes (and other organisms). J Ind Microbiol Biotechnol. 2012;39: 661–672.
119. Van Mellaert L, Mei L, Lammertyn E, Schacht S, Anné J. Site-specific integration of bacteriophage VWB genome into *Streptomyces venezuelae* and construction of a VWB-based integrative vector. Microbiology. 1998;144 (Pt 12): 3351–3358.
120. Deng Q, Zhou L, Luo M, Deng Z, Zhao C. Heterologous expression of Avermectins biosynthetic gene cluster by construction of a Bacterial Artificial Chromosome library of the producers. Synth Syst Biotechnol. 2017;2: 59–64.
121. Liu H, Jiang H, Haltli B, Kulowski K, Muszynska E, Feng X, et al. Rapid Cloning and Heterologous Expression of the Meridamycin Biosynthetic Gene Cluster Using a Versatile *Escherichia coli–Streptomyces* Artificial Chromosome Vector, pSBAC⊥. Journal of Natural Products. 2009. pp. 389–395. doi:10.1021/np8006149
122. Xu M, Wang Y, Zhao Z, Gao G, Huang S-X, Kang Q, et al. Functional Genome Mining for Metabolites Encoded by Large Gene Clusters through Heterologous Expression of a Whole-Genome Bacterial Artificial Chromosome Library in *Streptomyces* spp. Applied and Environmental Microbiology. 2016. pp. 5795–5805. doi:10.1128/aem.01383-16

123. Nah H-J, Woo M-W, Choi S-S, Kim E-S. Precise cloning and tandem integration of large polyketide biosynthetic gene cluster using *Streptomyces* artificial chromosome system. Microb Cell Fact. 2015;14: 140.
124. Clos J, Zander-Dinse D. Cosmid Library Construction and Functional Cloning. Methods Mol Biol. 2019;1971: 123–140.
125. Greunke C, Duell ER, D'Agostino PM, Glöckle A, Lamm K, Gulder TAM. Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential. Metab Eng. 2018;47: 334–345.
126. Lee NCO, Larionov V, Kouprina N. Highly efficient CRISPR/Cas9-mediated TAR cloning of genes and chromosomal loci from complex genomes in yeast. Nucleic Acids Res. 2015;43: e55.
127. Kouprina N, Larionov V. Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. Chromosoma. 2016;125: 621–632.
128. Fu J, Bian X, Hu S, Wang H, Huang F, Seibert PM, et al. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. Nat Biotechnol. 2012;30: 440–446.
129. Wang H, Li Z, Jia R, Yin J, Li A, Xia L, et al. ExoCET: exonuclease in vitro assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes. Nucleic Acids Res. 2018;46: 2697.
130. Jiang W, Zhao X, Gabrieli T, Lou C, Ebenstein Y, Zhu TF. Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. Nat Commun. 2015;6: 8101.
131. Wang J, Lu A, Liu J, Huang W, Wang J, Cai Z, et al. iCatch: a new strategy for capturing large DNA fragments using homing endonucleases. Acta Biochim Biophys Sin . 2019;51: 97–103.
132. Du D, Wang L, Tian Y, Liu H, Tan H, Niu G. Genome engineering and direct cloning of antibiotic gene clusters via phage ϕBT1 integrase-mediated site-specific recombination in *Streptomyces*. Sci Rep. 2015;5: 8740.
133. Jiang X, Palazzotto E, Wybraniec E, Munro LJ, Zhang H, Kell DB, et al. Automating Cloning by Natural Transformation. ACS Synth Biol. 2020;9: 3228–3235.
134. Hollywood KA, Schmidt K, Takano E, Breitling R. Metabolomics tools for the synthetic biology of natural products. Current Opinion in Biotechnology. 2018. pp. 114–120. doi:10.1016/j.copbio.2018.02.015

135. Mohimani H, Gurevich A, Shlemov A, Mikheenko A, Korobeynikov A, Cao L, et al. Dereplication of microbial metabolites through database search of mass spectra. Nat Commun. 2018;9: 4035.
136. Dührkop K, Shen H, Meusel M, Rousu J, Böcker S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. Proc Natl Acad Sci U S A. 2015;112: 12580–12585.
137. Ludwig M, Nothias L-F, Dührkop K, Koester I, Fleischauer M, Hoffmann MA, et al. Database-independent molecular formula annotation using Gibbs sampling through ZODIAC. Nature Machine Intelligence. 2020;2: 629–641.
138. Dührkop K, Fleischauer M, Ludwig M, Aksenov AA, Melnik AV, Meusel M, et al. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. Nat Methods. 2019;16: 299–302.
139. Senan O, Aguilar-Mogas A, Navarro M, Capellades J, Noon L, Burks D, et al. CliqueMS: a computational tool for annotating in-source metabolite ions from LC-MS untargeted metabolomics data based on a coelution similarity network. Bioinformatics. 2019;35: 4089–4097.
140. Dictionary of Natural Products 29.2 Chemical Search. [cited 16 Feb 2021]. Available: http://dnp.chemnetbase.com/
141. Laatsch H. AntiBase: the natural compound identifier. Wiley-Vch Weinheim; 2017.
142. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol. 2016;34: 828–837.
143. van Santen JA, Jacob G, Singh AL, Aniebok V, Balunas MJ, Bunsko D, et al. The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural Products Discovery. ACS Cent Sci. 2019;5: 1824–1833.
144. Sorokina M, Steinbeck C. Review on natural products databases: where to find data in 2020. J Cheminform. 2020;12: 20.
145. Rebets Y, Schmelz S, Gromyko O, Tistechok S, Petzke L, Scrima A, et al. Design, development and application of whole-cell based antibiotic-specific biosensor. Metab Eng. 2018;47: 263–270.
146. Sun Y-Q, Busche T, Rückert C, Paulus C, Rebets Y, Novakova R, et al. Development of a Biosensor Concept to Detect the Production of Cluster-Specific Secondary Metabolites. ACS Synth Biol. 2017;6: 1026–1033.

147. Xu W, Klumbys E, Ang EL, Zhao H. Emerging molecular biology tools and strategies for engineering natural product biosynthesis. Metab Eng Commun. 2020;10: e00108.
148. Mitchler MM, Garcia JM, Montero NE, Williams GJ. Transcription factor-based biosensors: a molecular-guided approach for natural product engineering. Curr Opin Biotechnol. 2021;69: 172–181.
149. Li J, Wang H, Kwon Y-C, Jewett MC. Establishing a high yielding *Streptomyces*-based cell-free protein synthesis system. Biotechnol Bioeng. 2017;114: 1343–1353.
150. Jian Li, He Wang, Michael C.Jewett. Expanding the palette of *Streptomyces*-based cell-free protein synthesis systems with enhanced yields. Biochem Eng J. 2018;130: 29–33.
151. Myronovskyi M, Welle E, Fedorenko V, Luzhetskyy A. Beta-glucuronidase as a sensitive and versatile reporter in actinomycetes. Appl Environ Microbiol. 2011;77: 5370–5383.
152. Roy S, Radivojevic T, Forrer M, Marti JM, Jonnalagadda V, Backman T, et al. Multiomics Data Collection, Visualization, and Utilization for Guiding Metabolic Engineering. Front Bioeng Biotechnol. 2021;9. doi:10.3389/fbioe.2021.612893
153. Radivojević T, Costello Z, Workman K, Garcia Martin H. A machine learning Automated Recommendation Tool for synthetic biology. Nat Commun. 2020;11: 4879.
154. Zhang J, Petersen SD, Radivojevic T, Ramirez A, Pérez-Manríquez A, Abeliuk E, et al. Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism. Nat Commun. 2020;11: 4880.
155. van Dissel D, Claessen D, Roth M, van Wezel GP. A novel locus for mycelial aggregation forms a gateway to improved *Streptomyces* cell factories. Microb Cell Fact. 2015;14: 44.
156. Zhang Z, Du C, de Barsy F, Liem M, Liakopoulos A, van Wezel GP, et al. Antibiotic production in is organized by a division of labor through terminal genomic differentiation. Sci Adv. 2020;6: eaay5781.
157. Jones SE, Ho L, Rees CA, Hill JE, Nodwell JR, Elliot MA. *Streptomyces* exploration is triggered by fungal interactions and volatile signals. Elife. 2017;6. doi:10.7554/eLife.21738

158. Nieselt K, Battke F, Herbig A, Bruheim P, Wentzel A, Jakobsen ØM, et al. The dynamic architecture of the metabolic switch in *Streptomyces coelicolor*. BMC Genomics. 2010;11: 10.
159. Hopwood DA, Others. *Streptomyces* in nature and medicine: the antibiotic makers. Oxford University Press; 2007.

Chapter 3

CRISPR-Cas9-Based, CRISPRi, and CRISPR-BEST-Mediated Genetic Manipulation in Streptomycetes

Yaojun Tong[1], **Christopher M. Whitford[1]**, Kai Blin[1], Tue S. Jørgensen[1], Tilmann Weber[1*], Sang Yup Lee[1,2*]

[1] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark.
[2] Metabolic and Biomolecular Engineering National Research Laboratory Department of Chemical and Biomolecular Engineering (BK21 Plus Program), Center for Systems and Synthetic Biotechnology, Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Republic of Korea.

[*]To whom correspondence should be addressed. T.W. or S.Y.L.

Published in Nature Protocols: doi.org/10.1038/s41596-020-0339-z

ABSTRACT

Streptomycetes are prominent sources of bioactive natural products, yet the metabolic engineering of natural products from these organisms is heavily hindered by relatively inefficient genetic manipulation approaches. New advances in genome editing techniques, particularly CRISPR-based tools have revolutionized genetic manipulation for many organisms, including actinomycetes. We have developed a comprehensive CRISPR toolkit, which includes several variants of "classical" CRISPR-Cas9, along with CRISPRi and CRISPR-base editing systems (CRISPR-BEST) for streptomycetes. Here, we provide step-by-step protocols for designing and constructing the CRISPR plasmids, transferring these plasmids into the target streptomycetes, and identifying correctly edited clones. Our CRISPR toolkit can be utilized to generate random-sized deletion libraries, introduce small indels, generate in-frame deletions of specific target genes, reversibly suppress gene transcription, and substitute single base pairs in streptomycete genomes. Furthermore, the toolkit includes a Csy4-based multiplexing option to introduce multiple edits in a single experiment. The toolkit can easily be extended to other actinomycetes. With our protocol, it takes less than 10 days to inactivate a target gene, which is much faster than alternative protocols.

INTRODUCTION

Bacteria of the genus *Streptomyces* and relatives have served as the most promising source of antimicrobials for more than 70 years. They provide more than 70% of our current in-use antibiotics[1,2]. However, the current natural products-based antibiotics discovery and development pipeline is suffering from diminishing returns. No new classes of antibiotics have been developed over the past decades[3], which worsens the global health threat of multidrug-resistant infectious diseases.

Fortunately, modern genome mining[4] reveals that the genomes of streptomycetes and other related actinomycetes still possess a very large unexploited potential of encoding novel natural products[5]. In order to unlock this potential genetically, advanced technologies to efficiently manipulate and engineer the genomes of the producer strains are required[5,6]. However, it is more difficult to edit their genomes than in other unicellular model microorganisms like *Escherichia coli* or *Saccharomyces cerevisiae*, due to their mycelial growth, intrinsic genetic instability, and very GC-rich (>70%) genome. Moreover, only a few genetic manipulation methods are available for actinomycetes.

The recent availability of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-associated (Cas)-based genome editing technologies have been revolutionizing the whole biotechnology field. CRISPR-Cas-mediated genetic manipulation approaches have been successfully applied to organisms across all domains of life[7,8] relatively quickly, efficiently, and with no scarring.

In order to develop advanced genetic manipulation methods for exploiting new bioactive natural products from streptomycetes, we first developed a toolkit using the “classic” DNA double-strand break (DSB) based CRISPR-Cas9 and its relatives[9]. Along with similar tools established by others[10-12], CRISPR-Cas9 dramatically increased the ease of genetically manipulating streptomycetes[9,13,14]. However, concerns of DSB-mediated genome instability and Cas9-related toxicity remained[13,14]. Accordingly, we then established a nearly single base pair resolution CRISPR base editor system, CRISPR-BEST[15], that provides comparable editing efficiency to the “classic” CRISPR-Cas9

systems without involving DSBs, with editing resolutions reaching up to single-nucleotide level[9,15,16]. To further simplify the sgRNA cloning step, we established an efficient single strand DNA (ssDNA) oligo bridging method (Fig. 1) and optimized the conjugation protocol. To meet the increasing demands of introducing multiple mutations in one single experiment, we included a multiplexing option[15], which uses Csy4 (also known as type I-F CRISPR-associated endoribonuclease Cas6f[17])-based sgRNA self-processing machinery. Together with the protospacer finder CRISPy-web[18], the CRISPR-Cas9 based toolkit[9] and the CRISPR-BEST toolkit[15] provide a versatile genetic manipulation system for streptomycetes (Fig. 2).

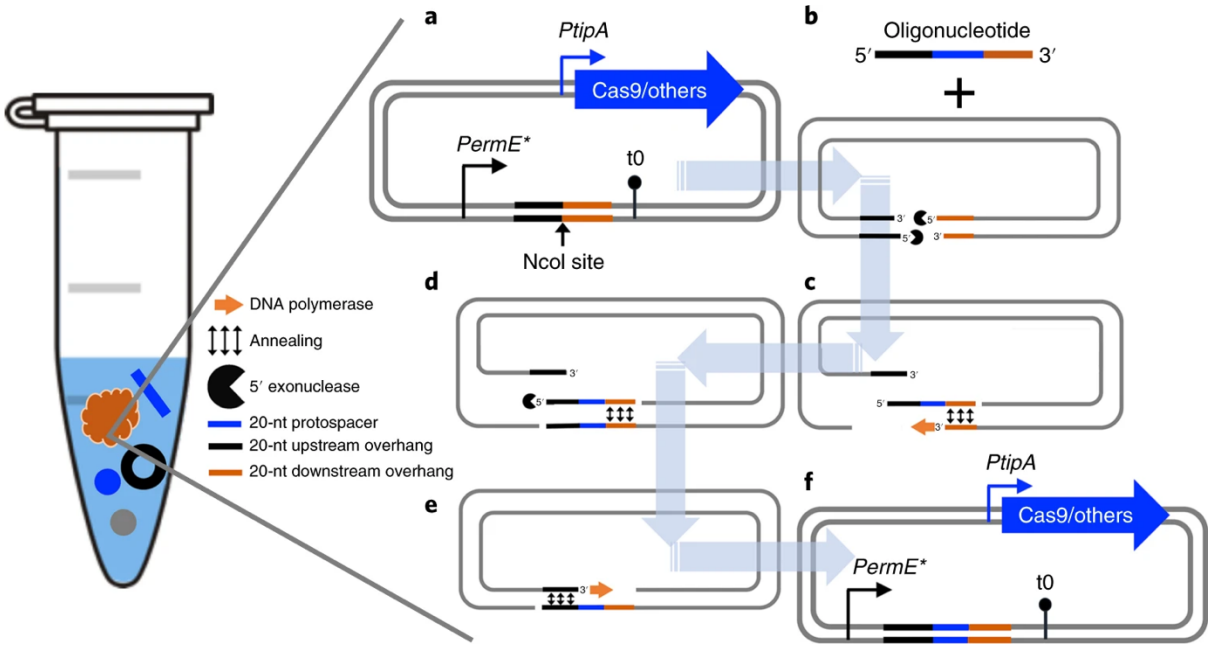


Figure 1: A mechanistic overview of the all-in-one pot ssDNA bridging method for sgRNA cloning. a. Linearize the plasmid with NcoI. **b–f**, Illustration of ssDNA bridging. **b**, The 5′ exonuclease included in the NEBuilder mix digests the 5′ ends of the double-stranded DNA (dsDNA), making the 3′ complementary ends available. **c**, The nondigested ssDNA oligonucleotide anneals to the exposed 3′ end of the plasmid backbone; the DNA polymerase extends the 3′ end, using the oligonucleotide as the template. **d**, The 5′ exonuclease then removes the original protospacer containing the ssDNA oligonucleotide. **e**, The other nicked strand of the plasmid anneals to the newly synthesized 3′ end of the complementary strand and then is extended by the DNA polymerase. **f**, The DNA ligase included in the NEBuilder mix then seals the nicks to form the desired protospacer-cloned CRISPR plasmid. P*ermE**, a widely used constitutive promoter; P*tipA*, thiostrepton-inducible promoter; t0, terminator. Blue arrows represent one of the key components of CRISPR plasmids; 'others' represents dCas9 or Cas9n–deaminase.

Comparison with other currently available genetic manipulation methods

Traditional mutagenesis methods were established in 1978, when for the first time external DNA was successfully transferred into streptomycetes[19]. For example, plasmids with temperature sensitive and unstable replicons were used for classical double-crossover based mutagenesis[20], which relies on natural occurring homologous recombination events. Later, PCR-targeting[21] became popular, which relies on identifying the editing target in a cosmid/fosmid library, followed by a λ-Red-mediated gene disruption or deletion of the target encoded on the cosmid/fosmid in *E. coli*, introducing the modified cosmid/fosmid into the native producer strain and, finally, screening for double-crossover events[20,21]. In order to recycle the resistance marker, it can be removed by Cre/lox, though that leaves a 81-bp scar[21]. However, this method is relatively intricate, since it requires a fully characterized genomic library of all target strains.

Besides the PCR-targeting method, a *Saccharomyces cerevisiae* meganuclease *I-SceI* based gene deletion method for streptomycetes was established in 2014[22]. The use of this method requires two conjugations, the first one is to integrate the 18-bp *I-SceI* recognition site together with the homologous recombination templates (editing templates) into the genome of the target strain by a single crossover event. In a second conjugation, the meganuclease *I-SceI* is introduced, which upon expression will cut the 18-bp *I-SceI* recognition site leading to a DSB in the chromosome. Subsequently, homologous recombination based DSB repair leads to either the desired deletion without a scar or the wild type genotype. Due to the selection pressure introduced by the DSB, the editing (double crossover) efficiency of this method is higher than PCR-targeting method. However, the protocol for this meganuclease *I-SceI* based method is still relatively difficult and time-consuming.

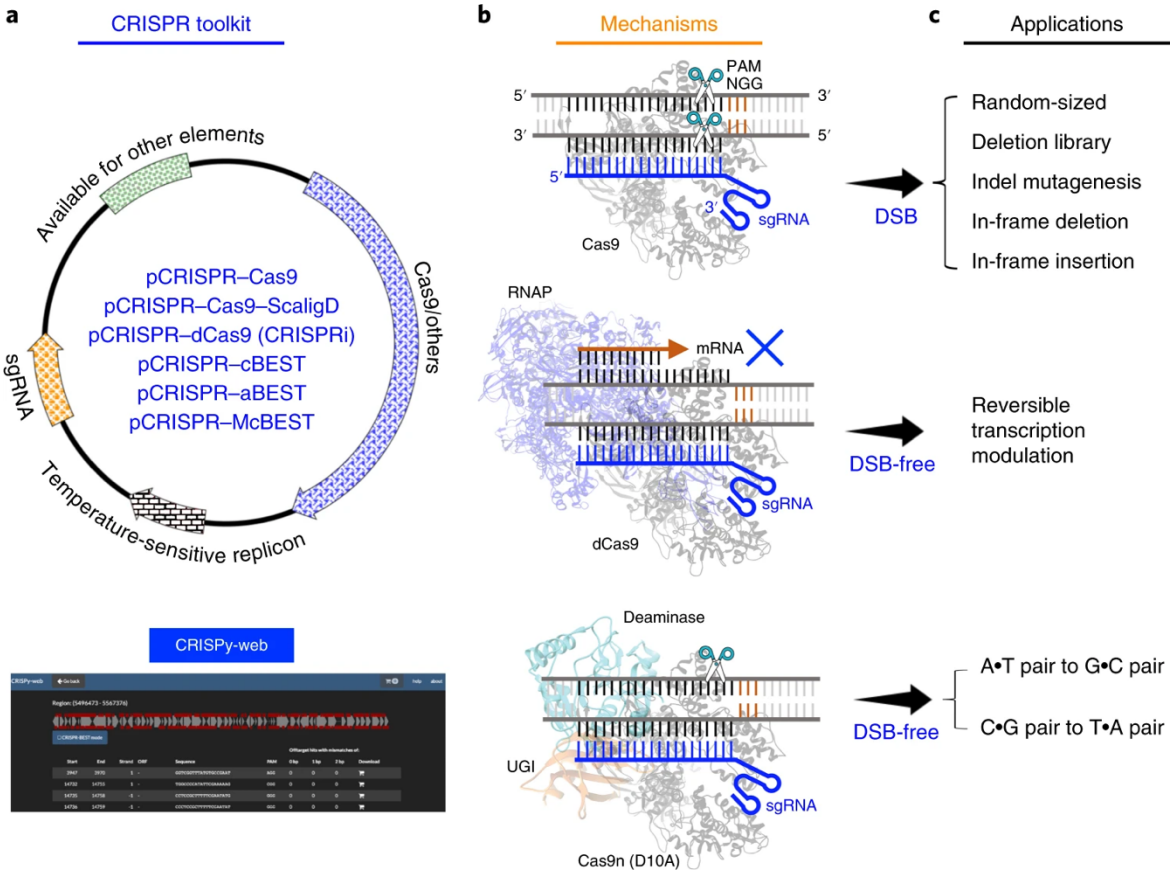


Figure 2: An overview of the CRISPR-based genetic manipulation system included in this protocol. a, The components of the CRISPR-based streptomycetes/actinomycetes genetic manipulation toolkit. The toolkit includes six plasmids (listed inside the plasmid map) and an sgRNA design tool, CRISPy-web. 'Available for other elements' means that there is space in this plasmid system that can be used for cloning additional elements, for example, the editing templates or Csy4. 'Cas9/others': represents the Cas9 component of the CRISPR plasmids; 'others' can include dCas9, Cas9n–adenosine deaminase, or UGI–Cas9n–cytidine deaminase; this part varies with the different plasmids in the toolkit. **b**, A simple mechanism of action for each type of Cas9 used in the CRISPR system is provided. pCRISPR–Cas9 and pCRISPR–Cas9–ScaligD: sgRNA binds to Cas9, forming a Cas9–sgRNA complex, and then guides the complex to find and bind to the target DNA region; then Cas9 cleaves the target DNA to introduce a DSB, and the DSB can be repaired by different ways, leading to the genome-editing events. pCRISPR–dCas9 (CRISPRi): The dCas9–sgRNA complex preserves DNA binding ability but has no DNA cleavage activity. It therefore blocks the transcription process mediated by the RNA polymerase when guided to the target sequence by the sgRNA. pCRISPR–aBEST, pCRISPR–cBEST, and pCRISPR–McBEST: A deaminase is fused to Cas9n, and then the whole complex is guided by sgRNA to the target DNA region, where adenosine deaminase can convert A•T pairs to G•C pairs, and cytidine deaminase can convert C•G pairs to T•A pairs, which are shown in c. **c**, The applications of different CRISPR tools within the CRISPR-based genetic manipulation system are summarized. PAM, protospacer- adjacent motif (in this protocol, all CRISPR plasmids use the same PAM, 5′-NGG-3′); RNAP, RNA polymerase; UGI, uracil glycosylase inhibitor.

Development of the protocol

The initial protocol for using our CRISPR-Cas9-based toolkit was described in Tong et al[16]. With years of practical usage and continuously extending the toolkit, the protocol has been updated accordingly for better output, which will be described further here. There are, however, some remaining concerns and challenges with this technology[13,14] including, for example, the problem that overexpression of active Cas9 is toxic in many streptomycetes, and may lead to unwanted off-target effects[13,14]. Moreover, most streptomycetes have a linear chromosome that is relatively unstable and susceptible of large-scale chromosomal deletions and rearrangements[23]. DSBs in the terminal inverted repeats of the chromosomes are speculated to be the main triggers of genome instability[15,24]. To address to the above-mentioned concerns, we then also developed a highly efficient, DSB-free, precise, and multiplex genome editing toolkit for streptomycetes, CRISPR-BEST[15], based on the deaminase-mediated base editing technology[25-27].

Given the complexity of intracellular metabolic networks, it is often necessary to engineer multiple genes simultaneously for both basic and applied studies.

With classical actinomycete genetic engineering methods, only one single genetic modification can be introduced per round of genetic manipulation. However, due to its simplicity and modularity, CRISPR inherently provides the capability for multiplexing. The current multiplexing applications in streptomycetes are still limited by the efficacy of individual sgRNA generation. This requires an independent promoter and terminator for each individual sgRNA transcription unit, increasing the risks of plasmid instability and uneven distribution of each sgRNA, along with the potential of triggering unwanted mutagenesis[28-30]. To address the aforementioned concerns, we therefore designed a Golden Gate Assembly-compatible Csy4[17]-based sgRNA self-processing system, which only requires a single promoter and terminator pair for multiple sgRNAs separated by the Csy4 recognition sites. A simple example of a three sgRNA multiplexed plasmid design and construct is shown in Fig. 3.

Overview of the procedure

A summary of all approaches is depicted in Fig. 4. The overall procedure can be divided into four (optionally, five) main stages: once the task (e. g. inactivation of a gene-of-interest) is chosen: **Stage I)** *in silico* design of sgRNA, primers, and (optional) editing templates (**Steps 1-7, 8(C)**); **Stage II)** Construct and validate the designed CRISPR plasmid (**Step 8-20**); **Stage III)** Conjugate the correctly constructed CRISPR plasmid into the target streptomycetes **(Steps 21-33)**; **Stage IV)** Validate the correctly edited target streptomycetes **(Step 34)**; **Stage V) (optional)** Removal of the CRISPR plasmids (in the following referred to as curing) from the correctly edited streptomycete strains to make a clean, CRISPR-plasmid free mutant that is ready for a second round of genetic manipulation (**Steps 35-39**).

Applications of the method

Primarily, we demonstrated the use of the CRISPR toolkit presented in this protocol in *S. coelicolor and S. collinus*[9,15], but over time it was also applied in many other streptomycetes, like *S. albus*[31], *S. sp.* SD85[32], and a series of environmental *Streptomyces* isolates[33]. Moreover, the application of this toolkit was extended to other actinomycetes, such as *Micromonospora chersina*[34], *Corynebacterium glutamicum*[35], *Saccharopolyspora erythraea*[36], and *Verrucosispora sp.* MS100137[37]. We furthermore believe that this toolkit can reach more species with minimal efforts, for example, by exchanging the replicon and/or related promoters.

Applications with CRISPR-Cas9 based plasmids

In previous work, we demonstrated that by using the plasmid pCRISPR-Cas9 from the CRISPR-Cas9 toolkit in a wildtype *S. coelicolor* A3(2) strain and many other streptomycetes[9], a library of differently sized deletions around the target site can be generated lacking the LigD component, due to the “incomplete” NHEJ pathway.

When adding a ~2 kb editing template composed of 1 kb each up- and downstream of the gene-of-interest to pCRISPR-Cas9, we can obtain an in-frame

deletion event with nearly 100% editing efficiency[9]. By including other DNA sequences, such as gene(s)-of-interest or promoters into the editing template, the system can also provide highly specific chromosomal insertions. This concept was reported by Zhang et al[38].

As all CRISPR plasmids described in this protocol follow the same design principle, the Csy4-mediated sgRNA multiplexing strategy[15] initially established for CRISPR-BEST can also be integrated into the original CRISPR-Cas9 toolkit. However, one needs to keep in mind that if multiple DSBs are simultaneously introduced into the genome of streptomycetes, the genomes may become unstable and thus the genome structure analysis and a genome-wide off-target evaluation (Step 34(C)) should be carried out for positive clones.

Applications with CRISPR-BEST based plasmids

Here, we provide a protocol for one application of CRISPR-BEST, the inactivation of coding genes by introducing a STOP codon in streptomycetes. The application of CRISPR-BEST toolkit can of course be easily extended to do *in vivo* protein engineering by customized amino acid substitutions.

Strengths and limitations of the presented methods

The protocols facilitate highly efficient, precise, and rapid generation of desired genetic modifications in streptomycetes, showing the following advantages over the conventional double crossover-based methods, including PCR-targeting-based strategies[21]: **i)** the protocol is easy to operate, in most cases only one highly efficient cloning step is required to construct a functional CRISPR plasmid; **ii)** the protocol does not require genome-mapped cosmid clones of the strain to be engineered (which is normally required by the PCR-targeting approach); **iii)** the protocol is relatively fast (approximately 10 days for inactivation of a gene); **iv)** the protocol is versatile and covers varied genetic manipulations including in-frame-knockout, in-frame-knockin, introduction of indels, single amino acid exchange and target gene knockdown; **v)** the protocol is capable of multiplexed genome editing with a single plasmid targeting multiple genes-of-interest simultaneously. This has been only demonstrated so far in base editing applications using pCRISPR-cBEST[15], but it can be easily applied to CRISPRi applications.

Similar to any other protocol using Cas9-based CRISPR genome editing technologies, there are certain limitations: the primary concern is off-target effects. There are two typical off-target effects, one causes visible phenotype changes, which can therefore be easily ruled out. The other does not link to any phenotype changes, which requires additional effort (for example, whole genome sequencing) to identify. As the Cas9 protein is an endonuclease, it is important to bear in mind that the expression of active nuclease proteins such as Cas9 or Cas12a (Cpf1) has been reported to be toxic[39,40] to some organisms. The toxicity would increase the off-target frequency in streptomycetes[41]. This problem can be addressed in different ways, for example, by using CRISPR-based approaches that don't require full endonuclease activity (i. e., CRISPR-base editing[25-27], and CRISPR-prime editing[42]).

In addition to these issues affecting CRISPR applications in general, several specific features of streptomycetes need to be considered: for example, the > 70% GC content makes routine molecular biological operations more difficult than for other organisms with lower GC content. Their linear chromosome is relatively unstable and can tolerate large deletions and genome rearrangements without displaying severe phenotypic effects under laboratory conditions. Therefore, any genetic engineering method involving DNA double-strand breaks (e.g. introduced by *I-SceI* or CRISPR-Cas9), can cause severe off-target effects, such as genome deletions and rearrangements[15,23,24], which need to be thoroughly investigated after the genome editing procedure by whole genome sequencing.
Most streptomycetes are compatible with the CRISPR plasmids presented here, meaning the plasmids can replicate, the resistance genes successfully confer antibiotic resistance in the target strain, and the promoters used to control Cas9/Cas9-variants and sgRNA are compatible with the machinery of the target strain. However, there may be individual strains or closely related genera, which may not accept the pSG5 replicon and/or the *tipA* promoter used. In these specific cases a simple solution is to exchange the plasmid backbone and/or promoter.

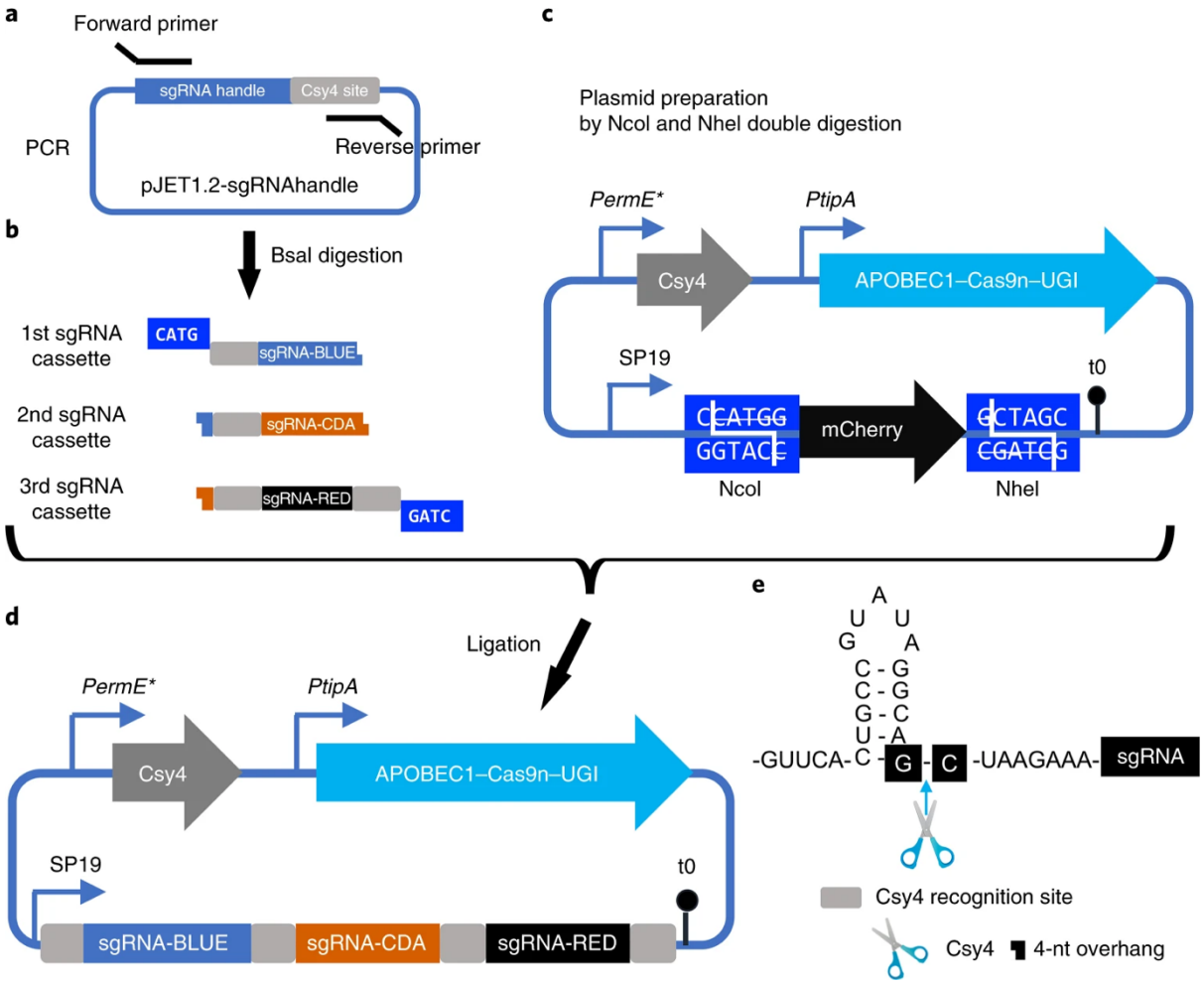


Figure 3: A three-sgRNA multiplexed example. **a**, PCR-amplify all sgRNA fragments; for detailed sgRNA design[15], refer to Step 8B(i) and Box 1. **b–d**, An overview of the two-step Golden Gate assembly. **b**, Preparation of the sgRNA fragments by BsaI digestion. **c**, Preparation of the plasmid backbone by NcoI and NheI digestion. **d**, Ligation of the prepared fragments from **b** and **c** to obtain the final, ready-to-use CRISPR plasmid. **e**, Illustration of the mechanism of action of Csy4-based RNA processing. SP19, synthetic promoter19 (ref. [51]); Csy4 site, Csy4 recognition and cleavage site; sgRNA-BLUE, the sgRNA targets the SCO5087 gene of the actinorhodin biosynthetic gene cluster; sgRNA-CDA, the sgRNA targets the SCO3230 gene of the calcium-dependent antibiotic (CDA) biosynthetic gene cluster, sgRNA-RED: the sgRNA targets the SCO5892 gene of the undecylprodigiosin (RED) biosynthetic gene cluster.

Experimental design

Plasmid design

Given the lack of efficient replicons and selection markers for streptomycetes[20], we assembled all essential components into a single pSG5 replicon[43]-based plasmid backbone, making an "all-in-one" system. Depending on the specific protocol (Fig. 4), these components include the Cas9/dCas9/Cas9n protein, deaminases, sgRNA, DNA repair components, editing templates, and sgRNA processing machinery. Only one cloning step, to achieve sgRNA assembly, is required to use most of the CRISPR plasmids presented in this protocol (an additional step for editing template assembly is however also required with in-frame deletion/insertion applications). Initially, a NcoI-SnaBI (Eco105I) based digestion-ligation protocol and a USER-cloning protocol were established for sgRNA assembly[9,16]. To further simplify the protocol and increase the cloning efficiency, an easy scalable PCR- and digestion-free single strand DNA oligo bridging system was set up and used for all described protocols here (Fig. 1). As the used pSG5[43] replicon is thermo-instable, the plasmids can be cured relatively easily by increasing the incubation temperature and re-used repetitively.

Multiplexed editing

To further extend the ability of multiplexed editing with the CRISPR toolkits, a Csy4[17] based sgRNA self-processing machinery was cloned into the "all-in-one" CRISPR plasmids[15]. Although the prototype multiplexing system was only successfully demonstrated in *S. coelicolor* and *S. griseofuscus* with pCRISPR-cBEST[15], it is reasonable to assume that it would be possible to extend this system to applications using the CRISPRi tool, but it would be more difficult to apply to the active Cas9 based tools due to the challenges associated with introduction of one, yet alone multiple DSBs. A Golden Gate assembly method was established for assembling multiple sgRNAs, the rules of primer design are: the overhang of the forward primers needs to consist of (from 5' to 3'): random 5-nt adapter (e.g. GATCG), the BsaI recognition site (GGTCTCN), CATG which after digestion matches with the NcoI restriction of the plasmid backbone (only for the first sgRNA), 28-nt Csy4 recognition site (only for the first sgRNA), 20-nt spacer (from Step 7); the overhang of the reverse primers

needs to consist of (from 5' to 3'): random 5-nt adapter (e.g. GATCG), the BsaI recognition site (GGTCTCN), first 4-nt of the next downstream sgRNA or CTAG for the last sgRNA in the array which after digestion matches with the NheI overhang of the plasmid backbone (an example is presented in Box 1).

sgRNA design

In order to provide a flexible sgRNA design platform, we developed the protospacer identification program CRISPy-web (https://crispy.secondarymetabolites.org)[18] that offers a user-friendly web interface and the possibility to upload user-specific genome sequences in which to find targets. In order to support the applications of the CRISPR-BEST toolkit, the software recently was updated[15] and now provides additional options to specifically design sgRNAs for base editing applications.

Modifications to increase efficiency

In order to reduce the experimental burden, we further established an easy method for *Streptomyces* colony PCR to avoid massive genomic DNA preparation, improved the classical spore conjugation method to increase the transformation efficiency, and modified the traditional protocol for making highly electro-competent *E. coli* ET12567/pUZ8002 cells[44], which to our best knowledge are not commercially available to date.

Controls

We recommend including two sets of controls: one is the empty plasmid control (no 20-bp spacer was cloned upstream of the sgRNA scaffold), involved in the interspecies conjugation (Steps 21-33); and the other one is the wild type strain control, involved in the editing evaluation (Step 34, Sanger sequencing and Illumina sequencing). The empty plasmid control indicates conjugation efficiency, and the editing efficiency of the selected spacers. The wild type strain control can be used to determine both on-target editing and off-target effects.

On-target evaluation of CRISPRi applications

Unlike the other CRISPR applications described in this protocol, the use of CRISPRi[45] / CRISPR-dCas9 will not cause chromosomal genetic changes, it will only block the transcription process, via preventing either of transcription initiation or transcription elongation steps by a road block effect of the nuclease activity defective Cas9 (dCas9):sgRNA:target-DNA complex (Fig. 2). Therefore, Sanger sequencing or whole genome sequencing cannot evaluate its effects. It requires validation from the transcriptional level and/or endpoint product level of a pathway. As the protocol presented here aims to affect the endpoint production quantity, i. e., the final molecule that is synthesized by a metabolic pathway, we describe the protocol for actinorhodin detection in *S. coelicolor*[9].

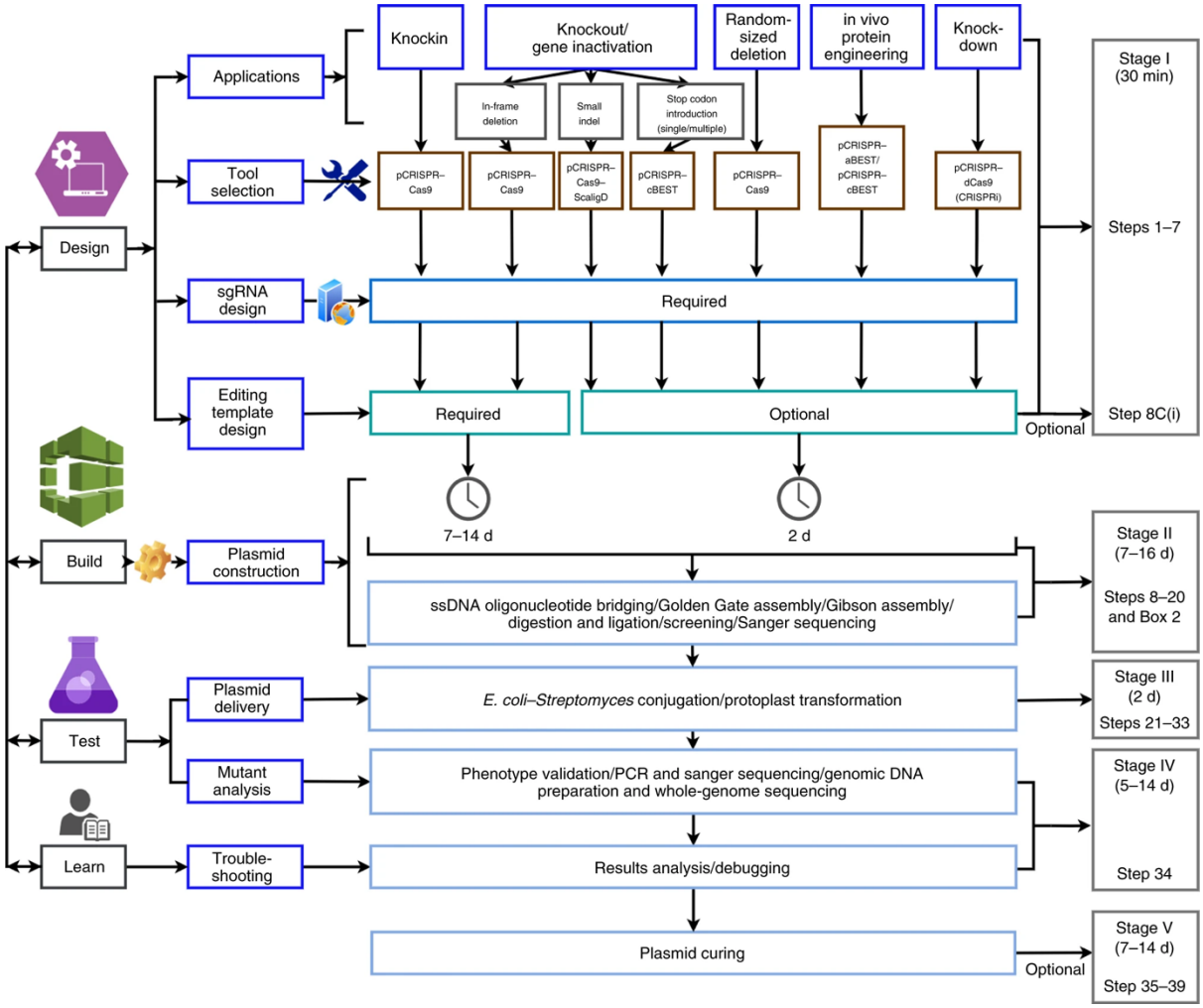


Figure 4: An overview of the entire protocol. Shown is a schematic representation of processes for using different CRISPR plasmids within this protocol to complete desired genetic manipulation in the target streptomycete strain.

MATERIALS

Biological materials

- One Shot™ Mach1™ T1 Phage-Resistant Chemically Competent *E. coli*, F^- ϕ80(lacZ)ΔM15 ΔlacX74 hsdR($r_K^-m_K^+$) ΔrecA1398 endA1 tonA (Thermo Fisher Scientific, cat. no. C862003).
- **!CAUTION** Competent cells should be stored at –80 °C before use.
- *E. coli* DH5α, F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 ϕ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17($r_K^-m_K^+$), λ^-. (Thermo Fisher Scientific, cat. no. 18265017). It is used for maintaining the pJET1.2-sgRNAhandle.
- *Escherichia coli* ET12567/pUZ8002 (ref. [44]), F^- dam-13::Tn9 dcm-6 hsdM hsdR zjj-202::Tn10 recF143 galK2 galT22 ara-14 lacY1 xyl-5 leuB6 thi-1 tonA31 rpsL136 hisG4 tsx-78 mtl-1 glnV44, Cm^r/Km^r. (Life Science Market, cat. no. S0052).
- **!CAUTION** Maintaining ET12567/pUZ8002 requires both chloramphenicol and kanamycin, as ET12567 has a chloramphenicol resistant marker and the carrying pUZ8002 has a kanamycin resistant marker.
- *Streptomyces coelicolor* A3(2) (ref. [46]) (DSMZ, DSM no. 40783).
- **!CAUTION** Over the years, the *S. coelicolor* WT maintained in our lab has accumulated several SNPs and a deletion of an integrated plasmid[9,15]. Therefore, we highly recommend users to sequence their own WT/parental strain when evaluating the genome editing results.

Reagents

- ddH_2O (Milli-Q filtered water)
- FastDigest BsaI/Eco31I (Thermo Fisher Scientific, cat. no. FD0294)
- FastDigest NcoI (Thermo Fisher Scientific, cat. no. FD0573)
- FastDigest NheI (Thermo Fisher Scientific, cat. no. FD0973)
- FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific, cat. no. FD0652)
- T4 Ligase (5U) (Thermo Fisher Scientific, cat. no. EL0014)
- Lysozyme from chicken egg white (Sigma-Aldrich, cat. no. L6876-10G)

- Protease K (>600 U/mL (~20 mg/mL; Thermo Fisher Scientific, cat. no. EO0491)
- Phusion High-Fidelity PCR Master Mix with HF Buffer (Thermo Fisher Scientific, cat. no. F531L)
- Phusion High-Fidelity PCR Master Mix with GC Buffer (Thermo Fisher Scientific, cat. no. F532L)
- OneTaq® 2× Master Mix with Standard Buffer (New England Biolabs, cat. no. M0482L)
- NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs, cat. no. E2621L)
- Q5® High-Fidelity 2× Master Mix (New England Biolabs, cat. no. M0492L)
- KAPA HyperPlus kit (Roceh, cat. no. 07962428001)
- NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, cat. no. 740609.250)
- NucleoSpin® Plasmid EasyPure Kit (Macherey-Nagel, cat. no. 740727.250)
- QIAamp DNA Blood Midi Kit (Qiagen, cat. no. 51183)
- Qubit™ dsDNA BR Assay Kit (Thermo Fisher Scientific, cat. no. Q32853)
- Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, cat. no. Q32854)
- Mix2Seq Kit OVERNIGHT (Eurofins Genomics)
- Dimethyl Sulfoxide (DMSO) (Sigma-Aldrich, cat. no. 472301-500ML)
- Ethanol Absolute (VWR, cat. no. 20821.310)
- **!CAUTION** Absolute ethanol is highly flammable. Store it in a fireproof cabinet and use it with caution and away from open flame.
- Ethanol ≥ 70% (v/v) (VWR, cat. no. 83801.290)
- Glycerol (≥ 99.5% (wt/vol); Sigma-Aldrich, cat. no. G9012-1GA)
- Isopropanol (Sigma, cat. no. I9516)
- **!CAUTION** Absolute isopropanol is highly flammable. Store it in a fireproof cabinet and use it with caution and away from open flame.
- Ampicillin sodium salt (Sigma-Aldrich, cat. no. A0166-25G)
- Apramycin sulfate salt (Sigma-Aldrich, cat. no. A2024-5G)
- Chloramphenicol (Sigma-Aldrich, cat. no. C1919-25G)
- Kanamycin sulfate (Sigma-Aldrich, cat. no. K1377-25G)
- Nalidixic Acid (Sigma-Aldrich, cat. no. N8878-5G)
- Thiostrepton (Sigma-Aldrich, cat. no. T8902-1G)
- LB Broth (Lennox) (Sigma-Aldrich, cat. no. L3022-1KG)

- LB Broth with agar (Miller) (Sigma-Aldrich, cat. no. L3147-1KG)
- ISP2 agar (ISP Medium 2, BD Difco, cat. no. DF0770-17-9)
- 6x DNA Gel Loading Dye (Thermo Fisher Scientific, cat. no. R0611)
- DNA stain (RedSafe™; iNtRON Biotechnology, cat. no. 21141)
- **!CAUTION** Though this product is claimed to be without mutagenesis concerns, we highly recommend using proper protection when handling it.
- GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific, cat. no. SM0311)
- GeneRuler 100 bp DNA Ladder (Thermo Fisher Scientific, cat. no. SM0242)
- 50x and 1x TAE buffer (in-house made, see Reagents Setup)
- Trizma base (Sigma-Aldrich, cat. no. 93362-1KG)
- 10x NEBuffer 2 (New England BioLabs, cat. no. B7002S)
- Sodium hydroxide (NaOH; pellets; Honeywell, cat. no. 30620-1KG)
- Sodium chloride (NaCl; Honeywell, cat. no. 13423-1KG-R)
- Acetic acid (Honeywell, cat. no. 33209-4x2.5L-M)
- **!CAUTION** Acetic acid is flammable, corrosive, and volatile. Use it inside a chemical fume hood with caution and away from open flame.
- Hydrochloric acid (HCl; VWR, cat. no. 20252.290)
- **!CAUTION** HCl is highly corrosive and volatile. Use it inside a chemical fume hood with caution.
- Magnesium chloride ($MgCl_2$; Sigma-Aldrich, cat. no. M4880-100G)
- OXOID™ Yeast Extract Powder (Thermo Fisher Scientific, cat. no. LP0021B)
- Malt Extract (Sigma-Aldrich, cat. no. 70167-500G)
- Dextrose (D-Glucose; Sigma-Aldrich, cat. no. D9434-1KG)
- Agar (Sigma-Aldrich, cat. no. 05040-1KG)
- D-Mannitol (≥ 98%; Sigma-Aldrich, cat. no. M4125-1KG)
- Hensel® Organic Soy Flour (Fat reduced; Violey, product no. 02004307)
- Standard Agarose – Type LE (BioNordika, cat. no. BN50004)
- Trizma® hydrochloride (Tris-Cl; Sigma-Aldrich, cat. no. T3253-500G)
- Trizma® base (Tris-base; Sigma-Aldrich, cat. no. 93362-500G)
- EDTA (Ethylenediaminetetraacetic acid disodium salt dihydrate; Sigma-Aldrich, cat. no. E5134-500G)

- Tween® 20 (for molecular biology, viscous liquid; Sigma-Aldrich, cat. no. P9416-100ML)
- Triton™ X-100 (Thermo Fisher Scientific, cat. no. AC215682500)
- Guanidine Hydrochloride (Guanidine HCl; Sigma-Aldrich, cat. no. G3272-500G)
- Buffer B1 (for streptomycetes genomic DNA isolation; in-house made; ingredients: $Na_2EDTA{\cdot}2H_2O$, Trizma base, Tween-20, and Triton X-100; see Reagents Setup for details)
- Buffer B2 (for streptomycetes genomic DNA isolation; in-house made; ingredients: guanidine HCl and Tween-20; see Reagents Setup for details)
- CloneJET PCR Cloning Kit (Thermo Fisher Scientific, cat. no. K1231)
- pCRISPR-Cas9, RRID: Addgene_125686
- pCRISPR-dCas9, RRID: Addgene_125687
- pCRISPR-Cas9-ScaligD, RRID: Addgene_125688
- pCRISPR-cBEST, RRID: Addgene_125689
- pCRISPR-aBEST, RRID: Addgene_131464

Equipment

- Sartorius mLINE® Pipettes (0.1-3 µL, 0.5-10 µL, 5-100 µL, and 100-1000 µL, Sartorius, cat. nos. 725010, 725120, 725130, and 725070, respectively)
- Sartorius Optifit Refill Tips (10 µL, 200 µL, and 1000 µL, Sartorius, cat. nos. 790012, 790202 and 791002, respectively)
- Sartorius Minisart® syringe filter (0.2 µm; single use; non-pyrogenic; Sartorius, cat. no. 50192103)
- Sartorius Quintix® Lab Scale (Sartorius, cat. no. QUINTIX213-1S)
- Pipetboy acu2 (Integra Biosciences, item no. 155017)
- Serological pipette, sterile, CELLSTAR® (1 mL, 10mL, and 25mL; Greiner bio-one, cat. nos. 604181, 608180, and 760180, respectively)
- Bacteriological Petri dish (94 mm × 16 mm; not tissue culture treated, without vents, aseptic; LabSolute, cat. no.7696401)
- Shake flasks (250 mL; wide-mouth flasks; VWR, cat. no. 214-1132)
- Bio-Rad C1000 Touch Thermal Cycler (Bio-Rad, cat. no. 1851148)
- Bio-Rad S1000 Thermal Cycler (Bio-Rad, cat. no. 1852196)

- Bio-Rad MicroPulser™ Electroporator (Bio-Rad, cat. no. 1652100)
- Molecular Imager® Bio-Rad Gel Doc™ XR+ Gel Documentation System with Image Lab™ Software (Bio-Rad, cat. no. 1708195)
- Gel electrophoresis tank and power source (Bio-Rad Mini-Sub and Wide Mini-Sub Cell GT Cells; Bio-Rad, cat nos. 1704466 and 1704468EDU; Bio-Rad PowerPac Basic Power Supply; Bio-Rad, cat. no. 1645050)
- Invitrogen Safe ImagerTM 2.0 (Invitrogen cat. no. G6600EU)
- Floor centrifuge Thermo Scientific® Heraeus Multifuge X3 FR centrifuge (Thermo Fisher Scientific, cat. no. 75004500)
- Rotors (Thermo Fisher Scientific; Fiberlite F14-6x250 LE, cat. no. 75003661 and Fiberlite F13-14x50cy LE, cat. no. 75003662)
- Benchtop centrifuge (Thermo Scientific® Heraeus Fresco17 centrifuge; Thermo Fisher Scientific, cat. no. 75002420)
- Desktop centrifuge (Micro Star 17; VWR, cat. no. 521-1646P)
- Desktop Microcentrifuge (VWR Microcentrifuge Mini Star Silverline; VWR, cat. no. 521-2845P)
- Labnet 311DS incubator (Labnet International, cat. no. I5311-DS)
- New Brunswick™ Innova® 42/42R shaker (Eppendorf, cat. no. M1335-0002)
- Polypropylene centrifuge bottles (250 mL; Corning, cat. no. 431841)
- Heating Block (Eppendorf Thermomixer Comfort MTP & 1.5 mL; Eppendorf, cat. no. 5355 000.011)
- UV spectrophotometer (NanoDrop; Thermo Fisher Scientific, model no. 2000)
- Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, cat. no. Q32866)
- Qubit™ Assay Tubes (Thermo Fisher Scientific, cat. no. Q32856)
- Gene Pulser® Cuvette/E. coli® Cuvette (0.1 cm electrode; Bio-Rad, cat. no. 165-2089)
- Magnetic stirrer (RCT basic; IKA, item. no. 0003810000)
- Magnetic stirring bar (IKAFLON® 155; IKA, item. no. 0001129000)
- Benchtop autoclave (Certoclav Multicontrol autoclave; CertoClav Sterilizer GmbH, cat. no. 8510298)
- Eppendorf Safe-Lock microcentrifuge tubes (1.5 mL; Eppendorf, cat. no. 0030120086)
- 50 mL Falcon tube (polypropylene; 50 mL; Sarstedt, cat. no. 62547254)

- 15 mL Falcon tube (CELLSTAR® Centrifuge tubes; polypropylene; 15 mL; Greiner Bio-One, cat. no. 188271)
- 8-strip PCR tubes (Sarstedt, cat. no. 72.985. 002)
- Flat Cap Strips (Cell Projects, cat. no. FC-08-CC/CP)
- Inoculating Loop (loop size 10 µL; Sigma-Aldrich, cat. no. I8388-500EA)
- Inoculation spreader (Sarstedt, cat. no. 86.1569.005)
- pH meter (827 pH lab; Metrohm, cat. no. 2.827.0214)
- Laminar flow hood (Biological Safety Cabinets Class 2; LaboGene, Mars-900, cat. no. 9.001.023.000; Mars-1200, cat. no. 9.001.020.000)
- Heratherm incubator (Thermo Fisher Scientific, item no. WZ-38800-16)
- Vortex-Genie® 2 (Scientific Industries, cat. no. SI-0236)
- CryoPure Tube 1.8 mL white (Sarstedt, non-pyrogenic, non-cytotoxic, non-mutagenic, cat. no. 72.379)
- Ultrospec 10 Classic (OD meter; Biochrom, cat. no. 84-741)
- OD Cuvettes (Semi-micro cuvette, PS, minimum filling volume 1,5 mL; Brand, cat. no. 759015)
- Cotton Swabs (Matas, item no. 732634)
- Cotton Pads (Matas, item no. 731144)
- Wooden toothpicks (Wooden Cocktail Sticks; Plastico, item code 304)
- Sera-Mag Select (GE healthcare, cat. no. 29343045)
- 5300 Fragment Analyzer System (Agilent Technologies, item no. M5311AA)
- Ice machine (Scotsman, item no. MISTERNAC176)
- Milli Q water system (Purelab flex2, Elga, cat. no. PF2XXXXM1-KIT)
- BlueCap bottle with blue GL45 lid (Buch-Holm, item nos. 100 mL, 9072331; 250 mL, 9072332; 500 mL, 9072334; 1000 mL, 9072335)
- Parafilm (Sigma-Aldrich, item no. P7793-1EA)
- Aluminum foil
- Paper towels
- Autoclave tape
- Tweezers
- Scissors

Software

- NEB Tm calculator (NEB), RRID:SCR_017969, https://tmcalculator.neb.com/#!/main
- CRISPy-web[18], RRID:SCR_017970, https://crispy.secondarymetabolites.org
- CLC Main Workbench, RRID:SCR_000354, (QIAGEN Bioinformatics), https://www.qiagenbioinformatics.com/products/clc-main-workbench/
- Image Lab Software (Bio-Rad, version 6.0.1), RRID:SCR_014210, https://www.bio-rad.com/en-us/product/image-lab-software?ID=KRE6P5E8Z
- merge-gbk-records: RRID:SCR_017968, https://github.com/kblin/merge-gbk-records
- AdapterRemoval v2[47], RRID:SCR_011834, https://github.com/MikkelSchubert/adapterremoval
- breseq[48] (version 0.33.2), RRID:SCR_010810, https://code.google.com/archive/p/breseq/
- FastQC, RRID:SCR_014583, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
- SnapGene, RRID:SCR_015052, https://www.snapgene.com

Reagent setup

MS agar plates

Make MS agar plates (also known as SFM or MSF plates) by mixing well (with a magnetic stirrer) 20 g/L D-mannitol, fat-reduced soya flour 20 g/L, and agar 20 g/L in normal tap water. Adjust pH to 8 prior to autoclaving with 1 M NaOH. After autoclaving at 121 °C for 20 min, add pre-autoclaved $MgCl_2$ to a final concentration of 10 mM, then using a magnetic stirrer to mix until the temperature lowers to around 50 °C for pouring into Petri dishes. The MS plates can be stored at 4 °C for one month.

!CAUTION Tap water from different suppliers might affect the efficiency of conjugation and/or sporulation of some streptomycetes; different fat composition of the soya flour might also affect the efficiency of conjugation and/or sporulation of some streptomycetes.

ISP2 plates for exconjugants screening

As all plasmids described in this protocol share the same selection maker of apramycin, we used the same type of antibiotic-containing ISP2 plates for screening and maintaining the exconjugants. Dissolve BD Difco pre-mixed ISP2 powder into normal tap water and adjust pH to 7.4 prior to autoclaving with 1 M NaOH. After autoclaving (121 °C, 20 min), add 50 μg/mL apramycin and 50 μg/mL nalidixic acid (optional, 0.5 μg/mL thiostrepton for induction) and mix well until the temperature drops to around 50 °C for pouring into Petri dishes. Plates can be stored at 4 °C for up to one month.
!CAUTION Thiostrepton is not always required for induction, as the tipA promoter is leaky in many strains, and the leaky expression of its controlled components normally is enough for achieving the desired genome editing events.
!CRITICAL STEP As streptomycetes have a relatively low growth speed, nalidixic acid is used for counterselecting the E. coli used for interspecific conjugation.

ISP2 broth for routine maintenance of streptomycetes

Make ISP2 broth by mixing 4 g/L of Bacto yeast extract, 10 g/L of Malt extract, and 4 g/L of Dextrose in normal tap water using a magnetic stirrer. Adjust pH to 7.4 prior to autoclaving (121 °C, 20 min) using 1 M NaOH. Autoclaved ISP2 broth can be stored at 4 °C for up to three months.

Ampicillin stock solution (100 mg/mL)

Use an electronic scale to weigh 500 mg apramycin sulfate salt, dissolve it in 10 mL ddH_2O. Filter the solution with a 0.2 μm filter in a laminar flow hood, then aliquot into sterilized 1.5 mL Eppendorf tubes. The stock can be stored at −20 °C for up to one year.

Apramycin stock solution (50 mg/mL)

Use an electronic scale to weigh 500 mg apramycin sulfate salt, dissolve it in 10 mL ddH_2O. Filter the solution with a 0.2 μm filter in a laminar flow hood, then aliquot into sterilized 1.5 mL Eppendorf tubes. The stock can be stored at −20 °C for up to one year.

Chloramphenicol stock solution (20 mg/mL)

Use an electronic scale to weigh 200 mg kanamycin sulfate, dissolve it in 10 mL absolute ethanol. Filter the solution with a 0.2 μm filter in a laminar flow hood, then aliquot into sterilized 1.5 mL Eppendorf tubes. The stock can be stored at −20 °C for up to one year.

Kanamycin stock solution (50 mg/mL)

Use an electronic scale to weigh 500 mg kanamycin sulfate, dissolve it in 10 mL ddH_2O. Filter the solution with a 0.2 μm filter in a laminar flow hood, then aliquot into sterilized 1.5 mL Eppendorf tubes. The stock can be stored at −20 °C for up to one year.

Nalidixic acid stock solution (50 mg/mL)

Use an electronic scale to weigh 500 mg nalidixic acid, dissolve it in 10 mL 0.1 N NaOH. Filter the solution with a 0.2 μm filter in a laminar flow hood, then aliquot into sterilized 1.5 mL Eppendorf tubes. The stock can be stored at −20 °C for up to one year.
!CAUTION NaOH is corrosive. Use it with caution.

Thiostrepton stock solution (5 mg/mL)

Use an electronic scale to weigh 10 mg apramycin sulfate salt, dissolve it in 2 mL DMSO. Filter the solution with a 0.2 μm filter in a laminar flow hood, then aliquot into sterilized 1.5 mL Eppendorf tubes. The stock can be stored at −20 °C for up to one year.

SOC broth for E. coli transformation

Make SOC broth by mixing 20 g/L of Bacto Tryptone, 5 g/L of Bacto yeast extract, 4.8 g/L of $MgSO_4.7H_2O$, 3.603 g/L of Dextrose, 0.5 g/L of NaCl, and 0.186 of g/L KCl in normal tap water using a magnetic stirrer. Adjust pH to 7.0 prior to autoclaving (121 °C, 20 min) with 1 M NaOH. Autoclaved SOC broth can be stored at 4 °C for up to three months.

2×YT broth

Make 2×YT broth by mixing 16 g/L of Bacto Tryptone, 10 g/L of Bacto yeast extract, and 5 g/L of NaCl in normal tap water using a magnetic stirrer. Adjust pH to 7.0 prior to autoclaving (121 °C, 20 min) with 1 M NaOH. Autoclaved 2xYT broth can be stored at 4 °C for up to three months.

Glycerol, 50% and 10% (vol/vol)

Mix 100 mL glycerol with 100 mL ddH_2O, autoclave it at 121 °C for 20 min. 10% Glycerol is made by mixing 100 mL of 50% glycerol with 400 mL of ddH_2O.Both glycerol solutions can be stored at room temperature up to one year.

0.5 M EDTA (pH 8.0)

Use an electronic scale to weigh 93.05 g EDTA, dissolve it in approximately 400 mL ddH_2O, and adjust the pH using NaOH pellets to 8.0. Top up the solution to a final volume of 500 mL, autoclave it at 121 °C for 20 min. Autoclaved EDTA solution can be stored at 4 °C for up to three months.

50×TAE buffer

Use an electronic scale to weigh 242 g Trizma base (Tris-base), dissolve it in approximately 700 mL ddH_2O, carefully add 57.1 mL 100% acetic acid and 100 mL of 0.5 M EDTA (pH 8.0), adjust the solution to a final volume of 1 L. pH does not need to be adjusted and should be about 8.5. 1×TAE buffer can be made by mixing 20 mL of 50×TAE buffer (vol/vol) with 980 mL of ddH_2O. Both TAE buffers can be stored at room temperature up to six months.

Buffer B1

Make Buffer B1 by dissolving 18.61 g $Na_2EDTA{\cdot}2H_2O$ and 6.06 g Trizma base in 800 mL ddH_2O. Then add 50 mL 10% Tween-20 solution and 50 mL 10% Triton X-100 solution. Adjust the pH to 8.0 with 1M HCl. Adjust the volume to 1 L with ddH_2O. Filter the solution with a 0.2 μm filter in a laminar flow good. It can be stored at room temperature up to one month.

Buffer B2

Make Buffer B2 by dissolving 286.59 g guanidine HCl in 700 mL ddH_2O. Add 200 mL of 100% Tween-20. Adjust the volume to 1 L with ddH_2O. pH does not need to be adjusted. Filter the solution with a 0.2 μm filter in a laminar flow hood. It can be stored at room temperature up to one month.

pJET1.2-sgRNAhandle

Clone the 82-nt sgRNA handle followed by a 28-nt Csy4 recognition site (GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTG AAAAAGTGGCACCGAGTCGGTGCTTTTTTGTTCACTGCCGTATAGGCAGCTAA GAAA) into pJET1.2 (from CloneJET PCR Cloning Kit).
!CRITICAL 100 μg/mL ampicillin is used to maintain the pJET1.2-sgRNAhandle carrying DH5alpha *E. coli.*

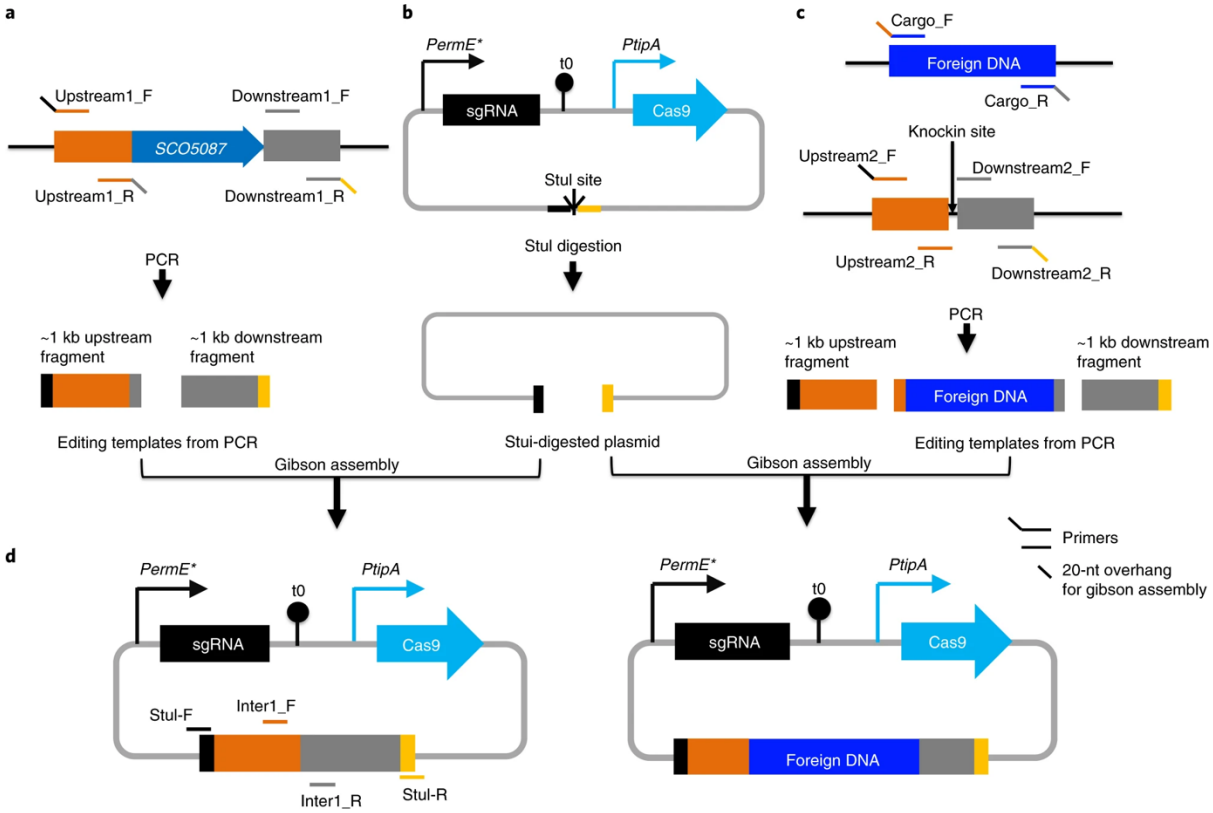


Figure 5: An overview of CRISPR–Cas9 plasmid construction for in-frame deletion or insertion of foreign DNA. **a**, Design primers and then PCR amplify ~1 kb up- and downstream DNA fragments of the target gene (SCO5087 gene is used as an example[9]). For Gibson assembly, the PCR product of the upstream fragment contains 20-bp overhangs from the CRISPR–Cas9 plasmid (shown in black) and the downstream fragment; the PCR product of the downstream fragment contains a 20-bp overhang from the CRISPR–Cas9 plasmid[9] (shown in yellow). **b**, Linearize the desired CRISPR–Cas9 plasmid correctly assembled to contain the sgRNA cassette (from Step 19) by StuI digestion. **c**, Design primers and then PCR-amplify the DNA of interest together with ~1 kb up- and downstream DNA fragments of the knockin site. For Gibson assembly, the PCR products of the up- and downstream fragments each contain a 20-bp overhang from the CRISPR–Cas9 plasmid (shown in black and yellow, respectively), whereas the PCR product of the DNA of interest contains 20-bp overhangs from the up- and downstream fragments. **d**, Use Gibson assembly to assemble fragments from a and b to construct the desired CRISPR–Cas9 plasmid for in-frame deletion[9], or from **b** and **c** for DNA-of-interest knockin. Primer sets of inter1_F and inter1_R and StuI-F and StuI-R are used for assembly validation. P*tipA*, thiostrepton-inducible promoter; P*ermE**, a widely used constitutive promoter; t0, terminator; StuI site: the restriction enzyme StuI recognition and cleavage site.

PROCEDURE

An overview of the entire procedure is shown in Fig. 4.

!CRITICAL All reagents and equipment are used according to the instructions provided by their manufacturers, unless the specific modifications are indicated. DNA oligos were synthesized by IDT, custom DNA fragments that were codon optimized (optional) were synthesized by Genscript. Diligently follow all waste disposal regulations of your institute/university/local government when disposing of waste materials. Living bacteria-related operations are carried out in a laminar flow hood to avoid contamination. All primer solutions are 20 μM in this protocol, unless the specific concentrations are indicated.

sgRNA design

Timing 30 min

1. Upload the genome-of-interest. Genome data can be uploaded to CRISPy-web at https://crispy.secondarymetabolites.org/ in GenBank format containing CDS feature annotations for the genes-of-interest. This can be achieved manually using option A, or directly via combination of the CRISPy-web and the antiSMASH genome mining tool[49], using option B.
   - **A.** Submit genome data to CRISPy-web manually
     - **i.** Use the "browse" button to find the input file on your local machine and hit "Start" to submit the file to the CRISPy-web server.
     - **ii.** Directly import genome data into CRISPy-web using antiSMASH
   - **B.** Directly fetch files using the antiSMASH job ID via the "Get sequence from antiSMASH" tab.

     !CAUTION Gene annotations are required to predict the codon changes introduced during CRISPR-BEST editing, so genome data in FASTA format or without gene annotations is not supported.

     !CAUTION CRISPy-web is designed to work on single-chromosome, single-contig bacterial genomes. In order to properly detect off target matches on multi-contig draft genomes, merge all contigs

into a single record (using a tool like https://github.com/kblin/merge-gbk-records).

?TROUBLESHOOTING

2. Select the region-of-interest. After a few seconds, CRISPy-web will load the genome data and present the region selection screen, which accepts genomic coordinates, locus tag names, or antiSMASH region numbers (if antiSMASH has found and annotated any biosynthetic regions-of-interest) as input. Genomic coordinates are the most flexible form of input. For example, if the region you want to edit spans 20 kb from the base 100,000 bp, you would enter a range of "100000-120000".

   !CRITICAL STEP The "Expert settings" panel allows you to tweak the editing window size and offset from the PAM, but unless you have evidence of the edit window properties being different in your strain-of-interest, it is recommended to keep the defaults.

   ?TROUBLESHOOTING

3. Click the "Find targets" button to start the protospacer identification process. Depending on the size of your uploaded genome and the current server load, the scan for protospacers will take several minutes. Once the scan is complete, your selected region is shown on top of a table with detailed protospacer descriptions. Genes are denoted as arrows pointing right for genes on the forward strand and pointing left for genes on the reverse strand. Protospacers are indicated by red boxes above and below the gene arrows, depending on the strand of the protospacer. If your region of interest is relatively large, it might be hard to visually distinguish different protospacers on a single gene. You can zoom into genes of interest by left-clicking on a gene arrow and selecting "Show results for this gene only". You can always go back to the overview by using the "Go back" button on the top left of the screen.

4. Select appropriate CRISPR spacers for the desired application. Initially, CRISPy-web shows all potential protospacers. Since these protospacers are scored by potential off-target calculation, one can directly select the desired spacers here for non-base editing applications. Click the "CRISPR-BEST mode" button to restrict the view to protospacers that can be used for CRISPR-BEST applications. This filters for protospacers that introduce amino acid changes when

used in CRISPR-BEST. Click the "Show only STOP mutations" button to further restrict the results to only show the protospacers that can be used to introduce stop codons.

5. To select a protospacer for download, simply click on its table row. The text color will change to blue and the counter on the "shopping cart" button on the upper right will increase by one. More than one protospacer can be selected this way.
6. Once you finish your selection of protospacers, click the "shopping cart" button on the top right. You will be shown an overview of selected protospacers.
7. Hit the "Download CSV file" to download the selected protospacers in a format compatible with spreadsheet applications. The 20 nucleotides in the "Sequence" column will be used for sgRNA assembly in Step 8.

   ?TROUBLESHOOTING

   !CRITICAL STEP In general, low numbers of off-target hits represent good protospacers. In general, it is advised to select protospacers with no "0 bp mismatches", and if possible, with no 1 bp and 2 bp mismatches for maximum specificity. However, for introducing stop codons, good protospacers strike a balance of a low number of off target hits and a location as close to the start codon as possible.

   !CRITICAL STEP Mutations are displayed as wildtype amino acid, coordinate in the protein sequence, and amino acid after mutation. For example, if the mutation is "L29F", the leucin at position 29 of the protein will be mutated to a phenylalanine. An amino acid code of "*" indicates a stop codon.

   !CRITICAL STEP Please note that CRISPy-web predicts protospacers that target both DNA strands. CRISPRi, however, only works with sgRNAs binding to the non-template DNA strand (coding strand or sense strand) when targeting within an open reading frame (ORF), or sgRNAs binding to both strands of the promoter region[9,45]. This makes protospacer selection for CRISPRi application a critical step. We recommend selecting protospacers in CRISPy-web that can bind to the non-template DNA strand (also known as coding strand, and sense strand), either binding to promoter or binding to an ORF region that is close to the start codon.

Construction and validation of the desired CRISPR plasmid

Timing 7-16 days

8. Prepare sgRNA cassettes and editing templates (optional). Follow option A if single-target CRISPR plasmids are going to be constructed, and option B if multiple-target CRISPR plasmids are going to be constructed. Construction of an editing template (option C), is only required for in-frame insertions or deletions.

A. Preparing 20-bp protospacer containing ssDNA oligos for single target CRISPR plasmid construction.

i. Design ssDNA oligos for the single-target CRISPR plasmid construction from the protospacer sequences obtained in Step 7. To facilitate the ssDNA oligo design for protospacer integration in all single target CRISPR plasmids, use the following template: CGGTTGGTAGGATCGACGGC-N20-GTTTTAGAGCTAGAAATAGC, N20 is the place to insert the custom 20-nt protospacer from Step 7.

ii. Directly order the ssDNA oligo from the supplier, for example IDT with standard desalting protocol.

iii. Resuspend the oligos into 100 μM stocks using 1x NEBuffer 2.

iv. Before starting Step 13(A) dilute the oligos to a working concentration of 0.2 μM using 1x NEBuffer 2.

¡PAUSE POINT

The prepared solutions, together with the stocks, can be stored at −20 °C for up to 6 months.

B. Preparing sgRNA cassettes for multiple-target CRISPR plasmid construction.

i. sgRNA cassette preparation for multiplexing-compatible CRISPR plasmids requires PCR to obtain the cassettes. Firstly, design and order primers for each sgRNA fragment according to the design rules described in the Experimental design section (an example is presented in Box1).

ii. Carry out the PCRs for each sgRNA unit in a 50 μL reaction system using pJET1.2-sgRNAhandle (see materials section) as template with the following set up and conditions:

Component	Amount (μL)	Final concentration
Forward primer (from Step 8(B)(i))	1	400 nM
Reverse primer (From Step 8(B)(i))	1	400 nM
Template DNA	0.5 (~50 ng)	
2×Phusion High-Fidelity PCR Master Mix with HF Buffer	25	1×
ddH_2O	22.5	
Total	50 (one reaction)	

!CRITICAL STEP We obtained equal efficiency by using NEB Q5® High-Fidelity 2× Master Mix kit and the 2×Phusion High-Fidelity PCR Master Mix with HF Buffer.

Cycle number	Denature	Anneal	Extend	Final
1	98 °C, 20 s			
2-31	98 °C, 10 s	60 °C (from Tm calculator), 30 s	72 °C, 10 s	
32			72 °C, 5 min	
33				10 °C, hold

!CRITICAL STEP The annealing temperature (Ta) in this work is calculated with a Tm Calculator from NEB https://tmcalculator.neb.com/#!/main. Other similar available Tm Calculators can also be used.

!CAUTION The extension time is calculated according to the used DNA Polymerase, i.e., 15-30 s/kb for Phusion High-Fidelity polymerase.

iii. Analyze 5 μL of each PCR reaction (add 1 μL of 6x DNA gel loading dye) along with the GeneRuler 100 bp DNA Ladder on an agarose gel (1%) with 1×TAE buffer. Run the gel at 100 V for 30 min and visualize the bands using a Bio-Rad Gel Doc XR+ Gel Documentation System. A band with approximately 150 bp is expected.

iv. Purify the positive fragments using NucleoSpin® Gel and PCR Clean-up kit, according to manufacturer's instructions.

v. Measure the concentration of each fragment using a NanoDrop 2000.

?TROUBLESHOOTING

vi. Pool all fragments, calculate the volume required of each fragment based on the measured concentrations (from Step 8(B)(v) and a desired molar ratio of 5:1 (insert:vector).

!CRITICAL STEP This equation is used for pmol-ng conversion. Pmols = (weight in ng) x 1,000 / (base pairs x 650 daltons)

vii. Digest the sgRNA fragments using FastDigest Eco31I (BsaI). Set up a 20 µL reaction using a mix of 1 µL FastDigest Eco31I and 1 µL FastAP thermosensitive alkaline phosphatase and incubate for 30 min at 37 °C. Heat inactivate for 10 min at 65 °C.

!CRITICAL STEP This reaction can be directly used for the following ligation, however, we recommend firstly purifying the fragments using NucleoSpin® Gel and PCR Clean-up kit, according to the manufacturer's instructions.

!CRITICAL STEP The amount of FastDigest Eco31I (BsaI) needs to be calculated based on the number of fragments, we recommend using 0.2 µL FastDigest BsaI for each fragment in a 20 µL reaction. In total, the amount of FastDigest Eco31I should not exceed 1 µL. If one has more than 5 fragments, we recommend using a NEB BsaI-HF®v2 kit to assemble the sgRNA array into the pGGA vector included in the kit, then apply a NcoI-NheI double digestion to isolate the pre-assembled sgRNA array.

!CRITICAL STEP As there are many BsaI sites in the pSG5 plasmid backbone, a two-step Golden Gate Assembly reaction needs to be set up to assemble the sgRNA fragments into the multiplexing-compatible CRISPR plasmids. Therefore, at this stage, removal of all restriction enzymes by proper heat inactivation or clean up of the fragments is of great importance.

¡PAUSE POINT

The prepared sgRNA fragments can be stored at −20 °C for up to 6 months.

Box 1: An example of a three-sgRNA array design[15].

The three-spacer sgRNA array was designed to simultaneously target the three key enzymes from three biosynthetic gene clusters (BGCs) in *S. coelicolor* A3(2), SCO5087 from the actinorhodin gene cluster (BLUE), SCO3230 from the calcium-dependent antibiotic gene cluster (CDA), and SCO5892 from the undecylprodigiosin gene cluster (RED). The sgRNA array is organized by the order of BLUE-CDA-RED. The RE handle is in black, the BsaI RE site is in lowercase letters, the 28-nt Csy4 recognition site is in blue, the 82-nt sgRNA handle is in red, the 20-nt spacer is in green, and the 4-nt overhang is underlined. All sequences are in 5' → 3'.

- **sgRNA-BLUE:**

  GATCGggtctccCATGGTTCACTGCCGTATAGGCAGCTAAGAAACCGTTCACAGGTCGCGGCGGGTTTTAGAGC TAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTT GTTCACTGCCGTATAGGCAGCTAAGAAAGCGGtgagaccCGATC

- **sgRNA-CDA:**

  GATCGggtctcaGCGGCGAACCAGCCCATCATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG TTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGTTCACTGCCGTATAGGCAGCTAAGAAACCCCt gagaccCGATC

- **sgRNA-RED:**

  GATCAggtctcaCCCCCAGGACGTGGAACAGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCC GTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGTTCACTGCCGTATAGGCAGCTAAGAAACTA GcgagaccTGATC

- **sgRNA-BLUE-F:**

  GATCAGGTCTCGCATGGTTCACTGCCGTATAGGCAGCTAAGAAACCGTTCACAGGTCGCGGCGGGTTTTAGA GCTAGAAATAGCAAGT

- **sgRNA-BLUE-R:** 5'- GATCGGGTCTCACCGCTTTCTTAGCTGCCTATACGG -3'
- **sgRNA-CDA-F:** 5'- GATCGGGTCTCAGCGGCGAACCAGCCCATCATGTTTTAGAGCTAGAAATAGCAAGT -3'
- **sgRNA-CDA-R:** 5'- GATCGGGTCTCAGGGGTTTCTTAGCTGCCTATACGG -3'
- **sgRNA-RED-F:** 5'-GATCAGGTCTCACCCCCAGGACGTGGAACAGAGTTTTAGAGCTAGAAATAGCAAGT -3'
- **sgRNA-RED-R:** 5'- GATCAGGTCTCGCTAGTTTCTTAGCTGCCTATACGG -3'

C. Preparing editing templates (optional).

Timing 5-12 d

CRITICAL: This step is required only for creating in-frame deletions/insertions. The Gibson Assembly in this step needs CRISPR-Cas9 plasmids constructed with sgRNAs in Step 13(A), however, the editing template preparation in this step can be carried out in parallel with CRISPR plasmid construction (Steps 9-19) to save time.

i. Design and order primer sets that can amplify two ~ 1kb DNA fragments flanking the DNA sequence of interest, (optionally, include the DNA sequence to be inserted) with a 20-nt overhang at the ends of each fragment for later Gibson Assembly purpose.

ii. Use direct *Streptomyces* colony PCR (see Step 34(B)) or a genomic DNA PCR for amplification of each of the required DNA fragments. In the latter case, a 50 μL PCR reaction is carried out as follows:

Component	Amount (μL)	Final concentration
Forward primer (from Step 8(C)(i))	1	400 nM
Reverse primer (from Step 8(C)(i))	1	400 nM
Template DNA	1 (~100 ng)	
2×Phusion High-Fidelity PCR Master Mix with GC Buffer	25	1×
DMSO	1.5	3%
ddH_2O	20.5	
Total	50 (one reaction)	

!CRITICAL STEP We recommend adding 3% DMSO if pure genomic DNA is used as the PCR template.

Cycle number	Denature		Extend	Final
1	98 °C, 20 s			
2-36	98 °C, 10 s	(from Tm calculator), 30 s	72 °C, 30 s	
37			72 °C, 5 min	
38				10 °C, hold

iii. Analyze 5 μL of each PCR reaction (add 1 μL of 6x DNA gel loading dye) along with the GeneRuler 1 kb DNA Ladder on an

agarose gel (1%) with 1×TAE buffer. Run the gel at 100 V for 30 min and visualize the bands using a Bio-Rad Gel Doc XR+ Gel Documentation System.

iv. Purify successfully amplified fragments either by PCR clean-up (if the bands are unique and sharp) or gel purification (if unspecific bands are observed) using a NucleoSpin® Gel and PCR Clean-up kit, according to the manufacturer's instructions.

v. Measure the concentration of each purified fragment using a NanoDrop 2000.

¡PAUSE POINT

The purified fragments can be stored at −20 °C for up to 3 months.

?TROUBLESHOOTING

vi. Digest the CRISPR-Cas9 plasmid bearing the desired sgRNA from Step 13(A), with FastDigest Eco147I (StuI). Ideally, use a 50 μL reaction system containing 2 μg plasmid DNA, 5 μL 10× FastDigest Buffer, 2 μL FastDigest Eco147I, and 1 μL FastAP thermosensitive alkaline phosphatase (1 U/μL). Incubate at 37 °C for 30 min.

vii. Analyze 2 μL of the reaction (add 1 μL of 6x DNA gel loading dye and 4 μL of ddH_2O) along with the GeneRuler 1 kb DNA Ladder on an agarose gel (1%) with 1×TAE buffer. Run the gel at 100 V for 30 min and visualize the bands using a Bio-Rad Gel Doc XR+ Gel Documentation System.

viii. Clean-up the gel confirmed linearized plasmid using PCR clean-up protocol with a NucleoSpin® Gel and PCR Clean-up kit, according to the manufacturer's instructions.

ix. Measure the concentration using a NanoDrop 2000. (Optionally, a gel purification step can also be used for getting the linearized plasmid fragment, however, it will normally give a much lower yield.)

¡PAUSE POINT

The linearized plasmids can be stored at −20 °C for up to 3 months.

x. Assemble the 2 (3, if the application is in-frame insertion) fragments obtained from Step 8(C)(v) into the StuI linearized CRISPR-Cas9 plasmid (from Step 8(C)(ix)) by Gibson Assembly

using NEBuilder® HiFi DNA Assembly Master Mix. The required volume of each fragment is calculated based on the Nanodrop 2000 measured concentrations and a recommended molar ratio of 3:1 (insert:vector). Use a 10 μL reaction system with 100 ng of the linearized plasmid for this Gibson Assembly reaction.

xi. Flick the tube 3-5 times with a fingertip, spin down the reaction, and incubate at 50 °C in a thermocycler for 60 min.

!CRITICAL STEP This equation is used for pmol-ng conversion.
Pmols = (weight in ng) x 1,000 / (base pairs x 650 daltons)

¡PAUSE POINT

The Gibson reaction can be stored at −20 °C for up to 3 months.

xii. Transform 50 μL of One Shot® Mach1™ T1 Phage-Resistant competent *E. coli* cells with the complete 10 μL reaction from the above step using the 42 °C heat shock transformation protocol (see step 13(B)(ii))

xiii. Screen the positive clones by *E. coli* colony PCR (see the procedures in steps 14-17) with a primer set of StuI-F (5'-GACAATGACAACAACCATCGCC-3') and StuI-R (5'-GGGAAGTCGTCGCTCTCTGG-3'), which flank the editing template.

!CRITICAL STEP Amplification of a > 2 kb high GC DNA fragment is relatively challenging. Therefore, we recommend using a custom primer set flanking a ~500 bp region of the joint site of the up-stream and down-stream homologous recombination templates.

xiv. Confirm the positive clones identified by colony PCR by Sanger sequencing using the same procedures from Step 19. Normally, both StuI-F and StuI-R and an additional sequencing primer that can bridge the reads of StuI-F and StuI-R are used to make sure the in-frame deletion/insertion can be read out.

xv. Make a 25% glycerol stock using the same procedure described in step 20.

9. Linearize the CRISPR plasmids by restriction enzyme digestion. Follow option A if non-multiplexing-compatible CRISPR plasmids are constructed, and option B for construction of multiplexing-compatible CRISPR plasmids.

!CRITICAL STEP As Thermo Fisher Scientific provides a compatible buffer system, both FastDigest restriction enzymes and FastAP thermosensitive alkaline phosphatase can be added at the same time for a total reaction time of 30 min. Instead of a 20 μL volume digestion, one can enlarge the volume to yield more linearized plasmid that can be stored at −20 °C for up to three months for future use.

A. Digesting non-multiplexing-compatible CRISPR plasmids.

i. Prepare a single digestion of the selected CRISPR plasmids (pCRISPR-Cas9; pCRISPR-Cas9-ScaligD; pCRISPR-dCas9; pCRISPR-cBEST; and pCRISPR-aBEST) (Fig. 1) for single editing applications with FastDigest NcoI in a 20 μL digestion system containing the following components:

Component	Amount (μL)	Final concentration
Plasmid DNA	10	40 ng/μL
FastDigest NcoI	1	
10× FastDigest buffer	2	1×
dd water	7	
Total volume	20	

Ideally, digest 800 ng of plasmid DNA. Incubate at 37 °C for 30 min. Then add 1 μL of FastAP thermosensitive alkaline phosphatase to the reaction and incubate for additional 10 min at 37 °C.

¡PAUSE POINT

The linearized plasmids can be stored at −20 °C for up to three months.

B. Digesting multiplexing-compatible CRISPR plasmids.

i. Use NcoI and NheI to digest the multiplexing compatible CRISPR plasmids (Fig. 3). Prepare a double digestion of the multiplexing-compatible CRISPR plasmid (pCRISPR-McBEST) with FastDigest NcoI and FastDigest NheI in a 20 μL digestion system containing the following components:

Component	Amount (μL)	Final concentration
Plasmid DNA	10	40 ng/μL
FastDigest NcoI	1	
FastDigest NheI	1	
10× FastDigest buffer	2	1×
dd water	6	
Total volume	20	

Ideally, digest 800 ng of plasmid DNA. Incubate it at 37 °C for 30 min, the add 1 μL of FastAP thermosensitive alkaline phosphatase to the reaction and incubate for additional 10 min at 37 °C. following an inactivation step at 75 °C for 10 min.

¡PAUSE POINT

The linearized plasmids can be stored at −20 °C for up to three months.

10. Analyze 2 μL of the above digestion reaction (add 1 μL of 6x DNA gel loading dye and 4 μL of ddH_2O) along with the GeneRuler 1 kb DNA Ladder on an agarose gel (1%) with 1×TAE buffer. Run the gel at 100 V for 30 min and visualize the bands using a Bio-Rad Gel Doc XR+ Gel Documentation System.

11. Clean-up the gel confirmed linearized plasmid with a NucleoSpin® Gel and PCR Clean-up kit, according to the manufacturer's instructions.

12. Measure the concentration using a NanoDrop 2000. (Optionally, a gel purification step can also be used for getting the linearized plasmid fragment, however, it will normally give a much lower yield. The gel purification step can be carried out with a NucleoSpin® Gel and PCR Clean-up kit, according to the manufacturer's instructions.)

13. To assemble the sgRNA cassettes (and optionally the editing template) into the digested CRISPR plasmid of choice, follow option A for construction of non-multiplexing-compatible CRISPR plasmids, or option B for construction of multiplexing-compatible CRISPR plasmids.

A. Construction of non-multiplexing-compatible CRISPR plasmids

i. To clone the ssDNA oligo containing the 20-nt spacer from Step 8(A)(iv) into the linearized single-target CRISPR plasmid using the PCR-free ssDNA oligo bridging method[18], firstly prepare a 20 μL reaction mix containing 30 ng of the linearized plasmid from Step 9(A)(i), 5 μL of the 0.2 μM oligo from Step 8(A)(iv), 10 μL of 2 x NEBuilder® HiFi DNA Assembly Master Mix and ddH_2O to 20 μL. Incubate the reaction for 1 h at 50 °C.

!CRITICAL STEP The reaction volume can be reduced proportionally to 10 μL in order to save reagents.

¡PAUSE POINT

The reaction can be stored at −20 °C for up to three months.

ii. Transfer 2 μL of the above reaction mixture into 50 μL in-house-made (or commercial) electroporation competent Mach1 *E. coli*. cells (see Box. 2 for making high-efficient electroporation competent *E. coli*) by the following procedures: take the 50 μL electroporation competent *E. coli* cells containing tubes out of –80 °C and thaw on ice (approximately 10 min). In the meantime, take the 1 mm electroporation cuvettes out of −20 °C and place them on ice. Carefully pipette the competent cells into the cuvettes. Mix the competent cells with 2 μL of the ssDNA oligo bridging from Step 13(A)(i) reaction by flicking with a fingertip 3-5 times (avoid bubble formation). Use the Ec1 program of a Bio-rad MicroPulser (alternatively, a similar electroporation program of 1.8 kV with a 1 mm electroporation cuvette of one time of pulse can be used). Immediately add 200 μL SOC into each cuvette, transfer the reaction into a sterilized 1.5 mL Eppendorf tube. Incubate it in a heating block at 37 °C and 800 rpm for 1 h.

iii. Plate 100 μL of the reaction onto a selective LB plate supplemented with 50 μg/mL apramycin. Incubate the plate overnight at 37 °C.

!CAUTION The transformation efficiency may differ with home-made competent cells. Therefore, we recommend using the commercial products if they are available.

!CRITICAL STEP Chemically competent cells can also be used in this step with a 42 °C heat shock protocol from Step 13(B)(ii).

B. Assemble the sgRNA cassettes into the multiplexing compatible CRISPR plasmids by a two-step Golden Gate Assembly.

i. Set up a 10 μL ligation reaction using T4 Ligase (5U), 100 ng of pre-digested plasmids from step 9(B) and the required volume of the prepared sgRNA fragments from step 8(B) (5:1 molar ratio, use directly the heat inactivated digestion reaction). Supplement the reaction with 2 μL of 50% PEG-4000 (included in the T4 Ligase kit) and incubate for 1 h at 22 °C.

¡PAUSE POINT

The reaction can be stored at −20 °C for up to three months.

ii. Transfer 5 μL of the above reaction into 50 μL One Shot™ Mach1™ T1 Phage-Resistant Chemically Competent *E. coli* using a heat

shock protocol as follows: take 50 μL aliquots of the chemically competent *E. coli* cells out of –80 °C and thaw on ice (approximately 10 min). Mix it with 5 μL of the reaction obtained from the above step by flicking with a fingertip (3-5 times). Incubate on ice for 20 min, followed by a 60 s heat shock at 42 °C using a water bath, steadily transfer the tubes into ice and incubate for another 5 min. Add 200 μL SOC into the tubes, incubate in a heating block under conditions of 37 °C, 800 rpm for 1 h.

iii. Plate all of the reaction on a selective LB plate supplemented with 50 μg/mL apramycin. Incubate the plate overnight at 37 °C.
!CRITICAL STEP Electroporation competent cells can also be used in this step with an electroporation transformation protocol from step 13(A)(ii).

14. Screen the clones using an *E. coli* colony PCR as follows: in the morning following Step 13, pick 12-24 *E. coli* colonies of each assembly into a 96-deep-well plate containing 300 μL LB broth supplemented with 50 μg/mL apramycin in each well using sterilized wooden toothpicks.

15. Incubate the plate at 37 °C at 300 rpm for 2 hours.

16. Directly use 1 μL of the culture as templates for colony PCR with the following set up and conditions:

Component	**Amount (μL)**	**Final concentration**
sgRNA-TEST-F	0.5	500 nM
sgRNA-TEST-R	0.5	500 nM
Template DNA	1	
OneTaq® 2× Master Mix with Standard Buffer	10	1×
ddH_2O	8	
Total	20 (one reaction)	

Primer sequences:
sgRNA-TEST-F: 5’- AATTGTACGCGGTCGATCTT-3’
sgRNA-TEST-R: 5’-TACGTAAAAAAAGCACCGAC-3’

Cycle number	**Denature**	**Anneal**	**Extend**	**Final**
1	94 °C, 3 min			
2-31	94 °C, 30 s	50 °C, 30 s	68 °C, 30 s	
32			68 °C, 5 min	
33				10 °C, hold

17. Analyze 5 µL of the above PCR reactions (add 1 µL of 6x DNA gel loading dye) along with the GeneRuler 1 kb DNA Ladder on a long (10 cm) agarose gel (3%) with 1×TAE buffer. Run the gel at 100 V for 60 min and visualize the bands using a Bio-Rad Gel Doc XR+ Gel Documentation System.

    !CRITICAL STEP As the size differences of the positive and the control is only 20 bp, it needs a > 2% agarose gel with 60 min to distinguish. We recommend using a 10 cm 3% agarose gel, and run the gel at 100 V for 60 min.

    ¡PAUSE POINT

    The *E. coli* culture in the 96-deep-well plate can be stored at 4 °C for up to one week.

18. Prepare overnight cultures of the above-obtained positive colonies in cultivation tubes containing 5 mL of LB broth supplemented with 50 µg/mL apramycin. Inoculate 50 µL of culture directly from the 96-deep-well plate.

19. Perform plasmid isolation using the NucleoSpin® Plasmid EasyPure Kit following the manufacturer's instructions the following day and submit the plasmids for Sanger sequencing using the sgRNA-TEST-F primer (5'- AATTGTACGCGGTCGATCTT-3') and Cas9-C-terminal-TEST primer (5'-GACCCTGATCCACCAGAGCA-3').

    !CRITICAL STEP 2×YT broth supplemented with 50 µg/mL apramycin performs better than LB broth, in general. We recommend isolating plasmids from a 5 mL overnight culture.

    !CRITICAL STEP We have observed some unknown factors that can trigger the instability of the pSG5 replicon-based shuttle plasmid in *E. coli*. Therefore, it is critical to confirm the integrity of the CRISPR plasmids. We observed several cases where the "hot region of instability" lies downstream of the *tipA-fd* fragment. Therefore, we recommend running an additional Sanger sequencing with sequencing primer Cas9-C-terminal. Alternatively, a NdeI-BglII double-digestion mapping can also indicate the integrity of the plasmids.

20. Freeze the *E. coli* strains with correct plasmids (ready-to-use CRISPR plasmids) confirmed by Sanger sequencing in 25% glycerol at −80 °C for long-term storage.

    ¡PAUSE POINT

The *E. coli* glycerol stock can be stored at −80 °C for at least 5 years.

Box 2: A modified electroporation competent cell preparation protocol.

TIMING 2 DAYS

!CRITICAL All reagents and equipment that will directly contact the *E. coli* cells need to be pre-sterilized by autoclaving at 121 °C for 20 min.

Day1

1. Put all required reagents and equipment (the centrifuge rotor, 2 L ddH_2O, 100 mL 10% glycerol (v/v), 500 mL LB broth, four 250 mL centrifuge bottles, four 50 mL canonical tubes and 50 1.5 mL Eppendorf tubes) to 4 °C cold room for overnight cooling.
2. Inoculate a fresh single colony of the required *E. coli* strain into a 250 mL shake flask containing 50 mL LB broth , incubate at 37 °C 200 rpm overnight.

Day2

3. Inoculate 5 mL of the overnight culture into a 2 L shake flask with 500 mL LB broth. Incubate it at 37 °C with 200 rpm shaking. Measure OD600 every hour and every 10 min once the OD600 reaches 0.2. After the inoculation, turn on the floor centrifuge and set up a 4 °C mode with the pre-chilled centrifuge rotor.
   !CRITICAL STEP Ensure that the centrifuge is properly cooled down to 4 °C. It is recommended to start cooling down the centrifuge around one hour before using it.
4. Once an OD600 of around 0.4 is observed, bury the whole shake flask (the part with culture) into ice to chill for 30 min with occasional swirling. Also bury the four 50 mL canonical tubes, four 250 mL centrifuge bottles, ddH_2O, and 10% glycerol into ice in this step.
   !CRITICAL STEP It is important not to let the OD get any higher than 0.4.
5. Subsequently, divide the culture equally into the four centrifuge bottles and then harvest cells by centrifugation at 1000×*g* for 20 min at 4 °C. **(!Spin No. 1)**
   !CRITICAL STEP If conditions permit, carry out the following operations at 4 °C cold room. Otherwise, keep all operations as fast as possible, and transport *E. coli* containing bottles/tubes on ice.
6. Carefully decant the supernatant without disturbing the cell pellets, and gently resuspend each pellet in 100 mL of ice-cold ddH_2O. Combine two resuspensions and then centrifugate again with the same condition of Step 5. **(!Spin No. 2)**
   !CRITICAL STEP Because the cells are becoming more and more competent and fragile, no pipetting is allowed to mix or suspend. The time that cell pellets spend out of ice needs to be as short as possible.
7. During the centrifugation time, rinse each of the 50 mL canonical tubes using 10 mL ice-cold 10% glycerol. Carefully decant the supernatant without disturbing the cell pellets, and gently resuspend each pellet in 40 mL of ice-cold 10% glycerol. Then transfer the resuspensions to the 50 mL canonical tubes and centrifugate again with the same conditions as of Step 5. **(!Spin No. 3)**.
8. Repeat Step 7. to wash the cell pellets once more with ice-cold 10% glycerol to make sure the competent cells are ion-free. **(!Spin No. 4)**.
9. During the centrifugation time, place a required number of pre-chilled Eppendorf tubes on ice. Prepare liquid nitrogen at this stage as well. Carefully remove and discard the supernatants, and then resuspend each pellet in 1 mL of ice-cold 10% glycerol by carefully swirling. Aliquot 50 μL of the resuspensions into ice-cold 1.5 mL microfuge tubes and snap frozen in liquid nitrogen. Then store the electrocompetent cells at −80 °C.
   !CAUTION Liquid nitrogen is an ultra-low temperature reagent. Wear safety glasses or a face shield when transferring and operating with liquid nitrogen.
   ¡PAUSE POINT
   The electrocompetent cell can be stored at −80 °C for up to six months.

Transfer of ready-to-use CRISPR plasmids into the target streptomycetes by interspecies conjugation (Supplementary Video 1)

Timing 2 d

21. Plate the target streptomycete strains onto MS plates for sporulation, it normally is done days ahead according to the growth speed, for example, it requires around 5 days for *S. ceolicolor* WT strain to fully sporulate on MS plates.

22. Transfer 200 ng of the correct plasmids from Step 19. into 50 μL in-house-made electroporation competent *E. coli* ET12567/pUZ8002[44] cells (in the following just referred to as ET) using the electroporation protocol described in step 13(A)(ii) with small modifications as follow: instead of plating onto selective LB plates, here we transfer all the reaction into a 50 mL Falcon tube with 20 mL selective LB broth supplemented with 50 μg/mL apramycin, 25 μg/mL kanamycin, and 12.5 μg/mL chloramphenicol. Incubate the tubes overnight at 37 °C at 200 rpm.

!CRITICAL STEP It is also possible to plate the transformation reaction onto a selective LB plate and then start the culture from a single colony the next day. We recommend using the procedures we provide based on the experience that the DNA methylation defective host ET12567/pUZ8002 strain increases the risk of plasmid instability; using a pool of transformants will lower the chance of picking a wrong clone and can save at least one day.

23. The following morning, prepare the ET cultures harboring the plasmids of interest by washing twice using the same volume (20 mL) antibiotics-free LB broth, and harvest the cells by centrifuging at 5000×*g*, 5 min at room temperature. Then re-suspend the cell pellets in 2 mL antibiotics-free LB broth per 20 mL culture.

!CRITICAL STEP We recommend using an amount of LB that is 10% of the initial culture volume for re-suspension of the ET cell pellets.

24. In the meantime, start collecting spores of the streptomycete-of-interest from Step 21 by carefully pipetting 10 mL 2×YT onto the surface of the spore lawn of a well sporulated MS plate (Fig. 5a).

25. Use a sterilized cotton swab to scrape off the spores (Fig. 5b and 5c).

26. Carefully put a sterilized cotton pad above the spore suspension with sterilized tweezers (Fig. 5d).

27. Use an Integra Pipetboy to get the spore suspension through the cotton pad to remove agar and mycelial debris (Fig. 5e), then transfer it into a 50 mL Falcon tube (Fig. 5f).

!CRITICAL STEP Repeat the above operation twice to maximize the amount of spores from one plate; for strains that only produce low number of spores it may be necessary to collect spores from more than one plate.

28. Collect the spores by centrifuging at 5000×*g*, 5 min at room temperature. Discard the supernatant carefully, re-suspend the spores in 2×YT (normally, use 2 mL per plate).

29. Heat shock the spore suspension for 10 min at 50 °C. The spore suspension now is ready for conjugation.

!CRITICAL STEP In order to achieve higher conjugation efficiency, we recommend leaving the spore suspension at 4 °C overnight (can be up to 3 days) for pre-germination.

¡PAUSE POINT

The spore suspension can be used for up to 2 weeks when it is stored at 4 °C.

30. Mix 500 μL ET suspension from Step 23 with 200 μL spore suspension in a sterilized 1.5 mL Eppendorf tube by pipetting (in total, 700 μL mix is obtained).

31. Plate 100 μL, 200 μL, and 400 μL onto three different MS plates, air dry in a laminar flow hood for 5 min and incubate the conjugation plates at 30 °C for around 24 h (overnight).

32. Overlay the conjugation plates with 1 mL sterilized H_2O containing 1 mg apramycin and 1 mg nalidixic acid. Air dry the plates in a laminar flow hood for 15 min.

!CAUTION It is not recommended to use a spreader for the overlay procedure. Instead, try to spread the 1 mL sterilized H_2O containing 1 mg apramycin and 1 mg nalidixic acid just by moving the plate. The 1 mL sterilized H_2O will form clouds after adding nalidixic acid stock.

33. Incubate the plates until the exconjugants can be picked (an example of the ready to pick exconjugants are shown in Fig. 6c), normally it takes 3-5 days if the streptomycetes-of-interest have a normal growth speed. Transfer the picked exconjugants to a fresh ISP2 plate

supplemented with 50 μg/mL apramycin and 50 μg/mL nalidixic acid and incubate for 3 to 5 days at 30 °C.

?TROUBLESHOOTING

!CRITICAL STEP Almost all currently available genetic manipulation methods for streptomycetes require that the streptomycetes strain can take up plasmids. This makes conjugation a critical step of genome editing. The efficiency of conjugation can be affected by many factors. Our protocols have been successfully used for various actinomycete strains[9,15,31-37].

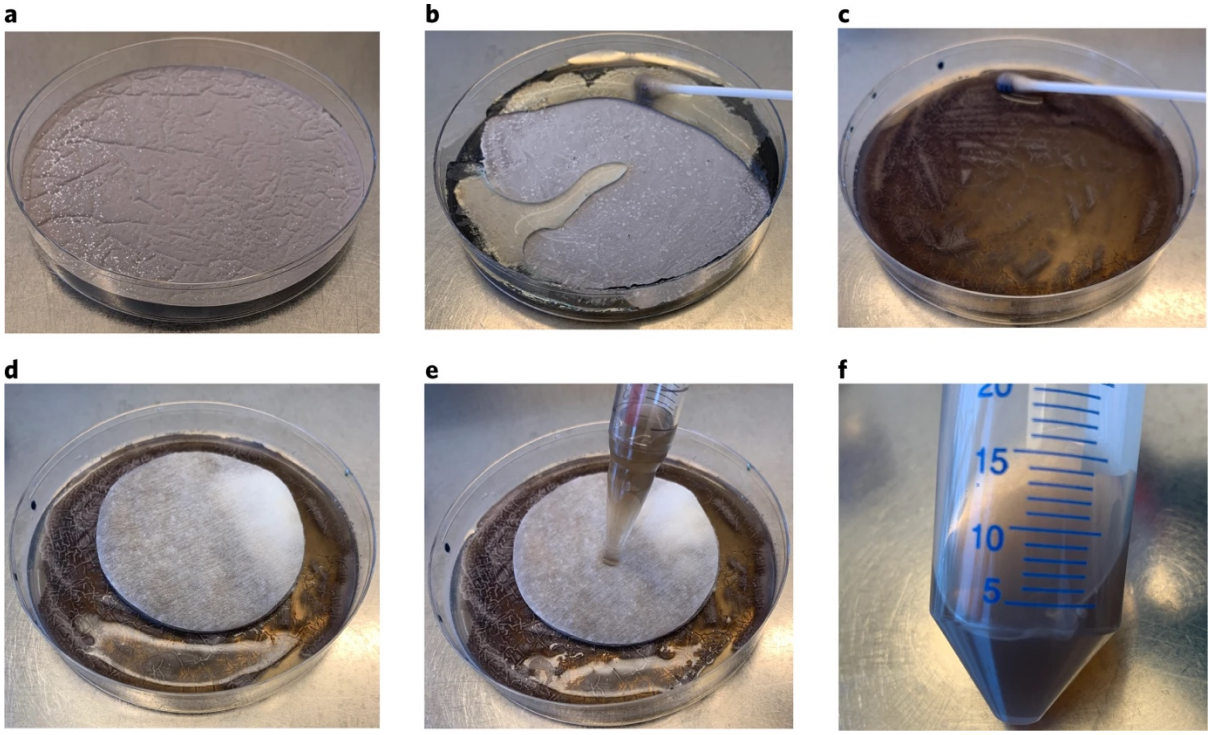


Figure 6: Experimental setup for collecting spores. A more detailed operation can be found in Supplementary Video 1. **a**, A well-sporulated *S. coelicolor* plate. **b**, Carefully pour 10 mL 2× YT on the spore lawn. **c**, Use a sterilized cotton swab to scrape off the spores. **d**, Carefully place a sterilized cotton pad above the spore suspension with sterilized tweezers. **e**, Use an Integra Pipetboy to aspirate the spore suspension through the cotton pad to remove agar and mycelial debris. **f**, Transfer the spore suspension to a 50-mL Falcon tube.

Evaluation of the successfully edited strains

Timing 5-14 d

34. Validate mutants. Use option A for evaluation of the non-genetically edited applications with CRISPRi, follow options B and option C for the chromosomally edited applications.

A. Evaluation of the non-genetically edited applications.

i. Make seed cultures by inoculating 3 randomly picked exconjugants (from Step 33) of each sgRNA and a non-treated control to individual shake flasks, each containing 50 mL selective ISP2 broth, and incubate at 30 °C, 180 rpm for 3 days.

ii. Normalize the above seed cultures with wet weight by spinning down 1 mL cultures at 10,000×*g*, 5 min at room temperature. This is to calculate equal inoculums for the main cultures as described in the next step.

iii. Inoculate 500 mg cell pellets (wet weight) from seed cultures of each exconjugant from Step 34(A)(i) to a fresh 50 mL selective ISP2 broth-containing shake flask, incubate at 30 °C, 180 rpm for 3-5 days.

iv. Extract and analyze the end-point product accordingly, a detailed example can be found in ref.[9].

!CAUTION As different sgRNAs normally have different suppression effects, we recommend targeting at least three different loci per gene by different sgRNAs. Besides directly analyzing the end-point product[9], transcription analysis can also be applied, i. e., qRT-PCR, and RNA-Seq etc.

?TROUBLESHOOTING

B. Evaluation of genetically edited applications by colony PCR.

i. Quick screening of the correctly edited clones by a *Streptomyces* colony PCR protocol: scratch around 4 square millimeters mycelia using a sterile wooden toothpick from the ISP2 plates from Step 33 incubated for 3-5 days into a PCR tube containing 20 μL DMSO; incubate for 15 min at 100 °C, 1000 rpm using a shaking/heating block. Transfer the tube to a −20 °C freezer and freeze for 30 min. Use 1.5 μL of the above-obtained solution as a

template for colony PCR. The primers flanking the target regions of around 500 bp are used. PCR is carried out as follows:

Component	Amount (μL)	Final concentration
Forward primer	1	400 nM
Reverse primer	1	400 nM
Template DNA	1.5	3% DMSO
2×Phusion High-Fidelity PCR Master Mix with HF Buffer	25	1×
ddH_2O	21.5	
Total	50 (one reaction)	

!CRITICAL STEP We obtained equal efficiency by using NEB Q5® High-Fidelity 2× Master Mix and 2×Phusion High-Fidelity PCR Master Mix with HF Buffer.

!CRITICAL STEP We recommend using the mycelia from a fast-growing stage (before you can see the sporulation) for *Streptomyces* colony PCR.

ii. Analyze 5 μL of the above PCR reactions (add 1 μL of 6x DNA gel loading dye) along with the GeneRuler 1 kb DNA Ladder on an agarose gel (1%) with 1×TAE buffer. Run the gel at 100 V for 30 min and visualize the bands using a Bio-Rad Gel Doc XR+ Gel Documentation System.

iii. Clean-up the above PCR products with clear bands using NucleoSpin® Gel and PCR Clean-up kit, following the manufacturer's instructions. Measure the concentration of each fragment using a NanoDrop 2000.

iv. Validate the edits by Sanger sequencing of 8-12 PCR products using the forward primers flanking the target regions.

!CRITICAL STEP As most of the streptomycetes are mycelial growing bacteria, they have certain growth stages with multiple chromosomes in one cell. This can sometimes lead to a mixed population of WT and edited cells. Taking into consideration that different sgRNAs often have different editing efficiencies, we recommend applying a single spore separation from the re-streak colonies of Step 33. A simple single spore isolation can be done by touching the sporulated colony using a sterilized toothpick, dipping it into 200 μL sterilized ddH_2O, and then plating 20 μL, 50 μL and 100 μL of the above-obtained spore

suspension onto three induction ISP2 plate, respectively. After 3-5 days of incubation at 30 °C, single colonies are expected to be seen from the plates.

?TROUBLESHOOTING

C. (**optional, but highly recommended**) Evaluation of genetically edited mutants by Illumina sequencing

i. Inoculate a single colony of both the Sanger sequencing validated positive strain and a wildtype parental strain into an independent 50 mL non-antibiotic ISP2 broth-containing shake flask and incubate at 30 °C, 180 rpm for 3-5 days.

ii. Isolate genomic DNA of the above strains using the Qiagen Blood & Cell Culture DNA Kit with a modified protocol showed in Box 3.

¡PAUSE POINT

The genomic DNA solution can be stored at −20 °C for up to 6 months.

?TROUBLESHOOTING

iii. Build the Illumina sequencing libraries using a KAPA HyperPlus kit, aiming for an insert size of ~ 600 nt using magnetic beads, following the manufacturer's instructions.

!CRITICAL STEP A kit with few PCR cycles is preferable to avoid a high proportion of clonal fragments. Similarly, because *Streptomyces* have a high GC content, a transposase-based sequencing library kit is highly discouraged as the fragment diversity would be low.

iv. Quantify the libraries using a fragment analyzer and a Qubit dsDNA HS Assay kit, following the manufacturer's instructions.

v. Sequence the samples on a suitable Illumina machine with a paired end protocol of 2×250 bp, such as a MiSeq or NextSeq500 to a coverage of ca. 100.

vi. After base calling the paired end reads, perform adapter and quality trimming of the reads, e.g. using Adapterremoval2[47] with the switches --trimns and --trimqualities.

vii. Run FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) to ensure that low quality data is not present in the trimmed dataset.

viii. Use breseq[48] (e.g. version 0.33.2) with a high quality reference genome sequence to estimate the changes in both the

wildtype and genome edited strains. We suggest running breseq without the switch --polymorphism-prediction as local low coverage and other technical noise can lead to false low frequency off target predictions. Evaluate the reports of both the wildtype and genome edited strains and report the predicted mutations from both datasets. Identical predicted "mutations" in the two datasets reflect mutations which happened before the genome editing experiments. Pay special attention to larger structural changes of the streptomycetes genomes, such as deletions in the chromosome ends, which would likely happen in DSB-based genome editing but rarely in DSB-free genome editing, i. e. CRISPR-BEST.

?TROUBLESHOOTING

!CAUTION The wildtype strain from the same origin can accumulate different spontaneous mutations when they are maintained and used in the laboratory over time. Therefore, we highly recommend including the parental strain when evaluating the on/off-target of the CRISPR genome editing using whole genome sequencing based techniques.

Plasmid curing (optional)

Timing 7-14 d

!CRITICAL: For a second round of editing using the same type of pSG5 based CRISPR plasmid, the CRISPR plasmids from the successfully edited streptomycete strains need to be eliminated by the following steps:

35. Inoculate a single colony of the successfully edited streptomycete strain from Step 34(B) into a 50 mL non-selective YEME broth with 3.4% sucrose-containing 250 mL shake flask, incubate at 40 °C, 180 rpm for 3-5 days to reach the exponential growth phase.

!CAUTION The incubation time depends on the target strain itself.

36. Plate an appropriately (1,000 to 10,000-fold, depending on the density of the obtained culture) diluted fraction of the above culture on a non-selective ISP2 plate, incubate at 30 °C for 3-5 days to obtain single colonies.

!CRITICAL STEP We recommend having around 100-200 single colonies on each standard 10-cm Petri dish.

37. Replicate the above plate containing single colonies onto another 50 μg/mL apramycin supplemented ISP2 plate.

!CAUTION Due to the vegetative growth feature of *Streptomyces,* it is relatively difficult to replicate colonies from one plate to another. We recommend directly picking and streaking the well-marked colonies using sterilized wooden toothpicks.

38. After 3-5 d incubation at 30 °C, some colonies with restored apramycin sensitivity can be observed. These are CRISPR plasmid cured strains.

39. For long-term storage, prepare a 25 % glycerol stock of a spore suspension or liquid culture at exponential phase of the correctly edited strain with the CRISPR plasmids (from step 34(B)) or without the CRISPR plasmids (from step 38) and store at −80 °C.

¡PAUSE POINT

The glycerol stock can be stored at −80 °C for at least 5 years.

Box 3: Modified protocol for using the Qiagen Blood & Cell Culture DNA Kit for streptomycetes.

1. Harvest the cell pellets from 10 mL of the 5 days old streptomycetes-of-interest culture by centrifugation at 10,000 g for 10 min at room temperature.
2. Discard supernatant and resuspend the above pellets in 3.5 mL Buffer B1 with 70 μL RNase A solution (10 mg/ml) by vortexing at top speed for 10 s.
3. Add 100 μL lysozyme stock solution (100 mg/mL), and 100 μL Protease K solution then incubate in a water bath at 37 °C for 60 min.
   !CAUTION The incubation time may need to be prolonged if a clear lysis cannot be observed. Overnight incubation at 4 °C is acceptable.
4. Add 1.2 mL Buffer B2, mix it well by inverting the tube 5-10 times, incubate in a water bath at 50 °C for 60 min.
5. During the last 10 min of the incubation, equilibrate a QIAGEN Genomic-tip 100/G with 4 mL Buffer QBT from the Blood & Cell Culture DNA Kit. Empty the QIAGEN Genomic-tip by gravity.
6. Vortex the lysis sample from Step 4 of this Box for 10 s at top speed and load it to the equilibrated QIAGEN Genomic-tip. Allow it to enter the resin by gravity flow.
7. Wash the QIAGEN Genomic-tip with 2 × 7.5 mL Buffer QC.
8. Elute the genomic DNA with 5 mL Buffer QF into a 15-mL Falcon tube.
9. Precipitate the DNA by adding 3.5 mL room temperature (22-25 °C) isopropanol, invert the tube 5-10 times.
10. Centrifuge the above solution immediately at 10,000×g for 20 min at 4 °C. Carefully remove and discard the supernatant by pipetting.
11. Wash the DNA pellets with 2 mL ice-cold 70% ethanol twice.
12. Carefully remove the last droplets by pipetting, and air dry the DNA pellets for 15 min at room temperature.
13. Dissolve the DNA pellets with ddH_2O (pH 8.5) at 55 °C for 1 h.
14. Run 1 μL of the above PCR reactions and GeneRuler 1 kb DNA Ladder on an agarose gel (0.7%) with 1×TAE. Run the gel at 100 V for 30 min and visualize the bands using a Bio-Rad Gel Doc XR+ Gel Documentation System to check the integrity of the isolated genomic DNA.
15. Measure the 260/280, and 260/230 ratios using a NanoDrop 2000, and measure the concentration with a Qubit 2.0 Fluorometer.
    !CAUTION A good genomic DNA sample for Illumina sequencing should have the 260/280 and 260/230 ratios at about 1.8 and 2.2, respectively.

Timing

The minimal timing is based on optimal conditions, i. e., receiving DNA oligos from provider the next day of ordering; receiving Sanger sequencing results from the Sanger sequencing provider the next day after sample submission; using target streptomycete strains like *S. coelicolor* WT with a relatively fast growth; and having all required reagents, including competent cells, prepared beforehand.

sgRNA design (Steps 1-7): **30 min**

Construction and validation of the desired CRISPR plasmid (Steps 8-20): **7-16 d**

Transfer of ready-to-use CRISPR plasmids into target streptomycetes by interspecies conjugation (Steps 21-33): **2 d**

Evaluation of the successfully edited strains (Step 34): **5-14 d**

Plasmid curing (optional) (Steps 35-39): **7-14 d**

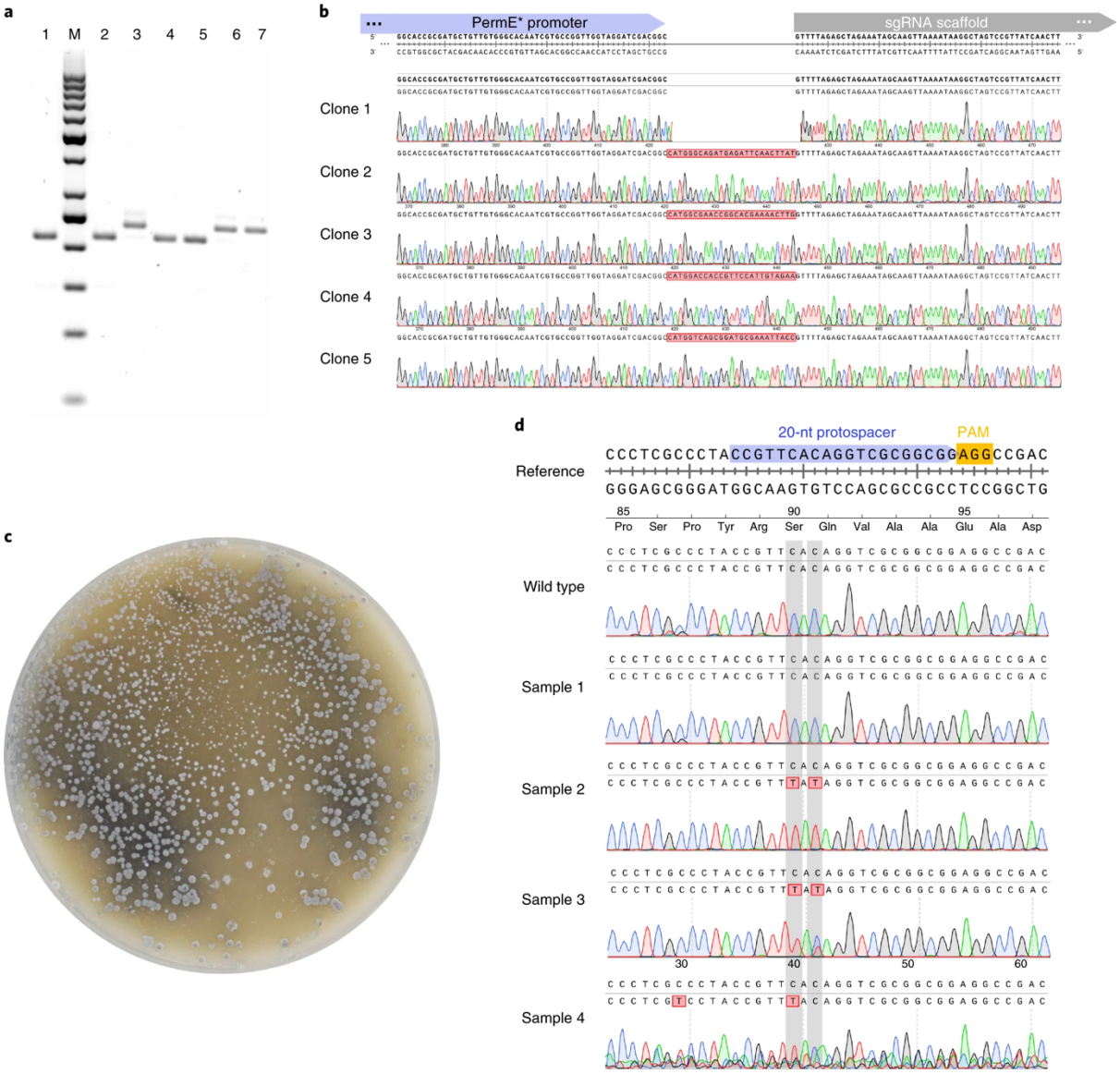


Figure 7: Anticipated results. a, Colony PCR screening of the 20-bp spacer cloning; the results are visualized by a 3% (wt/vol) agarose gel. Lane 1: negative control without the 20-bp spacer; lanes 3, 6, and 7: clones with 20-bp spacer inserted; lanes 2, 4, and 5: clones without 20-bp spacer inserted; lane M: GeneRuler 50-bp DNA Ladder, 13 bands (top to bottom) represent DNA sizes of 1,000, 900, 800, 700, 600, 500, 400, 300, 250, 200, 150, 100, and 50 bp, respectively. **b**, PCR screening and confirmation by Sanger sequencing of potentially positive clones. Clone 1 is a negative control without inserted spacer, whereas Clones 2–5 represent positive clones with a 20-bp; the different 20-bp spacers are highlighted in red boxes. **c**, Photograph of a successful conjugation experiment with >50 pickable colonies. **d**, Examples of the possible Sanger sequencing output of a base- editing application targeting the SCO5087 gene of S. coelicolor using pCRISPR–cBEST. The 20-nt protospacer sequence is highlighted in blue, whereas the 3-nt PAM sequence is shown in yellow. The two target cytosines and their expected edits are highlighted in gray. The variant nucleotides (as compared to the reference sequence) are highlighted in red.

Troubleshooting

Troubleshooting can be found in Table 1.

Table 1: Troubleshooting of the protocol.

STEP	PROBLEM	POSSIBLE REASON	SOLUTIONS
1(A)	CRISPy-web reports "Invalid input file" on upload	Uploaded file is not a valid GenBank file and/or was corrupted during upload	Verify the input file is a valid GenBank file and re-try the upload
2	On the overview page, the details box lists "0 genes"	No valid CDS records found in the GenBank file	Add gene annotations using a gene finding tool such as RAST (for an online service) or Prodigal (as a local tool)
2	On the overview page, no antiSMASH clusters are listed	1. Used the wrong input file 2. antiSMASH found no clusters	1. Check you either used the antiSMASH download or uploaded an antiSMASH result file 2. Select target region using genomic coordinates or locus tags instead
4(B) and 4(C)	No CRISPR-BEST protospacers shown on gene of interest	No combination of PAM and edit window introducing an amino acid change is present on the gene of interest	Select a different gene to edit
8(B)(iii)	No bands	Forgot to add primers or DNA template	Double check the components and re-run the PCR
8(B)(iii)	Unspecific bands	Given the large primer overhangs, especially of the first sgRNA, additional optimization of the reaction may be required.	Addition of 3% DMSO might improve the results. Successful elimination of unspecific bands was further achieved by running a touchdown PCR.
8(C)(iii)	No bands	Forgot to add primers or DNA template	Double check the components and re-run the PCR
8(C)(iii)	PCR unsuccessful, strong background	A direct colony PCR was used. The strong background of the PCR templates will increase the chances of unspecific amplifications or PCR failure.	1. Reduce the amount of the lysed *Streptomyces* templates. 2. Use purified genomic DNA as templates. 3. Use a touchdown PCR. 4. Use another high GC friendly PCR kit.
13(B)(i)	Ligation not successful	Unsuccessful ligation is likely caused by insufficient purity of the fragments and/or backbone, or by incompatible overhangs.	It is highly recommended to extract fragments from an agarose gel, unless the amplification was highly specific. Furthermore, after purification, the values of A260/A280 and A260/A230 should both be >1.8. Incompatible overhangs will also result in no ligation. It is recommended to double check and/or to perform in silico cloning to verify that the

			overhangs produced by BsaI digestion are compatible.
Box3	Low or no yield of genomic DNA	1. Overloaded tip. 2. Incomplete lysis reaction.	1. Reduce the volume by half for cell harvest. 2. Double the amount of the lysozyme, and/or prolong the incubation time.
Box3	Degraded genomic DNA samples	DNA degradation can be caused by a contamination of DNase, and/or too much physical force during preparation.	Replace the potentially contaminated reagents, and reduce the use of unnecessary physical force.
33	Too few exconjugants	Check concentration of spores and/or E. coli ET cultures.	Increase the amount of both the spores and the ET cultures.
33	Unsuccessful conjugation	Compromised CRISPR plasmid in ET or in the target streptomycete strain due to the plasmid instability.	Sanger sequence the region of pSG5 replicon of the suspicious plasmid from ET strain, to confirm its integrity. A simpler solution can also be using the plasmid from another clone, re-do the transformation to ET and the following conjugation.
33	Unsuccessful conjugation	pSG5 replicon not compatible with the target streptomycete strains.	If working with non-standard streptomycetes, it is recommended to check whether the PSG5 replicon is compatible. If not, the plasmid cannot be used without replacing the P_{SG5} replicon
33	Unsuccessful conjugation	Cross-contamination of E. coli ET12567 pUZ8002 with another E. coli strain.	Double check that the E. coli strain used for conjugations is E. coli ET12567 pUZ8002. Always use all three antibiotics kanamycin, chloramphenicol and apramycin.
33	Unsuccessful conjugation	Wrong media composition or preparation	Correct media preparation is crucial for successful conjugations. For specific non-standard streptomycetes we observed greatly reduced conjugation efficiencies when full fat soy flour was used. Furthermore, adding $MgCl_2$ before autoclaving can greatly reduce the number of sporulating exconjugants
34(A)(iv)	No or only little CRISPRi effects are observed	Non-functional and/or incorrect protospacers have been used.	Double check the protospacer selection, please refer to the sgRNA design section for more information. As CRISPRi is a knockdown effect, we recommend testing at least 3 protospacers from different locations.

			Instead of using the yield of the end point products as the evaluation, try to use qRT-PCR, or RNA-Seq to evaluate the suppression of the transcription.
34(A)(iv)	No or only little CRISPRi effects are observed	tipA promoter does not work or work with low efficiency in the strain-of-interest.	Replace the tipA promoter with an approved functional promoter.
34(A)(iv)	No or only little CRISPRi effects are observed	Inappropriate time for the analysis of the end point products.	As CRISPRi can only suppress the transcription, the end point products are still accumulating during time, if this time period is too long, a saturation effect of the end point products might mask the production differences between mutant and the wildtype. We recommend picking a good time for analysis the end point products or use qRT-PCR, or RNA-Seq to evaluate the suppression of the transcription.
34(B)(iv)	Mixed sequencing signal (Fig. 6d)	Multiple chromosomes present in the cell, mixed and only partially edited population	Re-streaking the exconjugant to obtain a single colony, followed by resequencing is recommended. For reads with mixed signals with < 20% of the unwanted signal, in most cases the edited phenotype is persistent. To be sure, re-streaking and resequencing is recommended.
33(B)(iv)	Low editing efficiency	A protospacer with low editing efficiency was used. This could, for example, be caused by sequence properties that allow forming of alternative RNA structure rather than the correct one.	1. Use an RNA structure prediction software to see if the used spacer is not good. 2. In some streptomycetes induction might be required. Re-streaking exconjugants on ISP2 plates supplemented with apramycin and thiostrepton and resequencing is recommended. 3. A simpler solution can be just repeat the experiment with another re-picked protospacer.
34(B)(iv)	Low editing efficiency	tipA promoter does not work or work with low efficiency in the strain-of-interest.	Replace the tipA promoter with an approved functional promoter.
34(B)(iv)	Low editing efficiency	For CRISPR-BEST, it could be the target C or target A is in an uncomfortable sequence context. For example, the target C is in a GC context.	Pick another protospacer.
34(C)(viii)	Unexpected large number of off-target effects is	The reference genome does not accurately reflect the	We suggest using the parental data set with a polishing tool such as Pilon[50] to change the database

	observed between the edited strain data set and the database reference genome	parental strain used for genome editing	reference genome to better reflect the parental strain. Then rerun breseq on both the parental and the edited strains.
34(C)(viii)	Unexpected large or important parts of the reference genome are not covered by the illumina data	A failure is likely happened during illumina sequencing process.	We highly recommend re-purifying the genomic DNA and rebuild the illumina library using fewer PCR cycles, as increased PCR cycles decreases the fragment diversity, particularly in high GC regions.

Anticipated results

sgRNA cloning efficiency

The ssDNA bridging method for 20 bp protospacer cloning, generally results in cloning efficiencies of > 50% (highly dependents on the sequence of the protospacers). The first step of validation could be colony PCR and the results can be visualized by 3-4% agarose gel (Fig. 6a). The obtained positive clones can then be confirmed by Sanger sequencing (Fig. 6b) as the second step of validation.

Conjugation efficiency

Using the optimized conjugation protocols described here, we are expecting to obtain > 200 exconjugants per 100 ng start plasmid DNA material on one MS plate (Fig. 6c).

A random sized deletion library construction with custom size using plasmid pCRISPR-Cas9

In order to achieve this application, the target streptomycete strain requires having a defective NHEJ pathway, i. e., the ligase component is not fully functional[9]. As the essential gene compromised clones cannot form visible colonies, all colonies from the selection plates are either edited or plasmids carrying but non-edited. Considering the high editing efficiency that we observed in *S. coelicolor* A3(2)[9], we are expecting to see > 80% of the picked colonies bearing random sized deletion of non-essential genetic regions (ref.[9]). The library size can be controlled by the amount of the start plasmid. The validation of the library requires whole genome sequencing.

Loss-of-function mutation

Within the toolkit described here, three plasmids can be used to inactivate a gene. i), When using the pCRISPR-Cas9-ScaligD, one can expect mutants bearing small indels around the DSB sites, the mutations can be validated by normal Sanger sequencing (80% editing efficiency is expected, ref.[9]); ii), When

using the pCRISPR-Cas9 with editing templates, one can expect a precise in-frame deletion mutant (with a well-designed editing template, an in-frame insertion mutant can be achieved), which also can be validated by normal Sanger sequencing (95% editing efficiency is expected, ref.[9]); iii), When using the pCRISPR-cBEST, one can expect mutants with stop codon introduction, again, the mutations can be validated by normal Sanger sequencing (90% editing efficiency is expected, ref.[15]). By visualizing the sequencing trace files in SnapGene or CLC Main Workbench, one can identify if the editing took place, bad sequencing traces due to PCR and/or sequencing processes, mixed population of edited and un-edited cells, and the expected clean editing (Fig. 6d).

Gene transcriptional modulation with CRISPRi (pCRISPR-dCas9)

By using the pCRISPR-dCas9, one can expect a decreased transcription of the gene-of-interest (ref.[9]).

Data availability

No new data were generated or analyzed with this protocol, the data used were from these two publications[9,15].

Accessions

The plasmids described in the protocols are available at Addgene (pCRISPR-Cas9, 125686; pCRISPR-dCas9, 125687; pCRISPR-Cas9-ScaligD, 125688; pCRISPR-cBEST, 125689; and pCRISPR-aBEST, 131464) or on request.

Author contributions statement

Y.T. designed and developed the protocol; K.B. designed the spacer identification software CRISPy-web; T.S.J. established the genome-wide off-targets evaluation pipeline. Y.T., and C.M.W. performed the experiments; T.W. and S.Y.L. supervised and steered the project; and Y.T., T.W., and S.Y.L. wrote the protocol.

Acknowledgments

We thank Simon Shaw for proofreading the manuscript. This work was supported by grants from the Novo Nordisk Foundation (NNF10CC1016517, NNF15OC0016226, NNF16OC0021746). S.Y.L. was also supported by the Technology Development Program to Solve Climate Changes on Systems Metabolic Engineering for Biorefineries (NRF-2012M1A2A2026556 and NRF-2012M1A2A2026557) from the Ministry of Science and ICT through the National Research Foundation (NRF) of Korea.

Competing interests

Y.T., T.W., and S.Y.L. are co-inventors on a patent application on actinomycete CRIRSPR application (WO2016150855) filed by Technical University of Denmark.

REFERENCES

1 Watve, M. G., Tickoo, R., Jog, M. M. & Bhole, B. D. How many antibiotics are produced by the genus *Streptomyces*? *Arch. Microbiol.* 176, 386-390, doi: 10.1007/s002030100345 (2001).

2 Berdy, J. Bioactive microbial metabolites. *J. Antibiot. (Tokyo)* 58, 1-26, doi:10.1038/ja.2005.1 (2005).

3 Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* 79, 629-661, doi:10.1021/acs.jnatprod.5b01055 (2016).

4 Blin, K. *et al.* antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* 45, W36-W41, doi:10.1093/nar/gkx319 (2017).

5 Weber, T. *et al.* Metabolic engineering of antibiotic factories: new tools for antibiotic production in actinomycetes. *Trends Biotechnol.* 33, 15-26, doi:10.1016/j.tibtech.2014.10.009 (2015).

6 Hwang, K. S., Kim, H. U., Charusanti, P., Palsson, B. O. & Lee, S. Y. Systems biology and biotechnology of *Streptomyces* species for the production of secondary metabolites. *Biotechnol. Adv.* 32, 255-268, doi:10.1016/J.Biotechadv.2013.10.008 (2014).

7 Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32, 347-355, doi:10.1038/nbt.2842 (2014).

8 Wang, H., La Russa, M. & Qi, L. S. CRISPR/Cas9 in genome editing and beyond. *Annu. Rev. Biochem.* 85, 227-264, doi:10.1146/annurev-biochem-060815-014607 (2016).

9 Tong, Y., Charusanti, P., Zhang, L., Weber, T. & Lee, S. Y. CRISPR-Cas9 based engineering of Actinomycetal genomes. *ACS Synth. Biol.* 4, 1020-1029, doi:10.1021/acssynbio.5b00038 (2015).

10 Cobb, R. E., Wang, Y. J. & Zhao, H. M. High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth. Biol.* 4, 723-728, doi:10.1021/sb500351f (2015).

11 Huang, H., Zheng, G. S., Jiang, W. H., Hu, H. F. & Lu, Y. H. One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta Biochim. Biophys. Sin.* 47, doi:10.1093/abbs/gmv007 (2015).

12 Zeng, H. *et al.* Highly efficient editing of the actinorhodin polyketide chain length factor gene in *Streptomyces coelicolor* M145 using

CRISPR/Cas9-CodA(sm) combined system. *Appl. Microbiol. Biotechnol.* 99, 10575-10585, doi:10.1007/s00253-015-6931-4 (2015).

13 Tong, Y., Weber, T. & Lee, S. Y. CRISPR/Cas-based genome engineering in natural product discovery. *Nat. Prod. Rep.* 36, 1262-1280, doi:10.1039/c8np00089a (2019).

14 Alberti, F. & Corre, C. Editing streptomycete genomes in the CRISPR/Cas9 age. *Nat. Prod. Rep.* 36, 1237-1248, doi:10.1039/c8np00081f (2019).

15 Tong, Y. *et al.* Highly efficient DSB-free base editing for streptomycetes with CRISPR-BEST. *Proc. Natl. Acad. Sci. USA* 116, 20366-20375, doi:10.1073/pnas.1913493116 (2019).

16 Tong, Y., Robertsen, H. L., Blin, K., Weber, T. & Lee, S. Y. CRISPR-Cas9 toolkit for actinomycete genome editing. *Methods Mol. Biol.* 1671, 163-184, doi:10.1007/978-1-4939-7295-1_11 (2018).

17 Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J. A. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* 329, 1355-1358, doi:10.1126/science.1192272 (2010).

18 Blin, K., Pedersen, LE., Weber, T., Lee, S. Y. CRISPy-web: An online resource to design sgRNAs for CRISPR applications. *Synth. Syst. Biotechnol.* 1, 4, doi:10.1016/j.synbio.2016.01.003 (2016).

19 Bibb, M. J., Ward, J. M. & Hopwood, D. A. Transformation of plasmid DNA into *Streptomyces* at high frequency. *Nature* 274, 398-400 (1978).

20 Kieser, T., Bibb, M., Buttner, M., Chater, K. & Hopwood, D. Practical *Streptomyces* genetics. *Norwich: The John Innes Foundation* (2000).

21 Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. USA* 100, 1541-1546, doi:10.1073/pnas.0337542100 (2003).

22 Fernandez-Martinez, L. T. & Bibb, M. J. Use of the meganuclease I-SceI of *Saccharomyces cerevisiae* to select for gene deletions in actinomycetes. *Sci. Rep.* 4, 7100, doi:10.1038/srep07100 (2014).

23 Volff, J. N. & Altenbuchner, J. Genetic instability of the *Streptomyces* chromosome. *Mol. Microbiol.* 27, 239-246, doi:10.1046/j.1365-2958.1998.00652.x (1998).

24 Hoff, G., Bertrand, C., Piotrowski, E., Thibessard, A. & Leblond, P. Genome plasticity is governed by double strand break DNA repair in *Streptomyces*. *Sci. Rep.* 8, 5272, doi:10.1038/s41598-018-23622-w (2018).

25 Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353, aaf8729, doi:10.1126/science.aaf8729 (2016).

26 Nicole M. Gaudelli, A. C. K., Holly A. Rees, Michael S. Packer, Ahmed H. Badran, David I. Bryson, David R. Liu. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 551, 464-471, doi:10.1038/nature24644 (2017).

27 Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420-424, doi:10.1038/nature17946 (2016).

28 Lovett, S. T. Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol. Microbiol.* 52, 1243-1253, doi:10.1111/j.1365-2958.2004.04076.x (2004).

29 Jack, B. R. *et al.* Predicting the genetic stability of engineered DNA sequences with the EFM calculator. *ACS Synth. Biol.* 4, 939-943, doi:10.1021/acssynbio.5b00068 (2015).

30 Brophy, J. A. & Voigt, C. A. Principles of genetic circuit design. *Nat. Methods* 11, 508-520, doi:10.1038/nmeth.2926 (2014).

31 Alberti, F. *et al.* Triggering the expression of a silent gene cluster from genetically intractable bacteria results in scleric acid discovery. *Chem. Sci.* 10, 453-463, doi:10.1039/c8sc03814g (2019).

32 Low, Z. J. *et al.* Identification of a biosynthetic gene cluster for the polyene macrolactam sceliphrolactam in a *Streptomyces* strain isolated from mangrove sediment. *Sci. Rep.* 8, 1594, doi:10.1038/s41598-018-20018-8 (2018).

33 Culp, E. J. *et al.* Hidden antibiotics in actinomycetes can be identified by inactivation of gene clusters for common antibiotics. *Nat. Biotechnol.* 37, 1149-1154, doi:10.1038/s41587-019-0241-9 (2019).

34 Cohen, D. R. & Townsend, C. A. A dual role for a polyketide synthase in dynemicin enediyne and anthraquinone biosynthesis. *Nat. Chem.* 10, 231-236, doi:10.1038/nchem.2876 (2018).

35 Cho, J. S. *et al.* CRISPR/Cas9-coupled recombineering for metabolic engineering of *Corynebacterium glutamicum*. *Metab. Eng.* 42, 157-167, doi:10.1016/j.ymben.2017.06.010 (2017).

36 Mo, J. *et al.* Efficient editing DNA regions with high sequence identity in actinomycetal genomes by a CRISPR-Cas9 system. *Synth. Syst. Biotechnol.* 4, 86-91, doi:10.1016/j.synbio.2019.02.004 (2019).

37 Wang, Q. *et al.* Dual-function chromogenic screening-based CRISPR/Cas9 genome editing system for actinomycetes. *Appl. Microbiol. Biotechnol.* 104, 225-239, doi:10.1007/s00253-019-10223-4 (2020).

38 Zhang, M. M. *et al.* CRISPR-Cas9 strategy for activation of silent *Streptomyces* biosynthetic gene clusters. *Nat. Chem. Biol.* 13, 607-609 doi:10.1038/nchembio.2341 (2017).

39 Dow, L. E. *et al.* Inducible in vivo genome editing with CRISPR-Cas9. *Nat. Biotechnol.* 33, 390-394, doi:10.1038/nbt.3155 (2015).

40 Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* 32, 670-676, doi:10.1038/nbt.2889 (2014).

41 Bu, Q. T. *et al.* Rational construction of genome-reduced and high-efficient industrial *Streptomyces* chassis based on multiple comparative genomic approaches. *Microb. Cell Fact.* 18, 16, doi:10.1186/s12934-019-1055-7 (2019).

42 Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149-157, doi:10.1038/s41586-019-1711-4 (2019).

43 Muth, G. The pSG5-based thermosensitive vector family for genome editing and gene expression in actinomycetes. *Appl. Microbiol. Biotechnol.* 102, 9067-9080, doi:10.1007/s00253-018-9334-5 (2018).

44 MacNeil, D. J. *et al.* Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* 111, 61-68, doi:10.1016/0378-1119(92)90603-m (1992).

45 Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173-1183, doi: 10.1016/J.Cell.2013.02.022 (2013).

46 Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417, 141-147, doi:10.1038/417141a (2002).

47 Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* 9, 88, doi:10.1186/s13104-016-1900-2 (2016).

48 Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol. Biol.* 1151, 165-188, doi:10.1007/978-1-4939-0554-6_12 (2014).

49 Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 47, W81-W87, doi:10.1093/nar/gkz310 (2019).

50 Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9, e112963, doi:10.1371/journal.pone.0112963 (2014).

51 Bai, C. *et al.* Exploiting a precise design of universal synthetic modular regulatory elements to unlock the microbial natural products in *Streptomyces*. *Proc. Natl. Acad. Sci. USA* 112, 12181-12186, doi:10.1073/pnas.1511027112 (2015).

Chapter 4

Systems Analysis of Highly Multiplexed CRISPR-Base Editing in Streptomycetes

Christopher M. Whitford[1], Tetiana Gren[1], Emilia Palazzotto[1,4], Sang Yup Lee[3], Yaojun Tong[2] & Tilmann Weber[1]

[1] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark.
[2] State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic and Developmental Sciences, School of Life Sciences and Technology, Shanghai Jiao Tong University, Shanghai 200240, China.
[3] Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering (BK21 Plus Program), Center for Systems and Synthetic Biotechnology, Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Republic of Korea.
[4] Present address: Department of Health, Promotion, Mother and Child Care, Internal Medicine and Medical Specialties, University of Palermo, 90133 Palermo PA, Italy.

To whom correspondence should be addressed:
yaojun.tong@sjtu.edu.cn or tiwe@biosustain.dtu.dk

This manuscript is currently under review at ACS Synthetic Biology

ABSTRACT

CRISPR tools, especially Cas9n-sgRNA guided cytidine deaminase base editors such as CRISPR-BEST, have dramatically simplified genetic manipulation of streptomycetes. One major advantage of CRISPR base editing technology is the possibility to multiplex experiments in genomically instable species. Here, we demonstrate scaled up Csy4 based multiplexed genome editing using CRISPR-mcBEST in *Streptomyces coelicolor*. We evaluated the system by simultaneously targeting 9, 18, and finally all 28 predicted specialized metabolite biosynthetic gene clusters in a single experiment. We present important insights into the performance of Csy4 based multiplexed genome editing at different scales. Using multi-omics analysis, we investigated the systems wide effects of such extensive editing experiments and revealed great potentials and important bottlenecks of CRISPR-mcBEST. The presented analysis provides crucial data and insights towards the development of multiplexed base editing as a novel paradigm for high throughput engineering of *Streptomyces* chassis and beyond.

INTRODUCTION

Actinomycetes, especially members of the *Streptomyces* genus, are among the most valuable sources of natural products with industrial and pharmaceutical importance. Advances in genome sequencing and genome mining revealed a vast majority of yet to be characterized biosynthetic gene clusters (BGCs)[1]. However, the activation of BGCs in laboratory conditions and the detection of the associated products remains challenging. The high GC content of actinomycete genomes, complex regulatory mechanisms controlling the expression of BGCs, poor genetic tractability, and limited availability of molecular tools greatly limit the exploration of the biosynthetic potential of actinomycetes.

To overcome this problem many research groups have turned to the expression of target BGCs in heterologous hosts (microbial cell factories) [2]. Several expression hosts were constructed with a simplified metabolic background and an increased productivity of the compounds encoded on the heterologous BGCs through deletion of BGCs from the host genome [3–5]. However, successful construction of genome minimized heterologous hosts using classical techniques remains a very labor-intensive process, usually spanning up to several years. With "classical" protocols, engineering just one genomic locus and obtaining a clean mutant can easily take over a month[6]. Typically, one iteration includes plasmid cloning, transformation and conjugation, screening, plasmid curing, sequencing, and data analysis [7]. In slow growing actinomycetes, this process usually results in substantial strain development times, easily reaching multiple years [8].

Such long development times are the result of iterative strain engineering cycles, in which only one target is edited per cycle. Furthermore, repeated plasmid curing might result in unwanted stress-induced effects, such as genome rearrangements or cyclization, resulting from the high genome plasticity of streptomycetes [9]. Tools for multiplexed genetic manipulations hence greatly increase the possibilities and throughput for advanced strain engineering in actinomycetes.

CRISPR-based genetic manipulation has become increasingly popular for streptomycete genetic engineering. Such tools include CRISPR-Cas9, CRISPR-

Cpf1, CRISPR interference (CRISPRi), and base editing systems such as CRISPR-BEST [6,10–13]. Besides high efficiencies, a major advantage of many CRISPR engineering systems lies in its intrinsic multiplexing compatibility. Multiplexing allows simultaneous introduction of multiple sgRNAs, greatly reducing the time needed for engineering of multiple genomic loci [14]. The native structure of the CRISPR array can be utilized for multiplexing based on separate expression of the crRNAs and tracrRNA and processing by RNAse III [15]. In addition to tracrRNA/RNAse III processing, expression of sgRNAs from individual, monocistronic expression cassettes, sgRNA array processing by ribozyme self-cleavage, RNase mediated processing of sgRNAs separated by tRNAs, Cas12a/Cas13a processing, and Csy4 mediated processing of sgRNAs can be used [14]. However, using individual promoter-sgRNA cassettes on a single plasmid complicates scale-up of the number of used sgRNAs. Furthermore, a high density of transcriptional units on one plasmid might result in promoter crosstalk effects, possibly negatively affecting[16] the expression of the sgRNAs. For tools intended for applications across multiple related species, reducing the dependencies on the hosts machinery for gRNA processing is beneficial, making system for multiplexed sgRNA delivery the approach of choice.

Csy4 based processing of sgRNA arrays has been demonstrated in a variety of organisms with promising results [17–20]. The type I-F CRISPR-associated endoribonuclease, also referred to as Cas6, recognizes a distinct 28 nt RNA sequence, which forms a stable hairpin structure, and cleaves it at a CG between positions 20 and 21 [21]. In a sgRNA array, Csy4 sites will release individual mature sgRNAs. We previously designed and demonstrated the feasibility of Csy4 based multiplexed cytosine base editing based on our CRISPR-BEST system (pCRISPR-mcBEST) in *Streptomyces coelicolor* and *Streptomyces griseofuscus* with three targets [5,12].

Here, we investigated the possibility to further scale up Csy4-based multiplexed engineering of streptomycetes by targeting biosynthetic genes of 9, 18 and 28 BGCs in *S. coelicolor* from a single cytosine base editing plasmid. We performed extensive analysis of the resulting mutants, including whole genome sequencing (WGS), growth profiling, untargeted proteomics, transcriptomics, and exometabolomics, to gain further insights into the potential, limitations, and bottlenecks of deep base editing mediated genome

perturbations in streptomycetes. We provide important insights into the performance of multiplexed base editing, including scalability and the relationship between the number of sgRNAs and SNPs. Our results highlight how multiplexed engineering can substantially speed up the Design-Build-Test-Learn cycle of metabolic engineering of streptomycetes and pave the way for high throughput engineering strategies of streptomycete chassis.

RESULTS

Development of a scalable Csy4-based multiplexed base editing system

Implementation of plasmid improvements and a new sgRNA array cloning strategy

Building on top of our previously published multiplexed base editing plasmid pCRISPR-mcBEST, we introduced several improvements. In the improved version of the plasmid, the SP19 promoter was exchanged with a stronger promoter *kasO*P*, to increase the transcription level of the sgRNA array. In the same step, *mCherry* was replaced with *sfGFP*, as SP19-BBa_B0034-*mCherry* did not produce a reliable signal in *Escherichia coli*, making screening for correct colonies difficult. The multiplexing systems utilizes the previously tested Csy4/Cas6 for array processing. We chose Csy4 as it allows the plasmid-based delivery of all necessary components for sgRNA array processing and reduces dependencies on the hosts processing machineries. It further allows easy cloning and integration of many sgRNAs, as opposed to individual sgRNA cassettes. To make this cloning step even easier, we further optimized the cloning procedure. Preassembly in pGGA using the established Golden Gate method[22] results not only in higher cloning efficiencies, but also allows easier reuse and modification of sgRNA arrays. The arrays are cut from the pGGA plasmid and ligated into the predigested pCRISPR-mcBEST destination vector.

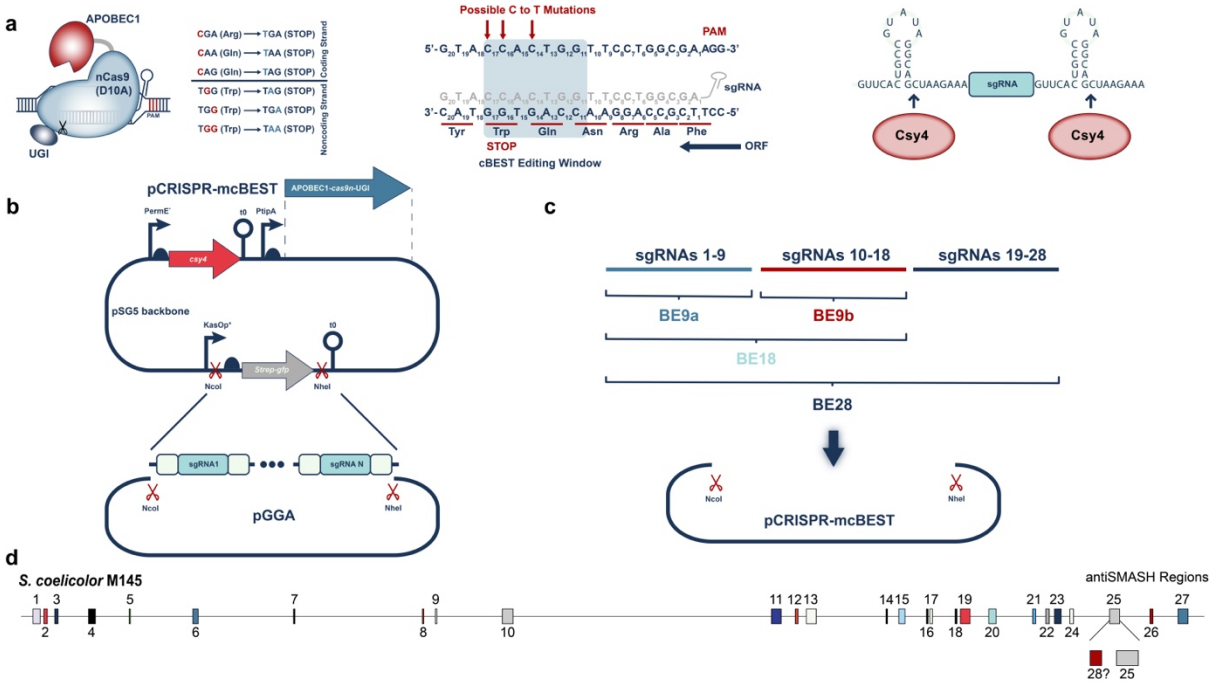


Figure 1: Principles of (multiplexed) cytosine base editing and experimental design. a. The base editor consists of codon optimized cytosine deaminase (APOBEC1) and an uracil glycosylation inhibitor (UGI) fused to a nicking Cas9 (D10A). Arg, Gln, or Trp codons (on the noncoding strand) can be edited into stop codons leading to truncated and thus non-functional proteins. Csy4 cleaves its recognition site downstream of the hairpin structure between positions 20 and 21. Flanking sgRNAs with Csy4 recognition sites enables multiplexing from a single transcriptional unit. **b.** The multiplexed pCRISPR-cBEST platform. sgRNA arrays are preassembled in pGGA using Golden Gate Assembly and cloned into the destination plasmid. **c.** Overview of the different sgRNA arrays constructed in this study. Arrays with 9, 18, and 28 sgRNAs targeting biosynthetic genes located in the 28 BGCs of *S. coelicolor* (shown in **d**) were constructed. BGCs were predicted by antiSMASH 6. The hypothesized additional BGC in region 25 was given the number 28.

Multiplexed Base Editing enables simultaneous engineering of an unprecedented number of genomic loci in streptomycetes

To investigate the scalability of Csy4 based multiplexed cytosine base editing beyond three targets, we attempted to simultaneously inactivate up to all 28 predicted biosynthetic gene clusters in *S. coelicolor* M145 through the introduction of premature stop codons. The *S. coelicolor* A3(2) genome (Accession: PRJNA557658) was analyzed using antiSMASH 6 [23] and CRISPy-web. antiSMASH 6 predicts 27 biosynthetic regions. Next to the germicidin BGC in region 25, there is a second, hypothetical BGC which we assigned the

number 28. 28 cBEST compatible protospacer sequences targeting key biosynthetic genes of the respective clusters were designed using CRISPy-web (see Supplementary Table 1) [24]. Protospacers were designed to be as close as possible to the translation start sites of the core biosynthetic genes, while displaying the lowest possible number of off-target sites. Where possible, unfavorable sequence contexts were avoided [12]. The selected protospacers were used for design of the synthetic Golden Gate based array construction (Supplementary Fig. 1). Four different arrays were constructed, one with sgRNAs 1-9, one with sgRNAs 10-18, one with the first 18, and one with all 28 sgRNAs. The 18 and 28 sgRNA arrays were built from smaller subarrays and cloned into pCRISPR-mcBEST using restriction and ligation cloning (as shown in Supplementary Fig. 1). To capture also editing events with low efficiencies, we initially screened and sequenced plasmid containing colonies. The sequenced colonies represent complex populations with mixed genotypes, which allowed us to obtain detailed information about the distribution of editing efficiencies across different target sites.

9 sgRNAs facilitate robust editing of 9 genomic loci

To test editing of up to 9 genomic targets, two strains were constructed using pCRISPR-mcBEST and sgRNAs 1-8 or sgRNAs 9-18, and referred to as *S. coelicolor* BE9a and BE9b, respectively. At least four sgRNAs target key biosynthetic genes involved in the production of specialized metabolites that are visible with the naked eye: sgRNA 8 (FQ762_13840, brownish pigment melanin), sgRNA 11 (FQ762_26280, dark blue actinorhodin), sgRNA 13 (FQ762_27465, spore pigment), and sgRNA 15 (FQ762_30470, bright red undecyprodigiosin). Based on the presence of targets that should result in a clear phenotypical change if successfully edited, three promising colonies were randomly selected of each strain for WGS and analyzed using breseq [25] for analysis of on- and off-target editing. For *S. coelicolor* BE9b (Fig. 2), editing events were observed for 7 out of 9, 8 out of 9, and 7 out of 9 targets, respectively. Often multiple editable cytosines were located within the respective editing windows, resulting in up to four editing events per target site. For colonies 3, 7, and 13, 3 out of 12, 6 out of 14, and 5 out of 10 of the detected mutations resulted in the introduction of premature stop codons.

Genome-wide SNP analysis revealed that the overwhelming majority of detected SNPs were characteristic C to T or G to A conversions, as expected for cytosine deaminase base editing. Most other mutations were also detected in the wildtype controls, highlighting that these are most likely the result of the normal background mutation rate (Supplementary Fig. 3).

Similar results were obtained for *S. coelicolor* BE9a as shown in Supplementary Fig. 2.

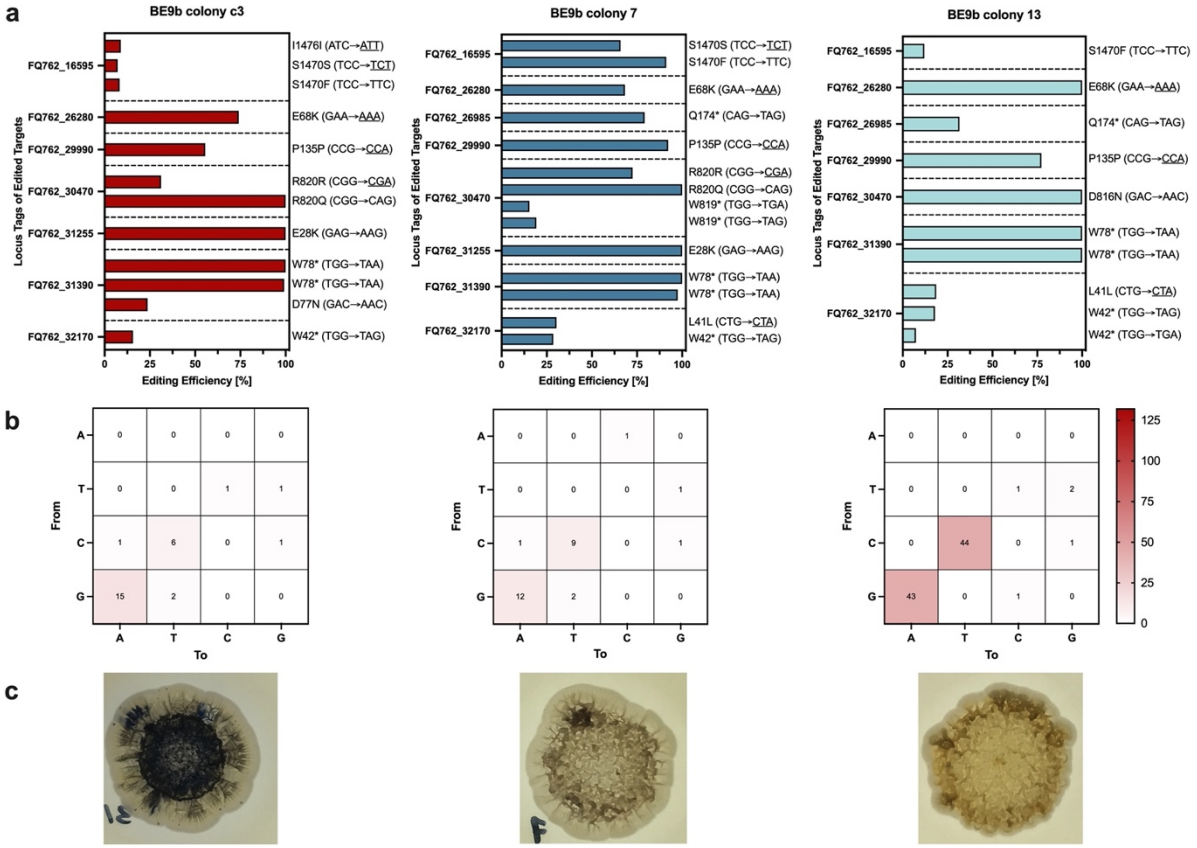


Figure 2: Editing outcomes for selected *S. coelicolor* BE9b colonies. **a**. 7 out of 9, 8 out of 9, and 7 out of 9 targeted loci were edited, resulting in 3, 6, and 5 premature stop codons for colonies 3, 7, and 13, respectively. Stop codons are indicated by *, and rare codons are underlined. **b**. Genome wide SNP analysis. Almost all SNPs were characteristic C to T, or G to A conversions, indicating cytosine deaminase activity as the primary driver of SNP introduction. **c**. Phenotypes of *S. coelicolor* BE9b mutants grown on ISP2 solid medium. Data for biological replicates is presented individually to better represent the large differences in observed outcomes.

18 sgRNAs facilitate robust editing despite minor plasmid instabilities

Strain *S. coelicolor* BE18 was constructed using plasmid p091, harboring a synthetic sgRNA array with sgRNAs 1-18. Editing of at least four of the targeted loci should result in clear phenotypical changes if edited successfully, as previously described. Colonies 1, 3, and 4 were selected for further analysis based on their phenotypical appearances. WGS and bioinformatic analysis revealed that in 11, 12, and 12 out of the 18 targeted genomic loci editing events were observed, respectively (Fig. 3). For strain BE18 colony 1, a total of 17 mutations were observed, out of which 5 resulted in successful introduction of premature stop codons. In strain BE18 colony 3, 20 different mutation events were detected, resulting in 8 premature stop codons. Finally, in strain BE18 colony 4, 6 stop codons were introduced, based on a total of 19 detected mutations within the target loci.

Analysis of the sequencing coverage of plasmid p091 resulted in the identification of a small deletion in the array in sgRNA10 and the downstream Csy4 recognition site. The functionality of sgRNA10, which targets the non-ribosomal peptide synthetase FQ762_16595, was therefore abolished. Accordingly, no mutations were detected at the targeted loci in FQ762_16595. The incomplete Csy4 recognition site upstream of sgRNA11 prevented the recognition of the stem loop and cleavage of the array at that position[21]. For the resulting prolonged sgRNA 11, editing events were only detected in BE18 colony 3 at low efficiencies.

The plasmid coverage extracted using samtools[26] varied noticeably between the three analyzed biological replicates. However, this seems to have had limited influence on the editing outcomes in the analyzed samples, indicating that even low copy numbers and low expression levels of the editing machinery can suffice for efficient editing. Genome wide SNP analysis again revealed dominant C to T or G to A mutations and can thus likely be attributed to the base editor. Interestingly, the number of SNPs accumulated varied noticeably between the three samples, without a great difference in the number of mutations in the targeted loci. This is especially visible when directly comparing BE18 colony 3 and colony 4.

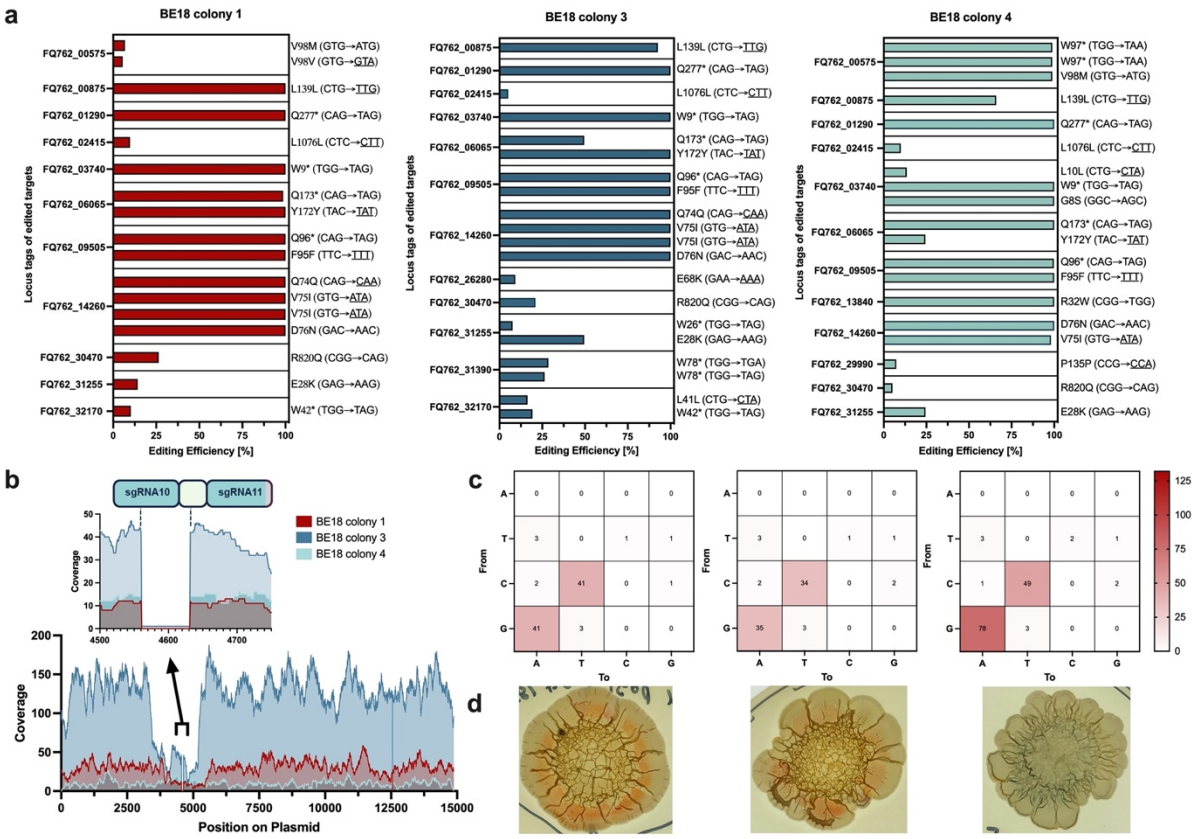


Figure 3: Editing outcomes for three sequenced colonies of *S. coelicolor* BE18. a. 11, 12, and 12 out of 18 target loci were edited in colonies 1, 3, and 4, respectively. Furthermore, 5, 8, and 6 of the resulting mutations introduced premature stop codons. Stop codons are indicated by *, and rare codons are underlined. **b.** Illumina sequencing coverage of plasmid p091. Within the sgRNA array, a small deletion was observed in all three colonies (highlighted region). **c.** Genome-wide SNP analysis. The overwhelming majority of the detected SNPs can be attributed to the cytosine base editor due to the introduction of C to T or G to A mutations (on the complementary strand). **d.** Phenotypes of the three sequenced colonies on ISP2 medium. All colonies were pale and almost colorless, except for some light red hues for colonies 1 and 3, and a light blue hue for colony 4. No sporulation was observed on ISP2 medium.

Limitations of Csy4 based multiplexing become apparent upon scaling up to 28 sgRNAs

To identify the upper limit of the number of targets that can be edited simultaneously, we attempted to introduce premature stop codons in all BGCs of *S. coelicolor*. Therefore, *S. coelicolor* M145 was conjugated with plasmid p085, containing the full 28 sgRNA array.

Editing events were detected in 10 out of the 28 targeted loci in strain BE28 colony 7. 16 different kinds of mutations were detected, including 7 premature

stop codons. The highest number of editing events was obtained for BE28 colony 8, with 17 edited loci. Of the 34 kinds of on-target mutations identified, 12 resulted in the introduction of premature stop codons. In strain BE28 colony 9, only 9 out of the 28 target loci displayed editing events, and out of the 14 identified mutations, only 4 resulted in premature stop codons. These results clearly demonstrate the scalability of Csy4 based multiplexed base editing, and to the best of our knowledge, 17 targets is the highest number of base edited genomic loci using Csy4 based multiplexing in bacteria to date, and the highest number of simultaneously edited genomic targets in actinomycetes using any technology.

Not unexpectedly, the analysis of the plasmid sequencing coverage unveiled large deletions within the synthetic sgRNA array. Unlike for BE18, the observed deletions appeared to be more heterogenous between the colonies. All appeared between sgRNAs 7 and 26. The largest deletions were observed for BE28 colonies 7 and 9, for which also lower numbers of editing events were observed. BE28 colony 8, which missed a noticeably smaller fragment of the sgRNA array, also displayed the highest number of observed editing events. This indicates that plasmid instability issues linked to the high number of repetitive sequences in the synthetic sgRNA array are one of the key issues in scale up of Csy4 based multiplexed engineering approaches in streptomycetes. Where larger sgRNA arrays are required, the use of nonrepetitive scaffold sequences could improve the sgRNA array stability and reduce the likelihood of recombination events[27].

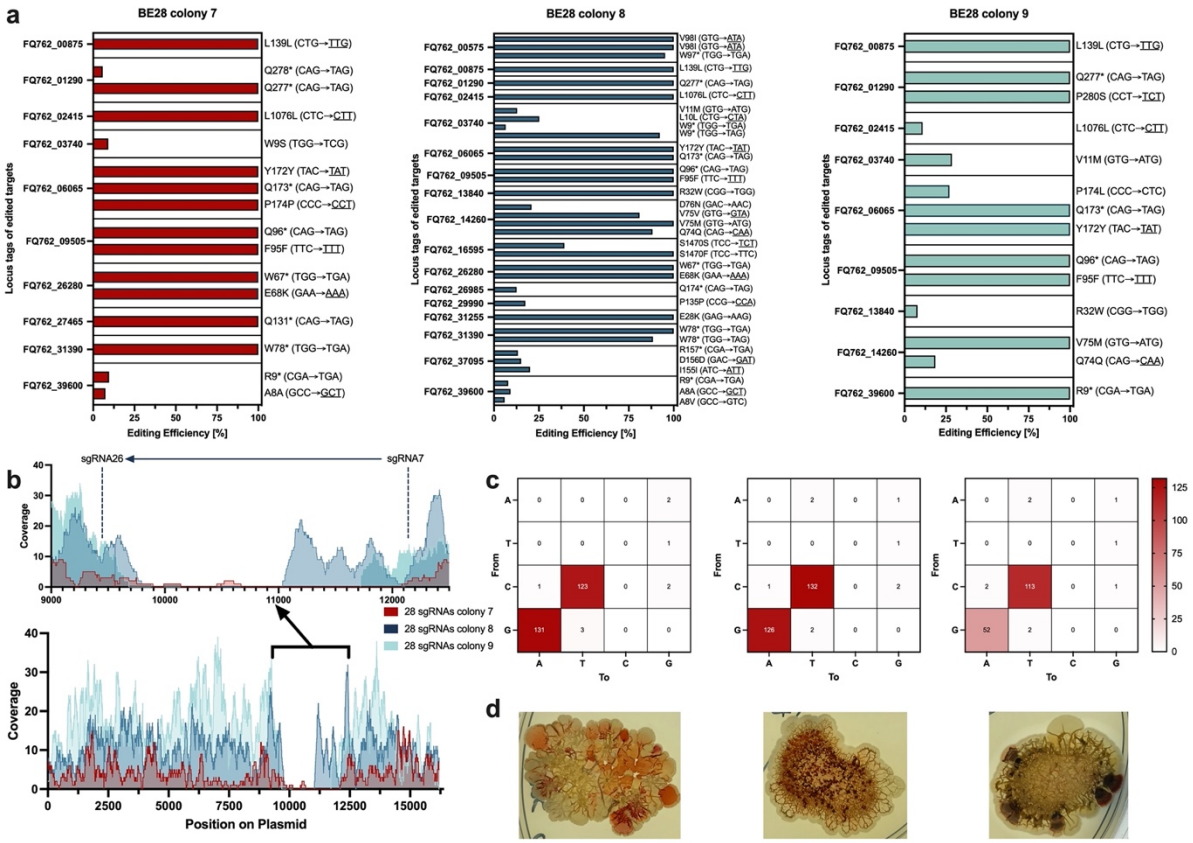


Figure 4: Editing outcomes for *S. coelicolor* BE28 using a 28 sgRNA array. **a**. Out of 28 targeted loci, 10, 17, and 9 were edited in colonies 7, 8, and 9, respectively. Out of the 16, 34, and 14 on-target mutations, 7, 12, and 4 resulted in the introduction of premature stop codons. Therefore, *S. coelicolor* BE28 colony 8 represents the best editing outcome for any of the analyzed plasmid containing strains. Stop codons are indicated by *, and rare codons are underlined. **b**. Sequencing coverage of plasmid p085. The highlight shows large deletions detected within the sgRNA array. Colony 8, which had the largest intact array, also had the highest number of edited target loci. **c**. Genome wide SNP analysis. The overwhelming majority of detected SNPs can be associated with the base editor. **d**. Phenotypes of the three sequenced colonies. All showed clear signs of mixed populations, with differently colored patches, as well as in parts unusual, wrinkled morphology. **c** and **d** from left to right: colonies 7, 8, 9.

Profiling of SNPs again revealed that most mutations can be attributed to the activity of the cytosine deaminase. In total, 254, 258, and 165 C to T mutations (including G to A) were detected for colonies 7, 8, and 9, respectively. Surprisingly, BE28 colony 9 displayed a substantially lower number of C to T mutations, even though it did not exhibit larger deletions than BE28 colony 7. Potentially, this points towards earlier recombination and loss of sgRNAs in BE28 colony 9 compared to BE28 colony 7 and 8. All three colonies showed

phenotypical differentiation, which highlights the need for plasmid curing to obtain stable genotypes (Fig. 4D).

Screening of plasmid cured strains results in isolation of heavily edited mutants

Since Csy4 mediated multiplexed cytosine base editing delivered robust editing, we next investigated plasmid cured colonies. All BE9a, BE9b, BE18, and BE28 strains were cured from the pCRISPR-mcBEST plasmids as described in the Methods and Materials section. A total of 33 plasmid free mutants were selected based on their phenotypical appearances (as shown in Supplementary Fig. 4). The wide range of different phenotypes highlights the size of the mutational landscape that can be generated by multiplexed base editing.

One colony of each BE9b, BE18, and BE28 was selected based on their phenotypes for further characterization. None of the strains produced any of the compounds normally visible during growth on agar plates (Fig. 5). While the initial samples were mixed populations, WGS of plasmid cured strains was expected to result in the identification of stable mutations. In the plasmid cured BE9b strain, only 4 out of the up to 8 target loci edited in the mixed population were edited in the selected colony. These 4 target loci were all edited with high efficiencies in the mixed population, which explains why these were easily recovered after plasmid curing.

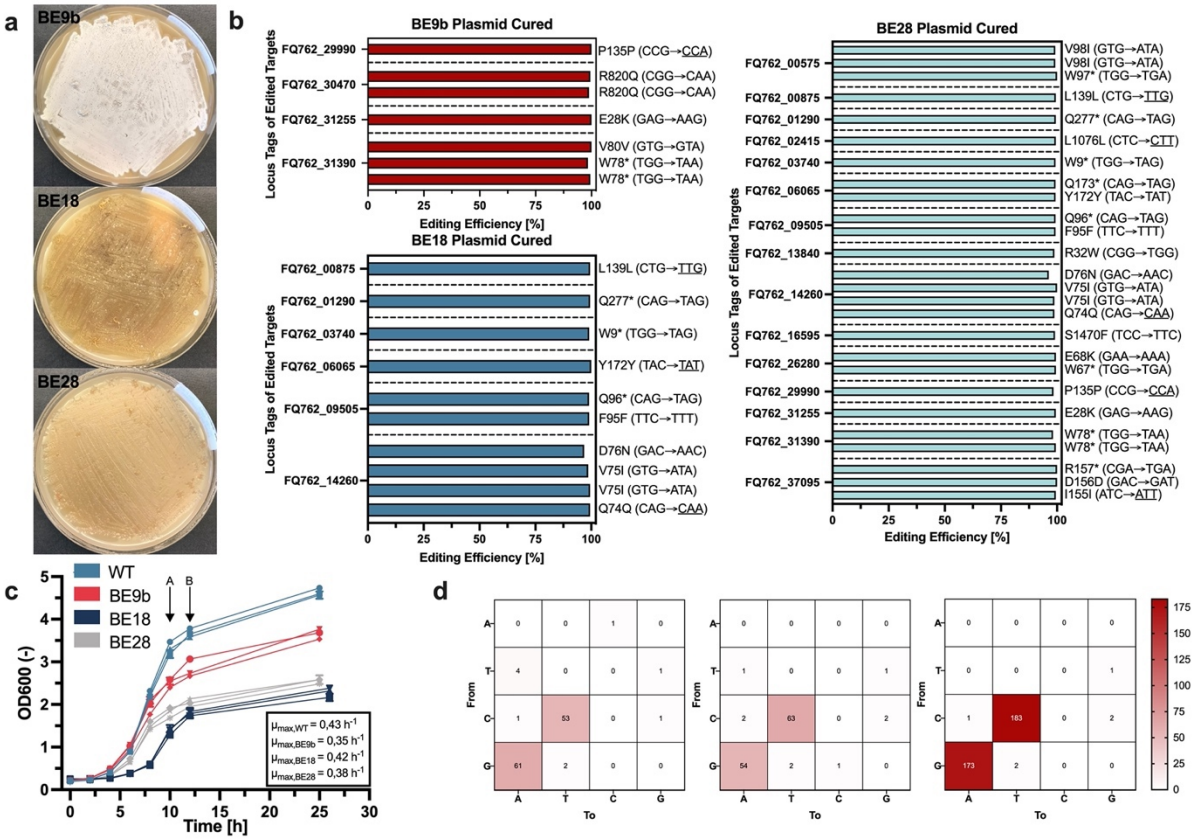


Figure 5: Analysis and characterization of plasmid cured strains. **a**. Phenotypes of the selected plasmid cured strains. Both *S. coelicolor* BE18 and BE28 did not sporulate anymore, and especially *S. coelicolor* BE28 appeared almost transparent on mannitol soy flour medium. **b**. Editing outcomes of the plasmid cured strains. Stable mutations (~100 % of reads edited) were obtained for all colonies. Stop codons are indicated by *, and rare codons were underlined. **c**. Growth curves of the plasmid cured strains and *S. coelicolor* M145. Clear differences in the maximum specific growth rate μ_{max} and the maximum OD600 were observed. Sampling time points are indicated with **A** and **B**. Data for three technical replicates of each strain, and the mean ± standard deviation of three repeated measurements is shown. **d**. Mutation heatmaps for the three plasmid cured mutants, showing a clear increase in C to T and A to G mutations for BE28.

Six target loci were edited in the plasmid cured BE18 strain, compared to up to 12 in the plasmid harboring strains. All detected mutations were stable, and the edited target loci detected in the plasmid cured strain were again among the target loci with the highest efficiencies in the plasmid containing strains.
In the plasmid cured BE28 strain, a total of 15 target sites were edited with stable mutations. Premature stop codons were introduced in 8 out of the 15 edited locus tags, corresponding to BGCs 1, 3, 5, 6, 7, 11, 17, and 28 (hypothetical BGC). As previously seen, all mutations observed in the plasmid cured strain were also dominant in the plasmid containing colonies,

highlighting selection challenges for edits with low efficiencies without either extensive screening or counterselection. In addition to the 8 premature stop codons, 5 rare codons (frequency <= 5 %) were introduced. Taken together, up to 13 key biosynthetic genes were inactivated through introduction of stop codons or potentially down regulated through introduction of rare codons.

We further characterized the plasmid cured strains by performing shake flask cultivations and observed noticeably different growth characteristics of the edited strains compared to the wildtype control (Fig. 5c). The wild type cultures reached a final OD600 of 4.63 ± 0.09, while the BE9b, BE18, and BE28 cultures reached final OD600s of 3.66 ± 0.12, 2.29 ± 0.11, and 2.55 ± 0.06, showing increased burden for the BE strains under standard cultivation conditions. BE18 further displayed a prolonged lag phase, lasting for 6 to 8 hours. The μ_{max} values (maximum specific growth rate) were determined as described in the methods section. For the wildtype, BE9b, BE18, and BE28, μ_{max} values of of 0.43 h^{-1}, 0.35 h^{-1}, 0.42 h^{-1}, and 0.38 h^{-1} were determined, respectively. The corresponding maximum doubling times were derived from the μ_{max} values as 1.63 h, 1.96 h, 1.65 h, and 1.86 hours for the wildtype, BE9b, BE18, and BE28, respectively.

Multiplexed base editing data from various scales provides key insights and allows identification of robust array sizes

Using the data obtained at various scales, we next aimed to identify the sweet spot for Csy4 based multiplexed base editing in streptomycetes. To this end, analyzed the SNP distributions and editing outcomes to gain deeper understanding of potential bottlenecks or optimal engineering approaches. A clear increase in the number of SNPs and amino acid (AA) changes was observed with an increasing number of sgRNAs (Fig. 6a). While there are some obvious outliers (colonies BE9a c3, BE9b c13, BE28 c9), the overall trend remains robust. The number of introduced AA changes is consistently around half of the SNPs, which is expected due to the redundant nature of the genetic code and off-targets outside of coding sequences. Plasmid cured strains carried a noticeably higher number of mutations compared to the plasmid harbouring strains, highlighting how plasmid curing and additional incubation steps can result in additional mutational load. Plotting of the number of SNPs against the number of sgRNAs further allowed determination of the approximate scaling

factor. The slope of the linear regression was calculated as 7.959 ± 1.042 SNPs/sgRNA, highlighting how the number of SNPs and AA changes scales linearly with the number of multiplexed sgRNAs. This further indicates a low frequency of Cas-independent off-target effects by the cytosine deaminase. We next looked at the number of edited target sites in the plasmid carrying BE strains (Fig. 6b).

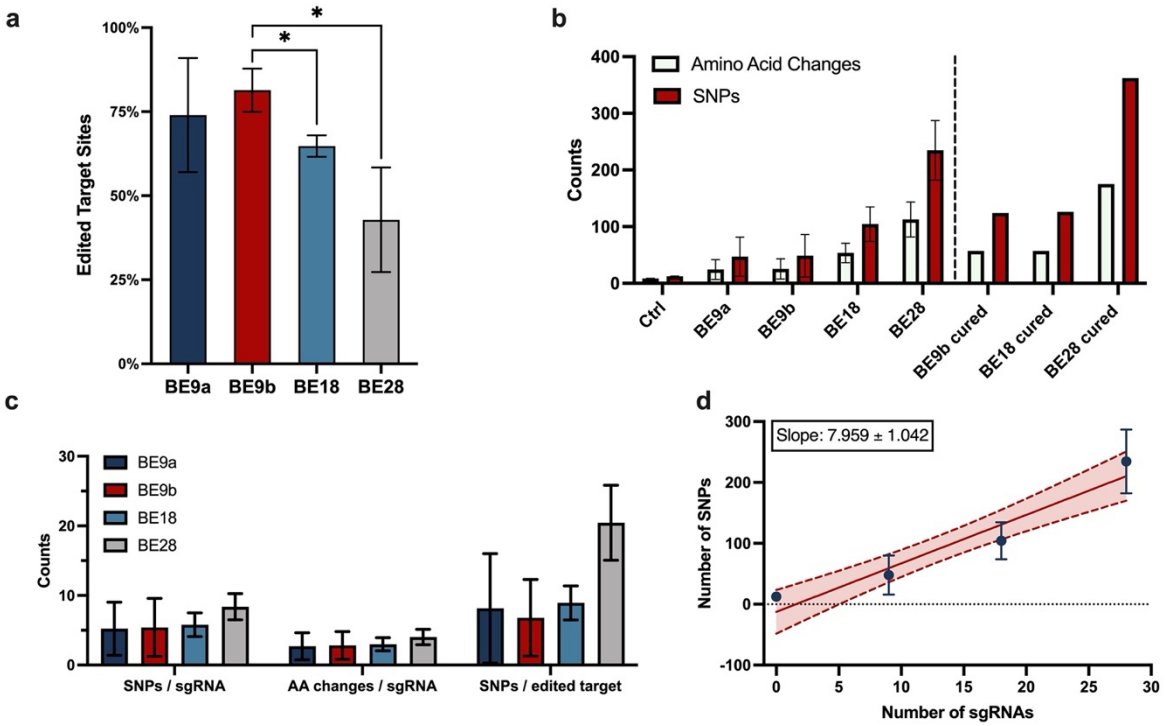


Figure 6: Detailed analysis of editing and SNP statistics. **a**. Overview of introduced SNPs and resulting amino acids changes for all generated strains. Addition of more sgRNAs resulted in a clear increase in SNPs and amino acid changes. Around half of all SNPs resulted in amino acid changes. Plasmid cured strains showed elevated levels of SNPs and amino acid changes (except for BE18 cured), likely resulting from the additional cultivation steps. **b**. Normalized SNP and amino acid changes. No significant difference was observed for the number of SNPS per sgRNA. The same was observed for amino acid changes. Interestingly, a clear difference was observed for SNPs per edited target, suggesting an increased off-target affinity for 28 targets, though large deviations were observed. **c**. Percentage of edited target sites for increasing array sizes. A significant drop was observed from 9 to 28 sgRNAs, showing how increasing array instability decreases successful editing outcomes. **d**. Plotting the number of sgRNAs against the number of SNPs, revealing linear scaling with a slope of around 8 SNPs per sgRNA. The data from both 9 sgRNA arrays was combined for the 9 sgRNA datapoint. The R squared was determined as 0.8179. Shown in **a**, **b**, **c**, and **d** are the means ± standard deviations of three biological replicates of plasmid harboring strains. Significance was tested using unpaired two tailed t-tests, where *P < 0.05.

In the BE9a and BE9b strains, editing events were observed on average in 74 % and 81.4 % of the target sites, respectively. For the BE18 strains, 64.8 % of all target sites displayed editing events, while for the BE28 strains this number dropped to 42.8 %. We further analyzed the number of SNPs and AA changes per sgRNA, and the number of SNPs per edited target. We observed no significant increase in the number of SNPs or AA changes per sgRNA while scaling from 9 to 28 sgRNAs. Interestingly however, we did see a significant increase in the number of SNPs per edited target when scaling from BE9a/b or BE18 to BE28. While scaling beyond 18 sgRNAs, on-target activity appeared to decrease, and off-target activity of the base editor increase.

Sequencing of mutants before and after plasmid curing allowed the identification of several bottlenecks for multiplexed base editing in *Streptomyces.* Most importantly, the off-target rate needs to be decreased, as around 8 SNPs are observed for each sgRNA. The used sgRNAs were designed using CRISPy-web and selected based on both high specificity and applicability for base editing. The off-target rate could likely be reduced by faster plasmid curing and reduction of leaky expression. Specificity increasing modifications to the sgRNAs such as bubble hairpins[28], as well as new Cas9 versions like SuperfiCas9 might also prove beneficial for *in vivo* application, though initial data suggests low on-target efficiency *in vivo*[29,30]. Novel base editors such as TadCBE might also be used to reduce off-targets[31]. Furthermore, counterselection or screening based on desired phenotypes might allow faster isolation of mutants of interest and help in recovering low-efficiency mutations, which often were not recovered after plasmid curing. All sgRNAs were selected with specificity as a key criterion. However, for some of the genes, only few suitable protospacer sequences enabling stop codon introductions were identified. For these genes, to retain a high specificity, protospacer sequences had to be selected that had Gs upstream of the target C, lowering the efficiency of base edits at those positions[12], and likely causing the low efficiencies observed for these targets. Our data further suggests that Cas-independent off-target mutations, if at all present, represent only a minor fraction of the total number of off-target mutations, as gRNA dependent Cas-activity appears to be the major driver of SNP introduction. The elevated ratio of SNPs per target for BE28 was surprising, as BE18 and BE28 are a combination of BE9a and BE9b, as well as the remaining 10 sgRNAs in the case of BE28. As previously shown, the number of SNPs per sgRNA scales linearly, and since

specificity was one of the key criteria for selection of sgRNAs, we did not expect the last 10 sgRNAs in the array to have a highly elevated off-target affinity compared to the first 18 sgRNAs. Secondly, large deletions were detected in the arrays of p085, which were acquired after conjugation as indicated by presence of mutations introduced by lost sgRNAs. Therefore, the mutational burden was expected to be reduced compared to BE9a/b and BE18. However, our view of these loci as mere targets overlooks the important role many of the specialized metabolites play in cell development, adaptation, and differentiation. Hence, the observed increase in off-target activity might be explained by the burden resulting from introduction of too many of the designed on-target edits for *S. coelicolor* M145, resulting in a selection bias towards off-target edits.

In conclusion, using Csy4 mediated multiplexed base editing data obtained at various scales, we were able to identify array sizes delivering the most robust outcomes. Array sizes of 9 sgRNAs were found to be the most robust for several reasons: First, cloning large synthetic sgRNA arrays (>12 sgRNAs) requires several complicated steps that can easily result in extended cloning times, especially considering the associated screening efforts. On the other hand, 9 sgRNAs can be easily assembled in a streamlined two-step protocol, with ligation into pCRISPR-mcBEST following preassembly in pGGA. Furthermore, we observed the highest percentage of edited target sites for smaller arrays, simplifying screening and selection of the *Streptomyces* mutants. Considering that the assembly of sgRNA arrays with up to 12 sgRNAs can still be performed in a simple two-step protocol with acceptable efficiencies (unpublished data), we propose that the best number of sgRNAs for multiplexed base editing in *Streptomyces* is between 8 and 12 sgRNAs, as it offers streamlined and robust generation of complex mutants.

Multi-omics analysis reveals extensive changes to central cellular processes and across metabolism

To gain further insights into the systems wide consequences of multiplexed base editing experiments, multi-omics analysis was performed. Samples were taken during cultivations at the highlighted time points (Figs. 5 and 7) and prepared as described in the methods section for untargeted proteomics and transcriptomics. The cultivations were continued after sampling and the supernatant was used for untargeted exometabolomics analysis after one week.

Focusing on only the targeted biosynthetic genes, highly significant negative fold changes were observed for the majority of targeted biosynthetic genes in ISP2 medium, especially if stop or rare codons were introduced (Fig. 7). Notable exceptions were observed in the transcriptomics data of *S. coelicolor* BE9b and BE28, showing how premature stop codons can result in overexpression of the target gene, likely based on the perturbed transcriptional regulatory network. Clear differences were observed between strains in DNPM medium. *S. coelicolor* BE18 and BE28 clearly stood out with little or no difference to the DNPM blank observable with the naked eye (Fig. 7d). In DNPM medium, the *m/z* of kalafungin, a shunt product of actinorhodin, was not detected in *S. coelicolor* B18 and BE28. The key biosynthetic gene of the actinorhodin BGC (FQ762_26280) was inactivated in BE28, while no mutation was detected in BE18. It is likely that altered high level regulation in response to other mutations resulted in abolished kalafungin production in BE18.

We further analyzed also wider changes in the untargeted exometabolomics, proteomics, and transcriptomics. Changes in the untargeted exometabolome were detected following base editing in ISP2 and DNPM liquid medium. Cultivations in ISP2 resulted in identification of the same major peaks across all wildtype and base edited strains, corresponding to germicidin B, germicidin A, though with differing peak ratios (Supplementary Fig. 8). In DNPM medium, much more distinct differences in the chromatograms were observed between the wildtype and base edited strains. The highest intensity peaks were identified as desferrioxamines A, B, D, E, and G, and germicidin B and A. Around 5.33 min, a major peak (C) was detected with a *m/z* 516.36, corresponding to desferrioxamine D2, which was much more dominant in the

supernatant of *S. coelicolor* BE18. Another peak was exclusively detected in *S. coelicolor* BE9a at 7.8 min and *m/z* 206.15, for which no match was found. However, molecular networking using GNPS[32] suggested a connection to the desferrioxamine network.

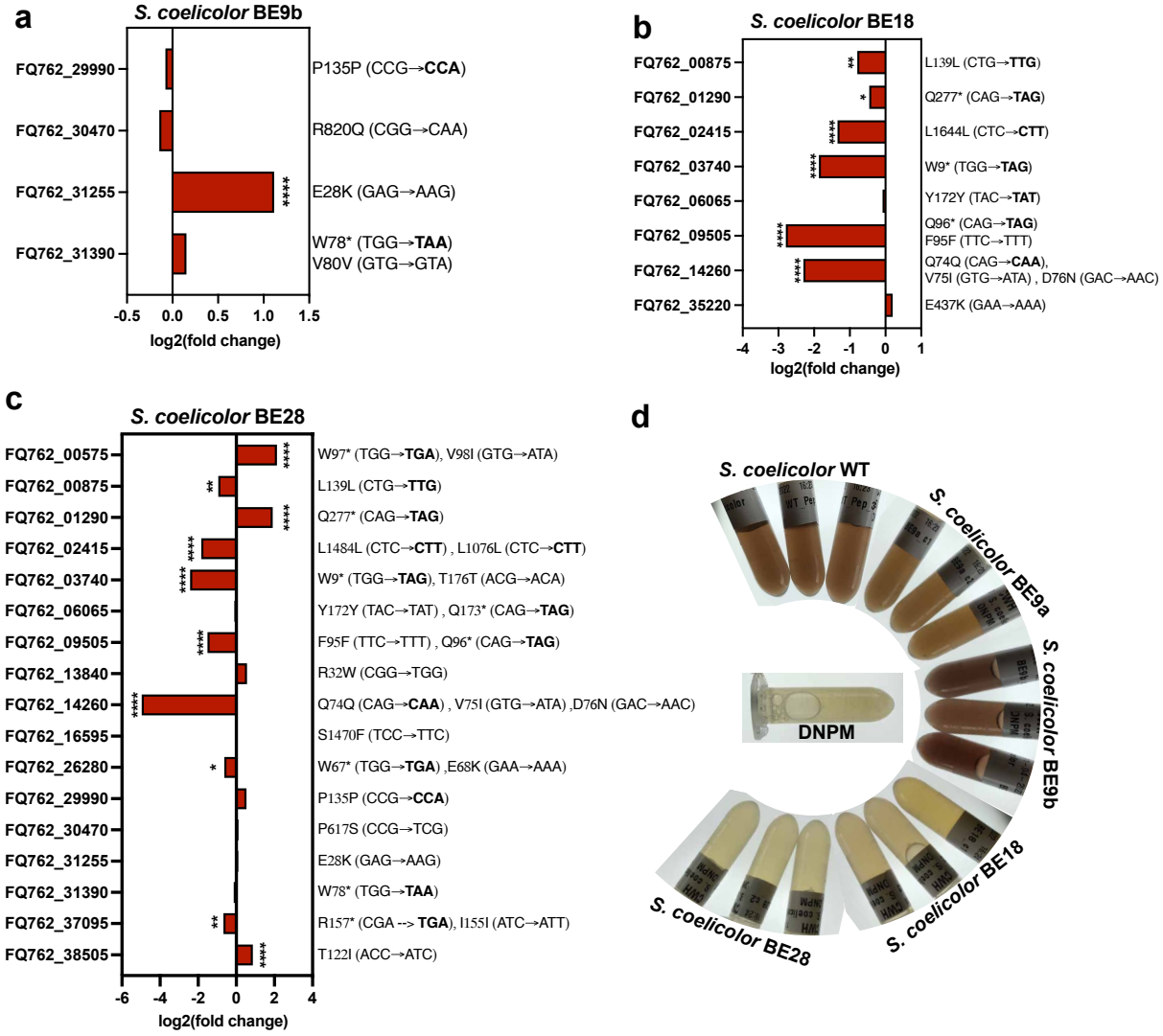


Figure 7: Selected transcriptomics result and supernatants of all mutants. a, **b**, **c**. Fold changes of targeted biosynthetic genes for *S. coelicolor* BE9b, BE18, and BE28, respectively. Highly significant changes in transcript levels were observed for the majority of edited targets, especially if stop codons or rare codons (highlighted bold) were introduced. Interestingly, editing also resulted in highly elevated transcript levels in some genes, potentially revealing characteristics of the underlying transcriptional regulatory networks. **d**. Supernatants of *S. coelicolor* WT, BE9a, BE9b, BE18, and BE28 DNPM cultures. The BE18 and BE28 supernatants were noticeably clearer in color compared to the wildtype or BE9a, BE9b. **e**. Extracted ion chromatograms for the m/z of kalafungin for all mutant strains, highlighting abolished production in *S. coelicolor* BE18 and BE28. In **a**, **b**, **c** the log2(fold change) of three technical replicates is shown. Significance was tested using unpaired two tailed t-tests, where *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

Both untargeted transcriptomics and proteomics revealed major changes to central cellular processes and both primary and secondary metabolism (Supplementary Figs. 5 and 6). The greatest changes (both positive and negative fold changes) were observed for genes with the COG categories amino acid metabolism and transport, unknown function, lipid metabolism, energy production and conversion, carbohydrate metabolism and transport, transcription, translation, and inorganic ion transport and metabolism. Interestingly, mapping the detected SNPs onto the highly changed genes revealed that only a small number of observed changes in transcript and peptide levels could be potentially explained by the SNPs. Instead, no mutations (on or off-target) were detected in the overwhelming majority of highly changed genes. The observed changes are likely the result of changes in regulation in response to the base edits and the strains adaptations to the mutational landscape, and as such were likely responses to *in trans* introduced mutations. Together with the observed highly significant changes in transcript levels of most of the target genes, we hypothesize that the observed system wide effects were primarily driven by the mutations introduced in the targeted biosynthetic genes. In that case, combining base editor mediated genomic perturbations with transcriptomics could potentially be applied to investigate regulatory mechanisms of BGCs by measuring variations in transcript levels in response to inactivation of biosynthetic genes.

Our results clearly show that multiplexed base editing can be used to perform complex rewiring and perturbation of cellular processes and pathways. However, given the high proportion of off-target mutations, a direct association of these observed systems wide changes to the on-target mutations remains difficult.

Conclusion

In the present study we developed and systematically investigated the performance and scalability of multiplexed base editing in *Streptomyces.* Multiplexed base editing has become a very popular method across various species for rapid engineering of complex phenotypes[33–36]. We performed a deep and systematic characterization of the editing outcomes to guide further improvements and developments of multiplexed base editing tools for non-model organisms, especially genetically recalcitrant ones. We show successful editing of up to 17 target sites and recovery of up to 15 stable edits after plasmid curing. To the best of our knowledge, this represents the highest number of simultaneous edits introduced in *Streptomyces* species to date. We analyzed off-target frequencies for all sgRNA array sizes, revealing a linear relationship between the number of SNPs and the number of sgRNAs. We further show that a significant drop in the percentage of edited target sites is observed with increasing array sizes. Multi-omics analysis revealed extensive changes across metabolism, most of which could not be directly associated with introduced mutations and are hence likely *in trans* acting responses to the introduced mutations.

In conclusion, we demonstrated for the first time highly multiplexed base editing in *Streptomyces* species and performed systems wide analysis of editing outcomes and consequences. We are convinced that this is a crucial step towards establishing highly multiplexed base editing as a standard tool for rapid strain construction and prototyping in *Streptomyces* species.

METHODS AND MATERIALS

Strains and Culture Conditions

All cloning work was performed using One Shot™ Mach1™ T1 Phage-Resistant Chemically Competent *E. coli* cells (ThermoFisher Scientific Inc., U.S.A). *E. coli* cells were cultivated on LB plates (10 g/l tryptone, 5 g/l yeast extract, 5 g/l sodium chloride, 15 g/l agar, to 1 l with MiliQ water) supplemented with the appropriate antibiotic (50 ng/µl apramycin or 25 ng/µl chloramphenicol) and incubated at 37 °C. 2xYT medium (16 g/l tryptone, 10 g/l yeast extract, 5 g/l sodium chloride, to 1 l with MiliQ water) and 15 ml culture tubes (VWR International, U.S.A) were used for liquid cultures (37 °C , 250 rpm). *Streptomyces* cultures were grown on mannitol soy flour (MS) (20 g/l fat reduced soy flour, 20 g/l mannitol, 20 g/l agar, to 1 l with tap water) plates supplemented with 10 mM $MgCl_2$ for conjugations or to obtain spores. Exconjugants were streaked on ISP2 plates (4 g/l yeast extract, 10 g/l malt extract, 4 g/l dextrose, 20 g/l agar, 333 ml tap water, 667 ml MiliQ water) supplemented with 50 ng/µl apramycin and 12.5 ng/µl nalidixic acid. Plates were incubated at 30 °C. *Streptomyces* liquid cultures were performed in 350 ml baffled shake flasks containing 50 ml of ISP2 medium supplemented with 25 ng/µl apramycin if needed. Cultivations were performed in shaking incubators at 30 °C and 180 rpm (311DS, Labnet International Inc., U.S.A). Plasmid curing was achieved through cultivation at 40 °C and parallel plating on selective and non-selective ISP2 plates for identification of plasmid-free clones.

In Silico Cloning and Protospacer Prediction

In silico cloning was performed using SnapGene 5.3 (GSL Biotech LLC, U.S.A). CRISPR protospacers were identified using CRISPy-web[24] using the CRISPR-BEST mode for C to T mutations and showing only STOP mutations. Golden Gate assemblies were simulated using the NEB Golden Gate Assembly Tool version 2.3.3 for BsaI-HFv2 (New England Biolabs Inc., U.S.A).

Plasmid Construction

All primers were ordered from IDT (Integrated DNA Technologies, U.S.A) and diluted to a working concentration of 10 µM. PCRs were performed using Q5® High-Fidelity DNA Polymerase with GC Enhancer (New England BioLabs Inc., U.S.A). Plasmid isolations were performed using NucleoSpin® Plasmid EasyPure Kit (Macherey-Nagel, Germany). PCR fragments were run on 1 % agarose gels at 100 V for 20-30 min using 6x DNA Gel Loading Dye and GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific Inc., U.S.A). Fragments were purified by gel purification using the NucleoSpin® PCR and Gel Clean Up Kit (Macherey-Nagel, Germany). DNA concentrations were measured on a NanoDrop™ 2000 (Thermo Fisher Scientific Inc., U.S.A). sgRNA arrays were assembled the BsaI-HFv2 GoldenGate kit (New England BioLabs Inc., U.S.A). The 28 sgRNA arrays were assembled in three pieces, the first containing sgRNAs 1-9, the second 10-18, and the third one 19-28. Smaller arrays were combined by restriction and ligation cloning. All fragments for GoldenGate assembly were DpnI digested and extracted from the gel. After chemical transformations following the protocol of the One Shot™ Mach1™ T1 Phage-Resistant Chemically Competent *E. coli* cells (ThermoFisher Scientific Inc., U.S.A), colony PCRs were performed using Q5® Mastermix with GC Enhancer (New England BioLabs Inc., U.S.A). Positive clones were verified by Sanger sequencing using Mix2Seq overnight kits (Eurofins Genomics Germany GmbH, Germany). The verified sgRNA arrays were cut from the pGGA plasmids using NcoI and NheI, gel purified, and ligated into digested and gel purified pCRISPR-mcBEST. Digestions were performed using Thermo Scientific™ FastDigest enzymes, using 1 h incubations at 37 °C and inactivation for 10 min at 75 °C. Ligations were performed using 5 U/µL T4 DNA Ligase (Thermo Fisher Scientific Inc., U.S.A) and incubation overnight at room temperature.

Interspecies Conjugations

Homemade chemically competent *E. coli* ET12567 pUZ8002 [37] cells were transformed with the plasmid of interest and plated on LB plates supplemented with 50 ng/µl apramycin, 25 ng/µl chloramphenicol, and 50 ng/µl kanamycin. Transformants were washed off the plate using 4 ml of LB

medium and 50 µl were used to inoculate 3 ml overnight cultures. 2 ml of the overnight cultures were harvested the next day and washed twice at 2000xg with 1 ml of LB medium. 500 µl were then mixed with 500 µL of spore suspension and spread on MS plates. Plates were overlayed with 1 ml of ddH_2O and 5 µl of 50 mg/ml apramycin the next day and incubated for another 4-5 days. Exconjugants were picked using wooden toothpicks and transferred to selective ISP2 plates. Spores for conjugations were prepared as described in Tong et al. [6].

Genomic DNA Extraction & Whole Genome Sequencing

Genomic DNA extractions were performed using Qiagen G20 columns with a Blood & Cell Culture DNA Mini Kit (Cat No./ID: 13323, Qiagen, Germany). Quality control was performed using a NanoDrop™ 2000 (Thermo Fisher Scientific Inc., U.S.A) and Qubit 4 Fluorometer for concentration determination (Thermo Fisher Scientific Inc., U.S.A). Genomic DNA was shipped to Novogene Co., Ltd (Bejing, China) for Illumina based WGS. Library preparation was performed based on the NEB Next® UltraTM DNA Library Prep Kit (New England Biolabs, U.S.A) with a target insert size of 350 nt and six PCR cycles.
Illumina data analysis was performed using trimgalore (https://github.com/FelixKrueger/TrimGalore) and breseq 0.33.2 [38]using the following commands:

trim_galore -j 8 --length 100 -o illumina --paired --quality 20 --fastqc --gzip file(s)

breseq -r (full genetic background reference).gb trimmed_1.fq.gz trimmed2.fq.gz -j 12

The *–polymorphism-frequency-cutoff* parameter was added with a value of 0.05 for sequencing of mixed populations.

Read coverage was extracted using samtools[26] using the breseq output. The optional parameter *-a* was added to include all positions.

samtools depth -a sample.bam > sample.coverage

Coverage for regions of interest was extracted for plotting/analysis:

awk '$1 == "Plasmid" {print $0}' sample.coverage > sample-plasmid.tsv

For annotation of rare codons, a cutoff of 5 % was used. Codons with a frequency of 5 % or less were regarded as rare codons.

Growth Curves & Sampling for multi-omics

Cultivations of plasmid free strains were performed in three technical replicates The OD600 of samples was measured using an Ultrospec 10 cell density meter (Amersham Biosciences, U.K.). Samples for proteomics (2 ml) and transcriptomics (2x 2ml) were taken in the mid to late exponential phase after 10 and 12 hours. The samples were centrifuged at 21,000xg for 1 min and the supernatant removed. Transcriptomics samples were immediately submerged in liquid nitrogen and stored at -80 °C until isolation. Proteomics samples were stored at -20 °C until analysis. The cultures were cultivated for a week after the last measurement, and then harvested and stored at -20 °C for metabolomics.

Metabolomics

Culture supernatants were collected by centrifugation of 2 ml of culture at 21,000xg for 10 min. Oasis® HLB 6cc (500mg) LP Extraction Cartridges (Waters Corp, U.S.A) were used for extractions in combination with a Positive Pressure-96 device (Waters Corp, U.S.A). The cartridges were conditioned with 600 µl of methanol, followed by 600 µL ddH_2O. 600 µl of supernatant were loaded. 600 µl of 5 % methanol was used for the washing step. A fresh capture plate was used for each of the elution steps. 600 µl of 50 % methanol were used for elution of polar compounds, followed by 600 µl of 100 % methanol for non-polar compounds. The elutions were dried under a nitrogen stream and redissolved using 50 µl ddH_2O (50 % methanol elutions) or 70 % methanol (100 % methanol elutions). The samples were subsequently analyzed using LC-UV-MS(MS) which was performed using a Vanquish Duo UHPLC binary system coupled to a PDA (Thermo Fisher Scientific Inc., USA) and to an Orbitrap IDX Tribrid Mass Spectrometer (Thermo Fisher Scientific Inc., USA). The LC separation was

achieved in reverse phase conditions as previously described [39] and the MS measurements were performed in positive-heated electrospray ionization (+HESI) mode with a voltage of 3500 V acquiring in full MS/MS spectra (Data dependent acquisition-driven MS/MS) in the mass range of 100-1500 Da. The MS settings were the following: automatic gain control (AGC) target value was set at 4e5 for the full MS and 5e4 for the MS/MS spectral acquisition, the mass resolution was set to 120,000 for full scan MS and 30,000 for MS/MS events. Precursor ions were fragmented by stepped High-energy collision dissociation (HCD) using collision energies of 20, 40, and 60.

Transcriptomics

RNA isolation was performed using the RNeasy kit (Qiagen, Germany). Concentration and purity of the samples was measured on a NanoDrop™ 2000. To check for DNA contaminations, 1.5 µl of each sample was digested with 1 µl of DNAse (Thermo Fisher Scientific Inc., U.S.A) and incubated at 37 °C for 20 min. DNAse digested samples were run on a gel in parallel to undigested samples. Remaining bands were indications of contaminations with DNA. The samples were analyzed for degradation using a Fragment Analyzer (Agilent Technologies, Inc., U.S.A). Library preparation, rRNA depletion, and sequencing was performed by Novogene Co., Ltd (Bejing, China) using provided total RNA.

All data analysis was performed in CLC Genomics 12.0.3 (Qiagen, Germany). A GenBank file of the reference strain (Accession: PRJNA557658) with all gene names removed was created to force CLC Genomics to use only locus tags, streamlining downstream data analysis. Read trimming and RNA seq analysis were performed with default parameters. PCAs, differential expression, heatmaps, Venn diagrams, and expression browsers were created using built-in functions. All data was exported for further processing and plotted using Prism 9 (GraphPad Software, U.S.A).

Proteomics

The cell pellets were thawed and remaining supernatant removed.
Two 3-mm zirconium oxide beads (Glen Mills, NJ, USA) and 100 µl of 95°C Guanidinium HCl (6 M Guanidinium hydrochloride (GuHCl), 5 mM tris(2-carboxyethyl)phosphine (TCEP), 10 mM chloroacetamide (CAA), 100 mM Tris–HCl pH 8.5) were added to all samples. A Mixer Mill (MM 400 Retsch, Haan, Germany) set at 25 Hz for 5 min at room temperature was used to achieve full cell disruption, followed by incubation in a thermomixer for 10 min at 95°C and 600 rpm. Cell debris was removed by centrifugation at 5000g for 10 min. For overnight tryptic digestion (constant shaking (400 rpm) for 8 h) a total of 100 ug protein were used, and 10 µl of 10% TFA were added afterwards. Subsequent desalting was performed using SOLAµ^{TM} SPE C18 plate (Thermo Fisher Scientific Inc., USA).

Desalted samples were injected into Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific Inc., USA) using a CapLC system (Thermo Fisher Scientific Inc., USA). Initial sample capturing was performed at a flow of 10 ul/min on a precolumn (µ-precolumn C18 PepMap 100, 5µm, 100Å), followed by peptide separation on a 15 cm C18 easy spray column (PepMap RSLC C18 2µm, 100Å, 150 µmx15cm) at a flow of 1.2 µl/min. Over a total of 60 minutes the applied gradient went from 4% acetonitrile in water to 76%. The instrument operated in data dependent mode while spraying the samples into the mass spectrometer using the following settings: MS-level scans were performed with Orbitrap resolution set to 60,000; AGC Target 3.0e6; maximum injection time 50 ms; intensity threshold 5.0e3; dynamic exclusion 25 sec. The Top 20 Speed mode with HCD collision energy set to 28% (AGC target 1.0e4, maximum injection time 22 ms, Isolation window 1.2 m/z) was used for data dependent MS2 selection.

Proteome discoverer 2.4 was used for analysis of the thermos raw files using the following settings: Fixed modifications: Carbamidomethyl (C) and variable modifications: oxidation of methionine residues. First search mass tolerance 20 ppm and a MS/MS tolerance of 20 ppm. Trypsin as enzyme and allowing one missed cleavage. FDR was set at 0.1%. The match between runs window was set to 0.7 min. Only unique peptides were used for the quantification and normalization between samples was performed using the total peptide

amount. A protein database consisting of all *S. coelicolor* M145 proteins was used for the searches. For both proteomics and transcriptomics volcano plots, cutoff values of a fold change of 2 (log2(FC)=1) and a p-value of 0.005 (-log10(0.005) = 2.3) were used. The conservative p-value was used to decrease the probability of false positive hits. Genes falling into the defined FC-p-value space were then annotated with their respective COG categories and searched for SNPs using the NGS data.

Data and Statistical Analysis

All data was reported as mean ± standard deviation as described where applicable. Statistical significance was tested using two-tailed unpaired t-tests, and significance was reported as *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001. Linear regressions were calculated in Prism 9 and plotted including the 95 % confidence interval. All experiments were performed in triplicates, either biological triplicates (e.g. different clones from the same conjugation) or technical triplicates (e.g. three separate cultivations of the same biological replicate).

Computational Analysis & figures & scripts

Data analysis and plotting was performed in Prism 9 (GraphPad Software, U.S.A), as well as Python 3 in combination with the data analysis library pandas v0.24.2, matplotlib v2.2.5, seaborn v0.9.1, and scipy v.1.2.3.

Data availability statement

The sequence datasets generated and analysed in this study are deposited in the NCBI sequence read archive (links to the individual files see BioProject: PRJNA934069) and DTU Data (10.11583/DTU.22100804 (will be activated once paper is accepted) // link for reviewers: https://figshare.com/s/6b8a026531f8395997b1). Jupyter Notebooks/scripts used to process the data are available on Github: https://github.com/NBChub/Multiplexing_Scripts.

ABBREVIATIONS

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats; CBE: cytosine base editing; sgRNA: single guide RNA; AA: amino acid; SNP: single nucleotide polymorphism

AUTHOR CONTRIBUTIONS

C.M.W, Y.T, and T.W conceived and design the project. C.M.W, Y.T, T.G, and E.P carried out laboratory experiments. C.M.W and Y.T analyzed data. C.M.W, Y.T, and T.W wrote the manuscript with input from all authors.

COMPETING INTERESTS

All authors have no competing interests to declare.

ACKNOWLEDGEMENTS

This work was supported by grants from the Novo Nordisk Foundation (NNF16OC0021746, NNF20CC0035580). Y.T. was also supported by the National Key Research and Development Program of China (2021YFC2100600 and 2021YFA0909500), the National Natural Science Foundation of China (32170080). S.Y.L. was also supported by the Technology Development Program to Solve Climate Changes on Systems Metabolic Engineering for Biorefineries (NRF-2012M1A2A2026556 and NRF-2012M1A2A2026557) from the Ministry of Science and ICT through the National Research Foundation (NRF) of Korea. We would like to thank Robert Samples for providing us with MS2 fragmentation spectra for kalafungin, and Eftychia Eva Kontou for help with analyzing the metabolomics datasets.

REFERENCES

(1) Gavriilidou, A.; Kautsar, S. A.; Zaburannyi, N.; Krug, D.; Müller, R.; Medema, M. H.; Ziemert, N. Compendium of Specialized Metabolite Biosynthetic Diversity Encoded in Bacterial Genomes. *Nat. Microbiol.* **2022**, *7* (5), 726–735. https://doi.org/10.1038/s41564-022-01110-2.

(2) Myronovskyi, M.; Luzhetskyy, A. Heterologous Production of Small Molecules in the Optimized: *Streptomyces* Hosts. *Nat. Prod. Rep.* **2019**, *36* (9), 1281–1294. https://doi.org/10.1039/c9np00023b.

(3) Bu, Q. T.; Yu, P.; Wang, J.; Li, Z. Y.; Chen, X. A.; Mao, X. M.; Li, Y. Q. Rational Construction of Genome-Reduced and High-Efficient Industrial *Streptomyces* Chassis Based on Multiple Comparative Genomic Approaches. *Microb. Cell Fact.* **2019**, *18* (1), 1–17. https://doi.org/10.1186/s12934-019-1055-7.

(4) Myronovskyi, M.; Rosenkränzer, B.; Nadmid, S.; Pujic, P.; Normand, P.; Luzhetskyy, A. Generation of a Cluster-Free *Streptomyces Albus* Chassis Strains for Improved Heterologous Expression of Secondary Metabolite Clusters. *Metab. Eng.* **2018**, *49* (July), 316–324. https://doi.org/10.1016/j.ymben.2018.09.004.

(5) Gren, T.; Whitford, C. M.; Mohite, O. S.; Jørgensen, T. S.; Kontou, E. E.; Nielsen, J. B.; Lee, S. Y.; Weber, T. Characterization and Engineering of *Streptomyces Griseofuscus* DSM 40191 as a Potential Host for Heterologous Expression of Biosynthetic Gene Clusters. *Sci. Rep.* **2021**, *11* (1), 18301. https://doi.org/10.1038/s41598-021-97571-2.

(6) Tong, Y.; Whitford, C. M.; Blin, K.; Jørgensen, T. S.; Weber, T.; Lee, S. Y. CRISPR–Cas9, CRISPRi and CRISPR-BEST-Mediated Genetic Manipulation in Streptomycetes. *Nat. Protoc.* **2020**, *15* (8), 2470–2502. https://doi.org/10.1038/s41596-020-0339-z.

(7) Whitford, C. M.; Cruz-Morales, P.; Keasling, J. D.; Weber, T. The Design-Build-Test-Learn Cycle for Metabolic Engineering of Streptomycetes. *Essays Biochem.* **2021**, *65* (2), 261–275. https://doi.org/10.1042/EBC20200132.

(8) Alberti, F.; Corre, C. Editing Streptomycete Genomes in the CRISPR/Cas9 Age. *Nat. Prod. Rep.* **2019**, *36* (9), 1237–1248. https://doi.org/10.1039/C8NP00081F.

(9) Hoff, G.; Bertrand, C.; Piotrowski, E.; Thibessard, A.; Leblond, P. Genome Plasticity Is Governed by Double Strand Break DNA Repair in

Streptomyces. *Sci. Rep.* **2018**, *8* (1), 5272. https://doi.org/10.1038/s41598-018-23622-w.

(10) Li, L.; Wei, K.; Zheng, G.; Liu, X.; Chen, S.; Jiang, W.; Lu, Y. CRISPR-Cpf1-Assisted Multiplex Genome Editing and Transcriptional Repression in *Streptomyces*. *Appl. Environ. Microbiol.* **2018**, *84* (18), 1–18. https://doi.org/10.1128/AEM.00827-18.

(11) Tong, Y.; Charusanti, P.; Zhang, L.; Weber, T.; Lee, S. Y. CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. *ACS Synth. Biol.* **2015**, *4* (9), 1020–1029. https://doi.org/10.1021/acssynbio.5b00038.

(12) Tong, Y.; Whitford, C. M.; Robertsen, H. L.; Blin, K.; Jørgensen, T. S.; Klitgaard, A. K.; Gren, T.; Jiang, X.; Weber, T.; Lee, S. Y. Highly Efficient DSB-Free Base Editing for Streptomycetes with CRISPR-BEST. *Proc. Natl. Acad. Sci.* **2019**, *116* (41), 20366–20375. https://doi.org/10.1073/pnas.1913493116.

(13) Cobb, R. E.; Wang, Y.; Zhao, H. High-Efficiency Multiplex Genome Editing of *Streptomyces* Species Using an Engineered CRISPR/Cas System. *ACS Synth. Biol.* **2015**, *4* (6), 723–728. https://doi.org/10.1021/sb500351f.

(14) McCarty, N. S.; Graham, A. E.; Studená, L.; Ledesma-Amaro, R. Multiplexed CRISPR Technologies for Gene Editing and Transcriptional Regulation. *Nat. Commun.* **2020**, *11* (1), 1281. https://doi.org/10.1038/s41467-020-15053-x.

(15) Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; Zhang, F. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (80-.).* **2013**, *339* (6121), 819–823. https://doi.org/10.1126/science.1231143.

(16) Hampf, M.; Gossen, M. Promoter Crosstalk Effects on Gene Expression. *J. Mol. Biol.* **2007**, *365* (4), 911–920. https://doi.org/10.1016/j.jmb.2006.10.009.

(17) Kurata, M.; Wolf, N. K.; Lahr, W. S.; Weg, M. T.; Kluesner, M. G.; Lee, S.; Hui, K.; Shiraiwa, M.; Webber, B. R.; Moriarity, B. S. Highly Multiplexed Genome Engineering Using CRISPR/Cas9 GRNA Arrays. *PLoS One* **2018**, *13* (9), e0198714. https://doi.org/10.1371/journal.pone.0198714.

(18) Nissim, L.; Perli, S. D.; Fridkin, A.; Perez-Pinera, P.; Lu, T. K. Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells. *Mol. Cell* **2014**, *54* (4), 698–710. https://doi.org/10.1016/j.molcel.2014.04.022.

(19) Liu, Y.; Gao, Y.; Gao, Y.; Zhang, Q. Targeted Deletion of Floral Development Genes in Arabidopsis with CRISPR/Cas9 Using the RNA Endoribonuclease Csy4 Processing System. *Hortic. Res.* **2019**, *6* (1), 99. https://doi.org/10.1038/s41438-019-0179-6.

(20) Ferreira, R.; Skrekas, C.; Nielsen, J.; David, F. Multiplexed CRISPR/Cas9 Genome Editing and Gene Regulation Using Csy4 in *Saccharomyces Cerevisiae*. *ACS Synth. Biol.* **2018**, *7* (1), 10–15. https://doi.org/10.1021/acssynbio.7b00259.

(21) Haurwitz, R. E.; Jinek, M.; Wiedenheft, B.; Zhou, K.; Doudna, J. A. Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease. *Science (80-.).* **2010**, *329* (5997), 1355–1358. https://doi.org/10.1126/science.1192272.

(22) Engler, C.; Kandzia, R.; Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS One* **2008**, *3* (11), e3647. https://doi.org/10.1371/journal.pone.0003647.

(23) Blin, K.; Shaw, S.; Kloosterman, A. M.; Charlop-Powers, Z.; van Wezel, G. P.; Medema, M. H.; Weber, T. AntiSMASH 6.0: Improving Cluster Detection and Comparison Capabilities. *Nucleic Acids Res.* **2021**, *49* (W1), W29–W35. https://doi.org/10.1093/nar/gkab335.

(24) Blin, K.; Pedersen, L. E.; Weber, T.; Lee, S. Y. CRISPy-Web: An Online Resource to Design SgRNAs for CRISPR Applications. *Synth. Syst. Biotechnol.* **2016**, *1* (2), 118–121. https://doi.org/10.1016/j.synbio.2016.01.003.

(25) Barrick, J. E.; Colburn, G.; Deatherage, D. E.; Traverse, C. C.; Strand, M. D.; Borges, J. J.; Knoester, D. B.; Reba, A.; Meyer, A. G. Identifying Structural Variation in Haploid Microbial Genomes from Short-Read Resequencing Data Using Breseq. *BMC Genomics* **2014**, *15* (1), 1039. https://doi.org/10.1186/1471-2164-15-1039.

(26) Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25* (16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

(27) Reis, A. C.; Halper, S. M.; Vezeau, G. E.; Cetnar, D. P.; Hossain, A.; Clauer, P. R.; Salis, H. M. Simultaneous Repression of Multiple Bacterial Genes Using Nonrepetitive Extra-Long SgRNA Arrays. *Nat. Biotechnol.* **2019**, *37* (11), 1294–1301. https://doi.org/10.1038/s41587-019-0286-9.

(28) Hu, Z.; Wang, Y.; Liu, Q.; Qiu, Y.; Zhong, Z.; Li, K.; Li, W.; Deng, Z.; Sun, Y. Improving the Precision of Base Editing by Bubble Hairpin Single Guide RNA. *MBio* **2021**, *12* (2). https://doi.org/10.1128/mBio.00342-21.

(29) Bravo, J. P. K.; Liu, M.-S.; Hibshman, G. N.; Dangerfield, T. L.; Jung, K.; McCool, R. S.; Johnson, K. A.; Taylor, D. W. Structural Basis for Mismatch Surveillance by CRISPR–Cas9. *Nature* **2022**, *603* (7900), 343–347. https://doi.org/10.1038/s41586-022-04470-1.

(30) Kulcsár, P. I.; Tálas, A.; Ligeti, Z.; Krausz, S. L.; Welker, E. SuperFi-Cas9 Exhibits Remarkable Fidelity but Severely Reduced Activity yet Works Effectively with ABE8e. *Nat. Commun.* **2022**, *13* (1), 6858. https://doi.org/10.1038/s41467-022-34527-8.

(31) Neugebauer, M. E.; Hsu, A.; Arbab, M.; Krasnow, N. A.; McElroy, A. N.; Pandey, S.; Doman, J. L.; Huang, T. P.; Raguram, A.; Banskota, S.; Newby, G. A.; Tolar, J.; Osborn, M. J.; Liu, D. R. Evolution of an Adenine Base Editor into a Small, Efficient Cytosine Base Editor with Low off-Target Activity. *Nat. Biotechnol.* **2022**. https://doi.org/10.1038/s41587-022-01533-6.

(32) Wang, M.; Carver, J. J.; Phelan, V. V; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V; Meehan, M. J.; Liu, W.-T.; Crüsemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderón, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C.-C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C.-C.; Yang, Y.-L.; Humpf, H.-U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; Boya P, C. A.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodríguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P.-M.; Phapale, P.; Nothias, L.-F.; Alexandrov, T.; Litaudon, M.;

Wolfender, J.-L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D.-T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Müller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. Ø.; Pogliano, K.; Linington, R. G.; Gutiérrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N. Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34* (8), 828–837. https://doi.org/10.1038/nbt.3597.

(33) Yuan, Q.; Gao, X. Multiplex Base- and Prime-Editing with Drive-and-Process CRISPR Arrays. *Nat. Commun.* **2022**, *13* (1), 2771. https://doi.org/10.1038/s41467-022-30514-1.

(34) Volke, D. C.; Martino, R. A.; Kozaeva, E.; Smania, A. M.; Nikel, P. I. Modular (de)Construction of Complex Bacterial Phenotypes by CRISPR/NCas9-Assisted, Multiplex Cytidine Base-Editing. *Nat. Commun.* **2022**, *13* (1), 3026. https://doi.org/10.1038/s41467-022-30780-z.

(35) Chen, Y.; Cheng, M.; Li, Y.; Wang, L.; Fang, L.; Cao, Y.; Song, H. Highly Efficient Multiplex Base Editing: One-Shot Deactivation of Eight Genes in *Shewanella Oneidensis* MR-1. *Synth. Syst. Biotechnol.* **2023**, *8* (1), 1–10. https://doi.org/10.1016/j.synbio.2022.09.005.

(36) Heo, Y. B.; Hwang, G.-H.; Kang, S. W.; Bae, S.; Woo, H. M. High-Fidelity Cytosine Base Editing in a GC-Rich *Corynebacterium Glutamicum* with Reduced DNA Off-Target Editing Effects. *Microbiol. Spectr.* **2022**. https://doi.org/10.1128/spectrum.03760-22.

(37) MacNeil, D. J.; Gewain, K. M.; Ruby, C. L.; Dezeny, G.; Gibbons, P. H.; MacNeil, T. Analysis of *Streptomyces Avermitilis* Genes Required for Avermectin Biosynthesis Utilizing a Novel Integration Vector. *Gene* **1992**, *111* (1), 61–68. https://doi.org/10.1016/0378-1119(92)90603-M.

(38) Deatherage, D. E.; Barrick, J. E. Identification of Mutations in Laboratory-Evolved Microbes from Next-Generation Sequencing Data Using Breseq. In *Methods in Molecular Biology*; 2014; pp 165–188. https://doi.org/10.1007/978-1-4939-0554-6_12.

(39) Kildegaard, K. R.; Arnesen, J. A.; Adiego-Pérez, B.; Rago, D.; Kristensen, M.; Klitgaard, A. K.; Hansen, E. H.; Hansen, J.; Borodina, I. Tailored Biosynthesis of Gibberellin Plant Hormones in Yeast. *Metab. Eng.* **2021**, *66*, 1–11. https://doi.org/10.1016/j.ymben.2021.03.010.

SUPPLEMENTARY INFORMATION

SI 1: Additional Information and experimental methods for pCRISPR-mcBEST

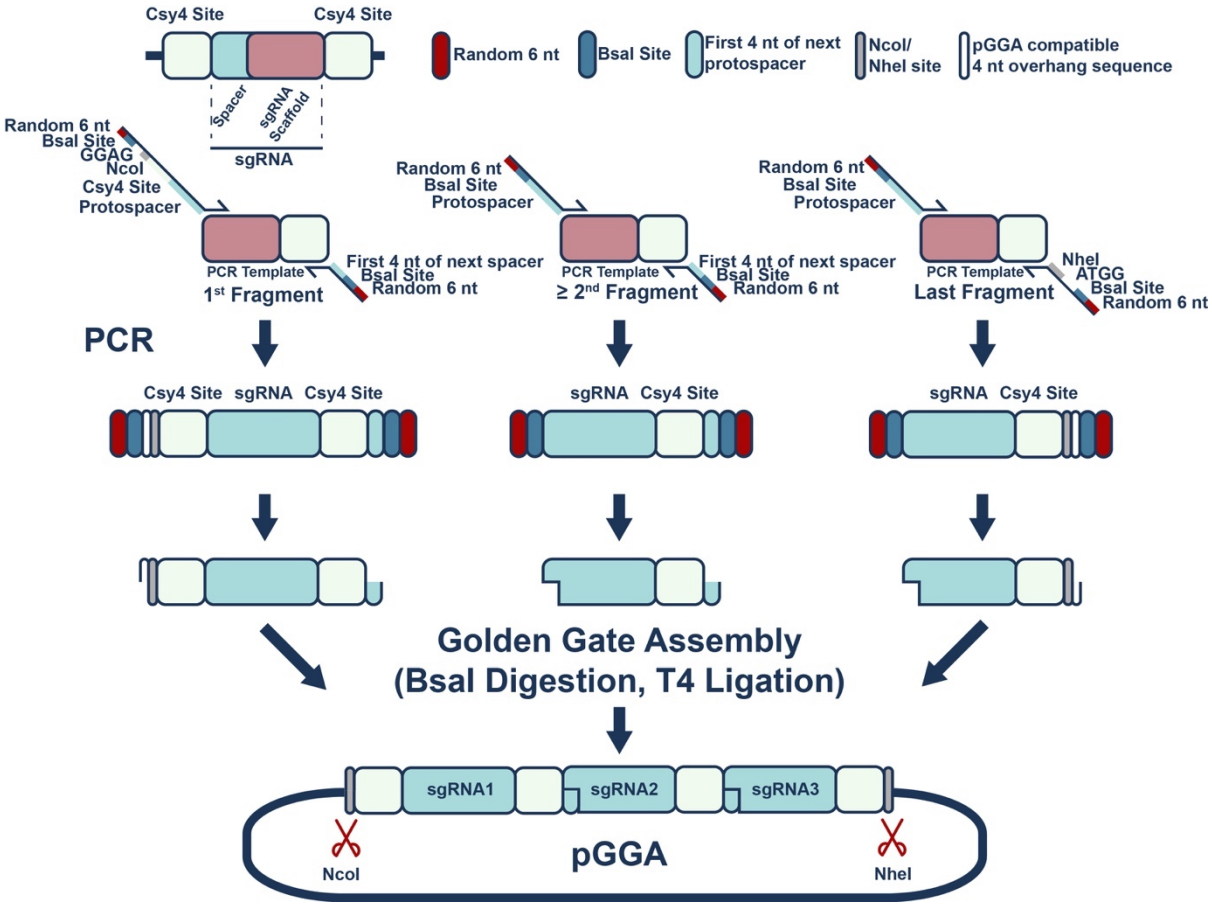
Development of an improved multiplexing plasmid

The initial plasmid (version 0) was constructed based on the pCRISPR-cBEST plasmid by replacing the sgRNA cassette with *csy4* from *Pseudomonas aeruginosa* PAO1. The codon usage of *csy4* from *P. aeruginosa* PAO1 was compared to the codon usage of *S. coelicolor* M145 using graphical codon usage analyzer and a mean difference of 19.69 % was found [1]. The mean difference was shown to be low enough to allow heterologous expression in *S. coelicolor* M145 without further codon optimization. Version 0 further carries a landing pad for the synthetic sgRNA array which replaces the *fd* terminator region present in pCRISPR-cBEST [2]. The original design of the landing pad consisted of a SP19 promoter of medium strength [3] followed by a NcoI restriction site, a strong RBS (Registry of Standard Biological Parts: BBa_B0034), the *mCherry* gene from pGM1192[4], followed by a NheI restriction site, and a t0 terminator. Replacing a selection marker with the synthetic sgRNA array enables simplified screening for correct colonies.

Construction of Plasmids for Multiplexed Engineering

Construction of synthetic sgRNA arrays is enabled by Golden Gate Assembly. Using a universal PCR template consisting of the sgRNA scaffold and one downstream Csy4 site, individual sgRNA-Csy4 fragments are amplified using high fidelity PCR. Three different fragment architectures are used for the assembly: the first fragment, last fragment, and middle fragment for array sizes greater than 2. Constant for all is the 6 nt random spacer sequence introduce at the 5' end of the primer overhang and the following BsaI recognition site. The forward primer of the first fragment introduces a pGGA compatible GGAG overhang sequence, and an NcoI restriction sequence, which can later be used for cloning into pCRISPR-mcBEST, another Csy4 sequence, and finally the 20 nt sequence of the first protospacer. The forward primer of the other architectures is missing the additional Csy4 site, but apart

from that have the same design. The reverse primer of the last fragment has a fixed pGGA compatible overhang sequence ATGG followed by a NheI recognition sequence. For the other fragment architectures, the first 4 nt of the following spacer are added in the overhang. All PCR reactions were DpnI digested and gel extracted, and subsequently used for a one pot Golden Gate reaction. The assembled pGGA fragments can then be verified by colony PCRs and sequencing and cut and ligated into the pCRISPR-mcBEST or pCRISPR-mdCas9 vectors. For construction of 18 and 28 sgRNA arrays, three such assembled intermediate arrays were combined by addition of HindIII (between sgRNAs 9 and 10) and XbaI (between sgRNAs 18 and 19) restriction sites and ligation of two or three sgRNA arrays into one.



Supplementary Figure 1: Construction scheme for synthetic sgRNA arrays. The arrays are constructed from PCR fragments based on one of the three potential architectures, based on the location of the fragment in the array. After DpnI digestion and gel clean-up, the fragments are then used for a one-pot assembly reaction using Golden Gate Assembly into the capture vector pGGA. The array can subsequently be cut and pasted into the pCRISPR-mcBEST plasmid using NcoI and NheI restriction sites.

SI 2: Table of the sgRNAs used in this study

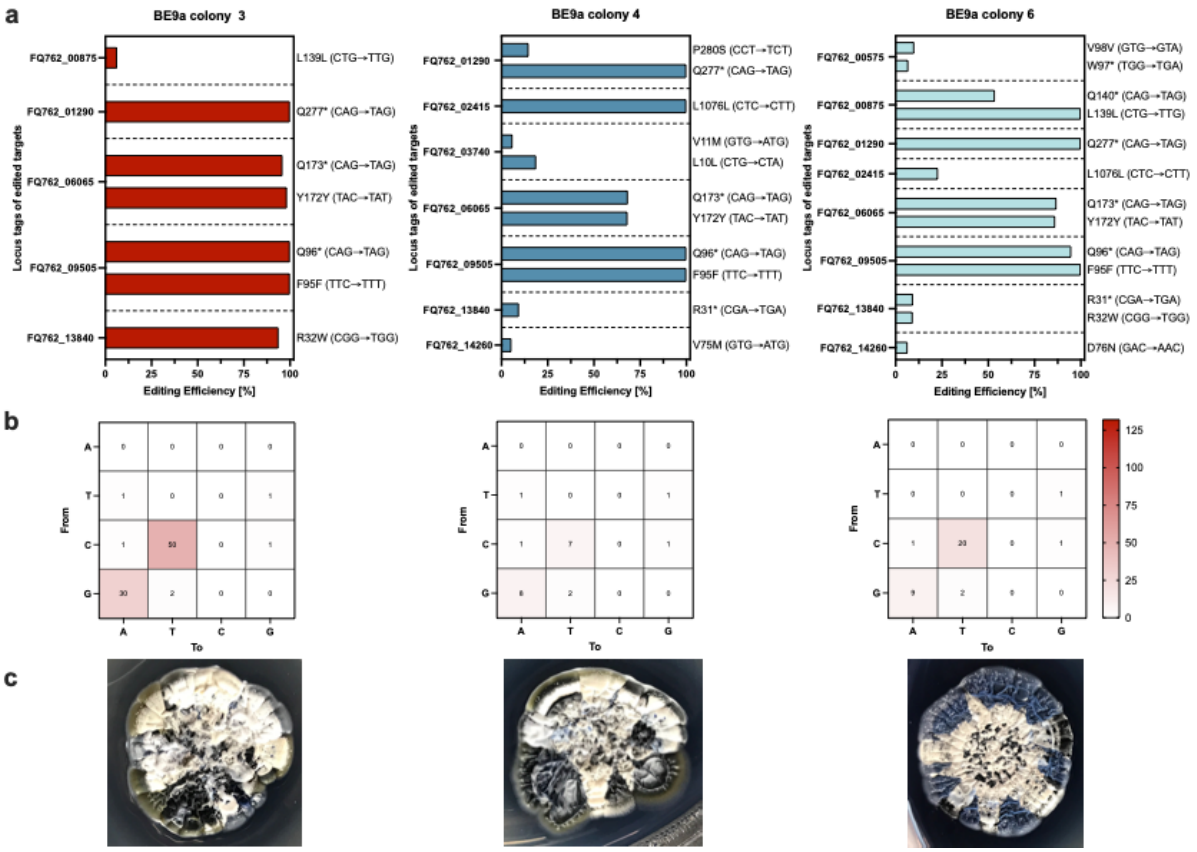
The targets used in this study are shown in supplementary Tab. 1. All protospacer sequences were predicted using CRISPy-web and selected based on specificity and compatibility with cytosine base editing.

Supplementary Table 1: List of base editing targets used in this study. The sgRNA numbers are the same as the locations of the respective sgRNAs in the final array. Both the locustag of the updated *S. coelicolor* assembly, as well as the corresponding SCO locustags are given. The listed BGC types were taken from antiSMASH, and all protospacer sequences are in 5' → 3' direction.

#sgRNA	Locustag	NC_003888 Locustag	BGC	Protospacer Sequence 5' → 3'
sg1	FQ762_00575	SCO0127	hglE-KS,T1PKS	TAGAGCACCCAGTGGAAGAG
sg2	FQ762_00875	SCO0187	isorenieratene	GATCTGCAGGCGTACGTACA
sg3	FQ762_01290	SCO0269	lanthipeptide	GGCATCAGCAGGATCCTCCG
sg4	FQ762_02415	SCO0492	coelichelin	CCAGGTGCTCTTCAACTATC
sg5	FQ762_03740	SCO0754	bacteriocin	CACCAGCCAGCCCTGGGGAC
sg6	FQ762_06065	SCO1206	T3PKS	TGCTACCAGCCCACCGACCT
sg7	FQ762_09505	SCO1865	ectoine	CCTTCCAGAACACCATCCTG
sg8	FQ762_13840	SCO2701	melanin	ACCCGACGGCAGGTGATGCG
sg9	FQ762_14260	SCO2785	siderophore	GTCCACCTGCCAGTGGTCCA
sg10	FQ762_16595	SCO3231	CDA	CGTGGGGCAGTTGGAAATCA
sg11	FQ762_26280	SCO5087	Actinorhodin	GAGCAGTTCCCAGAACTGCC
sg12	FQ762_26985	SCO5223	albaflavenone	GCGGGCAGTACATGGACGAG
sg13	FQ762_27465	SCO5317	Spore Pigment	CTGCAGAACCTGTGGGGGCA
sg14	FQ762_29990	SCO5800	siderophore	GTCGGGGACCAGTAGCGGAC
sg15	FQ762_30470	SCO5892	undecyprodigiosin	AGCAGCCGCCAGTAGCTGTC
sg16	FQ762_31255	SCO6045	bacteriocin	GCCTCGACCCACTCGATGCC
sg17	FQ762_31390	SCO6073	geosmin	CACGTACCAGTCGGTGATGA
sg18	FQ762_32170	SCO6226	siderophore	ACGCCACAGCCGCGACAGCA
sg19	FQ762_32415	SCO6275	coelimycin P1	GGGGCGCAGCGACCGGGCAT
sg20	FQ762_33215	SCO6432	thioamide-NRP,NRPS	AAGAGACACAACGGGGTGCG
sg21	FQ762_34460	SCO6680	SapB	GTACCACTGGTTCCTGGCGA
sg22	FQ762_34870	SCO6759	hopene	CGACGACCGACTGCTGATGC
sg23	FQ762_35220	SCO6827	arsono-polyketide	GGCTCGCAGTGGCCAGGCAT
sg24	FQ762_35750	SCO6930	lanthipeptide	GCAGGTTCAGGCCCTGTGCT
sg25	FQ762_37095	SCO7190	unknown	ATCGACCGATTCGCGATCAT
sg26	FQ762_38505	SCO7467	indole	GGCGGCCAGTTGCGAAGACT
sg27	FQ762_39600	SCO7683	coelibactin	CTGGCCCGATACGCGGACGA
sg28	FQ762_37265	SCO7221	germicidin	CGCGGTGCCCACGGAGCGTG

SI 3: Additional data for S. coelicolor BE9a

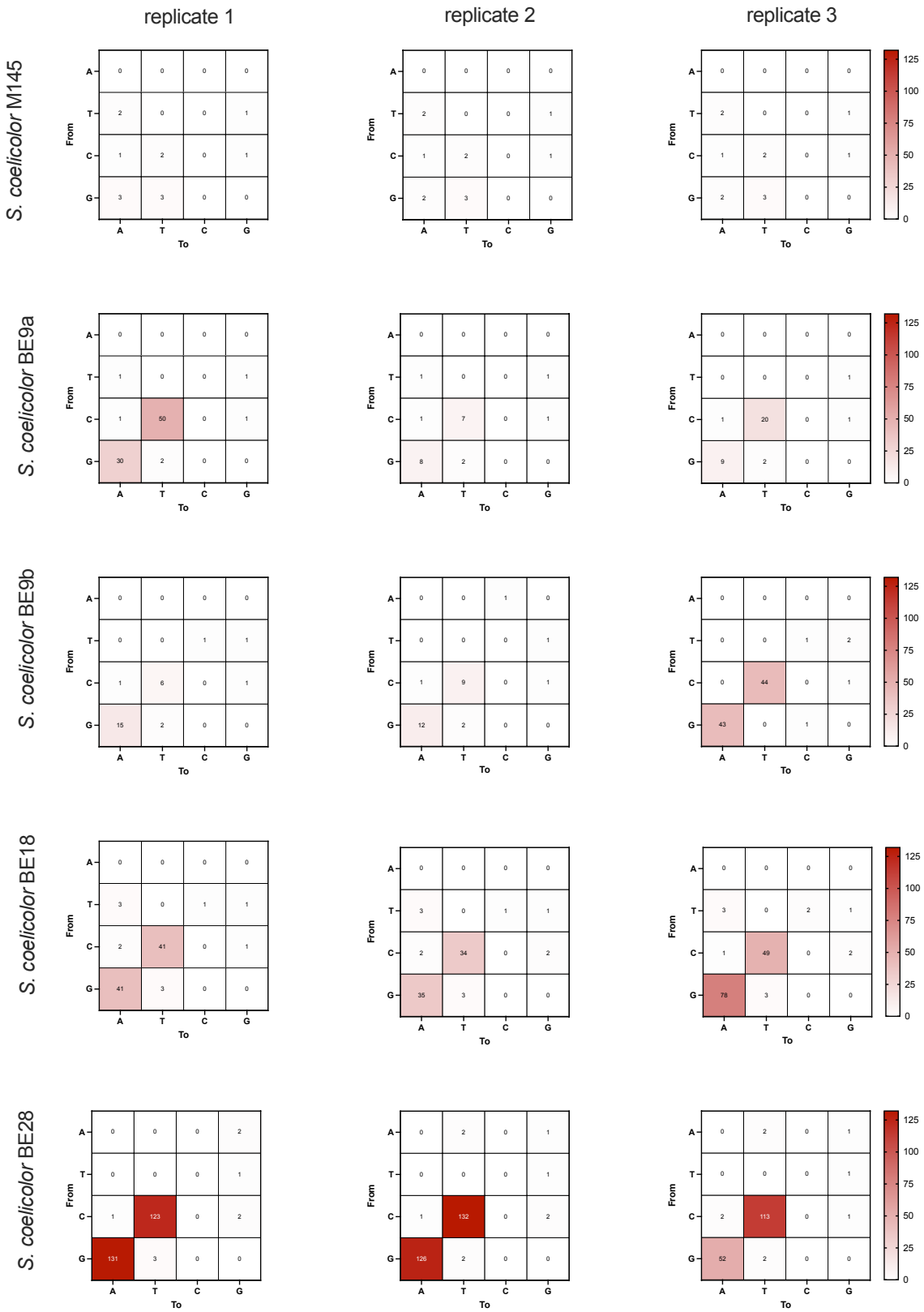
S. coelicolor BE9a mutants were constructed using sgRNAs 1-9, targeting biosynthetic genes in the first 9 BGCs of *S. coelicolor*. Apart from red flaviolin, which is a spontaneous reaction product of 1,3,6,8-tetrahydroxynaphthalene produced by *rppA* (FQ762_06065) in BGC 6[5], no direct clear phenotype was expected for any of the targeted BGCs as a result of successful editing. Between 5 and 8 out of the 9 targeted sites were successfully edited, however, with greatly varying efficiencies (Supplementary Fig. 2). In *S. coelicolor* BE9a colony 6, up to 6 stop codons were introduced in the targeted genes. Genome wide SNP analysis was performed using Illumina whole genome sequencing data. The majority of observed SNPs were either C to T or G to A mutations, indicating that these were introduced by the cytosine deaminase. Noticeably, colony 3, which has the fewest of the target sites edited, had a strongly elevated SNP count compared to colonies 4 and 6, which had 7 and 8 of the target sites edited. To assess phenotypical changes, drop of *S. coelicolor* BE9a colonies were plated on ISP2 plates. *S. coelicolor* BE9a colony 6 clearly stood out with its more structured and bright blue and white phenotype, indicating increased production of actinorhodin compared to colonies 3 and 4.



Supplementary Figure 2: Editing outcomes for three different colonies of *S. coelicolor* BE9a. **a**. Up to 8 out of 9 target sites were successfully edited, however with greatly varying editing efficiencies. Up to 6 stop codons were introduced in colony 6. **b**. Illumina-based genome wide SNP analysis. The majority of observed SNPs were C to T or G to A mutations, as expected for a cytosine deaminase. Interestingly, colony 3, which has the fewest edited target sites, has a highly elevated SNP count compared to colonies 4 and 6. **c**. Phenotypes of *S. coelicolor* BE9a mutants on ISP2 medium. While colonies 3 and 4 have a similar phenotype, colony 6 stands out with its patterned dark blue and white phenotype.

SI 4: SNP heatmaps for all sequenced strains

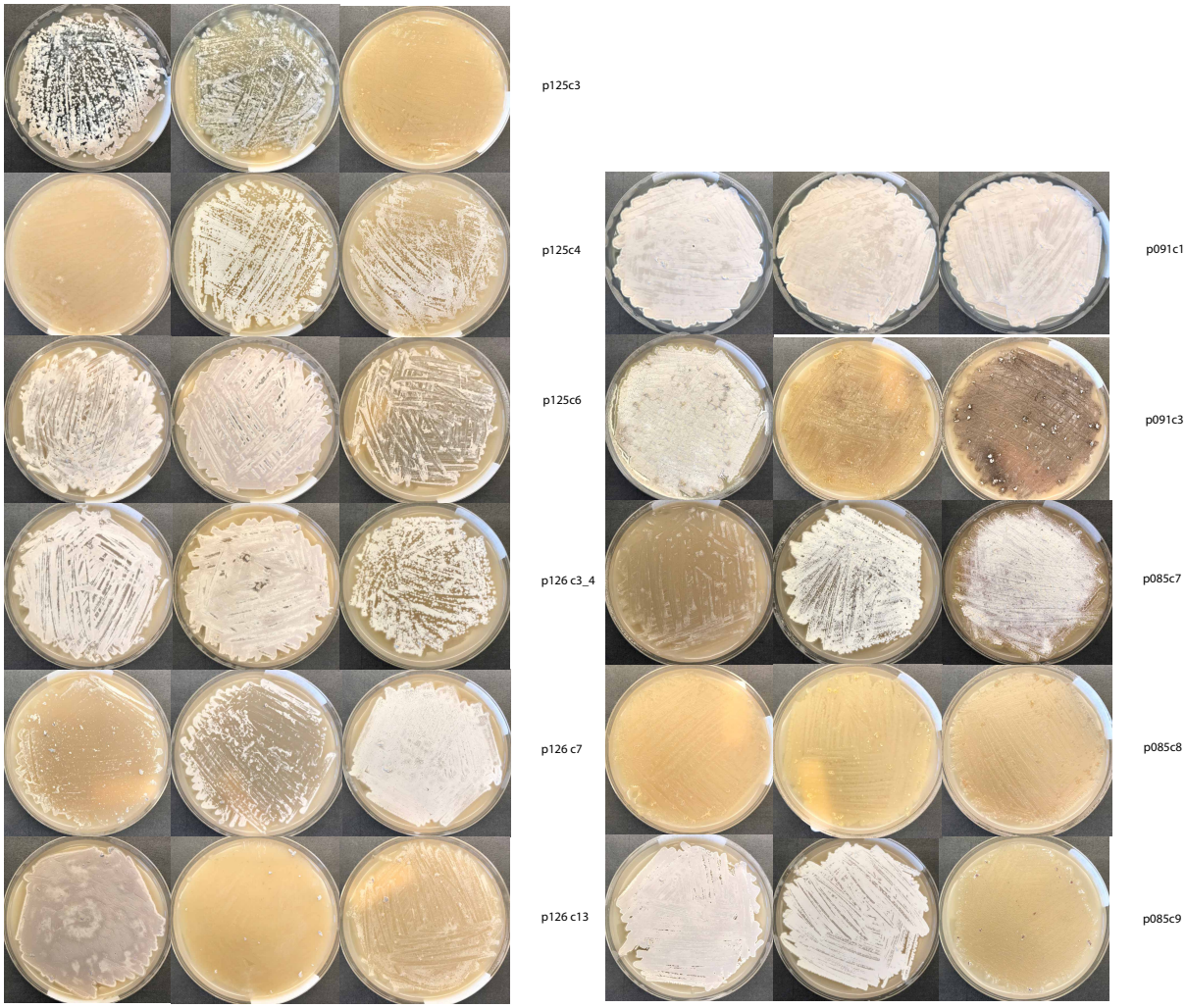
An overview of all heatmaps and detected SNPs in both the negative control and the mutant strains is shown in supplementary Fig. 3. A clear increase in the number of accumulated SNPs can be observed as the number of sgRNAs increases. However, these SNPs are primarily C to T or G to A mutations. The frequencies of other mutations do not differ noticeably from those observed in the wildtype.



Supplementary Figure 3: SNP heatmaps for all sequenced colonies before plasmid curing. A clear increase in the total number of C to T and G to A conversions can be observed as the number of sgRNAs increases. No substantial increase in other mutations was observed between mutant strains and wildtype replicates.

SI 5: Phenotypes of all cured strains

A total of 33 cured colonies were restreaked on MS plates, covering all plasmid variants. A great diversity in phenotypes was observed, most importantly absence of or weak sporulation, a colorless phenotype, and production of new compounds (Supplementary Fig. 4). All plates were fully grown. Three colonies each for 9sgRNAs, 18 sgRNAs, and 28 sgRNAs were selected for further characterizations based on their apparent reduction in production of specialized metabolites and greatly changed phenotypes.



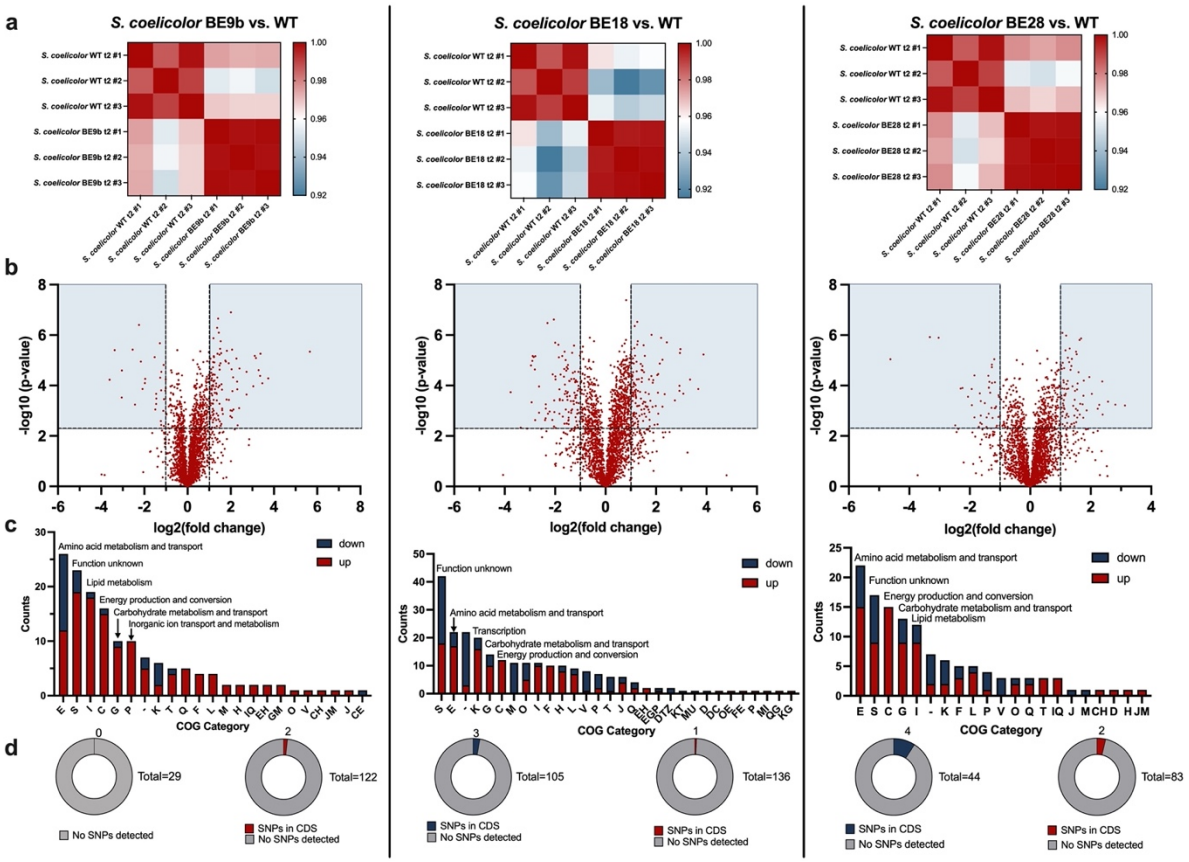
Supplementary Figure 4: Phenotypes of plasmid cured *S. coelicolor* mutants on mannitol soy flour medium. The number of the cured plasmid and the initial mutant colony the plasmid strains originated from are given for each row. Columns represent different colonies from a given plasmid cured strain. Clear differences in the phenotypes, especially in terms of sporulation, colony color, and metabolite production were observed.

SI 6: Proteomics data of three plasmid cured strains

To gain further insights into the systems wide effects of large multiplexed base editing experiments, samples were taken during cultivations at the highlighted time points (main test, Figs. 5 and 7) and prepared as described in the methods section. The cultivations were continued after sampling and the supernatant was used for metabolomics after one week.

Untargeted whole proteomics reveals changes to central cellular processes

Whole proteome comparison between the wild type and the base edited strains revealed many changes, primarily to central cellular processes. All proteomics analysis steps were performed with the shared core proteome across all samples, corresponding to complete datasets for 2401 out of 7548 CDS. Changes were especially observed between *S. coelicolor* BE18 and the wild type with whole proteome identities as low as 91.5 %. For further analysis genes were selected with at least an absolute 2-fold change in the observed peptide levels, as well as with a p value below 0.005 (to reduce the false positive rate). The hits falling into those spaces (highlighted in blue in Supplementary Fig. 5b) were then further characterized by annotation with the corresponding COG categories. The greatest changes were observed for all three strains in either amino acid metabolism and transport, or for genes with unknown function (Supplementary Fig. 5c). Further top categories were lipid metabolism, energy production and conversion, carbohydrate metabolism and transport, transcription, and inorganic ion transport and metabolism. Interestingly, in all three strains the number of proteins with higher observed abundances was greater than those with reduced abundances, indicating extensive rewiring and adaptation of the overall metabolism. Surprisingly, when matching the genes falling into the fold change-p-value space with the NGS data, only for a small fraction of the genes with highly changed protein abundances we could identify mutations in the coding sequences, suggesting that most of the observed changes are the result of *in trans* acting changes in regulatory cascades and downstream effects of mutations introduced in other loci.



Supplementary Figure 5: Whole proteome analysis of plasmid cured *S. coelicolor* BE9b, BE18, and BE28 strains. The shared core proteome was used as the basis for all analysis steps. **a**. Whole proteome correlation between mutant strains and wild type. Especially *S. coelicolor* BE18 showed great differences in the whole proteome correlation compared to the wild type. **b**. Volcano plots for each mutant. The chosen cutoffs were a log2(fold change) of 1/-1, and a p-value of 0.005 to reduce the number of false positives. The blue areas indicate the space between these two values. **c**. All genes falling into the defined spaces were categorized by COG categories. The highest-ranking categories were annotated. Most high-ranking COG categories were either part of the central metabolism, or of function unknown. **d**. The genes falling into the defined significance-fold change space were further matched with detected SNPs. Most of the highly changed genes did not have SNPs in the CDS, indicating that most changed were effects of wider regulatory changes. All shown data is based on three technical replicates. log2(fold change) of three technical replicates is shown. Significance was tested using unpaired two tailed t-tests, and p values are reported as -log10 (p-value).

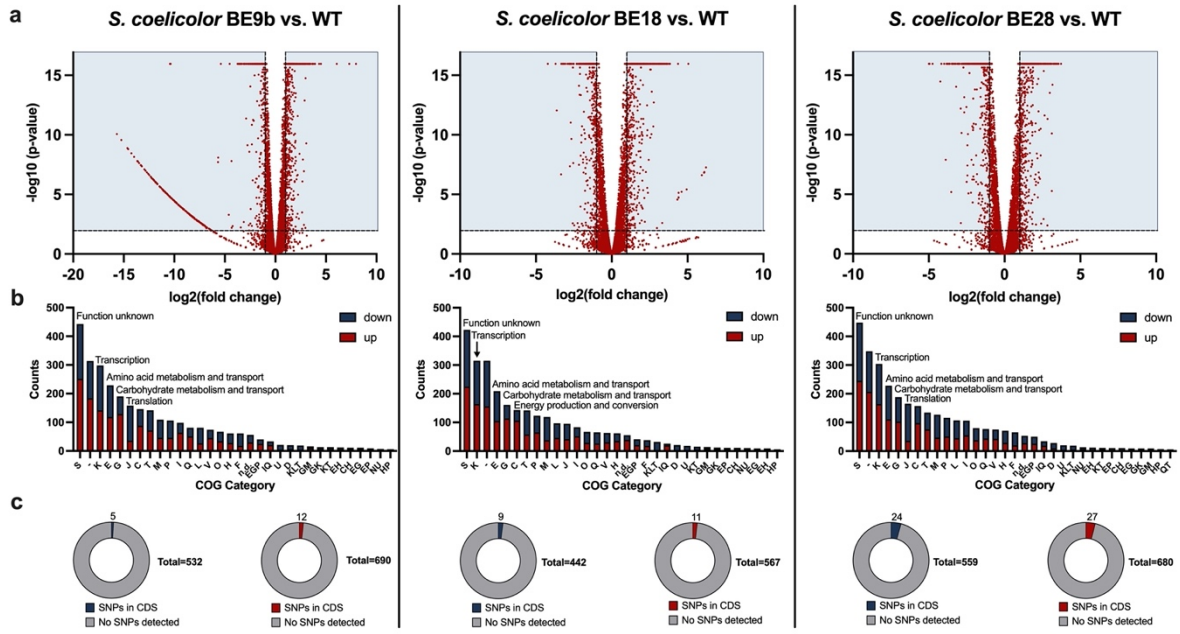
SI 7: Transcriptomics data of three plasmid cured strains

Following the same methodology, we analyzed transcriptomics data obtained from the same samples and time points. Base edits should not have a direct impact on the transcript level given that the effects of base editing only start having an effect during translation. However, we expected to see changes in the transcription levels likely resulting from feedback regulation. Further, we expected to be able to further validate the outcome of our proteomics analysis, showing great changes in central metabolism.

The whole transcriptome analysis overall mirrored the results of the proteomics analysis. All three mutants showed the greatest changes in COG categories belonging to central cellular processes and metabolism, including transcription, translation, amino acid metabolism and transport, carbohydrate metabolism and transport, as well as energy production and conversion. Function unknown was the top COG category for all three base editing strains for those genes falling into the defined fold change – p-value space. Matching these genes with the detected SNPs again revealed that the majority of changes could not be attributed directly to a SNP in a coding region. However, the percentage of matches between genes and SNPs increased from with the number of sgRNAs from BE9b to BE28.

Given the higher resolution of transcriptomics, we were able to obtain data for all targeted key biosynthetic genes (Supplementary Fig. 6b). We observed both positive and negative fold changes in expression in response to introduction of stop or rare codons. Given the overlapping sgRNA arrays, we were able to compare the fold changes of non-edited genes and edited genes. Examples of observed changes include FQ762_00875 with a rare TTG codon introduced in both BE18 and BE28, resulting in significant almost two-fold decrease om expression levels in both strains. Stop codons were introduced in FQ762_01290 in both BE18 and BE28, leading to significant reduction of expression levels in BE18, but to a significant increase in BE28, highlighting how the overall mutational landscape likely influences the cells response to gene inactivation through premature stop codon introduction. Other examples include FQ762_09505, which was inactivated through stop codon introduction in both BE18 and BE28, resulting in highly significantly reduced expression levels. FQ762_14260, the key biosynthetic gene of the desferrioxamine BGC, had

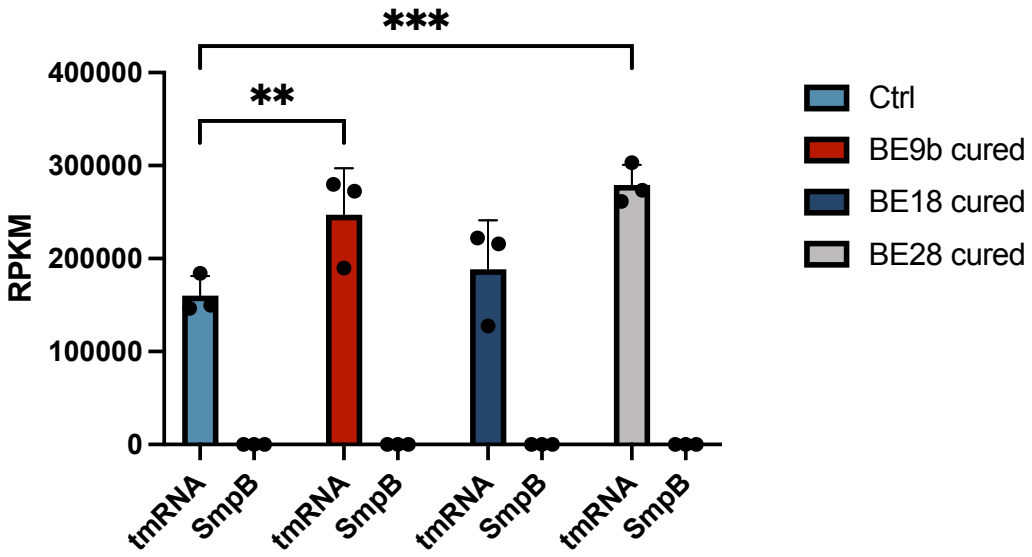
highly significantly reduced expression levels in all three strains, but was only edited in BE18 and BE28 through introduction two missense and one silent mutation, the latter of which resulting in the introduction of a rare CAA glutamine codon. These results highlight that not all stop codon introduction will also result in reduced expression levels. It is very likely that the cells response to gene-functionally detrimental base edits in key biosynthetic enzymes is not just influenced by possible feedback regulation (absence of functional enzyme resulting in increased expression), but also by trans-BGC metabolic adaptation. In all three mutants the majority of key biosynthetic genes showed reduced expression levels, indicating that the overall metabolic profile might also be reduced.



Supplementary Figure 6: Transcriptomics analysis of plasmid cured *S. coelicolor* BE9b, BE18, and BE28. a. Volcano plots for each mutant. Again, cutoff values of log2(fold change) of +1/-1 and a p-value of 0.005 were used. *S. coelicolor* BE9b appeared to have a large cluster of co-localized genes with high negative fold change values, indicating either a larger chromosomal deletion, or co-regulation. Given that no chromosomal deletion was detected in the NGS data, co-regulation appears to be most likely. **b**. COG categories of genes falling into the defined p-value-fold change space. The top categories identified in the transcriptomics data set mostly mirrors those identified in the proteomics data set. Function unknown is the dominant category for all mutants in the transcriptomics dataset, and translation was additionally identified in BE9b and BE28, but was absent from the top hits in the proteomics data. **c**. Number of genes within the defined p-value-fold change space to which SNPs could be assigned. The overall number of gene-SNP matches increased from BE9b to BE28. All shown data is based on three technical replicates. log2(fold change) of three technical replicates is shown. Significance was tested using unpaired two tailed t-tests, and p values are reported as -log10 (p-value).

SI8: tmRNA and SmpB RPKM values for the analyzed plasmid cured strains

Interestingly, we observed a significant increase in the RPKM values of tmRNA, which together with SmpB frees stalled ribosomes resulting from the absence of stop codons, matching tRNAs, or other translational problems (Supplementary Fig. 7)[6,7]. While these increases in RPKM values did not translate to significant fold changes, they did mirror the overall SNP profiles of the corresponding mutants, leading us to the hypothesis that the increased number of amino acid changes corresponds to an increased number of stalled ribosomes, potentially having further ripple effects given the intricate role this system plays in *Streptomyces*[6,7].



Supplementary Figure 7: Changes in tmRNA and SmpB RPKM values for *S. coelicolor* wildtype, BE9b, BE18, and BE28. While SpmB RPKM values were relatively constant between the different strains, tmRNA values differed significantly between wildtype and BE9b and BE28, suggesting increased occurrence of stalled ribosomes. Data of three technical replicates is shown. Significance was tested using unpaired two tailed t-tests, where *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

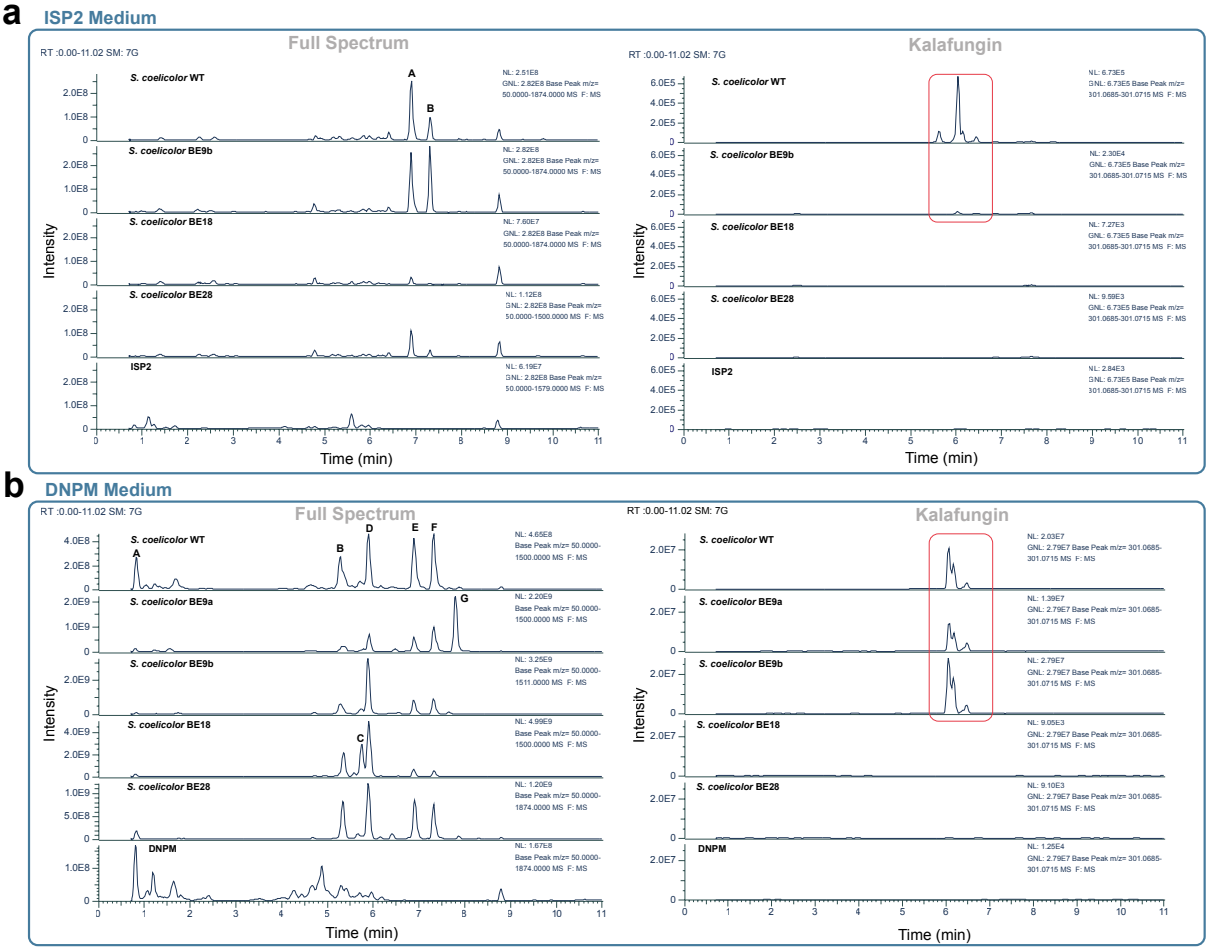
SI9: Metabolomics data for the analyzed plasmid cured strains

Finally, we performed untargeted metabolomics analysis of the ISP2 cultures from which samples were taken for both transcriptomics and proteomics analysis. Furthermore, we performed additional cultivations in DNPM production medium.

The major peaks observed were annotated in each chromogram (Fig. 8), with a focus on those that do not appear in the media controls. Cultivations of the BE strains in ISP2 medium did not reveal major changes compared to the wildtype cultivation with regards to dominant new or absent peaks. The two main peaks observed were germicidin B (A) and germicidin A (B) that were annotated through MS^2 matching using the GNPS library enriched with additional in-house annotated spectra. To highlight the effects of premature stop codon introduction on metabolite production, we further highlighted the spectrum for kalafungin. While we were not able to detect actinorhodin (which is typically quite difficult to produce in ISP2 medium), we putatively detected kalafungin, a precursor of actinorhodin biosynthesis[8]. Kalafungin was putatively detected only in significant amounts in the wild type, and in trace amounts in *S. coelicolor* BE9b. Interestingly, no major peak was detected in *S. coelicolor* BE18 and BE28, even though a stop codon was only introduced in BE28 and not BE18. It is worth mentioning, that the metabolomics acquisition was performed in positive mode, whereas kalafungin is better ionized and thus detected in negative mode.

We further performed cultivations in DNPM medium, a complex production medium[9]. The spectra for the DNPM cultivations were very different from the ISP2 spectra, and a higher number of major and differential peaks were detected. Peak A contains a matching mass-to-charge ratio (MS^1) for a chalcone, which was only detected in the wild type and in one replicate of *S. coelicolor* BE9b. In all other strains, replicates, and the media control, no MS^1 data that could match chalcone were observed. Peak B consists primarily of a compound with a *m/z* 517.33, with a putative SIRIUS[10] formula prediction for C23H41N5O7. The same peak is present in both *S. coelicolor* BE9a and BE9b, but not in BE18 and BE28. In *S. coelicolor* BE18 and BE28, the peak around 5.33 min consists primarily of desferrioxamine B with *m/z* 516.36. Peak C corresponds to desferrioxamine D2, which was highly abundant in the supernatant of *S.*

coelicolor BE18. Desferrioxamine E was detected in all samples with high relative abundance (contained in peak D), highlighting that the introduction of rare codons in the desferrioxamine BGC was not enough to abolish production and that the introduced mutations did not alter the cell's ability to produce desferrioxamines in high amounts. Peaks E and F were also observed in all samples and could again be mostly assigned to germicidin B and A, respectively.



Supplementary Figure 8: a. Qualitative LC-MS chromatograms for the base edited strains cultivated in ISP2 medium. Especially BE18 and BE28 showed clear differences to the wild type in terms of the observed intensities between the peaks. BE18 and BE28 also appeared to have a reduced metabolic profile. Further shown is the extracted ion chromatogram (EIC) for the m/z of kalafungin, a shunt product of actinorhodin. No detection was observed in BE18 and BE28, which is interesting, given that the key biosynthetic gene of the actinorhodin gene cluster was not edited in BE18, and only one silent mutation was introduced within the cluster. **b.** Qualitative LC-MS chromatograms of the base edited strains cultivated in DNPM medium. Included here is also the BE9a strain. Notice that global normalization was turned off to allow for better visualization of the peaks with greatly varying intensities. The EIC for the m/z of kalafungin showed more clearly the absence of production in the BE18 and BE28 strains, while clear production was observed in the wild type and BE9a and BE9b strains.

Finally, we detected a peak at a retention time of approximately 7.8 min (peak G) that only appeared in the supernatant of *S. coelicolor* BE9a. No MS^2 library hits were found for the dominant *m/z* 206.1538, but using GNPS[11] molecular networking established a connection to the deferrioxamine network. We hence hypothesized that the detected compound might be an intermediate or shunt product of the desferrioxamine biosynthesis.
Searching for the mass-to-charge of kalafungin revealed a putative production in the wild type, *S. coelicolor* BE9a, and *S. coelicolor* BE9b. No putative production was observed in *S. coelicolor* BE18 and BE28, mirroring the observations made in ISP2 medium.

SUPPLEMENTARY INFORMATION REFERENCES

(1) Fuhrmann, M.; Hausherr, A.; Ferbitz, L.; Schödl, T.; Heitzer, M.; Hegemann, P. Monitoring Dynamic Expression of Nuclear Genes in Chlamydomonas Reinhardtii by Using a Synthetic Luciferase Reporter Gene. *Plant Mol. Biol.* **2004**, *55* (6), 869–881. https://doi.org/10.1007/s11103-005-2150-1.

(2) Tong, Y.; Whitford, C. M.; Robertsen, H. L.; Blin, K.; Jørgensen, T. S.; Klitgaard, A. K.; Gren, T.; Jiang, X.; Weber, T.; Lee, S. Y. Highly Efficient DSB-Free Base Editing for Streptomycetes with CRISPR-BEST. *Proc. Natl. Acad. Sci.* **2019**, *116* (41), 20366–20375. https://doi.org/10.1073/pnas.1913493116.

(3) Myronovskyi, M.; Luzhetskyy, A. Native and Engineered Promoters in Natural Product Discovery. *Nat. Prod. Rep.* **2016**, *33* (8), 1006–1019. https://doi.org/10.1039/C6NP00002A.

(4) Muth, G. The PSG5-Based Thermosensitive Vector Family for Genome Editing and Gene Expression in Actinomycetes. *Appl. Microbiol. Biotechnol.* **2018**, *102* (21), 9067–9080. https://doi.org/10.1007/s00253-018-9334-5.

(5) Funa, N.; Ohnishi, Y.; Fujii, I.; Shibuya, M.; Ebizuka, Y.; Horinouchi, S. A New Pathway for Polyketide Synthesis in Microorganisms. *Nature* **1999**, *400* (6747), 897–899. https://doi.org/10.1038/23748.

(6) Yang, C.; Glover, J. R. The SmpB-TmRNA Tagging System Plays Important Roles in *Streptomyces Coelicolor* Growth and Development. *PLoS One* **2009**, *4* (2), e4459. https://doi.org/10.1371/journal.pone.0004459.

(7) Barends, S.; Zehl, M.; Bialek, S.; de Waal, E.; Traag, B. A.; Willemse, J.; Jensen, O. N.; Vijgenboom, E.; van Wezel, G. P. Transfer–Messenger RNA Controls the Translation of Cell-cycle and Stress Proteins in *Streptomyces*. *EMBO Rep.* **2010**, *11* (2), 119–125. https://doi.org/10.1038/embor.2009.255.

(8) Tahlan, K.; Ahn, S. K.; Sing, A.; Bodnaruk, T. D.; Willems, A. R.; Davidson, A. R.; Nodwell, J. R. Initiation of Actinorhodin Export in *Streptomyces Coelicolor*. *Mol. Microbiol.* **2007**, *63* (4), 951–961. https://doi.org/10.1111/j.1365-2958.2006.05559.x.

(9) Myronovskyi, M.; Brötz, E.; Rosenkränzer, B.; Manderscheid, N.; Tokovenko, B.; Rebets, Y.; Luzhetskyy, A. Generation of New Compounds through Unbalanced Transcription of Landomycin A Cluster. *Appl.*

Microbiol. Biotechnol. **2016**, *100* (21), 9175–9186. https://doi.org/10.1007/s00253-016-7721-3.

(10) Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A. A.; Melnik, A. V.; Meusel, M.; Dorrestein, P. C.; Rousu, J.; Böcker, S. SIRIUS 4: A Rapid Tool for Turning Tandem Mass Spectra into Metabolite Structure Information. *Nat. Methods* **2019**, *16* (4), 299–302. https://doi.org/10.1038/s41592-019-0344-8.

(11) Wang, M.; Carver, J. J.; Phelan, V. V; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V; Meehan, M. J.; Liu, W.-T.; Crüsemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderón, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C.-C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C.-C.; Yang, Y.-L.; Humpf, H.-U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; Boya P, C. A.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodríguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P.-M.; Phapale, P.; Nothias, L.-F.; Alexandrov, T.; Litaudon, M.; Wolfender, J.-L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D.-T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Müller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. Ø.; Pogliano, K.; Linington, R. G.; Gutiérrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N. Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34* (8), 828–837. https://doi.org/10.1038/nbt.3597.

Chapter 5

CASCADE-Cas3 Enables Highly Efficient Genome Engineering in *Streptomyces* Species

Christopher M. Whitford[1], Peter Gockel[1], David Faurdal[1], Tetiana Gren[1], Renata Sigrist[1], Tilmann Weber[1]

[1] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

Correspondence: tiwe@biosustain.dtu.dk

Published as a preprint on biorxiv: doi.org/10.1101/2023.05.09.539971

ABSTRACT

Type I CRISPR systems are widespread in bacteria and archaea. The main differences compared to more widely applied type II systems are the multi-effector CASCADE needed for crRNA processing and target recognition, as well as the processive nature of the hallmark nuclease Cas3. Given the widespread nature of type I systems, the processive nature of Cas3, as well as the recombinogenic overhangs created by Cas3, we hypothesized that Cas3 would be uniquely positioned to enable efficient genome engineering in streptomycetes. Here, we report a new type I based CRISPR genome engineering tool for streptomycetes. The plasmid system, called pCRISPR-Cas3, utilizes a compact type I-C CRISPR system and enables highly efficient genome engineering. pCRISPR-Cas3 outperforms pCRISPR-Cas9 and facilitates targeted and random-sized deletions, as well as substitutions of large genomic regions such as biosynthetic gene clusters. Without additional modifications, pCRISPR-Cas3 enabled genome engineering in several *Streptomyces* species at high efficiencies.

INTRODUCTION

Streptomycetes encode large untapped biosynthetic potential in cryptic BGCs

Streptomycetes and other filamentous actinomycetes represent some of the most gifted producers of complex specialized metabolites, with a wide diversity of applications in medicine, industry, and agriculture. Most importantly, many of the commonly used antibiotics are produced by members of the genus *Streptomyces* [1]. Many of these specialized metabolites play important ecological roles for the producer strains, to fend off competition for nutrients, communication, as well as for complex symbiotic relationships [2]. Due to the complex role specialized metabolites play in the natural environment, the expression of the corresponding biosynthetic gene clusters (BGC) is tightly regulated, and natural elicitors often remain unknown. Therefore, a majority of BGCs are not readily expressed under standard laboratory conditions, hindering the identification of the encoded specialized metabolites. While classical chemical screening resulted in the identification and utilization of high numbers of readily expressed BGCs, the advent of cheap whole genome sequencing revealed a large untapped potential of silent BGCs encoding the biosynthetic information for yet unknown specialized metabolites [3,4].

Expression of BGCs in engineered heterologous hosts is common practice for BGC activation

To overcome these limitations and to advance the exploration of the incredible biosynthetic potential of streptomycetes, several methods were developed to achieve expression of silent BGCs. Expression of silent BGCs in a heterologous host represents one of the commonly used approaches [5,6]. Since many actinomycetes are not (yet) genetically tractable, expression of BGCs cloned from isolated genomic DNA allows the study of BGCs independently of their source. Furthermore, using established expressions hosts streamlines experimental work, as no new methods need to be developed and each experiment has the potential to contribute to a growing knowledgebase. Commonly used hosts for the expression of heterologous BGCs include various

derivatives of *S. coelicolor*, *S. albidoflavus*, *S. lividans*, *S. avermitilis*, and *S. venezuelae* [7–15].

Genome reduction is desired for reduced metabolic background activity

To increase the success rate of heterologous expression of BGCs, hosts are usually engineered with a focus on three aspects. First, to aid the identification of compounds in metabolomics analysis, a simplified metabolic background is desired. This can be achieved through genome reduction with a special focus on native, readily expressed BGCs, as well as those usually silent under laboratory conditions. Second, to increase the chances of successful heterologous expression, the number of integration sites can be increased, to facilitate multicopy expression of the target BGC. Lastly, increasing the supply of precursors is also of interest to boost the production of a matching BGC. Examples of hosts constructed using these principles are *S. albidoflavus* BE4, *S. lividans* ΔYA11, or *S. avermitilis* SUKA22 [7,8,14,16].

Existing genome engineering methods for streptomycetes

A wide number of tools exist that can be used to achieve above mentioned host engineering [17]. Classically, genomic deletions were achieved through PCR targeting, which is based on homologous recombination and double crossovers [18]. While this method works in many *Streptomyces* strains, it requires having the editing site cloned in a cosmid, fosmid, or BAC and thus still is very labor and time intensive. Newer methods usually make use of CRISPR effectors, such as Cas9 or Cas12a. Several CRISPR-Cas9 based tools were developed for application in streptomycetes and were shown to facilitate small to medium-sized deletions with good efficiencies. Common vectors include pCRISPomyces, pCRISPR-Cas9, or pKCcas9dO[19–21]. Frequently observed toxicity of Cas9, especially when strongly expressed, has since led to the development of optimized plasmids allowing tighter control of the expression of Cas9 [22]. pKCCpf1 was developed to enable Cas12a (Cpf1) based deletions of one or two genomic loci, based on crRNA processing by Cas12a of its own crRNA array [23]. Cas12a further recognizes a T-rich PAM sequence, lowering the potential of off-target effects in the GC-rich streptomycete genome. Recently, base editing was developed for single and multiplexed inactivation of genes of interest through the introduction of premature stop codons [24]. While this

method has many benefits, the targeted genes remain present in the genome, and off-targets are more tolerated due to the absence of DSB introduction.

CASCADE-Cas3

Type I CRISPR systems are among the most widespread CRISPR systems. The signature effector of type I systems is the nuclease Cas3, which requires the assembly of a ribonucleoprotein complex called CASCADE, usually formed by Cas8, Cas7, Cas5, and the crRNA, to bind and cut DNA [25]. Hence, CASCADE Cas3 is a multi-effector CRISPR system that differs significantly from single-effector CRISPR systems like Cas9. The most widely studied type I systems are type I-E, of which the *Escherichia coli* system has been studied in great detail [26,27]. Type I and II CRISPR systems not only differ in their overall architecture, but also in the mode of action of the signature nuclease. While Cas9 acts only as an endonuclease by introducing double-strand breaks at the target site, Cas3 has a combined ATP-dependent nuclease-translocase activity and starts degrading DNA processively from 3' to 5' after cutting [28]. This results in the formation of long 5' overhangs. CASCADE Cas3 has previously been applied for genome engineering, but the large size of the complex has hindered widespread application using plasmid-based systems [29,30]. Recently, Csörgő et al. described a compact type I-C system and its application for plasmid-based genome engineering in *E. coli, Pseudomonas syringae,* and *Klebsiella pneumoniae* [31]. The described system requires only 4 genes (*cas3, cas5, cas8,* and *cas7*) to introduce efficient genomic deletions of up to half a megabase. The system uses a T-rich 5'-TTC-3' PAM sequence, making it an attractive system to use in high GC organisms. Furthermore, it was shown that homology-directed repair in combination with CASCADE Cas3 introduces genomic modifications with higher efficiencies than Cas9, likely due to the recombinogenic nature of the ssDNAse activity of Cas3.

pCRISPR-Cas3 is a novel tool for efficient genome engineering of streptomycetes

Here, we present a new CASCADE Cas3 based tool for streptomycetes facilitating highly efficient genomic deletions and integrations based on a previously published compact type I-C CRISPR system. We adapted the system for expression in streptomycetes and integrated it into our established CRISPR plasmid platform. We demonstrate highly efficient deletions of small, mid,

and large-sized deletions, and show that pCRISPR-Cas3 can facilitate deletions in several commonly used *Streptomyces* hosts. We show how pCRISPR-Cas3 can facilitate simultaneous deletions and integrations with superior efficiencies, allowing streamlined genome engineering in even recalcitrant *Streptomyces* strains. Finally, we demonstrate the application of genome engineering with pCRISPR-Cas3 by construction of a *S. coelicolor* expression host.

RESULTS

Distribution of Type I CRISPR Systems in Streptomycetes

Type I CRISPR systems are widespread in bacteria and archaea. However, only a few systems from *Streptomyces* have been characterized and described in detail so far [32]. To investigate the distribution of type I CRISPR systems vs. type II CRISPR systems in streptomycetes, we performed BLAST searches against 2401 high-quality publicly available *Streptomyces* genomes. The amino acid sequence of Cas9 from *Streptococcus pyogenes* and of Cas3 of the type I-E CRISPR system of *S. albidoflavus* (previously *albus*) J1074 were used as queries, as these are the hallmark nucleases from the two CRISPR types. Based on our search, type I CRISPR systems appear to be much more widespread in *Streptomyces* compared to type II CRISPR systems. Only two hits were obtained for *Spy*Cas9, while *Salb*Cas3 returned almost 1400 hits (Fig. 1). Of these, over 100 had a sequence similarity >50 % (Supplementary Information Table 1). Cas3 acts as a nuclease-translocase with an N-terminal HD phosphohydrolase and C-terminal helicase domain [33]. To obtain a more granular view of the distribution of type I systems in streptomycetes, and to reduce the number of false positive hits due to similarities to helicases, we performed another search with cblaster, which allows the search for clustered homologous sequences [34]. Using cblaster and the CASCADE from *S. albidoflavus* J1074 as input, we identified 472 strains carrying the entire CASCADE, comprising *cas3, casA, casB, cas7, cas5, and cas6* (Supplementary Information Table 2). Based on the high abundance of type I CRISPR systems in streptomycetes, we hypothesized that they might have a higher tolerance to genome engineering using type I CRISPR systems, as opposed to type II systems.

Construction of pCRISPR-Cas3, a plasmid-based compact CASCADE-Cas3 system

The previously characterized and utilized type I-C CRISPR system from *Pseudomonas aeruginosa* consists of only four genes, *cas3, cas5, cas8,* and *cas7,* totaling 5.6 kb. This allows plasmid-based expression of the system and application across various species. All four genes are arranged in an operon and can hence be expressed from a single promoter.

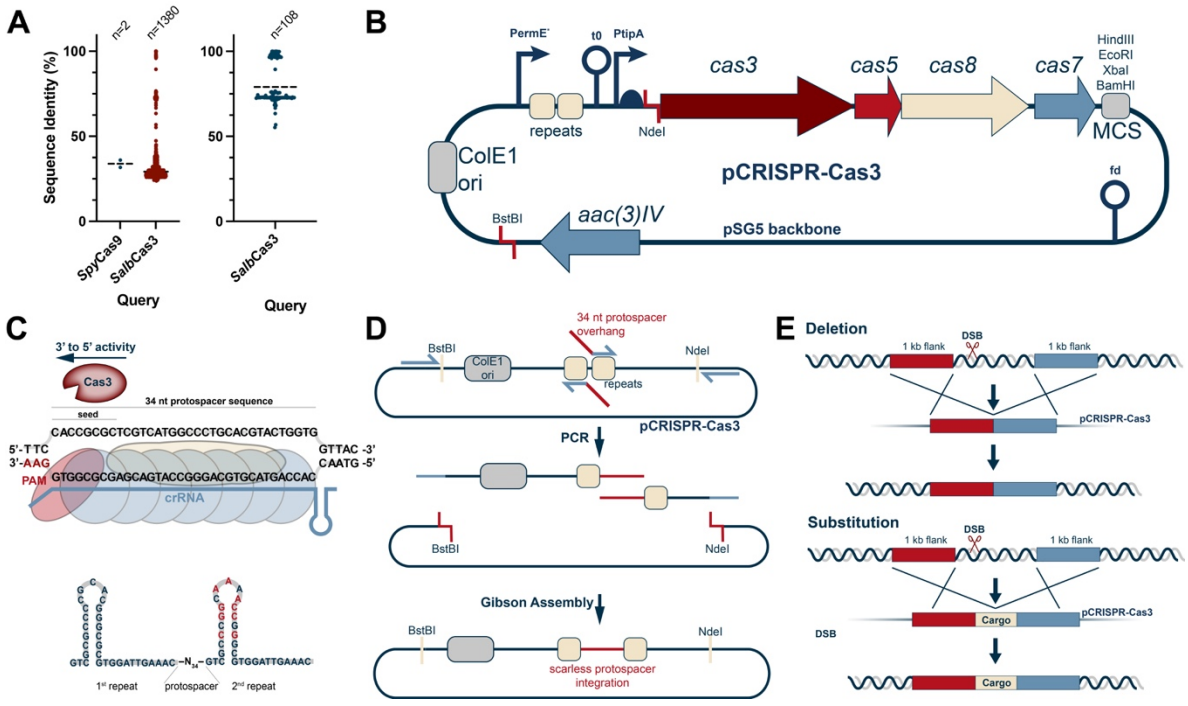


Figure 1: pCRISPR-Cas3 enables streamlined genome engineering of streptomycetes. **(A)** Distribution of type I and type II CRISPR systems in streptomycetes using *Spy*cas9 and *cas*3 from *S. albidoflavus* as references. The BLAST search was run on all high-quality publicly available *Streptomyces* genomes (n=2401). Type I CRISPR systems appear to be much wider distributed than type II CRISPR systems in streptomycetes. **(B)** Plasmid map of pCRISPR-Cas3. The plasmid is based on the pSG5 replicon and carries the codon-optimized type I-C minimal CASCADE under control of the inducible *tip*A promoter. The crRNA is cloned between two repeats in the cRNA cassette, which is controlled by the constitutive *ermE** promoter. Repair templates are cloned into a MCS in the backbone of the plasmid. **(C)** The second repeat has a modified sequence to prevent recombination between the two repeats. The CASCADE complex comprised of Cas5, Cas8, and Cas7 units binds the target sequence and recruits Cas3. Cas3 has a 3'-5' helicase nuclease activity, resulting in directionally biased deletions. **(D)** cRNAs are cloned between two repeats, posing some challenges due to sequence homologies. Since type IIS restriction enzymes cannot be used in high GC *Streptomyces* plasmids, a PCR and Gibson Assembly based cloning approach was established, allowing cloning of cRNAs with high efficiencies. **(E)** pCRISPR-Cas3 can be used for targeted deletions of large genomic regions, or for substitutions of such with a specified cargo. It can also be used for random-sized deletion experiments.

Given the differences in both codon usage and GC content, the previously published operon was codon optimized and synthesized. For codon optimization, the *S. coelicolor* codon usage table was used, as such optimized constructs have previously been successfully expressed in a wide variety of different streptomycetes [35]. The upstream region of each *cas* gene was designed

with a canonical RBS sequence (GGAGG or GGAGC). The 5.6 kb fragment was subsequently synthesized and delivered as a cloned plasmid. For expression in streptomycetes, our established CRISPR platform based on the pSG5 replicon was used. The Cas9 cassette from pCRISPR-Cas9 was removed by digestion with NdeI and HindIII and replaced by the synthesized CASCADE-Cas3 operon using Gibson Assembly. The Cas9 gRNA expression cassette was subsequently replaced by a dsDNA fragment containing the Cas3 repeats following digestion of the plasmid with NcoI and SnaBI and subsequent Gibson Assembly. The second repeat was modified as described by Csörgő et al. to prevent recombination events between the two repeats (Fig 1c). The resulting plasmid was named pCRISPR-Cas3 in accordance with our previous plasmid nomenclature.

Given the repetitive nature of the Cas3 crRNA hairpins, cloning spacer sequences in between these proved to be very challenging. Attempts to clone protospacer sequences using ssDNA oligo bridging by digestion of pCRISPR-Cas3 with NcoI (cutting between the two repeats) failed repeatedly. Given that common type IIS restriction enzymes such as BsaI and BstBI, commonly used for such cloning scenarios, cannot be used in high GC plasmids due to their omnipresent recognition sequences further complicated this cloning challenge. To achieve efficient cloning of user-defined protospacer sequences between the CRISPR repeats, we hence had to design an alternative cloning strategy. Using two fixed primers binding up and downstream from the protospacer integration site, two PCRs are set up using two primers binding one of the repeats each and carrying the desired 34 nt protospacer sequence in the overhangs. Digestion of pCRISPR-Cas3 with BstBI and NdeI removes the origin of replication, ensuring that only correctly assembled plasmids can replicate. Using this cloning method, we frequently achieved over 90 % cloning efficiencies (Supplementary Information Figure 1).

pCRISPR-Cas3 introduces highly efficient deletions with single nucleotide precision in S. coelicolor

To demonstrate the application of pCRISPR-Cas3 for genome engineering in streptomycetes, we attempted to delete the well-characterized 22 kb actinorhodin BGC in *S. coelicolor*. Deletion of the actinorhodin BGC leads to a clear red phenotype in *S. coelicolor*, resulting from expression of the

undecylprodigiosin BGC and enabling phenotypical screening for deletion mutants. Three different protospacers were selected, in the middle and towards each edge of the deletion region, to investigate potential differences in editing outcomes based on protospacer positions. Protospacers for pCRISPR-Cas3 were manually designed based on PAM+seed sequence queries within the deletion region, followed by genome-wide queries using both the seed and whole length sequences of putative protospacers to identify the most specific candidates. In *S. coelicolor,* 158,341 5'-TTC-3' PAM sequences were found, averaging one PAM for every 54.7 bp. In contrast, 1,574,641 5'-NGG-3' PAM sequences were found for Cas9, averaging one for every 5.5 bp, and highlighting the presence of up to an order of magnitude more potential off-target sites. Two flanking regions of 1 kb were selected as repair templates and cloned into pCRISPR-Cas3 using restriction cloning as described in the methods section. To compare editing outcomes to those achieved with Cas9, the same repair templates were cloned into pCRISPR-Cas9, and three different sgRNAs mirroring the positions of the Cas3 protospacers were selected. In parallel, all experiments were performed without repair templates for both pCRISPR-Cas3 and pCRISPR-Cas9.

Stark differences were observed between the different editing configurations. pCRISPR-Cas9 without repair templates failed to produce the distinct red phenotype, and with the "right" protospacer, only a handful of viable exconjugants were obtained on selective medium. Given that the pCRISPR-Cas9 plasmid without ScaLigD was used, higher toxicity was expected. As shown previously, coexpression of ScaLigD can greatly enhance NHEJ and reduce the toxicity in *S. coelicolor* [20]. pCRISPR-Cas3 without repair templates resulted in greatly varying outcomes, depending on the used protospacer. Both protospacers close to the edges of the deletion region appeared to be toxic. However, both surviving clones of the right protospacer had a weak red phenotype. pCRISPR-Cas3 without repair templates and the protospacer in the middle of the deletion region did not result in an observable loss in viability, but also did not produce a clear red phenotype. Clear deletion phenotypes were obtained with pCRISPR-Cas9 with repair templates and the spacer located in the middle of the deletion region. However, with the other two protospacers, either no clear deletion phenotypes were obtained, or only a small number of viable exconjugants could be obtained, which did not have the desired phenotype. The observed toxicities of pCRISPR-Cas9 might be predominantly

protospacer-specific. However, recent studies have shown that binding to off-target sites might be much more extensive, stabilized by as little as a few nucleotides of homology [36]. Using pCRISPR-Cas3 together with repair templates, clear deletion phenotypes were obtained for all spacers. With protospacers located towards the edges of the deletion region, clear deletion phenotypes were obtained for all picked exconjugants.

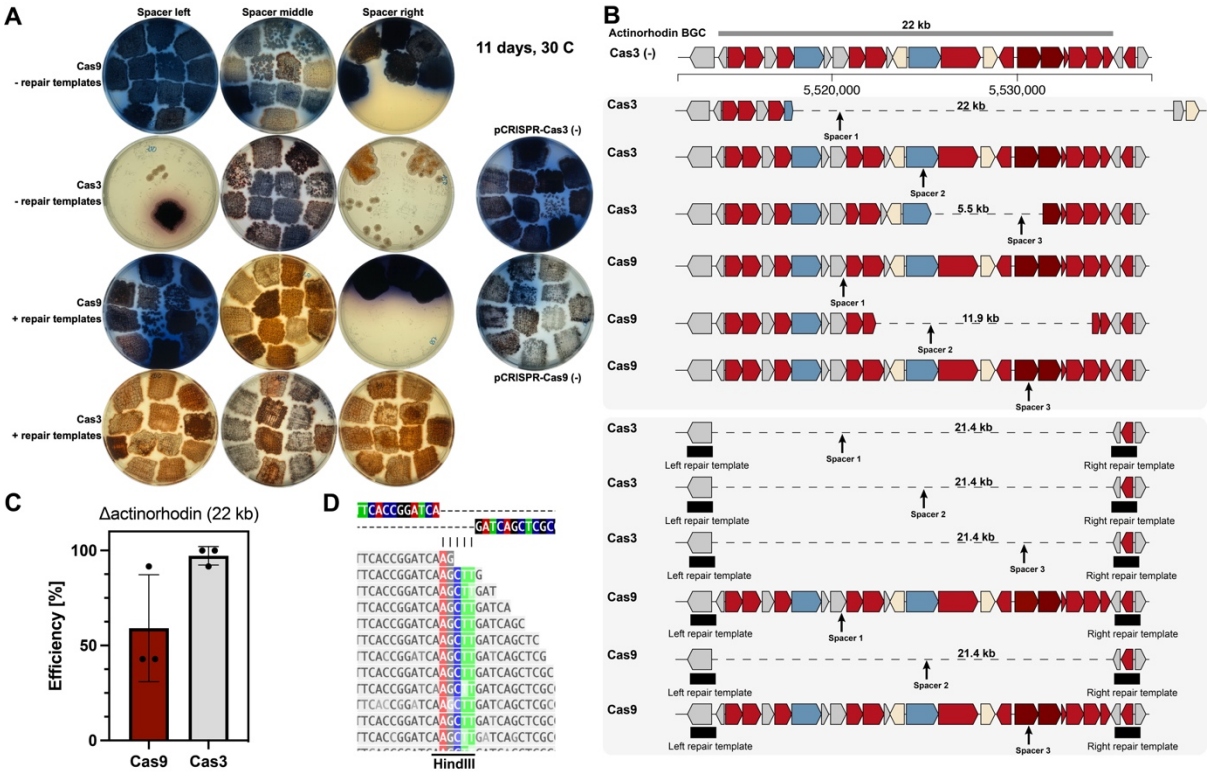


Figure 2: pCRISPR-Cas3 introduces genomic deletions with higher efficiencies than pCRISPR-Cas9. (A) Plate pictures of *S. coelicolor* mutants harboring pCRISPR-Cas9 or pCRISPR-Cas3 with protospacers targeting the actinorhodin BGC in three different locations, and with or without repair templates. pCRISPR-Cas3 displayed higher toxicity without repair templates, but resulted in more exconjugants overall and with the desired RED phenotype once repair templates were provided. **(B)** Representation of sequencing results of selected colonies, both for pCRISPR-Cas3 and pCRISPR-Cas9 with and without repair templates. Both pCRISPR-Cas3 and pCRISPR-Cas9 introduced random-sized deletions without repair templates. With repair templates, precise deletions were observed for both pCRISPR-Cas3 and pCRISPR-Cas9. (**C)** Efficiencies for actinorhodin deletions with pCRISPR-Cas9 and pCRISPR-Cas3. For pCRISPR-Cas3, the efficiencies were consistently high (> 90 %), while with pCRISPR-Cas3 the observed efficiencies were highly dependent on the used protospacer. **(D)** Read alignments for the junction site of the two homologous flanks. A HindIII site was integrated, demonstrating that the double-strand break was repaired using the repair templates cloned into pCRISPR-Cas3 using HindIII. Shown in (**C**) are the means ± standard deviations of three deletion experiments targeting the actinorhodin region with three different protospacers.

Only 5 out of 12 exconjugants showed the expected phenotype when the protospacer in the middle of the deletion region was used, with the remaining 7 having a greyish phenotype. Interestingly, these results appeared to be mirrored to pCRISPR-Cas9, which performed best with the protospacer located in the middle of the deletion region.

Phenotypical screening revealed an apparent superiority of pCRISPR-Cas3 compared to pCRISPR-Cas9, as all three protospacer configurations resulted in the expected deletion phenotype with high efficiencies. To verify these observations and to obtain more precise numbers for efficiencies of pCRISPR-Cas9 and pCRISPR-Cas3, the experiments with protospacers and repair templates were repeated. This time, conjugations were performed at larger scales to obtain more viable clones for pCRISPR-Cas9 with the protospacer located in the right flank of the deletion region. The deletion efficiencies were analyzed by performing colony PCRs on the targeted region with a primer binding inside of the repair template, and one binding outside of the repair template in the neighboring sequence (Supplementary Information Figure 2). For pCRISPR-Cas9, average deletion efficiencies of 60 % were obtained. These varied greatly depending on which protospacer was used. For pCRISPR-Cas3, an average efficiency of 97 % was obtained. The obtained efficiencies for pCRISPR-Cas3 were consistently high for all protospacers, suggesting a lower dependency on the protospacer sequence and location to achieve efficient deletions.

To verify the PCR results, Illumina-based whole genome sequencing was performed for one colony of each configuration (Fig. 2B). Random sized deletions were observed for both pCRISPR-Cas9 and pCRISPR-Cas3 when no repair template was provided. For pCRISPR-Cas3, the observed random-sized deletion were 5.5 kb and 22 kb in size, and for pCRISPR-Cas9 11.9 kb. No random sized deletions were observed for pCRISPR-Cas3 with the mid protospacer, and none for pCRISPR-Cas9 when the protospacers at the edges of the deletion region were used. Sequencing of repair template containing configurations showed that pCRISPR-Cas3 successfully introduced the designed mutations with single nucleotide prevision with all tested configurations. For pCRISPR-Cas9, only the protospacer located in the middle of the deletion region resulted in successful deletion of the designed region for the screened colonies.

As pCRISPR-Cas3 was able to introduce the designed deletions with robust efficiencies and single nucleotide precision, we next investigated whether pCRISPR-Cas3 could be used to install targeted deletions in other important *Streptomyces* species.

Genomic deletions were achieved using pCRISPR-Cas3 in S. venezuelae, S. albidoflavus, S. sp. NBC01270

Given the successful demonstration in *S. coelicolor*, we next attempted to install designed deletions in other strains of interest. Therefore, we selected *S. venezuelae* (ATCC 10712), an emerging production host and model species, *S. albidoflavus* J1074, a well-established host for expression of heterologous BGCs, and *S. sp.* NBC1270, an isolate from our own strain collection closely related to *S. albidoflavus* J1074.

In *S. venezuelae* ATCC 10712, region 22, as predicted by antiSMASH 6.0.1, was selected as a target for demonstration, given its size of 122 kb, and the high density of putative BGCs. Repair templates of 1 kb on each side of the target region were designed. Two protospacers were designed, one in the middle of the deletion region, and one approximately 5 kb from the left edge. Both combinations of repair template and protospacer installed the desired deletion with high efficiencies, with 9 out of 12, and 10 out of 12 exconjugants carrying the designed mutation, respectively (Fig. 3). The PCR-based screening results were further verified by whole genome sequencing, showing a clear drop in coverage of region 22.

Given the positive results in *S. venezuelae*, we next attempted to perform deletions in *S. albidoflavus* J1074 and the closely related strain *S. sp.* NBC1270. The strains are closely related and share the majority of BGCs as predicted by antiSMASH 6.0.1. At the far end of the chromosomal arm, both strains carry the same dense accumulation of 3 BGCs for production of antimycin, candicidin, and flaviolin. Given the total size of 318 kb, this BGC dense region was also interesting in order to investigate the possibility to delete hundreds of kb in a single step. Following several failed attempts to obtain the bands indicating successful introduction of the designed chromosomal deletion, some of the colonies for which no PCR bands were obtained were sequenced using Oxford Nanopore sequencing.

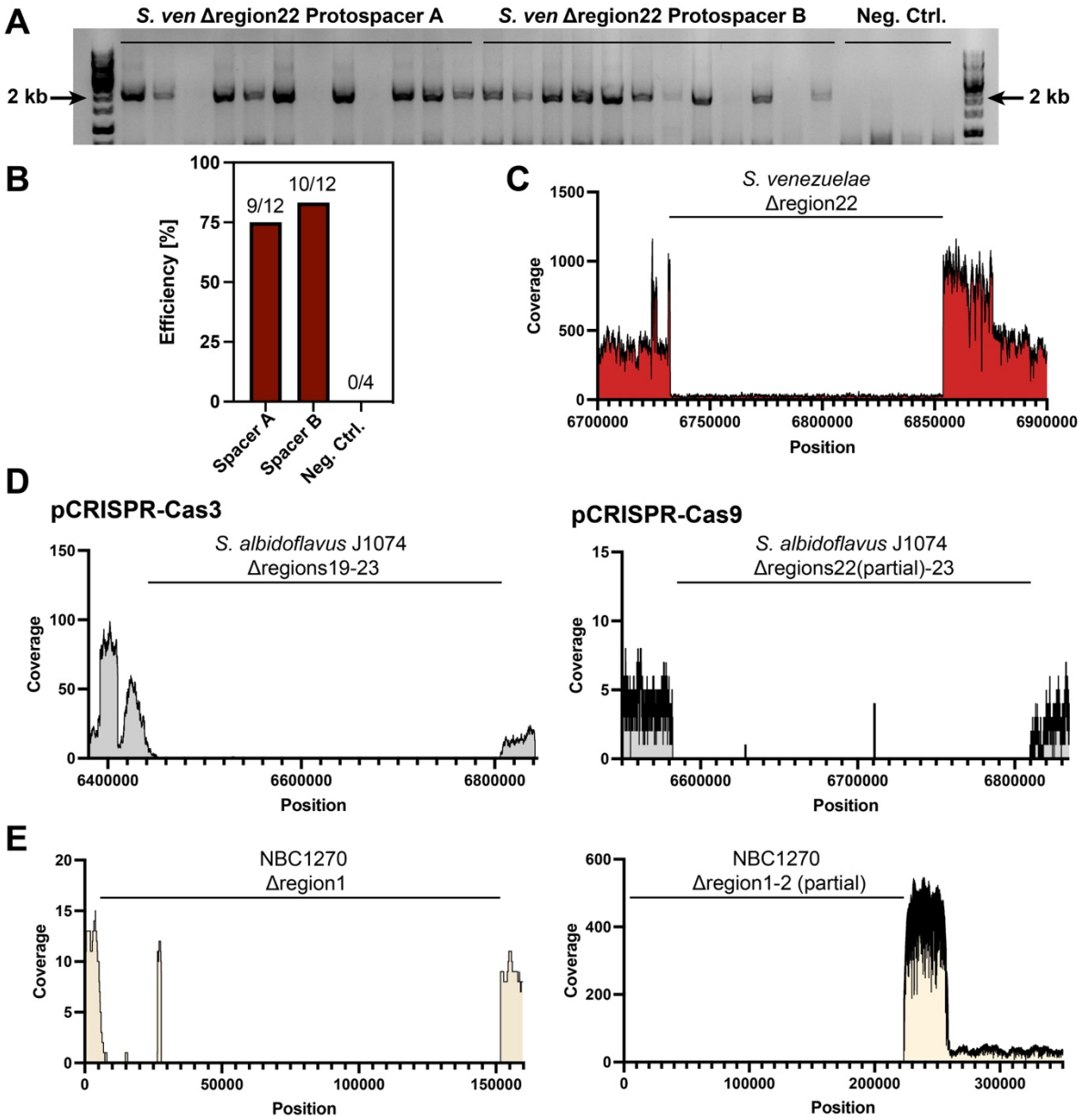


Figure 3: Application of pCRISPR-Cas3 in other *S. venezuelae*, *S. albidoflavus* J1074, and *S. sp.* NBC1270. (A) Gel picture for screening of successful deletions of the 122 kb region 22 in *S. venezuelae*. Bands were expected slightly above the 2 kb marker, as the primers were binding just outside of the repair templates. **(B)** pCRISPR-Cas3 was able to introduce the desired deletions with high efficiencies for both protospacers. **(C)** Illumina-based sequencing of a deletion mutants showing a clear and precise deletion of the targeted region. **(D)** In *S. albidoflavus* J1074 random-sized deletions were detected following targeting of regions 21-23 with pCRISPR-Cas3. However, similar random sized deletions were obtained when targeting the same region with the same repair templates using pCRISPR-Cas9. **(E)** In *S. sp.* NBC1270, a *Streptomyces* strain with high similarity to *S. albidoflavus* J1074, random sized deletions were obtained with pCRISPR-Cas3 while preserving the terminally inverted repeats. With pCRISPR-Cas9, targeting the same region resulted in random sized deletions and complete loss of the chromosomal end. A 33 kb sequence likely got replicated up to 12 times to prevent continuous degradation of the chromosomal end.

In both *S. albidoflavus* J1074 and *S. sp.* NBC1270, targeting of the far chromosomal end resulted in unspecific deletions approximately 380 kb and 140 kb in size, respectively. In *S. albidoflavus* J1074, this resulted in deletion of antiSMASH regions 19-23. Targeting of the same region with pCRISPR-Cas9 also resulted in extended deletions, suggesting that this was a general problem resulting from introducing double-strand breaks at the ends of the chromosomal arms [37]. Interestingly, targeting the same region in NCB1270 using pCRISPR-Cas9 resulted in the loss of the entire chromosomal end, as no reads were mapped against the terminally inverted repeats. A big spike in coverage was observed just at the end of the deletion region, suggesting that this 33 kb sequence stretch was replicated around 12 times to prevent continuous degradation of the chromosomal end. *De novo* assembly of the genome and subsequent visualization of the assembly graph further confirm this hypothesis (Supplementary Information Fig. 3). Both strains did not display any growth defects, and all desired BGCs were deleted in *S. albidoflavus* J1074, despite the unspecific nature of the deletion. These results demonstrate that pCRISPR-Cas3 can be used to introduce random sized deletions, especially when coupled with subsequent screening for desired geno- and phenotypes. Our results highlight that pCRISPR-Cas3 can enable efficient genome engineering even in non-model *Streptomyces* species.

Simultaneous deletions and integrations can be achieved through modification of the repair templates

Most sophisticated engineered *Streptomyces* hosts have not only a reduced metabolic background, achieved through deletion of readily expressed BGCs, but also added integration sites for either multicopy integrations or site-directed targeted integrations using different integrases [7,38,39]. Consequently, we wanted to demonstrate how pCRISPR-Cas3 can be used to delete genomic regions such as BGCs, and to simultaneously integrate additional integration sites instead. The *Streptomyces* bacteriophage PhiC31 integrase is a well-established system for integration of heterologous sequences in various *Streptomyces* species [40,41]. The consensus PhiC31 attB site from *S. coelicolor* [42] was chosen as cargo and cloned it in between the two flanks. As a proof of concept, we modified plasmid p129, previously used to delete the actinorhodin BGC in *S. coelicolor,* to carry the attB site between the two homologous repair template

arms. The obtained sequencing reads mapped perfectly against the *in silico* modified reference sequence (Fig. 4).

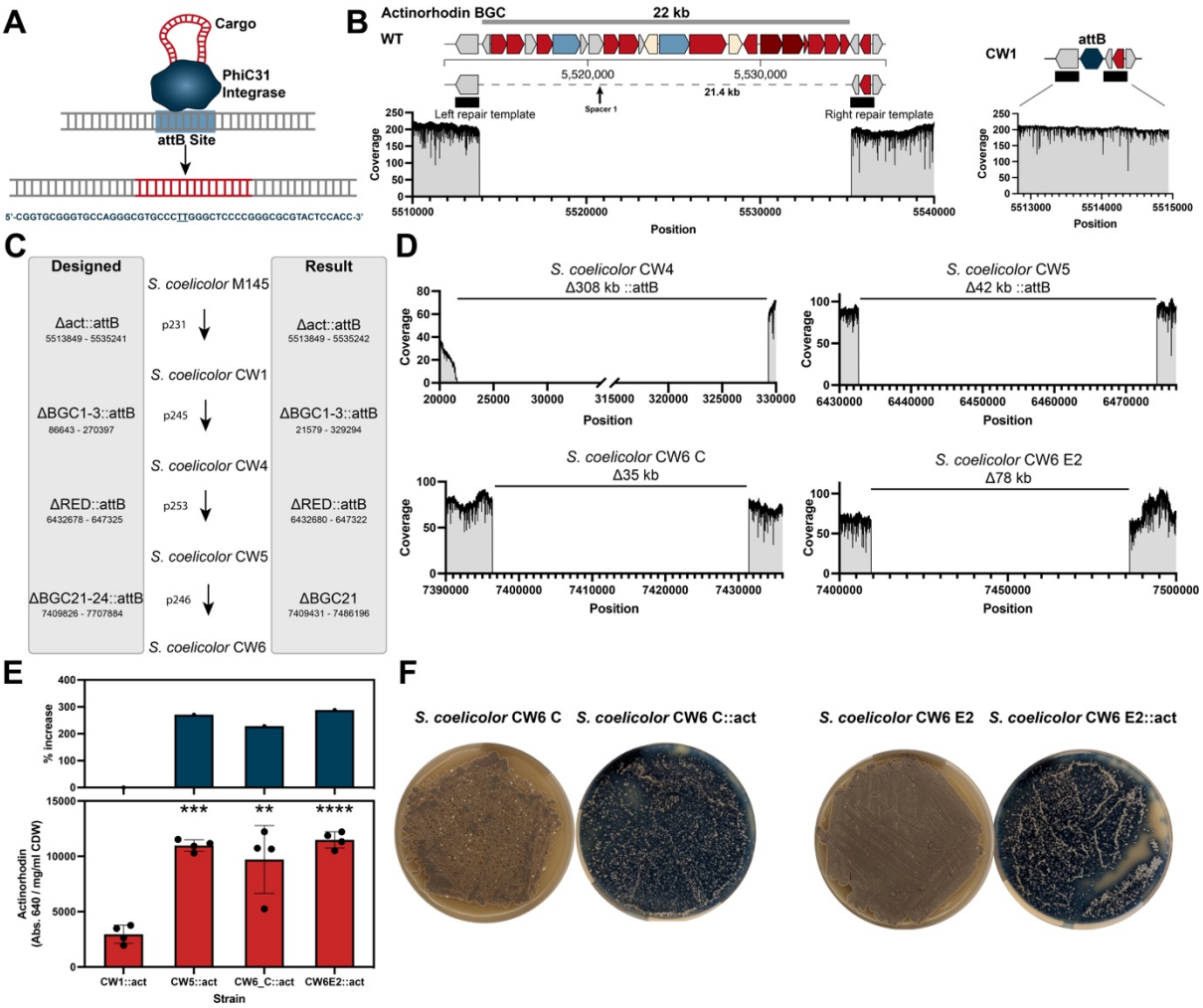


Figure 4: Simultaneous deletions and integrations enable streamlined genome engineering. (A) The PhiC31 *Streptomyces* integrase integrates cargo DNA into target *attB* sites. The consensus *attB* site from *S. coelicolor* is 51 bp long and features a central TT sequence where the cargo is integrated. **(B)** Substitution of the entire actinorhodin BGC with an additional *attB* site. The *attB* site was cloned between the repair templates. Coverage plots of mappings of ONT data against the wild type and the *in silico* generated substitution strain show precise genome engineering. **(C)** pCRISPR-Cas3 was used for the construction of a *S. coelicolor* expression host using both targeted and semi-targeted deletions and substitutions. **(D)** Oxford Nanopore sequencing results for all deletions based on minimap2 mappings to the reference genome. **(E)** The final strains *S. coelicolor* CW5 and CW6 both displayed > 200 % increase in actinorhodin production compared to the base strain *S. coelicolor* CW1 upon integration of an actinorhodin BGC BAC. **(F)** Phenotypes of *S. coelicolor* CW6 C and E2 without and with actinorhodin integrations. Shown in (**E**) are the means ± standard deviations of four biological replicates. Significance was tested using unpaired two tailed t-tests, where **P < 0.01, ***P < 0.001, ****P < 0.0001.

Cas3-based genome engineering enables rapid host construction in S. coelicolor

Motivated by these results, we wanted to expand on this approach to demonstrate how pCRISPR-Cas3 can enable streamlined genome engineering of *Streptomyces* hosts. *S. coelicolor* has previously been engineered as a host for heterologous expression of BGCs, and a number of derivatives were constructed in the process [43]. The most widely used ones, *S. coelicolor* M1152 and M1154 are both quadruplicate deletion strains (*Δact Δred Δcpk Δcda*) with varying additional mutations. To demonstrate how pCRISPR-Cas3 can be used for streamlined genome engineering experiments, we attempted to perform simultaneous genome reduction experiments and integrations of additional *attB* sites. It was previously shown that the integration site of BGCs can have great effects on product titers and that integrations in the chromosomal arms, where the BGC density is bigger, generally result in higher titers[44]. Therefore, deletion of BGCs and simultaneous integration of *attB* sites in place is likely to result in desirable production phenotypes.

Based on the *S. coelicolor* M145 Δact::*attB* strain, from here on referred to as *S. coelicolor* CW1, several rounds of pCRISPR-Cas3 mediated genome engineering experiments were performed. Given the processive nature of the Cas3 nuclease, we utilized a combination of targeted and semi-targeted deletions. As observed in *S. albidoflavus* J1074 and *S. sp.* NBC1270, Cas3 can install extending deletions if larger or unstable regions are targeted. We exploited that ability to not just delete single BGCs, but to also target larger regions with a high density of BGCs. The selected chromosomal regions contained antiSMASH regions 1-3 and 21-24, and encode several well-characterized BGCs, including those for isorenieratene, hopene, and arsono-polyketide. Targeting of regions 1-3 using p245 resulted in an extending deletion similar to what was observed in *S. albidoflavus* J1074 and *S. sp.* NBC1270. Nonetheless, the deleted region was substituted by an additional *attB* site, suggesting that the repair templates must have been used during the recombination event. The resulting strain with four deleted BGCs and two additional attB sites was named *S. coelicolor* CW4 and used as the basis for the next round of genome engineering. Using p253, the undecylprodigiosin BGC was deleted and substituted by a third additional *attB* site. The deletion was very precise and varied only by a few nucleotides from the *in silico* designed

deletion, again highlighting how pCRISPR-Cas3 can be used for both highly precise as well as semi-targeted random-sized deletions. The new strain was named *S. coelicolor* CW5 and used for the final round of genome engineering, targeting regions 21-23 with p246. The resulting deletion was smaller in size than designed and resulted in only the deletion of region 21. Additionally, no *attB* site was integrated, suggesting that this deletion was fully unspecific. The final strain, named *S. coelicolor* CW6 was reduced by 6 BGCs and carries a total of four *attB* sites, three of which were installed using pCRISPR-Cas3. In total, the genome was reduced by around 450 kb, corresponding to approximately 5 % of the genome. In addition to the pCRISPR-Cas3-induced deletions, several transposition events were detected. These were relatively stable across all sequenced strains, indicating that the majority of these came already from the parental *S. coelicolor* M145 strain.

To test the constructed *S. coelicolor* CW strains, we introduced the actinorhodin BGC encoded on an integrative plasmid via conjugation into all strains. *S. coelicolor* CW1 was used as the base strain, and *S. coelicolor* CW5 and *S. coelicolor* CW6 as final strains. For *S. coelicolor* CW6, two clones were tested. In addition to *S. coelicolor* CW6 E2, in which the deletion size in region 21 was 76.7 kb, we also tested *S. coelicolor* CW6 C, where the deletion was only 35 kb in size. After conjugation, four biological replicates were selected for each strain and cultivated in ISP2 medium supplemented with 25 µg/ml of apramycin for one week in 24 well plate as described in the methods section. The actinorhodin production was measured in the supernatant at 640 nm and normalized with the cell dry weight of the respective culture. *S. coelicolor* CW5 and CW6 strains all produced significantly more actinorhodin compared to the base strain *S. coelicolor* CW1. For *S. coelicolor* CW5, the specific product yield rose from 2956 $Abs._{640nm}$ /mg ml^{-1} CDW to 10974 $Abs._{640nm}$ /mg ml^{-1} CDW, an increase of 271 %. For *S. coelicolor* CW6 C, the specific production was 9713 $Abs._{640nm}$ /mg ml^{-1} CDW, corresponding to an increase of 228 % compared to *S. coelicolor* CW1. Finally, for *S. coelicolor* CW6 E2, a specific actinorhodin production of 11480 $Abs._{640nm}$ /mg ml^{-1} CDW was measured, an increase of 288 % over *S. coelicolor* CW1. These results highlight how the streamlined deletion of BGCs and integration of additional *attB* sites can be used to construct potential new expression hosts.

DISCUSSION

Here, we demonstrated for the first-time genome engineering of streptomycetes using a type I CRISPR system. Type I CRISPR systems are the most widespread systems in bacteria. In *Streptomyces,* type I CRISPR systems are widespread, while type II systems are very rare. We therefore hypothesized that *Streptomyces* might be more amendable to genome engineering facilitated by type I CRISPR systems. The processive nature of the hallmark nuclease Cas3, which results in long ssDNA overhangs results in recombinogenic overhangs. Given that *Streptomyces* have a high homologous recombination capability [45], we hypothesized that type I CRISPR systems are uniquely positioned to facilitate efficient genome engineering. The system used in this study is a previously characterized minimal type I-C CRISPR system [31].

The system, called pCRISPR-Cas3, is based on our established CRISPR platform, using a pSG5 replicon-based plasmid. The CASCADE is expressed from a single promoter as a polycistronic operon. Internal RBSs facilitate successful expression of all *cas* genes. The total plasmid size of almost 13 kb is the same as those of frequently used CRISPR plasmids in *Streptomyces,* suggesting that conjugative transfer and replication in *Streptomyces* should present no problem. To overcome cloning limitations resulting from the need to clone protospacer sequences between two CRISPR repeats, as well as the absence of established type IIS restriction enzymes for high GC contexts, we designed a PCR and Gibson Assembly based cloning workflow. This workflow robustly achieved high efficiencies, however, limits library-based applications due to the difficulties of cloning such with PCR-based approaches. This represents an obvious improvement for future iterations of the plasmid system.

pCRISPR-Cas3 introduced highly efficient deletions of the actinorhodin BGC in *S. coelicolor*. The strong red phenotype of deletion mutants simplified screening and assessment of efficiencies. The obtained results were verified by PCR and Illumina sequencing, showing that CASCADE-Cas3 is indeed more efficient than Cas9. Previous reports of pCRISPR-Cas9 based engineering reported higher efficiencies, however, were based on targeted screening of specific phenotypes [20]. Here, we followed an untargeted screening approach. Interestingly, successful deletions with pCRISPR-Cas3 appeared to be less

dependent on the selected protospacer compared to pCRISPR-Cas9. This suggests that the processive nature of the Cas3 nuclease might help to force the desired recombinations, and that Cas3 results in lower off-target-induced toxicity. The longer protospacer sequence, as well as the AT rich PAM sequence, likely minimizes off-target effects, thus reducing off-target associated toxicity of pCRISPR-Cas3. Direct toxicity of pCRISPR-Cas3 was only observed when protospacers were used alone and without repair templates, forcing repair through non-homologous end joining. Unspecific deletions were only obtained when targeting the highly plastic chromosomal arms and when targeting large extending chromosomal stretches with many hypothetical and potentially essential genes. However, the large extending deletions were observed for both pCRISPR-Cas9 and pCRISPR-Cas3 and were therefore likely just the result of targeting the highly unstable chromosomal arms.

pCRISPR-Cas3 was subsequently demonstrated to introduce genomic deletions in multiple model and non-model species. This suggests that pCRISPR-Cas3 can be easily implemented for genome engineering in many *Streptomyces* species. Finally, simultaneous deletions and integrations were demonstrated by integrating the PhiC31 *attB* site in place of BGCs. Through multiple rounds of genome engineering, the hosts *S. coelicolor* CW5 and CW6 were constructed, which produced significantly more actinorhodin compared to the base strain *S. coelicolor* CW1 after BGC integration. The observed increase in production was likely the result of both deletion of competing BGCs, a reduced genome size, as well as multicopy integrations of the BGC. This approach is likely to enable streamlined construction of overproduction strains in several *Streptomyces* species.

Given that pCRISPR-Cas3 is based on the established pCRISPR-Cas9 plasmid system, we hypothesize that most strains in which pCRISPR-Cas9 or CRISPR-BEST were established will also be amendable to genome engineering using pCRISPR-Cas3. Likely, pCRISPR-Cas3 will become a successor to pCRISPR-Cas9 for many applications, and enable highly efficient genome engineering, including targeted deletions, random sized deletions, as well as substitutions, even in previously difficult-to-engineer species. By enabling highly efficient genome engineering in more *Streptomyces* species, pCRISPR-Cas3 will likely

become a crucial tool for studies linking BGCs to compounds, host construction, and genome reduction studies.

MATERIALS AND METHODS

Strains and Culture conditions

All *E. coli* work for cloning and maintenance was performed in chemically competent One Shot™ Mach1™ T1 Phage-Resistant *E. coli* cells (ThermoFisher Scientific Inc., U.S.A) cells. Strains were cultivated on LB agar plates (10 g/l tryptone, 5 g/l yeast extract, 5 g/l sodium chloride, 15 g/l agar, to 1 l with MiliQ water) or in liquid 2xYT medium (16 g/l tryptone, 10 g/l yeast extract, 5 g/l sodium chloride, to 1 l with MiliQ water) and incubated at 37 °C. If required, medium was supplemented with the appropriate antibiotics (50 ng/µl apramycin, 50 ng/µl kanamycin, 25 ng/µl chloramphenicol). *Streptomyces* strains were cultivated at 30 °C on mannitol soy flour (MS) plates (20 g/l fat reduced soy flour, 20 g/l mannitol, 20 g/l agar, 10 mM $MgCl_2$, to 1 l with tap water) for spore generation. ISP2 plates (4 g/l yeast extract, 10 g/l malt extract, 4 g/l dextrose, 20 g/l agar, 333 ml tap water, 667 ml MiliQ water) were used for clean streaking. For liquid culture, 350 ml baffled shake flasks containing 50 ml of ISP2 (supplemented with the appropriate antibiotics if needed) were inoculated from spores. All shake flask cultivations were performed in shaking incubators at 30 °C and 180 rpm (311DS, Labnet International Inc., U.S.A). For actinorhodin production experiments, 24 well plates were used. 6 glass beads were added to each well, and cultivations were performed in 3.6 ml of ISP2 medium. Incubations were performed at 250 rpm and 30 °C. *Streptomyces* medium was supplemented with 12.5 ng/µl nalidixic acid and 50 ng/µl (solid medium) or 25 ng/µl (liquid medium) if required. For plasmid curing, plasmid harboring strains were cultivated at 40 °C in non-selective ISP2 liquid medium, streak on non-selective MS plates, and picked separately on selective and non-selective ISP2 plates for identification of plasmid-free clones.

Cloning Work

All *in silico* cloning was performed in SnapGene v6.2.1 (Dotmatics Limited, U.S.A). Primers were ordered from IDT (Integrated DNA Technologies, U.S.A). The CASCADE-Cas3 operon was synthesized by GenScript Biotech Corporation and delivered as a plasmid. PCRs for cloning and of high GC *Streptomyces*

elements were performed using Q5® High-Fidelity DNA Polymerase with GC Enhancer (New England BioLabs Inc., U.S.A). Colony PCRs for crRNA integrations into pCRISPR-Cas3 were performed using One*Taq*® 2X Master Mix with Standard Buffer Enhancer (New England BioLabs Inc., U.S.A).
Minipreps were performed using NucleoSpin® Plasmid EasyPure Kits (Macherey-Nagel, Germany). 1 % agarose gels were run at 100 V for 20-30 min using 6x DNA Gel Loading Dye and GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific Inc., U.S.A). Both gel purifications and PCR clean-ups were performed using the NucleoSpin® PCR and Gel Clean Up Kit (Macherey-Nagel, Germany). DNA concentrations and purities were measured on a NanoDrop™ 2000 (Thermo Fisher Scientific Inc., U.S.A). Restriction digestions were performed using Thermo Scientific™ FastDigest enzymes. Differing from the manufacturer's recommendations, the restriction digests were performed at 37 °C with extended incubation times of 2 hours, followed by inactivation for 10 min at 75 °C. 5 U/µL T4 DNA Ligase (Thermo Fisher Scientific Inc., U.S.A) was used for ligations overnight at room temperature. Gibson Assemblies were performed using NEBuilder® HiFi DNA Assembly Master Mix (New England BioLabs Inc., U.S.A) following the manufacturer's instructions. Sequence verifications were performed by Sanger sequencing using Mix2Seq overnight kits (Eurofins Genomics Germany GmbH, Germany).

Genome Mining & Protospacer Prediction

Prediction of biosynthetic gene clusters was performed using antiSMASH 6.0.1 [46]. Using the antiSMASH output, protospacers were predicted for pCRISPR-Cas9 based deletions using CRISPy-web [47]. Protospacers for pCRISPR-Cas3 were identified by searching for 5'-TTC-N_1-N_8-3' sequences within the target region with no mismatches in the genome. The "find similar DNA sequences" function of SnapGene was used to access potential off-targets with mismatches. The endogenous CRISPR system in *S. albidoflavus* J1074 was identified by searching CRISPRCasdb [48]. The presence of predicted CRISPR arrays and Cas genes was subsequently manually verified in the sequenced strain of *S. albidoflavus* J1074.

Interspecies Conjugations

Transfer of plasmid DNA into *Streptomyces* strains was performed through interspecies conjugations. Spores for conjugations were prepared as described by Tong et al. [35].
Homemade chemically competent or room temperature competent[49] *E. coli* ET12567 pUZ8002 [50] cells were transformed and plated on LB plates supplemented with 50 ng/µl apramycin, 25 ng/µl chloramphenicol, and 50 ng/µl kanamycin. All transformants were washed off using LB medium and 5 ml LB overnight cultures inoculated with 50 µl. 2 ml of the overnight cultures were harvested the next day and washed twice at 2000xg with 1 ml of fresh LB medium. 500 µl of resuspended *E. coli* ET12567 pUZ8002 cells were then mixed with 500 µl of filtered spore suspension and spread on MS + 10 mM $MgCl_2$ plates. The next day, overlays were performed with 1 ml of ddH_2O and 5 µl of 50 mg/ml apramycin. Exconjugants were picked using wooden toothpicks and transferred to selective ISP2 plates supplemented with nalidixic acid.

Streptomyces Colony PCRs

For colony PCRs on *Streptomyces* colonies, pieces of the colonies were scraped off the plates and transferred to PCR tubes containing 50 µl of 10 % DMSO. The tubes were sealed and boiled for 10 min at 99 °C, transferred to dry ice for 20 min, and then again boiled for 10 min. The entire process was repeated one more time, and the tubes were then spun down to separate the debris from the supernatant. 1 µl of the supernatant was used as template for colony PCRs. All *Streptomyces* colony PCRs were performed using Q5® High-Fidelity DNA Polymerase with GC Enhancer (New England BioLabs Inc., U.S.A).

Whole Genome Sequencing & Bioinformatic Analysis

Genomic DNA was isolated from *Streptomyces* plates or liquid cultures using the DNeasy PowerLyzer PowerSoil Kit (Qiagen, Germany). The bead beating was performed using a TissueLyser LT (Qiagen, Germany) at default settings for 7 min. Quality control was performed using a NanoDrop™ 2000 for purity

and Qubit 4 Fluorometer for concentration measurements (Thermo Fisher Scientific Inc., U.S.A). Genomic DNA was further run on 0.8 % agarose gels to assess fragmentation. Illumina sequencing was performed by Novogene Co., Ltd (Bejing, China). Libraries were prepared using the NEB Next® UltraTM DNA Library Prep Kit (New England Biolabs, U.S.A) with a target insert size of 350 nt and six PCR cycles.
trimgalore (https://github.com/FelixKrueger/TrimGalore) and breseq 0.33.2 [51] were used for read trimming and data analysis of Illumina sequencing data. The following commands were used:

trim_galore -j 8 --length 100 -o illumina --paired --quality 20 --fastqc --gzip file(s)

breseq -r (full genetic background reference).gb trimmed_1.fq.gz trimmed2.fq.gz -j 12

Using samtools[52], the read coverage was extracted from the breseq output as follows:

samtools depth -a sample.bam > sample.coverage

Oxford Nanopore sequencing was performed as described by Alvarez-Arevalo et al. [53] and basecalled using Guppy (5.0.17+99baa5b, client-server API version 7.0.0 or 6.3.8+d9e0f64, client-server API 13.0.0) applying the high-accuracy model and excluding reads shorter than 1 kb.

Minimap2 2.18 54was used to map reads and samtools was used to extract the reads as follows:

minimap2 -a reference.fa reads.fastq.gz > mapping.sam
samtools sort mapping.sam > mapping.sam.sort.bam
samtools depth -a mapping.sam.sort.bam > sample.coverage

Mappings were visualized using Artemis [55]. For plotting, the regions of interest were extracted from the .coverage files and plotted in Prim 9 (GraphPad Software, U.S.A). *De novo* assemblies were performed using flye 2.9 [56]. Assembly graphs were visualized using bandage 0.8.1[57].

Cell Dry Weight Measurements

The endpoint cell dry weight was measured during harvesting of cultures. 2 ml tubes were dried overnight at 60 °C. For each strain, three replicates were measured. The tubes were weighed and numbered. 2 ml of cell culture was added, spun down at max. speed, and washed with ddH_2O. After removing all supernatant, the tubes were dried overnight at 60 °C and measured the next day on a Sartorius Qunitix scale (1 mg). The weight of the specific empty tube was subsequently subtracted to give the cell dry weight.

Actinorhodin Measurements

For actinorhodin quantification, 2 ml of shake flask cultures were harvested by centrifugation for 10 min at max. speed. The supernatant was transferred to new tubes, and 100 µl were transferred to F-bottom, clear, 96 well microplates (Greiner Bio-One International GmbH, Austria). 50 µl of 3 M NaOH were added to each well and carefully mixed by pipetting up and down. The samples were then measured in a BioTek EPOCH2 microplate reader (Agilent Technologies, Inc.) at 640 nm.

DATA AVAILABILITY

All whole genome sequencing data is available in the sequence read archive under BioProject PRJNA966932. Additional supplementary information such as plasmid and primer lists are available at: https://figshare.com/s/e1bfb388df0ae502034e or doi: 10.11583/DTU.22786157

ACKNOWLEDGMENTS

This work was supported by grants of the Novo Nordisk Foundation (NNF16OC0021746, NNF20CC0035580).

AUTHOR CONTRIBUTIONS

C.M.W. and T.W conceived and design the project. C.M.W, P.G, T.G, and R.S. carried out laboratory experiments. C.M.W analyzed data. D.F. analyzed Oxford Nanopore sequencing data. C.M.W wrote the manuscript with input from all authors.

REFERENCES

(1) Katz, L.; Baltz, R. H. Natural Product Discovery: Past, Present, and Future. *J. Ind. Microbiol. Biotechnol.* **2016**, *43* (2–3), 155–176. https://doi.org/10.1007/s10295-015-1723-5.

(2) Tyc, O.; Song, C.; Dickschat, J. S.; Vos, M.; Garbeva, P. The Ecological Role of Volatile and Soluble Secondary Metabolites Produced by Soil Bacteria. *Trends Microbiol.* **2017**, *25* (4), 280–292. https://doi.org/10.1016/j.tim.2016.12.002.

(3) Bentley, S. D.; Chater, K. F.; Cerdeño-Tárraga, A.-M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C.-H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neil, S.; Rabbinowitsch, E.; Rajandream, M.-A.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B. G.; Parkhill, J.; Hopwood, D. A. Complete Genome Sequence of the Model Actinomycete *Streptomyces Coelicolor* A3(2). *Nature* **2002**, *417* (6885), 141–147. https://doi.org/10.1038/417141a.

(4) Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Shinose, M.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M.; Ōmura, S. Complete Genome Sequence and Comparative Analysis of the Industrial Microorganism *Streptomyces Avermitilis*. *Nat. Biotechnol.* **2003**, *21* (5), 526–531. https://doi.org/10.1038/nbt820.

(5) Whitford, C. M.; Cruz-Morales, P.; Keasling, J. D.; Weber, T. The Design-Build-Test-Learn Cycle for Metabolic Engineering of Streptomycetes. *Essays Biochem.* **2021**, *65* (2), 261–275. https://doi.org/10.1042/EBC20200132.

(6) Myronovskyi, M.; Luzhetskyy, A. Heterologous Production of Small Molecules in the Optimized: *Streptomyces* Hosts. *Nat. Prod. Rep.* **2019**, *36* (9), 1281–1294. https://doi.org/10.1039/c9np00023b.

(7) Myronovskyi, M.; Rosenkränzer, B.; Nadmid, S.; Pujic, P.; Normand, P.; Luzhetskyy, A. Generation of a Cluster-Free *Streptomyces Albus* Chassis Strains for Improved Heterologous Expression of Secondary Metabolite Clusters. *Metab. Eng.* **2018**, *49* (July), 316–324. https://doi.org/10.1016/j.ymben.2018.09.004.

(8) Komatsu, M.; Uchiyama, T.; Omura, S.; Cane, D. E.; Ikeda, H. Genome-Minimized *Streptomyces* Host for the Heterologous Expression of Secondary Metabolism. *Proc. Natl. Acad. Sci.* **2010**, *107* (6), 2646–2651. https://doi.org/10.1073/pnas.0914833107.

(9) Gomez-escribano, J. P.; Bibb, M. J. Engineering *Streptomyces Coelicolor* for Heterologous Expression of Secondary Metabolite Gene Clusters. **2011**, *4*, 207–215. https://doi.org/10.1111/j.1751-7915.2010.00219.x.

(10) Peng, Q.; Gao, G.; Lü, J.; Long, Q.; Chen, X.; Zhang, F.; Xu, M.; Liu, K.; Wang, Y.; Deng, Z.; Li, Z.; Tao, M. Engineered *Streptomyces Lividans* Strains for Optimal Identification and Expression of Cryptic Biosynthetic Gene Clusters. *Front. Microbiol.* **2018**, *9* (DEC), 1–15. https://doi.org/10.3389/fmicb.2018.03042.

(11) Komatsu, M.; Komatsu, K.; Koiwai, H.; Yamada, Y.; Kozone, I.; Izumikawa, M.; Hashimoto, J.; Takagi, M.; Omura, S.; Shin-ya, K.; Cane, D. E.; Ikeda, H. Engineered *Streptomyces Avermitilis* Host for Heterologous Expression of Biosynthetic Gene Cluster for Secondary Metabolites. *ACS Synth. Biol.* **2013**, *2* (7), 384–396. https://doi.org/10.1021/sb3001003.

(12) Gomez-Escribano, J. P.; Bibb, M. J. Heterologous Expression of Natural Product Biosynthetic Gene Clusters in *Streptomyces Coelicolor* : From Genome Mining to Manipulation of Biosynthetic Pathways. *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (2), 425–431. https://doi.org/10.1007/s10295-013-1348-5.

(13) Kallifidas, D.; Jiang, G.; Ding, Y.; Luesch, H. Rational Engineering of *Streptomyces Albus* J1074 for the Overexpression of Secondary Metabolite Gene Clusters. *Microb. Cell Fact.* **2018**, *17* (1), 25. https://doi.org/10.1186/s12934-018-0874-2.

(14) Ahmed, Y.; Rebets, Y.; Estévez, M. R.; Zapp, J.; Myronovskyi, M.; Luzhetskyy, A. Engineering of *Streptomyces Lividans* for Heterologous Expression of Secondary Metabolite Gene Clusters. *Microb. Cell Fact.* **2020**, *19* (1), 5. https://doi.org/10.1186/s12934-020-1277-8.

(15) Yin, S.; Li, Z.; Wang, X.; Wang, H.; Jia, X.; Ai, G.; Bai, Z.; Shi, M.; Yuan, F.; Liu, T.; Wang, W.; Yang, K. Heterologous Expression of Oxytetracycline Biosynthetic Gene Cluster in *Streptomyces Venezuelae* WVR2006 to Improve Production Level and to Alter Fermentation Process. *Appl. Microbiol. Biotechnol.* **2016**, *100* (24), 10563–10572. https://doi.org/10.1007/s00253-016-7873-1.

(16) Li, L.; Zheng, G.; Chen, J.; Ge, M.; Jiang, W.; Lu, Y. Multiplexed Site-Specific Genome Engineering for Overproducing Bioactive Secondary Metabolites in Actinomycetes. *Metab. Eng.* **2017**, *40*, 80–92. https://doi.org/10.1016/j.ymben.2017.01.004.

(17) Mitousis, L.; Thoma, Y.; Musiol-Kroll, E. M. An Update on Molecular Tools for Genetic Engineering of Actinomycetes—The Source of Important Antibiotics and Other Valuable Compounds. *Antibiotics* **2020**, *9* (8), 494. https://doi.org/10.3390/antibiotics9080494.

(18) Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F. PCR-Targeted *Streptomyces* Gene Replacement Identifies a Protein Domain Needed for Biosynthesis of the Sesquiterpene Soil Odor Geosmin. *Proc. Natl. Acad. Sci.* **2003**, *100* (4), 1541–1546. https://doi.org/10.1073/pnas.0337542100.

(19) Cobb, R. E.; Wang, Y.; Zhao, H. High-Efficiency Multiplex Genome Editing of *Streptomyces* Species Using an Engineered CRISPR/Cas System. *ACS Synth. Biol.* **2015**, *4* (6), 723–728. https://doi.org/10.1021/sb500351f.

(20) Tong, Y.; Charusanti, P.; Zhang, L.; Weber, T.; Lee, S. Y. CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. *ACS Synth. Biol.* **2015**, *4* (9), 1020–1029. https://doi.org/10.1021/acssynbio.5b00038.

(21) Huang, H.; Zheng, G.; Jiang, W.; Hu, H.; Lu, Y. One-Step High-Efficiency CRISPR/Cas9-Mediated Genome Editing in *Streptomyces*. *Acta Biochim. Biophys. Sin. (Shanghai).* **2015**, *47* (4), 231–243. https://doi.org/10.1093/abbs/gmv007.

(22) Ye, S.; Enghiad, B.; Zhao, H.; Takano, E. Fine-Tuning the Regulation of Cas9 Expression Levels for Efficient CRISPR-Cas9 Mediated Recombination in *Streptomyces*. *J. Ind. Microbiol. Biotechnol.* **2020**, *47* (4–5), 413–423. https://doi.org/10.1007/s10295-020-02277-5.

(23) Li, L.; Wei, K.; Zheng, G.; Liu, X.; Chen, S.; Jiang, W.; Lu, Y. CRISPR-Cpf1-Assisted Multiplex Genome Editing and Transcriptional Repression in *Streptomyces*. *Appl. Environ. Microbiol.* **2018**, *84* (18), 1–18. https://doi.org/10.1128/AEM.00827-18.

(24) Tong, Y.; Whitford, C. M.; Robertsen, H. L.; Blin, K.; Jørgensen, T. S.; Klitgaard, A. K.; Gren, T.; Jiang, X.; Weber, T.; Lee, S. Y. Highly Efficient DSB-Free Base Editing for Streptomycetes with CRISPR-BEST. *Proc. Natl. Acad. Sci.* **2019**, *116* (41), 20366–20375. https://doi.org/10.1073/pnas.1913493116.

(25) Makarova, K. S.; Wolf, Y. I.; Iranzo, J.; Shmakov, S. A.; Alkhnbashi, O. S.; Brouns, S. J. J.; Charpentier, E.; Cheng, D.; Haft, D. H.; Horvath, P.;

Moineau, S.; Mojica, F. J. M.; Scott, D.; Shah, S. A.; Siksnys, V.; Terns, M. P.; Venclovas, Č.; White, M. F.; Yakunin, A. F.; Yan, W.; Zhang, F.; Garrett, R. A.; Backofen, R.; van der Oost, J.; Barrangou, R.; Koonin, E. V. Evolutionary Classification of CRISPR–Cas Systems: A Burst of Class 2 and Derived Variants. *Nat. Rev. Microbiol.* **2020**, *18* (2), 67–83. https://doi.org/10.1038/s41579-019-0299-x.

(26) Brouns, S. J. J.; Jore, M. M.; Lundgren, M.; Westra, E. R.; Slijkhuis, R. J. H.; Snijders, A. P. L.; Dickman, M. J.; Makarova, K. S.; Koonin, E. V.; van der Oost, J. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science (80-.).* **2008**, *321* (5891), 960–964. https://doi.org/10.1126/science.1159689.

(27) Hayes, R. P.; Xiao, Y.; Ding, F.; Van Erp, P. B. G.; Rajashankar, K.; Bailey, S.; Wiedenheft, B.; Ke, A. Structural Basis for Promiscuous PAM Recognition in Type I-E Cascade from *E. Coli*. *Nature* **2016**, *530* (7591), 499–503. https://doi.org/10.1038/nature16995.

(28) He, L.; St. John James, M.; Radovcic, M.; Ivancic-Bace, I.; Bolt, E. L. Cas3 Protein—A Review of a Multi-Tasking Machine. *Genes (Basel).* **2020**, *11* (2), 208. https://doi.org/10.3390/genes11020208.

(29) Morisaka, H.; Yoshimi, K.; Okuzaki, Y.; Gee, P.; Kunihiro, Y.; Sonpho, E.; Xu, H.; Sasakawa, N.; Naito, Y.; Nakada, S.; Yamamoto, T.; Sano, S.; Hotta, A.; Takeda, J.; Mashimo, T. CRISPR-Cas3 Induces Broad and Unidirectional Genome Editing in Human Cells. *Nat. Commun.* **2019**, *10* (1). https://doi.org/10.1038/s41467-019-13226-x.

(30) Dolan, A. E.; Hou, Z.; Xiao, Y.; Gramelspacher, M. J.; Heo, J.; Howden, S. E.; Freddolino, P. L.; Ke, A.; Zhang, Y. Introducing a Spectrum of Long-Range Genomic Deletions in Human Embryonic Stem Cells Using Type I CRISPR-Cas. *Mol. Cell* **2019**, *74* (5), 936-950.e5. https://doi.org/10.1016/j.molcel.2019.03.014.

(31) Csörgő, B.; León, L. M.; Chau-Ly, I. J.; Vasquez-Rifo, A.; Berry, J. D.; Mahendra, C.; Crawford, E. D.; Lewis, J. D.; Bondy-Denomy, J. A Compact Cascade–Cas3 System for Targeted Genome Engineering. *Nat. Methods* **2020**. https://doi.org/10.1038/s41592-020-00980-w.

(32) Qiu, Y.; Wang, S.; Chen, Z.; Guo, Y.; Song, Y. An Active Type I-E CRISPR-Cas System Identified in *Streptomyces Avermitilis*. *PLoS One* **2016**, *11* (2), e0149533. https://doi.org/10.1371/journal.pone.0149533.

(33) Sinkunas, T.; Gasiunas, G.; Fremaux, C.; Barrangou, R.; Horvath, P.; Siksnys, V. Cas3 Is a Single-Stranded DNA Nuclease and ATP-Dependent

Helicase in the CRISPR/Cas Immune System. *EMBO J.* **2011**, *30* (7), 1335–1342. https://doi.org/10.1038/emboj.2011.41.

(34) Gilchrist, C. L. M.; Booth, T. J.; van Wersch, B.; van Grieken, L.; Medema, M. H.; Chooi, Y.-H. Cblaster: A Remote Search Tool for Rapid Identification and Visualization of Homologous Gene Clusters. *Bioinforma. Adv.* **2021**, *1* (1), 1–10. https://doi.org/10.1093/bioadv/vbab016.

(35) Tong, Y.; Whitford, C. M.; Blin, K.; Jørgensen, T. S.; Weber, T.; Lee, S. Y. CRISPR–Cas9, CRISPRi and CRISPR-BEST-Mediated Genetic Manipulation in Streptomycetes. *Nat. Protoc.* **2020**, *15* (8), 2470–2502. https://doi.org/10.1038/s41596-020-0339-z.

(36) Rostain, W.; Grebert, T.; Vyhovskyi, D.; Pizarro, P. T.; Tshinsele-Van Bellingen, G.; Cui, L.; Bikard, D. Cas9 Off-Target Binding to the Promoter of Bacterial Genes Leads to Silencing and Toxicity. *Nucleic Acids Res.* **2023**, 1–12. https://doi.org/10.1093/nar/gkad170.

(37) Hoff, G.; Bertrand, C.; Piotrowski, E.; Thibessard, A.; Leblond, P. Genome Plasticity Is Governed by Double Strand Break DNA Repair in *Streptomyces*. *Sci. Rep.* **2018**, *8* (1), 5272. https://doi.org/10.1038/s41598-018-23622-w.

(38) Ahmed, Y.; Rebets, Y.; Estévez, M. R.; Zapp, J.; Myronovskyi, M.; Luzhetskyy, A. Engineering of *Streptomyces Lividans* for Heterologous Expression of Secondary Metabolite Gene Clusters. *Microb. Cell Fact.* **2020**, *19* (1), 5. https://doi.org/10.1186/s12934-020-1277-8.

(39) Li, L.; Wei, K.; Liu, X.; Wu, Y.; Zheng, G.; Chen, S.; Jiang, W.; Lu, Y. AMSGE: Advanced Multiplex Site-Specific Genome Engineering with Orthogonal Modular Recombinases in Actinomycetes. *Metab. Eng.* **2019**, *52* (November 2018), 153–167. https://doi.org/10.1016/j.ymben.2018.12.001.

(40) Thorpe, H. M.; Smith, M. C. M. In Vitro Site-Specific Integration of Bacteriophage DNA Catalyzed by a Recombinase of the Resolvase/Invertase Family. *Proc. Natl. Acad. Sci.* **1998**, *95* (10), 5505–5510. https://doi.org/10.1073/pnas.95.10.5505.

(41) Myronovskyi, M.; Luzhetskyy, A. Genome Engineering in Actinomycetes Using Site-Specific Recombinases. *Appl. Microbiol. Biotechnol.* **2013**, *97* (11), 4701–4712. https://doi.org/10.1007/s00253-013-4866-1.

(42) Combes, P.; Till, R.; Bee, S.; Smith, M. C. M. The Streptomyces Genome Contains Multiple Pseudo-*AttB* Sites for the C31-Encoded Site-Specific

Recombination System. *J. Bacteriol.* **2002**, *184* (20), 5746–5752. https://doi.org/10.1128/JB.184.20.5746-5752.2002.

(43) Gomez-Escribano, J. P.; Bibb, M. J. Engineering *Streptomyces Coelicolor* for Heterologous Expression of Secondary Metabolite Gene Clusters. *Microb. Biotechnol.* **2011**, *4* (2), 207–215. https://doi.org/10.1111/j.1751-7915.2010.00219.x.

(44) Bilyk, B.; Horbal, L.; Luzhetskyy, A. Chromosomal Position Effect Influences the Heterologous Expression of Genes and Biosynthetic Gene Clusters in *Streptomyces* Albus J1074. *Microb. Cell Fact.* **2017**, *16* (1), 5. https://doi.org/10.1186/s12934-016-0619-z.

(45) Doroghazi, J. R.; Buckley, D. H. Widespread Homologous Recombination within and between *Streptomyces* Species. *ISME J.* **2010**, *4* (9), 1136–1143. https://doi.org/10.1038/ismej.2010.45.

(46) Blin, K.; Shaw, S.; Kloosterman, A. M.; Charlop-Powers, Z.; van Wezel, G. P.; Medema, M. H.; Weber, T. AntiSMASH 6.0: Improving Cluster Detection and Comparison Capabilities. *Nucleic Acids Res.* **2021**, *49* (W1), W29–W35. https://doi.org/10.1093/nar/gkab335.

(47) Blin, K.; Pedersen, L. E.; Weber, T.; Lee, S. Y. CRISPy-Web: An Online Resource to Design SgRNAs for CRISPR Applications. *Synth. Syst. Biotechnol.* **2016**, *1* (2), 118–121. https://doi.org/10.1016/j.synbio.2016.01.003.

(48) Pourcel, C.; Touchon, M.; Villeriot, N.; Vernadet, J.-P.; Couvin, D.; Toffano-Nioche, C.; Vergnaud, G. CRISPRCasdb a Successor of CRISPRdb Containing CRISPR Arrays and Cas Genes from Complete Genome Sequences, and Tools to Download and Query Lists of Repeats and Spacers. *Nucleic Acids Res.* **2019**, *48* (D1), D535–D544. https://doi.org/10.1093/nar/gkz915.

(49) Tu, Q.; Yin, J.; Fu, J.; Herrmann, J.; Li, Y.; Yin, Y.; Stewart, A. F.; Müller, R.; Zhang, Y. Room Temperature Electrocompetent Bacterial Cells Improve DNA Transformation and Recombineering Efficiency. *Sci. Rep.* **2016**, *6* (1), 24648. https://doi.org/10.1038/srep24648.

(50) MacNeil, D. J.; Gewain, K. M.; Ruby, C. L.; Dezeny, G.; Gibbons, P. H.; MacNeil, T. Analysis of *Streptomyces Avermitilis* Genes Required for Avermectin Biosynthesis Utilizing a Novel Integration Vector. *Gene* **1992**, *111* (1), 61–68. https://doi.org/10.1016/0378-1119(92)90603-M.

(51) Deatherage, D. E.; Barrick, J. E. Identification of Mutations in Laboratory-Evolved Microbes from Next-Generation Sequencing Data Using Breseq.

In *Methods in Molecular Biology*; 2014; pp 165–188. https://doi.org/10.1007/978-1-4939-0554-6_12.

(52) Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25* (16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

(53) Alvarez-Arevalo, M.; Sterndorff, E. B.; Faurdal, D.; Jørgensen, T. S.; Mourched, A. S.; Vuksanovic, O.; Saha, S.; Weber, T. Extraction and Oxford Nanopore Sequencing of Genomic DNA from Filamentous Actinobacteria. *STAR Protoc.* **2023**, *4* (1), 101955. https://doi.org/10.1016/j.xpro.2022.101955.

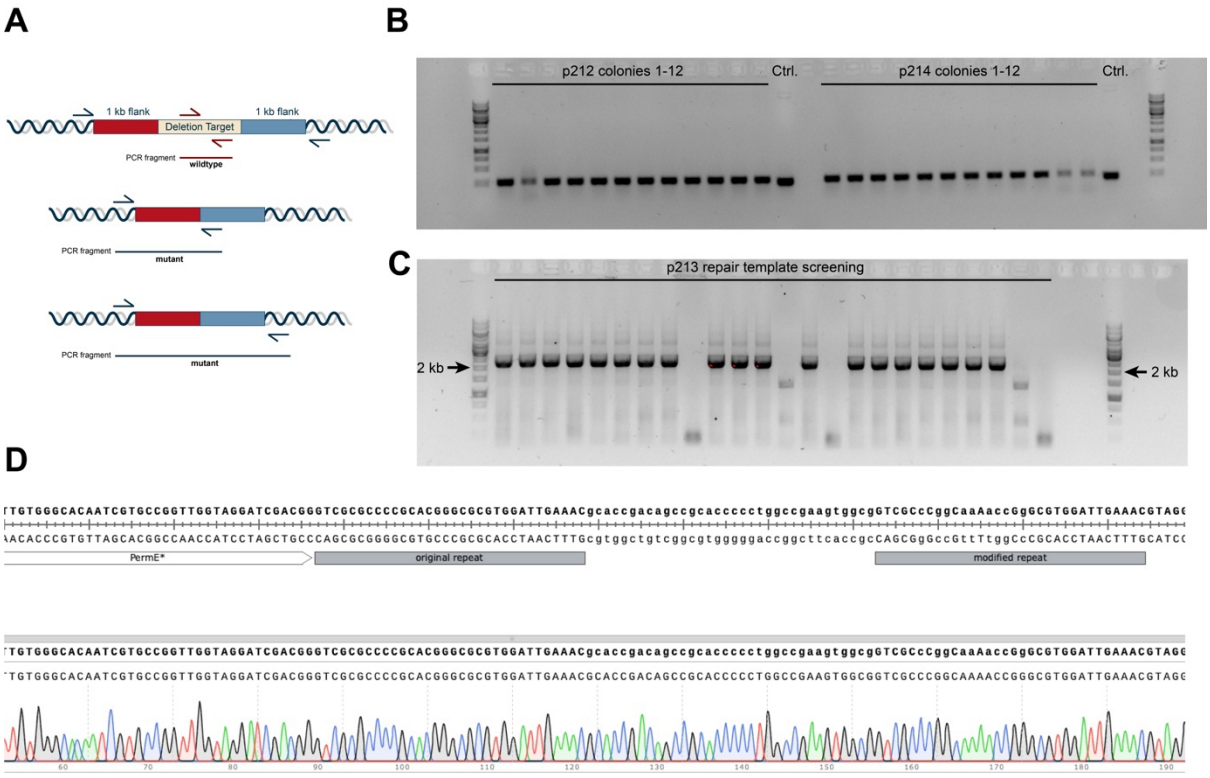
(54) Li, H. Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics* **2018**, *34* (18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

(55) Carver, T.; Harris, S. R.; Berriman, M.; Parkhill, J.; McQuillan, J. A. Artemis: An Integrated Platform for Visualization and Analysis of High-Throughput Sequence-Based Experimental Data. *Bioinformatics* **2012**, *28* (4), 464–469. https://doi.org/10.1093/bioinformatics/btr703.

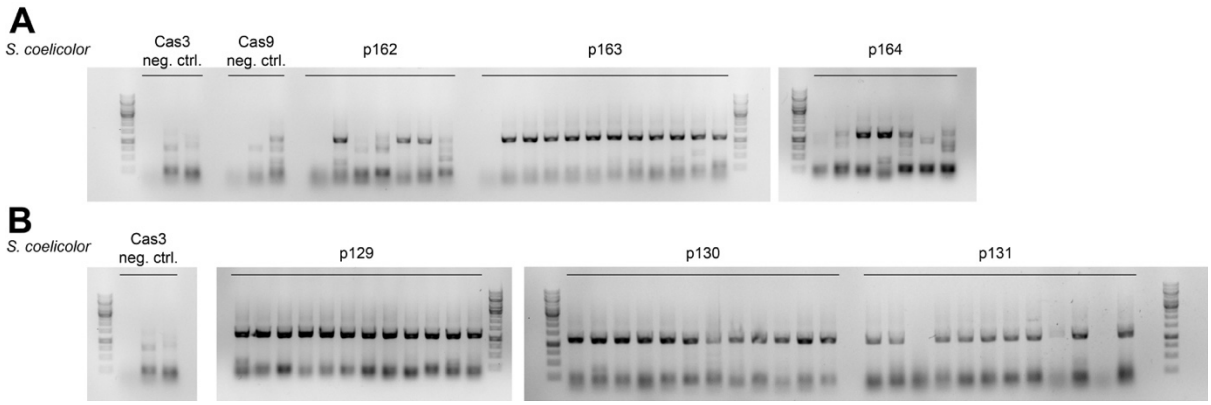
(56) Kolmogorov, M.; Yuan, J.; Lin, Y.; Pevzner, P. A. Assembly of Long, Error-Prone Reads Using Repeat Graphs. *Nat. Biotechnol.* **2019**, *37* (5), 540–546. https://doi.org/10.1038/s41587-019-0072-8.

(57) Wick, R. R.; Schultz, M. B.; Zobel, J.; Holt, K. E. Bandage: Interactive Visualization of de Novo Genome Assemblies. *Bioinformatics* **2015**, *31* (20), 3350–3352. https://doi.org/10.1093/bioinformatics/btv383.

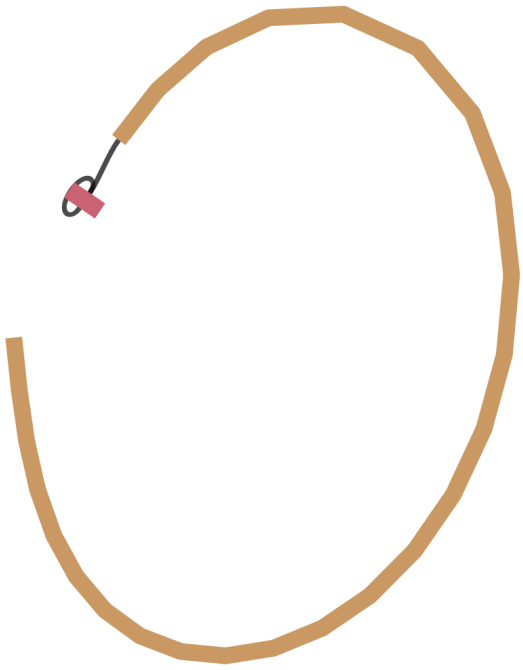
SUPPLEMENTARY INFORMATION



Supplementary Figure 5: A) Schematics for different screening approaches. Colony PCRs can be performed using primers binding inside the deletion region for screening of drop out mutants, pairs of primers with one binding inside the repair templates bridging the junction site and one primer binding in the chromosome outside, or with two primers both outside of the repair templates. Given the challenges associated with obtaining large high GC PCR fragments, the second approach is most widely used. **B & C)** Exemplary results for colony PCRs for protospacer (top) and repair template (bottom) integrations. The differences between the empty pCRISPR-Cas3 repeat region and the same region with an integrated protospacer is large enough to differentiate negative colonies from positives, as these lead to slightly higher bands in the gel. **D)** Example of an alignment of Sanger sequencing data of a successfully cloned protospacer.



Supplementary Figure 6: A) Gel pictures of colony PCR screening for pCRISPR-Cas9 mediated deletions of the actinorhodin biosynthetic gene cluster in *S. coelicolor* M145. p162-p164 carry the "left", "middle", and "right" protospacer, respectively. One of the negative controls produced a light band also around 1 kb in size, however upon closer inspection that band sits slightly higher than the expected band for deletions. Positive samples produced strong bands slightly lower than the unspecific light band. **B)** Colony PCRs for pCRISPR-Cas3 mediated deletions of the actinorhodin biosynthetic gene cluster. Strong bands slightly above the 1 kb ladder were observed for almost all colonies. p129-p131 carry the "left", "middle", and "right" protospacers, respectively.



Supplementary Figure 7: Assembly graph of a *de novo* assembly of NBC1270 after targeting BGCs 1-3 with pCRISPR-Cas9. The assembly was visualized using bandage. The graph highlights how a sequence stretch at the far chromosomal end was replicated multiple times. Based on the coverage, around 12 replications are expected.

Supplementary Information Table 1:

Blast results for *Salb*Cas3 and *Spy*Cas9. Can be downloaded here: https://figshare.com/s/e1bfb388df0ae502034e or doi: 10.11583/DTU.22786157

Supplementary Information Table 2:

Cblaster output for *Salb*CASCADE. Can be downloaded here: https://figshare.com/s/e1bfb388df0ae502034e or doi: 10.11583/DTU.22786157

Additional Data

Lists of primers, plasmids, protospacers, as well as antiSMASH results, can be found here: https://figshare.com/s/e1bfb388df0ae502034e or doi: 10.11583/DTU.22786157

Chapter 6

A Versatile Genetic Engineering Toolkit for *E. coli* Based on CRISPR-Prime Editing

Yaojun Tong[1,2,*], Tue S. Jørgensen[1], **Christopher M. Whitford[1]**, Tilmann Weber[1,*], Sang Yup Lee[1,3,*]

[1] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark.
[2] State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University (SJTU), Shanghai 200240, China.
[3] Department of Chemical and Biomolecular Engineering, BioProcess Engineering Research Center, BioInformatics Research Center, Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea.

*To whom correspondence should be addressed yaojun.tong@sjtu.edu.cn (Y.T.), tiwe@biosustain.dtu.dk (T.W.), or leesy@kaist.ac.kr (S.Y.L.).

Published in Nature Communications: doi.org/10.1038/s41467-021-25541-3

ABSTRACT

In most prokaryotes, missing or poorly active non-homologous end joining (NHEJ) DNA repair pathways heavily restrict the direct application of CRISPR-Cas for DNA double-strand break (DSB)-based genome engineering without providing editing templates. CRISPR base editors, on the other hand, can be directly used for genome engineering in a number of bacteria, including *E. coli*, showing advantages over CRISPR-Cas9, since they do not require DSBs. However, as the current CRISPR base editors can only engineer DNA by A to G or C to T/G/A substitutions, they are incapable of mediating deletions, insertions, and combinations of substitution, deletion, and insertion. To address these challenges, we developed a CRISPR-Prime Editing-based, DSB-free, versatile, and single-nucleotide resolution genetic manipulation toolkit for prokaryotes. This toolkit can be used to introduce substitutions, deletions, insertions, and the combination thereof, both in plasmids and the chromosome of *E. coli*. Notably, under optimal experiment conditions of plasmid DNA editing, 1-bp deletions that lead to frameshifts have editing-efficiencies of around 40%. In this study, we achieved deletion and insertion DNA fragments of up to 97 bp and 33 bp, respectively with the CRISPR-Prime Editing toolkit for *E. coli*. However, the efficiency drops sharply with the increase of the fragment size. By providing a second guide RNA, CRISPR-Prime Editing for *E. coli* can be used for multiplexed editing with a relatively low efficiency. Furthermore, a genome-wide off-target evaluation demonstrates the very high fidelity of our CRISPR-Prime Editing toolkit for *E. coli*. It is not only a useful addition to the genome engineering arsenal for *E. coli*, but also may be the basis for the development of similar toolkits for other bacteria.

MAIN

Advances in synthetic biology, metabolic engineering, multi-omics, high throughput DNA sequencing and synthesis, and computational biology have prompted a rapidly increasing demand for fast and robust genetic engineering methods to speed up the strain development in a Design-Build-Test-Learn cycle. The classic genetic engineering approaches in prokaryotes often use phage-derived RecET and lambda red recombinase-based recombineering[1, 2]. They employ the homology-directed integration/replacement of a donor double stranded DNA (dsDNA) or oligonucleotide for making insertions, deletions, and substitutions of the target DNA. For example, the "Multiplex Automated Genome Engineering" (MAGE)[3] is a method that can be used for simultaneous manipulation of genes across multiple chromosomal loci of *E. coli*. Possible mutations include mismatch mutation, insertion, and deletion, and editing efficiencies are usually below 20% for all types of edits[3]. Classical MAGE not only requires the synthesis and delivery of ssDNA oligos but also the expression of lambda (**λ**) red recombinase systems (Exo, Beta and Gam) in the target *E. coli* strain[3]. Several improved methods have been developed based on the classical MAGE to increase the editing efficiency and decrease the off-target effect. For example, the pORTMAGE system[4], using a dominant-negative mutant protein of the MMR pathway, not only achieves higher editing efficiency and lower off-target effect, but also works for different bacterial species other than *E. coli*. One step forward, an improved pORTMAGE system was built by discovery of new, highly active single-stranded DNA-annealing proteins (SSAP). The identified CspRec improved pORTMAGE editing efficiency to up to 50%[5]. Recently, retron library recombineering was introduced as a new method that achieves up to 90% editing efficiency by in vivo production of single-stranded DNA using the targeted reverse-transcription activity of retrons[6].

CRISPR-Cas (Clustered regularly interspaced short palindromic repeats-CRISPR associated (Cas) proteins) systems, originating from the bacterial adaptive immune system[7], have been engineered as genome editing tools for a variety of organisms[8]. Among these tools, the Class 2, type II CRISPR system CRISPR-Cas9 of *Streptococcus pyogenes* has been most widely studied and applied. The Cas9 nuclease can be guided by an engineered RNA (single guide RNA, sgRNA) to make DNA double strand breaks (DSBs) of the protospacer

adjacent motif (PAM)-containing target DNA[9]. Different types of genetic engineering can be achieved during the repair of DSBs. There are two major pathways for DSB repair *in vivo*, the non-homologous end joining (NHEJ) and the homology-directed repair (HDR)[10]. In most eukaryotes, NHEJ is the dominant way to repair DSBs. During NHEJ repair, small insertions and/or deletions (indels) are introduced at the lesion site, leading to gene disruptions in the target gene. In most bacteria, DSBs normally lead to cell death due to the lack of NHEJ[11]. In these organisms, DNA damage is primarily repaired via HDR with sister chromatids[12], where the template DNA replace the damaged DNA fragment by recombination[13].

The lack of NHEJ repair in most prokaryotes restricts the direct use of CRISPR-Cas9 without providing editing templates as a genome editing tool. However, the method is widely used for negative selection to eliminate wild-type cells in recombination-based engineering methods[14]. Unlike CRISPR-Cas9, "DSB-free" CRISPR base editing systems have successfully been applied for direct genome editing in a number of bacteria without providing editing templates[15-17]. As they rely on DNA deaminase reactions, CRISPR base editors can only make one type of changes to the DNA: the substitution (C to T/A/G, or A to G), and the target C or A has to be within the relatively narrow editing window. Hence, it soon becomes a bottleneck of applying CRISPR base editors for bacterial genome engineering. For the insertion of large DNA fragments, methods such as CRISPR-associated transposase (CAST)[18] and INsert Transposable Elements by Guide RNA-Assisted TargEting (INTEGRATE)[19] were developed by combining CRISPR-Cas systems and transposons. The INTEGRATE was successfully tested in *E. coli* for integrating a ~10.1 kb fragment into the chromosome[19].

Recently, reverse transcriptase-Cas9 H840A nickase (Cas9n)-mediated targeted prime editing (PE) has been demonstrated in human cells[20], rice and wheat cells[21] to directly knock-out, knock-in, and replace nucleotides at the target locus without introducing DSBs and requiring editing templates. The CRISPR-PE system uses the 3'-extension sequence of the modified sgRNA (herein named as PEgRNA) to provide a primer binding sequence (PBS) and a reverse transcription template (RTT) carrying the desired edits for reverse transcription with the reverse transcriptase that is allocated in the target locus by Cas9n:sgRNA. After DNA repair, designed mutations are introduced into the target locus. As the system only introduces a nick in one DNA strand, we

hypothesized that it may not cause cell death in bacteria and thus could be applicable in bacterial genome engineering as well. Here, we report the establishment and evaluation of the CRISPR Prime Editing toolkit for *E. coli*.

RESULTS

Design of CRISPR-Prime Editing system for E. coli

To evaluate if the reverse transcriptase-Cas9n-mediated DNA modification works in bacteria, we constructed a three-plasmid system (pCDF-GFPplus, pPEgRNA, and pCRISPR-PE). A fourth plasmid (pVRb_PEgRNA, Supplementary Fig. 3) is introduced for the multiplexed editing. Plasmid pCDF-GFPplus serves as the reporter plasmid harbouring a gene encoding an *E. coli* codon optimized fast folding GFP[22] under a constitutive promoter J23106 (Fig. 1a). Plasmid pPEgRNA carries the constitutive promoter J23119 driving PEgRNA transcription. The PEgRNA is composed of a 20-nt spacer and a 3 prime extension containing the PBS and RTT (Fig. 1b). The third plasmid pCRISPR-PE expresses an *E. coli* codon optimized fusion protein composed of an engineered reverse transcriptase M-MLV2 (moloney murine leukaemia virus variant[20]), a flexible linker, and a Cas9n (Cas9 nickase, the H840A mutant of SpyCas9) under a tetracycline-inducible promoter (Fig. 1c).

Validation of CRISPR-Prime Editing system on plasmid editing in E. coli

To assess the versatility of the CRISPR-Prime Editing system on plasmid DNA engineering in *E. coli,* we designed a full set of possible DNA engineering events, including insertions, deletions, substitutions, and combinations of these to introduce premature stop codons into the coding sequence of GFP. The loss of fluorescence enables easy screening and evaluation for desired editing events. We identified a protospacer located at positions 178-197 of the GFP coding sequence (Supplementary Table 3) that should allow the introduction of a stop codon by DNA engineering with designed PEgRNA. The testing was initiated following observations reported in human cells[20] with a 3'-extension consisting of 13 nt PBS and 13 nt RTT scaffold. In the case of insertion, the length of RTT equals the RTT scaffold size plus the designed insertion, for example the length of RTT for TAA insertion is 16 nt (Supplementary Table 3 and Supplementary Table 4). Designed edits were placed inside the potential editing window starting from the nick and continuing downstream[20] (Fig. 1a). The Cas9n-M-MLV2 fusion protein binds to the desired PEgRNA transcript, forming an RNA-protein complex, the Cas9n-

component of the complex subsequently finds its target DNA sequence and introduces a nick in the PAM containing DNA strand. The PBS within the 3 prime extension then binds to the flipped PAM containing DNA sequence, initiating the reverse transcription to elongate the nicked DNA sequence based on the sequence of the RTT (Fig. 1d). After the reverse transcription process, the nicked double stranded DNA hypothetically undergoes an equilibration between the edited 3′ flap and the unedited 5′ flap. The cleavage of the unedited 5’ flap then leads to the desired DNA editing[20] (Fig. 1e).

As a proof of concept, we transformed *E. coli* cells with CRISPR-Prime Editing systems programmed for TAA (3-bp) insertion, T to A (1-bp) substitution, T (1-bp) deletion, and the combination thereof. All of these edits will lead to stop codons that prematurely terminate translation to inactivate the target gene (Supplementary Table 4). After induction with 200 ng/mL anhydrotetracycline (ATc), we observed non-fluorescent (non-green) clones formed on all four plates of designed DNA editing events (Fig. 1f). In order to further confirm that GFP fluorescence loss was due to the designed DNA editing consequences, we randomly Sanger sequenced 24 non-fluorescent colonies from each induced plate. Results demonstrated that almost all of the non-fluorescent colonies were indeed carrying the designed stop codon edits (Fig. 2a). By extending the incubation time of the induction plates (for example to 3-5 days), we observed that the editing events were accumulating over time. This becomes visible when colonies from 24-h incubation are further incubated: non-green “sections” grow out of the original green colony, even surrounding it (Supplementary Fig. 1). Sanger sequencing confirmed that these no longer fluorescent strains were successfully edited. This result indicates that prolongation of the incubation time is one possibility to increase the number of correctly edited cells.

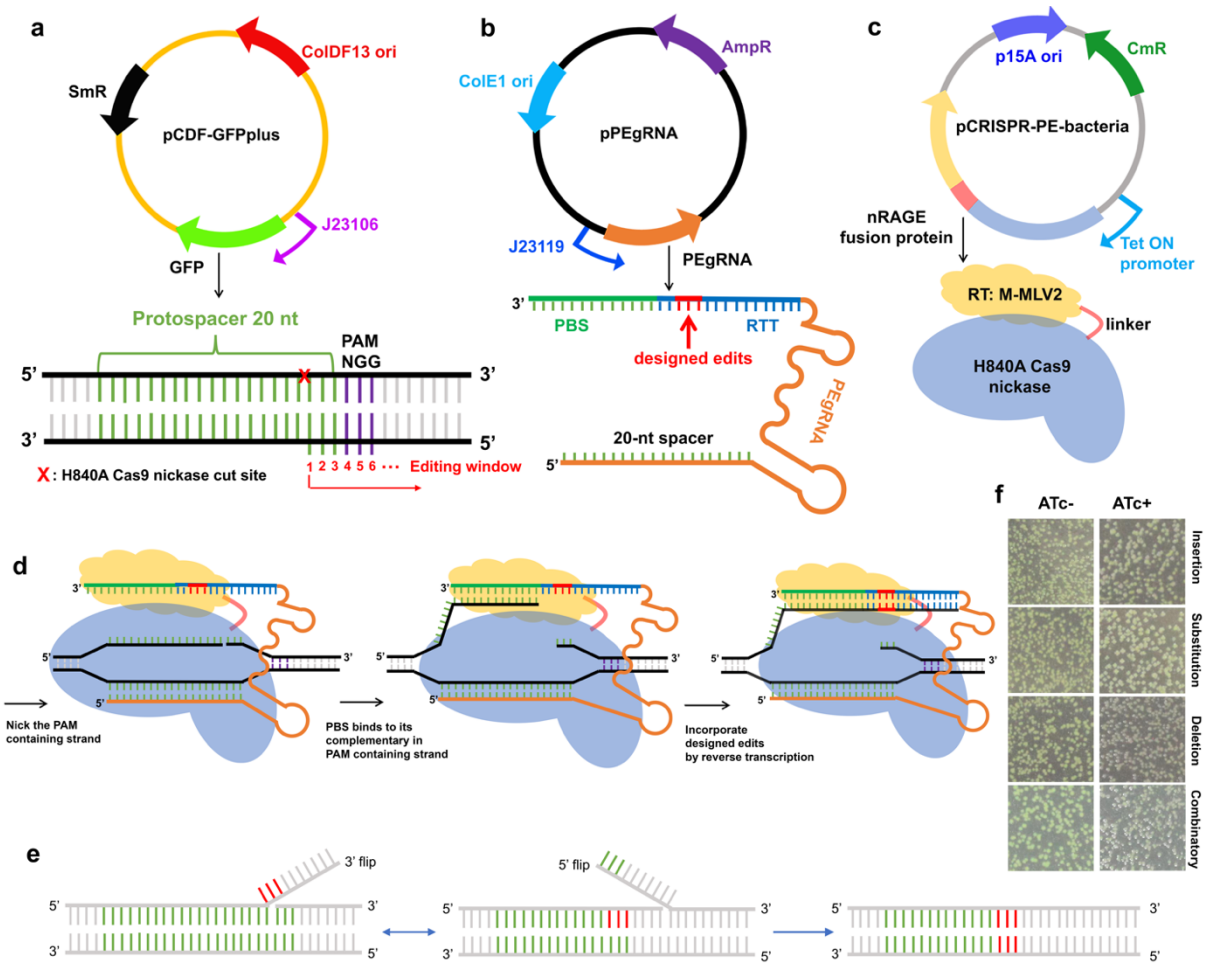


Figure 1: A three-plasmid system for evaluation of CRISPR-Prime Editing system in *E. coli*. **a**. The plasmid map of the reporter vector pCDF-GFPplus, which carries a constitutive promoter J23106 driving expression of fast folding GFP. The plasmid contains a spectinomycin-resistance (SmR) gene, and the ColDF13 origin (ori). An illustration of the target DNA composition is shown below the plasmid map. **b**. The plasmid map of the PEgRNA transcript bearing vector, which contains the ColE1 ori and an ampicillin-resistance (AmpR) gene for selection. The PEgRNA transcript is under control of the constitutive promoter J23119. A detailed structure is shown beneath, the 3 prime extension sequence is composed of a PBS (in green) and a RTT (in blue), which carries the intended edits (in red). **c**. Plasmid map of the CRISPR-PE-bacteria vector, which carries a p15A ori and a chloramphenicol-resistance gene (CmR) for selection. The *E. coli* codon optimized, tetracycline inducible promoter driven Cas9n-M-MLV2 fusion protein consists of a H840A Cas9 nickase (Cas9n), a 33-aa flexible linker, and a moloney murine leukaemia virus (M-MLV) variant M-MLV2, described previously[20] with the following mutations: D200N, L603W, T306K, W313F, and T330P compared to the WT M-MLV (GenBank: AAC82568.2). **d**. A schematic model for DNA engineering with CRISPR-Prime Editing system for *E. coli*. After being expressed, the Cas9n-M-MLV2:PEgRNA complex binds to the targeted DNA sequence in a sgRNA- and PAM-dependent manner. The Cas9n domain within the fusion protein nicks the PAM-containing strand, freeing the adjacent DNA sequence. Subsequently, this piece of single stranded DNA hybridizes to the PBS, then primes reverse transcription of new DNA containing the designed edits based on the RTT within the 3'-extension of the PEgRNA transcript. **e**. Two possible

consequences of CRISPR-Prime Editing. It normally has an equilibration between the edited 3 prime flap and the unedited 5 prime flap, only the cleavage of the 5 prime flap leads to the desired editing. **f**. Colony views of *E. coli* strains transformed with CRISPR-Prime Editing systems carrying designed edits of TAA insertion, T to A substitution, T deletion and the combinatorial edits with and without 200 ng/mL ATc induction using a Doc-It imaging station, non-green colonies appeared after 24 h induction of 200 ng/mL ATc.

In the 24 Sanger sequenced clones, we noticed that 8 clones harbouring indels near the target nucleotide of the T to A substitution editing event (Fig. 2b. A genome-wide off-target evaluation will be presented below), while no such indels were found in other types of editing. We classified this as target adjacent unintended edits, which only occur in physical proximity to the nick site. Besides the combinatorial editing of insertion, deletion and substitution, we also investigated the possibility of performing double substitutions from one construct. An edit replacing tyrosine at the position 66 of the GFP to histidine (Y66H) was designed by flipping the TAT codon to a CAC codon. We obtained eight non-fluorescent clones out of which six clones were correctly edited (Supplementary Fig. 2). To further assess whether CRISPR-Prime Editing system for *E. coli* is capable of multiplexed editing, we introduced another compatible plasmid pVRb carrying a second PEgRNA for a G deletion (pVRb_PEgRNA_312Gdel) into *E. coli* DH10B harbouring the 3-plasmid system for TAA insertion, T deletion, T to an A substitution, two Ts to two Cs substitution, and combinatory editing. We successfully identified all expected dual-editing events (Fig. 2c), however the editing efficiency was relatively low (< 1%).

Characterization of CRISPR-Prime Editing system in E. coli

As the Cas9n-M-MLV2 fusion protein is driven by the ATc inducible promoter, we evaluated the optimal condition of induction using eight different ATc concentrations. The editing efficiencies were defined by calculating the ratio of non-green colonies. We observed a dose dependent induction manner for all four designed DNA engineering events (Fig. 3a). For cases of 1-bp deletion, 3-bp insertion, 1-bp substitution, and the combinatorial editing, CRISPR-Prime Editing system can reach efficiencies up to 43.7%, 13.8%, 19.9%, and 2.1%, respectively with 1000 ng/mL of inducer (Fig. 3a).

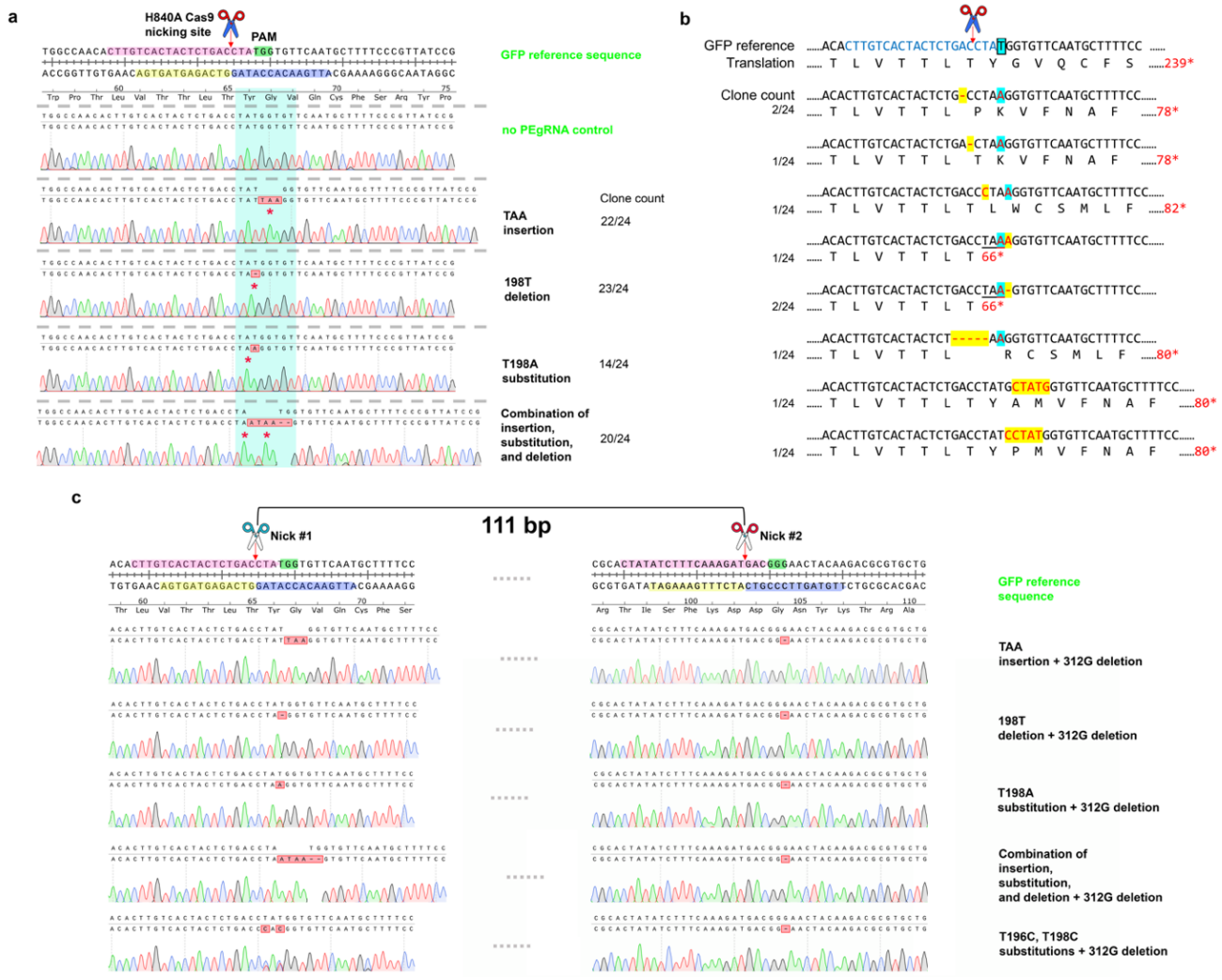


Figure 2: **Evaluation of plasmid-based editing by CRISPR-Prime Editing system using Sanger sequencing.** Stop codon is displayed as an asterisk under the nucleotide sequences, the potential Cas9 H840A nicking site is indicated by a red arrow. The 20 nt protospacer, the PAM sequence, the PBS, and the RTT are highlighted in pink, green, yellow, and blue, respectively; while the cyan masked Sanger sequencing traces show the sequence to be replaced by the RTT that will contain the different edits designed. **a.** 24 randomly picked colonies of each designed DNA engineering were Sanger sequenced and traces were aligned to the targeted locus of the GFP coding sequence. The correctly edited clone numbers and the total sequenced clone numbers are shown on the right of the figure. **b.** Shown are the recorded target-specific unintended edits of the 1-bp substitution. The target T is boxed and masked in light blue. The GFP reference DNA sequence and the translated amino acid sequences are show on the top row. The corrected edited nucleotide is in red and masked with light blue, while the unexpected mutations (off-target) are in red and masked with yellow. The recorded off-target clone numbers and the total sequenced clone numbers are shown on the left of the figure. **c**. Sanger sequencing traces of the successfully dual-edited clones by CRISPR-Prime Editing. Two nicks are 111 bp away, the left nick (nick #1) is introduced by pPEgRNA (ColE1 ori), and the right nick (nick #2) is introduced by pVRb_PEgRNA (pSC101 ori). The combinations of 3-bp insertion, 1-bp deletion, 1-bp substitution, 2-bp substitution, combinatory editing and 1-bp deletion are displayed.

Deletion and insertion of similar sized DNA fragments has an efficiency equal to, or higher than, MAGE[3]. Among the four DNA editing events, deletion showed the highest, while substitution showed the lowest efficiency (Fig. 3a). The editing efficiency did not change significantly with ATc concentrations of 100 to 500 ng/mL, and thus we decided to induce CRISPR-Prime Editing system with 200 ng/mL ATc in the following experiments unless specified otherwise.

Next, we evaluated the optimal length of PBS and RTT by measuring the frequency of the TAA-STOP codon insertion. Nearly no edits were observed when a PBS of 5 nt or 8 nt was used, while a 17 nt PBS showed editing efficiency equivalent to a 13 nt PBS (Fig. 3b). This indicated that the window of PBS for CRISPR-Prime Editing system for *E. coli* is 13 nt to 17 nt. For the RTT scaffold, we designed five different lengths, 4, 5, 8, 13, and 20 nt. We observed that too short (like < 10 nt) or too long (like > 20 nt) RTT reduced the editing efficiency, and the optimal length of the RTT scaffold was around 13 nt (Fig. 3c, and 3d). Results obtained in this study are consistent with previous reports in eukaryotes[20, 21] Moving forward, by using the optimal length of PBS and RTT scaffold, we systematically tested the capacity of both insertion and deletion with 200 ng/mL of inducer. We designed insertions of 3, 12, 18, and 33 bp, in which the 18 bp fragment is the mini-T7 promoter; and deletions of 1, 10, 23, 36, 49, and 97 bp. Though clones with all designed DNA engineering events could be successfully obtained, the editing efficiency dropped greatly with the increase of sizes (Fig. 3d, and 3e). For instance, under 200 ng/mL ATc, the editing efficiencies of 1-bp deletion and 10-bp deletion can reach 29.5% and 17.8%, respectively, while efficiencies dropped to below 2% with lengths of 23 bp-97 bp. For insertion, the efficiency was in general lower than deletion. The efficiency of 3-bp insertion was about 10% with 200 ng/mL ATc, and it dropped to below 1% when the size increased to 18 bp-33 bp (Fig. 3d, 3e, and Supplementary Fig. 5). In several of these cases, the editing efficiency was low. Many clones carrying the activated CRISPR-Prime Editing systems still showed GFP fluorescence. We randomly picked 10 of these “escapers”, together with four controls (Supplementary Table 1). The 14 strains were sequenced, and analyzed with our genome-wide SNP profiling approach that was used for the off/on-target evaluation as well. 7 out of the 10 “escapers” lost the 26-bp 3 prime extension sequence (Supplementary Table 5); except these deletions, the other parts of plasmids and the chromosome were intact. In 3 “escapers”, no mutations/SNPs were identified both on plasmids and chromosome that can

explain why no CRISPR-Prime Editing occurred (Supplementary Table 5). This indicates that besides mutating the guide RNA, yet-unknown escaping mechanisms are also present in *E. coli*.

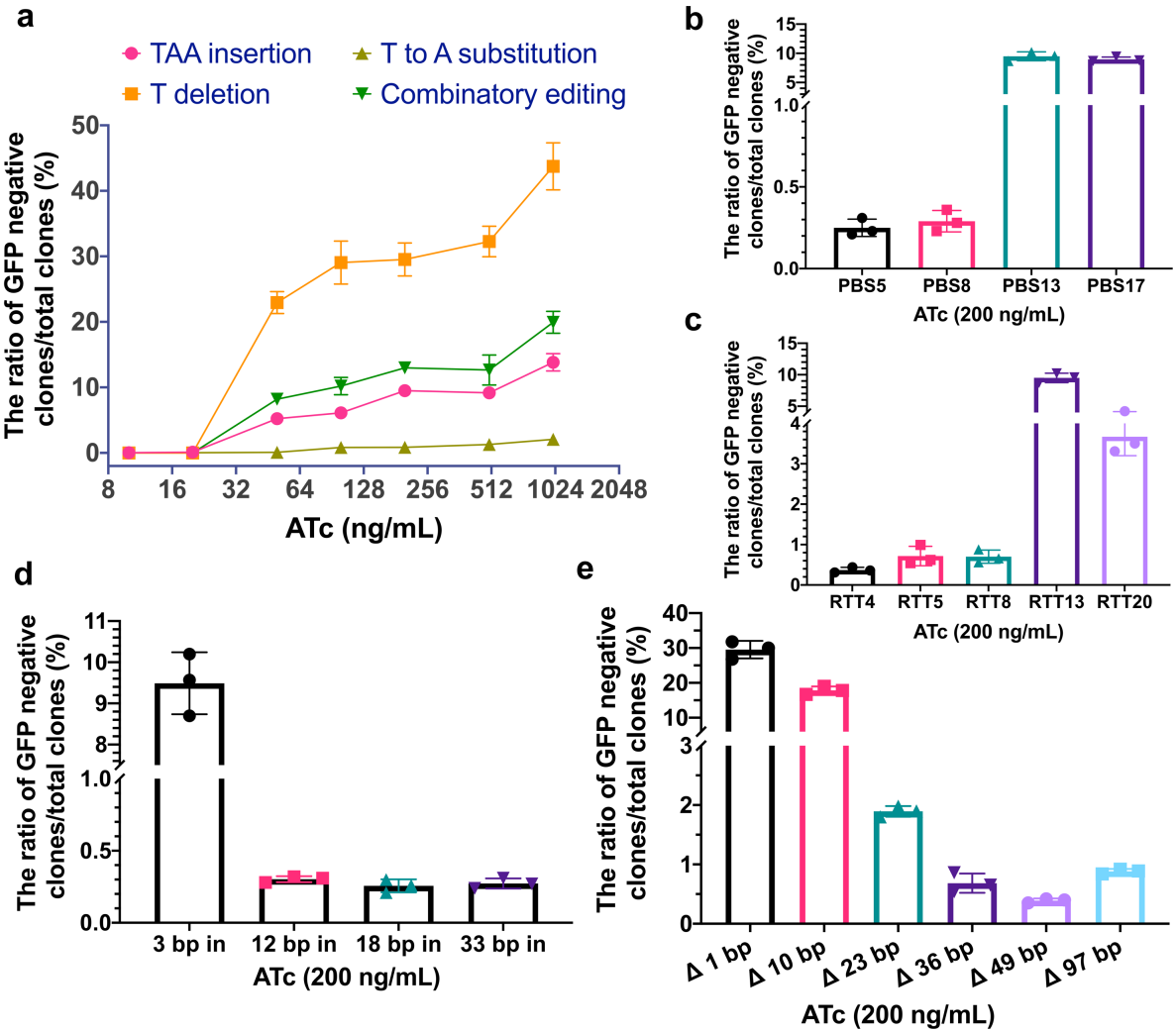


Figure 3: Characteristics of CRISPR-Prime Editing system for DNA engineering in *E. coli*. The editing efficiency was defined as ratio of white clones (GFP-negative)/ total clones on a screening plate. **a.** Eight different concentrations of ATc, ranging from 0 ng/mL to 1000 ng/mL (0, 10, 20, 50, 100, 200, 500, and 1000) were used to evaluate the induction of CRISPR-Prime Editing system on four DNA engineering events of 3-bp insertion, 1-bp deletion, 1-bp substitution, and the combinatorial editing. **b.** The evaluation of PBS length. **c.** The evaluation of RTT scaffold length. **d.** The capacity of DNA fragment insertion with different sizes. **e.** The capacity of DNA fragment deletion with different sizes. Mean ± s.d. of three biological replicates are shown. 200 ng/ml of ATc was used for **b.** to **e.**

Inspired by the observation that a second nick in the non-edited strand would increase the editing efficiency of CRISPR-Prime Editing in some mammalian[20] and plant cells[21], we designed and validated two strategies of the second nick

introduction in *E. coli* (Supplementary Note 1). There were almost no visible colonies after the second nick was introduced (Supplementary Fig. 4). This result indicates that in *E. coli,* which does not have a NHEJ pathway, introducing the second nick cannot increase the editing efficiency but only compromises the use of CRISPR-Prime Editing system.

Assessing the ability of chromosomal DNA editing with CRISPR-Prime Editing system in E. coli

Beyond editing plasmid DNA, we also assessed if CRISPR-Prime Editing system is capable of engineering chromosomal DNA in *E. coli*. To this end, two metabolic pathways for lactose and D-galactose degradation in *E. coli* MG1655 were selected. β-galactosidase, encoded by the *lacZ* gene within the lactose metabolic pathway, metabolizes X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside, an analog of lactose) into 5-bromo-4-chloro-indoxyl, which will form dark blue 5,5'-dibromo-4,4'-dichloro-indigo by oxidation. On the contrary, X-gal remains colorless if the *lacZ* gene is inactivated (Fig. 4a and 4b). An early stop codon was designed to be introduced into the *lacZ* gene of *E. coli* MG1655 by insertion of TAG, deletion of GC, and substitution of GT to TA. In general, the editing efficiencies similar to those for plasmid DNA engineering were observed by counting the white colonies out of the total formed colonies on X-gal supplemented LB plates (Fig. 4c). Editing efficiencies of substitution, insertion and deletion were 6.8%, 12.2%, and 26%, respectively. To validate the editing events, the targeted region of eight randomly picked non-blue colonies from each designed DNA engineering event were PCR amplified and further subjected to Sanger sequencing. All sequenced clones bore the expected edits (Fig. 4d-4f). Moreover, another gene, the *galK* gene from the Leloir pathway of D-galactose metabolism in *E. coli* MG1655 was tested. The loss of function of the *galK* gene can be positively selected by supplementing a galactose analogue 2-deoxy-D-galactose (2-DOG), as 2-DOG will be metabolized by galactokinase (encoded by the *galK* gene) to form a toxic compound 2-deoxy-galactose-1-phosphate, which cannot be further metabolized[23] (Fig. 4g and 4h). A TAA stop codon was designed to be inserted into the *galK* gene in *E. coli* MG1655 strain. Of the visible colonies on the 2-DOG supplemented M63 agar plate, all four that were randomly picked showed the expected insertion (Fig. 4i).

Genome-wide on/off-target evaluation of the CRISPR-Prime Editing system in E. coli

Precision is one of the most important requirements for DNA engineering. We applied a bacterial genome-wide SNP (Single Nucleotide Polymorphism) profiling approach[22] to systematically evaluate the on-target and potential off-target mutations caused by the CRISPR-Prime Editing system. We selected one clone of each designed DNA engineering events of both plasmid- and chromosome-based editing (Table 1 and Table S1) for on/off target mutation analysis using whole-genome sequencing. In order to assess the background noise of mutations, we also sequenced the two parental strains DH10B and MG1655 that were used in this work (Table S1). As expected, all designed editing events (on-target mutations) were present in the corresponding strains (Table 1). For potential off-target mutation analysis, we examined mutations on both, plasmids and chromosome, using breseq[24]. Only one single nucleotide substitution in the chromosome was identified in the edited strains of 1-bp deletion, 1-bp substitution, and the combinatorial editing (Table 1). No off-target mutations were found in the 3-bp insertion and 2-bp substitution edited strains (Table 1). As we observed some target specific off-target mutations in the editing event of 1-bp substitution (Fig. 2b), we wanted to further investigate if long-distance off-target mutations would be introduced in the target specific off-target mutation carrying strain. To this end, another clone (DH10B-plasmids-PE_1bpsub) of 1-bp substitution and a Sanger sequencing recorded A deletion 5-bp upstream of this designed substitution was subjected to whole-genome sequencing analysis (Table 1). Except the expected on-target 1-bp substitution and the target specific off-target mutation, no additional off-target mutations were found (Table 1).

For the chromosome DNA editing events, we also observed a high-fidelity of CRISPR-Prime Editing system. Except the excision of the *insB1–insA* fragment (a mobile element) in some strains, only one single nucleotide substitution was found in the chromosome of the designed 3-bp insertion clone (Table 1).

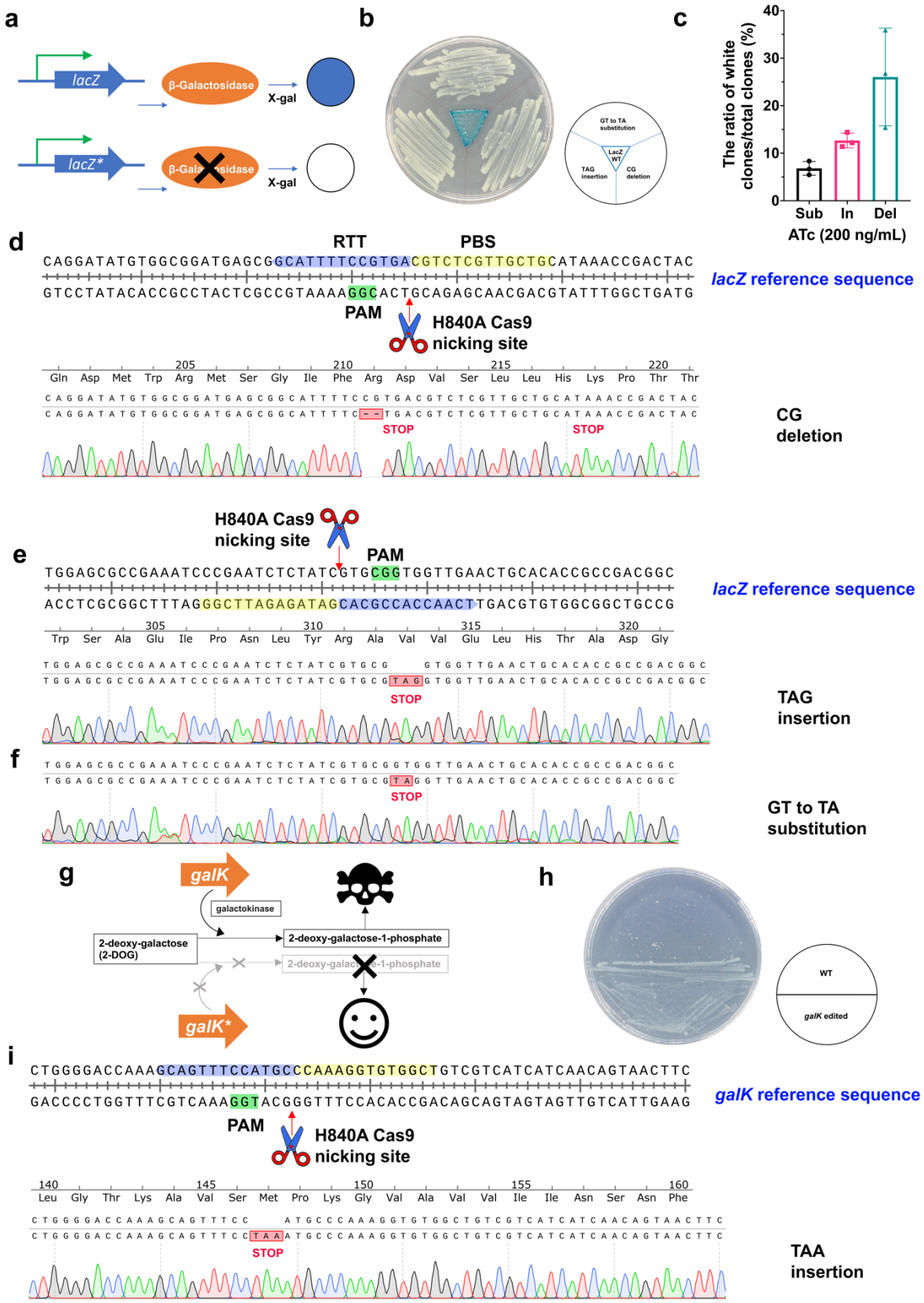


Figure 4: **CRISPR-Prime Editing for *E. coli* is capable of chromosomal DNA editing. a.** A graphic illustration of the function of *lacZ* gene. The star within the gene box represents a stop codon being introduced. **b.** Three clones of *E. coli* MG1655, where the inactivation of lacZ was confirmed by Sanger sequencing and a wild type *E. coli* MG1655 were re-streaked on

an agar plate with X-gal. **c.** A bar chart shows the editing efficiencies of chromosomal DNA engineering by 3-bp insertion, 2-bp deletion and 2-bp substitution by calculating the ratio of white clones/total clones on an induction LB plate with X-gal supplemented. **d.** to **f.** Eight non-blue colonies were picked and Sanger sequenced. Sequencing traces aligned with non-edited *lacZ* reference sequences are displayed. Shown are alignments of 2-bp deletion (d.); 3-bp (stop codon TAG) insertion (e.); and 2-bp substitution (f.). **g**. A graphic illustration of the function of *galK* gene. **h.** An agar plate view of the successful g*alK* gene inactivation in *E. coli* MG1655 strains by CRISPR-Prime Editing. **i.** Four colonies from the 2-DOG supplemented plate were picked and Sanger sequenced. Sequencing traces aligned with non-edited *galK* reference sequences are displayed. The alignment of 3-bp (stop codon TAA) insertion is shown. The potential Cas9 H840A nicking site is indicated by a red arrow. The 20 nt protospacer, the PAM sequence, the PBS, and the RTT are highlighted in pink, green, yellow, and blue, respectively. The translated amino acid sequences together with the introduced stop codons are labeled underneath each nucleotide sequence. Numbers in red on the right side show the correct and total Sanger sequenced colonies.

As it has been reported that MG1655 strain will lose the 776 bp insB1–insA fragment during cultivation[25], we therefore excluded this from the CRISPR-Prime Editing system related off-target effect, marking it as one of the parental variable mutations (Table 1). In summary, our results indicate a very high fidelity of CRISPR-Prime Editing system on both plasmid and chromosome DNA engineering in *E. coli*.

DISCUSSION

Due to the importance of *E. coli* in basic microbiological studies and biotechnological applications as a workhorse for the production of various bioproducts, there has been continued demand for novel and efficient DNA engineering tools. Widely used and versatile methods for genetic engineering of *E. coli* are RedET and lamda red-based recombineering[1, 2], or MAGE-based approaches[3-5]. Although much effort has been exerted to simplify and improve the recombineering protocol, it is either still relatively difficult to operate[1, 26], and it requires the target strain to have a specific genetic background, for example the deficiency of methyl-directed mismatch repair or RecA, the key enzyme for recombinational DNA repair[27]. The emergence of CRISPR-based genetic engineering has been revolutionizing biotechnology, however much less applications were reported in prokaryotes than in eukaryotes[8], partially because of the different dominant DSB repair pathways. As a result, in many bacteria, including *E. coli*, CRISPR-Cas9 has been widely employed as a tool for counter-selection to eliminate non-modified cells from a mixed population in homology-directed recombination methods such as the lambda red recombination systems[28]. It remains very challenging to engineer DNA at a single nucleotide level, even when combined with a powerful counter-selection system such as CRISPR-Cas9, the efficiency of making point mutations using oligonucleotide-directed mutagenesis is very low (< 3% before optimization)[29].

Recently, two types of CRISPR base editors were developed, which are capable of C to T conversion (CBE) by the cytosine deaminase (APOBEC1 or Target-AID)[30, 31], C to G or A substitution with engineered cytosine deaminases[32], and A to G conversion (ABE) by the adenosine deaminase (TadA)[16] without involving DSBs. Thus, they can be directly applied in bacteria for DNA manipulation. One of the main applications is gene inactivation using CBE to convert Arg, Gln, or Trp codons to a stop codon[15, 17]. There are also a few cases of using adenosine deaminase based base editing for *in vivo* protein engineering[15, 16]. However, so far, CRISPR base editing technology has not been as widely used in bacteria as expected due to the restriction of fixed substitutions (C to T/G/A, or A to G) and the relatively narrow editing window (5-7 nucleotides).

Table 1: Whole-genome sequencing-based analysis of on-target and off-target mutations.

Strain	Parental variable mutations	Recorded mutations				
		On-target				Off-target (Plasmids and/or chromsome)
		Genes	Regions in the target gene	References (coding strand)	Alleles	
DH10B-plasmids-PE_3bpin	0	GFP in plasmid: pCDF-GFPplus	283	-	+TAA	0
DH10B-plasmids-PE_1bpdel	0	GFP in plasmid: pCDF-GFPplus	283	T	-	1 (chr. Pos. 396,221. C→T)
DH10B-plasmids-PE_1bpsub	0	GFP in plasmid: pCDF-GFPplus	283	T	A	1 (pCDF-GFPplus pos. 278. -A)
DH10B-plasmids-PE_1bpsub-2	0	GFP in plasmid: pCDF-GFPplus	283	T	A	1 (chr. Pos. 2,198,023. A→G)
DH10B-plasmids-PE_2bpsub	0	GFP in plasmid: pCDF-GFPplus	281-283	TAT	CAC	0
DH10B-plasmids-PE_combo	0	GFP in plasmid: pCDF-GFPplus	283-284	TG	ATTA (T→A, +TAA, -G)	1 (chr. Pos. 4,047,371. G→T)
MG1655-chr-PE_3bpin	1 (chr. Pos. 1,976,527. Δ776 bp, insB1–insA)	lacZ in chromosome	364,594	-	+TAG	1 (chr. pos. 1,285,157. A→G)
MG1655-chr-PE_3bpin-2	1 (chr. Pos. 1,976,527. Δ776 bp, insB1–insA)	galK in chromosome	788,602	-	+TAA	0
MG1655-chr-PE_2bpdel	0	lacZ in chromosome	364,898-364,899	CG	-	0
MG1655-chr-PE_2bpsub	1 (chr. Pos. 1,976,527. Δ776 bp, insB1–insA)	lacZ in chromosome	364,592-364,593	GT	TA	0

We demonstrated in this study that CRISPR-Prime Editing system cannot only make substitutions but also insertion, deletion, and combinatorial editing at single base pair resolution in *E. coli* without requiring DSBs, editing templates, or homologous recombination. We observed a very high fidelity of using the

system in *E. coli*, while the mutation rates/off-target effects in the MAGE and other recombineering systems are much higher, and normally it requires pre-engineering of the host strains when using these systems[3-5]. CRISPR-Prime Editing system for *E. coli* has great potential in expanding the possibilities of DNA engineering, although further studies are required to further increase its editing efficiency. As a result of its high modularity and simple composition, CRISPR-Prime Editing for *E. coli* might be multiplexed by providing a PEgRNA self-processing machinery like Csy4[15] and consequently applied for high-throughput mutagenesis applications. However, it has to be noted that the editing efficiency was extremely low in our proof-of-concept multiplexing approach using the strategy of providing two PEgRNA delivery plasmids. CRISPR-Prime Editing also has the potential of being applicable to a wider range of bacteria including those previously considered difficult to be genetically engineered. As in the case of CRISPR-Cas9 and CRISPR base editing systems, the use of other Cas proteins and protein engineering will likely improve editing capabilities by expanding the selection of accepted PAMs and increasing efficiencies[33-36]. Different reverse transcriptases other than the M-MLV could also provide different features to increase the performance of CRISPR prime editing systems in applied organisms. Modulating intracellular DNA repair systems and better designed PEgRNAs could also be helpful in increasing the editing efficiency.

CRISPR-Prime Editing, a versatile DNA engineering system reported in this study, represents a powerful addition to the toolbox of genetic and metabolic engineers not only for *E. coli*, but other organisms. These tools are likely to substantially advance our understanding of basic life science and to increase capabilities for advanced microbial engineering for biotechnological purposes.

MATERIALS AND METHODS

Strains, plasmids, media and growth condition

All *Escherichia coli* strain and plasmids used in this study are listed in Supplementary Table 1. *E. coli* cultures were grown at 37 °C in LB (both broth and solid) (Sigma, US). Appropriate antibiotics were supplemented with the following working concentrations: spectinomycin (50 μg/mL), carbenicillin or ampicillin (100 μg/mL), chloramphenicol (25 μg/mL) kanamycin (50 μg/mL) and anhydrotetracycline (0 to 1 μg/mL). M63 minimal medium was used for positive selection of *galK* mutants. It is composed of 2 g/L $(NH_4)_2SO_4$, 13.6 g/L KH_2PO_4, 0.5 mg/L $FeSO_4$-$7H_2O$, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$ and 10 μg/mL thiamine, 0.2% glycerol and 0.1% 2-deoxy-D-galactose. 2% agar was supplemented when making agar plates. X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside) was used for screening *lacZ* mutants. Prior to use, each LB plate with appropriate antibiotics is plated with 40 μL of 20 mg/mL X-gal. All chemicals involved in this study were from Sigma, US.

General protocol of DNA manipulation

All primers, important sequences, spacers and 3 prime extensions used in this study are listed in Supplementary Tables 2, 3, and 4, respectively. Standard protocols were used for DNA (plasmids and genomic DNA) purification, PCR, and cloning. PCR was performed using Q5® High-Fidelity 2× Master Mix (New England Biolabs, US). The point mutation in dCas9 to create H840A Cas9n was made using Q5® Site-Directed Mutagenesis Kit (New England Biolabs, US). DNA assembly was done by using NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs, US) unless specified otherwise. DNA digestion was performed with FastDigest restriction enzymes (Thermo Fisher Scientific, US) unless specified otherwise. NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, Germany) was used for DNA clean-up both from PCR products and agarose gel extracts. NucleoSpin® Plasmid EasyPure Kit (Macherey-Nagel, Germany) was used for plasmid preparation. Sanger sequencing was carried out using Mix2Seq kit (Eurofins Scientific, Luxembourg). DNA fragments were synthesized by Genscript while oligonucleotides were synthesized by IDT (Integrated DNA Technologies, US).

All kits and enzymes were used according to the manufacturers' recommendations. We diligently followed all waste disposal regulations of our institute, university, and local government when disposing of waste materials.
Multi-plasmid system design and plasmid construction

All plasmids constructed in this study have been deposited to Addgene, individual Addgene plasmid number are listed below. Plasmids in the same testing system should be compatible with each other, and therefore they must have different origins of replication (ori). For this purpose, a combination of p15A ori, ColE1 ori, and ColDF13 ori was used.

Synthetic constitutive promoters J23119 (BBa_J23119) and J23106 (BBa_J23106), and the ribosome binding site (RBS) BBa_B0034 were obtained from the registry for standard biological parts in the iGEM Parts Registry (http://parts.igem.org/Main_Page).

The construction of GFP-based reporter plasmid: The plasmid was designed *in silico* to carry the GFP expression cassette, which is composed of a constitutive promoter J23106, a RBS BBa_B0034, a fast folding GFP variant GFP+[22], and a terminator T0. The GFP+ coding sequence was codon optimized to *E. coli*. The whole cassette was synthesized by Genscript and assembled into the pCDF-1b plasmid (ColDF13 ori, Millipore, US) replacing the MCS region by Gibson Assembly, and ended up with the plasmid pCDF-GFPplus.

The construction of CRISPR-Prime Editing plasmid: Firstly, we created pCas9n(H840A) from pdCas9-bacteria (p15A ori, Addgene plasmid #44249)[37] by site-specific mutation of 10A of dCas9 to 10D using Q5® Site-Directed Mutagenesis Kit (New England Biolabs, US). Secondly, we designed the 33a linker-M-MLV2 cassette *in silico*. Linker sequence: SGGSSGGSSGSETPGTSESATPESSGGSSGGSS. M-MLV2, a moloney murine leukaemia virus (M-MLV) variant from a previous study[20] with the following mutations D200N, L603W, T306K, W313F, and T330P compared to the WT M-MLV (GenBank: AAC82568.2). Thirdly, the cassette was codon optimized to *E. coli*, synthesized by Genscript, and then assembled into pCas9n(H840A) to replace the stop codon of Cas9n by Gibson Assembly, resulting in the plasmid pCRISPR-PE-bacteria (Addgene # 172715). The fusion protein (cargo) Cas9n-linker-M-MLV2 is under control by a tetracycline inducible promoter.

The construction of PEgRNA transcript carrying plasmid: The empty PEgRNA plasmid was modified from the pgRNA-bacteria (ColE1 ori, Addgene plasmid #44251)[37] by removing the 20 bp spacer, named as pPEgRNA (Addgene # 172716). For construction of functional pPEgRNA there were three steps: firstly, a spacer and 3 prime extensions were designed *in silico*; secondly, amplification of the functional PEGgRNA cassette using the pPEgRNA as a template was performed concurrently with amplifying the PEgRNA backbone fragment using the primer set (PEgRNA backbone_F and PEgRNA backbone_R); lastly, the functional PEgRNA cassette was assembled into the PEgRNA backbone. Sanger sequencing was used for validation. Spacers and 3 prime extensions were designed both manually and using PrimeDesign[38].

For introducing the second nick, we constructed pnsgRNA (pSC101 ori, kanR) by replacing the sfGFP expression cassette in pVRb20_992 (Addgene plasmid #49714)[39] with the sgRNA transcript cassette from pPEgRNA. We first amplified the plasmid backbone of pVRb20_992 and the sgRNA cassette with primer sets of pVRb_backbone_F and pVRb_backbone_R, and sgRNA_cassette_F and sgRNA_cassette_R from pVRb20_992 and pPEgRNA, respectively. Then these two fragments were Gibson assembled and later validated by Sanger sequencing, resulting in pnsgRNA plasmid (Addgene # 172717). Spacers for introducing the second nick in the nsgRNA paired with the related PEgRNA were designed using PrimeDesign[38]. This plasmid is also used for delivery of the second PEgRNA.

High throughput electroporation of multiple plasmids

In vivo assay of strains carrying multiple plasmids were performed from freshly transformed *E. coli* DH10β strains. A HT Nucleofector™ System (Lonza, Switzerland) together with 96-well Nucleocuvette plates (Lonza, Switzerland) were used for high throughput electroporation. Before electroporation, the 96-well Nucleocuvette plate was transferred from -20 °C to ice for 10 min. 20 μL of electrocompetent DH10β or MG1655 *E. coli* cells with 10% glycerol were added into each desired well, 0.5 μL of each plasmid (about 30 ng) was subsequently added. A total amount of plasmid DNA of <100 ng per transformation normally performed well. The program used in this study is X_bacteria_14, with the code GN-100. After electroporation, 180 μL of fresh LB broth were added into each

well. The cultures were then transferred into a 96-deep well plate containing 200 μL of fresh LB broth (making the transformation culture in total 400 μL) for recovery for 1h at 37 °C and 300 rpm.

Illumina deep sequencing-based genome-wide on/off-target evaluation and analysis of "escapers"

For on/off-target evaluation, one or two Sanger sequencing validated clones of each designed editing events were selected; while for escapers examination, ten clones with induced CRISPR-Prime Editing systems targeting pCDF-GFPplus, still showing GFP-fluorescence, were randomly picked. Together with necessary control strains, they were inoculated in a 50-mL tube (Greiner Bio-One, Germany) containing 10 mL LB broth without any antibiotics. After incubating at 37 °C, 300 rpm in an INNOVA 44R incubator shaker (Eppendorf, Germany) for 24 hours, 5 mL of the culture was used for genomic DNA plus intracellular plasmid DNA isolation with a Blood & Cell Culture DNA mini Kit (Cat No./ID: 13323, Qiagen, Germany). While a NucleoSpin® Plasmid EasyPure Kit (Macherey-Nagel, cat. no. 740727.250) was used for the WT pCDF-GFP plasmid isolation. The genomic library construction and illumina paired-end sequencing were carried out by Novogene Co., Ltd. (Beijing, China), using the NEB Next® Ultra™ DNA Library Prep Kit (New England Biolabs, US) with a target insert size of 350nt and 6 PCR cycles.

The illumina reads obtained from the sequenced samples were trimmed using Trim Galore (v. 0.6.4_dev, Cutadapt v. 2.10) with the switches --length 100 and --quality 20. All mutation calls were performed using breseq (v. 0.33.2, bowtie2 v. 2.3.4.1)[24,40] with default parameters. For plasmid-based editing, the *E. coli* DH10B genome sequence NC_010473 is used as the reference, while for chromosome-based editing, the *E. coli* MG1655 genome sequence NC_ U00096 is used as the reference, both along with the relevant plasmids. Mutation calls that existed in all samples as well as the parental strain were not counted as off-target effects.

Editing efficiency evaluation using a fluorescence-based colony counting assay

50 μL electroporation culture (400 μL in the cases of a second nick is introduced) of each strain was plated onto appropriate antibiotics containing LB agar plates supplemented with and without inducer, respectively. All plates were covered by aluminium foil and incubated at 37 °C for 24 h. After cultivation, total colonies were counted by a Doc-It imaging station (Fisher Scientific, US.) with a trisection protocol. Non-fluorescent colonies in each zone of all three zones were further counted with and without a Blue-Light Transilluminator (Safe Imager 2.0, Thermo Fisher Scientific, US). The editing efficiency was calculated as: the number of non-green colonies in each zone/total number of visible colonies in the same zone.

Editing events confirmation by Sanger sequencing

Eight to 24 primarily identified positive clones of each strain were picked, and inoculated into 5 mL LB broth with proper antibiotics. After overnight (~16 h) cultivation, cultures were subjected to plasmid isolation using the NucleoSpin® Plasmid EasyPure Kit (Macherey-Nagel, Germany) or colony PCR using Q5® High-Fidelity 2× Master Mix (New England Biolabs, US) if a chromosomal region was targeted. The isolated plasmids and the cleaned PCR products were Sanger sequenced using the Mix2Seq kit (Eurofins Scientific, Luxembourg) with proper primers. The obtained sequence traces were analyzed and visualized using SnapGene (GSL Biotech, US).

ACKNOWLEDGMENTS

We thank Simon Shaw for proofreading the manuscript. We thank Alexandra Hoffmeyer for excellent technical support in Illumina sequencing. This work was supported by grants from the Novo Nordisk Foundation (NNF10CC1016517, NNF15OC0016226, NNF16OC0021746). S.Y.L. was also supported by the Technology Development Program to Solve Climate Changes on Systems Metabolic Engineering for Biorefineries (NRF-2012M1A2A2026556 and NRF-2012M1A2A2026557) from the Ministry of Science and ICT through the National Research Foundation (NRF) of Korea.

COMPETING INTERESTS

The authors have no competing interests.

DATA AVAILABILITY

All data used in this study can be found in the manuscript and/or in the supplementary information. All Illumina sequencing data has been deposited to Dryad.
Data for on/off target evaluation: https://datadryad.org/stash/share/31zjCW-UpyxritX2r2TAL7VMkqPtXxzc9oRAp3PqMBw or doi:10.5061/dryad.1rn8pk0s9.

Data for editing escapers: https://datadryad.org/stash/share/MopTidFXxNNROC2mnIDxmasMImuFS2k_7Syy-W81RZk or doi: 10.5061/dryad.wstqjq2mj.

AUTHOR CONTRIBUTION

Y.T., T.W., and S.Y.L, conceived and designed the project. Y.T. and C.M.W. carried out the laboratory experiments and analyzed the data. T.S.J. performed computational analysis. Y.T., T.W. and S.Y.L. wrote the manuscript with input from all authors.

REFERENCES

1. Costantino, N. & Court, D.L. Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. U. S. A.* 100, 15748-15753 (2003).
2. Muyrers, J.P., Zhang, Y., Testa, G. & Stewart, A.F. Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res* 27, 1555-1557 (1999).
3. Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460, 894-898 (2009).
4. Nyerges, A. et al. A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proceedings of the National Academy of Sciences of the United States of America* 113, 2502-2507 (2016).
5. Wannier, T.M. et al. Improved bacterial recombineering by parallelized protein discovery. *Proc Natl Acad Sci U S A* 117, 13689-13698 (2020).
6. Schubert, M.G. et al. High-throughput functional variant screens via in vivo production of single-stranded DNA. *Proceedings of the National Academy of Sciences of the United States of America* 118 (2021).
7. Barrangou, R. et al. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* 315, 1709-1712 (2007).
8. Sander, J.D. & Joung, J.K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* 32, 347-355 (2014).
9. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821 (2012).
10. Kanaar, R., Hoeijmakers, J.H.J. & van Gent, D.C. Molecular mechanisms of DNA double-strand break repair. *Trends Cell Biol* 8, 483-489 (1998).
11. Cui, L. & Bikard, D. Consequences of Cas9 cleavage in the chromosome of *Escherichia coli*. *Nucleic Acids Res* 44, 4243-4251 (2016).
12. Cui, L. & Bikard, D. Consequences of Cas9 cleavage in the chromosome of *Escherichia coli*. *Nucleic Acids Res* 44, 4243-4251 (2016).
13. Hsu, P.D., Lander, E.S. & Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* 157, 1262-1278 (2014).
14. Cho, J.S. et al. CRISPR/Cas9-coupled recombineering for metabolic engineering of *Corynebacterium glutamicum*. *Metab Eng* 42, 157-167 (2017).
15. Tong, Y. et al. Highly efficient DSB-free base editing for streptomycetes with CRISPR-BEST. *Proc. Natl. Acad. Sci. U. S. A.* 116, 20366-20375 (2019).

16. Gaudelli, N.M. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 551, 464-471 (2017).
17. Banno, S., Nishida, K., Arazoe, T., Mitsunobu, H. & Kondo, A. Deaminase-mediated multiplex genome editing in *Escherichia coli*. *Nat Microbiol* 3, 423-429 (2018).
18. Strecker, J. et al. RNA-guided DNA insertion with CRISPR-associated transposases. *Science* 365, 48-53 (2019).
19. Klompe, S.E., Vo, P.L.H., Halpin-Healy, T.S. & Sternberg, S.H. Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. *Nature* 571, 219-225 (2019).
20. Anzalone, A.V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149-157 (2019).
21. Lin, Q. et al. Prime genome editing in rice and wheat. *Nat Biotechnol* 38, 582-585 (2020).
22. Scholz, O., Thiel, A., Hillen, W. & Niederweis, M. Quantitative analysis of gene expression with an improved green fluorescent protein. *Eur J Biochem* 267, 1565-1570 (2000).
23. Warming, S., Costantino, N., Court, D.L., Jenkins, N.A. & Copeland, N.G. Simple and highly efficient BAC recombineering using *gal*K selection. *Nucleic Acids Res* 33, e36 (2005).
24. Deatherage, D.E. & Barrick, J.E. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol Biol* 1151, 165-188 (2014).
25. Khare, A. & Tavazoie, S. Multifactorial Competition and Resistance in a Two-Species Bacterial System. *PLoS Genet* 11 (2015).
26. Lajoie, M.J., Gregg, C.J., Mosberg, J.A., Washington, G.C. & Church, G.M. Manipulating replisome dynamics to enhance lambda Red-mediated multiplex genome engineering. *Nucleic Acids Res* 40, e170-e170 (2012).
27. Wang, J. et al. An improved recombineering approach by adding RecA to lambda Red recombination. *Mol Biotechnol* 32, 43-53 (2006).
28. Pyne, M.E., Moo-Young, M., Chung, D.A. & Chou, C.P. Coupling the CRISPR/Cas9 System with Lambda Red Recombineering Enables Simplified Chromosomal Gene Replacement in *Escherichia coli*. *Appl Environ Microbiol* 81, 5103-5114 (2015).
29. Lee, H.J., Kim, H.J. & Lee, S.J. CRISPR-Cas9-mediated pinpoint microbial genome editing aided by target-mismatched sgRNAs. *Genome Res* (2020).

30. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. & Liu, D.R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420-424 (2016).
31. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353 (2016).
32. Zhao, D. et al. Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat Biotechnol* (2020).
33. Kim, Y.B. et al. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol* 35, 371-376 (2017).
34. Mitsunobu, H., Teramoto, J., Nishida, K. & Kondo, A. Beyond Native Cas9: Manipulating Genomic Information and Function. *Trends Biotechnol* 35, 983-996 (2017).
35. Walton, R.T., Christie, K.A., Whittaker, M.N. & Kleinstiver, B.P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* 368, 290-296 (2020).
36. Richter, M.F. et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat Biotechnol* (2020).
37. Qi, L.S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173-1183 (2013).
38. Hsu, J.Y. et al. PrimeDesign software for rapid and simplified design of prime editing guide RNAs.
39. Rhodius, V.A. et al. Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters. *Mol Syst Biol* 9, 702 (2013).
40. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357-359 (2012).

SUPPLEMENTARY INFORMATION

Supplementary note 1

A second nick does not increase the editing efficiency of CRISPR-Prime Editing system in E. coli

It has been reported that the editing efficiency of the reverse transcriptase-Cas9 H840A nickase (Cas9n)-mediated targeted prime editing in some mammalian cells[1] and plant cells[2] can be increased by introduction of a second nick to the complementary strand within around 100 bp from the site of the PEgRNA-induced nick. However, it also increases NHEJ-mediated indel formation because Cas9 nickase with paired sgRNAs within ~200 bp can introduce targeted DSBs[3,4]. As most bacteria do not possess a complete NHEJ system[5-7], it makes the DSB a lethal event if no homology-directed repair (HDR) template is provided. It was thus reasoned that it is impossible to apply the strategy of introducing a close nick in the complementary strand in the NHEJ-deficient bacteria, like *E. coli*. To evaluate our hypothesis, we introduced another plasmid to deliver the designed complementary strand nicking sgRNA (nsgRNA). Two approaches for the secondary nick introduction were tested: nicking the non-edited strand within ~200 bp from the first nick (CRISPR-PE3) and nicking the non-edited strand only after the first nicked strand (the complementary strand) has been edited (CRISPR-PE3b). We designed editing events accordingly both in the plasmid-based system and the chromosome. As expected, for all three designed chromosomal DNA engineering and one plasmid DNA deletion of CRISPR-Prime Editing, almost no visible colonies were observed after the second nick was introduced (Supplementary Fig. 3).

Supplementary Table 1: Plasmids and strains involved in this study.

Plasmid	Background	Reference
pdCas9-bacteria	*E. coli* codon optimized dCas9 is under control by a tetracycline inducible promoter; CamR; p15A ori	Addgene #44249[8]
pgRNA-bacteria	sgRNA transcript carrying plasmid, under control by a constitutive promoter J23119; AmpR; ColE1 ori	Addgene #44251[8]
pCDF-b1	SmR; CloDF13 ori	Millipore, US
pCDF-GFPplus	A fast folding GFP variant GFP+ is cloned into pCDF-b1 under control by a constitutive promoter	This study
pCRISPR-PE-bacteria	pdCas9- bacteria is used as the backbone. An *E. coli* codon optimized fusion protein of Cas9n-linker-M-MLV2 is cloned into pdCas9- bacteria by replacing the dCas9. The fusion protein is under control by a tetracycline inducible promoter.	This study
pPEgRNA	The 20 bp spacer was removed from pgRNA-bacteria. Therefore, this plasmid carries only a sgRNA scaffold without a spacer.	This study
pVRb20_992	KanR; pSC101 ori; carrying a cassette of PEcf20_992-sfGFP	Addgene #49714[9]
pnsgRNA	An sgRNA transcript cassette from pgRNA-bacteria was inserted to replace the PEcf20_992-sfGFP cassette	This study
pVRb_PEgRNA_312Gdel	pVRb_PEgRNA carries a 20 nt spacer targeting GFP coding gene, and a 3 prime extension for deleting the 312G, the length of both the reverse transcription template and the primer binding sequence is 13 bp	This study
pPEgRNA_GFP_TAAin	pPEgRNA carries a 20 nt spacer targeting GFP coding gene, and a 3 prime extension for inserting a stop codon TAA, the length of both the reverse transcription template and the primer binding sequence is 13 bp	This study
pPEgRNA_GFP_TAAin_T4	The same as pPEgRNA_GFP_TAAin, except that the length of the reverse transcription template is 4 bp	This study
pPEgRNA_GFP_TAAin_T5	The same as pPEgRNA_GFP_TAAin, except that the length of the reverse transcription template is 5 bp	This study
pPEgRNA_GFP_TAAin_T8	The same as pPEgRNA_GFP_TAAin, except that the length of the reverse transcription template is 8 bp	This study
pPEgRNA_GFP_TAAin_T20	The same as pPEgRNA_GFP_TAAin, except that the length of the reverse transcription template is 20 bp	This study

pPEgRNA_GFP_TAAin_PBS5	The same as pPEgRNA_GFP_TAAin, except that the length of the primer binding sequence is 5 bp	This study
pPEgRNA_GFP_TAAin_PBS8	The same as pPEgRNA_GFP_TAAin, except that the length of the primer binding sequence is 8 bp	This study
pPEgRNA_GFP_TAAin_PBS17	The same as pPEgRNA_GFP_TAAin, except that the length of the primer binding sequence is 17 bp	This study
pPEgRNA_GFP_T198A	pPEgRNA carries a 20 nt spacer targeting GFP coding gene, and a 3 prime extension for 198T to 198A substitution within the GFP coding gene, the length of both the reverse transcription template and the primer binding sequence is 13 bp	This study
pPEgRNA_GFP_T196C_T198C	pPEgRNA carries a 20 nt spacer targeting GFP coding gene, and a 3 prime extension for 196T，198T to 196C, 198C substitutions within the GFP coding gene, the length of both the reverse transcription template and the primer binding sequence is 13 bp	This study
pPEgRNA_GFP_198Tdel	pPEgRNA carries a 20 nt spacer targeting GFP coding gene, and a 3 prime extension for 198T deletion within the GFP coding gene, the length of both the reverse transcription template and the primer binding sequence is 13 bp	This study
pPEgRNA_GFP_combo	pPEgRNA carries a 20 nt spacer targeting GFP coding gene, and a 3 prime extension for 198T to 198A substitution, TAA insertion, and 199G deletion within the GFP coding gene, the length of both the reverse transcription template and the primer binding sequence is 13 bp	This study
pPEgRNA_GFP_12in	pPEgRNA carries a 20 nt spacer targeting GFP coding gene, and a 3 prime extension for insertion of a 12 bp fragment within the GFP coding gene, the length of both the reverse transcription template and the primer binding sequence is 13 bp	This study
pPEgRNA_GFP_18in	pPEgRNA carries a 20 nt spacer targeting GFP coding gene, and a 3 prime extension for insertion of a 18 bp fragment within the GFP coding gene, the length of both the reverse transcription template and the primer binding sequence is 13 bp	This study
pPEgRNA_GFP_33in	pPEgRNA carries a 20 nt spacer targeting GFP coding gene, and a 3 prime extension for insertion of a 33 bp fragment within the GFP coding gene, the length of both the reverse transcription template and the primer binding sequence is 13 bp	This study
pPEgRNA_GFP_10del	pPEgRNA carries a 20 nt spacer targeting GFP coding gene, and a 3 prime extension for deletion of a 10 bp fragment within the GFP coding gene, the length of both the reverse transcription template and the primer binding sequence is 13 bp	This study

pPEgRNA_GFP_23del	pPEgRNA carries a 20 nt spacer targeting GFP coding gene, and a 3 prime extension for deletion of a 23 bp fragment within the GFP coding gene, the length of both the reverse transcription template and the primer binding sequence is 13 bp	This study
pPEgRNA_GFP_36del	pPEgRNA carries a 20 nt spacer targeting GFP coding gene, and a 3 prime extension for deletion of a 36 bp fragment within the GFP coding gene, the length of both the reverse transcription template and the primer binding sequence is 13 bp	This study
pPEgRNA_GFP_49del	pPEgRNA carries a 20 nt spacer targeting GFP coding gene, and a 3 prime extension for deletion of a 49 bp fragment within the GFP coding gene, the length of both the reverse transcription template and the primer binding sequence is 13 bp	This study
pPEgRNA_GFP_97del	pPEgRNA carries a 20 nt spacer targeting GFP coding gene, and a 3 prime extension for deletion of a 97 bp fragment within the GFP coding gene, the length of both the reverse transcription template and the primer binding sequence is 13 bp	This study
pPEgRNA_lacZ_TAGin	pPEgRNA carries a 20 nt spacer targeting *lacZ* gene, and a 3 prime extension for TAG insertion into the *lacZ* gene for a stop codon introduction, the length for the reverse transcription template and the primer binding sequence is 16 bp and 13 bp, respectively	This study
pPEgRNA_lacZ_CGdel	pPEgRNA carries a 20 nt spacer targeting *lacZ* gene, and a 3 prime extension for CG deletion in the *lacZ* gene for a stop codon introduction, the length for the reverse transcription template and the primer binding sequence is 18 bp and 14 bp, respectively	This study
pPEgRNA_lacZ_GTtoTAsub	pPEgRNA carries a 20 nt spacer targeting *lacZ* gene, and a 3 prime extension for GT to TA substitution in the *lacZ* gene for a stop codon introduction, the length of both the reverse transcription template and the primer binding sequence is 14 bp	This study
pPEgRNA_galK_TAAin	pPEgRNA carries a 20 nt spacer targeting *galK* gene, and a 3 prime extension for TAG insertion into the *galK* gene for a stop codon introduction, the length for the reverse transcription template and the primer binding sequence is 16 bp and 13 bp, respectively	This study
pnsgRNA_GFP_TAAin	pnsgRNA carries a 20 nt spacer that pairs with pPEgRNA_GFP_TAAin to introduce the second nick	This study
pnsgRNA_lacZ_TAGin	pnsgRNA carries a 20 nt spacer that pairs with pPEgRNA_lacZ_TAGin to introduce the second nick	This study
pnsgRNA_lacZ_CGdel	pnsgRNA carries a 20 nt spacer that pairs with pPEgRNA_lacZ_CGdel to introduce the second nick	This study
pnsgRNA_lacZ_GTtoTAsub	pnsgRNA carries a 20 nt spacer that pairs with pPEgRNA_lacZ_GTtoTAsub to introduce the second nick	This study

Strain	Background	Reference
Escherichia coli DH10β (DH10B)	str. K F- mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ (ara, leu)7697 galU galK λ- rpsL nupG /pMON14272 / pMON7124. Whole-genome sequenced for parental strain characterization.	Thermo Fisher Scientific, US
Escherichia coli MG1655	Str. K F-, lambda-, rph-1. Whole-genome sequenced for parental strain characterization.	Maintained in lab
PE0001	*E. coli* DH10β carries pCDF-GFPplus	This study
PE0002	*E. coli* DH10β carries pCRISPR-PE-bacteria	This study
PE0003	*E. coli* DH10β carries pCRISPR-PE-bacteria and pCDF-GFPplus	This study
PE0004	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA	This study
PE0005	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_TAAin. Whole-genome sequenced for mutation analysis.	This study
PE0006	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_TAAin_T4	This study
PE0007	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_TAAin_T5	This study
PE0008	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_TAAin_T8	This study
PE0009	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_TAAin_T20	This study
PE0010	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_TAAin_PBS5	This study
PE0011	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_TAAin_PBS8	This study
PE0012	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_TAAin_PBS17	This study
PE0013	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_T198A. Clone #1. Whole-genome sequenced for mutation analysis.	This study
PE0014	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_T198A. Clone #2. Whole-genome sequenced for mutation analysis.	This study

PE0015	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_T196C_T198C. Whole-genome sequenced for mutation analysis.	This study
PE0016	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_198Tdel. Whole-genome sequenced for mutation analysis.	This study
PE0017	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_combo. Whole-genome sequenced for mutation analysis.	This study
PE0018	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_12in	This study
PE0019	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_18in	This study
PE0020	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_33in	This study
PE0021	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_10del	This study
PE0022	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_23del	This study
PE0023	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_36del	This study
PE0024	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_49del	This study
PE0025	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_97del	This study
PE0026	*E. coli* MG1655 carries pCRISPR-PE-bacteria and pPEgRNA_lacZ_TAGin. Whole-genome sequenced for mutation analysis.	This study
PE0027	*E. coli* MG1655 carries pCRISPR-PE-bacteria and pPEgRNA_lacZ_CGdel. Whole-genome sequenced for mutation analysis.	This study
PE0028	*E. coli* MG1655 carries pCRISPR-PE-bacteria and pPEgRNA_lacZ_GTtoTAsub. Whole-genome sequenced for mutation analysis.	This study
PE0029	*E. coli* MG1655 carries pCRISPR-PE-bacteria, pPEgRNA_galK_TAAin. Whole-genome sequenced for mutation analysis.	This study

PE0030	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, pPEgRNA_GFP_TAAin, and pVRb_PEgRNA_312Gdel	This study
PE0031	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, pPEgRNA_GFP_198Tdel, and pVRb_PEgRNA_312Gdel	This study
PE0032	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, pPEgRNA_GFP_T198A, and pVRb_PEgRNA_312Gdel	This study
PE0033	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, pPEgRNA_GFP_combo, and pVRb_PEgRNA_312Gdel	This study
PE0034	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, pPEgRNA_GFP_T196C_T198C, and pVRb_PEgRNA_312Gdel	This study
PE0035	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_198Tdel. Under induction, it is green (GFP expressed). Whole-genome sequenced for mutation analysis to find clues of editing escaping. Escaper #1.	This study
PE0036	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_198Tdel. Under induction, it is green (GFP expressed). Whole-genome sequenced for mutation analysis to find clues of editing escaping. Escaper #2.	This study
PE0037	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_198Tdel. Under induction, it is green (GFP expressed). Whole-genome sequenced for mutation analysis to find clues of editing escaping. Escaper #3.	This study
PE0038	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_198Tdel. Under induction, it is green (GFP expressed). Whole-genome sequenced for mutation analysis to find clues of editing escaping. Escaper #4.	This study
PE0039	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_198Tdel. Under induction, it is green (GFP expressed). Whole-genome sequenced for mutation analysis to find clues of editing escaping. Escaper #5.	This study
PE0040	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_198Tdel. Under induction, it is green (GFP expressed). Whole-genome sequenced for mutation analysis to find clues of editing escaping. Escaper #6.	This study
PE0041	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_198Tdel. Under induction, it is green (GFP expressed). Whole-genome sequenced for mutation analysis to find clues of editing escaping. Escaper #7.	This study

PE0042	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_198Tdel. Under induction, it is green (GFP expressed). Whole-genome sequenced for mutation analysis to find clues of editing escaping. Escaper #8.	This study
PE0043	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_198Tdel. Under induction, it is green (GFP expressed). Whole-genome sequenced for mutation analysis to find clues of editing escaping. Escaper #9.	This study
PE0044	*E. coli* DH10β carries pCRISPR-PE-bacteria, pCDF-GFPplus, and pPEgRNA_GFP_198Tdel. Under induction, it is green (GFP expressed). Whole-genome sequenced for mutation analysis to find clues of editing escaping. Escaper #10.	This study
PE0045	*E. coli* DH10β carries pCDF-GFPplus. Induced. Whole-genome sequenced as a control.	This study
PE0046	*E. coli* DH10β carries pCRISPR-PE-bacteria. Induced. Whole-genome sequenced as a control.	This study
PE0047	*E. coli* DH10β carries pCDF-GFPplus and pCRISPR-PE-bacteria. Induced. Whole-genome sequenced as a control.	This study
PE0048	*E. coli* DH10β carries pCDF-GFPplus, pCRISPR-PE-bacteria, and a pPEgRNA without the 26-bp 3 prime extension. Induced, the clone is green. Whole-genome sequenced as a control.	This study

Supplementary Table 2: Primers used in this study.

Names	Sequence (5' to 3')	Purpose
PEgRNA_GFP_F	gtcctaggtataatactagtCTTGTCACTACTCTGACCTAgttttagagctagaaatagc	For PCR amplification of desired PEgRNA functional cassettes. Sequences in lower case are overhangs for Gibson assembly.
PEgRNA_GFP_TAAin_R	gggcccaagcttcaaaaaaaTCACTACTCTGACCTATTAAGGTGTTCAAgcaccgactcggtgccactt	
PEgRNA_GFP_TAAin_T4_R	gggcccaagcttcaaaaaaaTCACTACTCTGACCTATTAAgcaccgactcggtgccactt	
PEgRNA_GFP_TAAin_T5_R	gggcccaagcttcaaaaaaaTCACTACTCTGACCTATTAAGgcaccgactcggtgccactt	
PEgRNA_GFP_TAAin_T8_R	gggcccaagcttcaaaaaaaTCACTACTCTGACCTATTAAGGTGgcaccgactcggtgccactt	
PEgRNA_GFP_TAAin_T20_R	gggcccaagcttcaaaaaaaTCACTACTCTGACCTATTAAGGTGTTCAATGCTTTTgcaccgactcggtgccactt	
PEgRNA_GFP_TAAin_PBS5_R	gggcccaagcttcaaaaaaCTGACCTATTAAGGTGTTCAAgcaccgactcggtgccactt	
PEgRNA_GFP_TAAin_PBS8_R	gggcccaagcttcaaaaaaaACTCTGACCTATTAAGGTGTTCAAgcaccgactcggtgccactt	
PEgRNA_GFP_TAAin_PBS17_R	gggcccaagcttcaaaaaaaACTCTGACCTATTAAGGTGTTCAAgcaccgactcggtgccactt	
PEgRNA_GFP_T198A_R	gggcccaagcttcaaaaaaaTCACTACTCTGACCTAAGGTGTTCAAgcaccgactcggtgccactt	
PEgRNA_GFP_T196C_T198C_R	gggcccaagcttcaaaaaaaTCACTACTCTGACCCACGGTGTTCAAgcaccgactcggtgccactt	
PEgRNA_GFP_198Tdel_R	gggcccaagcttcaaaaaaaTCACTACTCTGACCTAGGTGTTCAATgcaccgactcggtgccactt	
PEgRNA_GFP_combo_R	gggcccaagcttcaaaaaaaTCACTACTCTGACCTAATAAGTGTTCAAgcaccgactcggtgccactt	
PEgRNA_GFP_12in_R	gggcccaagcttcaaaaaaaTCACTACTCTGACCTATGTAATCTGTACAGGTGTTCAAgcaccgactcggtgccactt	
PEgRNA_GFP_18in_R	gggcccaagcttcaaaaaaaTCACTACTCTGACCTATTAATACGACTCACTATAGGGTGTTCAAgcaccgactcggtgccactt	

PEgRNA_GFP_33in_R	gggcccaagcttcaaaaaaaTCACTACTCTGACCTATTGTAGAATCAGCCCACGAAACCGAGCGGGCGAGGGTGTTCAAgcaccgactcggtgccactt	
PEgRNA_GFP_10del_R	gggcccaagcttcaaaaaaaTCACTACTCTGACATATCTTTTCAAAGgcaccgactcggtgccactt	
PEgRNA_GFP_23del_R	gggcccaagcttcaaaaaaaTCACTACTCTGACATATCTTTTCAAAGgcaccgactcggtgccactt	
PEgRNA_GFP_36del_R	gggcccaagcttcaaaaaaaTCACTACTCTGACATATCTTTTCAAAGgcaccgactcggtgccactt	
PEgRNA_GFP_49del_R	gggcccaagcttcaaaaaaaTCACTACTCTGACATATCTTTTCAAAGgcaccgactcggtgccactt	
PEgRNA_GFP_97del_R	gggcccaagcttcaaaaaaaTCACTACTCTGACATATCTTTTCAAAGgcaccgactcggtgccactt	
PEgRNA_GFP_312Gdel_F	gtcctaggtataatactagtCTATATCTTTCAAAGATGACgttttagagctagaaatagc	For PCR amplification of GFP 312G deletion PEgRNA for Gibson assembly.
PEgRNA_GFP_312Gdel_R	gggcccaagcttcaaaaaaaATCTTTCAAAGATGACGGAACTACAAGCACCGACTCGGTGCCACTT	
pVRb_PEgRNA_backbone_F	TTTTTTTGAAGCTTGGGCCC	For PCR amplification of plasmid backbone for Gibson assembly of PEgRNA.
pVRb_PEgRNA_backbone_R	ACTAGTATTATACCTAGGACTGAGC	
pPEgRNA_lacZ_TAGin_F	gtcctaggtataatactagtAATCCCGAATCTCTATCGTGgttttagagctagaaatagc	For PCR amplification of desired functional PEgRNA_lacZ_TAGincassettes. Sequences in lower case are overhangs for Gibson assembly.
pPEgRNA_lacZ_TAGin_R	gggcccaagcttcaaaaaaaCCGAATCTCTATCGTGCGTAGGTGGTTGAgcaccgactcggtgccactt	
pPEgRNA_lacZ_CGdel_F	gtcctaggtataatactagtTATGCAGCAACGAGACGTCAgttttagagctagaaatagc	For PCR amplification of desired functional PEgRNA_lacZ_CGde cassettes. Sequences in lower case are overhangs for Gibson assembly.
pPEgRNA_lacZ_CGdel_R	gggcccaagcttcaaaaaaaGCAGCAACGAGACGTCAGAAAATGCCGCTCATgcaccgactcggtgccactt	
pPEgRNA_lacZ_GToTAsub_F	gtcctaggtataatactagtGCGAGTTGCGTGACTACCTAgttttagagctagaaatagc	For PCR amplification of desired functional

pPEgRNA_lacZ_GTtoTAsub_R	gggcccaagcttcaaaaaaaAGTTGCGTGACTACTAACGGGTAACAGTgcaccgactcggtgccactt	PEgRNA_lacZ_GTtoTAsub cassettes. Sequences in lower case are overhangs for Gibson assembly.
pPEgRNA_galK_TAAin_F	gtcctaggtataatactagtGACAGCCACACCTTTGGGCAgttttagagctagaaatagc	For PCR amplification of desired functional PEgRNA_galK_TAAin cassettes. Sequences in lower case are overhangs for Gibson assembly.
pPEgRNA_galK_TAAin_R	gggcccaagcttcaaaaaaaGCCACACCTTTGGGCATTTAGGAAACTGCgcaccgactcggtgccactt	
lacZ_check_F	gatgaaagctggctacagga	For PCR amplification of the targeted region in *lacZ* gene. The forward primer is also used for Sanger sequencing.
lacZ_check_F	tgacggttaacgcctcgaat	
galK_check_F	caatgggctaactacgttcg	For PCR amplification of the targeted region in *galK* gene. The forward primer is also used for Sanger sequencing.
galK_check_F	gtcgccaatcacagctttga	
pnsgRNA_GFP_TAAin_F	AAAGCATTGAACACCttaATGTTTTAGAGCTAGAAATAGC	For PCR amplification of the nsgRNA_GFP_TAAin cassette.
pnsgRNA_GFP_TAAin_R	ATtaaGGTGTTCAATGCTTTactagtattatacctaggac	
pnsgRNA_lacZ_TAGin_F	CctaCGCACGATAGAGATTCGTTTTAGAGCTAGAAATAGC	For PCR amplification of the nsgRNA_lacZ_TAGin cassette.
pnsgRNA_lacZ_TAGin_R	GAATCTCTATCGTGCGtagGactagtattatacctaggac	
pnsgRNA_lacZ_CGdel_F	CTGGAGTGACGGCAGTTATCGTTTTAGAGCTAGAAATAGC	For PCR amplification of the nsgRNA_lacZ_CGdel cassette.
pnsgRNA_lacZ_CGdel_R	GATAACTGCCGTCACTCCAGactagtattatacctaggac	
pnsgRNA_lacZ_GTtoTAsub_F	CATTAAAGCGAGTGGCAACAGTTTTAGAGCTAGAAATAGC	For PCR amplification of the nsgRNA_lacZ_GTtoTAsub cassette.
pnsgRNA_lacZ_GTtoTAsub_R	TGTTGCCACTCGCTTTAATGactagtattatacctaggac	

pVRb_backbone_F	CTGTTGTTTGTCGGTGAACG	For PCR amplification of the pVRb backbone from pVRb_20_992.
pVRb_backbone_R	AAGGGCCTCGTGATACGCCT	
sgRNA_cassette_F	AGGCGTATCACGAGGCCCTTgaattctaaagatctttgac	For PCR amplification of the sgRNA cassette from pPEgRNA.
sgRNA_cassette_R	CGTTCACCGACAAACAACAGataaaacgaaaggcccagtc	
nsgRNA_seq	GCAATTCCGACGTCTAAG	For validation of the Gibson assembly of pnsgRNA. Also for validation of PEgRNA cloning of pVRb_PEgRNA.
PEgRNA_backbone_F	TTTTTTTGAAGCTTGGGCCC	For PCR amplification of the pPEgRNA backbone
PEgRNA_backbone_R	ACTAGTATTATACCTAGGAC	
PEgRNA_F	GTCCTAGGTATAATACTAGTGTTTTAGAGCTAGAAATAGCA	For removal of the 20 bp spacer from the plasmid pgRNA-bacteria.
PEgRNA_R	ACTAGTATTATACCTAGGACTGAGCTAGCT	
PEgRNA_Seq	AATAGGCGTATCACGAGGCA	For Sanger sequencing of correctly assembled PEgRNA plasmids.
pCDF-GFP_Seq	GAAATACTAGATGAGCAAAGGAGAAG	For screening of edited events using Sanger sequencing.
pCDF-1b_F	GTATATCTCCTTATTAAAGT	For PCR amplification of the pCDF-1b backbone.
pCDF-1b_R	ATTAACCTAGGCTGCTGCCA	
GFPplus_F	ACTTTAATAAGGAGATATACTTTACGGCTAGCTCAGTCCT	For PCR amplification of the J23106-GFPplus-T0 cassette.
GFPplus_R	TGGCAGCAGCCTAGGTTAATCGAACCGAACAGGCTTATGT	
GFPplus_check_F	ATGCGACTCCTGCATTAGGA	For screening of correctly assembled pCDF-GFPplus using PCR. The GFPplus_check_F is also
GFPplus_check_R	ACTAGTCGCCAGGGTTTTCC	

		used for Sanger sequencing validation.
A10D_F	TAGGCTTAGATATCGGCACAAA	For site-directed mutating of 10A to 10D with the dCas9 in the pdCas9-bacteria.
A10D_R	TTGAGTATTTCTTATCCATATG	
D10_Seq	GCGAGTTTACGGGTTGTTA	For screening of correct mutations of 10A to 10D using Sanger sequencing.
pCas9n_F	TAACTCGAGTAAGGATCTCC	For PCR amplification of the pCas9n(H840A) backbone, the stop codon of Cas9n is removed.
pCas9n_R	GTCACCTCCTAGCTGACTCA	
EcMMLV2_F	TGAGTCAGCTAGGAGGTGAC	For PCR amplification of the *E. coli* codon optimized linker-M-MLV2 fragment.
EcMMLV2_R	GGAGATCCTTACTCGAGTTA	
EcMMLV2_check_R	AACAAACAGCTCGAACGGCT	Together with EcMMLV2_F, this primer set is used for PCR screening of *E. coli* codon optimized linker-M-MLV2 fragment insertion.
EcMMLV2_seq	AACTGGATTGCCAACAGGGT	Together with EcMMLV2_F, these three primers are used to confirm the insertion of *E. coli* codon optimized linker-M-MLV2 fragment by Sanger sequencing.
dblTerm_R_seq	GAAGGTGAGCCAGTGTGACT	

Supplementary Table 3: Important sequences involved in this study.

Name	Sequence	Description
J23106-GFPplus-T0 cassette	TTTACGGCTAGCTCAGTCCTA GGTATAGTGCTAGCTACTAGA GAAAGAGGAGAAATACTAGAT GAGCAAAGGAGAAGAACTTTT CACTGGAGTTGTCCCAATTCTT GTTGAATTAGATGGTGATGTT AATGGGCACAAATTTTCTGTC AGTGGAGAGGGTGAAGGTGAT GCTACATACGGAAAACTCACC CTTAAATTTATTTGCACTACTG GAAAACTACCTGTTCCATGGC CAACACTTGTCACTACTCTGAC CTATGGTGTTCAATGCTTTTCC CGTTATCCGGATCACATGAAA CGGCATGACTTTTTCAAGAGT GCCATGCCCGAAGGTTATGTA CAGGAACGCACTATATCTTTC AAAGATGACGGGAACTACAAG ACGCGTGCTGAAGTCAAGTTT GAAGGTGATACCCTTGTTAAT CGTATCGAGTTAAAGGGTATT GATTTTAAAGAAGATGGAAAC ATTCTCGGACACAAACTAGAG TACAACTATAACTCACACAAT GTATACATCACGGCAGACAAA CAAAAGAATGGAATCAAAGCT AACTTCAAAATTCGCCACAAC ATTGAAGATGGTTCCGTTCAA CTAGCAGACCATTATCAACAA AATACTCCAATTGGCGATGGC CCTGTCCTTTTACCAGACAACC ATTACCTGTCGACACAATCTG CCCTTTCGAAAGATCCCAACG AAAAGCGTGACCACATGGTCC TTCTTGAGTTTGTAACTGCTGC TGGGATTACACATGGCATGGA TGAGCTCTACAAATGAAGCGC ATACCTGCAGGCATGCAAGCT TGCGGCCGCGTCGTGACTGGG AAAACCCTGGCGACTAGTCTT GGACTCCTGTTGATAGATCCA GTAATGACCTCAGAACTCCAT CTGGATTTGTTCAGAACGCTC GGTTGCCGCCGGGCGTTTTTT	A fast folding GFP variant GFP+ encoding gene (in blue), controlled by a constitutive promoter J23106 (in green) and ended with a T0 terminator (in orange).

	ATTGGTGAGAATCCAGGGGTC CCCAATAATTACGATTTAAATT TGACATAAGCCTGTTCGGTTC G	
E. coli codon optimized linker-M-MLV2 fragment	*TGAGTCAGCTAGGAGGTGAC*AG CGGCGGCAGCAGCGGCGGCAG CAGCGGCAGCGAAACCCCGGG CACCAGCGAAAGCGCGACCCC GGAAAGCAGCGGCGGCAGCAG CGGCGGTAGCAGCACCCTGAA CATCGAGGACGAGTATCGTCT GCATGAGACCAGCAAGGAGCC GGATGTTAGCCTGGGTAGCAC CTGGCTGAGCGACTTTCCGCA GGCGTGGGCGGAAACCGGCGG CATGGGTCTGGCGGTTCGCCA GGCGCCGCTGATCATTCCGCT GAAGGCGACCAGCACCCCGGT TAGCATCAAGCAGTATCCGAT GAGCCAGGAAGCGCGTCTGGG TATTAAGCCGCACATTCAACG TCTGCTGGACCAGGGTATTCT GGTGCCGTGCCAGAGCCCGTG GAATACCCCGCTGCTGCCGGT GAAGAAACCGGGTACCAATGA TTACCGTCCGGTGCAAGACCT GCGTGAGGTTAACAAGCGCGT TGAAGATATTCATCCGACCGT TCCGAACCCGTACAACCTGCT GAGCGGTCTGCCGCCGAGCCA CCAGTGGTATACCGTGCTGGA TCTGAAGGACGCGTTTTTCTG CCTGCGTCTGCACCCGACCAG CCAACCGCTGTTCGCGTTTGA ATGGCGTGACCCGGAAATGGG TATCAGCGGCCAACTGACCTG GACCCGTCTGCCGCAGGGCTT TAAAAACAGCCCGACCCTGTT CAACGAGGCGCTGCACCGTGA TCTGGCGGACTTCCGTATCCA ACACCCGGATCTGATCCTGCT GCAGTACGTGGACGATCTGCT GCTGGCGGCGACCAGCGAACT GGATTGCCAACAGGGTACCCG TGCGCTGCTGCAGACCCTGGG TAACCTGGGTTACCGTGCGAG	A 33-amino acid linker (in red) fused with the M-MLV2 encoding sequence (in black). 20 bp overhangs for Gibson assembly into pCas9n(H840A) is shown in gray italic.

	CGCGAAAAAGGCGCAAATTTG CCAGAAGCAAGTGAAGTATCT GGGCTACCTGCTGAAGGAAGG TCAACGCTGGCTGACCGAGGC GCGTAAGGAAACCGTTATGGG TCAGCCGACCCCGAAGACCCC GCGCCAACTGCGTGAGTTCCT GGGTAAAGCGGGTTTTTGCCG TCTGTTTATCCCGGGTTTCGCG GAAATGGCGGCGCCGCTGTAC CCGCTGACCAAACCGGGTACC CTGTTTAACTGGGGTCCGGAC CAGCAGAAAGCGTACCAAGAG ATCAAACAGGCGCTGCTGACC GCGCCGGCGCTGGGTCTGCCG GACCTGACCAAGCCGTTCGAG CTGTTTGTTGATGAAAAGCAG GGTTATGCGAAAGGCGTTCTG ACCCAGAAACTGGGTCCGTGG CGCCGTCCGGTTGCGTACCTG AGCAAGAAACTGGATCCGGTT GCGGCGGGCTGGCCGCCGTGC CTGCGTATGGTTGCGGCGATC GCGGTTCTGACCAAAGACGCG GGCAAGCTGACCATGGGTCAA CCGCTGGTGATTCTGGCGCCG CATGCGGTTGAAGCGCTGGTT AAGCAGCCGCCGGACCGTTGG CTGAGCAACGCGCGTATGACC CACTATCAAGCGCTGCTGCTG GATACCGACCGTGTTCAGTTC GGTCCGGTGGTTGCGCTGAAC CCGGCGACCCTGCTGCCGCTG CCGGAGGAAGGTCTGCAGCAT AACTGCCTGGACATTCTGGCG GAGGCGCACGGTACCCGTCCG GATCTGACCGACCAGCCGCTG CCGGACGCGGATCACACCTGG TATACCGACGGCAGCAGCCTG CTGCAAGAAGGCCAGCGTAAG GCGGGTGCGGCGGTTACCACC GAGACCGAAGTTATCTGGGCG AAAGCGCTGCCGGCGGGTACC AGCGCGCAGCGTGCGGAGCTG ATTGCGCTGACCCAAGCGCTG AAAATGGCGGAGGGCAAAAAG CTGAATGTTTATACCGATAGC	

	CGTTACGCGTTTGCGACCGCG CACATCCATGGTGAAATCTAC CGTCGTCGTGGTTGGCTGACC AGCGAAGGCAAAGAAATCAAA AATAAGGACGAGATTCTGGCG CTGCTGAAAGCGCTGTTCCTG CCGAAACGTCTGAGCATCATT CACTGCCCGGGTCACCAGAAA GGTCACAGCGCGGAGGCGCGT GGTAATCGCATGGCGGATCAA GCGGCGCGTAAAGCGGCGATT ACCGAAACCCCGGATACCAGC ACCCTGCTGATTGAAAATAGC AGCCCGTAA*TAACTCGAGTAA* *GGATCTCC*	
An example (TAAin) of functional PEgRNA cassette	ttgacagctagctcagtcctaggtataatac tagtCTTGTCACTACTCTGACCT AGTTTTAGAGCTAGAAATAGC AAGTTAAAATAAGGCTAGTCC GTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGCTTGAACA CC*TTA*ATAGGTCAGAGTAGTG Atttttt	Green: J23119 promoter Red: 20-nt spacer targeting GFPplus coding sequence Black: 76-nt gRNA scaffold Blue: RTT of the 3 prime extension, the target TAA is shown in bold, italic Orange: PBS of the 3 prime extension

Supplementary Table 4: Spacers and 3 prime extensions used in this study.

PEgRNA	Space (5'-3')	3 prime extension (5'-3')	RTT length (nt)	PBS length (nt)
PEgRNA_GFP_TAAin_T4	CTTGTCACTACTCTGACCTA	TTAATAGGTCAGAGTAGTGA	7	13
PEgRNA_GFP_TAAin_T5	CTTGTCACTACTCTGACCTA	CTTAATAGGTCAGAGTAGTGA	8	13
PEgRNA_GFP_TAAin_T8	CTTGTCACTACTCTGACCTA	CACCTTAATAGGTCAGAGTAGTGA	11	13
PEgRNA_GFP_TAAin	CTTGTCACTACTCTGACCTA	TTGAACACCTTAATAGGTCAGAGTAGTGA	16	13
PEgRNA_GFP_TAAin_T20	CTTGTCACTACTCTGACCTA	AAAAGCATTGAACACCTTAATAGGTCAGAGTAGTGA	23	13
PEgRNA_GFP_TAAin_PBS5	CTTGTCACTACTCTGACCTA	TTGAACACCTTAATAGGTCAG	16	5
PEgRNA_GFP_TAAin_PBS8	CTTGTCACTACTCTGACCTA	TTGAACACCTTAATAGGTCAGAGT	16	8
PEgRNA_GFP_TAAin_PBS17	CTTGTCACTACTCTGACCTA	TTGAACACCTTAATAGGTCAGAGTAGTGACAAG	16	17
PEgRNA_GFP_T198A	CTTGTCACTACTCTGACCTA	TTGAACACCTTAGGTCAGAGTAGTGA	13	13
PEgRNA_GFP_T196C_T198C	CTTGTCACTACTCTGACCTA	TTGAACACCGTGGGTCAGAGTAGTGA	13	13
PEgRNA_GFP_198Tdel	CTTGTCACTACTCTGACCTA	ATTGAACACCTAGGTCAGAGTAGTGA	13	13
PEgRNA_GFP_combo	CTTGTCACTACTCTGACCTA	TTGAACACTTATTAGGTCAGAGTAGTGA	15	13

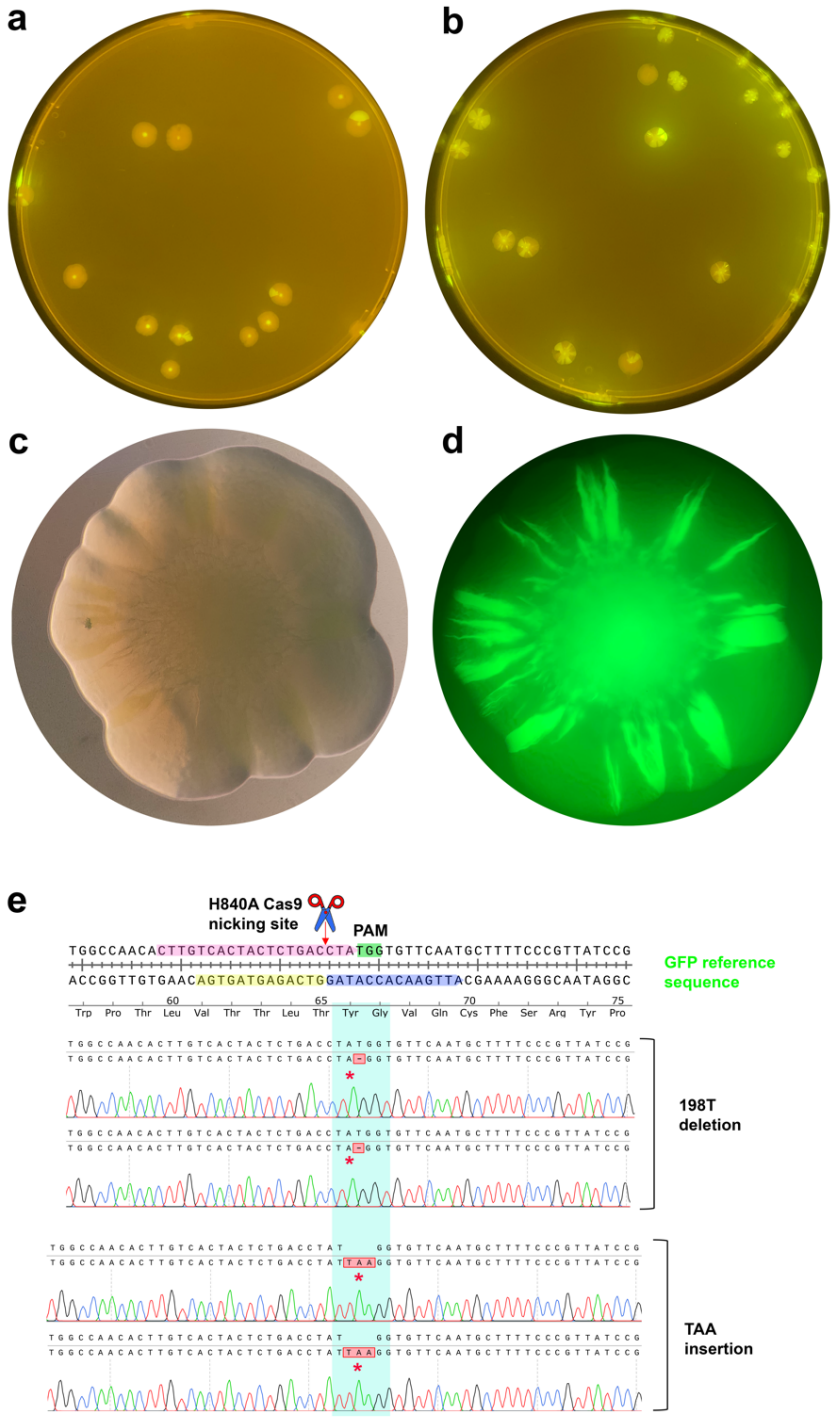
PEgRNA_GFP_12in	CTTGTCACTACTCTGACCTA	TTGAACACCTGTACAGATTACATAGGTCAGAGTAGTGA	25	13
PEgRNA_GFP_18in	CTTGTCACTACTCTGACCTA	TTGAACACCCTATAGTGAGTCGTATTAATAGGTCAGAGTAGTGA	31	13
PEgRNA_GFP_33in	CTTGTCACTACTCTGACCTA	TTGAACACCCTCGCCCGCTCGGTTTCGTGGGCTGATTCTACAATAGGTCAGAGTAGTGA	46	13
PEgRNA_GFP_10del	CTTGTCACTACTCTGACCTA	GGGAAAAGCATTGGTCAGAGTAGTGA	13	13
PEgRNA_GFP_23del	CTTGTCACTACTCTGACCTA	TGATCCGGATAACGTCAGAGTAGTGA	13	13
PEgRNA_GFP_36del	CTTGTCACTACTCTGACCTA	ATGCCGTTTCATGGTCAGAGTAGTGA	13	13
PEgRNA_GFP_49del	CTTGTCACTACTCTGACCTA	TCTTGAAAAAGTCGTCAGAGTAGTGA	13	13
PEgRNA_GFP_97del	CTTGTCACTACTCTGACCTA	CTTTGAAAGATATGTCAGAGTAGTGA	13	13
PEgRNA_lacZ_TAGin	AATCCCGAATCTCTATCGTG	TCAACCACCTACGCACGATAGAGATTCGG	16	13
PEgRNA_lacZ_CGdel	TATGCAGCAACGAGACGTCA	ATGAGCGGCATTTTCTGACGTCTCGTTGCTGC	18	14
PEgRNA_lacZ_GTtoTAsub	GCGAGTTGCGTGACTACCTA	ACTGTTACCCGTTAGTAGTCACGCAACT	14	14
PEgRNA_galK_TAAin	GACAGCCACACCTTTGGGCA	GCAGTTTCCTAAATGCCCAAAGGTGTGGC	16	13
PEgRNA_GFP_312Gdel	CTATATCTTTCAAAGATGAC	TTGTAGTTCCGTCATCTTTGAAAGAT	13	13
nsgRNA	PE3 or PE3b[1,10]	Space (5'-3')		

nsgRNA_GFP_TAAin	PE3b	AAAGCATTGAACACCttaAT		
nsgRNA_lacZ_TAGin	PE3b	CctaCGCACGATAGAGATTC		
nsgRNA_lacZ_CGdel	PE3	CTGGAGTGACGGCAGTTATC		
nsgRNA_lacZ_GTtoTAsub	PE3	CATTAAAGCGAGTGGCAACA		

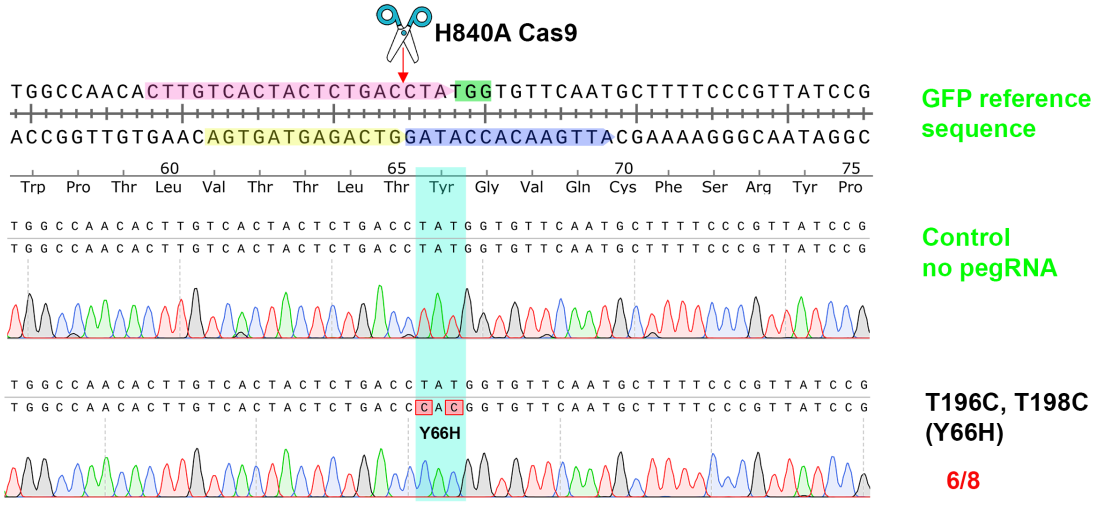
Supplementary Table 5: Whole-genome sequencing-based analysis of escapers from the CRISPR-Prime Editing

Strain	**Recorded mutations**[a]			
	On-plasmids			**On-chromosome**
	pCDF-GFPplus	pPEgRNA_GFP_198Tdel	pCRISPR-PE-bacteria	
PE0035	0	0	0	0
PE0036	0	2 (Pos. 126. -26 bp, the 3 prime extension was missing)	0	0
PE0037	0	2 (Pos. 126. -26 bp, the 3 prime extension was missing)	0	0
PE0038	0	0	0	0
PE0039	0	0	0	0
PE0040	0	0	0	0
PE0041	0	0	0	0
PE0042	0	0	0	0
PE0043	0	1 (Pos. 126. -26 bp, the 3 prime extension was missing)	0	0
PE0044	0	0	0	0
PE0045	0	-	-	0
PE0046	-	-	0	0
PE0047	0	-	0	0
PE0048	0	1 (Pos. 126. -26 bp, no 3 prime extension was cloned)	0	0

[a]: the shared mutations were listed here.
One shared mutation in pCRISPR-PE-bacteria is Pos. 8,584. A43A (GCC→GCA);
One shared mutation in pPEgRNA_GFP_198Tdel is Pos. 1,146. G→A, intergenic, 6XHis;
One shared mutation in the chromosome is Pos. 4,272,972. +T, intergenic.

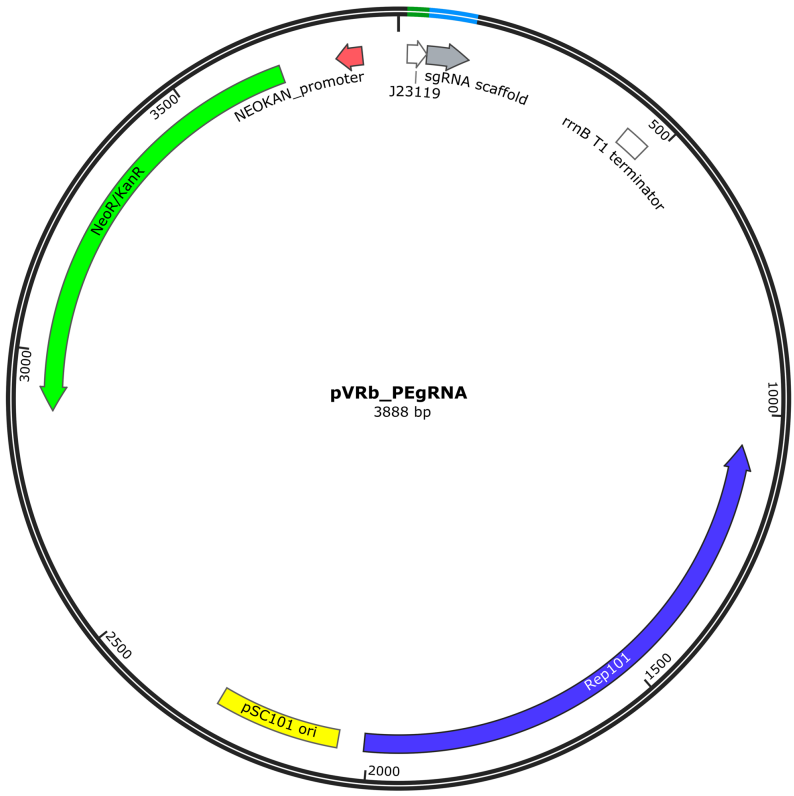


Supplementary Figure 1: Editing accumulates over time of using CRISPR-Prime Editing system in *E. coli*. a. A five-day old induction (200 ng/mL ATc) plate of GFP 1-bp deletion under a Blue-Light Transilluminator (Safe Imager 2.0, Thermo Fisher Scientific, US). **b.** A five-day old induction (200 ng/mL ATc) plate of GFP 3-bp insertion under a Blue-Light Transilluminator (Safe Imager 2.0, Thermo Fisher Scientific, US). **c.** The 40x optical view of a single colony from **b.** under a Leica DM4000 B Fluorescence Microscope (Leica Microsystems, Germany). **d.** The 40x GFP fluorescent view of the same colony as c. using the same microscope. **e.** Two of each outgrown colony from a. and b. were Sanger-sequenced. The in-figure legend is the same as Fig. 2a.



Supplementary Figure 2: A DNA editing of double substitutions by CRISPR-Prime Editing. Eight randomly picked colonies of each designed DNA engineering were Sanger sequenced and traces were aligned to the targeted locus of GFP coding sequence. The correctly edited colony numbers and the total sequenced numbers were shown in red. The in-figure legend is the same as Fig. 2a.

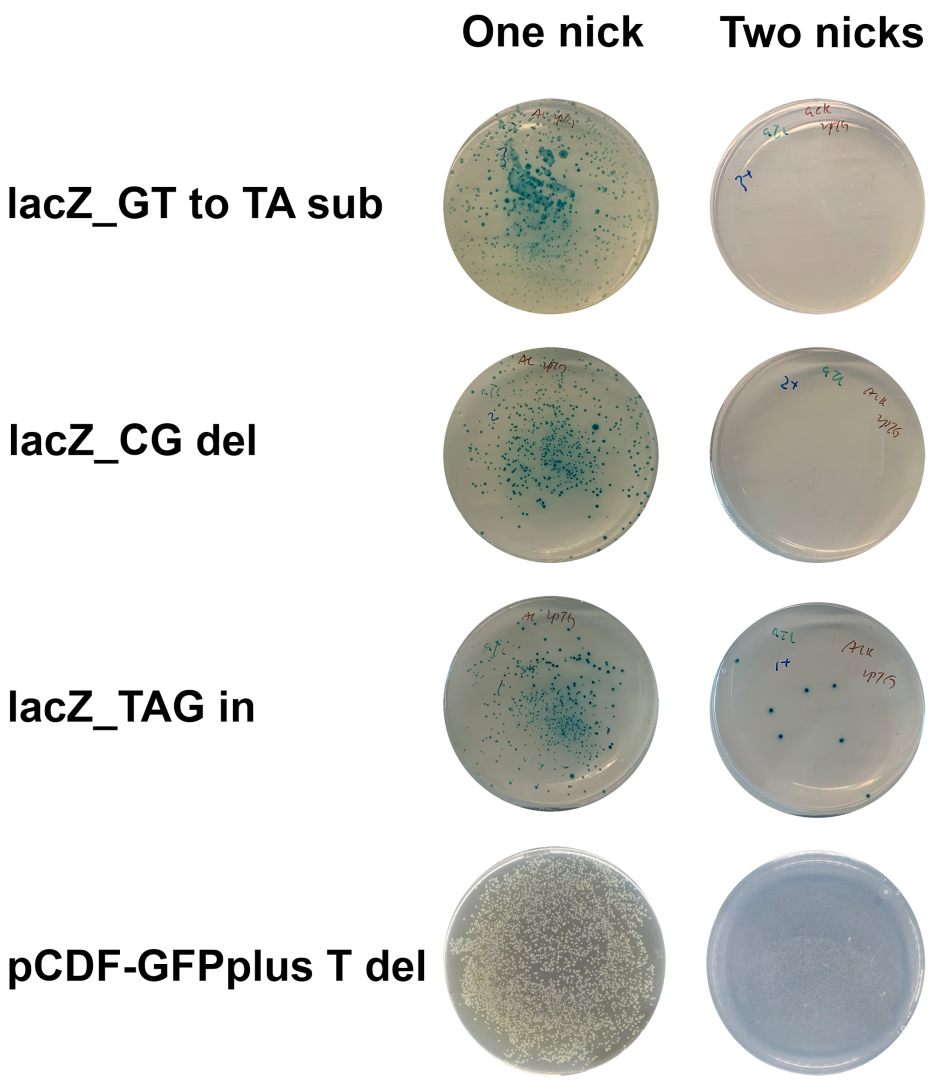
a



b

ttgacagctagctcagtcctaggtataatactagtGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
aactgtcgatcgagtcaggatccatattatgatcaCAAAATCTCGATCTTTATCGTTCAATTTTATTCCGATCAGGCAATAGTTGAACTTTTTCACCGTGGCTCAGCCACG
J23119
sgRNA scaffold

Supplementary Figure 3: The plasmid map of pVRb_PEgRNA. a. The plasmid map of the pVRb_PEgRNA plasmid, which is pSC101 ori, kanamycin resistant, and with a J23119 driven sgRNA scaffold. **b**. The DNA sequence of the J23119-sgRNA scaffold is displayed. The J23119 promoter sequence is in green, and the sgRNA scaffold is in blue.



Supplementary Figure 4: A second nick compromises the application of CRISPR-Prime Editing in *E. coli*. Plates showed the colony formation of transformants. For the one nick panel, 50 µl of transformation culture was plated onto appropriate antibiotics supplemented LB plates, while for the two nicks panel, 400 µl of transformation culture was plated. Photos were taken by a Doc-It imaging station after 24h incubation at 37 °C.

Supplementary Figure 5: Long DNA fragments deletion and insertions using CRISPR-Prime Editing in *E. coli*. As examples, Sanger sequencing traces of a 23-bp deletion, a 49-bp deletion, and a 18-bp (mini-T7 promoter) insertion traces were aligned to the targeted locus of GFP coding sequence. The in-figure legend is the same as Fig. 2a.

SUPPLEMENTARY REFERENCES:

1 Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149-157, doi:10.1038/s41586-019-1711-4 (2019).

2 Lin, Q. *et al.* Prime genome editing in rice and wheat. *Nat. Biotechnol.* **38**, 582-585, doi:10.1038/s41587-020-0455-x (2020).

3 Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833-838, doi:10.1038/nbt.2675 (2013).

4 Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380-1389, doi:10.1016/j.cell.2013.08.021 (2013).

5 Bowater, R. & Doherty, A. J. Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS Genet.* **2**, e8, doi:10.1371/journal.pgen.0020008 (2006).

6 Shuman, S. & Glickman, M. S. Bacterial DNA repair by non-homologous end joining. *Nat. Rev. Microbiol.* **5**, 852-861, doi:10.1038/nrmicro1768 (2007).

7 Tong, Y., Charusanti, P., Zhang, L., Weber, T. & Lee, S. Y. CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. *ACS Synth. Biol.* **4**, 1020-1029, doi:10.1021/acssynbio.5b00038 (2015).

8 Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183, doi:10.1016/j.cell.2013.02.022 (2013).

9 Rhodius, V. A. *et al.* Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters. *Mol. Syst. Biol.* **9**, 702, doi:10.1038/msb.2013.58 (2013).

10 Hsu, J. Y. *et al.* PrimeDesign software for rapid and simplified design of prime editing guide RNAs. doi:10.1101/2020.05.04.077750.

Chapter 7

Discussion & Outlook

Christopher M. Whitford[1]

[1] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby 2800, Denmark.

DISCUSSION

When I started working with streptomycetes, I primarily viewed these remarkable bacteria as the famous producers of antibiotics. Only later did I fully come to realize that viewing streptomycetes only as prolific producers of antibiotics falls incredibly short of acknowledging their true potential. A few years and a Ph.D. experience later, I now consider streptomycetes simply Nature's greatest chemists, and it is up to us to make the most of that potential, which reaches far beyond antibiotics.

It has also been a realization during my Ph.D. just how extensively the world around us, and our very human existence, has and continues to be, shaped by specialized metabolites. For thousands of years, specialized metabolites have improved our human experience, for example in the form of plant extracts. However, similar to fermentation, the underlying principles and "active ingredients" remained unknown for the majority of this time, and we are just now starting to fully grasp all the different ways in which specialized metabolites shape environments, microbiomes, and our very selves. In fact, we are just starting to be able to understand Nature's chemical language.

The modern era of specialized metabolites began in the early 20th century with the discovery of penicillin. The discovery of streptomycin in 1943, combined with the deployed methodology, would subsequently lay the groundwork for the "Golden Era" of antibiotic discovery. This period, from approximately 1950 to 1970, was a defining era for modern medicine and initiated an entire research field. The Waksman approach to specialized metabolite discovery is still widely used today and defined decades of research on *Streptomyces.* Certainly, these historical roots are why the *Streptomyces* research field has been so heavily biased towards the discovery of antibiotics, and why much less attention has so far been paid to other potential application areas.

Nonetheless, the emerging antimicrobial resistance crisis has resulted in renewed interest in antibiotic discovery from *Streptomyces.* However, given that traditional approaches have yielded very few antibiotics with novel scaffolds within the last few decades, it might be worth questioning, if the same decades-old methodologies will allow us to discover and utilize novel antibiotics within the required timeframe. Additionally, given the millions of

screened isolates over the last 70 years [1], it is a valid question to ask whether we have discovered the majority of the "low-hanging fruit" already. Novel approaches and methods, enabled by recent technological advances, are likely needed to discover the "higher hanging fruit", and to also utilize the biosynthetic potential of specialized metabolites for a much wider range of applications.

The speed and throughput of modern sequencing and analysis pipelines has resulted in more than a million computationally predicted biosynthetic gene clusters (BGCs) [2]. While sequencing and prediction of BGCs is a fast and streamlined process by now, in contrast, establishing just one BGC-to-molecule connection in the laboratory can require many months of tedious work. Even for a rather small strain collection such as the one in the Natural Products Genome Mining group, which consists of approximately 2000 strains encoding around 60,000 BGCs, connecting all BGCs to specialized metabolites would likely require hundreds of person-years using traditional methods. Obviously, this discrepancy is leading to an exponentially widening gap between computationally predicted and experimentally validated BGCs. While computational predictions can increasingly be used to direct experimental work and to ensure time, resources, and money are spent on the most promising BGCs, the promise of the entire field ultimately relies on being able to deliver molecules in vials. Therefore, our inability to experimentally access most of the biosynthetic potential encoded by streptomycetes remains the largest bottleneck of the field.

In this context, this Ph.D. thesis aimed to develop advanced CRISPR tools for simplified and streamlined engineering of *Streptomyces* species. CRISPR has proven to be a fantastic platform to build advanced genome engineering tools on and, as such, is a great starting point for expanding the genome engineering toolbox for *Streptomyces* species.

However, these better tools alone will not suffice to overcome the major experimental bottlenecks. It will require tight integration of computational, experimental, and analytical methods to achieve significant leaps in the exploration and ultimately utilization of the full biosynthetic potential of streptomycetes.

One widely used framework for this is the Design-Build-Test-Learn cycle, which was described in **Chapter 2**. Using the DBTL cycle, a tighter integration of the various stages of the experimental workflow can be achieved. Each stage equally depends on the previous one and informs the next stage. Therefore, compatibility of the various workflows and methods is of crucial importance. In traditional experimental workflows, too often learnings from previous experimental iterations are not properly used as a starting point for subsequent iterations. Using the DBTL cycle, this problem can be overcome and can result in significant improvements in terms of invested time, resources, and iterations needed to achieve a desired experimental outcome. By feeding machine learning algorithms with structured experimental data, designs for the next cycle iterations can be suggested, substantially cutting down overall development times [3]. By providing an overview of important resources for each cycle stage, we were aiming to enable research groups to move closer to well-defined and structured DBTL-centered workflows.

However, properly establishing the DBTL cycle for streptomycetes remains challenging. The major lack of standardized methods, parts, and plasmids, and the complicated biological features such as linear high GC chromosomes and filamentous growth, make it difficult to standardize experimental workflows and to quickly implement the required changes for the next cycle iteration. Furthermore, working with strain collections instead of single species means the rather static DTBL cycle might not allow for the needed flexibility in terms of experimental and analytical methods.

As a step towards standardized methods and to help other researchers establish CRISPR-based engineering of streptomycetes, we published a comprehensive protocol for using these tools (**Chapter 3**). The protocol covers all already established CRISPR tools for streptomycetes, including CRISPR-Cas9, CRISPR-dCas9, and CRISPR-BEST. By sharing these protocols, other groups can learn from our extensive experience in CRISPR-based engineering of *Streptomyces* species.

Even with streamlined workflows employing current CRISPR tools, engineering cycle times are painfully long and usually low throughput. There is an urgent need for tools that allow simultaneous engineering of multiple genomic loci to enable high throughput strain engineering.

Therefore, we systematically analyzed the scalability of multiplexed cytosine base editing, the results of which are described in **Chapter 4**. Multiplexed base editing is a very promising technology paving the way for high throughput strain prototyping. This allows strain engineering experiments to fail fast, resulting in more rapid iterations of the DBTL cycle, and to ultimately cut down strain engineering efforts significantly. We analyzed editing outcomes for sgRNA arrays consisting of 9, 18, and finally 28 sgRNAs. By Illumina sequencing all mutants, we were able to obtain systematic data about the editing performance of sgRNA arrays at various lengths, off-target frequencies, and genome-wide consequences of such large-scale base editing experiments. We identified key bottlenecks, namely off-target frequencies, array instabilities, and editing efficiencies. All these bottlenecks can likely be removed or reduced in the near future, paving the way for a truly robust technology for high throughput prototyping in non-model organisms.

Given that genes are not deleted but only inactivated through the introduction of premature stop codons, there are legitimate concerns that the introduced mutations are not stable and will just revert given enough time and evolutionary pressure. However, such cases were only rarely observed so far (unpublished data). Still, this concern underlines how multiplexed base editing might be primarily of interest for prototyping and identifying promising gene knockout targets for strain engineering.

Therefore, other tools are required to construct stable strains. So far, Cas9 based tools were most widely employed for these purposes. However, Cas9-based toxicity, as well as low efficiency in some strains have hindered widespread and streamlined application [4–6]. To overcome these limitations, we developed a new tool called pCRISPR-Cas3 based on a minimal type I-C CRISPR system (**Chapter 5).** Using this tool, we achieved superior deletion efficiencies compared to Cas9 and demonstrated streamlined genome engineering in multiple *Streptomyces* species. Construction of a new expression host was possible within a few months using pCRISPR-Cas3, demonstrating how highly efficient tools can streamline strain engineering projects and cut down development times significantly. Application of pCRISPR-Cas3 in other undisclosed non-model strains also resulted in very encouraging results, highlighting how type I CRISPR systems might generally be better suited for genome engineering of streptomycetes.

Even though the initial system proved to be very useful, further improvements could be implemented. For example, while the developed method for protospacer cloning works reliably, it is difficult to use this method for library-based approaches. Given the processivity of Cas3, pCRISPR-Cas3 is a very promising tool for random-sized deletion library generation. In this context, running genome-scale random-sized deletion experiments would be very interesting. Therefore, improved, library cloning compatible methods for integration of protospacers between the two repeats would allow expanding the potential applications of pCRISPR-Cas3. Furthermore, the function of Cas5, which is the same as that of Csy4 from **Chapter 4,** allows multiplexing of protospacers without further modifications of the plasmid system. Given the recombinogenic nature of the overhangs created by Cas3, Cas5 might enable efficient multiplexed genome engineering, even of unstable *Streptomyces* genomes. Nonetheless, the established pCRISPR-Cas3 system already adds important functionalities to the *Streptomyces* CRISPR toolbox and will likely become a new platform on which new methods can be built.

In **Chapter 6**, we reported the development of the first CRISPR-Prime tool for genetic engineering of *E. coli*. CRISPR-Prime is a very promising technology allowing the modification, deletion, and insertion of small sequences without the need to introduce double-stranded breaks. This is highly interesting for many bacterial species, but especially for species which cannot tolerate double strand breaks easily, such as *E. coli* and *Streptomyces species.* The potential applications of CRISPR-Prime editing are manifold, but among the most promising ones are promoter knock-ins, the introduction of complex mutations (e.g. in active sites), or the addition of integration sites. Furthermore, several tools have recently been developed based on CRISPR-Prime editing (such as twin prime editing or PASTE), which allow the targeted integration of large specified sequences [7,8]. The development of a CRISPR-Prime tool for *E. coli* served as a first stepping stone towards adapting the system for *Streptomyces*. However, the system for *Streptomyces* still needs significant optimizations before it can be used as a routine tool. Likely, our poor understanding of the DNA repair machinery in *Streptomyces,* as well as of the factors required to achieve successful prime editing, contribute to the observed difficulties. Coexpression of beneficial or auxiliary cellular components identified in mammalian cell lines might aid prime editing in *Streptomyces.* Furthermore, incorrect folding or assembly of the prime fusion

complex could also explain why prime editing has so far not been successful in *Streptomyces*. Splitting the fusion complex into individually expressed components might result in valuable insights [9]. Adapting CRISPR-Prime editing for *Streptomyces* will likely open up an entirely new suite of applications and methods, greatly advancing our genome engineering capabilities.

Taken together, this thesis expanded the CRISPR toolbox for *Streptomyces*, provided important resources to guide CRISPR-based metabolic engineering of streptomycetes, and laid the groundwork for next-generation CRISPR-based engineering of streptomycetes.

OUTLOOK

We are amid a planetary crisis due to accelerating climate change and biodiversity loss. Most human activities by now exceed the planetary boundaries, putting incredible pressure on the few remaining intact ecosystems. This will have far-reaching consequences for almost every aspect of modern life [10]. With both accelerating climate change and a growing population, the scale and significance of these crises will only increase in the decades to come.

For example, habitat loss and climate change increase the chances of more frequent pandemics [11]. Rising temperatures have been linked to an increased spread of deadly fungal infections [12]. Food insecurity will increase due to rising temperatures, extreme weather events, and biodiversity loss, especially as current agricultural methods put enormous pressure on already strained ecosystems. At the same time, antimicrobial resistance is spreading, due to overuse in animal husbandry, inadequate stewardship, and a dried-up R&D pipeline for novel antibiotics. Without intervention, antimicrobial resistance has been predicted to cause just as many deaths as cancer by the year 2050 [13]. Cancer on the other hand will increase in prevalence given an increasingly senior population [14].

For many of these problems, the most effective short-term solutions involve adequate political actions, but long term, many new scientific and technological solutions are needed to ensure decent living conditions for the generations to come.

In this context, the soil beneath our feet might hold answers to many of these pressing issues. From new antibiotics and antifungals, biologicals for agriculture, novel anticancer drugs, antivirals against future pandemics, to molecules for more sustainable chemical production, all of these might be already present within Nature's biosynthetic treasure chest.

To overcome the experimental bottlenecks and to ensure the biosynthetic potential of streptomycetes can contribute urgently needed solutions, much more systematic approaches are needed, based on stronger standardization of

workflows and tools, increased reproducibility, as well as stronger interoperability.

Experimentally, I see reproducibility, standardization, and modularity, scale (automatization & high-content experiments), and expanded screening platforms as the most pressing issues ahead.

Streptomyces are incredibly sensitive to changing parameters. Plenty of anecdotes circulate telling stories of the immense difficulties of getting strains to reliably produce specific compounds, to grow reliably in certain morphologies, or to conjugate strains with robust efficiencies. Many of these issues are at their core an issue of **lacking reproducibility**, in many cases due to many different experimental methods and schools of thought, high variations in experimental parameters, and experimental setups with generally little parameter control. This makes it both very difficult to reproduce results, and to draw the right conclusions from experiments.

Modular vector systems and standards, which the community can keep expanding, are urgently needed for *Streptomyces.* This is also part of a general lack of proper synthetic biology methods and standards for *Streptomyces*. While one modular plasmid system has previously been described [15], plasmid systems such as pSEVA[16] have yet to be developed for streptomycetes. The lack of such a community-driven plasmid-base is severely limiting the number of strains that can be efficiently engineered. We cannot expect to efficiently engineer entire strain collections based on single plasmid systems.

Automating workflows remains challenging, primarily due to the filamentous growth of most streptomycetes. While important work is ongoing that focusses on expanding the usability of automated systems for *Streptomyces* research, a long road ahead remains. Tasks such as colony picking will require extensive adaptation of current automated workflows, if not entirely new solutions.

Therefore, **library-based and high-content experiments** are very promising approaches slowly being adapted for *Streptomyces* research. These kinds of experiments usually do not require extensive automation to generate vast amounts of data and can easily be coupled to specific selection criteria to reduce experimental complexity. Once more widely applied, such

experimental approaches will rapidly expand our knowledge of *Streptomyces* genetics and of the underlying factors governing the production of specialized metabolites.

To deal with the incredible number of cryptic BGCs, emphasis should be put on experimental methods that can be **scaled robustly**. Unfortunately, this is not currently the case for most of the methods used to experimentally characterize silent or cryptic BGCs. Ideally, such methods do not require extensive *a priori* knowledge, and the information obtained from whole genome sequencing should suffice to inform expression experiments.

Furthermore, while millions of isolates were screened within the last 70 years, the majority of these were screened for antibiotic activity. However, the structural complexity of specialized metabolites confers a much wider array of bioactivities. Widening screening approaches to not just include antibiotic activity tests, but also assays for potential anticancer-, plant growth-promoting-, or antifungal-activity represents a rather straightforward approach to increase the utility and hit rate of specialized metabolite discovery research. By employing **multidimensional screening platforms**, the hit rate for novel lead compounds can likely be increased dramatically.

These are very exciting times, when computational, experimental, and analytical methods are all slowly coming together in a truly synergistic way. Computational platforms are increasingly allowing us to ask the right questions, better molecular biology tools allow us to engineer streptomycetes more efficiently, and constantly improving analytical chemistry methods allow us to decipher the complex metabolomes of these strains.

We may soon be able to systematically study and explore the vast dark biosynthetic matter of evolutionarily optimized specialized metabolites. This will rapidly expand our access to urgently needed solutions, while also expanding our understanding of the complex hidden world that is all around us.

A world that is worth preserving.

REFERENCES

1. Katz, L. & Baltz, R. H. Natural product discovery: past, present, and future. *Journal of Industrial Microbiology and Biotechnology* **43**, 155–176 (2016).
2. Kautsar, S. A., Blin, K., Shaw, S., Weber, T. & Medema, M. H. BiG-FAM: the biosynthetic gene cluster families database. *Nucleic Acids Research* **49**, D490–D497 (2021).
3. Radivojević, T., Costello, Z., Workman, K. & Garcia Martin, H. A machine learning Automated Recommendation Tool for synthetic biology. *Nat Commun* **11**, 4879 (2020).
4. Cho, S. *et al.* High-Level dCas9 Expression Induces Abnormal Cell Morphology in *Escherichia coli*. *ACS Synth. Biol.* **7**, 1085–1094 (2018).
5. Rostain, W. *et al.* Cas9 off-target binding to the promoter of bacterial genes leads to silencing and toxicity. *Nucleic Acids Research* **51**, 3485–3496 (2023).
6. Ye, S., Enghiad, B., Zhao, H. & Takano, E. Fine-tuning the regulation of Cas9 expression levels for efficient CRISPR-Cas9 mediated recombination in *Streptomyces*. *Journal of Industrial Microbiology and Biotechnology* **47**, 413–423 (2020).
7. Yarnall, M. T. N. *et al.* Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. *Nat Biotechnol* **41**, 500–512 (2023).
8. Anzalone, A. V. *et al.* Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. *Nat Biotechnol* **40**, 731–740 (2022).
9. Grünewald, J. *et al.* Engineered CRISPR prime editors with compact, untethered reverse transcriptases. *Nat Biotechnol* **41**, 337–343 (2023).
10. H.-O. Pörtner *et al. IPCC, 2022: Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. (Cambridge University Press.).
11. Carlson, C. J. *et al.* Climate change increases cross-species viral transmission risk. *Nature* **607**, 555–562 (2022).
12. Casadevall, A., Kontoyiannis, D. P. & Robert, V. On the Emergence of Candida auris: Climate Change, Azoles, Swamps, and Birds. **10**, (2019).

13. O'Neill, J. Tackling Drug-Resistant Infections Globally: Final Report and Recommendations the Review on Antimicrobial Resistance. https://amr-review.org/sites/default/files/160525_Final%20paper_with%20cover.pdf (2016).
14. Maddams, J., Utley, M. & Møller, H. Projections of cancer prevalence in the United Kingdom, 2010–2040. *Br J Cancer* **107**, 1195–1202 (2012).
15. Aubry, C., Pernodet, J.-L. & Lautru, S. Modular and Integrative Vectors for Synthetic Biology Applications in *Streptomyces* spp. *Appl Environ Microbiol* **85**, e00485-19 (2019).
16. Silva-Rocha, R. *et al.* The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Research* **41**, D666–D675 (2013).